XÁC SUẤT THỐNG KÊ

Tôn Thất Tú

Đà Nẵng, 2019

Chương 5. Ước lượng tham số

1. Ước lượng điểm

1.1 Định nghĩa

- Cho mẫu ngẫu nhiên $\{X_1, X_2, ..., X_n\}$ từ tổng thể có phân phối phụ thuộc vào tham số θ chưa biết. Khi đó, một hàm (thống kê) $\hat{\theta} = \hat{\theta}(X_1, ..., X_n)$ được gọi là một *ước lượng* của tham số θ.
- Với một mẫu giá trị cụ thể $\{x_1, x_2, ..., x_n\}$ ta thu được một giá trị cụ thể của $\hat{\theta}$. Khi đó, giá trị $\hat{\theta}(x_1, ..., x_n)$ được gọi là *ước lượng điểm* của tham số θ dựa trên mẫu giá trị $\{x_1, x_2, ..., x_n\}$.

1.2 Phân loại ước lượng

- Ước lượng $\hat{\theta} = \hat{\theta}(X_1, ..., X_n)$ được gọi là *ước lượng không chệch* của tham số θ nếu $E(\hat{\theta}) = \theta$.
Trong trường hợp ngược lại thì ta gọi $\hat{\theta}$ được gọi là *ước lượng chệch* và giá trị $b(\theta) = E(\hat{\theta}) - \theta$ được gọi là *độ chệch của ước lượng*.
- Ước lượng $\hat{\theta} = \hat{\theta}(X_1, ..., X_n)$ được gọi là *ước lượng không chệch tiệm cận* của tham số θ nếu

$$\lim_{n\to+\infty} E(\hat{\theta}) = \theta.$$

- Cho $\hat{\theta}_1$ và $\hat{\theta}_2$ là hai ước lượng không chệch của tham số θ.

$$D(\hat{\theta}_1) = E(\hat{\theta}_1 - E\hat{\theta}_1)^2 = E(\hat{\theta}_1 - \theta)^2 \text{ và } D(\hat{\theta}_2) = E(\hat{\theta}_2 - E\hat{\theta}_2)^2 = E(\hat{\theta}_2 - \theta)^2$$

Ta nói ước lượng $\hat{\theta}_1$ *hiệu quả hơn* ước lượng $\hat{\theta}_2$ nếu $D(\hat{\theta}_1) < D(\hat{\theta}_2)$.
- Ước lượng $\hat{\theta}$ của θ là ước lượng không chệch và có phương sai $D(\hat{\theta})$ bé nhất được gọi là *ước lượng tốt nhất*.

Ước lượng không chệch của trung bình và phương sai

Định lý: Cho biến ngẫu nhiên X của một tổng thể có $E(X) = \mu, D(X) = \sigma^2$. Giả sử $\{X_1, X_2, ..., X_n\}$ là mẫu ngẫu nhiên của X. Khi đó:

i) $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ là ước lượng không chệch của μ.

ii) $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ là ước lượng không chệch của σ^2.

Nhận xét: Có thể chứng minh được rằng:

$$E(S_*^2) = \frac{n-1}{n}\sigma^2 \text{ với } S_*^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Điều này có nghĩa S_*^2 là ước lượng chệch (không chệch tiệm cận) của σ^2, do đó nó «không được chọn» làm phương sai mẫu.

Ước lượng không chệch của tỉ lệ

Định lý: Cho biến ngẫu nhiên X có phân phối Bernoulli với tham số p. Gọi $(X_1, X_2, ..., X_n)$ là mẫu ngẫu nhiên của X. Khi đó:

$$\hat{P} = \frac{X_1 + X_2 + ... + X_n}{n}$$

là một ước lượng không chệch của tham số p.

Ý nghĩa: Nếu dấu hiệu A có tỉ lệ p chưa biết, và trong mẫu điều tra kích thước n có k cá thể mang dấu hiệu A, thì tần suất $f_n = k/n$ là một ước lượng không chệch của p.

Ví dụ 1

- Cho biến ngẫu nhiên X có kỳ vọng μ và phương sai σ^2 chưa biết. Để ước lượng cho μ và σ^2 người ta tiến hành điều tra mẫu kích thước 200 và tính được trung bình và phương sai chuẩn mẫu:

$$\bar{x} = 125,8; s^2 = 2,76$$

Khi đó, ta có thể xem $\mu \approx 125,8$ và $\sigma^2 \approx 2,76$.

- Để ước lượng tỉ lệ p cử tri ủng hộ cho ứng cử viên A, người ta khảo sát ngẫu nhiên 4000 người thì có 2640 người ủng hộ ứng cử viên này. Như vậy, ta có thể xem tỉ lệ ủng hộ ứng cử viên A xấp xỉ:

$$p \approx 2640/4000 = 0,66$$

1.3 Phương pháp hợp lý cực đại

Định nghĩa:

- Cho $\{k_1, k_2, ..., k_n\}$ là mẫu ngẫu nhiên kích thước n của phân phối rời rạc X có hàm xác suất $p(k, \theta)$ với θ là tham số chưa biết. Hàm hợp lí $L(\theta)$ là hàm được định nghĩa như sau:

$$L(\theta) = \prod_{i=1}^{n} p(k_i, \theta).$$

- Nếu $\{x_1, x_2, ..., x_n\}$ là mẫu ngẫu nhiên kích thước n của phân phối liên tục có hàm mật độ xác suất $f(x, \theta)$ với θ là tham số chưa biết thì hàm hợp lí $L(\theta)$ là hàm được định nghĩa như sau:

$$L(\theta) = \prod_{i=1}^{n} f(x_i, \theta).$$

Nhận xét: Hàm hợp lí $L(\theta)$ chính là hàm xác suất (mật độ xác suất) đồng thời của mẫu ngẫu nhiên.

