ƯỚC LƯỢNG KHOẢNG (interval estimation)

một đoạn thẳng $(b; c)$ chứa θ sao cho $\mathbb{P}\{\theta \in (b; c)\} = a\%$, trong đó $\hat{\theta}$ là ước lượng của tham số một chiều θ và $a \in (0; 100)$ là một số cho trước. Ước lượng khoảng được xét đến khi khó xác định chính xác giá trị thực của tham số θ mà phải dùng một hàm ước lượng nào đó để ước đoán giá trị thực của tham số. Việc ước đoán này thường gặp phải sai số và phải đánh giá mức độ sai sót của phép ước lượng. Nói cách khác, với một độ chính xác cho trước, cần kết luận phép ước lượng có cho ra kết quả chấp nhận được không. Ở đây, "độ chính xác" có thể hiểu là mức độ thu hẹp của lân cận chứa tham số cần ước lượng, "chấp nhận được" có nghĩa là miền lân cận đó có xác suất chứa giá trị đúng của tham số đủ gần 1, khả năng để tham số cần ước lượng nằm trong miền được chỉ ra là đủ lớn. Trong thống kê toán học, khái niệm "khoảng tin cậy" được đưa ra để diễn tả độ chính xác của ước lượng ứng với một xác suất nào đó. Khoảng tin cậy của ước lượng còn được gọi là *ước lượng khoảng* và thường được dùng trong thực tế, nhất là trong các bài toán kiểm định so sánh các giá trị của tham số.

Ví dụ. Giả sử X là một biến ngẫu nhiên phân bố chuẩn với kỳ vọng μ chưa biết (cần ước lượng) và phương sai σ^2 đã biết. Trung bình mẫu $\bar{X}$ là một ước lượng của tham số μ, khi đó $(\bar{X} - \mu) \cdot \sqrt{n}/\sigma$ là một biến ngẫu nhiên có phân bố chuẩn tắc $N(0; 1)$, khoảng tin cậy $a\%$ của $\bar{X}$ là đoạn thẳng

$$CI(a\%; \bar{X}) = \left(\bar{X} - Z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\, ; \bar{X} + Z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right),$$

trong đó $\alpha = (100 - a)/100$, còn Z_β là phân vị mức β của phân bố xác suất chuẩn tắc $N(0; 1)$. Chẳng hạn, với $\alpha = 5\%$ thì khoảng tin cậy 95% của $\bar{X}$ sẽ là

$$CI(95\%; \bar{X}) = \left(\bar{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}}\, ; \bar{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right).$$

Khi X có phân bố Bernoulli, nhận hai giá trị 1 và 0 với xác suất tương ứng là p và $1 - p$, $p \in (0; 1)$, rõ ràng $\hat{p} = \bar{X}$ là ước lượng của xác suất p. Theo Định lý Moivres – Laplace, $\bar{X}$ có phân bố tiệm cận phân bố chuẩn với kỳ vọng p và phương sai $p(1-p)/n$, khoảng tin cậy 95% của $\hat{p}$ là

$$CI(95\%; \hat{p}) = \left(\hat{p} - 1.96 \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}; \hat{p} + 1.96 \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\ \right).$$

Các khoảng tin cậy cho ước lượng của kỳ vọng cho biến ngẫu nhiên phân bố chuẩn và ước lượng của tham số xác suất cho biến ngẫu nhiên phân bố Bernoulli được sử dụng trong bài toán xác định cỡ mẫu của các nghiên cứu thực nghiệm. Đối với các ước lượng tham số khác, khoảng tin cậy cũng được xác định bằng cách tương tự khi biết được phân bố xác suất của ước lượng hoặc khi chỉ ra được dáng điệu tiệm cận về mặt phân bố xác suất của ước lượng tham số cần xét.

HỒ ĐĂNG PHÚC

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.