CƠ SỞ TOÁN
CHO MACHINE LEARNING

Chương 5. Một số vấn đề về tối ưu hóa

Tối ưu hóa là một lĩnh vực quan trọng của toán học có ảnh hưởng đến hầu hết mọi lĩnh vực khoa học – công nghệ và kinh tế – xã hội, bởi việc tìm giải pháp tối ưu cho một vấn đề nào đó chiếm một vai trò hết sức quan trọng. Machine Learning và tối ưu có quan hệ mật thiết với nhau. Trong tối ưu thì tối ưu lồi là quan trọng nhất.

5. 1. Bài toán tối ưu tổng quát và phân loại

5.1.1. Bài toán tối ưu tổng quát :Xét $\mathrm{f}\colon\ \mathrm{X} \to \mathbb{R}$, X là một không gian nào đó. Tìm $\mathrm{x}^* \in \mathrm{D} \subset \mathrm{X}\colon \mathrm{f}(\mathrm{x}^*) = (\max)\min\{\mathrm{f}(\mathrm{x})\colon \mathrm{x} \in \mathrm{D}\}$, hay: $\begin{cases} \mathrm{f}(\mathrm{x}) \to \max\ (\min) \\ \quad\quad \mathrm{x} \in \mathrm{D} \end{cases}$ (1)

Hay: Tìm: $\mathrm{x}^* = \arg\max\limits_{\mathrm{x}\in\mathrm{D}} \mathrm{f}(\mathrm{x})(\text{hoặc: } \mathrm{x}^* = \arg\min\limits_{\mathrm{x}\in\mathrm{D}} \mathrm{f}(\mathrm{x}))$

- Gọi $\mathrm{f}(\mathrm{x})$ là hàm mục tiêu, tập xác định X kí hiệu là *dom*(f) (*dom*: *domain*) đ/k: $\mathrm{x} \in \mathrm{D}$ là đ/kiện ràng buộc, D là miền ràng buộc hay tập p.án chấp nhận được, x^* là p.án (hay nghiệm) tối ưu toàn cục, $\mathrm{f}(\mathrm{x}^*)$ là giá trị tối ưu toàn cục.
- Nếu D = X thì (1) được gọi là bài toán tối ưu không ràng buộc
- Nếu $\mathrm{X} \setminus \mathrm{D} \neq \emptyset$ thì (1) được gọi là bài toán tối ưu có ràng buộc
- Nếu $\mathrm{x}^{(*)} = \arg\max\limits_{\mathrm{x}\in\mathrm{C}} \mathrm{f}(\mathrm{x})(\text{hoặc: } \mathrm{x}^{(*)} = \arg\min\limits_{\mathrm{x}\in\mathrm{C}} f(\mathrm{x}))\ (C \subset D)$ thì $\mathrm{X}^{(*)}$ gọi là p.án tối ưu địa phương của (1)
- Điều kiện ràng buộc $\mathrm{x} \in \mathrm{D}$ thường xuất hiện ở các dạng sau:

* Ràng buộc đẳng thức: F(x) = 0 (F là hàm xác định trên X)

* Ràng buộc bất đẳng thức: $\mathrm{f_i}(\mathrm{x}) \geq (\leq) 0, \mathrm{i} = \overline{1, \mathrm{m}}\ (\mathrm{f_i}\colon \mathrm{X} \to \mathbb{R}, \mathrm{i} = \overline{1, \mathrm{n}}\ )$

* Ràng buộc $\mathrm{x} \in \mathrm{D}$, với D là tập hợp cho trước trong X.

Ví dụ. (Bài toán Nhà xuất bản) Một NXB nhận được 2 đơn hàng về một cuốn sách: 800 cuốn tới tỉnh A, 600 cuốn tới tỉnh B. NXB hiện có 1000 cuốn ở địa điểm C và 800 cuốn ở địa điểm D. Đơn giá chuyển một cuốn từ C tới A là 50 ngàn đồng, tới B là 80 ngàn đồng. Đơn giá chuyển một cuốn từ D tới A là 100 ngàn đồng, tới B là 40 ngàn đồng. Tìm phương án vận chuyển 2 đơn hàng nói trên sao cho tổng chi phí chuyển phát là ít nhất.

5.1.2. Phân loại các bài toán tối ưu (hay bài toán quy hoach toán học): Quy hoạch tuyến tính, Quy hoạch phi tuyến, Quy hoạch rời rạc, Quy hoạch động, Quy hoạch đa mục tiêu, Quy hoạch ngẫu nhiên.

5.1.3. Ứng dụng: Nhiều bài toán thực tế có thể được giải quyết bằng phương pháp tối ưu toán học. Điều quan trọng là từ bài toán thực tế cần xây dựng một mh tối ưu để áp dụng các phương pháp tối ưu toán học và quy trình tính toán thích hợp để tìm ra lời giải. Các bước thực hiện là: **(1)** Khảo sát bài toán thực tế và vấn đề cần giải quyết; **(2)** Lựa chọn các biến quyết định, biểu diễn các điều kiện ràng buộc thông qua các ràng buộc đ/v các biến, biểu diễn hàm mục tiêu theo các biến và các định mục tiêu (min/max) và chỉ ra mh toán học; **(3)** Chọn phương pháp giải thích hợp; **(4)** Xác định quy trình giải/thuật toán; **(5)** Kiểm tra lại quy trình giải (p.pháp lựa chọn, thuật toán,...) để nếu có sai sót thì điều chỉnh; **(6)** Kiểm tra sự phù hợp giữa lời giải với thực tế, so sánh hiệu quả của phương án tìm được với các phương án đưa ra trước đây.

Với ví dụ trên, ta thiết lập MH toán học cho bài toán như sau:
- *Các biến cần thiết*: x_1 là số lượng cuốn sách cần chuyển phát từ C đến A, x_2 là số lượng cuốn sách cần chuyển phát từ C đến B, x_3 là số lượng cuốn sách cần chuyển phát từ D đến A, x_4 là số lượng cuốn sách cần chuyển phát từ D đến B. Mỗi phương án chuyển phát là: $x = (x_1, x_2, x_3, x_4) \in \mathbb{R}^4$
- Hàm mục tiêu: Tổng chi phí chuyển phát là:
$f(x) = 50.x_1 + 80.x_2 + 100.x_3 + 40.x_4$
- Điều kiện ràng buộc:
* Ràng buộc hàm: Số cuốn chuyển phát đến A: $x_1 + x_3 = 800$
Số cuốn chuyển phát đến B: $x_2 + x_4 = 600$
Số cuốn chuyển phát từ C: $x_1 + x_2 \leq 1000$
Số cuốn chuyển phát từ D: $x_3 + x_4 \leq 800$
* Ràng buộc biến: $x_1, x_2, x_3, x_4 \in \mathbb{N}$
MH toán học của bài toán là: Tìm $x = (x_1, x_2, x_3, x_4)$ sao cho:

$$f(x) = 50.x_1 + 80.x_2 + 100.x_3 + 40.x_4 \to \min$$

$$\begin{cases} x_1 + x_3 = 800 \\ x_2 + x_4 = 600 \\ x_1 + x_2 \leq 1000 \\ x_3 + x_4 \leq 800 \\ x_1, x_2, x_3, x_4 \in \mathbb{N} \end{cases}$$

5.1.3. Ứng dụng: Nhiều bài toán thực tế có thể được giải quyết bằng p.pháp tối ưu toán học. Điều quan trọng là từ bài toán thực tế cần xây dựng một mh tối ưu thích hợp để áp dụng các p.pháp tối ưu toán học và quy trình tính toán thích hợp để tìm ra lời giải.

Các bước cần thiết để thực hiện là: **(1)** Khảo sát bài toán thực tế và vấn đề cần giải quyết; **(2)** Lựa chọn các biến quyết định, biểu diễn các điều kiện ràng buộc thông qua các ràng buộc đ/v các biến, biểu diễn hàm mục tiêu theo các biến và các định mục tiêu (min/max) và chỉ ra mh toán học; **(3)** Chọn phương pháp giải thích hợp; **(4)** Xác định quy trình giải/thuật toán; **(5)** Kiểm tra lại quy trình giải (p.pháp lựa chọn, thuật toán,...) để nếu có sai sót thì điều chỉnh; **(6)** Kiểm tra sự phù hợp giữa lời giải với thực tế, so sánh hiệu quả của phương án tìm được với các phương án đưa ra trước đây.

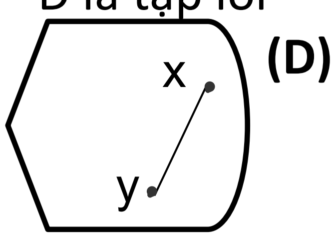
- Điểm $x^0 \in S$ được gọi là điểm cực biên của tập lồi S, nếu nó không là điểm trong của bất kỳ đoạn thẳng nào nối 2 điểm phân biệt trong S

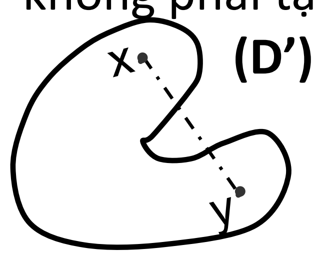
5.2. Tập lồi và hàm lồi. Trong toán tối ưu, chúng ta đặc biệt quan tâm tới những bài toán mà hàm mục tiêu là một hàm lồi, và miền ràng buộc là một tập lồi.

