Tổng Hợp Câu Hỏi Thống Kê Toán Nâng Cao

Kiến thức căn bản

Định lý Polya. Cho $F_n(x)$ và $F(x)$ là các hàm PPXS, nếu

$$\lim_{n\to\infty} F_n(x) = F(x), \quad \forall x \in C(F) = \mathbb{R}$$

thì

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |F_n(x) - F(x)| = 0$$

Định lý Lindeberg - Levy. Cho $\xi_1, \xi_2, ...$ là dãy các BSNN i.i.d có moment cấp hai hữu hạn. Khi đó

$$\mathbb{P}\left\{\frac{\sqrt{n}}{\sigma}\left(\frac{1}{n}\sum_{i=1}^{n}(\xi_i - \mu)\right) < x\right\} = \Phi(x), \quad \forall x \in \mathbb{R}$$

trong đó

$$\begin{aligned} \Phi(x) \quad &: \quad \text{Hàm Gaussian} \\ \mu = E\xi_1 \quad &\text{và} \quad \sigma = \sqrt{\text{Var}\xi_1} \end{aligned}$$

Định lý Glivenko. Cho $X_1, ..., X_n$ là n BSNN i.i.d với BNN X thì với

$$F_n(x) = \frac{1}{n}\#\left\{i \ : \ 1 \le i \le n \text{ và } X_i < x\right\} \text{\_ hàm PP mẫu}$$

và

$$F(x) = \mathbb{P}(X < x) \text{\_ là hàm PP lý thuyết}$$

ta sẽ có

$$\mathbb{P}\left[\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \to 0\right] = 1, \quad \text{as } n \to \infty$$

nói cách khác

$$F_n \longrightarrow F \text{ hầu chắc chắn}$$

Bài tập ôn tập

Câu 1. Từ KGXS $(\Omega, \mathcal{A}, \mathbb{P})$, phép thử là rút ngẫu nhiên một phần tử ω theo luật xác suất P. CMR nếu có hai phép thử độc lập ξ_1 và ξ_2 thì

$$\left[\begin{array}{l}\mathbb{P}(A)=0\\ \mathbb{P}(A)=1\end{array}\right., \quad \forall A \in \mathcal{A}$$

[CMR không thể xây dựng hai phép thử độc lập trong không gian cơ sở Ω]

Chứng minh :

Từ không gian $(\Omega, \mathcal{A}, \mathbb{P})$, lấy phép thử thứ nhất là

$\xi_1(\omega)=\omega$; ta có $\mathbb{P}(\xi_1 \in A)=\mathbb{P}(A)$

Ta lấy phép thử thứ hai

ξ_2 với $\xi_2(\omega)=\omega'$; tương tự $\mathbb{P}(\xi_2 \in A)=\mathbb{P}(A)$

Ta có thể xem $\xi_2 = f(\xi_1)$. Giả sử nếu ξ_2 độc lập với ξ_1 thì

$$\Leftrightarrow \xi_2 \text{ và } \xi_1 \text{ cảm sinh bởi hai } \sigma\text{- đại số độc lập}$$

Ta có : ξ_1 cảm sinh $\mathcal{A}$ bởi $\xi_1^{-1}(\mathcal{A})=\mathcal{A}$

ξ_2 cảm sinh σ- đại số con của $\mathcal{A}$, đặt là $\mathcal{A}_2$

Khi đó $\mathcal{A}_2 \subset \mathcal{A}$. Vậy $\mathcal{A}$ độc lập với $\mathcal{A}_2$ và do đó $\mathcal{A}_2$ độc lập với chính nó

Do đó

$$\begin{aligned}\mathbb{P}(\mathcal{A}_2 \cap \mathcal{A}_2) &= \mathbb{P}(\mathcal{A}_2).\mathbb{P}(\mathcal{A}_2)\\ \Rightarrow \mathbb{P}(\mathcal{A}_2) &= \mathbb{P}(\mathcal{A}_2).\mathbb{P}(\mathcal{A}_2) \Rightarrow \left[\begin{array}{l}\mathbb{P}(\mathcal{A}_2)=0\\ \mathbb{P}(\mathcal{A}_2)=1\end{array}\right.\end{aligned}$$

Mà với mọi $A \in \mathcal{A}_2 = \xi_2^{-1}(\mathcal{A})$ thì $\left[\begin{array}{l}\mathbb{P}(A)=0\\ \mathbb{P}(A)=1\end{array}\right.$

Lấy $A \in \mathcal{A}$, ta có

$$\mathbb{P}(A)=\mathbb{P}[\xi_2(\omega)\in A]=\mathbb{P}\{\omega:\ \xi_2(\omega)\in A\}=\mathbb{P}\left\{\xi_2^{-1}(A)\right\}$$

Do $\xi_2^{-1}(\mathcal{A}) \in \mathcal{A}_2$ nên

$$\left[\begin{array}{l}\mathbb{P}\left(\xi_2^{-1}(A)\right)=0\\ \mathbb{P}\left(\xi_2^{-1}(A)\right)=1\end{array}\right. \Rightarrow \left[\begin{array}{l}\mathbb{P}(A)=0\\ \mathbb{P}(A)=1\end{array}\right.$$

[***Kết luận :***

Nếu độ đo $\mathbb{P}$ là độ đo không tầm thường thì không thể xây dựng hai phép thử độc lập.

