

ĐẠI HỌC QUỐC GIA TP.HỒ CHÍ MINH  
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN  
KHOA VẬT LÝ - VẬT LÝ KỸ THUẬT  
CHUYÊN NGÀNH VẬT LÝ TIN HỌC

—oOo—

KHOÁ LUẬN TỐT NGHIỆP

Đề tài:

GIẢI TÍCH SỐ TÍCH PHÂN  
MONTE CARLO NHIỀU LỚP  
BẰNG GIEO ĐIỂM QUAN  
TRỌNG

SVTH: Huỳnh Thị Hạ Vy

CBHD: TS. Nguyễn Chí Linh

TP. HỒ CHÍ MINH - 2019

ĐẠI HỌC QUỐC GIA TP.HỒ CHÍ MINH  
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN  
KHOA VẬT LÝ - VẬT LÝ KỸ THUẬT  
CHUYÊN NGÀNH VẬT LÝ TIN HỌC

—————oOo—————

KHOÁ LUẬN TỐT NGHIỆP

Đề tài:

GIẢI TÍCH SỐ TÍCH PHÂN  
MONTE CARLO NHIỀU LỚP  
BẰNG GIEO ĐIỂM QUAN  
TRỌNG

SVTH: Huỳnh Thị Hạ Vy

CBHD: TS. Nguyễn Chí Linh

—————  
TP. HỒ CHÍ MINH - 2019

# Lời cảm ƠN

*TP. Hồ Chí Minh, tháng 7 năm 2019.*

Đầu tiên,

Huỳnh Thị Hạ Vy

# Mục lục

# Lời giới thiệu

Ngày nay,

- Chương ??: Kiến thức cơ sở xác suất và thống kê.....
- Chương ??: Tích phân Monte Carlo. Tích phân Monte Carlo.
- Chương ??: Thuật toán Vegas. Thuật toán Vegas.
- Chương ??: Kết luận và hướng phát triển. ....

---

# CHƯƠNG 1

---

## Kiến thức cơ sở xác suất và thống kê

### 1.1 Biến ngẫu nhiên và xác suất

Biến ngẫu nhiên là một hàm ánh xạ với đặc điểm nó gán một giá trị bằng số cho kết quả đầu ra của một phép thử ngẫu nhiên.

$$\mathbf{X}(\omega) = x \quad (1.1)$$

với  $\omega$  là đại diện cho đầu ra của một thực nghiệm,  $x$  là một số thực (hay sự kiện),  $\mathbf{X}$  là hàm ánh xạ (hay biến ngẫu nhiên). Tính ngẫu nhiên được thể hiện ở tham số đầu vào  $\omega$ . Điều này dẫn đến đầu ra của hàm là ngẫu nhiên.

Đây chưa phải là định nghĩa đầy đủ của một biến ngẫu nhiên. Khái niệm khác liên quan đến định nghĩa của một biến ngẫu nhiên là khái niệm xác suất.

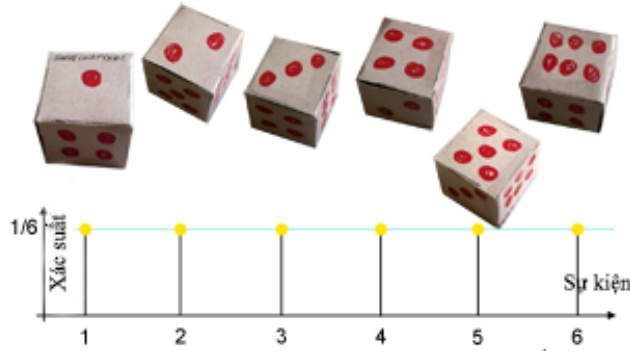
Xét một số thí dụ sau:

- Gieo một đồng xu trên mặt phẳng, đây là một phép thử. Kết quả có thể xảy ra là *Xuất hiện mặt sấp* hoặc *Xuất hiện mặt ngửa*. Như vậy xác suất cho các khả năng *Xuất hiện mặt sấp* và *Xuất hiện mặt ngửa* lần lượt là  $\frac{1}{2}$  và  $\frac{1}{2}$



Hình 1.1: Các sự kiện có thể xuất hiện khi gieo đồng xu

- Gieo một con súc sắc, đây là một phép thử. Kết quả có thể xảy ra là "Xuất hiện mặt  $k$  chấm" tương ứng với  $k = 1, 2, 3, \dots, 6$ . Xác suất cho mỗi sự kiện "Xuất hiện  $k$  chấm" đều là  $\frac{1}{6}$ .



