KHÁI NIỆM (G, ε, δ) – ĐẦY ĐỦ

Lê Ngọc Bường

Trong lý thuyết thống kê, khi nghiên cứu tính chất duy nhất của ước lượng không chệch có phương sai bé nhất, các nhà thống kê học đã đưa ra một khái niệm quan trọng gọi là thống kê đầy đủ. Theo định nghĩa thì một họ các phân bố xác xuất $P = (P_\theta, \theta \in \Theta)$ trong đó Θ là không gian tham số được gọi là đầy đủ nếu như từ đẳng thức

$$E_\theta \varphi = 0 \ \forall \ \theta \in \Theta$$

Thì suy ra

$$\varphi = 0 \quad (P-h.kn)$$

ở đây φ là hàm P – khả tích và không phụ thuộc θ còn toán tử E_θ là toán tử lấy kỳ vọng theo phân bố P_θ.

Cũng theo định nghĩa thì thống kê $T(x)$ được gọi là thống kê đầy đủ nếu họ các phân bố xác suất cảm sinh bởi T là một họ đầy đủ.

Trong thực tế, những họ phân bố xác suất đầy đủ không phải là phổ biến. Dưới đây, chúng tôi sẽ mở rộng khái niệm thống kê đầy đủ ra dạng thống kê (G, ε, δ) – đầy đủ và xét một số ví du cụ thể để mong đáp ứng phần nào việc xử dụng khái niệm đầy đủ một cách rộng rãi hơn trong lĩnh vực xác suất thống kê.

Định nghĩa 1: Giả sử $\mathcal{X}$ là không gian các đối tượng nghiên cứu, A là σ– trường Boren trên $\mathcal{X}$ còn $P = \{(P_\theta, \theta \in \Theta)\}$ là họ các độ đo xác suất cho trên $(\mathcal{X}, A)$, trong đó Θ là không gian tham số. Khi đó bộ ba $(\mathcal{X}, A, P)$ được gọi là một cấu trúc thống kê.

Định nghĩa 2: Cho $(\mathcal{X}, A, P)$ là một cấu trúc thống kê. G là một lớp con các hàm A – đo được và P – khả tích. Cấu trúc $(\mathcal{X}, A, P)$ được gọi là (G, ε, δ) – đầy đủ nếu từ bất đẳng thức:

$$|E_\theta \varphi| \leqslant \varepsilon \quad \forall \theta \in \Theta$$

Trong đó: $\varphi \in G$ thì suy ra

$$|\varphi| \leqslant \delta \quad (P - h.k.n)$$

Nếu G bao gồm tất cả các hàm A – đo được và P – khả tích thì ta nói rằng cấu trúc thống kê $(\mathcal{X}, A, P)$ là (ε, δ) – đầy đủ.

Nếu G bao gồm tất cả các hàm A – đo được, P – khả tích và giới nội thì ta nói cấu trúc thống kê $(\mathcal{X}, A, P)$ là (ε, δ) – đầy đủ giới nội.

Định nghĩa 3: Cho $(\mathcal{X}, A, P)$ là một cấu trúc thống kê.
G là một lớp con các hàm A – đo được và P – khả tích nào đó. Thống kê T ánh xạ từ $(\mathcal{X}, A, P)$ vào (Y, B, P^T) được gọi là (G, ε, δ) – đầy đủ nếu cấu trúc thống kê $(\mathcal{X}, T^{-1}(B), P)$ là (G, ε, δ) – đầy đủ.

Bổ đề 1: Cho g(t) là hàm đặc trưng có dạng:

$$g(t) = \int_{-\infty}^{+\infty} f(t)\, e^{itx} dx$$

trong đó f(x) là hàm đo được thì f (x) là hàm mật độ.

