

BỘ GIÁO DỤC VÀ ĐÀO TẠO  
TRƯỜNG ĐẠI HỌC MỎ-ĐỊA CHẤT

BÁO CÁO HỌC THUẬT

MỘT SỐ KHOẢNG CÁCH  
GIỮA CÁC BIẾN NGẪU NHIÊN  
(Phần 1)

ThS. Nguyễn Thu Hằng

HÀ NỘI, 6/2026

BỘ GIÁO DỤC VÀ ĐÀO TẠO  
TRƯỜNG ĐẠI HỌC MỎ-ĐỊA CHẤT

BÁO CÁO HỌC THUẬT

MỘT SỐ KHOẢNG CÁCH  
GIỮA CÁC BIẾN NGẪU NHIÊN  
(Phần 1)

Xác nhận của Bộ môn Toán

HÀ NỘI, 6/2026

# Mục lục

<b>1</b>	<b>Khoảng cách biến phân toàn phần và một số tính chất cơ bản</b>	<b>4</b>
1.1	Định nghĩa và tính chất . . . . .	4
1.2	Các ví dụ . . . . .	8
<b>2</b>	<b>Mối liên hệ của khoảng cách biến phân toàn phần với các khoảng cách khác</b>	<b>9</b>
2.1	Sự hội tụ yếu . . . . .	9
2.2	Khoảng cách Kolmogorov . . . . .	10
2.3	Khoảng cách entropy . . . . .	11

# MỞ ĐẦU

Ước lượng mức độ gần hay đo sự khác biệt giữa các phân phối xác suất là một trong những bài toán cơ bản, quan trọng của lý thuyết xác suất và thống kê toán học. Nó không chỉ cho chúng ta sự so sánh giữa các phân phối, giữa mô hình và hiện thực mà còn mang lại nhiều thông tin về sự hội tụ, ước tính sai số, độ rủi ro và tối ưu hóa trong thuật toán. Một số khoảng cách được sử dụng phổ biến như khoảng cách Kolmogorov, khoảng cách Wasserstein, khoảng cách biến phân toàn phần,  $L^p$  – khoảng cách... Trong đó, khoảng cách biến phân toàn phần (total variation distance viết tắt là  $d_{TV}$ ) là một trong những khoảng cách cơ bản nhất, dùng để đo khoảng cách giữa hai phân phối, được tính bằng

$$d_{TV}(P, Q) = \sup_{A \in \mathcal{F}} |P(A) - Q(A)|,$$

trong đó  $P, Q$  là hai độ đo trên không gian đo  $(\Omega, \mathcal{F})$ . Có thể hiểu rằng, khoảng cách biến phân toàn phần là độ lệch lớn nhất lấy trên các sự kiện của hai phân phối và thỏa mãn nhiều tính chất như tính bị chặn, là một metric đủ, bất biến đối với song ánh và thỏa mãn bất đẳng thức về xử lý thức dữ liệu (Data Processing Inequality)... Chính vì vậy, khoảng cách biến phân toàn phần có nhiều ứng dụng trong cả lý thuyết và thực hành, được sử dụng rộng rãi như một thước đo chính cho sự khác biệt giữa hai phân phối, được dùng trong nhiều lĩnh vực bao gồm xác suất thống kê, học máy và lý thuyết thông tin. Chẳng hạn như kiểm định giả thiết về phân phối xác suất [1], so sánh các thuật toán trong phép đếm [2], ước lượng mật độ, mã hóa xác suất [3]... Một số phương pháp nghiên cứu, ước lượng khoảng cách biến phân toàn phần thường dùng như tính toán giải tích đối với họ phân phối quen thuộc, dùng phương pháp Stein hay nghiên cứu gián tiếp thông qua các chặn trên như bất đẳng thức Pinsker,...

Mục tiêu của báo cáo này là tìm hiểu một số kiến thức cơ bản về khoảng cách biến phân toàn phần chẳng hạn như định nghĩa, một số tính chất của khoảng cách biến phân toàn phần, cách tính khoảng cách biến phân toàn phần với những phân phối thường gặp và một số liên hệ của khoảng cách biến phân toàn phần với các khoảng cách thường gặp khác.