Định nghĩa: Cho hàm hợp lí $L(\theta)$ với tham số θ chưa biết. Nếu θ_{ML} là giá trị tham số thỏa mãn $L(\theta_{ML}) \geq L(\theta)$ với mọi θ thì θ_{ML} được gọi là *ước lượng hợp lí cực đại* của tham số θ.

Nhận xét:
i) Trong thực hành để tìm ước lượng hợp lý cực đại ta giải phương trình:

$$L'(\theta) = 0 \text{ hoặc } [ln(L(\theta))]' = 0$$

ii) Trường hợp θ là một vectơ các tham số, tức là $\theta = (\theta_1, ..., \theta_m)$, để tìm ước lượng hợp lý cực đại ta giải hệ phương trình sau đây:

$$\begin{cases} \dfrac{\partial L(\theta)}{\partial \theta_1} = 0, \\ \cdots \\ \dfrac{\partial L(\theta)}{\partial \theta_m} = 0, \end{cases} \quad \text{hoặc} \quad \begin{cases} \dfrac{\partial \ln L(\theta)}{\partial \theta_1} = 0, \\ \cdots \\ \dfrac{\partial \ln L(\theta)}{\partial \theta_m} = 0. \end{cases}$$

Ví dụ 2

Tìm ước lượng tham số của phân phối Poisson bằng phương pháp hợp lí cực đại.

Giải. Phân phối Poisson với tham số λ có hàm xác suất:

$$f(x) = \frac{\lambda^x}{x!} e^{-\lambda}, x = 0, 1, ...$$

Giả sử $\{x_1, ..., x_n\}$ là một mẫu ngẫu nhiên từ phân phối trên. Khi đó, hàm hợp lý $L(\lambda)$ được xác định:

$$L(\lambda) = \prod_{i=1}^{n} f(x_i) = \frac{\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} e^{-n\lambda}.$$

Suy ra:

$$\ln L(\lambda) = \sum_{i=1}^{n} x_i \ln(\lambda) - \ln(\prod_{i=1}^{n} x_i!) - n\lambda.$$

Do đó:

$$\frac{d \ln L(\lambda)}{d\lambda} = \frac{1}{\lambda} \sum_{i=1}^{n} x_i - n = 0 \Leftrightarrow \lambda = \frac{1}{n} \sum_{i=1}^{n} x_i.$$

Vậy, ước lượng hợp lý cực đại λ_{ML} của λ là: $\lambda_{ML} = \frac{1}{n} \sum_{i=1}^{n} x_i$.

Ví dụ 3

Tìm ước lượng tham số của phân phối chuẩn $N(\mu;\sigma^2)$ bằng phương pháp ước lượng hợp lí cực đại.

Giải. Phân phối chuẩn $N(\mu,\sigma^2)$ có hàm mật độ xác suất:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

Giả sử $\{x_1,...,x_n\}$ là một mẫu ngẫu nhiên từ phân phối trên. Khi đó, hàm hợp lý $L(\mu,\sigma^2)$ được xác định:

$$L(\mu,\sigma^2)=\prod_{i=1}^{n}f(x_i)=\frac{1}{\sigma^n(\sqrt{2\pi})^n}\exp\left(-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}\right).$$

Suy ra:

$$\ln L(\mu,\sigma^2)=-\frac{n}{2}\ln(\sigma^2)-\frac{n}{2}\ln(2\pi)-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}.$$

Do đó:

$$\frac{\partial\ln L(\mu,\sigma^2)}{\partial\mu}=\frac{\sum_{i=1}^{n}(x_i-\mu)}{\sigma^2},$$
$$\frac{\partial\ln L(\mu,\sigma^2)}{\partial\sigma^2}=-\frac{n}{2\sigma^2}+\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^4}.$$

Giải hệ:

$$\begin{cases}\dfrac{\partial\ln L(\mu,\sigma^2)}{\partial\mu}=0\\ \dfrac{\partial\ln L(\mu,\sigma^2)}{\partial\sigma^2}=0,\end{cases}$$

theo μ,σ^2 ta được: $\mu=\frac{1}{n}\sum_{i=1}^{n}x_i=\bar{x}$ và $\sigma^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$.

Vậy ước lượng hợp lý cực đại (μ_{ML},σ^2_{ML}) của (μ,σ^2) là:

$$\mu_{ML}=\bar{x} \text{ và } \sigma^2_{ML}=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2.$$

2. Khoảng tin cậy

2.1 Định nghĩa

Cho biến ngẫu nhiên X có phân phối phụ thuộc vào tham số θ chưa biết và $\{X_1,X_2,...,X_n\}$ là một mẫu ngẫu nhiên từ phân phối này. Trên cơ sở mẫu ngẫu nhiên này ta xây dựng hai hàm $L=L(X_1,...,X_n)$ và $U=U(X_1,...,X_n)$ sao cho với một số $\alpha\in(0,1)$ cho trước ta có:

$$P(L<\theta<U)=1-\alpha. \qquad (*)$$

Khi đó, khoảng (L,U) được gọi là *khoảng tin cậy* của tham số θ với độ tin cậy $1-\alpha$. Giá trị α được gọi là *mức ý nghĩa*. Hiệu số $U-L$ được gọi là *độ dài khoảng tin cậy*.