Chứng minh: Vì g(t) là hàm đặc trưng nên $\int_{-\infty}^{+\infty} f(x)\, dx = 1$

Đặt

$$F(x) = \int_{-\infty}^{x} f(u)\, du$$

ta có $F(-\infty) = 0$ và $F(+\infty) = 1$

Hơn nữa

$$g(t) = \int_{-\infty}^{+\infty} e^{itx} dF(x) \tag{1}$$

Hàm F(x) là hàm liên tục trên R^1

Mặt khác, vì g(t) là hàm đặc trưng nên có tồn tại một hàm phân phối $\bar{F}(x)$ sao cho

$$g(t) = \int_{-\infty}^{+\infty} e^{itx}\, d\bar{F}(x) \tag{2}$$

Hàm $\bar{F}(x)$ là không giảm, liên tục phải và $\bar{F}(-\infty) = 0$, $\bar{F}(+\infty) = 1$.

Từ (1) và (2) ta thấy g(t) đồng thời là phép biến đổi Fourier của hàm F(x) và $\bar{F}(x)$ nên ta có $F(x) = \bar{F}(x)\ \forall x \in c(\bar{F})$ trong đó $c(\bar{F})$ là tập hợp tất cả các điểm liên tục của hàm $\bar{F}(x)$

Từ tính chất đã nêu ra ở trên của hai hàm F(x) và $\bar{F}(x)$ ta có $F(x) \equiv \bar{F}(x)\ \forall x$. Vậy f(x) chính là hàm mật độ

Bổ đề 2: Cho họ mũ 2k tham số $\eta = (\xi_1, \ldots \xi_k, \sigma_1, \ldots \sigma_k)$ trong đó

$$-\infty < \xi_i < +\infty,\ 0 < \sigma_i < 1 \qquad i = 1, 2, \ldots, k$$

với hàm mật độ cho bởi

$$f(x, \eta) = h(x) \exp\left\{ \sum_{i=1}^{k} \xi_i U_i(x) - \sum_{i=1}^{k} \frac{1}{2\sigma_i^2} U_i^2(x) + v(\eta) \right\}$$

Giả thiết thêm rằng hàm $\sum_{i=1}^{k} U_i(x)$ là hàm tăng thực sự, khả vi theo x, nhận mọi giá trị trên R^1. Khi đó từ bất đẳng thức

$$0 \leqslant E\eta\, \varphi < +\infty \quad \forall \eta$$

Suy ra

$$0 \leqslant \varphi \qquad (P - h.k.n)$$

Chứng minh: Giả sử $\varphi(x)$ là hàm bất kỳ thỏa mãn điều kiện

$$0 \leqslant E\eta\, \varphi < +\infty \qquad \forall \eta$$

Đặt

$$\tilde{\varphi}(x) = \varphi(x) \exp\left\{ -\frac{1}{2} \sum_{i=1}^{k} U_i^2(x) \right\}$$

$$\eta_0 = (\underbrace{0, \ldots, 0}_{k \text{ lần}}, \underbrace{1, \ldots, 1}_{k \text{ lần}})$$

Từ $E_{\eta_0} \varphi > 0$ ta suy ra $\int_{-\infty}^{+\infty} \tilde{\varphi}(x)\, h(x)\, dx > 0$ (1)

Đặt $\tilde{\sigma}_i^2 = \dfrac{\sigma_i^2}{1 - \sigma_i^2}$, $\theta_i = 2\zeta_i \tilde{\sigma}_i^2$ ta có $0 < \tilde{\sigma}_i^2 < +\infty$

Và từ $E\eta\varphi > 0$ $\forall\eta$ ta suy ra:

$$\int_{-\infty}^{+\infty} \tilde{\varphi}(x)\, h(x) \exp\left\{\sum_{i=1}^{k} \frac{(\theta_i U_i(x))^2}{2\tilde{\sigma}_i^2}\right\} dx > 0 \qquad (2)$$

Xét trường hợp thứ nhất: $\int_{-\infty}^{+\infty} \tilde{\varphi}(x)\, h(x)\, dx \neq 0$

Đặt
$$C^{-1} = \int_{-\infty}^{+\infty} \tilde{\varphi}(x)\, h(x)\, dx$$

$$\psi(\theta, \tilde{\sigma}) = \int_{-\infty}^{+\infty} \frac{C}{(2\pi)^{k/2} \prod_{i=1}^{k} \tilde{\sigma}_i}\, \tilde{\varphi}(x)\, h(x) \exp\left\{-\sum_{i=1}^{k} \frac{(U_i(x) - \theta_i)^2}{\tilde{\sigma}_i^2}\right\} dx$$

