GIỚI THIỆU PHƯƠNG PHÁP ƯỚC LƯỢNG BAYEST

Đỗ Văn Huân - Viện KHTK

Hiện nay, trong thực tiễn công tác thống kê phương pháp ước lượng Bayest được sử dụng rất rộng rãi, do điều kiện thu thập số liệu khó khăn, và sử dụng phương pháp ước lượng này để đánh giá độ tin cậy của số liệu thống kê. Các phương pháp ước lượng cổ điển chỉ dựa trên thông tin trực tiếp thu thập được (gọi là thông tin trực tiếp) vì vậy đôi khi kết quả chênh lệch rất nhiều so với các nguồn thông tin khác, như thông tin dự đoán trong quá khứ, thông tin từ chuyên gia,... (gọi là thông tin gián tiếp), như vậy, vô hình chung các phương pháp ước lượng cổ điển đã để "lãng phí" một nguồn thông tin quan trọng trong việc nghiên cứu, đánh giá hiện tượng. Bài viết này sẽ giới thiệu phương pháp ước lượng Bayest nhằm hạn chế nhược điểm của các phương pháp ước lượng cổ điển. Phương pháp ước lượng Bayest dựa trên cả hai nguồn thông tin.

Gọi X là đại lượng ngẫu nhiên cần nghiên cứu

$x_1, x_2, x_3, ... x_n$ là các giá trị quan sát của đại lượng X

x^0 là giá trị tiên nghiệm (thông tin có trước)

x^* là giá trị tính toán được từ các quan sát của đại lượng ngẫu nhiên X

Gọi tham số u là mức ước lượng độ lớn tham số cần được ước lượng theo quyền số đối với hai nguồn thông tin trên về giá trị độ lớn tham số của đại lượng ngẫu nhiên X

w là quyền số (xác định tầm quan trọng của mỗi nguồn thông tin)

Ta có $u = w. x^* + (1 - w). x^0$; (1)

với $0 \le w \le 1$

Khi đó phương sai của ước lượng u sẽ là

$V(u) = w^2 V(x^*) + (1 - w)^2 V(x^0)$

Vì x^* và x^0 là độc lập với nhau.

Ta lấy đạo hàm hai vế của phương trình V(u) theo w

$[V(u)]' = 2w\, V(x^*) - 2(1 - w)\, V(x^0)$

Để ước lượng u là tốt nhất thì phải tìm số w (với $0 \le w \le 1$) để phương sai của ước lượng u (V(u)) nhỏ nhất, để V(u) là nhỏ nhất thì $[V(u)]' = 0$

$2w\, V(x^*) - 2(1 - w)\, V(x^0) = 0$

$2w\, V(x^*) - 2V(x^0) + 2w\, V(x^0) = 0$

$w\, [V(x^*) + V(x^0)] = V(x^0)$

$$w = \frac{V(x^0)}{V(x^*) + V(x^0)} = \frac{1}{\dfrac{V(x^*)}{V(x^0)} + 1} \quad ; \quad (2)$$

thay w vào biểu thức (1) ta có

$$u = \frac{1}{\dfrac{V(x^*)}{V(x^0)} + 1}.x^* + \left(1 - \frac{1}{\dfrac{V(x^*)}{V(x^0)} + 1}\right).x^0$$

$$= \frac{\left(x^* + x^0.\dfrac{V(x^*)}{V(x^0)}\right)}{\dfrac{V(x^*)}{V(x^0)} + 1} \quad ; \quad (3)$$

- Áp trong trường hợp ước lượng số trung bình

Gọi X là đại lượng ngẫu nhiên cần nghiên cứu

Các quan sát tham số cần nghiên cứu của tổng thể $X = \{x_1, x_2, x_3, ... x_n\}$

$V(x) = \sigma^2$ là phương sai của x.

n là số quan sát của tổng thể X

x_i là giá trị đặc tính X quan sát được ở cá thể thứ i

Ta có $\overline{x} = x^* = \dfrac{\sum_{i=1}^{n} x_i}{n}$ là số trung bình của tổng thể, chính là ước lượng cổ điển cho kỳ vọng đại lượng X

Gọi x^0 là kỳ vọng tiên nghiệm (kỳ vọng có được từ thông tin có trước)

Và có $V(x^0) = \sigma_0^2$ là số đã biết

Từ $x^* = \dfrac{\sum_{i=1}^{n} x_i}{n}$ với x_i là những giá trị của quan sát thứ i là độc lập với nhau và mỗi quan sát x_i được quan sát từ cùng một tổng thể nên chúng có cùng một phương sai $V(x_i) = \sigma_i^2 = \sigma^2$ nên ta có

$$V(x^*) = \frac{1}{n^2}\sum_{i=1}^{n} V(x_i) = \frac{1}{n^2}\sum_{i=1}^{n} \sigma_i^2$$

$$= \frac{1}{n^2}.n\sigma^2 = \frac{1}{n}.\sigma^2$$

Thay $V^2(x^*)$ vào phương trình (2) ta có:

$$w = \frac{1}{\dfrac{\sigma^2}{n.\sigma_0^2} + 1}$$

Thay giá trị w vào (3) ta được

$$u = \frac{\overline{x} + x^0.\dfrac{\sigma^2}{n.\sigma_0^2}}{\dfrac{\sigma^2}{n.\sigma_0^2} + 1}$$

