1. Random variables

Một biến ngẫu nhiên (random variable) x là một đại lượng dùng để đo những đại lượng không xác định. Biến này có thể ký hiệu kết quả/đầu ra (outcome) của một thí nghiệm (ví dụ như tung đồng xu) hoặc một đại lượng biến đổi trong tự nhiên (ví dụ như nhiệt độ trong ngày). Nếu chúng ta quan sát rất nhiều đầu ra {x_i}_{i=1}^I của các thí nghiệm này, ta có thể nhận được những giá trị khác nhau ở mỗi thí nghiệm. Tuy nhiên, sẽ có những giá trị xảy ra nhiều lần hơn những giá trị khác. Thông tin về đầu ra được đo bởi phân phối xác suất (probability distribution) p(x) của biến ngẫu nhiên.

Một biến ngẫu nhiên có thể là rời rạc (discrete) hoặc liên tục (continuous). Một biến ngẫu nhiên rời rạc sẽ lấy giá trị trong một tập hợp cho trước. Ví dụ tung đồng xu thì có hai khả năng là head và tail (tên gọi này bắt nguồn từ đồng xu Mỹ, một mặt có hình mặt người, được gọi là head, trái ngược với mặt này được gọi là mặt tail, cách gọi này hay hơn cách gọi xấp ngửa vì ta không có quy định rõ ràng thế nào là xấp ngay ngửa). Tập các giá trị này có thể là có thứ tự (khi tung xúc xắc) hoặc không có thứ tự (unordered), ví dụ khi đầu ra là các giá trị nắng, mưa, bão, etc. Mỗi đầu ra có một giá trị xác suất tương ứng với nó. Các giá trị xác suất này không âm và có tổng bằng một:

if x is discrete:

$$\sum_x p(x) = 1 \quad (1)$$

Biến ngẫu nhiên liên tục lấy các giá trị là tập con của các số thực. Những giá trị này có thể là hữu hạn, ví dụ thời gian làm bài của mỗi thí sinh trong một bài thi 180 phút, hoặc vô hạn, ví dụ thời gian để chiếc xe bus tiếp theo tới. Không như biến ngẫu nhiên rời rạc, xác suất để đầu ra bằng chính xác một giá trị nào đó, theo lý thuyết, là bằng 0. Thay vào đó, ta có thể hình dung xác suất để đầu ra nằm trong một khoảng giá trị nào đó; và việc này được mô tả bởi hàm mật độ xác suất (probability density function - pdf). Hàm mật độ xác suất luôn cho giá trị dương, và tích phân của nó trên toàn miền possible outcome phải bằng 1.

if x is continuous:

$$\int p(x)\, dx = 1 \quad (2)$$

Để giảm thiểu ký hiệu, hàm mật độ xác suất của một biến ngẫu nhiên liên tục x cũng được ký hiệu là p(x).

Chú ý: Nếu x là biến ngẫu nhiên rời rạc, p(x) luôn luôn nhỏ hơn hoặc bằng 1. Trong khi đó, nếu x là biến ngẫu nhiên liên tục, p(x) có thể nhận giá trị dương bất kỳ, điều này vẫn đảm bảo là tích phân của hàm mật độ xác suất theo toàn bộ giá trị có thể có của x bằng 1. Với biến ngẫu nhiên rời rạc, p(x) được hiểu là mật độ xác suất tại x.

2. Joint probability

Xét hai biến ngẫu nhiên x và y. Nếu ta quan sát rất nhiều cặp đầu ra của x và y, thì có những tổ hợp hai đầu ra xảy ra thường xuyên hơn những tổ hợp khác. Thông tin này được biểu diễn bằng một phân phối được gọi là joint probability của x và y, và được viết là p(x, y). Dấu phẩy trong p(x, y) có thể đọc là và, vậy p(x, y) là xác suất của x và y. x và y có thể là hai biến ngẫu nhiên rời rạc, liên tục, hoặc một rời rạc, một liên tục. Luôn nhớ rằng tổng các xác suất trên mọi cặp giá trị có thể xảy ra (x, y) bằng 1.

