Chương 4
THỐNG KÊ VÀ ƯỚC LƯỢNG THAM SỐ

BỘ MÔN TOÁN ỨNG DỤNG[1]

VIỆN TOÁN ỨNG DỤNG VÀ TIN HỌC
ĐẠI HỌC BÁCH KHOA HÀ NỘI

SAMI.HUST – 2023

[1] Phòng BIS.201-D3.5

4.2. Ước lượng điểm

Khái niệm ước lượng điểm

Giả sử cần ước lượng tham số θ của biến ngẫu nhiên X. Từ X ta lập một mẫu ngẫu nhiên kích thước n, $W_X = (X_1, X_2, \ldots, X_n)$.

- Ước lượng điểm của θ là một thống kê $\widehat{\Theta} = g(X_1, X_2, \ldots, X_n)$, viết gọn là $\widehat{\Theta}$ và gọi là hàm ước lượng điểm cho θ.
- Tiến hành lập mẫu cụ thể $W_x = (x_1, x_2, \ldots, x_n)$. Giá trị cụ thể $\widehat{\theta} = g(x_1, x_2, \ldots, x_n)$ được gọi là một giá trị ước lượng điểm của θ.

Khái niệm ước lượng điểm

Ví dụ 16

(a) Giả sử X là biến ngẫu nhiên có phân phối chuẩn với kỳ vọng $E(X) = \mu$ chưa biết.

- Kỳ vọng mẫu ngẫu nhiên $\overline{X}$ là một hàm ước lượng điểm cho μ.
- Sau khi lựa chọn mẫu cụ thể, thì giá trị trung bình mẫu $\overline{x}$ sẽ là một ước lượng điểm của μ. Chẳng hạn, lấy ngẫu nhiên một mẫu kích thước $n = 10$ và nhận được bộ dữ liệu $x_1 = 13,8$, $x_2 = 10,4$, $x_3 = 9,7$, $x_4 = 12,6$, $x_5 = 14,1$, $x_6 = 10,8$, $x_7 = 15,1$, $x_8 = 9,5$, $x_9 = 13,1$, $x_{10} = 11,3$, thì một ước lượng điểm của μ là

$$\overline{x} = \frac{13,8 + 10,4 + 9,7 + 12,6 + 14,1 + 10,8 + 15,1 + 9,5 + 13,1 + 11,3}{10} = 12,04.$$

Khái niệm ước lượng điểm

Ví dụ 16 (tiếp theo)

(b) Nếu phương sai $V(X) = \sigma^2$ chưa biết thì phương sai mẫu ngẫu nhiên hiệu chỉnh S^2 là một hàm ước lượng điểm cho phương sai σ^2 và giá trị s^2 tính được từ dữ liệu mẫu là một ước lượng điểm của σ^2. Trong ví dụ trên,

$$\begin{aligned} s^2 &= \frac{1}{9}\big[(13,8-12,04)^2+(10,4-12,04)^2+(9,7-12,04)^2+(12,6-12,04)^2+(14,1-12,04)^2 \\ &\quad +(10,8-12,04)^2+(15,1-12,04)^2+(9,5-12,04)^2+(13,1-12,04)^2+(11,3-12,04)^2\big] \\ &\approx 4,8783 \end{aligned}$$

là một ước lượng điểm của σ^2.

Một số ước lượng điểm thông dụng

✍ Ta thường cần các ước lượng điểm cho các tham số sau đây:

- Kỳ vọng μ của tổng thể.
- Phương sai σ^2 của tổng thể; độ lệch chuẩn σ của tổng thể.
- Tỷ lệ p của các phần tử có tính chất A trong tổng thể.
- Sự khác nhau giữa hai kỳ vọng của hai tổng thể $\mu_1 - \mu_2$.
- Sự khác nhau về tỷ lệ của các phần tử có dấu hiệu A của hai tổng thể $p_1 - p_2$.

Một số ước lượng điểm thông dụng

✍ Các ước lượng điểm hợp lý của các tham số:

- Đối với μ, một ước lượng điểm là $\widehat{\mu} = \overline{x}$.
- Đối với σ^2, một ước lượng điểm là $\widehat{\sigma}^2 = s^2$.
- Đối với tỷ lệ p, một ước lượng điểm là tần suất mẫu $\widehat{p} = x/n$, trong đó x là số phần tử có dấu hiệu A trong mẫu kích thước n.
- Đối với $\mu_1 - \mu_2$, một ước lượng điểm là $\widehat{\mu}_1 - \widehat{\mu}_2 = \overline{x} - \overline{y}$, sự khác nhau giữa hai trung bình mẫu của hai mẫu ngẫu nhiên độc lập.
- Đối với $p_1 - p_2$, một ước lượng điểm là $\widehat{p}_1 - \widehat{p}_2$, sự khác nhau giữa hai tần suất mẫu được tính toán từ hai mẫu ngẫu nhiên độc lập.

