Bài 3 Chuẩn bị toán học

3.1 Xác suất (Probability)

3.2 Bất đẳng thức Chebyshev và luật yếu của số lớn

3.3 Tập lồi (Convex sets) và hàm lồi (convex functions), bất đẳng thức Jensen

3.4 Công thức Stirling

Xác suất

- Không gian mẫu (Sample space)
  - Là tập (hay không gian) tất cả các kết quả có thể có của một thí nghiệm. Thường được kí hiệu là *E* hay *S*. Nếu không gian mẫu là rời rạc thì *E* có thể được biểu diễn bằng $E = \{e_1, e_2, ..., e_n\}$
- Sự kiện (Event), sự kiện cơ bản (elementary event)
  - Mỗi tập con của *E* (không gian mẫu) được gọi là một sự kiện, đặc biệt mỗi phần tử của *E* được gọi là một sự kiện cơ bản.
- Ví dụ
  - Trong một thí nghiệm tung đồng xu thì *E* = {U (úp), N (ngửa)}. Nếu đồng tiền là đồng nhất thì xác suất *P*(U) = *P*(N) = 1/2.
  - Trong một thí nghiệm tung con xúc xắc thì $E = \{1, 2, 3, 4, 5, 6\}$. Nếu con xúc xắc là đồng nhất thì xác suất $P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = 1/6$, $P(2, 5) = 1/3$, $P(1, 3, 5) = 1/2$.

Xác suất (tt)

  - Lấy một văn bản tiếng Anh điển hình và nhặt một kí tự bất kỳ thì $E = \{a, b, c, ..., x, y, z\}$ và xác suất của các kí tự được phân bố như sau $P(a) = 0{,}0642$, ..., $P(e) = 0{,}103$, ..., $P(z) = 0{,}0005$.
- Biến ngẫu nhiên rời rạc (Discrete random variable)
  - Một biến ngẫu nhiên rời rạc x được định nghĩa bằng cách gán một số thực x_i tới mỗi sự kiện cơ bản e_i của không gian mẫu rời rạc E. Xác suất của x_i được định nghĩa là xác suất của sự kiện cơ bản tương ứng và được kí hiệu là $p(x_i)$.
- Trị trung bình (kỳ vọng) (average, expected value), phương sai (variance)
  - Trị trung bình và phương sai của biến ngẫu nhiên rời rạc x lần lượt được kí hiệu và định nghĩa như sau
  - $E(\mathrm{x}) \quad = \overline{\mathrm{x}} = \sum_i \mathrm{x}_i p(\mathrm{x}_i)$

Xác suất (tt)

- Var(**x**) $= E\left(\left(\mathrm{x}-\overline{\mathrm{x}}\right)^2\right) = \sum_i \left(\mathrm{x}_i - \overline{\mathrm{x}}\right)^2 p(\mathrm{x}_i)$

  $= E\left(\mathrm{x}^2\right) - \overline{\mathrm{x}}^2$

  trong đó $E(\mathbf{x}^2)$ là trị kỳ vọng của $\mathbf{x}^2$.
- Tổng quát, trị kỳ vọng của một hàm của **x**, chẳng hạn *f*(**x**), được định nghĩa bằng

$$E\left(f(\mathrm{x})\right) = \sum_i f(\mathrm{x}_i) p(\mathrm{x}_i)$$

- Xác suất đồng thời (joint probability), xác suất có điều kiện (conditional probability)
  - Một cặp biến ngẫu nhiên (x, y) liên kết với một thí nghiệm tạo thành một biến ngẫu nhiên nối (joint random variable). Nếu x, y là rời rạc, sự phân bố xác suất nối hay xác suất đồng thời được định nghĩa là

$$p_{ij} = \mathrm{P}(\mathrm{x} = x_i, \mathrm{y} = y_j)$$

Xác suất (tt)

- Xác suất của y trong điều kiện đã biết x được gọi là xác suất có điều kiện và được định nghĩa là

$$p(y_j | x_i) = \frac{p(x_i, y_j)}{p(x_i)}$$

trong đó **xác suất lề** (marginal probability) $p(x_i)$ được giả thiết là khác không.