both are discrete: $\sum_{x,y} p(x, y) = 1$

both are continuous: $\iint p(x, y)\, dx\, dy = 1$

x is discrete, y is continuous: $\sum_x \int p(x, y)\, dy = \int (\sum_x p(x, y))\, dy = 1$

Thông thường, chúng ta sẽ làm việc với các bài toán ở đó joint probability xác định trên nhiều hơn 2 biến ngẫu nhiên. Chẳng hạn, p(x, y, z) thể hiện joint probability của 3 biến ngẫu nhiên x, y và z. Khi có nhiều biến ngẫu nhiên, ta có thể viết chúng dưới dạng vector. Ta có thể viết p(x) để thể hiện joint probability của biến ngẫu nhiên nhiều chiều x = [x1, x2, …, xn]^T. Khi có nhiều tập các biến ngẫu nhiên, ví dụ x và y, ta có thể viết p(x, y) để thể hiện joint probability của tất cả các thành phần trong hai biến ngẫu nhiên nhiều chiều này.

3. Marginalization

Nếu biết joint probability của nhiều biến ngẫu nhiên, ta cũng có thể xác định được phân bố xác suất của từng biến bằng cách lấy tổng (rời rạc) hoặc tích phân (liên tục) theo tất cả các biến còn lại:

$p(x) = \sum_y p(x, y)$ (3)

$p(y) = \sum_x p(x, y)$ (4)

Và với biến liên tục:

$p(x) = \int p(x, y)\, dy$ (5)

$p(y) = \int p(x, y)\, dx$ (6)

Với nhiều biến hơn, chẳng hạn 4 biến rời rạc x, y, z, w:

$p(x) = \sum_{y,z,w} p(x, y, z, w)$ (7)

$p(x, y) = \sum_{z,w} p(x, y, z, w)$ (8)

Từ đây trở đi, nếu không nói gì thêm, tôi sẽ dùng ký hiệu $\sum$ để chỉ chung cho cả hai loại biến. Nếu biến ngẫu nhiên là liên tục, bạn đọc ngầm hiểu rằng dấu $\sum$ cần được thay bằng dấu tích phân $\int$, biến lấy vi phân chính là biến được viết dưới dấu $\sum$.

4. Conditional probability

Xác suất có điều kiện (conditional probability) của một biến ngẫu nhiên x biết rằng biến ngẫu nhiên y có giá trị y* được ký hiệu là $p(x \mid y = y^*)$. Conditional probability $p(x \mid y = y^*)$ có thể được tính dựa trên joint probability $p(x, y)$.

$p(x \mid y = y^*) = p(x, y = y^*) / \sum_x p(x, y = y^*) = p(x, y = y^*) / p(y = y^*)$ (9)

Thông thường, viết gọn: $p(x \mid y) = p(x, y) / p(y)$ (10)

Tương tự: $p(y \mid x) = p(y, x) / p(x)$

Quan hệ: $p(x, y) = p(x \mid y)\, p(y) = p(y \mid x)\, p(x)$ (11)

Khi có nhiều hơn hai biến:

$p(x, y, z, w) = p(x, y, z \mid w)\, p(w)$ (12)

$= p(x, y \mid z, w)\, p(z, w) = p(x, y \mid z, w)\, p(z \mid w)\, p(w)$ (13)

$= p(x \mid y, z, w)\, p(y \mid z, w)\, p(z \mid w)\, p(w)$ (14)

5. Quy tắc Bayes

Từ (11): $p(y \mid x)\, p(x) = p(x \mid y)\, p(y) \Rightarrow p(y \mid x) = p(x \mid y)\, p(y) / p(x)$ (15)

$= p(x \mid y)\, p(y) / \sum_y p(x, y)$ (16)

$= p(x \mid y)\, p(y) / \sum_y p(x \mid y)\, p(y)$ (17)

Ba công thức (15)–(17) thường được gọi là Quy tắc Bayes (Bayes' rule).