**Hình 1.2:** Các sự kiện có thể xuất hiện khi gieo súc sắc

Tổng quát, nếu một phép thử tạo ra  $n$  sự kiện khác nhau và khả năng xảy ra như nhau thì xác suất của mỗi sự kiện là  $\frac{1}{n}$  chúng ta có thể nói rằng nếu kết quả của một quá trình nào đó phải là một trong  $n$  kết quả khác nhau.

## 1.2 Phân phối xác suất

Phân phối xác suất xác định xác suất cho từng sự kiện có thể xảy ra của một phép thử ngẫu nhiên. Hàm phân phối xác suất của biến ngẫu nhiên  $\mathbf{X}$  được xác định như sau:

$$\mathbf{F}(x) = \mathbf{P}(\mathbf{X} \leq x) = \mathbf{P}\omega \in \Omega : \mathbf{X}(\omega) \leq x \quad (1.2)$$

với mọi  $x \in \mathbf{R}$  được gọi là hàm phân phối xác suất của biến ngẫu nhiên  $\mathbf{X}$ .

Đặc trưng của hàm phân phối xác suất hay gọi là hàm phân phối tích lũy (CDF) là lấy xác suất các biến ngẫu nhiên bên trái của một giá trị  $\mathbf{X}$  bất kì nào đó. Hàm này có đặc điểm là một hàm không giảm, tức là nếu  $a < b$  thì  $\mathbf{F}(a) \leq \mathbf{F}(b)$  vì sự kiện  $b$  đã bao gồm cả sự kiện  $a$  rồi.

### 1.2.1 Hàm khối xác suất của biến rời rạc

Hàm xác suất để xác định xác suất tại mỗi giá trị  $\mathbf{X}$  nào đó trong miền giá trị là bao nhiêu. Hàm xác suất như vậy được gọi là hàm khối xác suất (PMF) Đối với các



biến ngẫu nhiên rời rạc. Giả sử miền xác định của  $\mathbf{X}$  là  $\mathbf{D}$ , tức  $\mathbf{X}:\Omega \mapsto \mathbf{D}$  thì hàm khối xác suất được xác định như sau:

$$p(x) = p_{\mathbf{X}}(x) = \begin{cases} \mathbf{P}(\mathbf{X} = x) & \text{if } x \in \mathbf{D} \\ 0 & \text{if } x \notin \mathbf{D} \end{cases} \quad (1.3)$$

Như vậy ta có thể thấy rằng hàm khối xác suất thực chất cũng là một xác suất nên nó mang đầy đủ tất cả các tính chất của xác suất như:

- $0 \leq p(x) \leq 1$
- $\sum_{x_i \in \mathbf{D}} p(x_i) = 1$

### 1.2.2 Hàm mật độ xác suất của biến liên tục

Hàm mật độ xác suất (PDF - Probability Density Function) đối với các biến ngẫu nhiên liên tục, dùng để ước lượng độ tập trung xác suất tại lân cận điểm nào đó. Hàm mật độ xác suất  $f(x)$  tại điểm  $x$  được xác định bằng cách lấy đạo hàm của hàm phân phối tích lũy  $\mathbf{F}(x)$  tại điểm đó:

$$f(x) = \mathbf{F}'(x) \quad (1.4)$$

Như vậy thì nơi nào  $f(x)$  càng lớn thì ở đó mức độ tập xác suất càng cao. Từ đây ta cũng có thể biểu diễn hàm phân phối tích lũy như sau:

$$\mathbf{F}(x) = \int_{-\infty}^x f(t) dt \quad (1.5)$$

Xác suất trong 1 khoảng  $(\alpha, \beta)$  cũng có thể được tính bằng hàm mật độ xác suất:

$$\mathbf{F}(x) = \int_{\alpha}^{\beta} f(x) dx \quad (1.6)$$

Hàm mật độ xác suất cũng có 2 tính chất như xác suất như sau:

- Không âm:  $f(x) \geq 0, \forall (x) \in \mathbf{R}$
- Tổng toàn miền bằng 1:  $\int_{-\infty}^{\infty} f(x) dx = 1$

### 1.3 Kỳ vọng và phương sai

Giá trị kỳ vọng của  $X$  là tổng “sự kiện” nhân với xác suất của sự kiện đó. Như đã nhắc đến phần ??, xác suất được tính theo hàm PMF cho biến ngẫu nhiên rời rạc và hàm PDF cho biến ngẫu nhiên liên tục. Hay nói cách khác giá trị kỳ vọng là giá trị trung bình có trọng số.