]

Câu 2. CMR η hội tụ hầu chắc chắn đến ρ [ƯLVM] trong đó

$$\eta = \frac{m_{11}}{\sqrt{m_{02}.m_{20}}} \text{ và } \rho = \frac{\mu_{11}}{\sqrt{\mu_{02}.\mu_{20}}}$$

trong đó η là hệ số tương quan mẫu và ρ là hệ số tương quan lý thuyết.

Chứng minh Ta có

$$m_{ij} = \sum_{k=1}^{n} \frac{1}{n} \left(X_k - \overline{X}\right)^i \left(Y_k - \overline{Y}\right)^j \text{ và } \mu_{ij} = \mathbb{E}\left[(X - \mathbb{E}X)^i (Y - \mathbb{E}Y)^j\right]$$

Xét

$$m_{11} = \sum_{k=1}^{n} \frac{1}{n} \left(X_k - \overline{X}\right)\left(Y_k - \overline{Y}\right) = \sum_{k=1}^{n} \frac{1}{n} X_k Y_k - \left(\sum_{k=1}^{n} \frac{1}{n} X_k\right)\left(\sum_{k=1}^{n} \frac{1}{n} Y_k\right) \quad (1)$$

$$m_{20} = \sum_{k=1}^{n} \frac{1}{n} \left(X_k - \overline{X}\right)^2 = \sum_{k=1}^{n} \frac{1}{n} X_k^2 - \left(\sum_{k=1}^{n} \frac{1}{n} X_k\right)^2 \quad (2)$$

$$m_{02} = \sum_{k=1}^{n} \frac{1}{n} \left(Y_k - \overline{Y}\right)^2 = \sum_{k=1}^{n} \frac{1}{n} Y_k^2 - \left(\sum_{k=1}^{n} \frac{1}{n} Y_k\right)^2 \quad (3)$$

Mà (X_k, Y_k) là các quan trắc độc lập cho điểm ngẫu nhiên (X, Y) nên

$X_1, X_2, ..., X_n$ độc lập và có cùng phân phối với X

$Y_1, Y_2, ..., Y_n$ độc lập và có cùng phân phối với Y

$X_1Y_1, .., X_nY_n$ độc lập và có cùng phân phối với XY

Vậy theo luật mạnh số lớn Kolmogorov ta có

$$\sum_{k=1}^{n} \frac{1}{n} X_k \longrightarrow \mathbb{E}X \qquad \text{h.c.c} \quad (4)$$

$$\sum_{k=1}^{n} \frac{1}{n} Y_k \longrightarrow \mathbb{E}Y \qquad \text{h.c.c} \quad (5)$$

$$\sum_{k=1}^{n} \frac{1}{n} X_k Y_k \longrightarrow \mathbb{E}XY \qquad \text{h.c.c} \quad (6)$$

Từ đó ta được

$$m_{11} \to \mathbb{E}XY - \mathbb{E}X.\mathbb{E}Y = \mu_{11} \qquad \text{h.c.c}$$

Tương tự

$$m_{20} \to \mathbb{E}\left(X^2\right) - (\mathbb{E}X)^2 = \mu_{20} \qquad \text{h.c.c}$$

$$m_{02} \to \mathbb{E}\left(Y^2\right) - (\mathbb{E}Y)^2 = \mu_{02} \qquad \text{h.c.c}$$

Vậy

$$\eta = \frac{m_{11}}{\sqrt{m_{02}.m_{20}}} \to \frac{\mu_{11}}{\sqrt{\mu_{02}.\mu_{20}}} = \rho \qquad \text{h.c.c}$$

Biện luận phương sai mẫu.

Nếu Var$X > 0$ và Var$Y > 0$ thì ρ_{XY} nhỏ nhất sinh bởi HCN [σ-đs Borel]. Khuếch trương độ đo trên tất cả các σ- đs Borel. xác định và khi đó $P\left(\lim\limits_{n\to\infty} r_{XY} = \rho_{XY}\right) = 1$

Hơn nữa, do Var$X = \mathbb{E}\left(X-\alpha\right)^2$ với $\alpha = \mathbb{E}X$ nên nếu

Var$X > 0$ tức là $\mathbb{E}\left(X-\alpha\right)^2 > 0$.

