

# Bài tập về nhà số 2

## Nền tảng toán học của các mô hình tạo sinh – PIMA

### Chủ đề: Phương pháp suy diễn

Người giải: Võ Hoàng Nhật Khang

1. Cho dãy các biến ngẫu nhiên độc lập và tuân theo cùng phân phối  $\mathcal{P}$ , tức là  $(Y_k : \Omega \rightarrow \mathbb{R})_{k \geq 1}$  với  $Y_k \sim \mathcal{P}$  và  $\mathbb{E}[|Y|] < \infty$ . Chứng minh rằng:

$$\mathbb{P} \left( \left\{ \omega \in \Omega \mid \lim_{N \rightarrow \infty} \frac{Y_1(\omega) + Y_2(\omega) + \dots + Y_N(\omega)}{N} = \mathbb{E}[Y] \right\} \right) = 1.$$

#### Lời giải:

Theo định lý Luật số lớn (mạnh), nếu  $(X_k)_{k \geq 1}$  là dãy các biến ngẫu nhiên độc lập và có cùng phân phối với  $\mathbb{E}[|X_1|] < \infty$ , thì:

$$\frac{X_1 + X_2 + \dots + X_N}{N} \xrightarrow{\text{a.s.}} \mathbb{E}[X_1] \quad \text{khi } N \rightarrow \infty.$$

Ở đây,  $(Y_k)_{k \geq 1}$  thỏa mãn các điều kiện: độc lập, cùng phân phối  $\mathcal{P}$ , và  $\mathbb{E}[|Y|] < \infty$ . Do đó:

$$\frac{Y_1 + Y_2 + \dots + Y_N}{N} \xrightarrow{\text{a.s.}} \mathbb{E}[Y].$$

Điều này tương đương với xác suất tập hợp các  $\omega$  mà tại đó trung bình cộng hội tụ về  $\mathbb{E}[Y]$  bằng 1, tức là:

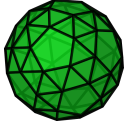
$$\mathbb{P} \left( \left\{ \omega \in \Omega \mid \lim_{N \rightarrow \infty} \frac{Y_1(\omega) + Y_2(\omega) + \dots + Y_N(\omega)}{N} = \mathbb{E}[Y] \right\} \right) = 1.$$

#### Cách tính xấp xỉ $\hat{Y}_N$ :

Dựa trên kết quả trên, ta có thể xấp xỉ  $\mathbb{E}[Y]$  bằng trung bình cộng của  $N$  mẫu:

$$\hat{Y}_N = \frac{Y_1 + Y_2 + \dots + Y_N}{N}.$$

Khi  $N$  lớn,  $\hat{Y}_N$  hội tụ hầu chắc chắn đến  $\mathbb{E}[Y]$ . ■



2. Giả định thêm  $0 < \text{Var}(Y) < \infty$ . Chứng minh rằng:

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

**Lời giải:**

Theo định lý giới hạn trung tâm, nếu  $(X_k)_{k \geq 1}$  là dãy các biến ngẫu nhiên độc lập, cùng phân phối, với  $\mathbb{E}[X_1] < \infty$  và  $0 < \text{Var}(X_1) < \infty$ , thì:

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

Đặt  $S_N = Y_1 + Y_2 + \dots + Y_N$ . Với  $(Y_k)_{k \geq 1}$  thỏa mãn các điều kiện đã cho, ta có:

$$\frac{S_N - N \cdot \mathbb{E}[Y]}{\sqrt{N \cdot \text{Var}(Y)}} \xrightarrow{w} \mathcal{N}(0, 1).$$