Ý nghĩa: Nếu ta thực hiện việc lấy mẫu nhiều lần và mỗi lần ta đều tính khoảng tin cậy (L,U) dựa trên mẫu thu được,thì tỉ lệ khoảng tin cậy chứa giá trị chính xác θ xấp xỉ $1-\alpha$.

Phương pháp chung xây dựng khoảng tin cậy với độ tin cây cậy $1-\alpha$:
i) chọn hai số $\alpha_1, \alpha_2 \in (0,1)$ sao cho $\alpha_1 + \alpha_2 = \alpha$.
ii) xây dựng thống kê:

$$G_n = G_n(X_1, ..., X_n, \theta)$$

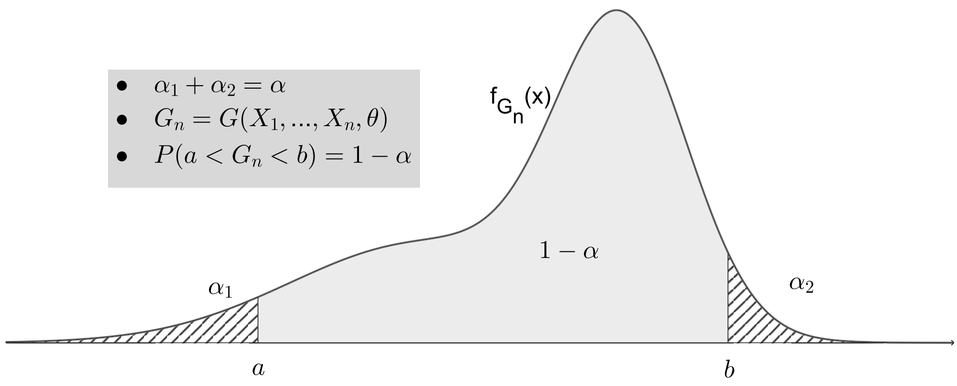
sao cho phân phối của G_n hoàn toàn xác định (hoặc phân phối giới hạn của G_n khi $n \to \infty$ xác định) và không phụ thuộc vào tham số θ. Ngoài ra, G_n là một hàm đơn điệu theo θ.
iii) Chọn hai số a và b sao cho $P(G_n \le a) = \alpha_1$ và $P(G_n < b) = 1 - \alpha_2$.
iv) Giải bất phương trình $a < G_n(X_1, ..., X_n, \theta) < b$ theo θ ta được khoảng tin cậy cần tìm.

Nhận xét:
i) Trong một số trường hợp, các giá trị L, U trong định nghĩa có thể được chọn sao cho:

$$P(L < \theta < U) \ge 1 - \alpha.$$

ii) Nếu $\alpha_1 = 0$ hoặc $\alpha_2 = 0$ ta được các *khoảng tin cậy 1 phía*.

- $\alpha_1 + \alpha_2 = \alpha$
- $G_n = G(X_1, ..., X_n, \theta)$
- $P(a < G_n < b) = 1 - \alpha$



2.2 Khoảng tin cậy cho kì vọng khi đã biết phương sai

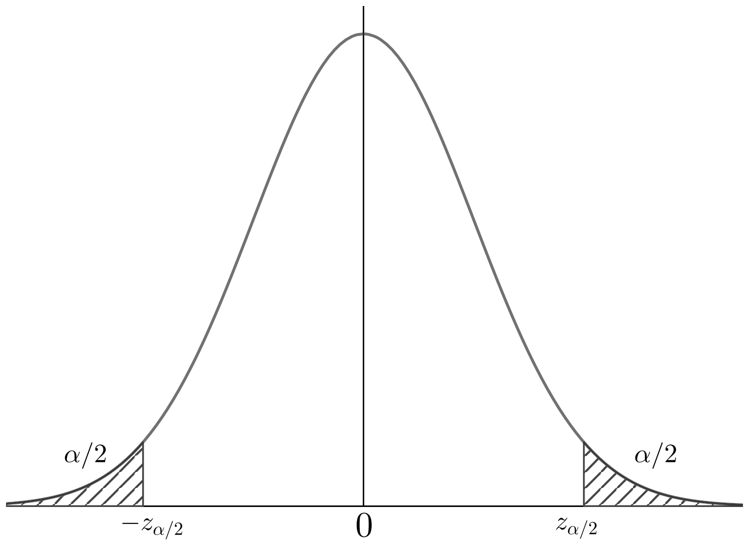
Bài toán. Cho biến ngẫu nhiên X của một tổng thể có phân phối chuẩn $N(\mu; \sigma^2)$ với μ chưa biết và σ^2 đã biết. Tìm khoảng tin cậy cho μ với mức ý nghĩa α.
Gợi ý. Nếu $\{X_1, X_2, ..., X_n\}$ là mẫu ngẫu nhiên của X thì

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0; 1).$$

- Với $\alpha \in (0; 1)$, chọn $\alpha_1 = \alpha_2 = \alpha/2$ và $z_{\alpha/2} = \Phi^{-1}(1 - \frac{\alpha}{2})$ là **giá trị tới hạn** mức $\alpha/2$ của phân phối chuẩn tắc, giải bất phương trình:

$$-z_{\alpha/2} < Z < z_{\alpha/2} \Leftrightarrow \overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

ta được khoảng tin cậy đối xứng.
- Nếu chọn $\alpha_1 = 0$ hoặc $\alpha_2 = 0$. Lúc này ta thay $z_{\alpha/2}$ bởi z_α và thu được khoảng tin cậy một phía.