Nếu $\mathbb{E}\left(X-\alpha\right)^2 = 0$, đặt $Z = \left(X-\alpha\right)^2; Z \geq 0$ thì tồn tại dãy các hàm đơn giản không âm Z_k sao cho

$$0 \leq Z_1 \leq Z_2 \leq ... \leq Z_k \leq ... \text{ và } Z = \lim_{k\to\infty} Z_k$$

hay $0 \leq \mathbb{E}Z_k \uparrow \mathbb{E}Z$

Vậy $\mathbb{E}Z = 0 \Leftrightarrow \mathbb{E}Z_k = 0, \qquad \forall k$ với $Z_k = c_1 I_{A_1} + ... + c_m I_{A_m}$ và $c_1, ..., c_m > 0$

Ta có

$$\begin{aligned} \mathbb{E}Z_k &= c_1 P\left(A_1\right) + ... + c_m P\left(A_m\right) = 0 \\ \Rightarrow &\quad P\left(A_1\right) = ... = P\left(A_m\right) = 0 \\ \Rightarrow &\quad Z_k = 0 \text{ h.c.c} \qquad \forall k \\ \Rightarrow &\quad Z = 0 \text{ h.c.c} \end{aligned}$$

Vậy Var$Z = 0 \Rightarrow \mathbb{E}\left(X-\alpha\right)^2 = 0$ h.c.c nên $\left(X-\alpha\right) = 0$ h.c.c do đó X là hằng [Vô lý]

Kết luận : Nếu X và Y không phải hằng số thì hệ số tương quan mẫu sẽ hội tụ đến hệ số tương quan lý thuyết.

Câu 3. Cho $X_1, ..., X_n$ là dãy i.i.d BSNN với

$$\alpha_1 = \mathbb{E}X_1,\ \alpha_2 = \mathbb{E}X_1^2;\ \sigma^2 = \mathrm{Var}X_1;\ \overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ và } m_2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X_n}\right)^2$$

trong đó α_1 là kỳ vọng lý thuyết, $\overline{X}_n$ là kỳ vọng mẫu (n BSNN) và m_2 là phương sai mẫu

CMR hàm phân phối của $\dfrac{\sqrt{n-1}\left(\overline{X_n} - \alpha_1\right)}{\sqrt{m_2}}$ hội tụ đều đến $\Phi(x)$ trên $\mathbb{R}$.

Từ đó suy ra KTC xấp xỉ cho α_1 với hệ số tin cậy $1 - \dfrac{P}{100}$

Chứng minh :

Ta sẽ ch.m qua các bước như sau:

Viết lại

$$\frac{\sqrt{n-1}\left(\overline{X_n} - \alpha_1\right)}{\sqrt{m_2}} = \frac{\overline{X_n} - \alpha_1}{\sigma/\sqrt{n}} : \frac{\sqrt{\frac{n}{n-1} m_2}}{\sigma}$$

ta thấy m_2 =Momen mẫu cấp 2 −(Momen mẫu cấp 1)^2 hội tụ đến $\left(\alpha_2 - \alpha_1^2\right) = \sigma^2$

Vậy $\sqrt{m_2} \longrightarrow \sqrt{\alpha_2 - \alpha_1^2}$ (h.c.c) nên $\dfrac{\sigma}{\sqrt{m_2}} \longrightarrow 1$ (h.c.c) (1)

Dễ thấy $\lim\limits_{n\to\infty}\sqrt{\dfrac{n}{n-1}} = 1$ (2)

Vậy

$$\sqrt{\frac{n}{n-1}} : \frac{\sigma}{\sqrt{m_2}} \to 1 \text{ (h.c.c)}$$

Mặt khác theo "Định lý giới hạn trung tâm" Linderberg-Levy với $X_1, .., X_n$ i.i.d, ta được

$\mathbb{P}\left(\dfrac{\sqrt{n}}{\sigma}\left(\overline{X_n} - \alpha_1\right) < x\right)$ hội tụ yếu đến hàm Gaussian $\Phi(x)$ (3)

where $\alpha_1 = \mathbb{E}X_1$ và $\sigma = \sqrt{\mathrm{Var}X_1}$

Áp dụng định lý hội tụ cho $(1), (2)$ và (3) ta có

$$\left.\begin{array}{l} \dfrac{\overline{X_n} - \alpha_1}{\sigma/\sqrt{n}} \to \Phi(x) \\ \dfrac{\sqrt{\left(\frac{n}{n-1}\right) m_2}}{\sigma} \to 1 \end{array}\right\} \Rightarrow \frac{\sqrt{n-1}\left(\overline{X_n} - \alpha_1\right)}{\sqrt{m_2}} \text{ h.t. yếu } \Phi(x)$$

