Chương 6

ƯỚC LƯỢNG ĐIỂM

6.1 Một số khái niệm tổng quan về ước lượng điểm

Các suy luận thống kê hầu hết xoay quay việc rút ra các kết luận về một hay một vài tham số của tổng thể. Để thực hiện được điều này ta thường tiến hành một điều tra quan sát mẫu của tổng thể về vấn đề đang quan tâm. Các kết luận được rút ra từ việc tính toán các giá trị của mẫu quan sát được.

Ví dụ ta ký hiệu μ là chiều cao trung bình của tất cả sinh viên nữ của Trường Đại học Sư phạm Kỹ thuật. Quan sát một mẫu ngẫu nhiên gồm chiều cao của $n = 200$ sinh viên nữ của trường ta, thu được các giá trị $x_1, x_2, ..., x_{200}$. Trung bình mẫu $\overline{x}$ sẽ được sử dụng để đưa ra kết luận về μ.

Một cách tổng quát ta ký hiệu θ là tham số đang quan tâm, chưa biết và cần rút ra kết luận về θ. Giả sử ta cần ước lượng θ dựa trên mẫu ngẫu nhiên cỡ n là $X_1, X_2, ..., X_n$. Một thống kê trên mẫu nhiên này là một hàm của các x_i, ví dụ như trung bình mẫu $\overline{X} = \dfrac{1}{n}\sum_{x=1}^{n} X_i$ cũng là một biến ngẫu nhiên.

Điều này cũng đúng với trường hợp dữ liệu quan sát gồm nhiều mẫu. Ví dụ như có 2 mẫu nhiên $X_1, X_2, ..., X_n$ và $Y_1, Y_2, ..., Y_m$; $\overline{X} - \overline{Y}$ là 1 hàm thống kê cho suy luận thống kê về sự khác biệt giữa 2 trung bình của 2 tổng là $\mu_1 - \mu_2$.

Định nghĩa 6.1: Một giá trị ước lượng điểm của tham số θ là một số được xem như là 1 giá trị hợp lý của θ. Một giá trị ước lượng điểm thu được bằng cách chọn 1 thống kê phù hợp và tính giá trị của thống kê trên dữ liệu của mẫu. Hàm thống kê này được gọi là **ước lượng điểm** của θ.

Hàm thống kê là ước lượng điểm của θ tính toán trên mẫu thực nghiệm, kí hiệu là $\widehat{\theta}$, gọi là giá trị ước lượng điểm.

Ví dụ 6.1: Quan sát thời gian sử dụng của máy điện thoại A (đơn vị: giờ) sau 10 lần sạc đầy pin ta có mẫu

$$x_1 = 5.5, x_2 = 5.33, x_3 = 6.2, x_4 = 6.5, x_5 = 7.0, x_6 = 5.66, x_7 = 4.5, x_8 = 6.5, x_9 = 4.8, x_{10} = 5.0.$$

Gọi μ là thời gian sử dụng trung bình của máy điện thoại A sau sạc đầy pin.

$\widehat{\mu} = \overline{X} = \dfrac{1}{10}\sum_{x=1}^{10} X_i$ là một ước lượng điểm của μ.

$\overline{x} = \dfrac{1}{10}(5.5+5.33+6.2+6.5+7.0+5.66+4.5+6.5+4.8+5.0) = 5.699$ (giờ) là giá trị ước lượng điểm của μ ứng với ước lượng điểm $\overline{X}$.

Gọi Y là số lần máy điện thoại A có thời gian sử dụng từ 6 giờ trở lên sau n lần sạc pin đầy.

p là khả năng phải máy điện thoại A có thời gian sử dụng ít nhất là 6 giờ.

$\widehat{p} = \dfrac{Y}{n}$ là một ước lượng điểm của p.

$\dfrac{y}{n} = \dfrac{5}{10} = 0.5$ là một giá trị ước lượng điểm của p.

$\widetilde{X} = \dfrac{\min_i X_i + \max_i X_i}{2}$ cũng là một ước lượng điểm khác của trung bình μ.

$\widetilde{x} = \dfrac{4.5+7.0}{2} = 5.75$ cũng là một giá trị ước lượng điểm của trung bình μ tương ứng với ước lượng điểm $\widetilde{X}$.

$\widehat{\sigma}^2 = S^2 = \dfrac{\sum(X_i - \overline{X})^2}{n-1}$ là một ước lượng điểm của phương sai σ^2.

$s^2 = \dfrac{\sum(x_i - \overline{x})^2}{n-1} = 0.6820544444$ là một giá trị ước lượng điểm của phương sai σ^2 ứng với ước lượng điểm S^2.