5.2.1. Các khái niệm về tập lồi

a. ĐN. Tập $D \subset \mathbb{R}^n$ được gọi là một tập lồi nếu mọi đọan thẳng nối 2 điểm bất kỳ của D đều nằm trọn trong D, tức là: $\forall x, y \in D$ thì: $\boldsymbol{\lambda}.x + (1-\boldsymbol{\lambda}).y \in D, \forall \boldsymbol{\lambda} \in [0,1]$

Minh họa: D là tập lồi D' không phải tập lồi





b. Các ví dụ: Cho $a, b \in \mathbb{R}^n, a = (a_1\ a_2 \dots a_n)^T, b = (b_1\ b_2 \dots b_n)^T, r > 0$
Các tập hợp sau là các tập lồi:
$D_1 = \{x = (x_1\ x_2 \dots x_n)^T \in \mathbb{R}^n\!:\ a^T.x = b\}$**:** Siêu phẳng trong $\mathbb{R}^n$ (*Hyerplane*)
$D_2 = \{x = (x_1\ x_2 \dots x_n)^T \in \mathbb{R}^n\!:\ a^T.x \le (\ge) b\}$**:** Nửa không gian trong $\mathbb{R}^n$ (*Halfspace*)
$B(a, r) = \{x = (x_1\ x_2 \dots x_n)^T \in \mathbb{R}^n\!:\ \|x - a\|_2 \le r\}$**:** Hình cầu đóng tâm a, bán kính r
c. Tổ hợp lồi, đa diện lồi. Trong $\mathbb{R}^n$ cho k điểm $x^1, x^2, \dots, x^k$ và k số không âm: $t_1, t_2, \dots, t_k$ với: $t_1 + t_2 + \cdots + t_k = 1$, khi đó:
* $x = t_1.x^1 + t_2.x^2 + \cdots + t_k.x^k$ gọi là một tổ hợp lồi của k điểm $x^1, \dots, x^k$
+ Tập tất cả các tổ hợp lồi của k điểm $x^1, x^2, \dots, x^k$ được gọi là đa diện lồi sinh bởi hệ k điểm $x^1, x^2, \dots, x^k$.
d. Chú ý. (1). Giao của một họ các tập lồi là một tập lồi; **(2).** Nếu S là một tập lồi thì S chứa mọi tổ hợp lồi của một họ điểm bất kỳ trong S; **(3).** Mỗi đa diện lồi là một tập lồi.

** Trong thực tế, tập lồi rất được quan tâm bởi tính chất ưu việt của nó: Với một căn phòng là hình lồi thì khi đặt bóng đèn ở bất kỳ điểm nào trong phòng thì mọi vị trí của căn phòng đều được chiếu sáng; trong một đất nước có bản đồ hình lồi thì hai thành phố bất kỳ đều có thể kết nối bởi một tuyến đường thẳng, nó cho phép thiết lập các tuyến đường bay nội địa là các đường bay thẳng (tiết kiệm được thời gian và chi phí cho mỗi chuyến bay),...*

5.2.2. Hàm lồi

a. ĐN. Hàm số f: $X(\subset \mathbb{R}^n) \to \mathbb{R}$ được gọi là một hàm lồi (lõm), nếu thỏa các đ/k:

a1. dom(f) = X là một tập lồi

a2. $f(t.u + (1-t).v) \le (\ge)\ t.f(u) + (1-t).f(v), \forall u, v \in X, \forall t \in [0, 1]$

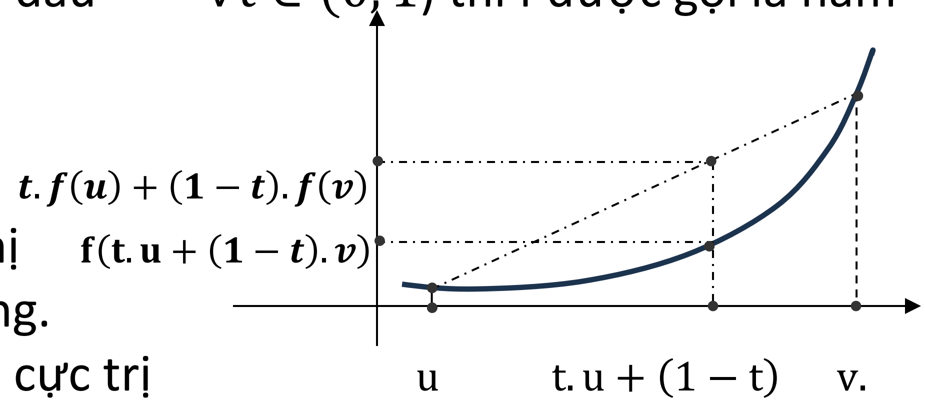
Nếu trong bất đẳng thức (a2) không có dấu "= " $\forall t \in (0, 1)$ thì f được gọi là hàm lồi (lõm) chặt.

b. Ý nghĩa hình học:

Xét hàm lồi một biến số: $y = f(x)$.

Khi đó: với mọi dây cung thì cung đồ thị tương ứng luôn nằm phía dưới dây cung.

- Nếu f(x) là hàm số lồi chặt và có điểm cực trị thì điểm cực trị đó là duy nhất và do đó là cực trị toàn cục



c. Các tính chất: Cho $f(x), f_1(x), f_{2(x)} \dots, f_m(x)$ là các hàm lồi và số thực a. Khi đó:

c1. Với a > 0 thì $a.f(x)$ là hàm lồi và với a < 0 thì $a.f(x)$ là hàm lõm.

c2. $F(x) = f_1(x) + f_2(x) + \dots + f_m(x)$ là một hàm lồi.

c3. $h(x) = \max\{f_1(x), f_{2(x)} \dots, f_m(x)\}$ là một hàm lồi.

5.2.3. Phương pháp kiểm tra tính lồi của hàm số nhờ đạo hàm:

Xét hàm lồi $f(x)$ xác định và có đạo hàm trên tập lồi $X \subset \mathbb{R}^n$

a. $f(x)$ là hàm lồi $\Leftrightarrow f(u) \ge f(v) + \nabla f(v)^T.(u - v), \forall u, v \in dom(f) = X$

$f(x)$ là hàm lồi chặt $\Leftrightarrow f(u) > f(v) + \nabla f(v)^T.(u - v), \forall u, v \in X, u \ne v$

b. $f(x)$ có đạo hàm cấp 2 là hàm lồi (hoặc lồi chặt) nếu Hesse tại mọi điểm là ma trận bán xác định dương (hoặc xác định dương), tức là:

$$\nabla^2 f(x) \geq 0 (> 0), \forall x \in \text{dom}(f)$$

Ý nghĩa hình học: Hàm f(x) là lồi khi và chỉ khi đồ thị của nó không nằm dưới mọi siêu phẳng tiếp xúc.

b. Một số ví dụ quan trọng:

b1. Một số hàm lồi một biến:

(1) y = a.x + b (a, b là các hằng số)

(2) $y = ax^2 + b.x + c$ $(a > 0)$

(3) $y = e^{\alpha.x}$ (α là hằng số bất kỳ)

(4) $y = x^\alpha, x > 0$ (α là hằng số $\notin (0,1)$)

(5) $y = x.\log x$ (hàm entropy âm)

b2. Một số hàm lồi nhiều biến:

(1) Các hàm affin:

$f(x) = a^T.x + b, x \in \mathbb{R}^n$ (a cho trước $\in \mathbb{R}^n$, b là hằng số)

$f(X) = \text{trace}(A^T.X) + b, X \in \mathbb{R}^{m\times n}$ (A cho trước $\in \mathbb{R}^{m\times n}$, b là hằng số)

(2) Các dạng toàn phương:

$$f(x) = x^T.A.x + b^T.x + c, x \in \mathbb{R}^n$$

(A cho trước $\in \mathbb{R}^{n\times n}$, b cho trước $\in \mathbb{R}^n$, c là hằng số)

(3) Các chuẩn trong các không gian định chuẩn.

b3. Một số hàm lõm: (1) f(x) = a.x + b; (2) $y = ax^2 + b.x + c\ (a < 0)$; (3) $y = x^{\alpha}, x > 0\ (\alpha \text{ là } hằng\ số\text{: } 0 \le \alpha \le 1)$; (4) $y = \log x,\ x > 0$.

b4. Các ví dụ về các hàm không lồi, không lõm: (1) $y = ax^3 + bx^2 + c.x + d, a \neq 0$ (a, b, c, d là các hằng số); (2) $z = x^2 - y^2$ (hàm hyperbolic); ...

5.3. Bài toán tối ưu lồi (Quy hoạch lồi).

5.3.1. Các khái niệm

a. ĐN: Bài toán tối ưu lồi là bài toán tối ưu có dạng:

$$x^* = \arg\min_x f(x)$$

$$\begin{cases} f_i(x) \le 0, i = \overline{1, m} \\ h_j(x) = a_j^T.x - b_j = 0, j = \overline{1, n} \end{cases}$$

Trong đó $f, f_1, \dots, f_m$ là các hàm lồi

b. Nhận xét: Trong bài toán tối ưu lồi, hàm mục tiêu được tối ưu trên một tập lồi (trong các bất đẳng thức ràng buộc: các hàm $f_1, \dots, f_m$ là các hàm lồi, trong các đẳng thức ràng buộc: các hàm h_j là các hàm affine).

5.3.2. Các tính chất của bài toán tối ưu lồi

a. *Cực trị địa phương của bài toán tối ưu lồi là cực trị toàn cục* (t/c quan trọng nhất của bài toán tối ưu lồi)

b. *Điều kiện tối ưu cho hàm mục tiêu khả vi:* Phương án $x_0 \in D$ (tập tất cả các phương án) là phương án tối ưu cho bài toán tối ưu lồi $\Leftrightarrow \nabla f(x_0)^T(x - x_0) \ge 0, \forall x \in D$.

