Lớp: Xác suất & thống kê thầy Việt Anh

Nguyễn Mạnh Linh

Ngày 10 tháng 5 năm 2020

1 Một số phân bố rời rạc quan trọng

1.1 Phân bố Bernoulli

Ta nói X là một biến ngẫu nhiên Bernoulli nếu nó chỉ nhận hai giá trị 0 hoặc 1.
Giả sử X là một biễn ngẫu nhiên Bernoulli. Đặt $p = \mathbb{P}(X = 1)$. Ta nói X có *phân bố Bernoulli* với tham số p và viết $X \sim \mathrm{Ber}(p)$. $p \in [0, 1]$.
Tính chất:

1. $\mathbb{E}X = p$.
2. $\operatorname{Var} X = p(1 - p)$.

1.2 Phân bố nhị thức

Ví dụ 1. Tung một đồng xu n lần liên tiếp. Giả sử đồng xu này có xác suất xuất hiện mặt sấp bằng p. Gọi X là số lần suất hiện mặt sấp trong n lần tung. X là một biến ngẫu nhiên, nhận các giá trị $0, 1, \ldots, n$.
Với $i = 1, ..., n$, gọi X_i là biến ngẫu nhiên sao cho:

- $X_i = 1$ nếu lần tung xu thứ i ra mặt sấp.
- $X_i = 0$ nếu lần tung xu thử i ra mặt ngửa.

Khi đó $X = X_1 + \cdots + X_n$.
Nhận xét: $X_i \sim \mathrm{Ber}(p)$, và $X_1, \ldots, X_n$ độc lập.
Ta nói X có *phân bố nhị thức* với tham số (n, p) và ký hiệu $X \sim \mathrm{Bin}(n, p)$.
Ta tính phân bố của X. Xét $k \in \{0, 1, \ldots, n\}$, ta có

$$\mathbb{P}(X = k) = C_n^k p^k (1 - p)^{n-k}.$$

Tính chất:

1. $\mathbb{E}X = np$.
2. $\operatorname{Var} X = np(1 - p)$.

Chú ý. Nếu X, Y độc lập thì $\operatorname{Var}(X + Y) = \operatorname{Var} X + \operatorname{Var} Y$.

Ví dụ 2. Một học sinh phải làm một bài kiểm tra trắc nghiệm gồm 10 câu hỏi. Mỗi câu có 4 phương án trả lời và chỉ có một phương án đúng. Học sinh đó sẽ đỗ nếu trả lời lời đúng ít nhất 6 câu.

1. Nếu học sinh đó trả lời ngẫu nhiên cả 10 câu hỏi thì xác suất học sinh đó đỗ là bao nhiêu?

2. Nếu học sinh đó chắc chắn trả lời đúng 2 câu thì xác suất học sinh đó đỗ là bao nhiêu?

Lời giải. 1. Gọi X là số câu trả lời đúng của học sinh đó. Ta có $n = 10$, $p = 0.25$ và $X \sim \mathrm{Bin}(10, 0.25)$. Xác suất học sinh đó đỗ là

$$\begin{aligned}\mathbb{P}(X \geq 6) &= \mathbb{P}(X = 6) + \mathbb{P}(X = 7) + \mathbb{P}(X = 8) + \mathbb{P}(X = 9) + \mathbb{P}(X = 10) \\ &= C_{10}^{6} \cdot 0.25^{6} \cdot 0.75^{4} + C_{10}^{7} \cdot 0.25^{7} \cdot 0.75^{3} + C_{10}^{8} \cdot 0.25^{8} \cdot 0.75^{2} \\ &\quad + C_{10}^{9} \cdot 0.25^{9} \cdot 0.75^{1} + C_{10}^{10} \cdot 0.25^{10} \cdot 0.75^{0} \\ &= 0.0197 \approx 1.97\%.\end{aligned}$$