$S'^2 = \dfrac{\sum(x_i - \overline{x})^2}{n}$ cũng là một ước lượng điểm khác của phương sai σ^2 ứng với ước lượng điểm S'^2.

$s'^2 = \dfrac{\sum(X_i - \overline{X})^2}{n} = 0.613849$ là một giá trị ước lượng điểm của phương sai σ^2.

Như vậy tương ứng với một tham số θ ta có thể có nhiều ước lượng điểm. Tối ưu nhất ta cần tìm ước lượng điểm mà $\widehat{\theta} = \theta$. Tuy nhiên $\widehat{\theta}$ là một biến ngẫu nhiên nên giá trị ước lượng điểm có thể thay đổi trên các mẫu thực nghiệm khác nhau. Vậy làm thế nào lựa chọn một ước lượng điểm tốt nhất trong nhóm hữu hạn các ước lượng điểm, ta có một số tính chất mà một ước lượng điểm cần có như sau.

6.1.1 Ước lượng không chệch

Định nghĩa 6.2: Một ước lượng điểm $\widehat{\theta}$ được gọi là một ***ước lượng không chệch*** của θ nếu $E(\widehat{\theta}) = \theta$ với mọi giá trị có thể có của θ. Nếu $\widehat{\theta}$ là một ước lượng có chệch thì $E(\widehat{\theta}) - \theta$ được gọi là **sai số hệ thống của** θ.

Như vậy $\widehat{\theta}$ là một ước lượng không chệch của θ nếu phân phối xác suất của nó luôn có "trung tâm" tại giá trị đúng của tham số θ. "Trung tâm" ở đây là kỳ vọng của phân phối $\widehat{\theta}$.

Mệnh đề

X là biến ngẫu nhiên có phân phối nhị thức với hai tham số n và p. Tỷ lệ mẫu $\widehat{p} = \dfrac{X}{n}$ là một ước lượng không chệch của p.

Thật vậy, $E(\widehat{p}) = E\left(\dfrac{X}{n}\right) = \dfrac{1}{n}E(X) = \dfrac{1}{n}(np) = p$.

Mệnh đề

Cho $X_1, X_2, ..., X_n$ là các biến ngẫu nhiên có cùng phân phối với trung bình μ và phương sai σ^2. Khi đó ước lượng $\widehat{\sigma}^2 = S^2 = \dfrac{\sum(X_i - \overline{X})^2}{n-1}$ là một ước lượng không chệch của σ^2.

Chứng minh

Với Y là một biến ngẫu nhiên, $V(Y) = E(Y^2) - [E(Y)]^2$. Do đó ta có $E(Y^2) = V(Y) + [E(Y)]^2$.

$$\begin{aligned}
S^2 &= \frac{1}{n-1}\left[\sum X_i^2 - \frac{(\sum X_i)^2}{n}\right] \\
E(S^2) &= \frac{1}{n-1}\left\{\sum E(X_i^2) - \frac{1}{n}E(\sum X_i)^2\right\} \\
&= \frac{1}{n-1}\left\{\sum(\sigma^2 - \mu^2) - \frac{1}{n}\left\{V(\sum X_i) + [E(\sum X_i)]^2\right\}\right\} \\
&= \frac{1}{n-1}\left\{n\sigma^2 + n\mu^2 - \frac{1}{n}n\sigma^2 - \frac{1}{n}(n\mu)^2\right\} = \frac{1}{n-1}\left\{n\sigma^2 - \sigma^2\right\} = \sigma^2
\end{aligned}$$

Do đó S^2 là ước lượng không chệch của σ^2. Mặt khác $E(S'^2) = E\left(\dfrac{n-1}{n}S^2\right) = \dfrac{n-1}{n}\sigma^2 \neq \sigma^2$ nên $S'^2 = \dfrac{\sum(X_i - \overline{X})^2}{n}$ là một ước lượng chệch của σ^2.

Khi chọn một ước lượng điểm trong nhiều ước lượng điểm của cùng một tham số ta chọn ước lượng không chệch.

$\widehat{\mu}_1 = \overline{X} = \dfrac{1}{n}\sum_{i=1}^n X_i$ là một ước lượng điểm không chệch của μ.

$$\begin{aligned}
\text{Thật vậy}, E(\overline{X}) &= E\left(\frac{1}{n}\sum_{i=1}^n X_i\right) \\
&= \frac{1}{n}\sum_{i=1}^n E(X_i) = \frac{1}{n}\sum_{i=1}^n \mu = \frac{1}{n}n\mu = \mu
\end{aligned}$$

