Luật số lớn và định lí giới hạn trung tâm

Giảng viên: Hoàng Thị Điệp
Khoa CNTT – Đại học Công Nghệ

Nội dung

» Luật số lớn
» Định lí giới hạn trung tâm

Luật số lớn
(law of large numbers)

» Khi chúng ta tiến hành một thí nghiệm nhiều lần, thì kết quả trung bình nhận được sẽ gần với giá trị kì vọng, và càng ngày càng gần hơn khi chúng ta tiến hành thêm nhiều lần nữa.

Ví dụ: Gieo 1 con súc sắc với 6 mặt đều nhau có giá trị 1,2,3,4,5,6. Giá trị kì vọng của số ở mặt trên:

$$EX = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} = 3.5$$

Nếu ta tung con súc sắc nhiều lần, giá trị trung bình của số ở mặt trên sẽ rất gần giá trị kì vọng 3.5.

- **H là chiều cao (đơn vị m) của 1 sinh viên chọn ngẫu nhiên trong lớp**
- H là biến ngẫu nhiên; 1 <= H <= 3
- Bạn Công chọn H là bnn liên tục, phân bố đều
- f(x) = 1/2 với 1<=x<=3
- f(x) = 0 với x khác
- **H1, H2, …., H10 là chiều cao (đơn vị m) của 10 sinh viên chọn ngẫu nhiên trong lớp**

Bất đẳng thức Chebyshev (dạng hệ quả)

Giả sử X là ĐLNN với μ =EX. Khi đó với mọi $\varepsilon > 0$:

$$P\{|X-\mu| > \varepsilon\} \leq \frac{DX}{\varepsilon^2}$$

Ứng dụng:

- Đánh giá cận trên hoặc cận dưới xác suất để X nhận giá trị sai lệch không quá ε so với kì vọng
  - từ đó lý giải cho các sai số trong đo lường vật lý

Thí dụ 6

Để xác định giá trị của một đại lượng vật lí nào đó, người ta thường tiến hành đo n lần một cách độc lập và lấy trung bình cộng các kết quả đo ấy làm giá trị thực của đại lượng cần đo. Giả sử kết quả n lần đo là $X_1, X_2, \ldots, X_n$.

Chúng là các ĐLNN độc lập với nhau và cùng phân bố.

Kì vọng $EX_i = \mu$ chính là giá trị thực của đại lượng vật lí đó.

$DX_i = \sigma^2$ đặc trưng cho độ chính xác của phép đo (Thiết bị đo càng chính xác thì σ^2 càng nhỏ).

Sai số khi lấy trung bình cộng của n lần đo là :

$$\frac{X_1 + \ldots + X_n}{n} - \mu.$$

Áp dụng hệ quả của bất đẳng thức Trebusép ta có

$$P\left\{\left|\frac{\sum_{i=1}^{n} X_i}{n} - \mu\right| > \varepsilon\right\} \leqslant \frac{D\left(\sum_{i=1}^{n} X_i\right)}{n^2\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}$$

vì $$E\left(\frac{\sum X_i}{n}\right) = \mu \text{ và } D\left(\frac{\sum X_i}{n}\right) = \frac{D(\sum X_i)}{n^2} = \frac{n\sigma^2}{n^2}.$$

Cho trước sai số được phép ε. Bài toán đặt ra là cần tiến hành bao nhiêu phép đo để với xác suất 0,99 sai số của trung bình cộng so với giá trị thực không vượt quá ε. Từ đánh giá trên ta suy ra để :

$$P\left\{\left|\frac{\sum_{i=1}^{n} X_i}{n} - \mu\right| > \varepsilon\right\} \leqslant 0{,}01$$

ta cần có $\dfrac{\sigma^2}{n\varepsilon^2} \leqslant 0{,}01$ hay $n > \dfrac{100\sigma^2}{\varepsilon^2}$.