Nội dung báo cáo gồm hai phần:

- Phần 1 trình bày tóm tắt một số kiến thức cơ bản về khoảng cách biến phân toàn phần như định nghĩa, một số tính chất của khoảng cách biến phân toàn phần và các ví dụ tính toán khoảng cách biến phân toàn phần với những phân phối thường gặp.
- Phần 2 trình bày mối liên hệ giữa khoảng cách biến phân toàn phần với một số khoảng cách khác như  $L^1$ —khoảng cách, khoảng cách Kolmogorov, khoảng cách entropy.

# NỘI DUNG BÁO CÁO

## 1 Khoảng cách biến phân toàn phần và một số tính chất cơ bản

### 1.1 Định nghĩa và tính chất

**Định nghĩa 1.1.** Xét không gian đo  $(\Omega, \mathcal{F})$  và  $P, Q$  là hai độ đo trên  $(\Omega, \mathcal{F})$ . Khoảng cách biến phân toàn phần ( $d_{TV}$ ) giữa  $P, Q$  được định nghĩa bởi

$$d_{TV}(P, Q) = \sup_{A \in \mathcal{F}} |P(A) - Q(A)|.$$

Nói cách khác, nếu  $X, Y$  là hai biến ngẫu nhiên thì

$$d_{TV}(X, Y) = \sup_{D \in \mathcal{B}(\mathbb{R})} |P(X \in D) - P(Y \in D)|.$$

Từ định nghĩa, có thể suy ra một số tính chất của  $d_{TV}$  như sau:

1.  $d_{TV}$  là một Metric đủ.
2.  $0 \leq d_{TV}(P, Q) \leq 1$ . Hơn nữa,  $d_{TV} = 0$  khi và chỉ khi  $P = Q$  và  $d_{TV} = 1$  khi và chỉ khi tồn tại  $A \in \mathcal{F}$  sao cho  $P(A) = 1, Q(A) = 0$  hoặc  $P(A) = 0, Q(A) = 1$ .
3. Nếu  $X, Y$  là hai biến ngẫu nhiên thì  $d_{TV}(X, Y) \leq P(X \neq Y)$ .

Ngoài ra, tính chất 4,5 trong Mệnh đề sau cho phép ta tính toán  $d_{TV}$ .

**Mệnh đề 1.2.** 4. Nếu  $X, Y$  là các biến ngẫu nhiên liên tục có hàm mật độ lần lượt là  $f(x)$  và  $g(x)$  thì

$$d_{TV}(X, Y) = \frac{1}{2} \int_{\mathbb{R}} |f(x) - g(x)| dx.$$

5. Nếu  $X, Y$  là các biến ngẫu nhiên rời rạc nhận giá trị trong  $\mathbb{Z}$  thì

$$d_{TV}(X, Y) = \frac{1}{2} \sum_{k \in \mathbb{Z}} |P(X = k) - P(Y = k)|.$$

*Chứng minh.* Đầu tiên ta chứng minh 4.

Giả sử  $B = \{x \in \mathbb{R} : f(x) > g(x)\}$ . Khi đó

$$\begin{aligned} \int_{\mathbb{R}} |f(x) - g(x)| dx &= \int_B (f(x) - g(x)) dx + \int_{\mathbb{R} \setminus B} (g(x) - f(x)) dx \\ &\leq 2 \sup_{A \in \mathcal{B}(\mathbb{R})} \left| \int_A (f(x) - g(x)) dx \right|. \quad (1) \end{aligned}$$

Ta lại có

$$\int_{\mathbb{R}} (f(x) - g(x)) dx = \int_{\mathbb{R}} f(x) dx - \int_{\mathbb{R}} g(x) dx = 0$$

Suy ra

$$\int_B (f(x) - g(x)) dx = \int_{B^c} (g(x) - f(x)) dx.$$

Với mọi  $A \in \mathcal{B}(\mathbb{R})$  ta có

$$\begin{aligned} \left| \int_A (f(x) - g(x)) dx \right| &= \max \left\{ \int_A (f(x) - g(x)) dx, \int_A (g(x) - f(x)) dx \right\} \\ &= \max \left\{ \int_{A \cap B} (f(x) - g(x)) dx + \int_{A \cap B^c} (f(x) - g(x)) dx, \right. \\ &\quad \left. \int_{A \cap B^c} (g(x) - f(x)) dx + \int_{A \cap B} (g(x) - f(x)) dx \right\} \\ &\leq \max \left\{ \int_{A \cap B} (f(x) - g(x)) dx, \int_{A \cap B^c} (g(x) - f(x)) dx \right\} \\ &\leq \max \left\{ \int_B (f(x) - g(x)) dx, \int_{B^c} (g(x) - f(x)) dx \right\} \\ &= \int_B (f(x) - g(x)) dx = \frac{1}{2} \int_{\mathbb{R}} |f(x) - g(x)| dx. \end{aligned}$$