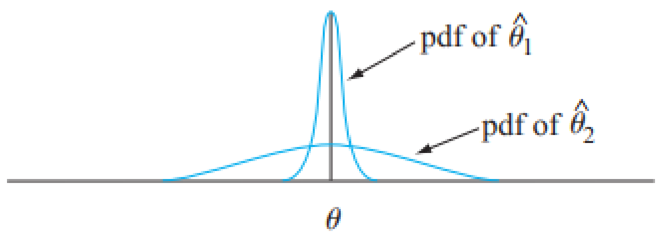
$\widehat{\mu}_2 = \dfrac{\min\limits_i X_i + \max\limits_i X_i}{2}$ cũng là một ước lượng điểm không chệch của μ.

$$\begin{aligned}
\text{Vì}, E(\widehat{\mu}_2) &= E\left(\frac{\min\limits_i X_i + \max\limits_i X_i}{2}\right) \\
&= \frac{1}{2}\left(E(\min_i X_i) + E(\max_i X_i)\right) \\
&= \frac{1}{2}\{\mu + \mu\} = \mu
\end{aligned}$$

Vậy chúng ta cần chọn ước lượng điểm không chệch nào trong lớp các ước lượng điểm không chệch của θ? Để trả lời câu hỏi này ta xét mục tiếp theo.

6.1.2 Ước lượng hiệu quả

Giả sử $\widehat{\theta}_1$ và $\widehat{\theta}_2$ là hai ước lượng điểm không chệch của θ. Phân phối của hai ước lượng này đều có trung tâm tại giá trị đúng θ, tuy nhiên độ phân tán các giá trị có thể có của $\widehat{\theta}_1$, $\widehat{\theta}_2$ xung quanh giá trị θ có thể khác nhau. Nếu phương sai của $\widehat{\theta}_1$ nhỏ hơn phương sai của $\widehat{\theta}_2$ có thể có của $\widehat{\theta}_1$ tập trung gần θ hơn các giá trị có thể có của $\widehat{\theta}_2$.



Hình 6.1: Đồ thị hai hàm phân phối của hai ước lượng điểm không chệch.

Định nghĩa 6.3: Trong lớp các ước lượng điểm không chệch của θ, ước lượng điểm $\widehat{\theta}$ có phương sai nhỏ nhất được ước lượng hiệu quả của θ so với các ước lượng không chệch khác.

Ước lượng hiệu quả có phương sai nhỏ nhất trong lớp tất cả các ước lượng không chệch của θ, hầu hết các giá trị của có thể có của ước lượng hiệu quả sẽ tập trung gần giá trị đúng θ.

Định lý

Cho $X_1, X_2, ..., X_n$ là các biến ngẫu nhiên có phân phối chuẩn với hai tham số μ và σ^2. Khi đó ước lượng điểm $\widehat{\mu} = \overline{X}$ là một ước lượng hiệu quả của μ.

6.1.3 Sai số chuẩn

Định nghĩa 6.4: ***Sai số chuẩn*** của ước lượng $\widehat{\theta}$ là độ lệch chuẩn của nó

$$\sigma_{\widehat{\theta}} = \sqrt{V(\widehat{\theta})}$$

Sai số chuẩn là tham số đo độ lệch chuẩn giữa một ước lượng điểm và giá trị đúng θ.

Nếu sai số chuẩn có chứa tham số chưa biết mà giá trị tham số đó có thể ước lượng thì ta sẽ ước lượng sai số chuẩn của một ước lượng. Ước lượng sai số chuẩn ký hiệu là $\widetilde{\sigma}_{\widehat{\theta}}$ hay $s_{\widehat{\theta}}$.

Ví dụ 6.2: Cho biến ngẫu nhiên X có phân phối chuẩn với trung bình μ và phương sai σ^2.

Quan sát một mẫu ngẫu nhiên cỡ $n = 20$ ta thu được mẫu thực nghiệm:

1.65	1.71	1.66	1.69	1.72	1.68	1.64	1.66	1.71	1.74
1.62	1.64	1.61	1.69	1.73	1.76	1.7	1.69	1.67	1.68

$\widehat{\mu} = \overline{X}$ là một ước lượng điểm tốt nhất của μ với giá trị ước lượng là $\overline{x} = 1.6825$

Nếu ta biết $\sigma = 0.05$ thì sai số chuẩn của $\overline{X}$ là:
$\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{0.05}{\sqrt{20}} = 0.0111803399.$

Nếu σ chưa biết, giá trị ước lượng của σ là $\widehat{\sigma} = s = 0.03971941061$ thì ta có ước lượng sai số chuẩn của ước lượng $\overline{X}$ là:
$\widehat{\sigma}_{\overline{x}} = s_{\overline{X}} = \dfrac{s}{\sqrt{n}} = 8.881530214.10^{-3}.$

Giả sử X là biến ngẫu nhiên có phân phối nhị thức với 2 tham số n và p. Tỉ lệ mẫu $\widehat{p} = \dfrac{X}{n}$ là ước lượng điểm tốt cho p với sai số chuẩn.