Luật số lớn
(law of large numbers)

Vận dụng bất đẳng thức Chebyshev

$$P\left(\left|\frac{X1+X2+\ldots+Xn}{n}-\mu\right|\geq\varepsilon\right)\leq\frac{\sigma^2}{n\varepsilon^2}$$

Biến đổi toán học: xem Thí dụ 6 (tr. 157)

Định lí giới hạn trung tâm
Central limit theory

Cử 10 sinh viên đi khảo sát mức độ hài lòng (cho điểm từ 0-100) của sinh viên ĐHCN. Mỗi bạn sinh viên đi khảo sát chọn 1 tập mẫu (sample) 20 bạn sinh viên khác nhau và tính số điểm trung bình của mức độ hài lòng.

» Có 10 mẫu (sample)

» Kích thước mẫu là 20 (số sinh viên trong mỗi mẫu là 20)

	1	2	3	4	5	6	7	8	9	10
Điểm trung bình	70	76	78	80	79	81	83	72	85	87

Nhận xét: Giá trị trung bình của các mẫu có phân bố chuẩn (không cần biết phân bố của không gian mẫu).

Định lí giới hạn trung tâm
Central limit theory

Giả sử S={X_1, X_2,…,X_n} là một mẫu, hay một dãy các biến ngẫu nhiên độc lập có cùng phân bố với kì vọng μ và phương sai σ^2. Trung bình cộng

$$S_n = \frac{X_1+X_2+\cdots+X_n}{n}$$

Theo luật số lớn S_n sẽ tiến gần đến μ theo xác suất. S_n có phân bố chuẩn với kì vọng μ và phương sai σ^2/n.

Lưu ý: $T_n = (X_1 + X_2 + \cdots + X_n)$ có phân bố chuẩn với kì vọng nμ và phương sai $n\sigma^2$.

n lớn bao nhiêu thì xấp xỉ là tốt?

- Nếu phân bố của $X_1, X_2,\ldots,X_n$ là đối xứng
  - n>20 là đủ
- Nếu không
  - cần n lớn hơn

Ví dụ

0. Gieo con súc sắc 1 lần. Gọi Y là số nốt ở mặt trên.

a) Lập bảng phân bố của Y

b) Tính EY, DY

1. Gieo con súc sắc 30 lần. Tìm xác suất để tổng số nốt xuất hiện

a) lớn hơn 135.

b) nhỏ hơn 90.

Ví dụ

1. Gieo con súc sắc 30 lần. Tìm xác suất để tổng số nốt xuất hiện

b) nhỏ hơn 90.

c) lớn hơn 150.

2. Khối lượng của người VN có phân bố chuẩn, trung bình là 65kg với độ lệch chuẩn là 5kg. Một thang máy cho phép đi tối đa 10 người, và có khối lượng không quá 700kg. Tính xác suất để 10 người bất kì đi vào thang máy không bị quá tải.

3. Một nông dân nhận thấy sản lượng trung bình của 1 cây khoai tây là 1,82kg, với độ lệch chuẩn là 0,34kg. Năm nay ông bón thêm một loại phân mới và thu được 1395kg trên 750 cây. Hỏi rằng loại phân mới này có thực sự làm thay đổi sản lượng của cây khoai tây? Biết mức ý nghĩa = 0.01.

Ví dụ 3

Một nông dân nhận thấy sản lượng trung bình của 1 cây khoai tây là 1,82kg, với độ lệch chuẩn là 0,34kg. Năm nay ông bón thêm một loại phân mới và thu được 1395kg trên 750 cây. Hỏi rằng loại phân mới này có thực sự làm thay đổi sản lượng của cây khoai tây? Biết mức ý nghĩa = 0.01.

- Giả sử phân bón mới không làm thay đổi sản lượng cây khoai tây => sản lượng của mỗi cây khoai tây bón bằng phân bón mới có thể biểu diễn bằng biến ngẫu nhiên S_i
- ....
- P(S>=1395) = 1 – 0.9994 < mức ý nghĩa
- Suy ra, điều giả sử không đáng tin cậy. Tức là phân bón mới CÓ làm thay đổi sản lượng khoai tây.
- Tham khảo thêm: Thí dụ 15, tr.172; Thí dụ 21. tr.179

Hiệu chỉnh xác suất

» Xem trang 176