- Các xác suất lề được định nghĩa như sau:

$$p(x_i) = \sum_j p(x_i, y_j)$$

$$p(y_j) = \sum_i p(x_i, y_j)$$

Ví dụ

- Thí nghiệm tung đồng thời một đồng xu và con xúc xắc.
- Từ kết quả trên ta thấy

  P(U, 5) = 1/18

  P(Đồng xu = U) = 5/9

  P(Đồng xu = N) = 4/9

  P(Xúc xắc = 5) = 7/72

  P(Xúc xắc = 5 đã biết Đồng xu = U)

Xúc xắc \ Đồng xu	U	N
6	1/12	1/12
5	1/18	1/24
4	1/9	1/24
3	1/9	1/6
2	1/9	1/18
1	1/12	1/18

Xác suất (tt)

- Sự độc lập (Independence)
  - Hai biến ngẫu nhiên x và y được gọi là độc lập nếu

$$p(x_i, y_j) = p(x_i)p(y_j)\ \forall\ i, j.$$

  - Chúng ta thấy nếu hai biến x và y độc lập thì

$$p(y_j|x_i) = \frac{p(x_i, y_j)}{p(x_i)} = \frac{p(x_i)p(y_j)}{p(x_i)} = p(y_j)$$

    có nghĩa là xác suất y_j trong điều kiện có x_i xảy ra hay không xảy ra đều như nhau, không thay đổi, và ngược lại.
  - Cũng từ sự độc lập chúng ta suy ra một kết quả mà hay được sử dụng sau này

$$E(\mathrm{xy}) = \mathrm{E(x)\ E(y)} = \overline{\mathrm{x}}\,\overline{\mathrm{y}}$$

Xác suất (tt)

- Sự tương quan (correlation)
  - Sự tương quan C giữa hai biến x và y được định nghĩa là trị kỳ vọng của $(x - \overline{x})(y - \overline{y})$:

$$C(x, y) = E((x - \overline{x})(y - \overline{y})) = E(xy) - \overline{x}\,\overline{y}$$

  - Trong trường hợp x và y là độc lập chúng ta suy ra $C(x, y) = 0$. Tuy nhiên điều ngược lại thì không đúng.

Bất đẳng thức Chebyshev và luật yếu của số lớn

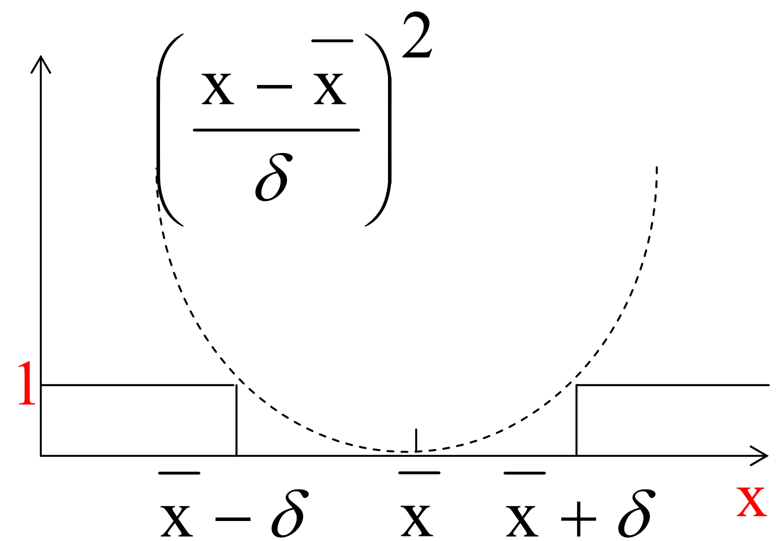
- Bất đẳng thức Chebyshev
  - Cho một biến ngẫu nhiên **x** có trị trung bình là $\overline{x}$ và phương sai là δ_x^2, bất đẳng thức Chebyshev đối với một số dương tuỳ ý δ là