a. Kết quả

Với độ tin cậy $1-\alpha$:
- Khoảng tin cậy đối xứng của μ:

$$\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}},$$

- Khoảng tin cậy tối đa của μ:

$$\mu < \overline{x} + z_{\alpha}\frac{\sigma}{\sqrt{n}}.$$

- Khoảng tin cậy tối thiểu của μ:

$$\mu > \overline{x} - z_{\alpha}\frac{\sigma}{\sqrt{n}}.$$

Ví dụ 4

Khối lượng (kg) của một thiết bị có phân phối chuẩn $N(\mu;\sigma^2)$ với $\sigma = 0.2$ (kg). Chọn ngẫu nhiên 25 thiết bị người ta tính được trung bình mẫu $\overline{x} = 65,1$ (kg). Với độ tin cậy 95% hãy tìm khoảng tin cậy (đối xứng) cho khối lượng trung bình của thiết bị này. Cho biết $z_{0,025} = 1,96$.

Giải.
- Các số đặc trưng mẫu: $n = 25; \bar{x} = 65,1$.
- Độ tin cậy: $1-\alpha = 0,95 \Rightarrow \alpha = 0,05; z_{\alpha/2} = z_{0,025} = 1,96$.
- Sai số ước lượng:

$$\varepsilon = z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 1,96 * \frac{0,2}{\sqrt{25}} = 0,0784$$

- Khoảng tin cậy cho khối lượng trung bình μ của thiết bị này:

$$\bar{x} - \varepsilon < \mu < \bar{x} + \varepsilon \Leftrightarrow 65,02 < \mu < 65,18$$

b. Vấn đề cỡ mẫu

Từ công thức khoảng tin cậy cho μ ta thấy rằng sai số của ước lượng $|\overline{x} - \mu|$ bé hơn hoặc bằng $z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$.

- Sai số ước lượng:

$$\varepsilon = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

- Điều kiện hạn chế: $\varepsilon < \Delta$ với $\Delta > 0$ cho trước.

$$\varepsilon = z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \Delta$$

- Giải bất phương trình theo n:

$$n > \left(\frac{z_{\alpha/2}\sigma}{\Delta}\right)^2.$$

Ví dụ 5

Trở lại với Ví dụ 4 nếu yêu cầu sai số ước lượng không vượt quá 0,05 thì với độ tin cậy 98% ta cần chọn tối thiểu bao nhiêu thiết bị để khảo sát? Cho biết $z_{0,01} = 2,326$.

Giải.

- Sai số ước lượng:

$$\varepsilon = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

- Điều kiện hạn chế: $\varepsilon < \Delta = 0,05$.
- Suy ra:

$$n > \left(\frac{z_{\alpha/2}\sigma}{\Delta}\right)^2.$$

- Độ tin cậy: $1 - \alpha = 0,98 \Rightarrow \alpha = 0,02; z_{\alpha/2} = z_{0,01} = 2,326$.

Do đó:

$$n > \left(\frac{2,326 * 0,2}{0,05}\right)^2 = 86,56$$

Vậy, kích thước mẫu tối thiểu cần khảo sát: $n = 87$

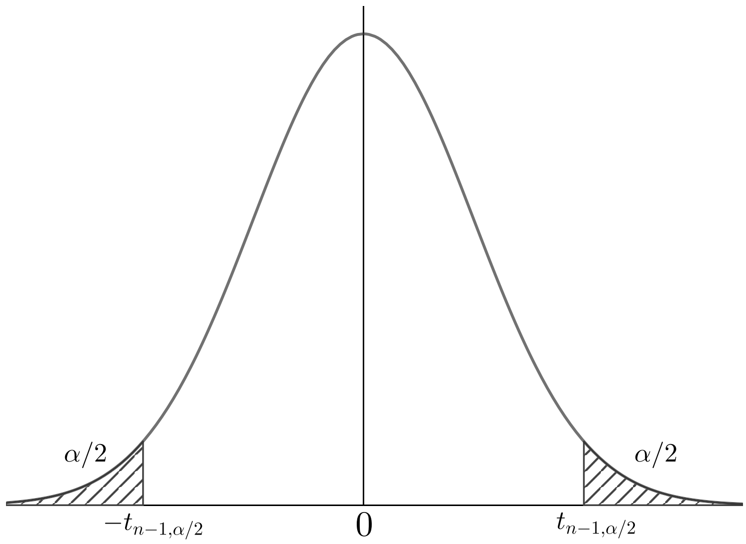
2.3 Khoảng tin cậy cho kì vọng khi chưa biết phương sai

Bài toán: Cho biến ngẫu nhiên X của một tổng thể có phân phối chuẩn $N(\mu; \sigma^2)$ với μ và σ^2 chưa biết. Tìm khoảng tin cậy cho μ với độ tin cậy $1 - \alpha$.

Gợi ý. Nếu $\{X_1, X_2, ..., X_n\}$ là mẫu ngẫu nhiên của X thì

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim T_{n-1}$$

- Với mức ý nghĩa α, chọn $\alpha_1 = \alpha_2 = \alpha/2$ và lấy giá trị $t_{n-1;\alpha/2}$ là **giá trị tới hạn** của phân phối T_{n-1}, ta được khoảng tin cậy đối xứng.
- Chọn $\alpha_1 = 0$ hoặc $\alpha_2 = 0$. Lúc này ta thay $t_{n-1;\alpha/2}$ bởi $t_{n-1;\alpha}$ và thu được khoảng tin cậy một phía.