4.2. Ước lượng điểm

Một số tiêu chuẩn lựa chọn hàm ước lượng

✍ Có thể có một số lựa chọn hàm ước lượng điểm khác nhau cho một tham số.

Ví dụ 17

Với dữ liệu trong Ví dụ 16, ta có thể chọn

- giá trị trung bình mẫu
$$\overline{x} = 12,04,$$
- giá trị trung vị mẫu
$$\widetilde{x} = \frac{x_5 + x_6}{2} = \frac{11,3 + 12,6}{2} = 11,95,$$
- giá trị trung bình mẫu sau khi đã bỏ đi giá trị nhỏ nhất và lớn nhất
$$\overline{x}_{-(\max + \min)} = \frac{9,7 + 10,4 + 10,8 + 11,3 + 12,6 + 13,1 + 13,8 + 14,1}{8} = 11,975$$

làm ước lượng điểm cho μ.

Một số tiêu chuẩn lựa chọn hàm ước lượng

✍ Tại sao phải lựa chọn hàm ước lượng?

- Chú ý rằng θ là một số chưa biết, còn $\widehat{\Theta} = g(X_1, X_2, \ldots, X_n)$ là một biến ngẫu nhiên. Như vậy, ta đã lấy một biến ngẫu nhiên để xấp xỉ cho một số.
- Câu hỏi đặt ra là:

1. Ước lượng đưa ra có "tốt" không?
2. "Ước lượng tốt" được hiểu theo nghĩa nào?

Ước lượng không chệch

- Ước lượng $\widehat{\Theta}$ cho θ được gọi là ước lượng không chệch của θ nếu

$$E(\widehat{\Theta}) = \theta. \tag{32}$$

- Nếu ước lượng là có chệch, nghĩa là $E(\widehat{\Theta}) \neq \theta$, thì

$$E(\widehat{\Theta}) - \theta \tag{33}$$

  được gọi là độ chệch của ước lượng $\widehat{\Theta}$.

✍ Có thể có nhiều ước lượng không chệch cho một tham số của tổng thể.

Ước lượng không chệch cho μ

- Giả sử biến ngẫu nhiên gốc X có $E(X) = \mu$ chưa biết, μ được xem là kỳ vọng của tổng thể.
- Từ X ta lập mẫu ngẫu nhiên kích thước n, $W_X = (X_1, X_2, \ldots, X_n)$.
- Chọn $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ là hàm ước lượng điểm cho kỳ vọng μ.
- Vì $E(\overline{X}) = \mu$, nên $\overline{X}$ là một ước lượng không chệch cho μ.
- Khi có một mẫu cụ thể $W_x = (x_1, x_2, \ldots, x_n)$ thì $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ là một ước lượng điểm không chệch của μ.

Ước lượng không chệch cho μ

Ví dụ 18

Với dữ liệu trong Ví dụ 16, thì

- trung bình mẫu
$$\overline{x} = 12,04,$$
- trung vị mẫu
$$\widetilde{x} = 11,95,$$
- trung bình mẫu sau khi đã bỏ đi giá trị nhỏ nhất và lớn nhất
$$\overline{x}_{-(\max + \min)} = 11,975$$

đều là các ước lượng không chệch cho μ.

Ước lượng không chệch cho σ^2

- Giả sử biến ngẫu nhiên X có $V(X) = \sigma^2$ chưa biết, σ^2 được xem là phương sai của tổng thể.
- Lập mẫu ngẫu nhiên kích thước n, $W_X = (X_1, X_2, \ldots, X_n)$. Xuất phát từ công thức tính phương sai, thống kê

$$\widetilde{S}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

  được xem xét để dùng làm hàm ước lượng điểm cho σ^2.
- Tuy nhiên

$$E(\widetilde{S}^2) = \frac{n-1}{n}\sigma^2 \neq \sigma^2,$$

  nên $\widetilde{S}^2$ là một hàm ước lượng điểm có chệch cho σ^2.