$$\mathbf{E}[\mathbf{X}] = \sum_{i=0}^{N-1} X_i pmf(X_i) \quad (1.7)$$

Trong đó  $N$  là số “sự kiện” có thể xảy ra của một biến ngẫu nhiên rời rạc  $\mathbf{X}$ . Để mở rộng cho biến ngẫu nhiên liên tục  $\mathbf{X}$ :

$$\mathbf{E}[\mathbf{X}] = \int_{-\infty}^{\infty} \mathbf{X} pdf(\mathbf{X}) \quad (1.8)$$

Giả sử  $\mathbf{X}$  là một biến ngẫu nhiên có tập giá trị sau 1, 2, 3, 4, 5, 6 và  $\mathbf{F}[\mathbf{X}] = (\mathbf{X} - 3)^2$ . Chú ý rằng  $\mathbf{X}$  là biến ngẫu nhiên thì  $\mathbf{E}[\mathbf{X}]$  cũng là một biến ngẫu nhiên. Như vậy:

- $\mathbf{X} = 1, \mathbf{F}(1) = (1 - 3)^2 = 4$
- $\mathbf{X} = 2, \mathbf{F}(2) = (2 - 3)^2 = 1$
- $\mathbf{X} = 3, \mathbf{F}(3) = (3 - 3)^2 = 0$
- $\mathbf{X} = 4, \mathbf{F}(4) = (4 - 3)^2 = 1$
- $\mathbf{X} = 5, \mathbf{F}(5) = (5 - 3)^2 = 4$
- $\mathbf{X} = 6, \mathbf{F}(6) = (6 - 3)^2 = 9$

Nhận thấy, xác suất của một kết quả  $\mathbf{Y}$  từ  $\mathbf{F}[\mathbf{X}]$  bằng tổng xác suất của bất kỳ  $X$  nào thỏa  $\mathbf{F}[\mathbf{X}] = \mathbf{Y}$ .

$$\begin{aligned} \mathbf{P}(\mathbf{F}(0)) &= \frac{1}{6} \\ \mathbf{P}(\mathbf{F}(1)) &= \frac{1}{6} + \frac{1}{6} = \frac{2}{6} \\ \mathbf{P}(\mathbf{F}(4)) &= \frac{1}{6} + \frac{1}{6} = \frac{2}{6} \\ \mathbf{P}(\mathbf{F}(9)) &= \frac{1}{6} \end{aligned}$$

Trong toán học có hai phương pháp để tính giá trị kỳ vọng của  $\mathbf{F}[\mathbf{X}]$

- Phương pháp 1: Nếu phân phối xác suất của  $\mathbf{F}[\mathbf{X}]$  đã biết trước, thì giá trị kỳ vọng được tính như sau:

$$\mathbf{E}[\mathbf{F}[\mathbf{X}]] = \mathbf{E}[\mathbf{Y}] = \sum Y_i \mathbf{P}(Y_i) \quad (1.9)$$

- Phương pháp 2: Nếu phân phối xác suất của  $\mathbf{F}[\mathbf{X}]$  chưa biết trước, thì giá trị kỳ vọng được tính như sau: Đối với biến ngẫu nhiên rời rạc

$$\mathbf{E}[\mathbf{F}(\mathbf{X})] = \sum \mathbf{F}(\mathbf{X}) \mathbf{P}(\mathbf{X}) \quad (1.10)$$

Đối với biến ngẫu nhiên liên tục

$$\mathbf{E}[\mathbf{F}(\mathbf{X})] = \int \mathbf{F}(\mathbf{X}) \mathbf{P}(\mathbf{X}) d\mathbf{X} \quad (1.11)$$

Phương pháp 2 đóng vai trò quan trọng trong phương pháp Monte Carlo. Bởi vì hàm phân phối xác suất  $\mathbf{F}(\mathbf{X})$  có thể không được biết trước. Phương sai được xác định như sau

$$\sigma^2[\mathbf{Y}] = E[(\mathbf{Y} - \mathbf{E}[\mathbf{Y}])^2] \quad (1.12)$$

trong đó  $\sigma$ , độ lệch chuẩn, là căn bậc hai của phương sai. Từ những định nghĩa này, nó là dễ dàng cho thấy rằng với bất kỳ hằng số  $a$

$$\mathbf{E}[a\mathbf{Y}] = a\mathbf{E}[\mathbf{Y}] \quad (1.13)$$

$$\sigma^2[a\mathbf{Y}] = a^2 \sigma^2[\mathbf{Y}] \quad (1.14)$$

Hơn nữa, giá trị kỳ vọng của một tổng các biến ngẫu nhiên  $Y_i$  là tổng giá trị kỳ vọng của  $Y_i$ :

$$E\left[\sum_i Y_i\right] = \sum_i E[Y_i] \quad (1.15)$$

Từ các tính chất này, có thể rút ra biểu thức đơn giản hơn cho phương sai:

$$\sigma^2[\mathbf{Y}] = \mathbf{E}[\mathbf{Y}^2] - \mathbf{E}[\mathbf{Y}]^2 \quad (1.16)$$