$$P(|\mathbf{x} - \overline{x}| \geq \delta) \leq \frac{\delta_x^2}{\delta^2}$$

- Chứng minh
  - Định nghĩa một hàm *f*(*x*) như sau $f(x) = \begin{cases} 1, & |x - \overline{x}| \geq \delta \\ 0, & |x - \overline{x}| < \delta \end{cases}$
  - Thì

$$P(|\mathbf{x} - \overline{x}| \geq \delta) = \Sigma f(x_i) p(x_i)$$

Bất đẳng thức Chebyshev (tt)



- Dựa trên hình chúng ta có

$$f(x) \le \left(\frac{x-\bar{x}}{\delta}\right)^2$$

- Vì vậy,

$$P\left(\left|x-\bar{x}\right| \ge \delta\right) \le \sum_i \left(\frac{x-\bar{x}}{\delta}\right)^2 p(x_i) = \frac{\delta_x^2}{\delta^2}$$

Luật yếu của số lớn (tt)

- Xét một thí nghiệm nhị phân trong đó các kết quả của thí nghiệm là 0 và 1 với các xác suất tương ứng là p_0 và $1-p_0$.
- Thí nghiệm này được lặp lại N lần một cách độc lập, và kết quả trung bình được định nghĩa là $\mathbf{y}_N$; tức là, $\mathbf{y}_N$ bằng tổng số các số 1 trong N lần thí nghiệm chia cho N.
- Rõ ràng, $\mathbf{y}_N$ là một biến ngẫu nhiên có không gian mẫu là {0, 1/N, 2/N, ..., 1}.
- Định nghĩa $\mathbf{x}^{(n)}$ là biến ngẫu nhiên tương ứng với kết quả của lần thí nghiệm thứ n, chúng ta có

$$\mathrm{y}_N = \frac{1}{N}\sum_{n=1}^{N} \mathrm{x}^{(n)}$$

Luật yếu của số lớn (tt)

$$\overline{y}_N = \frac{1}{N}\sum_{n=1}^{N} E\left(x^{(n)}\right) = \frac{1}{N}\sum_{n=1}^{N} \overline{x} = \overline{x}$$

$$\delta_y^2 = E\left(\left(y_N - \overline{y}_N\right)^2\right) = E\left(\left[\frac{1}{N}\sum_{n=1}^{N} x^{(n)} - \overline{x}\right]^2\right)$$

$$= E\left(\left(\frac{1}{N}\left[\sum_{n=1}^{N} x^{(n)} - N\overline{x}\right]\right)^2\right) = \frac{1}{N^2} E\left(\left[\sum_{n=1}^{N}\left(x^{(n)} - \overline{x}\right)\right]^2\right)$$

$$= \frac{1}{N^2}\sum_{n=1}^{N} E\left(\left(x^{(n)} - \overline{x}\right)^2\right) = \frac{1}{N^2} N\delta_x^2 = \frac{\delta_x^2}{N}$$

Luật yếu của số lớn (tt)

- Đối với một số nguyên dương tuỳ ý ε, theo bất đẳng thức Chebyshev chúng ta có

$$P\left(\left|\,\mathrm{y}_N - \overline{\mathrm{y}}_N\,\right| \geq \varepsilon\right) \leq \frac{\delta_{\mathrm{y}}^2}{\varepsilon^2}$$

  từ đây chúng ta dẫn ra được luật yếu của số lớn

$$P\left(\left\|\left[\frac{1}{N}\sum_{n=1}^{N}\mathrm{x}^{(n)}\right] - \overline{\mathrm{x}}\right| \geq \varepsilon\right) \leq \frac{\delta_{\mathrm{x}}^2}{N\varepsilon^2}$$

- Chú ý rằng vế phải tiến tới 0 khi N tiến ra vô cùng.
- Luật yếu của số lớn vì vậy khẳng đinh rằng trị trung bình mẫu của **x** tiếp cận trị trung bình thống kê với xác suất cao khi $N \to \infty$.