Kết quả

Với độ tin cậy $1-\alpha$:
- Khoảng tin cậy đối xứng của μ:

$$\overline{x} - t_{n-1;\alpha/2}\frac{s}{\sqrt{n}} < \mu < \overline{x} + t_{n-1;\alpha/2}\frac{s}{\sqrt{n}},$$

- Khoảng tin cậy tối đa của μ:

$$\mu < \overline{x} + t_{n-1;\alpha}\frac{s}{\sqrt{n}}.$$

- Khoảng tin cậy tối thiểu của μ:

$$\overline{x} - t_{n-1;\alpha}\frac{s}{\sqrt{n}} < \mu.$$

Nhận xét:
i) Khi $n > 30$: $t_{n-1;\alpha/2} \approx z_{\alpha/2}$.
ii) Khi kích thước mẫu lớn ($n > 30$) theo định lý giới hạn trung tâm kết quả trên vẫn áp dụng được dù thiếu giả thiết về điều kiện phân phối chuẩn của biến ngẫu nhiên X.

Ví dụ 6

Một mẫu 16 pin dùng cho smartphone được chọn ngẫu nhiên của công ty A có tuổi thọ trung bình tính trên mẫu $\overline{x} = 24308$ (giờ) và độ lệch chuẩn mẫu $s = 727$ (giờ). Giả sử rằng tuổi thọ pin smartphone có phân phối chuẩn. Với độ tin cậy 95%, hãy tìm khoảng tin cậy cho tuổi thọ trung bình của pin smartphone được sản xuất bởi công ty A. Cho biết $t_{15;0,025} = 2,1314$.

Giải. Theo giả thiết:

$$n = 16; \bar{x} = 24308; s = 727$$

Độ tin cậy: $1 - \alpha = 0,95 \Rightarrow \alpha = 0,05; t_{n-1;\alpha/2} = t_{15;0,025} = 2,1314$
Sai số:

$$\varepsilon = t_{n-1;\alpha/2}\frac{s}{\sqrt{n}} = 2,1314 * \frac{727}{\sqrt{16}} = 387,382$$

Khoảng tin cậy cho tuổi thọ trung bình của pin smartphone:

$$\bar{x} - \varepsilon < \mu < \bar{x} + \varepsilon \Leftrightarrow 23920.618 < \mu < 24695.382$$

Ví dụ 7

Năng suất (tạ/ha) của một loại cây trồng tuân theo luật phân phối chuẩn. Hãy tìm ước lượng khoảng đối xứng năng suất trung bình của một loại cây trồng này với độ tin cậy 95% trên cơ sở bảng số liệu điều tra sau đây:

Năng suất(tạ/ha)	*42-47*	*47-52*	*52-57*	*57-62*	*62-67*
Số điểm thu hoạch	*2*	*5*	*14*	*10*	*5*

Cho biết $z_{0,05} = 1,645; z_{0,025} = 1,96; z_{0,02} = 2,054; z_{0,01} = 2,326.$

Giải. Dạng thu gọn:

Năng suất(tạ/ha)	44,5	49,5	54,5	59,5	64,5
Số điểm thu hoạch	2	5	14	10	5

Các số đặc trưng mẫu:

$$n = 36; \bar{x} = 56,028; s = 5,321$$

Độ tin cậy: $1 - \alpha = 0,95 \Rightarrow \alpha = 0,05; t_{n-1;\alpha/2} \approx z_{\alpha/2} = z_{0,025} = 1,96$
Sai số:

$$\varepsilon = t_{n-1;\alpha/2}\frac{s}{\sqrt{n}} = 1,96 * \frac{5,321}{\sqrt{36}} = 1,738$$

Khoảng tin cậy cho năng suất trung bình:

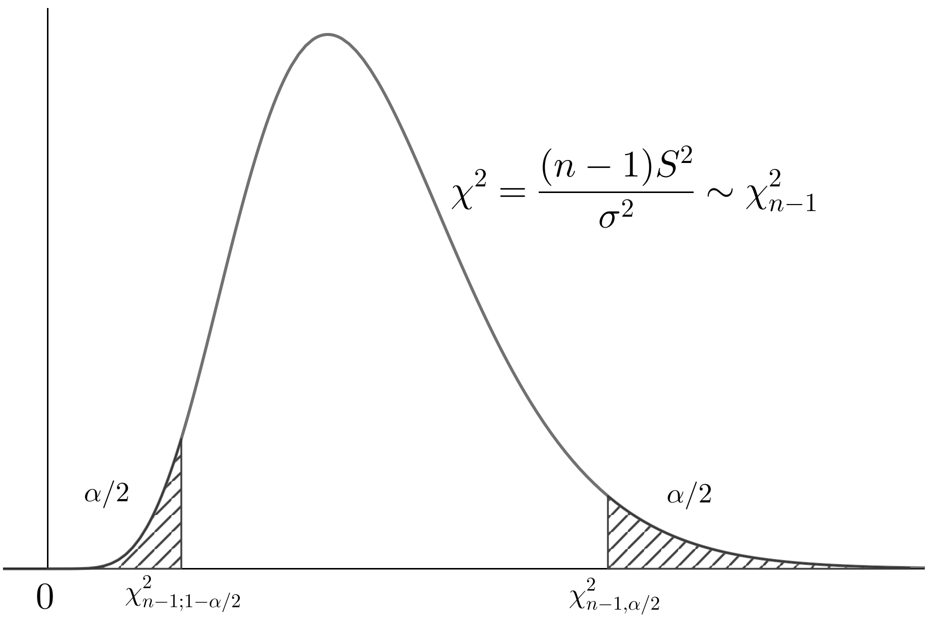
$$\bar{x} - \varepsilon < \mu < \bar{x} + \varepsilon \Leftrightarrow 54,29 < \mu < 57,766$$