$$\sigma_{\widehat{p}} = \sqrt{V(\frac{X}{n})} = \sqrt{\frac{V(X)}{n^2}} = \sqrt{\frac{np(1-p)}{n^2}} = \sqrt{\frac{p(1-p)}{n}}$$

Khi một ước lượng điểm có phân phối xấp xỉ chuẩn, điều này thường xảy ra khi n lớn thì giá trị đúng của θ có nhiều khả năng thuộc vào khoảng với độ rộng là 2 sai số chuẩn của θ. Ví dụ như với mẫu có $n = 36$ có giá trị trung bình mẫu $\overline{x} = 29.35$ và độ lệch chuẩn mẫu $s = 0.36$ khi đó tính được ước lượng sai số chuẩn của $\widehat{\mu}$ là $\dfrac{s}{\sqrt{n}} = \dfrac{0.36}{\sqrt{36}} = 0.6$. Khi đó μ nhiều khả năng thuộc khoảng $29.35 \pm (2)(0.6) = (28.15; 30.55)$

Nếu ước lượng không chệch $\widehat{\theta}$ không xấp xỉ chuẩn ta không thể sử dụng công thức trên để đưa ra khoảng giá trị để dự đoán θ thuộc vào. Để khắc phục tình huống này gần đây trong thống kê hiện đại có một phương pháp gọi là bootstrap. Phương pháp bootstrap là phương pháp coi mẫu gốc ban đầu đóng vai trò tổng thể mà từ đó nó được rút ra. Từ mẫu ban đầu lấy lại các mẫu ngẫu nhiên cùng cỡ với mẫu gốc bằng phương pháp lấy mẫu có hoàn lại, gọi là mẫu bootstrap. Với mỗi mẫu lấy lại ta tính được giá trị tham số thống kê quan tâm gọi là tham số bootstrap. Sự phân bố của các tham số thống kê mẫu bootstrap là phân phối bootstrap.

Lấy mẫu có hoàn lại có nghĩa là sau khi chúng ta rút ra ngẫu nhiên một quan sát từ mẫu ban đầu, ta đặt nó trở lại trước khi lấy quan sát tiếp theo. Điều này cũng giống như lấy một số từ một chiếc mũ, sau đó đặt nó trở lại trước khi rút lại. Kết quả là, bất kỳ số có thể được rút ra một lần, nhiều hơn một lần, hoặc không được rút ra lần nào.

Với mẫu ban đầu $x = (x_1, x_2, ..., x_n)$, ta có mẫu bootstrap $x^* = (x_1^*, x_2^*, \ldots, x_n^*)$ với mỗi giá trị x_i^* được lấy ngẫu nhiên từ tập các giá trị $x_1, x_2, \ldots, x_n$ với xác suất $\frac{1}{n}$.

Tương ứng với mỗi mẫu bootstrap x^* ta có mô phỏng bootstrap của $\widehat{\theta}$ là $\widehat{\theta}^* = T(x^*)$, với hàm thống kê $T(x^*)$ tương tự với hàm thống kê $T(x)$ tác động lên mẫu x.

Mẫu lấy lại thứ nhất: $x_1^* = (x_{1_1}^*, x_{2_1}^*, ..., x_{n_1}^*)$ ta tính được giá trị $\widehat{\theta}_1^2$.

Mẫu lấy lại thứ hai: $x_2^* = (x_{1_2}^*, x_{2_2}^*, ..., x_{n_2}^*)$ ta tính được giá trị $\widehat{\theta}_2^*$.

...

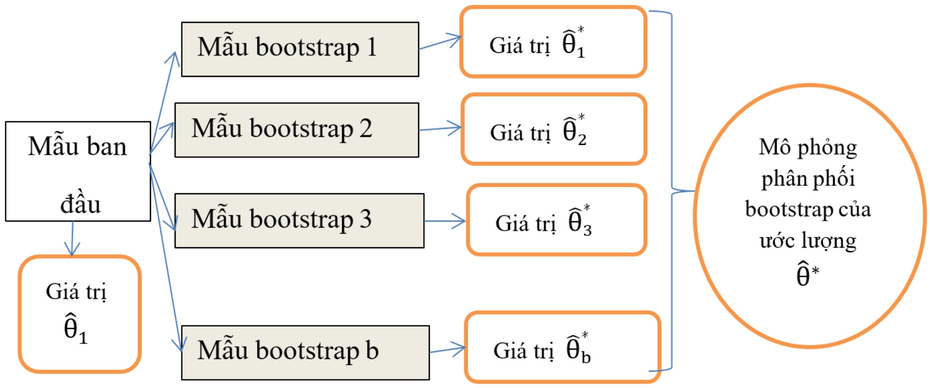
Mẫu lấy lại thứ B: $X_B^* = (x_{1_B}^*, x_{2_B}^*, ..., x_{n_B}^*)$ ta tính được giá trị $\widehat{\theta}_B^*$.

