Chương 3
BIẾN NGẪU NHIÊN NHIỀU CHIỀU

BỘ MÔN TOÁN ỨNG DỤNG[1]

VIỆN TOÁN ỨNG DỤNG VÀ TIN HỌC
ĐẠI HỌC BÁCH KHOA HÀ NỘI

SAMI.HUST – 2023

[1] Phòng BIS.201-D3.5

Ví dụ

Ví dụ 23

(a) Gieo n lần một đồng xu cân đối đồng chất. Gọi x là số lần xuất hiện mặt sấp trong n lần gieo. Khi đó, tỷ số x/n được gọi là tần suất xuất hiện mặt sấp. Người ta thấy rằng nếu số lần gieo càng lớn thì x/n càng gần tới $1/2$ (xem Chương 1).

(b) Lặp lại n lần đo độc lập biến ngẫu nhiên X trong cùng một điều kiện như nhau, kết quả của các lần đo là $x_1, x_2, \ldots, x_n$. Thực nghiệm cho thấy rằng trung bình số học $\frac{x_1+x_2+\cdots+x_n}{n}$ gần với số $E(X)$ (xem Chương 2).

✍ Câu hỏi đặt ra:

1. "Với điều kiện nào thì x/n hội tụ về xác suất $p = P(A)$?"
2. "Với điều kiện nào thì $\frac{x_1+x_2+\cdots+x_n}{n}$ hội tụ về $E(X)$?"

3.4. LUẬT SỐ LỚN

Hội tụ theo xác suất

Định nghĩa 12

Xét dãy biến ngẫu nhiên $\{X_n\}_{n=1}^{\infty}$ và biến ngẫu nhiên X trong cùng một phép thử. Dãy các biến ngẫu nhiên $\{X_n\}_{n=1}^{\infty}$ được gọi là hội tụ theo xác suất về biến ngẫu nhiên X khi $n \to \infty$ nếu với mọi $\varepsilon > 0$,

$$\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0. \tag{31}$$

✍ Nếu dãy các biến ngẫu nhiên $\{X_n\}_{n=1}^{\infty}$ hội tụ theo xác suất về biến ngẫu nhiên X thì với n đủ lớn, thực tế gần như chắc chắn ta có thể coi rằng X_n không khác mấy so với X.

Hội tụ theo phân phối

Định nghĩa 13

Dãy các biến ngẫu nhiên $\{X_n\}_{n=1}^{\infty}$ được gọi là hội tụ theo phân phối về biến ngẫu nhiên X khi $n \to \infty$ nếu dãy các hàm phân phối xác suất $\{F_{X_n}(x)\}_{n=1}^{\infty}$ hội tụ về hàm phân phối $F_X(x)$ khi $n \to \infty$. Tức là

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x). \qquad (32)$$

3.4. LUẬT SỐ LỚN

Bất đẳng thức Chebyshev

Định lý 15 (Bất đẳng thức Markov)

Cho Y là biến ngẫu nhiên không âm có kỳ vọng hữu hạn. Khi đó, với $\varepsilon > 0$ tùy ý cho trước, ta có

$$P(Y \geq \epsilon) \leq \frac{E(Y^2)}{\epsilon^2}. \tag{33}$$

Định lý 16 (Bất đẳng thức Chebyshev)

Cho X là biến ngẫu nhiên có kỳ vọng $E(X) = \mu$ và phương sai $V(X) = \sigma^2$ hữu hạn. Khi đó, với $\varepsilon > 0$ tùy ý cho trước, ta có