2.4 Khoảng tin cậy cho phương sai (đọc thêm)

Bài toán: Cho biến ngẫu nhiên X của một tổng thể có phân phối chuẩn $N(\mu;\sigma^2)$ với μ và σ^2 chưa biết. Tìm khoảng tin cậy cho σ^2 với độ tin cậy $1-\alpha$.
Kết quả: Nếu $\{X_1, X_2, ..., X_n\}$ là mẫu ngẫu nhiên của X thì biến ngẫu nhiên:

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}.$$

- Với mức ý nghĩa α, chọn $\alpha_1 = \alpha_2 = \alpha/2$ và lấy giá trị $\chi^2_{n-1;\alpha/2}, \chi^2_{n-1;1-\alpha/2}$ là **giá trị tới hạn** mức $\alpha/2$ và $1-\alpha/2$ của phân phối χ^2_{n-1}, ta được khoảng tin cậy.
- Chọn $\alpha_1 = 0$ hoặc $\alpha_2 = 0$. Lúc này ta thay $\chi^2_{n-1;\alpha/2}$ bởi $\chi^2_{n-1;\alpha}$, $\chi^2_{n-1;1-\alpha/2}$ bởi $\chi^2_{n-1;1-\alpha}$ và thu được khoảng tin cậy một phía.



Kết quả

Với độ tin cậy $1-\alpha$:
- Khoảng tin cậy hai phía cho σ^2:

$$\frac{(n-1)s^2}{\chi^2_{n-1;\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{n-1;1-\alpha/2}},$$

- Khoảng tin cậy tối đa cho phương sai σ^2:

$$0 < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{n-1;1-\alpha}}.$$

- Khoảng tin cậy tối thiểu cho phương sai σ^2:

$$\frac{(n-1)s^2}{\chi^2_{n-1;\alpha}} < \sigma^2 < +\infty.$$

Ví dụ 8

Một dây chuyền tự động đóng gạo vào bao được kiểm tra. Cân thử 25 bao gạo do dây chuyền này thực hiện thì tính được độ lệch chuẩn khối lượng gạo mỗi bao $s = 0,15$ *(kg). Với độ tin cậy 95% tìm khoảng tin cậy cho phương sai của khối lượng gạo được đóng cho mỗi bao bởi dây chuyền này. Giả thiết khối lượng gạo được đóng cho mỗi bao tuân theo luật phân phối chuẩn. Cho biết:* $\chi^2_{24;0.025} = 39,364$, $\chi^2_{24;0.975} = 12,401$.

Đáp số: $0,0137 < \sigma^2 < 0,0435$.

2.5 Khoảng tin cậy cho tỷ lệ

Bài toán: Xét dấu hiệu A trong tổng thể có tỉ lệ là p (chưa biết). Để khảo sát giá trị p ta tiến hành điều tra một mẫu kích thước n. Giả sử trong mẫu này có k phần tử mang dấu hiệu A. Hãy tìm khoảng tin cậy cho p.

Kết quả:

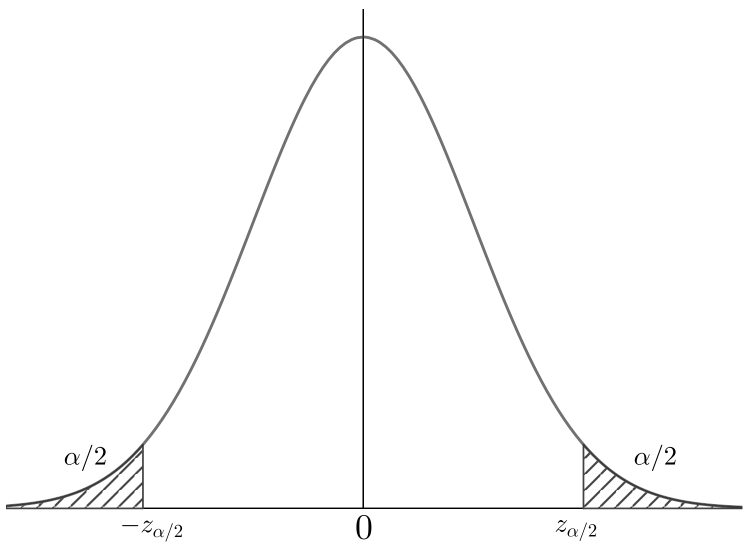
- Gọi X là số phần tử mang dấu hiệu A trong mẫu kích thước n, ta có $X \sim B(n,p)$. Theo định lý giới hạn tích phân Moivre-Laplace, biến ngẫu nhiên:

$$Z = \frac{X - np}{\sqrt{np(1-p)}} \text{ có phân phối xấp xỉ chuẩn tắc } N(0,1).$$

- Với $\alpha \in (0;1)$ cho trước, chọn $\alpha_1 = \alpha_2 = \alpha/2$ ta được khoảng:

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

với $\hat{p} = X/n = k/n$.