Áp dụng Định lý Polya. Đặt

$$F_n(x) = \mathbb{P}\left\{\left(\frac{\sqrt{n-1}\left(\overline{X_n} - \alpha_1\right)}{\sqrt{m_2}}\right) < x\right\} \text{ và } F(x) = \Phi(x)$$

là các hàm PPXS; ta được

$$\lim_{n\to\infty}\sup_{x\in\mathbb{R}}|F_n(x) - F(x)| = 0$$

hàm phân phối của $\dfrac{\sqrt{n-1}\left(\overline{X_n}-\alpha_1\right)}{\sqrt{m_2}}$ hội tụ đều đến hàm Gaussian $\Phi\left(x\right).$

Step 5. Từ đó, ta có

$$\mathbb{P}\left(\left|\frac{\sqrt{n-1}\left(\overline{X_n}-\alpha_1\right)}{\sqrt{m_2}}\right|\leq z_{P/100}\right)=\Phi\left(z_{P/100}\right)=1-\frac{P}{100}$$

với $z_{p/100}$ là phân vị mức $p/100$.

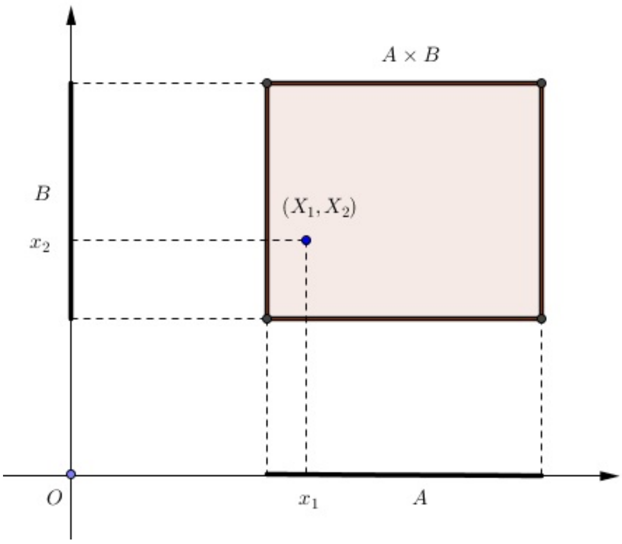
Vậy KTC với hệ số tin cậy $1-p/100$ cho α_1 là

$$\left[\overline{X_n}-\frac{\sqrt{m_2}}{\sqrt{n-1}}.z_{P/100},\ \overline{X_n}+\frac{\sqrt{m_2}}{\sqrt{n-1}}.z_{P/100}\right]$$

Câu 4. Cho X_1 và X_2 là các BSNN độc lập cùng phân phối có hàm mật độ f đối với độ đo Lebesgue trên đường thẳng. Dùng thống kê thứ tự biến (X_1, X_2) thành (X_1^*, X_2^*) với $X_1^* \leq X_2^*$. Hãy tìm hàm mật độ của điểm (X_1^*, X_2^*) đối với độ đo Lebesgue trong mặt phẳng.

Giải.

Ta có thống kê thứ tự $T(X_1, X_2) = (X_1^*, X_2^*)$ với $X_1^* \leq X_2^*$. Trong đó (X_1, X_2) là một điểm ngẫu nhiên trong mặt phẳng