Ta có
$$\underbrace{\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty}}_{k \text{ lần}} \psi(\theta, \tilde{\sigma}) \prod_{i=1}^{k} d\theta_i = 1 \qquad \forall(\tilde{\sigma}_1, \ldots, \tilde{\sigma}_k)$$

Vậy $\psi(\theta, \tilde{\sigma})$ chính là hàm mật độ xác suất của các biến $(\theta_1, \ldots, \theta_k)$ và tham số là $(\tilde{\sigma}_1, \ldots, \tilde{\sigma}_k)$. Từ đó hàm đặc trưng tương ứng với $\psi(\theta, \tilde{\sigma})$ được cho bởi:

$$\zeta_{\tilde{\sigma}}(t_1, \ldots, t_k) = \underbrace{\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty}}_{k \text{ lần}} \exp\left\{i \sum_{j=1}^{k} \theta_j t_j\right\} \psi(\theta, \tilde{\sigma}) \prod_{i=1}^{k} d\theta_i =$$

$$= c \exp\left\{-\frac{1}{2} \sum_{j=1}^{k} \tilde{\sigma}_j^2 t_j^2\right\} \cdot \int_{-\infty}^{+\infty} \tilde{\varphi}(x)\, h(x) \exp\left\{i \sum_{j=1}^{k} t_j U_j(x)\right\} dx$$

Đặc biệt lấy $t_1 = t_2 = \ldots = t_k = t$ thì

$$\zeta_{\tilde{\sigma}}(t, \ldots, t) = c \exp\left\{-\frac{t^2}{2} \sum_{j=1}^{k} \tilde{\sigma}_j^2\right\} \cdot \int_{-\infty}^{+\infty} \tilde{\varphi}(x)\, h(x) \exp\left\{i t \sum_{j=1}^{k} U_j(x)\right\} dx$$

Vì hàm $\zeta_{\tilde{\sigma}}(t, \ldots, t)$ liên tục tại $t = 0$ và vì hàm $\exp\left[-\dfrac{t^2}{2} \sum_{i=1}^{k} \tilde{\sigma}_i^2\right]$ liên tục tại $t = 0$ nên ta có

$$\int_{-\infty}^{+\infty} \tilde{\varphi}(x)\, h(x) \exp\left[it \sum_{j=1}^{k} U_j(x) \right] dx$$ cũng liên tục tại $t = 0$

$\lim \zeta_{\tilde{6}}(t,.., t)$

do đó $\sum_{j=1}^{k} \tilde{6}_j^2 \to 0 = C\int_{-\infty}^{+\infty} \tilde{\varphi}(x)\, h(x) \exp\{ it \sum_{j=} U_j(x)\}\, dx$

là hàm đặc trưng.

Đặt $U(x) = \sum_{i=1}^{k} U_i(x)$. Từ giả thiết có $U'(x) > 0$ và như vậy

$$C\int_{-\infty}^{+\infty} \tilde{\varphi}(U^{-1}(x))\, h(U^{-1}(x))\, e^{itx} \frac{dx}{U'(x)}$$

là hàm đặc trưng. Theo bổ đề 1 ta có

$\frac{1}{U'(x)} \tilde{\varphi}(U^{-1}(x))\, h(U^{-1}(x))$ là hàm mật độ. Vậy

$$\tilde{\varphi}(U^{-1}(x))\, h(U^{-1}(x)) \geqslant 0 \quad \forall x$$

Từ đó $\tilde{\varphi}(x) \geqslant 0 \quad \forall x \Rightarrow \varphi(x) \geqslant 0 \quad \forall x$ trừ trên tập đếm được

Trường hợp 2 : $\int_{-\infty}^{+\infty} \tilde{\varphi}(x)\, h(x)\, dx = 0$

Ta lấy dãy $\tilde{\varphi}_n(x)$ thỏa mãn :

1. $\lim_{n\to\infty} \tilde{\varphi}_n(x) = \tilde{\varphi}(x) \qquad \forall x$