Với $B = 1000$ hay 2000.

Với mẫu bootstrap ngẫu nhiên $X^* = (X_1^*, X_2^*, \ldots, X_n^*), \widehat{\theta}^* = T(X^*)$ là một thống kê trên mẫu bootstrap, khi đó $F^*(t) = P(\widehat{\theta}^* \leq t)$ là phân phối bootstrap của $\widehat{\theta}^*$.

Các bước dùng phương pháp bootstrap ước lượng sai số tiêu chuẩn của $\widehat{\theta}$ từ một mẫu gốc ban đầu:

Bước 1: Lấy theo phương pháp có hoàn lại từ mẫu gốc ban đầu được B mẫu bootstrap



Hình 6.2: Sơ đồ mô phỏng phân phối bootstrap.

độc lập cùng cỡ với mẫu gốc $x_k^* = (x_(k_1)^*, x_(k_2)^*, \ldots, x_(k_n)^*), k = 1, 2, \ldots, B$.

Bước 2: Với mỗi mẫu bootstrap có được ở bước 1 ta tính giá trị thống kê $\widehat{\theta}_k^* = T(x_k^*), k = 1, 2, \ldots, B$.

Bước 3: Tính độ lệch chuẩn của B giá trị tính được ở bước 2.

$$S_{\widehat{\theta}} = \sqrt{\frac{1}{B-1}\sum(\widehat{\theta}_i^* - \overline{\theta}^*)^2}$$

với $\overline{\theta}^* = \dfrac{1}{B}\sum \widehat{\theta}_i^*$.

Độ lệch tiêu chuẩn này là ước lượng bootstrap của sai số tiêu chuẩn $\sigma_{\widehat{\theta}}$. Sai số chuẩn của ước lượng bootstrap là độ lệch chuẩn mẫu của giá trị $\widehat{\theta}_i^*$.

Ta có giá trị $S_{\widehat{\theta}}$ xấp xỉ $\sigma_{\widehat{\theta}}$ khi số lượng mẫu bootstrap B là lớn.

6.2 Các phương pháp ước lượng điểm

6.2.1 Phương pháp Moment

Định nghĩa 6.5: Cho $X_1, X_2, ..., X_n$ là một mẫu ngẫu nhiên có cùng phân phối xác suất.

Moment thứ k của X (moment tổng thể hay moment lý thuyết) là $E(X^k)$.

Moment mẫu thứ k của X là $\dfrac{1}{n}\sum_{i=1}^n X_i^k$.

Cụ thể ta có:

Moment thứ nhất của X là $E(X) = \mu$.

Moment mẫu thứ nhất của X là $\overline{X} = \dfrac{1}{n}\sum_{i=1}^n X_i$.

Moment thứ hai của X là $E(X^2)$.

Moment mẫu thứ hai của X là $\dfrac{1}{n}\sum_{i=1}^n X_i^2$.

Định nghĩa 6.6: Cho $X_1, X_2, ..., X_n$ là một mẫu ngẫu nhiên có cùng phân phối xác suất có hàm xác suất hay mật độ xác suất $f(x; \theta_1, \theta_2, ..., \theta_m)$ phụ thuộc vào m tham số chưa biết là $\theta_1, \theta_2, ..., \theta_m$. Ước lượng moment của m tham số này thu được bằng cách giải m phương trình m moment cấp k đầu tiên bằng m mẫu moment cùng bậc tương ứng.

Ví dụ 6.3: Cho X là biến ngẫu nhiên có phân phối Bernoulli với tham số p chưa biết. X ít có hàm xác suất

$$p_X(x) = \begin{cases} p & \text{với } x = 1, \\ 1-p & \text{với } x = 0. \end{cases}$$

Quan sát một mẫu ngẫu nhiên có cỡ $n = 10$ ta có bộ số liệu $(1, 1, 0, 1, 0, 1, 1, 1, 0, 0)$. Hãy tìm ước lượng moment của p.

Giải:

Moment thứ nhất của X là $E(X) = p$

Moment mẫu thứ nhất của X là $\overline{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i = 0.6$

Giải phương trình $E(X) = \overline{X}$ ta có $\widehat{p} = \overline{X}$ là ước lượng moment của p với giá trị ước lượng là 0.6.