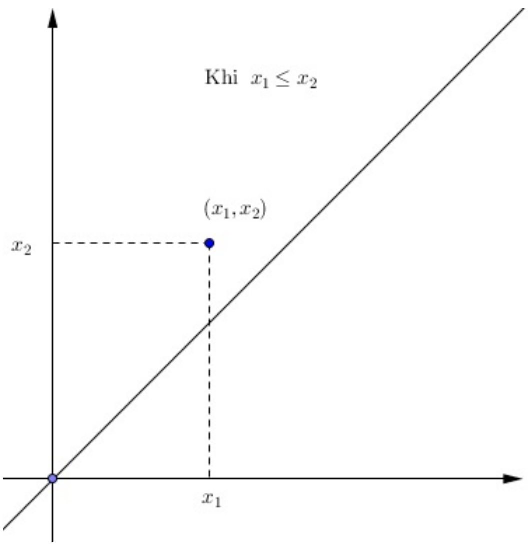


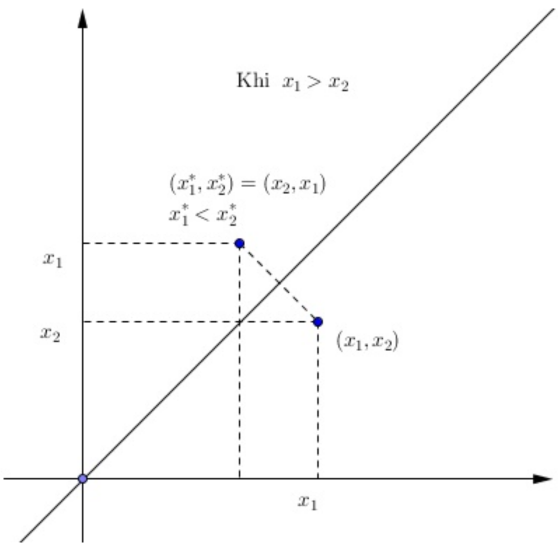
Xét tất cả các σ đại số sinh bởi những HCN bằng cách lấy $\sigma-$đs nhỏ nhất sinh bởi HCN [σ-đs Borel]. Khuếch trương độ đo trên tất cả các σ- đs Borel.

Xét hình chữ nhật $A \times B$, ta có $\mathbb{P}\{(X_1, X_2) \in A \times B\} = \mathbb{P}[X_1 \in A, X_2 \in B] = \mathbb{P}_{X_1} \otimes \mathbb{P}_{X_2}$

Xét ánh xạ $T : \mathbb{R}^2 \to \mathcal{D} = \{(X_1, X_2) : X_1 \leq X_2\}$ xác định bởi

$$T(x_1, x_2) = \begin{cases} (x_1, x_2) & \text{nếu} \quad x_1 \leq x_2 \\ (x_2, x_1) & \qquad\quad x_1 > x_2 \end{cases}$$





Giả sử $\mathbb{P}_{X_1} \ll l$ và $\mathbb{P}_{X_2} \ll l$ khi đó theo Đl R-N.

$$\frac{d\mathbb{P}_{X_1}}{dl} = f_1\left(x_1\right) \text{ và } \frac{d\mathbb{P}_{X_1}}{dl} = f_2\left(x_2\right)$$

Ta có

$$\begin{aligned}
\mathbb{P}\left\{\left(X_1, X_2\right) \in \square\right\} &= \mathbb{P}\left[\left(x_1 \leq X_1 \leq x_1 + dx_1\right) \cap \left(x_2 \leq X_2 \leq x_2 + dx_2\right)\right] \\
&= \mathbb{P}\left(x_1 \leq X_1 \leq x_1 + dx_1\right).\mathbb{P}\left(x_2 \leq X_2 \leq x_2 + dx_2\right) \\
&= f_1\left(x_1\right) dx_1.f_2\left(x_2\right) dx_2 = f_1\left(x_1\right) f_2\left(x_2\right) dx_1 dx_2
\end{aligned}$$

Do đó, ta có

$$\begin{aligned}
\mathbb{P}\left\{\left(X_1, X_2\right) \in \square\right\} &= f_1\left(x_1\right) f_2\left(x_2\right) \times \text{Diện tích HCN vi phân} \\
\Rightarrow \frac{dP\left(X_1, X_2\right)}{dl^2} &= f_1(x_1) f_2(x_2)
\end{aligned}$$

Xét

$$\left\{\left(X_1^*, X_2^*\right) \in \square\right\} = \left\{\left(X_1, X_2\right) \in \square\right\} \sqcup \left\{\left(X_2, X_1\right) \in \square\right\}$$

Vậy

$$\begin{aligned}
\mathbb{P}\left\{\left(X_1^*, X_2^*\right) \in \square\right\} &= \mathbb{P}\left\{\left(X_1, X_2\right) \in \square\right\} + \mathbb{P}\left\{\left(X_2, X_1\right) \in \square\right\} \\
&= f_1\left(x_1^*\right) f_2\left(x_2^*\right) S_\square + f_1\left(x_2^*\right) f_2\left(x_1^*\right) S_\square
\end{aligned}$$

Do các HCN trong biểu thức là các HCN vi phân nên theo định lý Randon Nicodym, ta được

$$\frac{d\mathbb{P}\left\{\left(X_1^*, X_2^*\right)\right\}}{dl^2}.S_\square = \mathbb{P}\left\{\left(X_1^*, X_2^*\right) \in \square\right\}$$

trong đó l^2 là độ đo Lebesgue trên mặt phẳng. Do đó

$$\frac{d\mathbb{P}\left\{\left(X_1^*, X_2^*\right)\right\}}{dl^2}\left(x_1^*, x_2^*\right) = f_1\left(x_1^*\right) f_2\left(x_2^*\right) + f_1\left(x_2^*\right) f_2\left(x_1^*\right)$$

Mà X_1, X_2 độc lập có cùng phân phối nên (X_1, X_2) và (X_2, X_1) sẽ có phân phối như nhau

$$\frac{d\mathbb{P}\left\{\left(X_1^*, X_2^*\right)\right\}}{dl^2}\left(x_1^*, x_2^*\right) = 2f\left(x_1^*\right) f\left(x_2^*\right)$$

Vậy $\mathbb{P}_{X_1^*, X_2^*}$ có hàm mật độ theo độ đo l^2 là $2f\left(x_1^*\right) f\left(x_2^*\right)$ với f là hàm mật độ của X_1 hoặc X_2theo l.

