PHÂN PHỐI MẪU

BỘ MÔN THỐNG KÊ TOÁN HỌC
KHOA TOÁN - TIN HỌC
ĐẠI HỌC KHOA HỌC TỰ NHIÊN TP.HCM

Tháng 9 năm 2021

Outline

Outline

Mẫu ngẫu nhiên

Định nghĩa 1

Các biến ngẫu nhiên $X_1, X_2, \ldots, X_n$ là một **mẫu ngẫu nhiên** kích thước n nếu

(i) X_i là các biến ngẫu nhiên độc lập nhau.
(ii) Mọi X_i đều có cùng một phân phối xác suất.

Thống kê

Định nghĩa 2

Một **thống kê** (statistic) là một hàm bất kì của các quan sát trong một mẫu ngẫu nhiên.

Ví dụ 3 (Các thống kê thường dùng)

Nếu $X_1, X_2, \ldots, X_n$ là một mẫu ngẫu nhiên kích thước n, thì

- Trung bình mẫu: $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
- Phương sai mẫu: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$
- Độ lệch chuẩn mẫu: $S = \sqrt{S^2}$
- Trung vị mẫu: Sắp xếp mẫu theo thứ tự tăng dần rồi lấy giá trị ở vị trí trung tâm làm trung vị. Nếu kích thước mẫu là chẵn thì trung vị là giá trị trung bình của hai giá trị trung tâm.

đều là các thống kê.

Phân phối mẫu

Bởi vì một thống kê là một biến ngẫu nhiên, nên nó có phân phối xác suất

Định nghĩa 4

Phân phối xác suất của một thống kê được gọi là một **phân phối mẫu**

Ví dụ 5

Phân phối xác suất của $\bar{X}$ được gọi là **phân phối mẫu của trung bình**

Nhận xét 6

Phân phối mẫu của một thống kê phụ thuộc vào phân phối của tổng thể, kích thước mẫu, và phương pháp chọn mẫu.

Outline

Phân phối mẫu của trung bình và phương sai

Trường hợp tổng thể có phân phối chuẩn

Định lí 7

Giả sử $(X_1, X_2, \ldots, X_n)$ là một mẫu ngẫu nhiên được lấy từ một tổng thể có phân phối chuẩn với trung bình μ và phương sai σ^2. Khi đó,

(i) $\bar{X}$ *và* S^2 *độc lập với nhau.*

(ii) $\bar{X} \sim N(\mu, \sigma^2/n)$.

(iii) $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$.

(iv) $\dfrac{\bar{X}-\mu}{S/\sqrt{n}} \sim T(n-1)$.

Phân phối mẫu của trung bình và phương sai

Trường hợp tổng thể có phân phối xác suất chưa biết

Định lí 8

Xét mẫu ngẫu nhiên $X_1, X_2, \cdots, X_n$ lấy từ một phân phối có trung bình μ hữu hạn và phương sai dương σ^2. Ta có biến ngẫu nhiên $Y_n = \sqrt{n}\left(\bar{X} - \mu\right)/\sigma$ và $Z_n = \sqrt{n}\left(\bar{X} - \mu\right)/S$ đều có phân phối xấp xỉ với phân phối chuẩn $N(0;1)$.

Trong thực hành khi mẫu có kích thước đủ lớn ($n \geq 30$), ta có các phân phối xấp xỉ chuẩn sau:

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \approx N(0,1)$$

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \approx N(0,1)$$

Phân phối mẫu của tỷ lệ

Giả sử cần khảo sát đặc trưng $\mathcal{A}$ của tổng thể, khảo sát n phần tử và đặt

$$X_i = \begin{cases} 1 & \text{nếu thỏa } \mathcal{A} \\ 0 & \text{nếu khác} \end{cases}$$

thu được mẫu ngẫu nhiên $X_1, \ldots, X_n$ với $X_i \sim B(1,p)$, với p là tỷ lệ phần tử thỏa đặc trưng $\mathcal{A}$.

Khi đó, $\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \equiv \hat{p}$ được gọi là tỷ lệ mẫu. Đây là một ước lượng của tỷ lệ tổng thể p.

Hơn nữa,

$$\mathbb{E}(\bar{X}) = p, \quad \mathbb{V}ar(\bar{X}) = \frac{p(1-p)}{n}$$

Phân phối mẫu của tỉ lệ

tt

Định lí 9

Xét mẫu ngẫu nhiên $X_1, X_2, \cdots, X_n$ lấy từ một phân phối Bernoulli $B(1;p)$. Ta có các biến ngẫu nhiên $\frac{(\hat{p}-p)\sqrt{n}}{\sqrt{p(1-p)}}$ và $\frac{(\hat{p}-p)\sqrt{n}}{\sqrt{\hat{p}(1-\hat{p})}}$ có phân phối xấp xỉ với phân phối chuẩn $N(0;1)$.