Ước lượng không chệch cho σ^2

- Để thu được ước lượng không chệch cho σ^2 ta sử dụng phương sai mẫu ngẫu nhiên hiệu chỉnh

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \frac{n}{n-1}\widetilde{S}^2,$$

vì

$$E(S^2) = E\left(\frac{n}{n-1}\widetilde{S}^2\right) = \sigma^2.$$

- Khi có mẫu cụ thể $W_x = (x_1, x_2, \ldots, x_n)$ ta tính được giá trị

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2,$$

đây là một ước lượng điểm không chệch của σ^2.

Ước lượng không chệch cho σ^2

Ví dụ 19

Với dữ liệu trong Ví dụ 16, thì một ước lượng điểm không chệch cho σ^2 là

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 \approx 4,8783.$$

Ước lượng không chệch cho p

- Gọi p là tỷ lệ các phần tử có dấu hiệu A trong tổng thể, p chưa biết. Ta thực hiện n quan sát độc lập và gọi X là số cá thể có dấu hiệu A trong mẫu quan sát. Khi đó, tần suất mẫu ngẫu nhiên

$$\widehat{P} = \frac{X}{n}$$

  là một hàm ước lượng điểm cho p.
- Vì $E(\widehat{P}) = p$, nên $\widehat{P}$ là hàm ước lượng không chệch của p.
- Khi có mẫu cụ thể, ta tính được giá trị cụ thể của $\widehat{P}$, ký hiệu là $\widehat{p}$, đây là một ước lượng điểm không chệch cho p.

Nhận xét về ước lượng không chệch

✍

- Khi một ước lượng là không chệch thì "độ chệch" bằng 0, nghĩa là $E(\widehat{\Theta}) - \theta = 0$. Tính chất này có nghĩa là ước lượng $\widehat{\Theta}$ không có sai số hệ thống mà chỉ có sai số ngẫu nhiên.
- Điều kiện $E(\widehat{\Theta}) = \theta$ của ước lượng không chệch có nghĩa là trung bình các giá trị của $\widehat{\Theta}$ bằng θ. Tuy nhiên, không có nghĩa là mọi giá trị của $\widehat{\Theta}$ đều trùng khít với θ mà từng giá trị của $\widehat{\Theta}$ có thể sai lệch rất lớn so với θ.

✍ Chọn hàm ước lượng nào? **Ta cần tìm ước lượng không chệch sao cho độ sai lệch trung bình bình phương là bé nhất.**

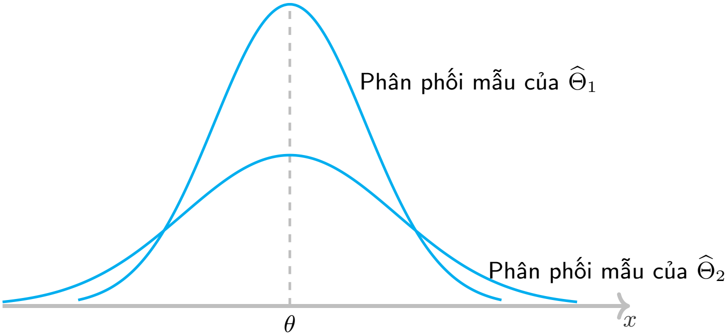
Phương sai của ước lượng

✍

- Ví dụ 18 cho thấy, ước lượng không chệch cho θ là không duy nhất. Vấn đề đặt ra là, trong số các ước lượng không chệch này ta nên chọn ước lượng nào?
- Giả sử $\widehat{\Theta}_1$ và $\widehat{\Theta}_2$ là các hàm ước lượng không chệch cho θ, tức là $E(\widehat{\Theta}_1) = \theta$ và $E(\widehat{\Theta}_2) = \theta$. Tuy nhiên phương sai của chúng lại có thể khác nhau (xem Hình 6).

  Vì phương sai của $\widehat{\Theta}_1$ nhỏ hơn phương sai của $\widehat{\Theta}_2$ nên $\widehat{\Theta}_1$ có khả năng là một ước lượng gần với giá trị θ hơn.

Phương sai của ước lượng



Hình 6: Phân phối mẫu của hai ước lượng không chệch $\widehat{\Theta}_1$ và $\widehat{\Theta}_2$

Phương sai của ước lượng

Định lý 7

Nếu $W_X = (X_1, X_2, \ldots, X_n)$ là một mẫu ngẫu nhiên kích thước n được xây dựng từ biến ngẫu nhiên X có phân phối chuẩn $\mathcal{N}(\mu; \sigma^2)$, thì trung bình mẫu $\overline{X}$ là hàm ước lượng không chệch có phương sai bé nhất cho μ.