Mặt khác, do $\hat{p}$ là một ước lượng của p nên ta có thể xem $p \approx \hat{p}$ khi n lớn.
Điều kiện: $n\hat{p} > 5$ và $n(1-\hat{p}) > 5$.

a. Kết quả

Với độ tin cậy $1-\alpha$:
- Khoảng tin cậy đối xứng cho p:

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

- Khoảng tin cậy tối đa cho p:

$$p < \hat{p} + z_{\alpha}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

- Khoảng tin cậy tối thiểu cho p:

$$p > \hat{p} - z_{\alpha}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

Ví dụ 9

Gieo 400 hạt đậu xanh thì có 50 hạt không nảy mầm. Với độ tin cậy 98%, hãy tìm:
a) Khoảng tin cậy đối xứng cho tỉ lệ hạt nảy mầm.
b) Khoảng tin cậy tối thiểu cho tỉ lệ hạt không nảy mầm.
Cho biết $z_{0,05} = 1,645; z_{0,025} = 1,96; z_{0,02} = 2,054; z_{0,01} = 2,326.$

Giải.
a. - Tỉ lệ mẫu: $\hat{p} = \frac{k}{n} = \frac{400-50}{400} = 0,875.$
- Độ tin cậy: $1-\alpha = 0,98 \Rightarrow \alpha = 0,02; z_{\alpha/2} = z_{0,01} = 2,326$
- Sai số: $\varepsilon = z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 2,326 * \sqrt{\frac{0,875*(1-0,875)}{400}} = 0,0385$
- Khoảng tin cậy cho tỉ lệ hạt nảy mầm p:

$$\hat{p} - \varepsilon < p < \hat{p} + \varepsilon \Leftrightarrow 0,8365 < p < 0,9135$$

b. - Tỉ lệ mẫu:

$$\hat{p} = \frac{k}{n} = \frac{50}{400} = 0,125$$

- Độ tin cậy: $1-\alpha = 0,98 \Rightarrow \alpha = 0,02; z_{\alpha} = z_{0,02} = 2,054$
- Sai số:

$$\varepsilon = z_{\alpha}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 2,054 * \sqrt{\frac{0,125*(1-0,125)}{400}} = 0,034$$

- Khoảng tin cậy tối thiểu cho tỉ lệ hạt không nảy mầm p:

$$p > \hat{p} - \varepsilon \Leftrightarrow p > 0,091$$

b. Nhận xét

i) Khoảng tin cậy cho số lượng cá thể

Từ khoảng tin cậy cho tỉ lệ ta có thể suy ra khoảng tin cậy cho số lượng cá thể mang dấu hiệu nghiên cứu hoặc số lượng cá thể của tổng thể. Cụ thể:

- Gọi K và N lần lượt là số lượng cá thể mang dấu hiệu nghiên cứu và số lượng cá thể của tổng thể. Khi đó, tỉ lệ dấu hiệu nghiên cứu:

$$p = \frac{K}{N} \quad (1)$$

- Dựa trên mẫu điều tra, ta ước lượng được khoảng tin cậy cho tỉ lệ p:

$$p_1 < p < p_2 \quad (2)$$

Từ (1) và (2) suy ra khoảng tin cậy cho giá trị K hoặc N.

Ví dụ 10

Trong một cuộc bầu cử ở một địa phương có 10000 cử tri. Người ta phỏng vấn ngẫu nhiên 300 cử tri thì thấy có 156 người ủng hộ cho ứng cử viên A. Với độ tin cậy 95% tìm khoảng tin cậy (đối xứng) cho số lượng cử tri ở địa phương này ủng hộ cho ứng cử viên A.

Giải. Gọi K là số lượng cử tri ở địa phương này ủng hộ ứng cử viên A. Ta có tỉ lệ ủng hộ ứng cử viên A: $p = K/10000$.
Tỉ lệ mẫu: $\hat{p} = k/n = 156/300 = 0,52$
Độ tin cậy: $1 - \alpha = 0,95 \Rightarrow \alpha = 0,05; z_{\alpha/2} = z_{0,025} = 1,96$
Sai số:

$$\varepsilon = z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1,96 * \sqrt{\frac{0,52 * 0,48}{300}} = 0,0565$$

Khoảng tin cậy cho tỉ lệ ủng hộ:

$$\hat{p} - \varepsilon < p < \hat{p} + \varepsilon \Leftrightarrow 0,4635 < p < 0,5765$$

Từ đó, suy ra:

$$0,4635 < K/10000 < 0,5765 \Leftrightarrow 4635 < K < 5765$$

Ví dụ 11

Để ước lượng số lượng cá có trong hồ, người ta làm như sau. Bắt ngẫu nhiên 500 con cá, sau đó đánh dấu vào các con cá đã được bắt và thả chúng xuống hồ. Sau đó bắt ngẫu nhiên 200 con cá để kiểm tra thì thấy có 30 con có đánh dấu. Với độ tin cậy 90% tìm khoảng tin cậy đối xứng cho số lượng cá có trong hồ.

Giải. Gọi N là số lượng cá trong hồ. Ta có tỉ lệ cá được đánh dấu: $p = 500/N$.
Tỉ lệ mẫu: $\hat{p} = k/n = 30/200 = 0,15$
Độ tin cậy: $1 - \alpha = 0,9 \Rightarrow \alpha = 0,1; z_{\alpha/2} = z_{0,05} = 1,645$
Sai số:

$$\varepsilon = z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1,645 * \sqrt{\frac{0,15 * 0,85}{200}} = 0,0415$$

Khoảng tin cậy cho tỉ lệ cá được đánh dấu:

$$\hat{p} - \varepsilon < p < \hat{p} + \varepsilon \Leftrightarrow 0,1085 < p < 0,1915$$