Lấy supremum cả hai vế theo  $A \in \mathbb{R}$  ta được

$$\sup_{A \in \mathcal{B}(\mathbb{R})} \left| \int_A (f(x) - g(x)) dx \right| \leq \frac{1}{2} \int_{\mathbb{R}} |f(x) - g(x)| dx. \quad (2)$$

Từ (1) và (2) suy ra

$$d_{TV}(X, Y) = \frac{1}{2} \int_{\mathbb{R}} |f(x) - g(x)| dx.$$

Tương tự, ta chứng minh 5.

Đặt  $E = \{k \in \mathbb{Z} : P(X = k) > P(Y = k)\}$ . Ta có

$$P(X \in E) - P(Y \in E) = \frac{1}{2} \sum_{k \in \mathbb{Z}} |P(X = k) - P(Y = k)|.$$

Suy ra  $d_{TV}(X, Y) \geq \frac{1}{2} \sum_{k \in \mathbb{Z}} |P(X = k) - P(Y = k)|$ . Mặt khác,

$$\begin{aligned} & \sum_{k \in \mathbb{Z}} |P(X = k) - P(Y = k)| \\ &= \sum_{k \in E} |P(X = k) - P(Y = k)| + \sum_{k \in E^c} |P(X = k) - P(Y = k)| \\ &\leq 2 \sup_{A \subset \mathbb{Z}} \sum_{k \in A} |P(X = k) - P(Y = k)| = 2d_{TV}(X, Y). \end{aligned}$$

Suy ra

$$d_{TV}(X, Y) = \frac{1}{2} \sum_{k \in \mathbb{Z}} |P(X = k) - P(Y = k)|.$$

□

Đặt  $\mathcal{G}$  là không gian các hàm đo được  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  thỏa mãn

$$\|\varphi\|_\infty = \sup_{x \in \mathbb{R}} |\varphi(x)| \leq 1.$$

Ta có Mệnh đề sau:

**Mệnh đề 1.3.** *Cho  $X, Y$  là các biến ngẫu nhiên. Khi đó*

$$d_{TV}(X, Y) = \frac{1}{2} \sup_{\varphi \in \mathcal{G}} |E[\varphi(X)] - E[\varphi(Y)]|.$$

*Chứng minh.* Với  $B \in \mathcal{B}(\mathbb{R})$  bất kì ta có

$$\begin{aligned} P(X \in B) - P(Y \in B) &= \frac{1}{2} (E[(\mathbf{1}_B - \mathbf{1}_{B^c})(X)] - E[(\mathbf{1}_B - \mathbf{1}_{B^c})(Y)]) \\ &= \frac{1}{2} E[\varphi(X)] - E[\varphi(Y)], \end{aligned}$$

trong đó  $\varphi := \mathbf{1}_B - \mathbf{1}_{B^c}$ . Dễ thấy  $\varphi$  thỏa mãn  $\|\varphi\|_\infty \leq 1$ . Do đó

$$d_{TV}(X, Y) \leq \frac{1}{2} \sup_{\varphi \in \mathcal{G}} |E[\varphi(X)] - E[\varphi(Y)]|. \quad (3)$$



Để chứng minh chiều ngược lại, lấy  $\varphi \in \mathcal{G}$  cố định. Xét hàm tập  $B \mapsto \alpha(B) := P(X \in B) - P(Y \in B)$ . Ta có

$$|E[\varphi(X)] - E[\varphi(Y)]| \leq \int_{\mathbb{R}} |h(x)| d|\alpha|(x) \leq |\alpha|(C),$$

trong đó,  $C$  là giá của  $\varphi$ ,  $|\alpha|$  là biến phân toàn phần của  $\alpha$  và

$$|\alpha|(C) = \sup \{\alpha(B) : B \subset C\} - \inf \{\alpha(B) : B \subset C\}.$$