Tập lồi

- Trong không gian Ơclit, một tập S được gọi là lồi (convex cap ($\cap$)) nếu đối với một cặp điểm P_1, P_2 thuộc S thì mọi điểm thuộc đoạn P_1P_2 cũng thuộc S.

(a) (b)

- Nếu $P_1 = (x_1, x_2, ..., x_n)$ và $P_2 = (y_1, y_2, ..., y_n)$ là các điểm trong không gian Ơclit n chiều, thì đoạn thẳng nối chúng được biểu diễn bằng tập các điểm P, trong đó

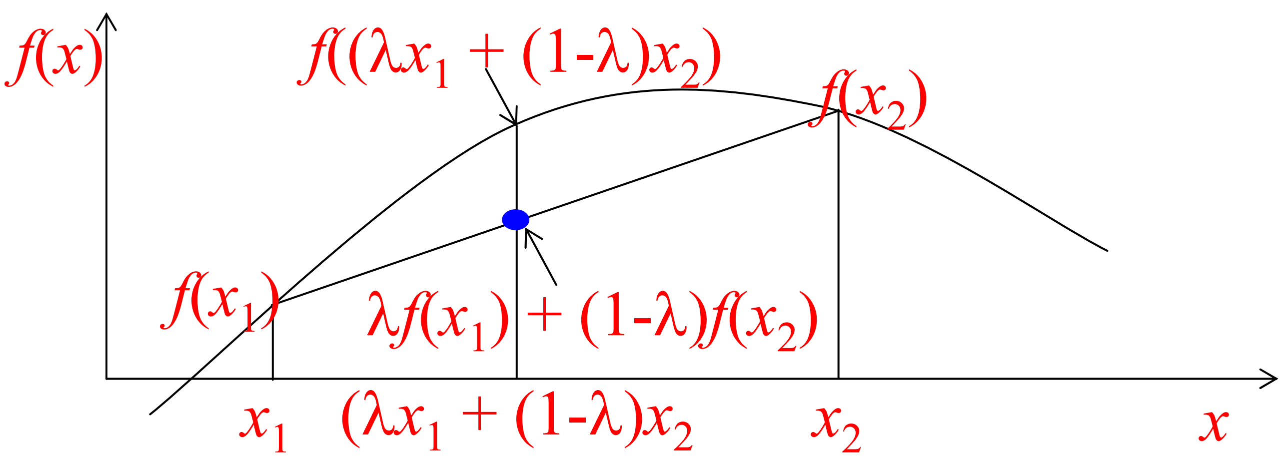
$$P = \lambda P_1 + (1-\lambda)P_2$$

$$= (\lambda x_1 + (1-\lambda)y_1, \lambda x_2 + (1-\lambda)y_2, ..., \lambda x_n + (1-\lambda)y_n) \text{ và } \lambda \in [0, 1].$$

Hàm lồi

- Một ví dụ quan trọng của tập lồi là tập tất cả các điểm (p_1, p_2, ..., p_n) trong đó (p_1, p_2, ..., p_n) là một sự phân bố xác suất (tức là các $p_i \in [0, 1]$ và $\Sigma p_i = 1$).
- Một hàm thực $f(P)$, được định nghĩa trên tập lồi S, được gọi là lồi nếu $\forall$cặp điểm P_1, $P_2 \in S$, và $\forall\ \lambda \in [0, 1]$ bất đẳng thức sau đây đúng:

$$f(\lambda P_1 + (1-\lambda)P_2) \geq \lambda f(P_1) + (1-\lambda)f(P_2)$$



Định lý, bất đẳng thức Jensen

- Nếu $\lambda_1, ..., \lambda_N$ là các số không âm có tổng bằng 1 thì đối với mọi tập điểm $P_1, ..., P_N$ trong miền xác định của hàm lồi $f(P)$ bất đẳng thức sau đây đúng

$$f\left(\sum_{n=1}^{N} \lambda_n P_n\right) \geq \sum_{n=1}^{N} \lambda_n f(P_n)$$