6. Independence

Nếu x và y độc lập: $p(x \mid y) = p(x)$ (18), $p(y \mid x) = p(y)$ (19)

Khi đó: $p(x, y) = p(x)\, p(y)$ (20)

7. Kỳ vọng

Kỳ vọng (expectation) của một biến ngẫu nhiên:

$E[x] = \sum_x x\, p(x)$ nếu x rời rạc (21)

$E[x] = \int x\, p(x)\, dx$ nếu x liên tục (22)

Với hàm f(.): $E[f(x)] = \sum_x f(x)\, p(x)$ (23)

Với joint probability: $E[f(x, y)] = \sum_{x,y} f(x, y)\, p(x, y)$ (24)

Ba quy tắc:

$E[\alpha] = \alpha$ (25)

$E[\alpha x] = \alpha E[x]$ (26); $E[f(x) + g(x)] = E[f(x)] + E[g(x)]$ (27)

Nếu x, y độc lập: $E[f(x)\, g(y)] = E[f(x)]\, E[g(y)]$ (28)

8. Một vài phân phối thường gặp

8.1 Bernoulli distribution

Bernoulli distribution: $x \in \{0,1\}$, tham số $\lambda \in [0,1]$ là xác suất để x=1.

$p(x=1) = \lambda$, $p(x=0) = 1 - \lambda$

Viết gọn: $p(x) = \lambda^x (1 - \lambda)^{1 - x}$ (29)

Ký hiệu: $p(x) = Bern_x[\lambda]$ (30)

8.2 Categorical distribution

Categorical distribution với K lớp, tham số $\lambda = [\lambda 1, \ldots, \lambda K]$, $\sum_k \lambda k = 1$.

$p(x = k) = \lambda_k$; viết gọn: $p(x) = Cat_x[\lambda]$

Biểu diễn one-hot: $x \in \{e1, \ldots, eK\}$

$p(x = e_k) = \prod_{j=1}^K \lambda_j^{x_j} = \lambda_k$ (31)

8.3 Univariate normal distribution

$p(x) = 1 / \sqrt{(2\pi\sigma^2)} \cdot \exp(-(x - \mu)^2 / (2\sigma^2))$ (32)

Ký hiệu: $p(x) = Norm_x[\mu, \sigma^2]$

8.4 Multivariate normal distribution

$p(x) = 1 / ((2\pi)^{D/2} |\Sigma|^{1/2}) \cdot \exp(-1/2 (x-\mu)^T \Sigma^{-1} (x-\mu))$ (33)

Ký hiệu: $p(x) = Norm_x[\mu, \Sigma]$

8.5 Beta distribution

$p(\lambda) = \Gamma(\alpha+\beta) / (\Gamma(\alpha)\, \Gamma(\beta)) \cdot \lambda^{\alpha-1} (1-\lambda)^{\beta-1}$ (34)

Trong đó $\Gamma(z) = \int_0^\infty t^{z-1} \exp(-t)\, dt$, và $\Gamma[z] = (z-1)!$ nếu z là số tự nhiên.

Ký hiệu: $p(\lambda) = Beta_\lambda[\alpha, \beta]$

8.6 Dirichlet distribution

$$p(\lambda_1, \ldots, \lambda_K) = \frac{\Gamma(\sum_{k=1}^{K} \alpha_k)}{\prod_{k=1}^{K} \Gamma(\alpha_k)} \cdot \prod_{k=1}^{K} \lambda_k^{\alpha_k - 1} \quad (35)$$

Ký hiệu: $p(\lambda_1, \ldots, \lambda_K) = \mathrm{Dir}_{\{\lambda_1,\ldots,\lambda_K\}}[\alpha_1, \ldots, \alpha_K]$

9. Thảo luận

Về Xác suất thống kê, còn rất nhiều điều cần lưu ý. Tạm thời, phần này ôn tập lại các kiến thức xác suất cơ bản để phục vụ cho các bài viết tiếp theo. Khi nào có phần nào cần nhắc lại, sẽ tiếp tục ôn tập bổ sung.