✍ Khi không biết có tồn tại phương sai bé nhất của các hàm ước lượng không chệch hay không, để chọn ước lượng tốt ta sẽ so sánh độ lệch tiêu chuẩn hay phương sai của các hàm ước lượng. Ước lượng không chệch có độ lệch tiêu chuẩn hay phương sai nhỏ hơn sẽ "tốt hơn".

Ước lượng hiệu quả

✍ **Một nguyên tắc hợp lý của ước lượng điểm là lựa chọn ước lượng điểm có phương sai nhỏ nhất.**

Định nghĩa 10

Hàm $\widehat{\Theta}$ được gọi là ước lượng hiệu quả cho tham số θ nếu nó là hàm ước lượng không chệch cho θ và phương sai của nó nhỏ hơn bất kỳ phương sai của một hàm ước lượng không chệch nào khác cho θ.

Sai số tiêu chuẩn của ước lượng

- Độ lệch chuẩn của ước lượng điểm $\widehat{\Theta}$, ký hiệu là $\sigma_{\widehat{\Theta}}$, được gọi là sai số tiêu chuẩn.
- Ước lượng điểm của sai số tiêu chuẩn được ký hiệu là $\widehat{\sigma}_{\widehat{\Theta}}$.
- Giả sử $W_X = (X_1, X_2, \ldots, X_n)$ là một mẫu ngẫu nhiên kích thước n được xây dựng từ biến ngẫu nhiên X có phân phối chuẩn $\mathcal{N}(\mu; \sigma^2)$. Khi đó, phân phối mẫu của $\overline{X}$ là phân phối chuẩn $\mathcal{N}\left(\mu; \sigma^2/n\right)$, sai số tiêu chuẩn của $\overline{X}$ là

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}. \tag{34}$$

- Nếu σ chưa biết, ta thay độ lệch chuẩn hiệu chỉnh mẫu S vào phương trình trên, thì sai số tiêu chuẩn ước lượng của $\overline{X}$ là

$$\widehat{\sigma}_{\overline{X}} = \frac{S}{\sqrt{n}}. \tag{35}$$

Sai số tiêu chuẩn của ước lượng

Ví dụ 20

Với dữ liệu của Ví dụ 16,

- Một ước lượng điểm không chệch cho μ là trung bình mẫu $\overline{x} = 12,04$.

  Sai số tiêu chuẩn của trung bình mẫu là $\sigma_{\overline{x}} = \sigma/\sqrt{n}$, trong đó $n = 10$ là kích thước mẫu. Ước lượng điểm cho sai số tiêu chuẩn của trung bình mẫu là

$$\widehat{\sigma}_{\overline{x}} = \frac{s}{\sqrt{n}} = \frac{4,8783}{\sqrt{10}} \approx 1,5427.$$

- Nếu ta dùng $\widehat{\mu} = \dfrac{x_1 + x_2}{2}$ là một ước lượng điểm không chệch khác cho μ thì sai số tiêu chuẩn là $\sigma_{\widehat{\mu}} = \sigma/\sqrt{2}$ và ước lượng điểm cho sai số tiêu chuẩn này là

$$\widehat{\sigma}_{\widehat{\mu}} = \frac{s}{\sqrt{2}} = \frac{4,8783}{\sqrt{2}} \approx 3,4495.$$

- Rõ ràng ước lượng $\widehat{\mu}$ không hiệu quả bằng ước lượng $\overline{x}$.

Sai số bình phương trung bình của ước lượng

- Đôi khi cần sử dụng độ chệch của ước lượng. Trong những trường hợp như vậy, sai số bình phương trung bình của ước lượng là quan trọng. Sai số bình phương trung bình của hàm ước lượng $\widehat{\Theta}$ được định nghĩa là kỳ vọng của bình phương độ lệch giữa $\widehat{\Theta}$ và θ.
- Đại lượng $\widehat{\Theta} - \theta$ được gọi là sai số (nó là một biến ngẫu nhiên) còn giá trị trung bình của sai số $E(\widehat{\Theta} - \theta) = E(\widehat{\Theta}) - \theta$ được gọi là độ chệch. Ta mong muốn tìm được ước lượng sao cho sai số bình phương trung bình

$$MSE(\widehat{\Theta}) = E(\widehat{\Theta} - \theta)^2 \tag{36}$$

nhỏ nhất có thể.
- Tuy nhiên, trong nhiều trường hợp tìm ước lượng thỏa mãn điều kiện này khá khó. Lưu ý rằng

$$E(\widehat{\Theta} - \theta)^2 = [E(\widehat{\Theta}) - \theta]^2 + V(\widehat{\Theta}) = (\text{Độ chệch})^2 + V(\widehat{\Theta}).$$

Vì vậy ta thường chỉ tìm ước lượng tốt nhất trong các ước lượng không chệch tức là ước lượng có phương sai $V(\widehat{\Theta})$ nhỏ nhất trong các ước lượng không chệch.