Ta lại có

$$\sup \{\alpha(B) : B \subset C\} \leq \sup_{B \subset C} |P(X \in B) - P(Y \in B)| \leq d_{TV}(X, Y)$$

và

$$\inf \{\alpha(B) : B \subset C\} \leq \sup_{B \subset C} |P(X \in B) - P(Y \in B)| \leq d_{TV}(X, Y).$$

Do đó,

$$|E[\varphi(X)] - E[\varphi(Y)]| \leq |\alpha|(C) \leq 2d_{TV}(X, Y).$$

Suy ra,

$$d_{TV}(X, Y) \geq \frac{1}{2} \sup_{\varphi \in \mathcal{G}} |E[\varphi(X)] - E[\varphi(Y)]|.$$

Ta được đẳng thức

$$d_{TV}(X, Y) = \frac{1}{2} \sup_{\varphi \in \mathcal{G}} |E[\varphi(X)] - E[\varphi(Y)]|.$$

□

Với  $f$  là hàm đo được bị chặn bất kì, đặt  $M = \|f\|_{\infty}$  và  $g = \frac{f}{M}$  thì  $g \in \mathcal{G}$ . Do đó, ta có

$$|E[g(X_n)] - E[g(X)]| \leq 2d_{TV}(X_n, X).$$

Suy ra

$$|E[f(X_n)] - E[f(X)]| \leq 2\|f\|_{\infty} d_{TV}(X_n, X).$$

Từ đó, nếu  $d_{TV}(X_n, X) \rightarrow 0$  khi  $n \rightarrow \infty$  thì  $E[f(X_n)] \rightarrow E[f(X)]$ .

## 1.2 Các ví dụ

**Ví dụ 1.** Nếu  $X, Y$  là các biến ngẫu nhiên có phân phối 0-1 với tham số lần lượt là  $p, q$  với  $p, q \in (0, 1)$  và  $p > q$  thì

$$d_{TV}(X, Y) = \frac{1}{2} (p - q + (1 - q) - (1 - p)) = p - q.$$

**Ví dụ 2.** Nếu  $X, Y$  là hai biến ngẫu nhiên có phân phối đều liên tục  $X \sim U[0, a]$  và  $Y \sim U[0, b]$  với  $0 < a < b$  thì

$$d_{TV}(X, Y) = \frac{1}{2} \left( \int_0^a \left( \frac{1}{a} - \frac{1}{b} \right) dx + \int_a^b \frac{1}{b} dx \right) = \frac{b - a}{b}.$$

**Ví dụ 3.** Nếu  $X$  có phân phối 0 – 1 với tham số  $p$  và  $Y$  có phân phối Poisson cũng tham số  $p$  thì

$P(Y = k) = \frac{e^{-p} p^k}{k!}$ . Mặt khác, theo công thức khai triển Maclaurin ta có

$$e^{-p} = 1 - p + \frac{e^{-c}}{2!} p^2 \geq 1 - p; 0 < c < p.$$

Suy ra

$$\begin{aligned} d_{TV}(X, Y) &= \frac{1}{2} \left[ e^{-p} - (1 - p) + p - pe^{-p} + e^{-p} \left( \frac{p^2}{2!} + \frac{p^3}{3!} + \dots \right) \right] \\ &= \frac{1}{2} \left[ e^{-p} \left( 1 + p + \frac{p^2}{2!} + \frac{p^3}{3!} + \dots \right) + 2p - 2pe^{-p} - 1 \right] \\ &= \frac{1}{2} (e^{-p} e^p + 2p - 2pe^{-p} - 1) = p(1 - e^{-p}). \end{aligned}$$

**Ví dụ 4.** Giả sử  $X, Y$  là hai biến ngẫu nhiên có phân phối chuẩn,  $X \sim N(\mu_1, 1), Y \sim N(\mu_2, 1)$  với  $\mu_2 > \mu_1$ .

Đặt  $a = \frac{\mu_1 + \mu_2}{2}$ . Ta có

$$\begin{aligned} d_{TV}(X, Y) &= \frac{1}{2} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \left| e^{-\frac{(x-\mu_1)^2}{2}} - e^{-\frac{(x-\mu_2)^2}{2}} \right| dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^a \left( e^{-\frac{(x-\mu_1)^2}{2}} - e^{-\frac{(x-\mu_2)^2}{2}} \right) dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mu_1 - \varepsilon}^{\mu_1 + \varepsilon} e^{-\frac{(x-\mu_1)^2}{2}} dx = 2\phi_0(\varepsilon). \end{aligned}$$

trong đó  $\phi_0$  là hàm Laplace và  $\varepsilon = \frac{\mu_2 - \mu_1}{2}$ .