- Cho biến ngẫu nhiên x lấy các giá trị $x_1, ..., x_n$ với các xác suất $p_1, ..., p_n$. Cho $f(x)$ là một hàm lồi có miền xác định chứa $x_1, ..., x_n$. Chúng ta có $E(\mathrm{x}) = \sum_i p_i x_i$ và $E(f(\mathrm{x})) = \sum_i p_i f(x_i)$.
- Áp dụng định lý trên chúng ta có

$$f(E(\mathrm{x})) \geq E(f(\mathrm{x}))$$

Đây được gọi là bất đẳng thức Jensen.

Bài 4 Lượng tin

4.1 Lượng tin

4.2 Lượng tin trung bình

Vấn đề cơ bản của truyền thông là việc tái sinh tại một điểm hoặc chính xác hoặc gần đúng một thông báo được chọn tại một điểm khác.

(Claude Shannon 1948)

Lượng tin

- Lượng tin (measure of information) dùng để so sánh định lượng các tin tức với nhau.
- **Một tin đối với người nhận đều mang hai nội dung, một là độ bất ngờ của tin, hai là ý nghĩa của tin**.
- Khía cạnh ngữ nghĩa chỉ có ý nghĩa đối với con người.
- Khía cạnh quan trọng nằm ở chỗ tin thật sự là một cái được chọn từ một **tập các tin** (tập các khả năng) có thể.
- Nếu số tin trong tập tin càng nhiều thì sẽ mang lại một "lượng tin" càng lớn khi nhận được một tin (giả sử các tin là bình đẳng như nhau về khả năng xuất hiện).
- Để sự truyền tin đạt hiệu quả cao chúng ta không thể đối đãi các tin như nhau nếu chúng xuất hiện ít nhiều khác nhau.

Lượng tin

- Xét một tin *x* có xác suất xuất hiện là *p*(*x*), thì chúng ta có thể xem tin này như là một tin trong một tập có 1/*p*(*x*) tin với các tin có xác suất xuất hiện như nhau.
- Nếu *p*(*x*) càng nhỏ thì 1/*p*(*x*) càng lớn và vì vậy "lượng tin" khi nhận được tin này cũng sẽ càng lớn.
- Vậy "lượng tin" của một tin tỉ lệ thuận với số khả năng của một tin và tỉ lệ nghịch với xác suất xuất hiện của tin đó.
- Xác suất xuất hiện của một tin tỉ lệ nghịch với độ bất ngờ khi nhận được một tin.

"lượng tin" ↑ số khả năng ↑ độ bất ngờ ↓ xác suất

- Một tin có xác suất xuất hiện càng nhỏ thì có độ bất ngờ càng lớn và vì vậy có lượng tin càng lớn.

Lượng tin (tt)

- Xét một nguồn $A = \{a_1, a_2,\ldots, a_m\}$ với các xác suất xuất hiện là $p(a_i)$ $i = 1, ..., m$.
- Kí hiệu lượng tin trong mỗi tin a_i là $I(a_i)$. Vậy hàm f dùng để biểu thị lượng tin phải thoả mãn những điều kiện gì?
- Phản ánh được các tính chất thống kê của tin tức.
  - Ví dụ có hai nguồn K, L với số tin tương ứng là k, l (giả thuyết đều là đẳng xác suất). Nếu $k > l$, thì độ bất ngờ khi nhận một tin bất kỳ của nguồn K phải lớn hơn độ bất ngờ khi nhận một tin bất kỳ của nguồn L, vậy $f(k) > f(l)$
- Hợp lý trong tính toán.
  - Giả thiết hai nguồn độc lập K và L với số tin tương ứng là k và l. Cho việc nhận một cặp k_i và l_j bất kỳ đồng thời là một tin của nguồn hỗn hợp KL. Số cặp k_il_j mà nguồn này có là $k*l$.