Biểu thức  $\frac{S_N}{N}$  có thể được viết lại như sau:

$$\frac{S_N}{N} = \mathbb{E}[Y] + \frac{1}{N} \cdot \left( \frac{S_N - N \cdot \mathbb{E}[Y]}{\sqrt{N \cdot \text{Var}(Y)}} \cdot \sqrt{N \cdot \text{Var}(Y)} \right) = \mathbb{E}[Y] + \frac{\sqrt{\text{Var}(Y)}}{\sqrt{N}} \cdot Z_N,$$

trong đó  $Z_N \xrightarrow{w} \mathcal{N}(0, 1)$ . Do đó, phân phối của  $\frac{S_N}{N}$  xấp xỉ:

$$\mathcal{N}\left(\mathbb{E}[Y], \left(\frac{\sqrt{\text{Var}(Y)}}{\sqrt{N}}\right)^2\right) = \mathcal{N}\left(\mathbb{E}[Y], \frac{\text{Var}(Y)}{N}\right).$$

Vậy:

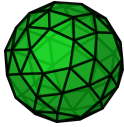
$$\frac{Y_1 + Y_2 + \dots + Y_N}{N} \xrightarrow{w} \mathcal{N}\left(\mathbb{E}[Y], \frac{\text{Var}(Y)}{N}\right).$$

**Nhận xét về sai số  $|\hat{Y}_N - \mathbb{E}[Y]|$ :**

Sai số  $|\hat{Y}_N - \mathbb{E}[Y]|$  phụ thuộc vào  $\text{Var}(Y)$  và  $N$ . Phương sai của  $\hat{Y}_N$  là:

$$\text{Var}(\hat{Y}_N) = \text{Var}\left(\frac{S_N}{N}\right) = \frac{\text{Var}(S_N)}{N^2} = \frac{N \cdot \text{Var}(Y)}{N^2} = \frac{\text{Var}(Y)}{N}.$$

Độ lệch chuẩn của  $\hat{Y}_N$  là  $\frac{\sqrt{\text{Var}(Y)}}{\sqrt{N}}$ , nghĩa là sai số kỳ vọng giảm theo  $\frac{1}{\sqrt{N}}$ . Để giảm sai số, ta cần tăng  $N$  hoặc giảm  $\text{Var}(Y)$  (nếu có thể). ■



3. Giả sử  $Y_k = F(X_k)$ , với  $X_k \sim \mathcal{U}[0, 1]$  và  $F$  khả tích tuyệt đối trên  $[0, 1]$ , tức  $\int_0^1 |F(x)| dx < \infty$ . Chứng minh rằng:

$$\frac{F(X_1) + F(X_2) + \cdots + F(X_N)}{N} \xrightarrow{\text{a.s.}} \int_0^1 F(x) dx.$$

**Lời giải:**

Vì  $X_k \sim \mathcal{U}[0, 1]$ , ta có:

$$\mathbb{E}[F(X_k)] = \int_0^1 F(x) \cdot 1 dx = \int_0^1 F(x) dx,$$

và:

$$\mathbb{E}[|F(X_k)|] = \int_0^1 |F(x)| dx < \infty.$$

Các  $X_k$  độc lập, nên  $Y_k = F(X_k)$  cũng độc lập. Áp dụng Luật số lớn (mạnh) cho dãy  $(Y_k)_{k \geq 1}$ :

$$\frac{Y_1 + Y_2 + \cdots + Y_N}{N} = \frac{F(X_1) + F(X_2) + \cdots + F(X_N)}{N} \xrightarrow{\text{a.s.}} \mathbb{E}[Y_1] = \int_0^1 F(x) dx.$$

**Tính xấp xỉ tích phân**  $\int_0^1 \frac{1}{x^5 + 1} dx$ :

Dựa trên kết quả trên, ta có thể dùng phương pháp Monte Carlo:

$$\hat{I}_N = \frac{1}{N} \sum_{k=1}^N \frac{1}{X_k^5 + 1},$$

với  $X_k \sim \mathcal{U}[0, 1]$ .

```
1 import numpy as np
2
3 def monte_carlo_integral(N):
4     X = np.random.uniform(0, 1, N)
5     F_X = 1 / (X**5 + 1)
6     I_hat = np.mean(F_X)
7     return I_hat
8
9 # N = 1000000 steps
10 N = 1000000
11 approx_integral = monte_carlo_integral(N)
12 print(f"Approximation with N = {N}: {approx_integral}")
```

Kết quả được làm tròn đến chữ số thập phân thứ 4 là: 0.8883. ■