5.3.3. Một số dạng bài toán tối ưu lồi.

a. Quy hoạch tuyến tính: Hàm mục tiêu f và các hàm ràng buộc $f_1, \dots, f_m$ đều là các hàm affine

a1. QHTT dạng tổng quát:

$$x^* = \arg\min_x (c^T.x + d)$$

$$\begin{cases} G.x \preccurlyeq h \\ A.x = b \end{cases}$$

Trong đó: $G \in \mathbb{R}^{m \times n}, A \in \mathbb{R}^{k \times n}, b \in \mathbb{R}^k, c, x \in \mathbb{R}^n, d \in \mathbb{R}$

QHTT dạng tổng quát có thể đưa về dạng chuẩn sau đây.

a2. QHTT dạng chuẩn:

$$x^* = \arg\min_x (c^T.x + d)$$

$$\begin{cases} A.x = b \\ \quad x \succcurlyeq 0 \end{cases}$$

Ví dụ 1 (Bài toán cắt vật liệu): Cần mua những thanh vật liệu có cùng độ dài L để cắt thành: n_1 đoạn có độ dài L_1, n_2 đoạn có độ dài L_2, ..., n_k đoạn có độ dài L_k ($L_i \leq L, \forall i = \overline{1, k}$). Hãy lập phương án cắt vật liệu sao cho tổng số vật liệu dư thừa là ít nhất.

- Thiết lập MH toán học: Gọi n là số cách cắt một thanh thành các đoạn con theo yêu cầu trên (để xác định n, cần lập bảng liệt kê các cách cắt cho từng bài toán cụ thể). Gọi x_j là số thanh được cắt theo cách thứ j , d_j là độ dài phần dư của một thanh được cắt theo cách thứ j, $j = \overline{1, n}$ ($d_j \leq \min\{L_1, L_2, \dots, L_k\}$), a_{ij} là số đoạn có độ dài L_i cắt ra từ một thanh được cắt theo cách thứ j, $i = \overline{1, k}$, $j = \overline{1, n}$. Khi đó mỗi $x = (x_1, x_2, \dots, x_n)^T$ là một phương án cắt và tổng độ dài của các đoạn dư thừa là $\sum_{j=1}^{n} d_j \,.\, x_j$;. MH toán học:

$$f(x) = \sum_{j=1}^{n} d_j \,.\, x_j \to \min$$

$$\begin{cases} a_{i1}\,.\,x_1 + a_{i2}\,.\,x_2 + \cdots + a_{in}\,.\,x_n = n_i, i = \overline{1, n} \\ x_j \geq 0, j = 1,2, \dots, n \end{cases}$$

Ví dụ (Bài toán vận tải): Trong một khu vực kinh tế có m điểm A1, A2,..., Am cung cấp một loại hàng với khối lượng tương ứng là S1, S2,..., Sm (tấn) tương ứng và có n điểm tiêu thụ loại hàng này là B1, B2,..., Bn với nhu cầu tương ứng là R1, R2, ..., Rn (tấn). Hãy lập kế hoạch vận chuyển hàng hóa sao cho các điểm cung giao hết hàng, nếu S1 + S2 +... + Sm ≤ R1 + R2 + ... + Rn, hoặc các điểm tiêu thụ nhận đủ hàng nếu S1 + S2 +... + Sm ≥ R1 + R2 + ... + Rn, với tổng chi phí vận chuyến là thấp nhất.

- Gọi A1, A2,..., Am là các điểm phát, B1, B2,..., Bn là các điểm thu, S1 + S2 +... + Sm = S là tổng phát, và R1 + R2 + ... + Rn = R là tổng thu. Nếu S = R, ta có bài toán vận tải cân bằng thu – phát. Nếu $S \neq R$, ta có bài toán vận tải không cân bằng thu – phát

- Thiết lập MH cho BTVT: Kí hiệu c_{ij} là chi phí vận chuyển một tấn hàng nói trên từ điểm phát Ai đến điểm thu Bj, ta có ma trận C = (c_{ij}) gọi là ma trận cước phí. Kí hiệu x_{ij} (tấn) là lượng hàng vận chuyển từ Ai đến Bj, ta có ma trận $x = (x_{ij})$ là phương án vận chuyển. Khi đó tổng lượng hàng chuyển đi từ điểm phát Ai là $\sum_j x_{ij}$, tổng lượng hàng chuyển tới Bj là $\sum_i x_{ij}$ và tổng chi phí vận chuyển là: $f(x) = \sum_i \sum_j c_{ij}.x_{ij}$ (hàm mục tiêu)

1. *MH cho toán học cho BTVT đóng* (S = R): $f(x) = \sum_i \sum_j c_{ij}.x_{ij} \to \min$

$$\begin{cases} \sum_{j=1}^{n} x_{ij} = S_i, i = 1,2,\dots,m \\ \sum_{i=1}^{m} x_{ij} = R_j, j = 1,2,\dots,n \\ x_{ij} \geq 0, \forall i = \overline{1,m}, \forall j = \overline{1,n} \end{cases}$$

-Chú thích: BTVT đóng có thể được giải bằng phương pháp thế vị.

2. ***MH cho toán học cho BTVT mở khi*** **S < R:** Bổ sung thêm một điểm phát giả thứ (m + 1) với lượng hàng phát: $S_{m+1} = R - S$, và mọi cước phí vận chuyển từ điểm phát giả đều = 0, tức là: $c_{m+1,j} = 0, \forall j = \overline{1,n}$, nhận được BTVT đóng

3. ***MH cho toán học cho BTVT mở khi*** **S > R:** Bổ sung thêm một điểm thu giả thứ (n + 1) với lượng hàng thu: $R_{n+1} = S - R$, và mọi cước phí vận chuyển đến điểm thu giả đều = 0, tức là: $c_{i,n+1} = 0, \forall i = \overline{1,m}$, nhận được BTVT đóng

b. Dạng toàn phương: Sự khác biệt của dạng toàn phương so với QHTT là hàm mục tiêu có dạng toàn phương.

b1. MH toán học:

$$x^* = \arg\min_{x\in\mathbb{R}^n}\left(\frac{1}{2}x^T.P.x + q^T.x + c\right)$$

$$\begin{cases} G.x \preccurlyeq h \\ A.x = b \end{cases}$$

(Trong đó: cho trước P là ma trận vuông cấp n, bán xác định dương, các ma trận $G \in \mathbb{R}^{m\times n}, A \in \mathbb{R}^{k\times n}, q \in \mathbb{R}^n, b \in \mathbb{R}^k, h \in \mathbb{R}^m$ và c là một hằng số)

Như vậy QH dạng toàn phương là một dạng toàn phương được tối thiểu trên một đa diện lồi D trong $\mathbb{R}^n$.

b2. Ví dụ (*Bài toán tìm đường đi ngắn nhất từ một điểm tới một đa diện lồi*)

Cần xác định hướng đi của một con thuyền đang ở ngoài biển tới một hòn đảo có hình tiếp nước là đa giác lồi sao cho đường đi ngắn nhất.

- Gọi a là vị trí hiện tại của con thuyền trên biển, x là điểm trên thiết diện tiếp nước D của hòn đảo (D là đa giác lồi: xác định bởi: $G.x \preccurlyeq h$ với G là một ma trận vuông cấp 2, $h \in \mathbb{R}^2$). Bình phương hoảng cách từ a tới x là:$\|x - a\|_2^2 = (x-a)^T.(x - a) = x^T.I.x - a^T.x + a^T.a$. Vậy MH toán học của bài toán là:

$$x^* = \arg\min_{x\in\mathbb{R}^n}\left(\frac{1}{2}x^T.I.x - \frac{1}{2}a^T.x + \frac{1}{2}a^T.a\right)$$

$$G.x \preccurlyeq h$$

c. Dạng QH hình học

Mô hình toán học: $$x^* = \arg\min_{x} f(x)$$

$$\begin{cases} f_i(x) \le 1, i = \overline{1,m} \\ h_j(x) = 1, j = \overline{1,k} \\ x \succ 0 \end{cases}$$

(với: $f(x), f_1(x), f_2(x), \dots, f_m(x)$là các đa thức, $h_1(x), \dots, h_k(x)$ là các đơn thức)

- Hàm số f: $\mathbb{R}^n_{++} \to \mathbb{R}$ được gọi là một đơn thức (*monomial*), nếu:

$$f(x) = c.x_1^{s_1}.x_2^{s_2} \dots .x_n^{s_n}, \forall x = (x_1, x_2, \dots, x_n)^T \in \mathbb{R}^n_{++}$$

trong đó: hằng số c > 0, $s_1, s_2, \dots, s_n$ là các hằng số bất kỳ, $\mathbb{R}^n_{++}$ gồm các véc tơ của $\mathbb{R}^n$có tất cả các tọa độ đều > 0: $\mathbb{R}^n_{++} = \{x \in \mathbb{R}^n : x \succ 0\}$.

- Hàm số f(x) có dạng là tổng các đơn thức: $f(x) = \sum_{j=1}^{k} c_j.x_1^{s_{1j}}.x_2^{s_{2j}} \dots .x_n^{s_{nj}}$ được gọi là một đa thức (*posynomial*).