4.2. Ước lượng điểm

Phương pháp mô men

Giả sử $W_X = (X_1, X_2, \ldots, X_n)$ là mẫu ngẫu nhiên kích thước n được xây dựng từ biến ngẫu nhiên X có phân phối xác suất $F_X(x)$.

- Mô men lý thuyết bậc k (hay mô men phân phối) là $E(X^k)$, $k = 1, 2, \ldots$.
- Tương ứng mô men mẫu bậc k là $\frac{1}{n}\sum_{i=1}^{n} X_i^k$, $k = 1, 2, \ldots$

✍ **Ý tưởng của phương pháp**

- Ý tưởng chung của phương pháp mô men là cân bằng các mô men lý thuyết, được xác định dựa trên giá trị kỳ vọng với mô men mẫu tương ứng.
- Mô men lý thuyết là hàm của các tham số chưa biết. Giải hệ phương trình này tìm các ước lượng của tham số chưa biết.

Phương pháp mô men

Định nghĩa 11

Giả sử $W_X = (X_1, X_2, \ldots, X_n)$ là mẫu ngẫu nhiên kích thước n được xây dựng từ biến ngẫu nhiên X có phân phối $F_X(x; \theta_1, \theta_2, \ldots, \theta_m)$, trong đó các tham số $\theta_1, \theta_2, \ldots, \theta_m$ chưa biết. Khi đó, ước lượng mô men $\widehat{\Theta}_1$, $\widehat{\Theta}_2$, $\ldots$, $\widehat{\Theta}_m$ được xác định bằng cách cân bằng m mô men lý thuyết bậc k với m mô men mẫu bậc k, $k = 1, 2, \ldots, m$.

Ví dụ 21

Mô men lý thuyết bậc nhất là $E(X) = \mu$ và mô men mẫu bậc nhất là $\frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$. Do đó, ta tìm được ước lượng mô men cho μ là

$$\widehat{\mu}_{\text{MoL}} = \overline{X}.$$

Ước lượng mô men phân phối mũ

Giả sử $W_X = (X_1, X_2, \ldots, X_n)$ là mẫu ngẫu nhiên kích thước n được xây dựng từ biến ngẫu nhiên gốc X có phân phối mũ với tham số λ chưa biết, cần ước lượng.

- Để tìm ước lượng mô men cho tham số λ ta cân bằng $E(X)$ với $\overline{X}$, tức là

$$\frac{1}{\lambda} = \overline{X}.$$

- Do đó, ước lượng mô men cho λ là

$$\widehat{\lambda}_{\mathsf{MoL}} = \frac{1}{\overline{X}}.$$

Ước lượng mô men phân phối mũ

Ví dụ 22

Giả sử tuổi thọ (tháng) của một thiết bị điện tử là biến ngẫu nhiên có phân phối mũ. Lấy ngẫu nhiên 10 đơn vị thiết bị điện tử và kiểm tra, thấy tuổi thọ (tính bằng tháng) như sau

$$11,96;\ 5,03;\ 67,40;\ 16,07;\ 31,50;\ 7,73;\ 11,10;\ 22,38;\ 20,10;\ 23,20.$$

Vì

$$\overline{x} = \frac{11,96+5,03+67,40+16,07+31,50+7,73+11,10+22,38+20,10+23,20}{10} = 21,747 \text{ (tháng)}$$

nên ước lượng mô men của λ là

$$\widehat{\lambda}_{\mathrm{MoL}} = \frac{1}{\overline{x}} = \frac{1}{21,747} \approx 0,0460.$$

Ước lượng mô men phân phối chuẩn

Giả sử $W_X = (X_1, X_2, \ldots, X_n)$ là mẫu ngẫu nhiên kích thước n được xây dựng từ biến ngẫu nhiên gốc X có phân phối chuẩn $\mathcal{N}(\mu; \sigma^2)$ với các tham số μ và σ^2 chưa biết, cần ước lượng.