Ngoài ra, nếu các biến ngẫu nhiên độc lập thì:

$$\sigma^2\left[\sum_i Y_i\right] = \sum_i \sigma^2[Y_i] \quad (1.17)$$

---

## CHƯƠNG 2

---

### Tích phân Monte Carlo

#### 2.1 Tổng quát

Giá trị kỳ vọng của một hàm biến ngẫu nhiên  $f(\mathbf{X})$  có thể được định nghĩa là:

$$\mathbf{E}[f(\mathbf{X})] = \int f(\mathbf{X})P(\mathbf{X})d\mathbf{X} \quad (2.1)$$

Với  $P(\mathbf{X})$  là phân phối xác suất của biến ngẫu nhiên  $\mathbf{X}$ . Phương sai của biến ngẫu nhiên  $\mathbf{X}$  được xác định:

#### 2.2 Tích phân Monte Carlo với hàm phân bố đều

#### 2.3 Tích phân Monte Carlo với hàm phân bố bất kỳ

#### 2.4 Giảm phương sai

---

# CHƯƠNG 3

---

## Thuật toán Vegas

### 3.1 Giới thiệu

Phương pháp Monte Carlo sử dụng các thuật toán để giải quyết các bài toán trên máy tính bằng cách lấy mẫu ngẫu nhiên. Kết quả của phương pháp Monte Carlo này càng chính xác khi số lượng mẫu càng tăng. Theo quy luật số lớn khi kích thước mẫu càng lớn thì tốc độ hội tụ của tích phân càng lớn. Nhưng khi mở rộng số chiều của tích phân thì số lượng mẫu cũng phải tăng theo. Điều này cũng làm ảnh hưởng đến tốc độ tính toán của máy tính. Nếu cố định số lượng mẫu thì tốc độ hội tụ của tích phân sẽ giảm khi số chiều tăng lên. Nhìn chung, tốc độ hội tụ của phương pháp Monte Carlo là khá thấp. Vì vậy, phương pháp tích phân Monte-Carlo sử dụng thuật toán Vegas sẽ giải quyết hiệu quả với các tích phân nhiều chiều.

Đặc điểm thuật toán:

1. Phương sai chính xác cho tích phân thì được tính toán dễ dàng.
2. Hàm tính tích phân không cần liên tục khi sử dụng thuật toán mới này. Đặc biệt, các hàm bước nhảy cũng không có vấn đề gì. Vì thế để tính tích phân nhiều chiều là việc đơn giản.
3. Tốc độ hội tụ không phụ thuộc vào số chiều của tích phân.
4. Đây là thuật toán thích nghi, nó tự động tập trung tính toán tích phân tại những vùng có tích phân quan trọng nhất.

Đặc điểm (1) và (3) là phổ biến trong tất cả phương pháp Monte Carlo. Với (4) là một tính năng quan trọng nhất trong thuật toán này. Vấn đề chính trong tính toán các tích phân nhiều chiều là việc tăng số lượng mẫu theo cấp số nhân khi tăng số chiều. Do đó, mục đích chung của thuật toán để tính toán tích phân nhiều chiều là thích nghi.

## 3.2 Tích phân Monte Carlo sử dụng thuật toán Vegas

Monte Carlo thật ra là một phương pháp khá đơn giản. Nhưng vấn đề giảm phương sai cần được quan tâm. Để giảm phương sai, tất cả những gì có thể làm là tăng số lượng mẫu  $N$ . Muốn giảm phương sai xuống hai lần thì phải tăng gấp bốn lần số mẫu. Đối với các hàm tích phân nhiều chiều, số lượng mẫu tăng theo cấp số nhân khi số chiều tăng lên. Do đó, có nhiều phương pháp đã được nghiên cứu để giảm phương sai mà không cần tăng số lượng mẫu  $N$ . Như vậy, phương pháp lấy mẫu trọng và phương pháp lấy mẫu phân tầng được thuật toán Vegas sử dụng để giảm phương sai.

### 3.2.1 Lấy mẫu trọng

### 3.2.2 Lấy mẫu phân tầng

### 3.2.3 Thuật toán Vegas

---

## CHƯƠNG 4

---

### Kết luận và hướng phát triển

#### 4.1 Kết luận

## Phụ lục A

### Định nghĩa đầy đủ của câu lệnh



## Phụ lục B

Trích đoạn tập tin đầu vào  $\text{\LaTeX}$  của đề tài