Tổng quát hơn, nếu  $X \sim N(\mu_1, \sigma), Y \sim N(\mu_2, \sigma)$  với  $\mu_2 > \mu_1$  thì đặt  $c = \frac{\mu_1 + \mu_2}{2}$ . Khi đó

$$d_{TV}(X, Y) = P(X \leq c) - P(Y \leq c),$$

trong đó

$$P(X \leq c) = \Phi\left(\frac{c - \mu_1}{\sigma}\right) = \Phi\left(\frac{\mu_2 - \mu_1}{2\sigma}\right)$$

và

$$P(Y \leq c) = \Phi\left(\frac{c - \mu_2}{\sigma}\right) = \Phi\left(-\frac{\mu_2 - \mu_1}{2\sigma}\right).$$

Chú ý rằng  $\Phi(-x) = 1 - \Phi(x)$  nên ta được

$$d_{TV}(X, Y) = 2\Phi(a) - 1 \text{ với } a = \frac{\mu_2 - \mu_1}{2\sigma}.$$

Tổng quát,

$$d_{TV}(X, Y) = 2\Phi\left(\frac{|\mu_1 - \mu_2|}{2\sigma}\right) - 1.$$

Nếu  $\delta = \frac{|\mu_1 - \mu_2|}{2\sigma}$  nhỏ, dùng khai triển Maclaurin của  $\Phi$  ta được

$$\Phi(\delta) = \frac{1}{2} + \phi(0)\delta + o(\delta),$$

với  $\phi(x)$  là hàm mật độ của phân phối chuẩn tắc có  $\phi(0) = \frac{1}{\sqrt{2\pi}}$ . Thay lại  $\delta$  ta suy ra

$$d_{TV}(X, Y) \sim \frac{|\mu_1 - \mu_2|}{\sqrt{2\pi}\sigma}, \text{ khi } |\mu_1 - \mu_2| \rightarrow 0.$$

## 2 Mối liên hệ của khoảng cách biến phân toàn phần với các khoảng cách khác

### 2.1 Sự hội tụ yếu

Từ định nghĩa của  $d_{TV}$  và từ Mệnh đề 1.3 trong chương 1 ta có kết quả sau đây về mối liên hệ giữa  $d_{TV}$  sự hội tụ yếu.

**Mệnh đề 2.1.** 1. Cho dãy biến ngẫu nhiên liên tục  $\{F_i\}_{i=1}^{\infty}$  và biến ngẫu nhiên liên tục  $F$ . Khi đó,  $d_{TV}(F_n, F) \xrightarrow{n \rightarrow \infty} 0$  tương đương  $F_n \xrightarrow{L_1} F$ .

2. Cho dãy biến ngẫu nhiên  $\{F_i\}_{i=1}^{\infty}$  và biến ngẫu nhiên  $F$ . Khi đó,  $d_{TV}(F_n, F) \xrightarrow{n \rightarrow \infty} 0$  suy ra  $F_n \xrightarrow{law} F$ .

Chú ý rằng, điều ngược lại của 2. chưa chắc đã đúng. Nếu  $F_n \xrightarrow{law} F$  thì chưa chắc  $d_{TV}(F_n, F) \xrightarrow{0}$ . Thật vậy, Lấy  $F_n$  là dãy biến ngẫu nhiên liên tục có hàm mật độ  $f_n(x) = \frac{2}{\pi} \cos^2(nx) \mathbb{1}_{[0, \pi]}(x)$  và  $F$  là biến ngẫu nhiên có phân phối đều liên tục trên  $[0, \pi]$ . Khi đó,  $F_n \xrightarrow{law} F$ . Tuy nhiên

$$\begin{aligned} d_{TV}(F_n, F) &= \frac{1}{2} \int_0^\pi \frac{1}{\pi} |2 \cos^2(nx) - 1| dx = \frac{1}{2\pi} \int_0^\pi |\cos(2nx)| dx \\ &= \frac{4n}{2\pi} \int_0^{\frac{\pi}{4n}} \cos(2nx) dx = 1 > 0, \forall n. \end{aligned}$$