Chú thích: Có thể đưa mh trên về bài toán tối ưu lồi, bằng cách đặt $x_i = \exp(y_i)$

5.3.4. Chú ý:

a. Bài toán tối ưu lồi nằm trong bài toán chung: Tìm cực trị có điều kiện của hàm nhiều biến, điều kiện là các ràng buộc hàm và ràng buộc biến.

b. Bài toán tối ưu lồi có thể giải theo phương pháp nhân tử Lagrange, phương pháp xấp xỉ (Thuật toán Gradient descent, thuật toán Newton, thuật toán Frank-Wolfe)

c. CVXOPT là một thư viện miễn phí trên Python. Có thể lập trình để giải các bài toán tối ưu lồi trên thư viện CVXOPT này

Chương 6. CÁC ĐẶC TRƯNG MẪU VÀ CÁC PHƯƠNG PHÁP ƯỚC LƯỢNG

5.1. Mẫu và các đặc trưng trên mẫu (Ôn tập lại)

5.1.1. Các khái niệm về mẫu

- **Tổng thể đ.tra:** Khi n/c một h.tượng hay g.quyết một c.việc, ta cần có những t.tin về h.tượng hay c.việc cần g.quyết. Tập hợp tất cả các đ.tượng mà ta q.tâm đ.tra, x.xét để có được t.tin nói trên được gọi là tổng thể đ.tra, (gọi tắt là tổng thể),ký hiệu là: Ω
- **Tiêu chuẩn đ.tra:** Những t.tin về p.tử của tổng thể p.vụ cho m.đích đ.tra, n/c gọi là t.chuẩn đ.tra, hay gọi tắt là t.chuẩn.
- **Biến quan sát:** Tập hợp tất cả các g.trị của một tiêu chuẩn đ.tra gọi là một biến quan sát, ký hiệu là: X, Y, Z,…

* Nếu tập g.trị của biến q.sát X là một tập hợp các số, tức là trên mỗi phần tử đ.tra thông tin được cho dưới dạng con số thì tiêu chuẩn đ.tra gọi là tiêu chuẩn số lượng. Khi đó gọi X là biến định lượng.

* Nếu trên mỗi p.tử được đ.tra, thông tin được thể hiện không phải là số (chẳng hạn: xanh, đỏ,...hay: trung bình, khá, giỏi,…) thì tiêu chuẩn đ.tra được gọi là t.chuẩn chất lượng. Khi đó gọi biến q.sát X tương ứng là biến định tính.

NX: G.sử X là biến q.sát trên tổng thể Ω. Việc chọn một p.tử từ tổng thể để q.sát, đ.tra thực chất là việc thực hiện một phép thử, mỗi p.tử của tổng thể là một kết cục và g.trị của X phụ thuộc vào p.tử được chọn, tức là phụ thuộc vào kết cục của phép thử. Vậy mỗi biến q.sát X là một biến ngẫu nhiên.

- **Mẫu**: Mỗi tập hợp con của tổng thể được lấy ra để quan sát gọi là mẫu điều tra hay gọi tắt là mẫu.

Y/c đặt ra đ.với mẫu là mẫu phải có t/c đ.diện cho tổng thể. Muốn vậy khi chọn mẫu, phải đ.bảo tính k.quan, không cố ý, không thiên vị,..., những y/c này gọi một cách đ.giản là chọn n.nhiên. Mẫu thu được theo cách đó gọi là mẫu n.nhiên. Số lượng p.tử được chọn vào mẫu gọi là kích thước mẫu hay cỡ mẫu, ký hiệu: n, m,...
Ở đây ta chỉ xét mẫu n.nhiên và khi nói tới mẫu mà không giải thích gì thêm thì hiểu rằng đó là mẫu n.nhiên.

5.1.2. Mẫu n.nhiên về biến q.sát: Xét mẫu n.nhiên k.thước n để q.sát X.
Ký hiệu X_j là g.trị của X ở p.tử thứ j của mẫu, (j = 1, 2,..., n) thì X_j là biến n.nhiên cùng p.phối x.suất với X. Hơn nữa do các p.tử được chọn vào mẫu một cách độc lập nhau nên là các biến n.nhiên đ.lập nhau.
Gọi $(X_1, X_2, \dots, X_n)$ là *mẫu ngẫu nhiên kích thước n về biến quan sát X*: đó là một véc tơ n.nhiên n chiều với các t.phần độc lập nhau và có cùng p.phối x.suất với biến q.sát X.

5.1.3. Một số đặc trưng quan trọng trên mẫu
G.sử $(X_1, X_2, \dots, X_n)$ là mẫu n.nhiên k.thước n về biến q.sát X trên t.thể Ω. Để p.biệt với các đ.trưng mẫu dưới đây, ta gọi:
$\mu = EX$ là t.bình t.thể (về biến qsát X); $\sigma^2 = DX$ là phương sai tổng thể;
$p = P(A)$ là tỷ lệ t/c A trong t.thể hay tần suất t.quát về A

a. Trung bình mẫu

a1. Đ/n: T.bình mẫu của biến q.sát X là đại lượng: $\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$

a2. *Các t/c của phép lấy t.bình mẫu:* (1) Nếu $X = c = const$ thì: $\bar{X} = c$; **(2)** Nếu $X \geq 0$ thì: $\bar{X} \geq 0$; **(3)** Với mọi hằng số k thì: $\overline{k.X} = k.\bar{X}$; **(4):** Với hai biến q.sát X, Y ta có:
4a. $\overline{X+Y} = \bar{X} + \bar{Y}$; **4b.** Nếu X và Y độc lập nhau thì: $\overline{X.Y} = \bar{X}.\bar{Y}$

(5) Với g(x) là một hàm số thì: $\overline{g(X)} = \frac{1}{n}\sum_{j=1}^{n} g(X_j)$. Đặc biệt: $\overline{X^2} = \frac{1}{n}\sum_j X_j^2$

(6) $\bar{X}$ là biến ngẫu nhiên có các t/chất: 6a. $\mathrm{E}\bar{X} = \mu; D\bar{X} = \frac{\sigma^2}{n}$; 6b. Khi n đủ lớn ta có:

$\bar{X} \approx \mu$, và $\bar{X}$ có p.phối x.xỉ chuẩn $\mathrm{N}(\mu, \frac{\sigma^2}{n})$ hay: $\frac{(\bar{X}-\mu)\sqrt{n}}{\sigma}$ có p.phối x.xỉ chuẩn $\mathrm{N}(0,1)$

b. Phương sai mẫu. Phương sai mẫu của biến q.sát X là: $\boldsymbol{S^2(X) = \frac{1}{n}\sum_{j=1}^{n}\left(X_j - \bar{X}\right)^2}$

Các t/c của phép lấy p.sai mẫu:

T/c 1: $S^2(X) \geq 0$, gọi: $S(X) = \sqrt{S^2(X)}$ là độ lệch mẫu của X

T/c 2: $\boldsymbol{S^2(X) = \overline{X^2} - (\bar{X})^2}$

T/c 3: Với mọi hằng số *a,* ta có: $S^2(a.X) = a^2.S^2(X)$

T/c 4: Nếu là các biến q.sát X, Y độc lập nhau thì:

$$\boldsymbol{S^2(X \pm Y) = S^2(X) + S^2\ (Y)}$$

T/c 5: $S^2(X)$ là một biến ngẫu nhiên mà: $\boldsymbol{ES^2(X) = \frac{n-1}{n}\sigma^2}$

Khi kích thước mẫu n đủ lớn ta có: $\boldsymbol{S^2(X) \approx \sigma^2}$

c. Tần suất mẫu. Xét mẫu n.nhiên k.thước n để q.sát t/c A nào đó của các p.tử.

c1. Đ.n: Ký hiệu m(A) là số p.tử có t/c A trên mẫu. Khi đó tỷ lệ t/c A trên mẫu là $\boldsymbol{f(A) = \frac{m(A)}{n}}$, còn gọi là tần suất mẫu của A.

c2. Các t/c: **(1)** $\boldsymbol{n.f(A) \sim B(n, p)}$ và vì thế: $\boldsymbol{Ef(A) = p; Df(A) = \frac{p(1-p)}{n}}$

(2) Khi n đủ lớn thì: a/ $\boldsymbol{f(A) \approx p}$ **;** b/ $f(A)$ có p.phối x.xỉ chuẩn $\boldsymbol{N(p; \frac{p(1-p)}{n})}$

5.1.4. Mẫu cụ thể - Trình bày số liệu đ.tra

Cho X là biến q.sát trên t.thể Ω và $(X_1, X_2, \dots, X_n)$ là mẫu n.nhiên k.thước n về X. Khi ta chọn n p.tử cụ thể vào mẫu thì n p.tử đó gọi là một mẫu cụ thể hay một g.trị x.định của mẫu. Với một mẫu cụ thể thì $(X_1, X_2, \dots, X_n)$ có một g.trị x.định là n g.trị của X q.sát được, ta gọi đó là tập số liệu hay k.quả đ.tra.

a. Trình bày bảng số liệu điều tra

a1 T.hợp 1: Khi số liệu đ.tra là n số liệu x.định gồm k g.trị khác nhau: $x_1, x_2, \dots, x_k$. Ký hiệu: n_i là số số liệu có g.trị là x_i. Ta gọi n_i là tần số của x_i. Khi đó số liệu được t.bày bằng bảng sau:

X	x_1	x_2	$\cdots$	x_k	Σ
Tần số n_i	n_1	n_2	$\cdots$	n_k	n

(1)

a2. T.hợp 2: Khi n số liệu không được cho cụ thể mà được chỉ ra trong k khoảng rời nhau: $I_1, I_2, \dots, I_k$, trong đó khoảng I_i chứa n_i số liệu.Ta thay tất cả các số liệu có trong khoảng I_i bởi trị t.tâm x_i(hay trị đ.diện) là điểm giữa của khoảng . Khi đó số liệu đ.tra được t.bày bởi bảng sau:

Các khoảng X	x_1	x_2	$\cdots$	x_k	Σ
Trị đại diện	x_1	x_2	$\cdots$	x_k	Σ
Tần số n_i	n_1	n_2	$\cdots$	n_k	n

b. Tính g.trị các đ.trưng mẫu từ tập s.liệu

G.sử biến q.sát X có k.quả đ.tra được cho bởi bảng (1) hoặc (2). Khi đó từ c.thức của các đ.trưng mẫu ta nhận được:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i;\ \overline{X^2} = \frac{1}{n}\sum_{i=1}^{k} n_i \mathrm{x}_\mathrm{i}^2; S^2(X) = \overline{X^2} - (\bar{X})^2$$

Vd 1: Đ.tra về mức t.nhập X(triệu đ/ tháng) của các hộ g.đình ở đ.p B, có các khoảng t.nhập: [4,8 – 5), [5 – 5,2), [5,2 – 5,4), [5,4 – 5,6), [5,6 – 5,8), [5,8 – 6), [6 – 6,2), [6,2 – 6,4], với số hộ tương ứng: 5, 20, 65, 105, 100, 50, 15, 5. Hãy tính mức t.nhập bq và đ.lệch mẫu về mức t.nhập của các hộ này.