2. Gọi Y là số câu trả lời đúng trong 8 câu còn lại. Khi đó $Y \sim \mathrm{Bin}(8, 0.25)$. Xác suất học sinh đó đỗ là

$$\begin{aligned}\mathbb{P}(Y \geq 4) &= \mathbb{P}(Y = 4) + \mathbb{P}(Y = 5) + \mathbb{P}(Y = 6) + \mathbb{P}(Y = 7) + \mathbb{P}(Y = 8) \\ &= C_{8}^{4} \cdot 0.25^{4} \cdot 0.75^{4} + C_{8}^{5} \cdot 0.25^{5} \cdot 0.75^{3} + \cdots \\ &= 0.8862 \approx 88.62\%.\end{aligned}$$

1.3 Phân bố hình học

Tung một đồng xu liên tiếp cho đến khi nào đồng xu xuất hiện mặt sấp. Biết xác suất xuất hiện mặt sấp ở mỗi lần tung là p. Gọi X là lần tung đầu tiên mà thu được mặt sấp. Khi đó $X \in \mathbb{N}^*$.

$$\mathbb{P}(X = n) = (1-p)^{n-1}p.$$

Ta nói X có *phân bố hình học* với tham số p và viết $X \sim \mathrm{Geom}(p)$. Nhận xét (giả sử $p \in (0, 1)$):

1. Ta có

$$\mathbb{P}(X < \infty) = \sum_{n=1}^{\infty} \mathbb{P}(X = n) = p \sum_{n=1}^{\infty} (1-p)^{n-1} = p \cdot \frac{1}{1-(1-p)} = 1.$$

   Ý nghĩa: Một biến cố nếu có xác suất xảy ra dương thì nó hầu như chắc chắn sẽ xảy ra.

2. $\mathbb{E}X = 1/p$. Chứng minh: Ta có

$$\sum_{n \geq 0} (1-p)^n = \frac{1}{p}.$$

   Đạo hàm hai vế theo p ta thu được điều phải chứng minh.

3. $\operatorname{Var} X = \dfrac{p}{(1-p)^2}$. (Chứng minh: tương tự như trên, nhưng đạo hàm hai lần).

4. $\mathbb{P}(X = n+k | X \geq k) = \mathbb{P}(X = n)$ với mọi $n, k \in \mathbb{N}^*$ (Tính mất trí nhớ).

1.4 Phân bố Poisson

Ta nói biến ngẫu nhiên $X \in \mathbb{N}$ có *phân bố Poisson* với tham số $\lambda > 0$ nếu

$$\forall k \in \mathbb{N}, \qquad \mathbb{P}(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}.$$

Ta viết $X \sim \operatorname{Poi}(\lambda)$.

Ví dụ 3. Thống kê cho thấy một cái công ty giao hàng A nhận được trung bình λ cuộc gọi mỗi ngày. Gọi X là số cuộc gọi nhận được trong một ngày. Khi đó $X \in \mathbb{N}$ là một biến ngẫu nhiên.
Cho trước $n \in \mathbb{N}^*$. Ta chia một ngày thành n khoảng thời gian bằng nhau. Gọi X_i là số cuộc gọi nhận được trong khoảng thời gian thứ i ($i = 1, \ldots, n$). Khi đó

$$X = X_1 + \cdots + X_n.$$

Trung bình trong mỗi khoảng thời gian trên, công ti nhận được λ/n cuộc gọi. Khi $n \to \infty$ thì các khoảng thời gian rất nhỏ, nên mỗi khoảng chỉ có nhiều nhất một cuộc gọi, hay $X_i \in \{0, 1\}$ với $i = 1, \ldots, n$. Vậy X_i là một biến ngẫu nhiên Bernoulli. Mà $\mathbb{E}X_i = \lambda/n$ nên $X_i \approx \operatorname{Ber}(\lambda/n)$.
Vì thế $X = X_1 + \cdots + X_n \approx \operatorname{Bin}(n, \lambda/n)$. Do đó

$$\forall k \in \mathbb{N}, \qquad \mathbb{P}(X = k) = \lim_{n\to\infty} C_n^k \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = \lim_{n\to\infty} \frac{n!}{k!(n-k)!} \cdot \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k}.$$