Câu 5. Trong cấu trúc thống kê $(\mathcal{X}, \mathcal{B}, \mathbb{P}_{X,\theta}, \theta \in \Theta)$; giả sử với một độ đo σhữu hạn μ trên $\mathcal{B}$, $P_{X,\theta} \ll \mu, \quad \forall\theta \in \Theta$.

CMR nếu T là thống kê đủ cho Θ, thì ước lượng hợp lý cho tham ẩn θ sẽ là một hàm của thống kê đủ T và lượng thông tin Fisher $I_X(\theta)$ chỉ phụ thuộc vào phân phối của T. Áp dụng cho dãy các quan trắc về gieo một đồng tiền n lần có xác suất xuất hiện trong một lần gieo là $\theta \in (0,1)$; cmr số lần sấp ν là một thống kê đủ; từ đó để tìm KTC cho θ, ta chỉ cần xét tần suất $\frac{\nu}{n}$ và KTC sẽ là xấp xỉ của tần suất này.

Chứng minh Từ cấu trúc thống kê $(\mathcal{X}, \mathcal{B}, \mathbb{P}_{X,\theta}, \theta \in \Theta)$; với mọi $\theta \in \Theta$ $P_{X,\theta} \ll \mu$ trong đó μ là độ đo hữu hạn trên $\mathcal{B}$ và

$T : (\mathcal{X}, \mathcal{B}) \to (\mathcal{J}, \mathcal{C})$ thì $T^{-1}(\mathcal{C}) \subset \mathcal{B}$

Theo tiêu chuẩn tách Neyman, khi T là TK đủ cho Θ thì ta sẽ có

$$\frac{dP_{X,\theta}}{d\mu}(x) = h(x)g_\theta[T(x)], \qquad \forall\theta \in \Theta$$

với $h(x)$ là hàm không âm trên $\mathcal{B}$ đo được và hàm này không phụ thuộc vào θ.

$g_\theta[T(x)]$ là hàm không âm $\mathcal{C}$ đo được.

Theo đạo hàm Randon-Nikodym thì $\frac{dP_{X,\theta}}{d\mu}$ chính là hàm MĐXS của phân phối XS P_X theo μ tức là

$$\frac{dP_{X,\theta}}{d\mu}(x) = f(x,\theta)$$

Khi đó, $f(x,\theta) = h(x)\, g_\theta[T(x)]$ ước lượng hợp lý cho θ

$$\begin{aligned} \log f(x,\theta) &= \log h(x) + \log g_\theta[T(x)] \\ \Rightarrow \frac{\partial}{\partial\theta}\log f(X,\theta) &= \frac{\frac{\partial}{\partial\theta} g_\theta[T(X)]}{g_\theta[T(X)]} \end{aligned}$$

Khi $\frac{\partial}{\partial\theta}\log f(X,\theta) = 0$ thì $\frac{\partial}{\partial\theta} g_\theta[T(X)] = 0$

Khi cố định X và cho θ chạy thì ước lượng $\hat{\theta}$ là hàm theo T.

Vậy ƯL hợp lý cho θ sẽ là một hàm cho TK đủ T.

Lượng tin Fisher $I_X(\theta) = \mathbb{E}_\theta\left[\frac{\partial}{\partial\theta}\log f(X,\theta)\right]^2 \Rightarrow \mathbb{E}_\theta\left[\frac{\partial}{\partial\theta}\log g_\theta[T(X)]\right]^2$

Do θ là tham số chưa biết nên $\mathbb{E}_\theta\left[\frac{\partial}{\partial\theta}\log g_\theta[T(X)]\right]$ chỉ phụ thuộc vào phân phối của T. Hay $I_X(\theta)$ chỉ phụ thuộc vào phân phối của T.

Áp dụng. Gieo n lần 1 đồng tiền có XS xuất hiện sấp trong 1 lần gieo là $\theta \in (0,1)$

Gọi $X_i = \begin{cases} 1 & \text{"xuất hiện mặt sấp với xác suất } \theta\text{"} \\ 0 & \text{" xuất hiện mặt ngửa với xác suất } 1-\theta\text{"} \end{cases}$ và

$\nu = \sum_{i=1}^{n} X_i$ là tổng số lần xuất hiện mặt sấp trong n lần gieo.

