Bài tập về nhà số 2
Nền tảng toán học của các mô hình tạo sinh – PiMA

Chủ đề: Phương pháp suy diễn

1 Bài tập lý thuyết: Phương pháp Monte Carlo

Các biến ngẫu nhiên được hiểu là một ánh xạ từ không gian mẫu vào tập các số thực $\mathbb{R}$.

Khi nói về sự hội tụ của các biến ngẫu nhiên, ta luôn hiểu rằng kết quả hội tụ luôn là một biến ngẫu nhiên. Tuy nhiên có thể giản lược về mặt ký hiệu. Ví dụ,

- các biến ngẫu nhiên $X_1, X_2, \dots$ hội tụ về a (một giá trị thực) khi $N \to \infty$, viết là $\lim_{N\to\infty} X_N = a$ hay $X_N \to a$ khi $N \to \infty$. Ta hiểu rằng $X_1, X_2, \dots$ hội tụ về một biến ngẫu nhiên $X\colon \Omega \to \mathbb{R}$, trong đó X là ánh xạ $\Omega \ni \omega \mapsto a$;
- các biến ngẫu nhiên $X_1, X_2, \dots$ hội tụ về $\mathcal{D}$ (một phân phối xác suất) khi $N \to \infty$, viết là $\lim_{N\to\infty} X_N = \mathcal{D}$ hay $X_N \to \mathcal{D}$ khi $N \to \infty$. Ta hiểu rằng $X_1, X_2, \dots$ hội tụ về một biến ngẫu nhiên X, trong đó X tuân theo phân phối $\mathcal{D}$.

Một câu hỏi nhỏ: Một dãy các biến ngẫu nhiên $X_1, X_2, \dots$ có thể cùng lúc hội tụ về một giá trị thực a và một phân phối $\mathcal{D}$ được hay không?

Ta xem xét hai định lý dưới đây phục vụ cho phần bài tập lý thuyết.

Định lý 1 (Luật số lớn (mạnh)) *Xét $(X_k)_{k\geq 1}$ là dãy các biến ngẫu nhiên độc lập và có cùng phân phối**. *Khi đó trung bình cộng của các biến ngẫu nhiên X_k hội tụ hầu chắc chắn*† *tới $\mathbb{E}[X_1]$ khi $N \to \infty$,*

$$\frac{X_1 + X_2 + \ldots + X_N}{N} \xrightarrow{a.s.} \mathbb{E}[X_1],$$

khi và chỉ khi $\mathbb{E}[|X_1|] < \infty$.

Định lý 2 (Định lý giới hạn trung tâm) *Xét $(X_k)_{k\geq 1}$ là dãy các biến ngẫu nhiên độc lập và có cùng phân phối. Giả sử $\mathbb{E}[X_1] < \infty$ và $\mathrm{Var}(X_1) \in (0, \infty)$. Khi đó dãy các biến ngẫu nhiên $\left(\frac{X_1+X_2+\ldots+X_N-N\mu}{\sigma\sqrt{N}}\right)_{N\geq 1}$ hội tụ yếu*‡ *tới phân phối chuẩn tắc khi $N \to \infty$. Nghĩa là*

$$\frac{X_1 + X_2 + \ldots + X_N - N \cdot \mathbb{E}[X_1]}{\sqrt{N \cdot \mathrm{Var}(X_1)}} \xrightarrow{w} \mathcal{N}(0, 1).$$

*Independent and identically distributed random variables
†Almost sure convergence
‡Weak convergence. Đôi khi được gọi là hội tụ theo phân phối (convergence in distribution)

Mục tiêu của một lớp các bài toán trong Thống kê là tính được các đặc trưng của các biến ngẫu nhiên. Ta đã quen với việc tính toán các đặc trưng này khi có thông tin về hàm phân phối. Tuy nhiên thực tế không phải lúc nào phân phối tường minh của một biến ngẫu nhiên cũng được biết trước. Dù vậy nhưng trong nhiều trường hợp, có thể sinh ra các mẫu từ phân phối chưa biết đó. Cụ thể, trong phần này, ta sẽ tìm hiểu cơ sở lý thuyết của việc tính kỳ vọng của một biến ngẫu nhiên (bản chất là xấp xỉ một tích phân) tuân theo một phân phối chưa biết (nhưng lấy mẫu được) bằng phương pháp Monte Carlo.

Xét biến ngẫu nhiên $Y \sim \mathcal{P}$ với $\mathcal{P}$ chưa biết (nhưng có thể lấy mẫu).

Bài 1. Xét $(Y_k\colon \Omega \to \mathbb{R})_{k\geq 1}$ là dãy các biến ngẫu nhiên độc lập và tuân theo phân phối $\mathcal{P}$. Giả sử $\mathbb{E}[|Y|] < \infty$. Chứng minh rằng

$$\mathbb{P}\left(\left\{\omega \in \Omega \;\middle|\; \lim_{N\to\infty} \frac{Y_1(\omega) + Y_2(\omega) + \ldots + Y_N(\omega)}{N} = \mathbb{E}[Y]\right\}\right) = 1.$$

(gợi ý: dựa vào định nghĩa hội tụ hầu chắc chắn và **Luật số lớn (mạnh)**).

Từ đó hãy nêu một cách tính $\hat{Y}$ là xấp xỉ của $\mathbb{E}[Y]$.

Bài 2. Giả định rằng $\mathbb{E}[|Y|] < \infty$ và $0 < \mathrm{Var}(Y) < \infty$. Chứng minh rằng

$$\frac{Y_1 + Y_2 + \ldots + Y_N}{N} \xrightarrow{w} \mathcal{N}\left(\mathbb{E}[Y], \frac{\mathrm{Var}(Y)}{N}\right).$$

(gợi ý: sử dụng **Định lý giới hạn trung tâm**).

Từ **Bài 1** ta thấy rằng $\hat{Y}_N$ là một xấp xỉ cho $\mathbb{E}[Y]$. Có nhận xét gì về sai số $\left|\hat{Y}_N - \mathbb{E}[Y]\right|$ của phép xấp xỉ này (xem xét sự phụ thuộc vào các đại lượng)?

Bài 3. Giả sử thêm rằng $Y_k = F(X_k)$, $k = 1, 2, \ldots$ (và $Y = F(X)$) với F khả tích tuyệt đối trên $[0, 1]$, $\int_0^1 |F(x)|\, \mathrm{d}x < \infty$, trong đó $X, X_1, X_2, \ldots, X_N$ tuân theo phân phối đều trên đoạn $[0, 1]$. Chứng minh rằng

$$\frac{F(X_1) + F(X_2) + \ldots + F(X_N)}{N} \xrightarrow{a.s.} \int_0^1 F(x)\, \mathrm{d}x.$$

Dựa vào tính hội tụ trên, viết một chương trình máy tính để tính xấp xỉ $\int_0^1 \frac{1}{x^5+1}\, \mathrm{d}x$ (làm tròn đến chữ số thập phân thứ 4).

2 Bài tập lập trình

Miêu tả bài tập và trình chấm tự động đã được ghi sẵn trong Python Notebook đi kèm, hãy hoàn thành bài tập trên Google Colab hoặc tải xuống thiết bị để hoàn thành bài tập.