2. $\int_{-\infty}^{+\infty} \tilde{\varphi}_n(x)\, h(x) \exp\left\{ -\frac{1}{2} \sum_{j=1}^{k} \frac{(U_j(x) - \theta_j)^2}{2\tilde{6}_j^2} \right\} dx > 0 \quad \forall_n \geqslant N$

Có thể chỉ ra một dãy cụ thể ví dụ như ta lấy dãy $\varepsilon_n \to 0$ và hàm $g(x)$ sao cho $\int_{-\infty}^{+\infty} h(x)\, g(x)\, dx < +\infty$ khi đó lấy

$$\tilde{\varphi}_n(x) = \tilde{\varphi}(x) + \varepsilon_n\, g(x)$$

Theo chứng minh trên thì $\tilde{\varphi}_n(x) \geqslant 0 \quad \forall_n \geqslant N, \forall x$

Vậy $\tilde{\varphi}(x) \geqslant 0\ \forall x \Rightarrow \varphi(x) \geqslant 0 \quad \forall x$ trừ trên tập đếm được.

Bổ đề 3 : Cho họ ε – mũ 2k tham số $\eta = (\xi_1,.., \xi_k, 6_1,.., 6_k)$

Trong đó $-\infty < \zeta_i < +\infty,\quad 0 < 6_i < 1 \qquad i = 1, 2,.., k$

Với hàm mật độ cho bởi :

$$f(x, \eta) = h(x) \exp\left\{ \sum_{i=1}^{k} \zeta_i U_i(x) - \sum_{i=1}^{k} \frac{U_i^2(x)}{2 6_i^2} + v(\eta) \right\} . w(x, \eta)$$

trong đó $\quad | w(x, \eta) - 1 | \leqslant (\varepsilon) \quad \forall (x, \eta)$

Giả thiết thêm rằng hàm $\sum_{i=1}^{k} U_i(x)$ là hàm tăng thực sự, khả vi theo **x** và nhận mọi giá trị trên R^1. Khi đó từ bất đẳng thức $E_\eta \varphi > 0\ \forall\eta$ và $|\varphi| \leqslant M\ \forall x$ ta có:

$$\varphi(x) \geqslant -\frac{2M\varepsilon}{1-\varepsilon} \quad \forall x \text{ trừ trên tập đếm được}$$

Chứng minh:

Đặt $$\bar{f}(x,\eta) = h(x)\exp\left\{\sum_{i=1}^{k}\xi_i U_i(x) - \sum_{i=1}^{k}\frac{U_i^2(x)}{2\sigma_i^2} + \bar{v}(\eta)\right\}$$

trong đó $\bar{v}(\eta)$ được xác định sao cho $\int_{-\infty}^{+\infty}\bar{f}(x,\eta)\,dx = 1$

Dễ dàng kiểm tra rằng: $\left|e^{v(\eta)-\bar{v}(\eta)} - 1\right| \leqslant \frac{\varepsilon}{1-\varepsilon}$,

$$\left|e^{\bar{v}(\eta)-v(\eta)} - 1\right| \leqslant \frac{\varepsilon}{1-\varepsilon}$$

Và $E_\eta\varphi = \int_{-\infty}^{+\infty}\varphi(x)\bar{f}(x,\eta)\,dx \geqslant \frac{-2M\varepsilon}{1-\varepsilon}$ trong đó $\varphi(x)$ là hàm sao cho $E_\eta\varphi \geqslant 0$

Vậy theo bổ đề 1 ta có:

$$\varphi(x) \geqslant -\frac{2M\varepsilon}{1-\varepsilon} \quad \forall x \text{ trừ trên tập đếm được.}$$

Định lý 1: Cho họ mũ 2k tham số $\eta = (\xi_1,.., \xi_k, \sigma_1,.., \sigma_k)$

Trong đó $-\infty < \xi_i < +\infty$, $0 < \sigma_i < 1$, $i = 1, 2,..., k$, với hàm mật độ cho bởi

$$f(x,\eta) = h(x)\exp\left\{\sum_{i=1}^{k}\xi_i U_i(x) - \sum_{i=1}^{k}\frac{U_i^2(x)}{2\sigma_i^2} + v(\eta)\right\}$$