Lượng tin (tt)

- Độ bất ngờ khi nhận được một cặp như vậy phải bằng tổng lượng tin của khi nhận được k_i và l_j. Vì vậy chúng ta phải có:

$$f(kl) = f(k) + f(l)$$

- Khi nguồn chỉ có một tin, lượng tin chứa trong tin duy nhất đó phải bằng không.

$$f(1) = 0$$

- Định nghĩa
  - Lượng đo thông tin của một tin được đo bằng logarit của độ bất ngờ của tin hay nghịch đảo xác suất xuất hiện của tin đó.

$$I(x) = \log\frac{1}{p(x)} = -\log p(x)$$

Lượng tin (tt)

- Lượng tin chứa trong một dãy $x = a_1 a_2 \ldots a_n$ với $a_i \in A$ là

$$I(x) = \log \frac{1}{p(x)} = -\sum_{i=1}^{n} \log p(a_i)$$

- Trong trường hợp m kí hiệu của nguồn đẳng xác suất với nhau tức $p(a_i) = 1/m$ thì

$$I(a_i) = \log \frac{1}{p(a_i)} = \log m$$

Nếu $x = a_1 a_2 \ldots a_n$ với $a_i \in A$

$$I(x) = n \log m$$

Lượng tin trung bình

- Đơn vị của lượng tin
  - Nếu cơ số là 2 thì đơn vị là bits (cho các kí số nhị phân); nếu cơ số là *e* thì đơn vị là nats (cho đơn vị tự nhiên), nếu cơ số là 10 thì đơn vị là Hartley.
- Định nghĩa
  - Lượng tin trung bình của một nguồn tin *A* là lượng tin trung bình chứa trong một kí hiệu bất kỳ của nguồn tin. Nó thường được kí hiệu là $I(A)$ và được tính bằng công thức sau

$$I(A) = \sum_{a_i \in A} p(a_i) I(a_i) = -\sum_{a_i \in A} p(a_i) \log p(a_i)$$

Ví dụ

- Cho một nguồn tin U bao gồm 8 tin $U = \{u_0, u_1, u_2, u_3, u_4, u_5, u_6, u_7\}$, với các xác suất xuất hiện như sau:

$p(u_0)$	$p(u_1)$	$p(u_2)$	$p(u_3)$	$p(u_4)$	$p(u_5)$	$p(u_6)$	$p(u_7)$
1/4	1/4	1/8	1/8	1/16	1/16	1/16	1/16

  Hãy cho biết lượng tin riêng của mỗi tin và lượng tin trung bình của nguồn này trong đơn vị bits.

- Giải
  - Lượng tin riêng của mỗi tin là

$I(u_0)$	$I(u_1)$	$I(u_2)$	$I(u_3)$	$I(u_4)$	$I(u_5)$	$I(u_6)$	$I(u_7)$
2	2	3	3	4	4	4	4

Ví dụ (tt)

- Lượng tin trung bình của nguồn là

  $I(U) = (1/4) \times 2 + (1/4) \times 2 + (1/8) \times 3 + (1/8) \times 3 + (1/16) \times 4 + (1/16) \times 4 + (1/16) \times 4 + (1/16) \times 4 = 2{,}75$ bits.
- Điều này nói lên một ý nghĩa quan trọng rằng, chúng ta có thể biểu diễn mỗi tin trong nguồn *U* bằng một chuỗi có chiều dài trung bình là 2,75 bits. Nó sẽ tốt hơn so với trong trường hợp chúng ta không chú ý đến cấu trúc thông kê của nguồn. Lúc đó chúng ta sẽ biểu diễn mỗi tin trong 8 tin của nguồn bằng các chuỗi có chiều dài là 3 bits.