Ví dụ 6.4: Thời gian hoạt động của loại thiết bị A là biến ngẫu nhiên X (đơn vị: giờ) có phân phối mũ với tham số λ. Quan sát thời gian hoạt động của một số thiết bị A ta có số liệu

$$(15.6, 16.0, 16.5, 15.8, 16.2, 17.5, 18.5, 17.8, 18.2, 16.8, 17.4, 17.9, 16.8, 16.0, 16.6)$$

Hãy tìm ước lượng của moment của λ.

Giải:

Moment thứ nhất của X là $E(X) = \dfrac{1}{\lambda}$.

Giá trị moment mẫu thứ nhất của X là $\overline{x} = 16.9$ (giờ)

$$E(X) = \overline{X} \Rightarrow \widehat{\lambda} = \frac{1}{\overline{X}} \text{ là một ước lượng moment của } \lambda.$$

Vậy giá trị ước lượng moment của λ là $\dfrac{1}{16.9}$.

Ví dụ 6.5: Điểm thi xác suất của sinh viên trường Đại học M là biến ngẫu nhiênS X có phân phối chuẩn $N(\mu, \sigma^2)$. Quan sát một mẫu ngẫu nhiên gồm $n = 1018$ bài thi của sinh viên trường M ta có bảng số liệu:

Điểm số	0	1	2	3	4	5	6	7	8	9	10
Số bài	2	24	50	98	151	193	189	162	95	46	8

Hãy tìm ước lượng moment của μ và σ^2.

Giải:

Moment cấp 1 của X là $E(X) = \mu$ với moment mẫu tương ứng là $\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$.

Moment cấp 2 của X là $E(X^2) = \sigma^2 + \mu^2$ với moment mẫu tương ứng là $\overline{X^2} = \dfrac{1}{n}\sum_{i=1}^{n} X_i^2$.

Giá trị moment mẫu cấp 1 của X là $\overline{x}_n = 5.411591356$.

Giá trị moment mẫu cấp 2 của X là $\dfrac{1}{n}\sum_{i=1}^{n} x_i^2 = 33.09921415$.

Giải hệ phương trình $\begin{cases} \mu = \overline{X} \\ \sigma^2 + \mu^2 = \dfrac{1}{n}\sum_{i=1}^{n} x_i^2 \end{cases}$

Ta có ước lượng moment của μ là $\widehat{\mu} = \overline{X}$ và giá trị ước lượng moment của μ bằng 5.411591356

Ước lượng moment của phương sai σ^2 là $\widehat{\sigma}^2 = \dfrac{1}{n}\sum_{i=1}^{n} X_i^2 - (\overline{X})^2$ và giá trị ước lượng moment của σ^2 bằng 3.813893141.

6.2.2 Phương pháp ước lượng hợp lý cực đại

Định nghĩa 6.7: Cho $X_1, X_2, ..., X_n$ có cùng phân phối $f(x_1, x_2, ..., x_n, \theta_1, \theta_2, ..., \theta_m)$ với $\theta_1, \theta_2, ..., \theta_m$ là các tham số chưa biết. Hàm $f(x_1, x_2, ..., x_n, \theta_1, \theta_2, ..., \theta_m)$ là hàm của các giá trị quan sát được $x_1, x_2, ..., x_n$ và $\theta_1, \theta_2, ..., \theta_m$ được gọi là hàm hợp lý. Ước lượng $\widehat{\theta}_1, \widehat{\theta}_2, ..., \widehat{\theta}_m$ làm cực đại hàm hợp lý gọi là ước lượng hợp lý cực đại của $\theta_1, \theta_2, ..., \theta_m$.

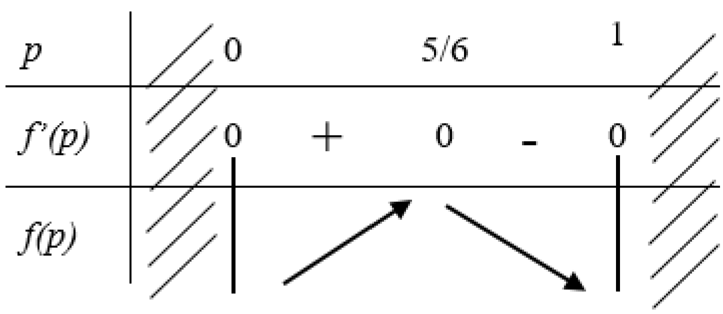
Ví dụ 6.6: Tỷ lệ sản phẩm đạt chuẩn của nhà máy B là p chưa biết. Một người kiểm tra từng sản phẩm cho đến khi được 10 sản phẩm đạt chuẩn thì dừng. Biết người này dừng lại ở lần kiểm tra thứ 12. Hãy tìm ước lượng hợp lý cực đại của p.

Giải:
Gọi X là số lần kiểm tra cho đến khi được 10 sản phẩm đạt chuẩn thì dừng $\Rightarrow X$ có phân phối nhị thức âm với hai hàm số là $r = 10$ và p.