X	[4,8 - 5)	[5 – 5,2)	[5,2-5,4)	[5,4-5,6)	[5,6-5,8)	[5,8-6)	[6-6,2)	[6,2-6,4}	Σ
Tần số	5	20	65	105	100	50	15	5	365
x_i	4,9	5,1	5,3	5,5	5,7	5,9	6,1	6,3	

Chú ý: Nếu biến quan sát X là véc tơ k chiều thì dữ liệu thu được ở mẫu kích thước n là n véc tơ k chiều: $\mathrm{X}_1, \mathrm{X}_2, \dots, \mathrm{X}_n$. Khi đó véc tơ trung bình mẫu $\bar{X}$ và ma trận hiệp phương sai mẫu $S^2(X)$ có công thức tính tương tự như một chiều:

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \dots + X_n) = \frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{X}_\mathrm{i}\ ;\ \mathrm{S}^2(\mathrm{X}) = \frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}\left(\mathrm{X}_\mathrm{i} - \bar{\mathrm{X}}\right).\left(\mathrm{X}_\mathrm{i} - \bar{\mathrm{X}}\right)^{\mathrm{T}} = \frac{1}{\mathrm{n}}\widehat{\mathrm{X}}.\left(\widehat{\mathrm{X}}\right)^{\mathrm{T}}$$

Trong đó các véc tơ được viết theo ma trận cột, các phép toán cộng, nhân là cộng, nhân các ma trận, $\left(\widehat{\mathrm{X}}\right)^{\mathrm{T}} = \left(\widehat{\mathrm{X}}_1, \widehat{\mathrm{X}}_2, \dots, \widehat{\mathrm{X}}_\mathrm{n}\right)$, với $\widehat{\mathrm{X}}_\mathrm{i} = \mathrm{X}_\mathrm{i} - \bar{X}$

5.2. Ước lượng khoảng tin cậy (KTC) cho các tham số
5.2.1. Các khái niệm
a. Đặt v.đề: G.sử θ là một g.trị chân thực chưa biết, nhưng cần biết mà lại k.thể biết c.xác. Khi đó phải tìm một đ.lượng $\hat{\theta}$ để x.xỉ cho θ, ta nói $\hat{\theta}$ là một ư.lượng cho θ. Vậy tìm $\hat{\theta}$ như thế nào? Y/cầu: $\hat{\theta}$ phải p.hợp với vai trò và ý nghĩa cũng như những t.tin có được về θ .
b. P.pháp chung g.quyết v.đề: Trước hết cần phải có thông tin về θ, nên ta cần chỉ ra biến q.sát, ta ký hiệu là X, có l.quan đến θ mà θ đóng vai trò là tham số của p.phối x.suất. Việc lấy t.tin ở đây có nghĩa là lập mẫu đ.tra về biến q.sát X, những t.tin này cũng là những t.tin về θ. Trên mẫu ta có đ.lượng $\hat{\theta}$ tương ứng với θ(Vai trò của $\hat{\theta}$ trên mẫu tương tự như vai trò của θ đối với biến q.sát X). Ta dùng đ.lượng $\hat{\theta}$ để ư.lượng cho θ.
Chú ý: $\hat{\theta}$ p.thuộc vào mẫu: $\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$ nên là biến n.nhiên p.thuộc cỡ mẫu n.
c. Ước lượng không chệch và ước lượng vững
c1. Tính không chệch: Ước lượng $\hat{\theta}$ được gọi là không chệch cho θ nếu: $E\hat{\theta} = \theta$
c2. Tính vững: $\hat{\theta}$ được gọi là vững cho θ nếu: $\forall \varepsilon > 0, \lim\limits_{n\to\infty} P\left(\left|\hat{\theta} - \theta\right| < \varepsilon\right) = 1$
Chú ý: $S^2(X)$ là ước lượng chệch cho σ^2, vì: $ES^2(X) = \frac{n-1}{n}\sigma^2 \neq \sigma^2$. Tuy nhiên, nếu xét đại lượng: ${S'}^2(X) = \frac{n}{n-1}S^2(X)$, thì ${S'}^2(X)$ là ước lượng vững, không chệch cho phương sai tổng thể σ^2. Ta gọi ${S'}^2(X)$ là *phương sai mẫu điều chỉnh* của biến quan sát X.
* Lưu ý, trong các phần mềm và một số tài liệu lại đ.nghĩa phương sai mẫu của biến quan sát X là $S^2(X) = \frac{n}{n-1}\left\{\overline{X^2} - (\bar{X})^2\right\}$, tuy nhiên đại lượng này lại không có tính chất tương tự như phương sai tổng thể.

d. Khoảng tin cậy cho tham ẩn

Nếu ta chỉ ra một khoảng ngẫu nhiên: $(\hat{\theta}_1, \hat{\theta}_2)$ mà tham ẩn θ có thể rơi vào với xác suất $\gamma = \mathrm{P}\{\theta \in (\hat{\theta}_1, \hat{\theta}_2)\}$ đủ lớn, thì ta gọi $(\hat{\theta}_1, \hat{\theta}_2)$ là KTC cho θ với độ tin cậy γ và bán kính của KTC được gọi là độ chính xác của ước lượng

5.2.2. Ước lượng KTC cho các tham số của tổng thể

G.sử $(X_1, X_2, \dots, X_n)$ là mẫu n.nhiên k.thước n về biến q.sát X trên tổng thể Ω. Trong đó trung bình tổng thể $\mu = EX$, phương sai tổng thể $\sigma^2 = \mathrm{Var}X$ và tỷ lệ t/c A trong t.thể là $p = P(A)$ chưa biết, cần tìm KTC với độ tin cậy $\gamma = 1 - \alpha$.

a. KTC cho trung bình tổng thể $\mu = EX$

a1. KTC một phía: $(-\infty;\ \overline{\boldsymbol{X}} + \boldsymbol{\varepsilon})$; (KTC bên trái, $\overline{\boldsymbol{X}} + \boldsymbol{\varepsilon}$ gọi là ước lượng tối đa cho μ)

$(\overline{\boldsymbol{X}} - \boldsymbol{\varepsilon};\ +\infty)$ (KTC bên trái, $\overline{\boldsymbol{X}} - \boldsymbol{\varepsilon}$ gọi là ước lượng tối thiểu cho μ)

$$\text{Trong đó: } \boldsymbol{\varepsilon} = \begin{cases} \boldsymbol{u}(\boldsymbol{\alpha})\dfrac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{n}}}, \text{nếu biết p. sai tổng thể } \sigma^2 \\ \boldsymbol{t}_{\boldsymbol{n-1}}(\boldsymbol{\alpha})\dfrac{\boldsymbol{S}(\boldsymbol{X})}{\sqrt{\boldsymbol{n-1}}}, \boldsymbol{n}\text{ếu chưa biết } \sigma^2 \text{ và cỡ mẫu n} < 30 \\ \boldsymbol{u}(\boldsymbol{\alpha})\dfrac{\boldsymbol{S}(\boldsymbol{X})}{\sqrt{\boldsymbol{n-1}}}, \text{nếu chưa biết } \sigma^2 \text{ và cỡ mẫu n} \geq 30 \end{cases}$$

với: $\boldsymbol{u}(\boldsymbol{\lambda})$ là phân vị mức $1 - \boldsymbol{\lambda}$ của phân phối chuẩn; $\boldsymbol{t}_{\mathrm{k}}(\boldsymbol{\lambda})$: phân vị mức $1 - \boldsymbol{\lambda}$ của p.phối Student với k bậc tự do. $S(X) = \sqrt{S^2(X)}$ $(S^2(X)$ *là phương sai mẫu*$)$

$$a2.\ \textit{KTC đối xứng: } (\overline{\boldsymbol{X}} - \boldsymbol{\varepsilon}, \overline{\boldsymbol{X}} + \boldsymbol{\varepsilon})\ (\boldsymbol{\varepsilon} = \begin{cases} u\left(\dfrac{\alpha}{2}\right).\dfrac{\sigma}{\sqrt{n}}, \textit{nếu biết phương sai tổng thể } \sigma^2 \\ t_{n-1}\left(\dfrac{\alpha}{2}\right)\dfrac{S(X)}{\sqrt{n-1}}, \textit{nếu chưa biết } \sigma^2, \textit{với cỡ mẫu } n < 30) \\ u\left(\dfrac{\alpha}{2}\right).\dfrac{S(X)}{\sqrt{n-1}}, \textit{nếu chưa biết } \sigma^2, \textit{với cỡ mẫu } n \geq 30 \end{cases}$$