Từ đó, suy ra:

$$0,1085 < 500/N < 0,1915 \Leftrightarrow 2610,9 < N < 4608,3$$

ii) Vấn đề về cỡ mẫu

- Sai số ước lượng:

$$\varepsilon = z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

- Điều kiện hạn chế: $\varepsilon \leq \Delta$, $\Delta > 0$ cho trước.
- Giả sử giá trị $\hat{p}$ không biến động lớn khi tính tính trên các mẫu khác nhau, khi đó:

$$n \geq \left(\frac{z_{\alpha/2}}{\Delta}\right)^2 \hat{p}(1-\hat{p}).$$

- Trong trường hợp không có dữ liệu để tính giá trị $\hat{p}$, sử dụng bất đẳng thức $\hat{p}(1-\hat{p}) \leq 1/4$ ta thu được bất đẳng thức:

$$n \geq \frac{1}{4}\left(\frac{z_{\alpha/2}}{\Delta}\right)^2.$$

Ví dụ 12

Một cuộc khảo sát sức khỏe cộng đồng đang được lên kế hoạch trong một khu vực đô thị lớn với mục đích ước tính tỷ lệ trẻ em từ 0 đến 14 tuổi chưa được tiêm chủng bại liệt đầy đủ. Các nhà tổ chức của dự án muốn tỷ lệ mẫu của trẻ không được tiêm chủng đầy đủ k/n phải nằm trong khoảng của tỷ lệ thực p với sai số $\pm 0,05$ với xác suất tối thiểu 98%. Hỏi kích thước mẫu điều tra tối thiểu là bao nhiêu?
Cho biết $z_{0,05} = 1,645; z_{0,025} = 1,96; z_{0,02} = 2,054; z_{0,01} = 2,326$.

Giải. Sai số ước lượng:

$$\varepsilon = z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Điều kiện: $\varepsilon \leq \Delta = 0,05$. Suy ra:

$$n \geq \left(\frac{z_{\alpha/2}}{\Delta}\right)^2 \hat{p}(1-\hat{p}).$$

Vì $\hat{p}$ chưa biết nên áp dụng bất đẳng thức $\hat{p}(1-\hat{p}) \leq 1/4$ ta thu được:

$$n \geq \frac{1}{4}\left(\frac{z_{\alpha/2}}{\Delta}\right)^2.$$

Độ tin cậy: $1-\alpha = 0,98 \Rightarrow \alpha = 0,02; z_{\alpha/2} = z_{0,01} = 2,326$. Do đó:

$$n \geq \frac{1}{4}\left(\frac{2,326}{0,05}\right)^2 = 541,0276$$

Vậy, $n = 542$.

Ví dụ 13

Sản lượng gạo bán ra trong 1 ngày có phân phối chuẩn. Điều tra sản lượng bán ra trong 120 ngày, ta được bảng số liệu sau:

Sản lượng (kg)	*400-440*	*440-480*	*480-520*	*520-560*	*560-600*	*600-640*	*640-680*
Số ngày	*4*	*9*	*15*	*23*	*30*	*22*	*17*

a) Với độ tin cậy 95% tìm khoảng tin cậy đối xứng cho số tiền bán ra trung bình trong một ngày. Biết giá mỗi kg gạo là 15.000 đồng.
b) Ngày có sản lượng gạo bán ra không nhỏ hơn 560 kg được gọi là ngày cao điểm. Với độ tin cậy 98%, tìm khoảng tin cậy đối xứng cho số ngày cao điểm trong 1000 ngày bán.
Cho biết $z_{0,05} = 1,645; z_{0,025} = 1,96; z_{0,02} = 2,054; z_{0,01} = 2,326.$

Giải.
a. Dạng thu gọn:

Sản lượng (kg)	420	460	500	540	580	620	660
Số ngày	4	9	15	23	30	22	17

Các số đặc trưng mẫu: $n = 120; \bar{x} = 566,67; s = 64$
Độ tin cậy: $1 - \alpha = 0,95 \Rightarrow \alpha = 0,05; t_{n-1;\alpha/2} = t_{119;0,025} \approx z_{0,025} = 1,96.$
Sai số:

$$\varepsilon = t_{n-1;\alpha/2}\frac{s}{\sqrt{n}} = 1,96 * \frac{64}{\sqrt{120}} = 11,45$$

Khoảng tin cậy cho khối lượng gạo trung bình bán được trong 1 ngày:

$$\bar{x} - \varepsilon < \mu < \bar{x} + \varepsilon \Leftrightarrow 555,22 < \mu < 578,12$$

Khoảng tin cậy cho số tiền T (ngàn đồng) trung bình bán được trong 1 ngày:

$$555,22 * 15 < T < 578,12 * 15 \Leftrightarrow 8328,3 < T < 8671,8$$

b. Gọi K là số ngày cao điểm trong 1000 ngày. Ta có tỉ lệ ngày cao điểm: $p = K/1000.$
Tỉ lệ mẫu ngày cao điểm:

$$\hat{p} = \frac{k}{n} = \frac{30 + 22 + 17}{120} = 0,575$$

Độ tin cậy: $1 - \alpha = 0,98 \Rightarrow \alpha = 0,02; z_{\alpha/2} = z_{0,01} = 2,326.$
Sai số:

$$\varepsilon = z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 2,326 * \sqrt{\frac{0,575 * (1 - 0,575)}{120}} = 0,105$$

Khoảng tin cậy cho tỉ lệ ngày cao điểm p:

$$\hat{p} - \varepsilon < p < \hat{p} + \varepsilon \Leftrightarrow 0,47 < p < 0,68$$

Suy ra:

$$0,47 < K/1000 < 0,68 \Leftrightarrow 470 < K < 680$$