$P(X = 12) = C_{11}^{10} p^{10}(1-p)^2 = f(p)$ làm hàm hợp lý cực đại của p.

$$\begin{aligned} f'(p) &= 11[10p^9(1-p)^2 - 2p^{10}(1-p)] \\ &= 11p^9(1-p)[10(1-p) - 2p] \\ &= 11p^9(1-p)(10-12p) \end{aligned}$$

$f'(p) = 0$ tại $p = 0$ hoặc $p = 1$ hoặc $p = \dfrac{10}{12} = \dfrac{5}{6}$



$f(p)$ đạt cực đại tại 5/6. Vậy giá trị ước lượng hợp lý cực đại của p tương ứng với quan sát người kiểm tra dừng lại ở lần kiểm tra thứ 12 là 5/6.

Ví dụ 6.7: Giả sử $X_1, X_2, ..., X_n$ là một mẫu ngẫu nhiên có phân phối mũ với tham số λ

$X_1, X_2, ..., X_n$ độc lập nên ta có hàm hợp lý

$$\begin{aligned} f(x_1, x_2, ..., x_n; \lambda) &= (\lambda e^{-\lambda x_1})...(\lambda e^{-\lambda x_n}) \\ &= \lambda^n e^{-\lambda \sum x_i} \end{aligned}$$

$$ln[f(x_1, x_2, ..., x_n; \lambda)] = nln(\lambda) - \lambda \sum x_i$$

Đạo hàm $ln[f(x_1, x_2, ..., x_n; \lambda)]$ theo λ và cho bằng 0 ta có

$$\frac{n}{\lambda} - \sum x_i = 0 \Leftrightarrow \lambda = \frac{n}{\sum x_i} = \frac{1}{\overline{x}}$$

Suy ra ta có ước lượng hợp lý cực đại của λ là

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

Kết quả này tương tự với kết quả thu được bằng phương pháp moment.

Tuy nhiên vì $E(\frac{1}{\overline{X}}) \neq \frac{1}{E(\overline{X})}$ nên đây là một ước lượng chệch của λ.

Bài tập 6.1

1. Quan sát điểm môn Toán của sinh viên trường Đại học A ta có bảng số liệu

5.9	7.3	6.8	7.8	8.1	8.8	6.5	6.3	7.6
4.5	8.2	7.9	7.6	6.9	5.6	5.2	5.4	7.3
4.8	9.0	9.1	8.3	8.4	4.8	5.5	5.6	6.5
6.4	3.5	9.5	4.0	6.8	7.0	7.3	7.8	7.9
5.2	5.1	6.8	7.5	7.3	7.1	7.9	9.8	5.0

   (a) Tìm một giá trị ước lượng điểm trung bình môn Toán của sinh viên trường Đại học A, chỉ ra ước lượng điểm tương ứng.

   (b) Tìm một giá trị của ước lượng điểm của tỷ lệ sinh viên trường Đại học A có điểm Toán từ 5 trở lên, chỉ ra ước lượng điểm tương ứng.

   (c) Tìm một giá trị ước lượng điểm của độ lệch chuẩn σ có điểm thi môn Toán sinh viên trường Đại học A, chỉ ra ước lượng điểm tương ứng.

   (d) Tìm một giá trị ước lượng điểm của tham số $\frac{\sigma}{\mu}$, chỉ ra ước lượng điểm tương ứng.

2. Quan sát số xe bán ra trong một ngày tại một số cửa hàng ta có bộ số liệu:

Số xe bán ra	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Số cửa hàng	3	5	7	10	11	15	18	21	19	17	15	13	12	10

   (a) Tìm một giá trị ước lượng điểm cho số xe trung bình bán ra trong một ngày tại các cửa hàng, chỉ ra ước lượng điểm tương ứng.

   (b) Tìm một giá trị ước lượng điểm cho tỉ lệ cửa hàng có số xe bán ra trong một ngày từ 10 xe trở lên, chỉ ra ước lượng điểm tương ứng

(c) Tìm một giá trị của những điểm cho phương sai cho số xe bán ra trong một ngày tại các cửa hàng, chỉ ra ước lượng điểm tương ứng.

(d) Tìm một giá trị ước lượng điểm cho trung vị của số xe bán ra trong một ngày tại các cửa hàng, chỉ ra ước lượng điểm tương ứng.

(e) Xác định sai số chuẩn của ước lượng trong câu (d).

3. (a) Quan sát số lượng ga (đơn vị: therm) sử dụng tại 10 ngôi nhà được chọn ngẫu nhiên tại vùng A trong tháng 1 , ta có bộ số liệu $103, 156, 118, 89, 125, 147, 122, 99, 138, 90$. Ký hiệu μ là lượng ga trung bình sử dụng tại mỗi nhà có sử dụng ga trong toàn bộ vùng A. Tính một giá trị ước lượng điểm của μ.