## 2.2 Khoảng cách Kolmogorov

**Định nghĩa 2.2.** *Khoảng cách Kolmogorov giữa hai biến ngẫu nhiên  $X, Y$  được định nghĩa bởi*

$$d_K(X, Y) := \sup_{x \in \mathbb{R}} |P(X \leq x) - P(Y \leq x)|.$$

Do các tập  $(-\infty, x]$  chỉ là một lớp con của mọi tập đo được nên

$$d_K(X, Y) \leq d_{TV}(X, Y).$$

Tiếp theo, ta lấy ví dụ cụ thể để thấy sự khác biệt giữa khoảng cách biến phân toàn phần và khoảng cách Kolmogorov

**Ví dụ 5.** Xét  $X \sim N(0, 1)$  và  $Y \sim N(a, 1)$ . Theo tính toán trong mục 1, ta có

$$d_{TV}(X, Y) = 2\Phi\left(\frac{a}{2}\right) - 1.$$

Còn khoảng cách Kolmogorov

$$d_K(X, Y) = \sup_x |\Phi(x) - \Phi(x - a)|.$$

Xét hàm  $g(x) = \Phi(x) - \Phi(x - a)$  có  $g'(x) = \phi(x) - \phi(x - a) = 0$  khi  $x = \frac{a}{2}$  và  $g(x)$  đạt cực đại tại  $x = \frac{a}{2}$ . Do đó

$$d_K(X, Y) = \Phi\left(\frac{a}{2}\right) - \Phi\left(\frac{-a}{2}\right) = 2\Phi\left(\frac{a}{2}\right) - 1.$$

Trong trường hợp này  $d_{TV}(X, Y) = d_K(X, Y)$ .

Tiếp theo, xét  $X \sim U[0, 1]$  và  $Y_n$  có phân phối đều rời rạc trên tập  $\{\frac{1}{n}, \frac{2}{n}, \dots, \frac{n}{n}\}$ . Khi đó

$$d_K(X, Y_n) = \frac{1}{n}$$

và do đó  $d_K(X, Y_n) \rightarrow 0$  khi  $n \rightarrow \infty$  trong khi xét tập  $A = \{\frac{1}{n}, \frac{2}{n}, \dots, \frac{n}{n}\}$  thì  $P(Y_n \in A) = 1$  và  $P(X \in A) = 0$ . Suy ra

$$d_{TV}(X, Y_n) \geq 1.$$

Do đó

$$d_{TV}(X, Y_n) = 1.$$

### 2.3 Khoảng cách entropy

**Định nghĩa 2.3.** 1. Cho  $X, Y$  là hai biến ngẫu nhiên rời rạc với  $P(X = x_i) = p_i$  và  $P(Y = y_i) = q_i$  thì khoảng cách entropy (độ lệch entropy tương đối) từ  $X$  tới  $Y$  định nghĩa bởi

$$D_{KL}(X\|Y) = \sum_i p_i \log \frac{p_i}{q_i},$$

với quy ước  $0 \log 0 = 0$ .

2. Cho  $X, Y$  là hai biến ngẫu nhiên liên tục có hàm mật độ lần lượt là  $p(x)$  và  $q(x)$ . Khi đó khoảng cách entropy (độ lệch entropy tương đối) từ  $X$  tới  $Y$  cho bởi

$$D_{KL}(X\|Y) = \int_{\mathbb{R}} p(x) \log \left( \frac{p(x)}{q(x)} \right) dx.$$

**Ví dụ 6.** Giả sử  $X \sim N(\mu_1, \sigma^2)$  và  $Y \sim N(\mu_2, \sigma^2)$ . Ta có

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu_1)^2}{2\sigma^2}}, \quad q(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu_2)^2}{2\sigma^2}}$$

$$\begin{aligned} \log \left( \frac{p(x)}{q(x)} \right) &= -\frac{(x-\mu_1)^2}{2\sigma^2} + \frac{(x-\mu_2)^2}{2\sigma^2} \\ &= \frac{2(\mu_1 - \mu_2)x + (\mu_2^2 - \mu_1^2)}{2\sigma^2} \end{aligned}$$