Chú ý: Nếu dùng phương sai mẫu điều chỉnh: $S'^2(X) = \frac{n}{n-1}.S^2(X)$ (hay có: $S^2(X) = \frac{n-1.}{n}S'^2(X)$ thì trong công thức của ε, biểu thức $\frac{S(X)}{\sqrt{n-1}} = \frac{S'(X)}{\sqrt{n}}$

b. KTC cho tỷ lệ: Lập mẫu n.nhiên k.thước n để q.sát t/c A, trên đó ta có: $\mathrm{f} = f(A)$ là tỷ lệ t/c A trên mẫu. với độ tin cậy $\gamma = 1 - \alpha$, bài toán tìm KTC cho p = P(A) là tỷ lệ t/c A trong t.thể Ω chỉ xét dựa trên mẫu có k.thước n khá lớn và được g.quyết trong các tr.hợp như sau:

1/ Khi n lớn và f không quá gần 0 hoặc 1: KTC cho p là:

$$(\boldsymbol{f} - \boldsymbol{\varepsilon}, \boldsymbol{f} + \boldsymbol{\varepsilon}), \text{ (với } \varepsilon = u\left(\frac{\alpha}{2}\right).\sqrt{\frac{f(1-f)}{n}})$$

2/ *Khi n lớn và f gần 0 hoặc 1,với độ tin cậy* 95%*:*

* KTC cho $\lambda = np$ là: $(\boldsymbol{np_1}; \boldsymbol{np_2})$: Với m từ mẫu, tra bảng p.lục, tìm được: $np_1; np_2$. Suy ra KTC cho p là: $(\boldsymbol{p_1}; \boldsymbol{p_2})$

c. KTC cho phương sai tổng thể $\boldsymbol{\sigma}^2$: Với độ tin cậy $\gamma = 1 - \alpha$, KTC cho σ^2 là:

* KTC đối xứng cho σ^2 là: $\left(\frac{\boldsymbol{nS^2(X)}}{\boldsymbol{\chi^2_{n-1}\left(\frac{\alpha}{2}\right)}}, \frac{\boldsymbol{nS^2(X)}}{\boldsymbol{\chi^2_{n-1}(1-\frac{\alpha}{2})}}\right)$

* Các KTC một phía cho phương sai là:

$$\left(\boldsymbol{0}, \frac{\boldsymbol{nS^2(X)}}{\boldsymbol{\chi^2_{n-1}(1-\alpha)}}\right) \; (bên\ trái); \; \left(\frac{\boldsymbol{nS^2(X)}}{\boldsymbol{\chi^2_{n-1}(\alpha)}}; \; +\infty\right) \text{(bên phải)}$$

trong đó $\chi^2_k(\lambda)$ là giá trị tới hạn mức λ của phân phối khi – bình phương với k bậc tự do, tra từ bảng phụ lục của phân phối khi – bình phương

5.3. Một số phương pháp ước lượng tham số của mô hình

Nhiều mô hình Machin Learning được xây dựng dựa trên các mh thống kê. Các mh thống kê dựa vào các mh p.phối xác suất được đề cập trong chương 4. Gọi θ tập các tham số của một mh thống kê. Chẳng hạn với p.phối Poisson, tham $\theta = \lambda$ là trung bình tổng thể, với p.phối chuẩn n chiều, tham θ gồm véc tơ kì vọng μ và ma trận hiệp phương sai. Learning chính là quá trình ước lượng các tham số θ sao cho mh ước lượng có được phù hợp nhất với phân phối của dữ liệu.

5.3.1. Phương pháp bình phương tối thiểu (OLS: *Ordinary Least Squares*)

Được biết biến Y có quan hệ hàm số với X (một biến số hoặc một biến véc tơ): Y = Y(X) (Y(X) chưa biết) và qua quan sát có dữ liệu: $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ Ta muốn xấp xỉ Y(X) với một dạng hàm $f(x, a_1, a_2, \dots, a_m)$ đã biết, trong đó $a_1, a_2, \dots, a_m$ là các tham số chưa biết (ví như dạng đa thức bậc 3: $f(x) = ax^3 + bx^2 + cx + d$ có 4 tham số a, b, c, d chưa biết). P.pháp OLS dựa vào dữ liệu q/s để tìm ước lượng cho các tham số $a_1, a_2, \dots, a_m$ sao cho tổng bình phương các sai số giữa các giá trị ước lượng và các giá trị q/s được của hàm Y(x) là bé nhất, tức là: $F(a_1, a_2, \dots, a_m) = \sum_{j=1}^{n} \{f(x_j, a_1, a_2, \dots, a_m) - y_j\}^2 \longrightarrow \min$

Vậy các ước lượng cho các tham số $a_1, a_2, \dots, a_m$ là nghiệm của hệ ph.trình:

$$\frac{\partial F}{\partial a_k} = 0, \forall k = 1,2, \dots, m \qquad (1)$$

<u>VD1</u>. Giả sử từ bảng n dữ liệu quan sát về mối quan hệ hàm của biến Y và biến số X được cho ở trên, ta tìm hàm số f(x) = a + bx xấp xỉ cho hàm Y(x).

Giải. Có: $F(a,b) = \sum_{j=1}^{n}\{f(x_j,a,b) - y_j\}^2 = \sum_{j=1}^{n}\{a + bx_j - y_j\}^2$. Theo OLS, ta tìm a, b từ hệ p.trình: $\begin{cases} \frac{\partial F}{\partial a} = 0 \\ \frac{\partial F}{\partial b} = 0 \end{cases} \Leftrightarrow \begin{cases} \sum_{j=1}^{n}\{a + bx_j - y_j\} = 0 \\ \sum_{j=1}^{n}\{a + bx_j - y_j\}x_j = 0 \end{cases} \Leftrightarrow \begin{cases} a = \hat{a} = \overline{Y} - \hat{b}.\overline{X} \\ b = \hat{b} = \frac{\overline{X.Y} - \overline{X}.\overline{Y}}{\overline{X^2} - (\overline{X})^2} \end{cases}$ (*)

Thay a, b từ (*) vào, ta có: $A = \frac{\partial^2 F}{\partial a^2} > 0;\ \Delta = \frac{\partial^2 F}{\partial a^2}.\frac{\partial^2 F}{\partial b^2} - \frac{\partial^2 F}{\partial a \partial b} > 0$, nên (a,b) là điểm cực tiểu duy nhất của F, tức là F đạt trị nhỏ nhất tại (a,b) xác định bởi (*). Do đó hàm $f(x) = a + bx$, với (a,b) xác định từ (*) là hàm ước lượng cần tìm.

5.3.2. Phương pháp Hợp lý cực đai ML *(MLE: Maximum Likelihood Estimate)*

Giả sử biến *X* có phân phối xs phụ thuộc vào tham véc tơ $\theta = (\theta_1, \theta_2, \dots, \theta_k)$ chưa biết, cần tìm ước lượng cho θ. Với $W_n = (X_1, X_2, \dots, X_n)$ là mẫu ngẫu nhiên của biến q/s *X*. Ký hiệu $q(x,\theta)$ là xs để $W_n = (X_1, X_2, \dots, X_n)$ nhận giá trị $x = (x_1, \dots, x_n)$: $q(x,\theta) = P(X_1 = x_1, \dots, X_n = x_n) = P(X_1 = x_1). \dots P(X_n = x_n)$ nếu *X* là biến ngẫu nhiên rời rạc, và $q(x,\theta)$, là giá trị hàm mật độ xác suất của véc tơ $W_n = (X_1, \dots, X_n)$ tại x nếu *X* là biến liên tục. Ta gọi hàm $L(\theta) = q(X_1, X_2, \dots, X_n, \theta)$ là hàm hợp lý và ước lượng $\hat{\theta}$ mà tại đó hàm hợp lý $L(\theta) = q(X_1, X_2, \dots, X_n, \theta)$ đạt trị lớn nhất, là ước lượng hợp lý cực đại hay ước lượng hợp lý nhất, và nó là nghiệm hệ phương trinh sau mà ta gọi là hệ p.trình hợp lý:

$$\frac{\partial l(\theta_1, \theta_2, \dots, \theta_k)}{\partial \theta_j} = 0, \forall j = 1,2, \dots, k \qquad (2a)$$

Phương pháp hợp lý cực đại, ký hiệu là p.pháp ML (Maximum Liklihood) là p.pháp tìm ước lượng $\hat{\theta}$ cho tham ẩn θ là tìm nghiệm $\hat{\theta}$ của hệ phương trình hợp lý sao cho tại đó hàm hợp lý đạt trị lớn nhất.

Vì $\mathrm{L}(\theta)$ cùng tính đơn điệu với hàm Log $\mathrm{L}(\theta)$, nên hệ p.trình hợp lý được thay bởi:
$$\frac{\partial \mathrm{Log}\, l(\theta_1,\theta_2,\ldots,\theta_k)}{\partial \theta_j} = 0, \forall j = 1,2,\ldots,k \qquad \text{(2b)}$$

***VD2**. Với* $\mathrm{W_n} = (\mathrm{X_1}, \mathrm{X_2}, \ldots, \mathrm{X_n})$ là mẫu ngẫu nhiên của biến quan sát $\mathrm{X} \sim \mathrm{N}(\mu, \sigma^2)$, trong đó μ và σ^2 là các tham số chưa biết. Cần tìm ước lượng ML cho μ và σ^2

Giải. Đặt $\mathrm{t} = \sigma^2$, cần tìm ước lượng hợp lý cực đại cho tham véc tơ: $\theta = (\mu, \mathrm{t})$.