   (b) Giả sử có 10000 hộ sử dụng ga tại vùng A. Ký hiệu T là tổng lượng ga được sử dụng tại vùng A trong tháng 1. Ước lượng T với dữ liệu đã cho trong phần (a) và chỉ ra ước lượng điểm đã dùng.

   (c) Sử dụng dữ liệu trong phần (a), ước lượng tỉ lệ p là tỷ lệ hộ sử dụng ít nhất 100(therm)ga trong tháng 1.

4. Quan sát ngẫu nhiên n_1 nam giới hút thuốc, có X_1 người hút thuốc có đầu lọc. Quan sát ngẫu nhiên n_2 nữ giới hút thuốc, có X_2 người hút thuốc có đầu lọc. Ký hiệu p_1 và p_2 là tỷ lệ nam, nữ hút thuốc có đầu lọc trong số các nam giới, nữ giới có thuốc.

   (a) Chỉ ra rằng $\left(\dfrac{X_1}{n_1}\right) - \left(\dfrac{X_2}{n_2}\right)$ là một ước lượng không chệch của $p_1 - p_2$.

   (b) Xác định sai số chuẩn của ước lượng trong phần (a). Sử dụng các giá trị x_1, x_2 là các giá trị quan sát của các biến ngẫu nhiên X_1, X_2 để tính giá trị sai số chuẩn của ước lượng này.

   (c) Cho $n_1 = n_2 = 200, x_1 = 126, x_2 = 176$ sử dụng ước lượng trong phần a tính giá trị ước lượng của $p_1 - p_2$.

5. Cho mẫu ngẫu nhiên $X_1, .., X_n$ có cùng phân phối xác suất, xác định bởi hàm mật độ xác suất

$$f(x, \theta) = 0.5(1 + \theta x) \qquad -1 \le x \le 1$$

với $-1 \le \theta \le 1$. Chỉ ra rằng $\widehat{\theta} = 3\overline{X}$ là một ước lượng không chệch của θ.

Bài tập 6.2

1. Một xét nghiệm được áp dụng cho n người chắc chắn không bệnh. Gọi X là số xét nghiệm dương tính (tức là có bệnh) trong n kết quả tức là X là số kết quả dương tính sai. Và p là tỉ lệ người không có bệnh bị kết luận là có bệnh sau xét nghiệm.

   (a) Xác định ước lượng hợp lý cực đại của p. Nếu $n = 20$ và $x = 3$. Xác định giá trị của hợp lý cực đại của p.

   (b) Ước lượng điểm trong phần (a) có là ước lượng không chệch không?

   (c) Nếu $n = 20$ và $x = 3$ hãy xác định ước lượng hợp lý cực đại của xác suất $(1 - p)^5$ là không có kết quả dương tính nào trong 5 xét nghiệm đối với 5 người không có bệnh.

2. Cho biến ngẫu nhiên có hàm mật độ xác suất

$$f(x,\theta) = \begin{cases} (\theta+1)x^{\theta} & 0 \le x \le 1, \\ 0 & \text{trường hợp ngược lại.} \end{cases}$$

   với $\theta > -1$. Quan sát một mẫu ngẫu nhiên cỡ n của X ta thu được bộ số liệu

   $x_1 = 0.92; x_2 = 0.79; x_3 = 0.65; x_4 = 0.9; x_5 = 0.86; x_6 = 0.47; x_7 = 0.73; x_8 = 0.97; x_9 = 0.76.$

   (a) Sử dụng phương pháp moment tìm một ước lượng điểm của θ và tính giá trị của ước lượng này theo số liệu trên.

   (b) Tìm ước lượng hợp lý cực đại của θ rồi tính giá trị của lượng này theo số liệu trên.

3. X_1, X_2 là các biến ngẫu nhiên có phân phối Poisson với tham số λ_1, λ_2. Xác định ước lượng hợp lý cực đại của λ_1, λ_2 và $\lambda_1 - \lambda_2$.

4. Cho một mẫu ngẫu nhiên $X_1, X_2, ..., X_n$ có hàm mật độ xác suất

$$f(x,\lambda,\theta) = \begin{cases} \lambda e^{-\lambda(x-\theta)} & x \ge 0, \\ 0 & \text{trường hợp ngược lại.} \end{cases}$$

   (a) Xác định ước lượng hợp lý cực đại của θ và λ.

   (b) Với mẫu thực nghiệm $3.11, 0.64, 2.55, 2.20, 5.44, 3.42, 10.39, 8.93, 17.82$ và 1.3 tính giá trị ước lượng của θ và λ.