Vì $X_1,...,X_n$ là n BSNN độc lập và có cùng phân phối của X_i cho bởi

$$P\left(X_i=1\right)=\theta \text{ và } P\left(X_i=0\right)=1-\theta$$

Cho nên phân phối xác suất của $X=\left(X_1,...,X_n\right)$ là

$$\begin{aligned}
\frac{dP_{X,\theta}}{d\mu}\left(x\right)=P\left(X=x\right) &= P\left(X_1=x_1,X_2=x_2,...,X_n=x_n\right)\\
&= P\left(X_1=x_1\right)....P\left(X_n=x_n\right)\\
&= \theta^{x_1}\left(1-\theta\right)^{1-x_1}....\theta^{x_n}\left(1-\theta\right)^{1-x_n}\\
&= \theta^{x_1+...+x_n}\left(1-\theta\right)^{n-(x_1+...+x_n)}\\
&= \left(\frac{\theta}{1-\theta}\right)^{x_1+...+x_n}\left(1-\theta\right)^n\\
&= h\left(x\right).g_\theta\left(T\left(x\right)\right)
\end{aligned}$$

trong đó $h\left(x\right)\equiv 1$,

$g_\theta\left[T\left(x\right)\right]=\left(1-\theta\right)^n\left(\frac{\theta}{1-\theta}\right)^{x_1+...+x_n}, \quad \forall\theta\in\Theta$ và

$T\left(x\right)=\sum x_i$ với $x=\left(x_1,...,x_n\right)$

Vậy tiêu chuẩn tách Neyman được thỏa và $\nu=T\left(X\right)$ là thống kê đủ.

Do $X_1,...,X_n$ độc lập có cùng phân phối nên

$$\begin{aligned}
\mathbb{E}X_1 &= ...=\mathbb{E}X_n=\theta\\
\mathrm{Var}X_1 &= ...=\mathrm{Var}X_n=\theta\left(1-\theta\right)
\end{aligned}$$

Áp dụng định lý giới hạn trung tâm Linderberg-Levy, ta được

$$\begin{aligned}
P\left(\sqrt{\frac{n}{\theta\left(1-\theta\right)}}\left(\frac{X_1+...+X_n}{n}-\theta\right)<x\right) &\rightarrow \Phi\left(x\right) \text{ khi } n\rightarrow\infty\\
\Rightarrow P\left(\sqrt{\frac{n}{\theta\left(1-\theta\right)}}\left(\frac{\nu}{n}-\theta\right)<x\right)\rightarrow\Phi\left(x\right) & \qquad \text{khi } n\rightarrow\infty
\end{aligned}$$

Vậy KTC cho θ là một hàm của tần suất $\dfrac{\nu}{n}$ và

$$P\left(-z_p\le\sqrt{\frac{n}{\theta\left(1-\theta\right)}}\left(\frac{\nu}{n}-\theta\right)\le z_p\right)=1-p$$

do đó với mức ý nghĩa p phần trăm thì ta được KTC cho θ là nghiệm của BPT

$$\nu^2-2\nu n\theta-n^2\theta^2\le n\theta\left(1-\theta\right)z_p^2$$

trong đó z_p là phân vị mức p phần trăm.

Câu 6. Cho $X_1, ..., X_n$ là n quan trắc ĐL cùng PP với hàm mật độ $f(x) = \left[\pi\left(1 + (x-\mu)^2\right)\right]^{-1}$. CMR $\overline{X} = n^{-1}\sum_1^n X_i$ không phải là UL vững cho μ. Nhắc lại : với hàm mật độ $\left[\pi\left(1+x^2\right)\right]^{-1}$ thì hàm đặc trưng là $e^{-|t|}$.

Giải.

Áp dụng hàm đặc trưng của BNN ξ là $\varphi(t) = \mathbb{E}e^{it\xi}$. $\xi_1, ..., \xi_n$ là n BSNN độc lập thì hàm đặc trưng

$$\varphi_{\xi_1+...+\xi_n}(t) = \prod_{i=1}^n \varphi_{\xi_i}(t)$$

Ta có hàm mật độ ứng với hàm đặc trưng $\varphi(t) = e^{-|t|}$ là $\dfrac{1}{\pi(1+x^2)}$

Nếu ξ có hàm đặc trưng là $\varphi(t) = \mathbb{E}e^{it\xi}$ thì $\lambda\xi + \mu$ có hàm đặc trưng là $\mathbb{E}e^{i\mu t + it\lambda\xi}$ với μ, λ là các hằng số.