$$P(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2} \tag{34}$$

hay tương đương

$$P(|X - \mu| < \varepsilon) \geq 1 - \frac{\sigma^2}{\varepsilon^2}. \tag{35}$$

Bất đẳng thức Chebyshev

✍

1. Nếu đặt $\varepsilon = k\sigma$ với k là một số thực dương tùy ý thì bất đẳng thức (35) có dạng

$$P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2}. \tag{36}$$

2. Bất đẳng thức Chebyshev đưa ra một ước lượng quan trọng về xác suất mà một biến ngẫu nhiên nhận giá trị sai lệch so với giá trị trung bình không quá k lần độ lệch chuẩn, với bất kỳ số thực dương k nào.
   Chẳng hạn, $k = 2$, Định lý 16 nói rằng xác suất để biến ngẫu nhiên X nằm trong khoảng $(\mu - 2\sigma; \mu + 2\sigma)$ là không nhỏ hơn $3/4$. Tương tự, định lý cũng nói rằng ít nhất 8/9 các quan sát của bất kỳ phân phối nào nằm trong khoảng $(\mu - 3\sigma; \mu + 3\sigma)$.

Bất đẳng thức Chebyshev

Ví dụ 24

Cho biến ngẫu nhiên X có kỳ vọng $\mu = 8$, phương sai $\sigma^2 = 9$ và không biết phân phối xác suất. Đánh giá các xác suất sau:

(a) $P(-4 < X < 20)$.

(b) $P(|X-8| \geq 6)$.

Giải. Áp dụng bất đẳng thức Chebysshev (36),

(a) $P(-4 < X < 20) = P(8-(4)(3) < X < 8+(4)(3)) \geq \frac{15}{16}$.

(b) $P(|X-8| \geq 6) = 1 - P(|X-8| < 6) = 1 - P(8-(2)(3) < X < 8+(2)(3)) \geq \frac{1}{4}$.

Bất đẳng thức Chebyshev

✍

- Bất đẳng thức Chebyshev phù hợp với bất kỳ phân phối nào của các quan sát và vì lý do này, kết quả thường yếu.
- Giá trị được đưa ra bởi bất đẳng thức chỉ là một giới hạn dưới. Chẳng hạn, ta biết rằng xác suất để một biến ngẫu nhiên X nhận giá trị trong khoảng $(\mu - 2\sigma; \mu + 2\sigma)$ là không nhỏ hơn $3/4$, nhưng không biết giá trị thực sự của nó có thể lớn hơn bao nhiêu.

Luật số lớn Chebyshev

Định lý 17

Nếu dãy các biến ngẫu nhiên $X_1, X_2, \ldots, X_n, \ldots$ độc lập, có kỳ vọng hữu hạn và phương sai bị chặn đều (nghĩa là, $V(X_i) \leq C$, $\forall i = 1, 2, \ldots$, C là hằng số dương), thì với $\varepsilon > 0$ tùy ý cho trước ta có:

$$\lim_{n \to +\infty} P\left(\left| \frac{1}{n} \sum_{i=1}^{n} X_i - \frac{1}{n} \sum_{i=1}^{n} E(X_i) \right| < \epsilon \right) = 1. \tag{37}$$

Hệ quả 3

Nếu dãy các biến ngẫu nhiên $X_1, X_2, \ldots, X_n, \ldots$ độc lập, có cùng kỳ vọng hữu hạn ($E(X_i) = \mu$, $i = 1, 2, \ldots$) và phương sai bị chặn đều, thì với $\varepsilon > 0$ tùy ý cho trước ta có:

$$\lim_{n \to +\infty} P\left(\left| \frac{1}{n} \sum_{i=1}^{n} X_i - \mu \right| < \varepsilon \right) = 1. \tag{38}$$

Luật số lớn Chebyshev

✍

- Định lý Chebyshev chứng tỏ rằng trung bình số học của các biến ngẫu nhiên độc lập hội tụ theo xác suất về trung bình số học của kỳ vọng tương ứng của nó.
- Như vậy, mặc dù từng biến ngẫu nhiên độc lập có thể nhận giá trị khác nhiều so với kỳ vọng của chúng, song trung bình số học của một số lớn các biến ngẫu nhiên lại nhận giá trị gần bằng trung bình số học các kỳ vọng của chúng. Điều đó cho phép dự đoán giá trị trung bình số học của các biến ngẫu nhiên.
- Chẳng hạn, gieo một con xúc xắc cân đối đồng chất. Gọi X là số chấm xuất hiện ở mặt trên con xúc xắc thì $E(X) = 3,5$. Một nhà thống kê đã gieo một con xúc xắc cân đối đồng chất một triệu lần và ghi lại số chấm xuất hiện ở mặt trên con xúc xắc. Số chấm trung bình của một triệu lần gieo được tìm thấy là $\dfrac{x_1 + x_2 + \cdots + x_{10^6}}{10^6} \approx 3,500867.$