X có hàm mật độ xác suất là : $\mathrm{f(x)} = \frac{1}{\sqrt{2\pi \mathrm{t}}}.\mathrm{e}^{-\frac{1}{2}\left(\frac{\mathrm{x}-\mu}{\mathrm{t}}\right)^2}$, do đó hàm hợp lý là:
$$\mathrm{L}(\theta) = \mathrm{L}(\mu, \mathrm{t}) = \prod_{\mathrm{j}=1}^{\mathrm{n}} \mathrm{f}(\mathrm{x_j}) = (2\pi)^{-\frac{\mathrm{n}}{2}} \mathrm{t}^{-\frac{\mathrm{n}}{2}}.\exp\left\{-\frac{1}{2\mathrm{t}} \sum_{\mathrm{j}=1}^{\mathrm{n}} (\mathrm{x_j} - \mu)^2\right\}$$
$$\mathrm{Log}\big(\mathrm{L}(\theta)\big) = -\frac{\mathrm{n}}{2}\log(2\pi) - \frac{\mathrm{n}}{2}\log \mathrm{t} - \frac{1}{2\mathrm{t}}\sum_{\mathrm{j}=1}^{\mathrm{n}} (\mathrm{x_j} - \mu)^2$$

Hệ phương trình hợp lý:
$$\begin{cases} \dfrac{\partial \mathrm{Log}\, l(\theta)}{\partial \mu} = 0 \\ \dfrac{\partial \mathrm{Log}\, l(\theta)}{\partial \mathrm{t}} = 0 \end{cases} \Leftrightarrow \begin{cases} \frac{1}{\mathrm{t}} \sum_{\mathrm{j}=1}^{\mathrm{n}} (\mathrm{x_j} - \mu) = 0 \\ -\frac{\mathrm{n}}{2\mathrm{t}} + \frac{1}{2\mathrm{t}^2} \sum_{\mathrm{j}=1}^{\mathrm{n}} (\mathrm{x_j} - \mu)^2 = 0 \end{cases} \Leftrightarrow \begin{cases} \boldsymbol{\mu} = \widehat{\boldsymbol{\mu}} = \overline{\mathrm{X}} = \frac{1}{\mathrm{n}} \sum_{\mathrm{j}=1}^{\mathrm{n}} \mathrm{x_j} \\ \mathrm{t} = \hat{\mathrm{t}} = \mathrm{S}^2(\mathrm{X}) = \overline{\mathrm{X}^2} - \left(\overline{\mathrm{X}}\right)^2 \end{cases}$$

Như vậy đối với biến ngẫu nhiên có phân phối chuẩn thì trung bình mẫu $\overline{\mathrm{X}}$ là ước lượng hợp lý nhất cho giá trị trung bình tổng thể $\mu = \mathrm{EX}$ và phương sai mẫu $\mathrm{S}^2(\mathrm{X})$ là ước lượng hợp lý nhất cho phương sai tổng thể $\sigma^2 = \mathrm{Var(X)}$.

VD3. Tìm ước lượng ML cho véc tơ kỳ vọng μ và ma trận hiệp phương sai Λ của véc tơ ngẫu nhiên X có phân phối chuẩn k chiều.

Giải. X có mật độ chuẩn k chiều:

$$f(x) = \mathrm{f}(\mathrm{x}_1, \mathrm{x}_2, \dots, \mathrm{x}_\mathrm{k}) = \frac{1}{(2\pi)^{\mathrm{k}/2}.(\det\Lambda)^{1/2}} exp\left\{\frac{1}{2}.(x-\mu)^T\Lambda^{-1}.(x-\mu)\right\}$$

Khi đó từ mẫu ngẫu nhiên kích thước n $\mathrm{W} = (\mathrm{X}_1, \mathrm{X}_2, \dots, \mathrm{X}_\mathrm{n})$, ta có hàm hợp lý:

$$\mathrm{p}(\mathrm{X}_1, \mathrm{X}_2, \dots, \mathrm{X}_\mathrm{n}) = \prod_{\mathrm{j}=1}^{\mathrm{n}} f(\mathrm{X}_\mathrm{j}) = \frac{1}{(2\pi)^{\mathrm{kn}/2}.(\det\Lambda)^{\mathrm{n}/2}} \prod_{\mathrm{j}=1}^{\mathrm{n}} exp\left\{\frac{1}{2}.(\mathrm{X}_\mathrm{j}-\mu)^T\Lambda^{-1}.(\mathrm{X}_\mathrm{j}-\mu)\right\}$$

$$\ln \mathrm{p}(\mathrm{X}_1, \mathrm{X}_2, \dots, \mathrm{X}_\mathrm{n}) = \ln\frac{1}{(2\pi)^{\mathrm{kn}/2}.(\det\Lambda)^{\mathrm{n}/2}} + \sum_{\mathrm{j}}\left\{\frac{1}{2}.(\mathrm{X}_\mathrm{j}-\mu)^T\Lambda^{-1}.(\mathrm{X}_\mathrm{j}-\mu)\right\}$$

Hệ p.trình hợp lí: $\begin{cases} \nabla_\mu \ln\mathrm{p}(\mathrm{X}_1, \mathrm{X}_2, \dots, \mathrm{X}_\mathrm{n}) = 0 \\ \nabla_{\Lambda^{-1}} \ln\mathrm{p}(\mathrm{X}_1, \mathrm{X}_2, \dots, \mathrm{X}_\mathrm{n}) = 0 \end{cases}$ (*)

Tính: $\nabla_\mu\left\{(\mathrm{X}_\mathrm{j}-\mu)^T\Lambda^{-1}.(\mathrm{X}_\mathrm{j}-\mu)\right\} = -\left(\Lambda^{-1} + (\Lambda^{-1})^\mathrm{T}\right).(\mathrm{X}_\mathrm{j}-\mu)$

$$\nabla_{\Lambda^{-1}}\left\{(\mathrm{X}_\mathrm{j}-\mu)^T\Lambda^{-1}.(\mathrm{X}_\mathrm{j}-\mu)\right\} = (\mathrm{X}_\mathrm{j}-\mu)(\mathrm{X}_\mathrm{j}-\mu)^T$$

$$\nabla_{\Lambda^{-1}}\left\{\ln\frac{1}{(2\pi)^{\mathrm{kn}/2}.(\det\Lambda)^{\mathrm{n}/2}}\right\} = \nabla_{\Lambda^{-1}}\left(\frac{\mathrm{n}}{2}.\ln(\det\Lambda^{-1})\right) = \frac{\mathrm{n}}{2}.\Lambda$$

$$(*) \Leftrightarrow \begin{cases} \left(\Lambda^{-1} + (\Lambda^{-1})^\mathrm{T}\right)\left\{\sum_\mathrm{j} \mathrm{X}_\mathrm{j} - \mathrm{n}\mu\right\} = 0 \\ \frac{\mathrm{n}}{2}.\Lambda + \frac{1}{2}\sum_\mathrm{j}(\mathrm{X}_\mathrm{j}-\mu)(\mathrm{X}_\mathrm{j}-\mu)^T = 0 \end{cases} \Leftrightarrow \begin{cases} \mu = \hat{\mu} = \frac{1}{n}\sum_\mathrm{j} \mathrm{X}_\mathrm{j} = \overline{X} \\ \Lambda = \widehat{\Lambda} = \frac{1}{n}\sum_\mathrm{j}(\mathrm{X}_\mathrm{j}-\mu)(\mathrm{X}_\mathrm{j}-\mu)^T \end{cases} \quad (**)$$

Vậy ước lượng ML cho véc tơ kì vọng μ và ma trận hiệp phương sai Λ của phân phối chuẩn nhiều chiều là $\hat{\mu}$ và $\widehat{\Lambda}$ được cho bởi (**)

VD4. Tìm ước lượng ML cho các tham số $\lambda_1, \lambda_2, \dots, \lambda_{\mathrm{k}}$ của phân phối phân loại (categorical distribution)

Giải. Xét X là biến ngẫu nhiên có phân phối phân lọai với các tham số $\lambda_1, \lambda_2, \dots,$ λ_{k}**:** $\boldsymbol{P(X = i) = \lambda_i, i = 1,2, \dots, k}$. Việc q/s biến ngẫu nhiên X có thể được mô hình hóa bởi việc tung một khối đa diện lồi có k mặt được đánh số từ 1 đến k với xác suất để mặt i (mặt có chữ số i) tiếp xúc với mặt phẳng nền khi rơi xuống là $\boldsymbol{\lambda_i, i = 1,2, \dots, k}$**.** Tiến hành n lần tung khối đa diện này. kí hiệu $\mathrm{X_i}$ là số lần mặt i tiếp xúc với mặt phẳng sàn, thì $\mathrm{X_i} \sim \mathrm{B}(\mathrm{n}, \boldsymbol{\lambda}_{\mathrm{i}})$ và véc tơ $(X_1, X_2, \dots, X_k)$ có phân phối đa thức với các tham số n, $\lambda_1, \lambda_2, \dots, \lambda_{\mathrm{k}-1}$**,** tức là ta có hàm hợp lí:

$$\mathrm{L}(\lambda_1, \lambda_2, \dots, \lambda_{\mathrm{k}-1}) = \frac{\mathrm{n}!}{X_1! X_2! \dots, X_k!} . \lambda_1^{X_1} \lambda_2^{X_2} \dots \lambda_{\mathrm{k}}^{X_k} \text{ (Lưu ý: } \sum_{\mathrm{j}=1}^{\mathrm{k}} \lambda_{\mathrm{j}} = 1,\ \sum_{\mathrm{j}=1}^{\mathrm{k}} \mathrm{X_j} = \mathrm{n})$$

$$\log \mathrm{L}(\lambda_1, \lambda_2, \dots, \lambda_{\mathrm{k}-1}) = \log \frac{\mathrm{n}!}{X_1! X_2! \dots, X_k!} + \sum_{\mathrm{j}=1}^{\mathrm{k}-1} \mathrm{X_j} . \log \lambda_{\mathrm{j}} + \mathrm{X_k} . \log\left(1 - \sum_{\mathrm{j}=1}^{\mathrm{k}-1} \lambda_{\mathrm{j}}\right)$$

