

## ƯỚC LƯỢNG KHÔNG CHỆCH (unbiased estimation)

cho tham số  $\theta$  của biến ngẫu nhiên  $X$  là hàm ước lượng  $n$  biến  $\hat{\theta}^{(n)} : \mathbb{R}^n \rightarrow \mathbb{R}$  sao cho  $\mathbb{E}(\hat{\theta}^{(n)}(X_1, \dots, X_n)) = \theta$ , đối với mọi mẫu ngẫu nhiên  $(X_1, \dots, X_n)$ , bao gồm  $n$  bản sao độc lập cùng phân bố với  $X$ . Trong nghiên cứu thống kê, tính "không chệch" thường được dùng để đánh giá chất lượng của một hàm ước lượng cho tham số thống kê  $\theta$  cần quan tâm, theo nghĩa trung bình của hàm ước lượng đó trên toàn không gian mẫu có giá trị đúng bằng  $\theta$ .

Ví dụ, cho  $(X_1, \dots, X_n)$  là mẫu của biến ngẫu nhiên  $X$  với kỳ vọng  $\mu$ . Dễ dàng thấy các hàm ước lượng  $\hat{\theta}_1^{(n)}(x_1, \dots, x_n) = x_1$ ,  $\hat{\theta}_2^{(n)}(x_1, \dots, x_n) = (x_1 + x_2)/2$ ,  $\dots$ ,  $\hat{\theta}_n^{(n)}(x_1, \dots, x_n) = \bar{X} = (x_1 + \dots + x_n)/n$  đều là các ước lượng không chệch của kỳ vọng  $\mu$ . Trong số đó, trung bình mẫu  $\bar{X}$  là ước lượng tốt nhất khi ước lượng đó có phương sai nhỏ nhất.

Theo định nghĩa thông thường, phương sai mẫu  $s^2 = \sum_{i=1}^n (x_i - \bar{X})^2/n$  không phải là một ước lượng không chệch cho phương sai  $\sigma^2$  của biến ngẫu nhiên  $X$ . Tuy nhiên, nếu chỉnh sửa lại một chút, phương sai mẫu hiệu chỉnh  $S^2 = \sum_{i=1}^n (x_i - \bar{X})^2/(n-1)$  lại là một ước lượng không chệch cho phương sai  $\sigma^2$  của biến ngẫu nhiên  $X$ .

Chú ý rằng, phương sai mẫu hiệu chỉnh sai chỉ khác phương sai mẫu một hệ số bằng  $(n-1)/n$ . Hệ số này dần tới 1 khi  $n$  tiến tới  $\infty$ . Do vậy, phương sai mẫu là ước lượng tiệm cận không chệch của tham số phương sai. Ta có định nghĩa của khái niệm này: Hàm ước lượng  $\hat{\theta}^{(n)}$  là một *ước lượng tiệm cận không chệch* của tham số  $\theta$ , nếu khi cỡ mẫu  $n$  tăng tới vô cùng thì  $\mathbb{E}(\hat{\theta}^{(n)}(X_1, \dots, X_n)) \rightarrow \theta$ .

Các ước lượng không chệch có ý nghĩa quan trọng trong các nghiên cứu thống kê. Có thể xác định được phân bố xác suất của nhiều hàm thống kê lập từ các ước lượng không chệch, trong khi nếu thiếu điều kiện không chệch thì không thể tìm được phân bố (thậm chí là phân bố tiệm cận) của các thống kê đó.

### **Tài liệu tham khảo**

1. H. Cramér, *Mathematical Methods of Statistics*, Princeton Univ. Press, NJ, 1946.
2. R. A. Fisher, *Statistical Methods and Scientific Inference*, Hafner, London, 1973.
3. E. L. Lehmann, *Testing Statistical Hypotheses*, Wiley, New York, 1986.