3.4. LUẬT SỐ LỚN

Luật số lớn Bernoulli

✍ Áp dụng luật số lớn Chebyshev với trường hợp $X_i \sim \mathcal{B}(p)$, với X_i là biến ngẫu nhiên chỉ số lần xuất hiện A ở phép thử thứ i, $i = 1, 2, \ldots, n$, ta có luật số lớn Bernoulli.

Định lý 18

Giả sử ta có n phép thử Bernoulli với $P(A) = p$ và x là số lần xuất hiện A trong n phép thử đó. Khi đó, với $\varepsilon > 0$ tùy ý cho trước, ta có

$$\lim_{n\to+\infty} P\left(\left|\frac{x}{n} - p\right| < \varepsilon\right) = 1. \tag{39}$$

✍ Định lý Bernoulli chỉ ra rằng tần suất xuất hiện của sự kiện A trong n phép thử độc lập sẽ hội tụ theo xác suất về xác suất của sự kiện đó khi số lần thử tăng lên vô hạn. Chính vì vậy định lý Bernoulli là cơ sở lý thuyết của định nghĩa thống kê về xác suất ở Chương 1 "khi $n \to +\infty$ thì $\frac{x}{n} \to p$".

Luật số lớn Bernoulli

SAMI
1956

Ví dụ 25

- Giả sử p là tỷ lệ cử tri sẽ bầu cho ứng cử viên A. Để ước lượng trước tỷ lệ này người ta phỏng vấn ngẫu nhiên n cử tri. Có thể coi kết quả bầu của cử tri thứ i là biến ngẫu nhiên X_i có phân phối Bernoulli tham số p.
- Ta có $E(X_i) = p$ và phương sai $V(X_i) = p(1-p)$. Đặt $K_n = \frac{X_1+X_2+\cdots+X_n}{n}$, áp dụng bất đẳng thức Chebyshev ta có

$$P(|K_n - p| \geq \varepsilon) \leq \frac{p(1-p)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2} \quad \text{hay} \quad P(|K_n - p| < \varepsilon) \geq 1 - \frac{1}{4n\varepsilon^2}.$$

- Chẳng hạn, nếu $\varepsilon = 0,1$ và $n = 100$ thì $P(|K_{100} - p| \geq 0,1) \leq \dfrac{1}{4 \times 100 \times (0,1)^2} = 0,25$. Nói cách khác, nếu phỏng vấn 100 cử tri và lấy kết quả này để ước lượng cho tỷ lệ cử tri sẽ bỏ phiếu cho ứng cử viên A thì sai số vượt quá 0,1 có xác suất không vượt quá 0,25.

Luật số lớn Bernoulli

Ví dụ 25 (tiếp theo)

- Nếu muốn ước lượng tin cậy hơn (chẳng hạn xác suất lớn hơn 95%) và chính xác hơn (với sai số 0,01) thì $P(|K_n - p| < 0,01) \geq 1 - \dfrac{1}{4n(0,01)^2}$.
- Vậy số cử tri phải phỏng vấn thỏa mãn

$$1 - \frac{1}{4n(0,01)^2} \geq 0,95 \quad \text{suy ra} \quad n \geq 50000.$$

- Như vậy để ước lượng với độ tin cậy cao và độ chính xác cao thì cần phải lấy mẫu với số lượng lớn. Tuy nhiên, ở đây ta chỉ dựa vào bất đẳng thức Chebyshev để giải quyết bài toán.
- Trong phần thống kê ta sẽ nghiên cứu về bài toán ước lượng này và bằng phương pháp khác ta sẽ chỉ ra số cử tri phải phỏng vấn nhỏ hơn kết quả trên.