Giả thiết thêm rằng hàm $\sum_{i=1}^{k} U_i(x)$ là tăng thực sự, khả vi theo **x** và nhận mọi giá trị trên đường thẳng thực. Khi đó thống kê

$$T(x) = (U_1(x), U_2(x),..., U_k(x))$$

là $(\varepsilon, \varepsilon)$ — đầy đủ với $\varepsilon > 0$ bất kỳ

Chứng minh: Giả sử $\varphi(x)$ là hàm bất kỳ mà

$$|E_\eta \varphi| \leqslant \varepsilon \quad \forall\eta$$

Khi đó $E_\eta(\varphi + \varepsilon) \geqslant 0\ \forall\eta$ và $E_\eta(\varphi - \varepsilon) \leqslant 0\ \forall\eta$

Áp dụng bổ đề 1 ta có $\varepsilon + \varphi \geqslant 0$ (P. h. k. n) và $\varepsilon - \varphi \geqslant 0$ (P. h. k. n)

Vậy $|\varphi(x)| \leqslant \varepsilon$ (P. h. k. n)

Định lý 2: Cho họ ε — mũ 2k tham số $\eta = (\xi_1,.., \xi_k, \sigma_1,.., \sigma_k)$

trong đó $-\infty < \xi_i < +\infty$, $0 < \sigma_i < 1$, $i = 1, 2,.., k$, với hàm mật độ cho bởi :

$$f(x, \eta) = h(x) \exp \sum_{i=1}^{k} \left\{ \xi_i U_i(x) - \sum_{i=1}^{k} \frac{U_i^2(x)}{2\sigma_i^2} + v(\eta) \right\} W(x, \eta)$$

trong đó

$$| W(x, \eta) - 1 | \leqslant \varepsilon \qquad \forall (x, \eta)$$

Giả thiết thêm rằng hàm $\sum_{i=1}^{k} U_i(x)$ là hàm tăng thực sự, khả vi theo x và nhận mọi giá trị trên R^1. Khi đó nếu gọi G_M là lớp tất cả các hàm $\varphi(x)$ đo được thỏa mãn

$$| \varphi(x) | \leqslant M \quad \forall x$$

thì thống kê

$$T(x) = (U_1(x), U_2(x), \ldots, U_k(x))$$

là $\left(G_M, \varepsilon, \frac{2(1+M)\varepsilon}{1-\varepsilon}\right)$ — đầy đủ

Chứng minh : Từ bất đẳng thức $| E_\eta \varphi | \leqslant \varepsilon \quad \forall \eta$

ta suy ra $E_\eta (\varepsilon + \varphi) \geqslant 0 \quad \forall \eta$ và $E_\eta (\varepsilon - \varphi) \geqslant 0 \quad \forall \eta$

Vì $| \varphi(x) | \leqslant M \quad \forall x$ nên $| \varphi(x) + \varepsilon | \leqslant M + \varepsilon$ và $| \varepsilon - \varphi(x) | \leqslant M + \varepsilon$

Vậy theo bổ đề 3 ta có :

$$\varphi(x) + \varepsilon \geqslant - \frac{2(M + \varepsilon)\varepsilon}{1 - \varepsilon}$$

$$\varepsilon - \varphi(x) \geqslant - \frac{2(M + \varepsilon)\varepsilon}{1 - \varepsilon}$$

Vậy

$$| \varphi(x) | \leqslant \frac{2(M + 1)\varepsilon}{1 - \varepsilon} \quad \text{(P. h. k. n)}$$

TÀI LIỆU THAM KHẢO

1. М. Поэв. Теория вероятностей. Изд. иностранной литературы Москва. 1962.

2. А. М. Катан, Н, В. Линник, С. Р. Рао. Характеризационные задачи математической статистики. Изд. наука. 1972.

Ле Нгок Быонг.

(G, ε, δ). ПОЛНАЯ СТАТИСТИКА

В этом статье мы дали определение (G, ε, δ) — полной статистики и доказали, что совокупность экспоненциальных функций, плотности которых имеют следующие формы:

(xem trang 24)