$\dfrac{\boldsymbol{\partial} \log \mathrm{L}(\lambda_1, \lambda_2, \dots, \lambda_{\mathrm{k}-1})}{\boldsymbol{\partial} \lambda_{\mathrm{i}}} = \dfrac{\mathrm{X_i}}{\lambda_{\mathrm{i}}} - \mathrm{X_k} . \dfrac{1}{\left(1 - \sum_{\mathrm{j}=1}^{\mathrm{k}-1} \lambda_{\mathrm{j}}\right)}, \mathrm{i} = 1,2, \dots, \mathrm{k} - 1$**.** Có hệ p.t hợp lí:

$\dfrac{\boldsymbol{\partial} \log \mathrm{L}(\lambda_1, \lambda_2, \dots, \lambda_{\mathrm{k}-1})}{\boldsymbol{\partial} \lambda_{\mathrm{i}}} = 0, \left(\forall \mathrm{i} = \overline{1, \mathrm{k} - 1}\right) \Leftrightarrow \mathrm{X_k} . \lambda_{\mathrm{i}} = \mathrm{X_i} . \lambda_{\mathrm{k}}, \forall \mathrm{i} = \overline{1, \mathrm{k} - 1}$ **(*)**

Từ (*) lấy tổng 2 vế theo i, nhận đươc: $\mathrm{X_k} . (1 - \lambda_{\mathrm{k}}) = (\mathrm{n} - \mathrm{X_k}) . \lambda_{\mathrm{k}}$**,** nhận được: $\lambda_{\mathrm{k}} = \dfrac{\mathrm{X_k}}{\mathrm{n}}$ **(**).** Thay (**) vào (*), nhận được: $\lambda_{\mathrm{i}} = \dfrac{\mathrm{X_i}}{\mathrm{n}}, \forall \mathrm{i} = \overline{1, \mathrm{k}}$

Bài tập. Tìm ước lượng hợp lý cực đại (ML) cho tham số p trong phân phối nhị thức B(n, p), tham số λ của phân phối P_λ, tham số λ của phân phối mũ E_λ

5.3.3. Phương pháp ước lượng hậu nghiệm cực đại

P.pháp MLE ước lượng tập tham số θ trong phân bố của biến q/s X là cực đại hóa hàm hợp lí dựa trên mẫu điều tra về X. Vì chỉ phụ thuộc mẫu về X nên khi mẫu không đủ tính đại diện (chẳng hạn cỡ mẫu bé) thì ước lượng MLE kém hiệu quả. Trong khi phương pháp ước lượng hậu nghiệm (MAP: *Maximum A Posteriori Esitimation*) coi θ là một biến ngẫu nhiên,không những dựa vào mẫu về X mà còn dựa vào thông tin đã biết về phân bố của θ là p(θ). Khi đã có mẫu $(X_1, X_2, \dots, X_\mathrm{n})$ thì thông tin về θ là $p(\theta|(X_1, X_2, \dots, X_n))$, biểu thức này gọi là xác suất hậu nghiệm. P.pháp tìm ước lượng cho θ là cực đại xác suất hậu nghiệm gọi là p.pháp ước lượng hậu nghiệm cực đại: $\theta = \arg\max p(\theta|(X_1, X_2, \dots, X_n))$

Ta có: $$p(\theta|(X_1, X_2, \dots, X_n)) \to \max \Leftrightarrow \frac{\mathrm{p}\big((X_1, X_2, \dots, X_n)\big|\theta\big)\mathrm{p}(\theta)}{(X_1, X_2, \dots, X_n)} \to \max$$

$$\Leftrightarrow \mathrm{p}((X_1, X_2, \dots, X_n)|\theta)\mathrm{p}(\theta) \to \max \Leftrightarrow \prod_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{p}(X_\mathrm{i}|\theta) \,.\, \mathrm{p}(\theta) \to \max$$

a. Chú ý:

(1) $\mathrm{p}((X_1, X_2, \dots, X_n)|\theta) = \mathrm{L}(\theta)$ là hàm hợp lí, còn p(θ) được gọi là xác suất tiên nghiệm.
(2) Khác biệt giữa MAP và MLE là hàm mục tiêu của MAP có thêm phân phối $\mathrm{p}(\theta)$ của θ, phân phối này chính là những thông tin biết trước về θ, nên gọi là tiên nghiệm.

b. Chọn tiên nghiệm: Để chọn tiên nghiệm, cần lưu ý các k/niệm sau:

b1. Tiên nghiệm liên hợp: Nếu phân phối hậu nghiệm $p(\theta|(X_1, X_2, \dots, X_n))$ có cùng dạng với phân phối tiên nghiệm $\mathrm{p}(\theta)$ thì hai p.phối này được gọi là cặp p.phối liên hợp và $\mathrm{p}(\theta)$ được gọi là tiên nghiệm liên hợp của hàm hợp lí $\mathrm{p}((X_1, X_2, \dots, X_n)|\theta)$

- Nếu hàm hợp lí và tiên nghiệm cho véc tơ kỳ vọng là các p.phối chuẩn thì p.phối hậu nghiệm cũng là p.phối chuẩn. Ta nói p.phối chuẩn liên hợp với chính nó (tự liên hợp).
- Nếu hàm hợp lí là p.phối chuẩn và tiên nghiệm cho phương sai là p.phối Gamma, p.phối

hậu nghiệm cũng là p.phối chuẩn thì ta nói p.phối Gamma là tiên nghiệm liên hợp cho phương sai của p.phối chuẩn.
- P.phối Beta là liên hiệp của p.phối Bernoulli.
- P.phối Dirichlet là liên hợp của p.phối Catergorical.

b2. Siêu tham số: Khi p.phối tiên nghiệm p(θ) lại phụ thuộc vào các tham số $\alpha, \beta, \ldots$ khác thì các tham số mới này được gọi là các siêu tham số

VD5. Cần ước lượng cho tham số θ là xác suất thành công trong của biến X có p.phối Bernoulli: $\theta = \mathrm{P}(\mathrm{X} = 1)$. Với mẫu cỡ n về biến q/sát X, gọi m là số thành công (số lần xuất hiện (X = 1)). MLE cho ta ước lượng của θ là $\hat{\theta} = \frac{\mathrm{m}}{\mathrm{n}}$. Ta cần tìm ước lượng MAP cho θ.
P.phối Bernoulli có mật độ xác suất: $\mathrm{p}(\mathrm{x}|\theta) = \theta^{\mathrm{x}}(1-\theta)^{1-\mathrm{x}}, \mathrm{x} \in \{0,1\}$. Liên hợp của nó là p.phối Beta có hàm mật độ xác suất p(θ) = $\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha).\Gamma(\beta)}.\theta^{\alpha-1}.(1-\theta)^{\beta-1}$. Như vậy thành phần chứa θ của p(θ) cùng dạng với p.phối Bernoulli và nếu dung p.phối Beta làm tiên nghiệm cho tham θ, không kể hằng số $\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha).\Gamma(\beta)}$ thì $\mathrm{p}(\theta|\mathrm{x})$ tỉ lệ với $\mathrm{p}(\mathrm{x}|\theta)$. p($\theta$) hay tỉ lệ với $\theta^{\mathrm{x}+\alpha-1}.(1-\theta)^{(1+\beta-\mathrm{x})-1}$ có dạng một p.phối Bernoulli. Vậy p.phối Beta là một tiên nghiệm liên hợp của p.phối Bernoulli và α, β là các siêu tham số.
- Để ước lượng cho θ trong p.phối Bernoulli của biên X theo MAP, với tiên nghiệm p(θ) là hàm mật độ p.phối Beta (với 2 siêu tham số α, β), ta có bài toán tối ưu:

$$\mathrm{p}(\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_\mathrm{n}|\theta).\mathrm{p}(\theta) \to \max \Leftrightarrow \prod_{\mathrm{i}=1}^{\mathrm{n}} \theta^{\mathrm{x}_\mathrm{i}}(1-\theta)^{1-\mathrm{x}_\mathrm{i}}.\theta^{\alpha-1}.(1-\theta)^{\beta-1} \to \max$$

$$\Leftrightarrow \theta^{\mathrm{m}+\alpha-1}.(1-\theta)^{\mathrm{n}-\mathrm{m}+\beta-1} \to \max \Leftrightarrow \theta = \hat{\theta} = \frac{\mathrm{m}+\alpha-1}{n+\alpha+\beta-2}$$

NX: Khi $\alpha = \beta = 1$, nhận được $\hat{\theta} = \frac{\mathrm{m}}{\mathrm{n}}$ là ước lượng MLE của θ.