$$\varphi_{\lambda\xi+\mu}(t) = e^{i\mu t}\varphi_\xi(\lambda t)$$

Nếu ξ có hàm đ trưng là $e^{-|t|}$ và có hàm mật độ là $\dfrac{1}{\pi(1+x^2)}$ thì $\lambda\xi + \mu$ có hàm đặc trưng là $e^{i\mu t - \lambda|t|}$ và có hàm pp

$$F_{\lambda\xi+\mu}(x) = P(\lambda\xi + \mu < x) = P\left(\xi < \frac{x-\mu}{\lambda}\right)$$

do $\lambda > 0$. Vậy

$$F_{\lambda\xi+\mu}(x) = F_\xi\left(\frac{x-\mu}{\lambda}\right)$$

nên

$$f_{\lambda\xi+\mu}(x) = \frac{1}{\lambda} f_\xi\left(\frac{x-\mu}{\lambda}\right)$$

Do đó $\lambda\xi + \mu$ có hàm MĐ là

$$\frac{1}{\lambda\pi\left[1 + \left(\dfrac{x-\mu}{\lambda}\right)^2\right]} = \frac{\lambda}{\pi\left[\lambda^2 + (x-\mu)^2\right]}$$

là hàm MĐ của PP Cauchy. Hàm đặc trưng chung của $X_1, ..., X_n$ là $e^{i\mu t - \lambda|t|}$ nên

$$\begin{aligned} \varphi_{X_1+...+X_n}(t) &= e^{i\mu n t - \lambda|nt|} \\ \varphi_{\overline{X}}(t) &= e^{i\mu t - \lambda|t|} \end{aligned}$$

suy ra $\overline{X}$ và các X_i có cùng hàm pp.

Cần CM $\overline{X}$ không phải là hàm UL yếu cho μ.

Theo Luật yếu số lớn Khintchine thì

$$\lim_{n\to\infty} P\left(\left|\overline{X}-\mu\right|<\epsilon\right)=1, \qquad \forall\epsilon>0$$

Bằng phản chứng, ta xét : Giả sử $\overline{X}$ là ULVY cho μ, i.e

$$\lim_{n\to\infty} P\left(\left|\overline{X}-\mu\right|<\epsilon\right)=1, \qquad \forall\epsilon>0$$

Với mọi $k=1,2,...$ thì $\lim\limits_{n\to\infty} P\left(\left|\overline{X}-\mu\right|<\dfrac{1}{k}\right)=1$.

Ta có

$$\left|\overline{X}-\mu\right|=0 \Leftrightarrow \left|\overline{X}-\mu\right|<\frac{1}{k}, \qquad \forall \text{k là số nguyên dương.}$$

nên

$$\left\{\left|\overline{X}-\mu\right|=0\right\}=\cap_{k=1}^{\infty}\left\{\left|\overline{X}-\mu\right|<\frac{1}{k}\right\}$$

Áp dụng Đ.lý : Nếu $A_1\supset A_2\supset ...\supset A_k\supset ....$ thì $\lim\limits_{k\to\infty} P\left(A_k\right)=P\left(\cap_{k=1}^{\infty}A_k\right)$ với

$$A_k=\left\{\left|\overline{X}-\mu\right|<\frac{1}{k}\right\}\subset\Omega$$

trong đó $\Omega\in\sigma$ đại số $\mathcal{A}\Rightarrow A_k$ là một biến cố.

Nếu $\left|\overline{X}-\mu\right|<\dfrac{1}{k+1}$ thì $\left|\overline{X}-\mu\right|<\dfrac{1}{k}$

Do đó A_{k+1} xảy ra thì A_k xảy ra hay $A_{k+1}\subset A_k$ và

$$\begin{aligned} P\left\{\overline{X}-\mu=0\right\} &= P\left(\cap_{k=1}^{\infty}\left\{\left|\overline{X}-\mu\right|<\frac{1}{k}\right\}\right) \\ &= \lim_{n\to\infty} P\left\{\left|\overline{X}-\mu\right|<\frac{1}{k}\right\}=1 \end{aligned}$$

Vậy $P\left(\overline{X}=\mu\right)=1$.

Mà $P\left(X_1=\mu\right)=P\left(\overline{X}=\mu\right)$ do $X_1,..,X_n$ và $\overline{X}$ có cùng phân phối suy ra $P\left(X_1=\mu\right)=1$

$\Rightarrow X_1$ là hằng số hầu chắc chắn mà X_1 có PP Cauchy nên X_1 không phải hằng số $\Rightarrow\overline{X}$ ko phải ULVY cho $\mu\Rightarrow\overline{X}$ ko phải UL vững cho μ [ko vững yếu cũng sẽ ko vững mạnh].