Ví dụ: Trong điều tra sản lượng lúa của một địa phương (có N thửa ruộng có diện tích như nhau), giả sử gọi X là sản lượng lúa, với giả thiết X là đại lượng ngẫu nhiên tuân theo quy luật phân phối chuẩn. Để tính sản lượng lúa trung bình trên mỗi thửa ruộng của địa phương đó, người ta gặt thí điểm n thửa ruộng và kết quả cho được như sau: (xem bảng trang bên)

Thửa ruộng thứ (i)	1	2	3	4	5	6	7	8	9	10	11
Sản lượng (kg) (x_i)	450	465	784	526	458	465	486	520	472	761	453

Và có nguồn thông tin từ chuyên gia cho rằng sản lượng lúa của địa phương trên trung bình là x^0 = 500 kg/thửa với độ lệch tiêu chuẩn là 20.

Từ kết quả gặt thí điểm trên ta tính ra được sản lượng trung bình trên mỗi thửa ruộng là 475kg. Tức là kỳ vọng của sản lượng trên mỗi thửa rộng ở địa phương trên là 475kg, như vậy kết quả chênh lệch tương đối lớn so với kết quả của chuyên gia, nếu lấy w = 0,5 tức là cộng hai số trung bình lại rồi chia cho hai, đây cũng là một cách nhưng không có cơ sở khoa học, người ta dùng phương pháp ước lượng Bayest để ước lượng sản lượng lúa trung bình của địa phương trên.

Từ kết quả gặt thí nghiệm ta tính phương sai cho những điểm gặt đó:

$$\sigma^2 = \frac{\sum_{i=1}^{11}(x_i - \overline{x})^2}{11} = \frac{6746}{11} = 613{,}273$$ thay các giá trị vào để tính w

$$w = \frac{1}{\dfrac{\sigma^2}{n.\sigma_0^2} + 1} = \frac{1}{\dfrac{613{,}273}{11.20^2} + 1} = 0{,}8776$$

u = 0,8776 . 475 + (1- 0,8776). 500 = 478 (kg)

- Áp trong trường hợp ước lượng số tỷ lệ

Gọi X là đại lượng ngẫu nhiên tuân theo quy luật không – một (ký hiệu là A(p)), tức là X chỉ có thể nhận giá trị 1 hoặc 0.

n_A: Số cá thể có thuộc tính cần nghiên cứu

n: Số cá thể được quan sát

$p = \dfrac{\sum_{i=1}^{n} x_i}{n} = \dfrac{n_A}{n}$ là tỷ lệ số cá thể có thuộc tính cần nghiên cứu (Đại lượng ngẫu nhiên tuân theo quy luật A(p) thì số tỷ lệ cũng chính là số bình quân)

$$V(x_i) = \sigma_i^2 = \sigma^2 = p(1-p)$$

f^0 : là kỳ vọng tiên nghiệm về tỷ lệ thuộc tính cần nghiên cứu

và có $V(f^0) = \sigma_0^2$

Ta có

$$V(p) = V\left(\frac{\sum_{i=1}^{n} x_i}{n}\right) = \frac{n.p(1-p)}{n^2} = \frac{p(1-p)}{n}$$

Thay V(p) và $V(f^0)$ vào (2) ta có

$w = \dfrac{1}{\dfrac{p(1-p)}{n.\sigma_0^2} + 1}$ thay w vào (3) ta

có: $u = \dfrac{\left[p + f^0.\dfrac{p(1-p)}{n.\sigma^2}\right]}{\dfrac{p(1-p)}{n.\sigma_0^2} + 1}$

Ví dụ: Khi đánh giá chất lượng của một dây truyền sản xuất mới đưa vào ứng dụng, người ta sản xuất thử n sản phẩm và kiểm tra chất lượng. Sản phẩm chỉ có thể là đạt yêu cầu hoặc không đạt yêu cầu, chẳng hạn khi ta kiểm tra 250 sản phẩm thì thấy có 25 sản phẩm không đạt yêu cầu, nhưng theo thiết kế thì các chuyên gia cho rằng dây truyền sản xuất của nhà máy khi đưa vào sản xuất thì tỷ lệ sản phẩm không đạt yêu cầu chỉ là 5%, với độ đồng đều (độ lệch chuẩn) là 0,02

Từ kết quả kiểm tra thực tế ta có tỷ lệ sản phẩm không đạt yêu cầu là:

$p = \dfrac{n_A}{n} = \dfrac{25}{250} = 10\,(\%)$, phương sai của p là

$\sigma^2 = \dfrac{p(1-p)}{n} = \dfrac{0{,}9.0{,}1}{250} = 0{,}00036$ Ta có

$$w = \frac{1}{\dfrac{p(1-p)}{n.\sigma_0^2} + 1} = \frac{1}{\dfrac{0{,}1.0{,}9}{250.0{,}0004} + 1} = 0{,}526$$

u = 0,526. 0,1 + (1 – 0,526). 0,05

= 0,0763 hay 7,63% sản phẩm không đạt yêu cầu ■

Nguồn: *Báo cáo sinh hoạt khoa học tháng 8/2002*