- Để tìm ước lượng mô men cho tham số μ và σ^2 ta cân bằng $E(X) = \mu$ và $E(X^2) = \mu^2 + \sigma^2$ với $\overline{X}$ và $\frac{1}{n}\sum_{i=1}^{n} X_i^2$ tương ứng

$$\mu = \overline{X}; \quad \mu^2 + \sigma^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2.$$

- Giải các phương trình này ta nhận được các ước lượng mô men cho μ và σ^2 là

$$\widehat{\mu}_{\text{MoL}} = \overline{X}; \quad \widehat{\sigma}^2_{\text{MoL}} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \Big(\frac{1}{n}\sum_{i=1}^{n} X_i\Big)^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

✍ Ước lượng mô men cho σ^2 không phải là ước lượng không chệch.

Phương pháp hợp lý cực đại

- Một trong những phương pháp tốt nhất để tìm ước lượng điểm cho một tham số là phương pháp hợp lý cực đại. Kỹ thuật này được phát triển vào những năm 1920 bởi nhà thống kê nổi tiếng người Anh, R.A. Fisher.
- Như tên của phương pháp, ước lượng điểm sẽ là giá trị của tham số làm tối đa hóa hàm hợp lý.

Hàm hợp lý

- Giả sử ta đã biết phân phối xác suất của biến ngẫu nhiên gốc X, chẳng hạn, hàm mật độ xác suất $f_X(x;\theta)$ (có thể hiểu $f_X(x,\theta)$ là công thức xác suất nếu X rời rạc).
- Lập một mẫu cụ thể $W_x = (x_1, x_2, \ldots, x_n)$ kích thước n. Hàm của đối số θ

$$L(x_1, x_2, \ldots, x_n, \theta) = f_X(x_1, \theta) f_X(x_2, \theta) \ldots f_X(x_n, \theta) \tag{37}$$

  gọi là hàm số hợp lý của tham số θ.
- Giá trị của hàm hợp lý chính là xác suất hay mật độ xác suất tại điểm $(x_1, x_2, \ldots, x_n)$. Giá trị $\widehat{\theta} = g(x_1, x_2, \ldots, x_n)$ được gọi là ước lượng hợp lý cực đại của θ nếu ứng với giá trị này của θ hàm hợp lý đạt cực đại.

Phương pháp hợp lý cực đại

Các bước tiến hành

Do hàm L và hàm $\ln L$ đạt cực đại tại cùng một giá trị của θ nên ta có thể tìm giá trị của θ để $\ln L$ đạt cực đại theo các bước sau.

1. Tìm đạo hàm bậc nhất của hàm $\ln L$ theo θ.
2. Lập phương trình $\dfrac{d\ln L}{d\theta} = 0$. Phương trình này gọi là phương trình hợp lý. Giả sử nó có nghiệm $\theta := \widehat{\theta} = g(x_1, x_2, \ldots, x_n)$.
3. Tìm đạo hàm bậc hai $\dfrac{d^2 \ln L}{d\theta^2}$. Nếu tại điểm $\theta = \widehat{\theta}$ đạo hàm bậc hai âm thì tại điểm này hàm $\ln L$ đạt cực đại. Khi đó, $\widehat{\theta} = g(x_1, x_2, \ldots, x_n)$ là ước lượng điểm hợp lý cực đại cần tìm.

Ước lượng hợp lý cực đại phân phối Bernoulli

Ví dụ 23

Giả sử X là biến ngẫu nhiên có phân phối Bernoulli với $P(X=x)=p^x(1-p)^{1-x}$, $x=0,1$ và p là tham số cần ước lượng. Hàm hợp lý của một mẫu $W_x=(x_1,x_2,\ldots,x_n)$ kích thước n là

$$\begin{aligned} L(p) := L(x_1,x_2,\ldots,x_n,p) &= p^{x_1}(1-p)^{1-x_1}p^{x_2}(1-p)^{1-x_2}\ldots p^{x_n}(1-p)^{1-x_n} \\ &= \Pi_{i=1}^n p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}. \end{aligned}$$

Nếu $\widehat{p}$ làm cực hàm $L(p)$ thì nó cũng làm cực đại hàm $\ln L(p)$ với

$$\ln L(p) = \left(\sum_{i=1}^n x_i\right)\ln p + \left(n - \sum_{i=1}^n x_i\right)\ln(1-p).$$