Lấy kì vọng 2 vế với chú ý rằng  $E[X] = \mu_1$  ta được

$$\begin{aligned} D_{KL}(X\|Y) &= \frac{2(\mu_1 - \mu_2)\mu_1 + (\mu_2^2 - \mu_1^2)}{2\sigma^2} \\ &= \frac{\mu_1^2 - 2\mu_1\mu_2 + \mu_2^2}{2\sigma^2} = \frac{(\mu_1 - \mu_2)^2}{2\sigma^2}. \end{aligned}$$

**Mệnh đề 2.4.**

$$d_{TV}(X, Y) \leq \sqrt{\frac{1}{2} D_{KL}(X \| Y)}. \quad (4)$$

*Chứng minh.* Báo cáo trình bày chứng minh của Clément L. Canonne trong [4].

Xét  $X$  và  $Y$  là các phân phối 0 – 1 với tham số lần lượt là  $p, q$ . Theo ví dụ 1 của mục 1 ta có

$$d_{TV}(X, Y) = |p - q|$$

và

$$D_{KL}(X \| Y) = p \log \frac{p}{q} + (1 - p) \log \frac{1 - p}{1 - q}.$$

Xét hàm  $f(x) : (0, 1) \rightarrow \mathbb{R}$  cho bởi  $f(x) = p \log x + (1 - p) \log(1 - x)$  thì ta có thể viết

$$\begin{aligned} f(p) - f(q) &= \int_q^p f'(x) dx = \int_q^p \frac{p - x}{x(1 - x)} dx \\ &\leq 4 \int_q^p (p - x) dx = 4 \cdot \frac{1}{2} (p - q)^2. \end{aligned}$$

Ở trên, ta đã sử dụng bất đẳng thức  $x(1 - x) \leq 1/4$  với  $x \in (0, 1)$ . Từ đó suy ra

$$f(p) - f(q) = p \log \frac{p}{q} + (1 - p) \log \frac{1 - p}{1 - q} \leq 2(p - q)^2.$$

Nên bất đẳng thức 4 đúng trong trường hợp này. Trong trường hợp tổng quát, giả sử  $X, Y$  là biến ngẫu nhiên với trên miền  $\Omega$  tùy ý tương ứng với độ đo xác suất  $P, Q$  trên  $\Omega$ . Cố định tập con đo được  $S \subset \Omega$ . Hai biến ngẫu nhiên  $X' = \mathbb{1}_S(X)$  và  $Y' = \mathbb{1}_S(Y)$  tương ứng là các phân phối 0 – 1 với tham số  $p' = P(S)$  và  $q' = Q(S)$ . Do đó

$$2(P(S) - Q(S))^2 = 2d_{TV}(X', Y')^2 \leq D_{KL}(X', Y') \leq D_{KL}(X \| Y),$$

trong đó bất đẳng thức thứ 2 là nhờ bất đẳng thức data processing. Lấy sup trên các tập  $S$  ta được

$$2d_{TV}(X, Y)^2 = 2 \sup_{S \in \Omega} 2(P(S) - Q(S))^2 \leq D_{KL}(X \| Y).$$

□

## KẾT LUẬN

Báo cáo đã tổng kết một số kiến thức cơ bản về khoảng cách biến phân toàn phần giữa các biến ngẫu nhiên. Từ đó, tính toán khoảng cách biến phân toàn phần giữa một số biến ngẫu nhiên cơ bản. Báo cáo cũng tổng hợp một số mối quan hệ giữa khoảng cách biến phân toàn phần và một số khoảng cách thường dùng khác. Kết quả của báo cáo nhằm mục đích làm tài liệu tham khảo chuyên ngành xác suất thống kê và hỗ trợ giảng dạy cho sinh viên ngành toán.

## TÀI LIỆU THAM KHẢO

- [1] Clément Canonne. Topics and techniques in distribution testing. Preprint at <https://ccanonne.github.io/files/misc/main-survey-fnt.pdf>, 2022.
- [2] Mark Jerrum. Counting, sampling and integrating: algorithms and complexity. In *Lectures in Mathematics – ETH Zürich*. Birkhauser, Berlin, 2003.
- [3] Shafi Goldwasser and Silvio Micali. Probabilistic encryption. *J. Comput. Syst. Sci.*,28(2):270–299, 1984.
- [4] Clément L. Canonne. A short note on an inequality between KL and TV, <https://arxiv.org/abs/2202.07198v2>