Giải phương trình $\frac{d\ln L(p)}{dp} = \frac{\sum_{i=1}^n x_i}{p} - \frac{n-\sum_{i=1}^n x_i}{1-p} = 0$ ta nhận được nghiệm $p := \widehat{p} = \frac{1}{n}\sum_{i=1}^n x_i$. Vì $\frac{d^2\ln L}{dp^2}\Big|_{p=\frac{1}{n}\sum_{i=1}^n x_i} < 0$, nên hàm ước lượng hợp lý cực đại của p là

$$\widehat{P}_{\mathrm{MLE}} = \frac{1}{n}\sum_{i=1}^n X_i.$$

Ước lượng hợp lý cực đại phân phối chuẩn

Ví dụ 24

Hàm ước lượng hợp lý cực đại của các tham số μ và σ của biến ngẫu nhiên có luật phân phối chuẩn $\mathcal{N}(\mu; \sigma^2)$ là

$$\widehat{\mu}_{\text{MLE}} = \overline{X} \quad \text{và} \quad \widehat{\sigma}^2_{\text{MLE}} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

(Phần chứng minh xem như là bài tập).

4.2. Ước lượng điểm

Bài tập

Bài 13

Cho tổng thể có trung bình là 1065 và độ lệch chuẩn là 500. Nếu rút ra một mẫu ngẫu nhiên có kích thước $n = 100$ từ tổng thể đó thì kỳ vọng của trung bình mẫu ngẫu nhiên bằng bao nhiêu?

A. 100

B. 1065

C. 500

D. 500/100

Bài 14

Cho tổng thể có trung bình là 1065 và độ lệch chuẩn là 500. Nếu rút ra một mẫu ngẫu nhiên có kích thước $n = 100$ từ tổng thể đó thì phương sai của trung bình mẫu ngẫu nhiên bằng bao nhiêu?

A. 1000

B. 1065

C. 500

D. $500^2/100$

Bài tập

Bài 15

Để xác định số người chơi chứng khoán ở thành phố X, người ta điều tra 1250 người, thấy có 20 người chơi chứng khoán. Hãy ước lượng số lượng người chơi chứng khoán ở thành phố X biết thành phố có 2 triệu dân.

A. 32000
B. 30000
C. 22000
D. 25000

Bài 16

Để xác định số cá thể A trong một quần thể, người ta bắt ngẫu nhiên 100 cá thể A và đánh dấu rồi thả lại về quần thể. Sau một thời gian người ta bắt ngẫu nhiên 500 cá thể A thấy có 20 cá thể bị đánh dấu. Hãy ước lượng số lượng cá thể A trong quần thể.

A. 2000
B. 3000
C. 2200
D. 2500

Bài tập

Bài 17

Kiểm tra lượng tạp chất trong 9 sản phẩm được ta thu được mẫu như sau (đơn vị tính là gam):

$$18{,}2 \quad 13{,}7 \quad 15{,}9 \quad 17{,}4 \quad 21{,}8 \quad 16{,}6 \quad 12{,}3 \quad 18{,}8 \quad 16{,}2$$

(a) Ước lượng điểm cho trọng lượng trung bình của tạp chất trong sản phẩm là bao nhiêu?

(b) Hãy tìm ước lượng điểm không chệch cho phương sai của trọng lượng của tạp chất trong sản phẩm.

(c) Hãy tìm ước lượng điểm cho sai số tiêu chuẩn của trung bình mẫu.

Bài tập

Bài 18

Điều tra doanh thu (triệu/tháng) của một chuỗi cửa hàng bán lẻ, người ta thu được bảng kết quả như sau

Doanh thu	6-10	10-14	14-18
Số cửa hàng	6	9	9

(a) Ước lượng điểm cho doanh thu trung bình của chuỗi cửa hàng trên là bao nhiêu?

(b) Hãy tìm ước lượng điểm cho mức độ không đồng đều về doanh thu (độ lệch chuẩn) của các cửa hàng.

(c) Hãy tìm ước lượng điểm cho sai số tiêu chuẩn của trung bình mẫu.

Bài tập

Bài 19

Xác suất để một mạng điện gặp sự cố trong một ngày là p. Quan sát mạng điện đó trong n ngày (n lớn), gọi X là số ngày mạng điện đó gặp sự cố trong n ngày quan sát.

(a) Giải thích tại sao có thể sử dụng $\frac{X}{n}$ để ước lượng cho p?

(b) Trình bày cách tính xấp xỉ xác suất sự sai khác giữa $\frac{X}{n}$ và p nhỏ hơn ϵ? Áp dụng cho $n = 365$, $p = 0,1$ và $\epsilon = 0,01$.

Bài tập

Bài 20

(a) Hãy chứng minh trung bình mẫu ngẫu nhiên $\overline{X}$ là ước lượng không chệch của trung bình tổng thể μ.

(b) Hãy chứng minh tần suất mẫu ngẫu nhiên $\widehat{P}$ là ước lượng không chệch của tỷ lệ trong tổng thể p.

(c) Hãy chứng minh phương sai mẫu ngẫu nhiên $\tilde{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ là ước lượng chệch của phương sai tổng thể σ^2.

Bài 21

Hãy chứng minh sai số bình phương trung bình của ước lượng $\widehat{\Theta}$ của tham số θ có thể được biểu diễn như sau:

$$MSE(\widehat{\Theta}) = E[(\widehat{\Theta} - \theta)^2] = (\text{độ lệch})^2 + V(\widehat{\Theta})$$

Bài tập

Bài 22

Giả sử $W_X = (X_1, X_2, \ldots, X_n)$ là một mẫu ngẫu nhiên kích thước n lấy từ tổng thể có phân phối Bernoulli với tham số $p \in (0; 1)$.

(a) Hãy tìm ước lượng mô men của tham số p.

(b) Hãy tìm ước lượng hợp lý cực đại của tham số p.

(c) Ước lượng mô men và ước lượng hợp lý cực đại của p ở trên có chệch hay không? Tại sao?

Bài 23

Giả sử $W_X = (X_1, X_2, \ldots, X_n)$ là một mẫu ngẫu nhiên kích thước n lấy từ tổng thể có phân phối Poisson với tham số $\lambda > 0$.

(a) Hãy tìm ước lượng mô men của tham số $\lambda > 0$.

(b) Hãy tìm ước lượng hợp lý cực đại của tham số $\lambda > 0$.

(c) Ước lượng mô men và ước lượng hợp lý cực đại của $\lambda > 0$ ở trên có chệch hay không? Tại sao?

Bài 24

Giả sử $W_X = (X_1, X_2, \ldots, X_n)$ là một mẫu ngẫu nhiên kích thước n lấy từ tổng thể có phân phối chuẩn với tham số μ và σ^2.

(a) Hãy tìm ước lượng mô men của các tham số μ và σ^2.

(b) Hãy tìm ước lượng hợp lý cực đại của các tham số μ và σ^2.

(c) Ước lượng mô men và ước lượng hợp lý cực đại của μ và σ^2 ở trên có chệch hay không? Tại sao?

Bài tập

Bài 25

Giả sử $W_X = (X_1, X_2, \ldots, X_n)$ là một mẫu ngẫu nhiên kích thước n lấy từ tổng thể có hàm mật độ xác suất với tham số $\lambda > 0$ như sau:

$$f(x;\lambda) = \begin{cases} \frac{1}{\lambda}e^{-\frac{x}{\lambda}}, & \text{nếu } x \geq 0, \\ 0, & \text{nếu } x < 0. \end{cases}$$

(a) Hãy tìm ước lượng mô men của tham số λ.

(b) Hãy tìm ước lượng hợp lý cực đại của tham số λ.

(c) Ước lượng mô men và ước lượng hợp lý cực đại của λ ở trên có chệch hay không? Tại sao?

Bài tập

Bài 26

Giả sử X là một biến ngẫu nhiên rời rạc có bảng phân phối xác suất như sau:

X	0	1	2	3
$P(X = x)$	$2\theta/3$	$\theta/3$	$2(1-\theta)/3$	$(1-\theta)/3$

Xét một mẫu được lấy từ tổng thể X như sau: $(3, 0, 2, 1, 3, 2, 1, 0, 2, 1)$.

(a) Hãy tìm ước lượng mô men của tham số θ.

(b) Hãy tìm ước lượng hợp lý cực đại của tham số θ.

Bài tập

Bài 27

Giả sử $W_X = (X_1, X_2, \ldots, X_n)$ là một mẫu ngẫu nhiên kích thước n lấy từ tổng thể X có hàm mật độ xác suất như sau:

$$f(x;\sigma) = \frac{1}{2\sigma}\exp\Big(-\frac{|x|}{\sigma}\Big).$$

(a) Hãy tìm ước lượng hợp lý cực đại của tham số σ.

(b) Ước lượng ước lượng hợp lý cực đại của σ ở trên có chệch hay không? Tại sao?