Ta có

$$\mathbb{P}(X = k) = \lim_{n\to\infty} \frac{\lambda^k}{k!} \cdot \frac{n(n-1)\cdots(n-k+1)}{n^k} \cdot \left(1 - \frac{\lambda}{n}\right)^{n-k}.$$

Mà

$$\lim_{n\to\infty} \cdot \frac{n(n-1)\cdots(n-k+1)}{n^k} = \lim_{n\to\infty} = 1\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\cdots\left(1 - \frac{k-1}{n}\right) = 1,$$

$$\lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^n = \lim_{n\to\infty} \left(\left(1 - \frac{\lambda}{n}\right)^{-n/\lambda}\right)^{-\lambda} = e^{-\lambda},$$

$$\lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{-k} = 1,$$

nên

$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

Chú ý. Ta có

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{\lambda} \cdot e^{-\lambda} = 1.$$

Tính chất:

1. $\mathbb{E}X = \lambda$.
2. $\mathrm{Var}(X) = \lambda$.
3. Nếu $X \sim \mathrm{Poi}(\lambda)$, $Y \sim \mathrm{Poi}(\mu)$ và X, Y độc lập thì $X + Y \sim \mathrm{Poi}(\lambda + \mu)$.

Ví dụ 4. Một trạm bưu điện chuyển điện trong khoảng thời gian 10^{-5} giây. Trong quá trình truyền, có các tiếng ồn ngẫu nhiên. Số tiếng ồn ngẫu nhiên trung bình trong một giây là khoảng 10^4. Tính xác suất để trong khoảng thời gian truyền có nhiễu.

Lời giải. Gọi X là số tiếng ồn trong khoảng thời gian truyền tin 10^{-5} giây. Khi đó $X \sim \mathrm{Poi}(10^4 \cdot 10^{-5}) = \mathrm{Poi}(0.1)$. Xác suất có nhiễu trong thời gian truyền trên là

$$\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X = 0) = 1 - \frac{e^{-0.1} \cdot 0.1^0}{0!} = 1 - e^{-0.1} = 0.0952\ldots.$$

Ứng dụng khác của phân bố Poisson: Xấp xỉ phân bố nhị thức.

Công thức De Moivre - Laplace: Nếu $n > 30$, $np > 5$, $n(1-p) > 5$ và $X \sim \mathrm{Bin}(n, p)$ thì ta có thể coi $X \approx \mathrm{Poi}(np)$.

Ví dụ 5. Trong một kho lúa có khoảng 20% hạt lép.

1. Phải lấy ra bao nhiêu hạt lúa để xác suất có ít nhất một hạt lép $\geq 95\%$.
2. Lấy ngẫu nhiên 100 hạt lúa. Tính xác suất để có 25 hạt lép.

Lời giải.

1. Gọi n là số hạt cần lấy. Xác suất để không có hạt lép nào là 0.8^n. Ta cần có $1 - 0.8^n \geq 0.95 \Leftrightarrow 0.8^n \leq 0.05 \Leftrightarrow n \geq \log_{0.8}(0.05)$.
2. Gọi X là số hạt lép trong 100 hạt. Thế thì $X \sim \mathrm{Bin}(100, 0.2)$. Ta có

$$n = 100 > 30, \qquad np = 100 \cdot 0.2 > 5, \qquad n(1-p) = 100 \cdot 0.8 > 5,$$

vì thế ta có thể coi $X \approx \mathrm{Poi}(20)$. Từ đó

$$\mathbb{P}(X = 25) \approx \frac{e^{-20} 20^{25}}{25!} = 0.0446\ldots.$$

Chú ý. Nếu tính theo công thức

$$\mathbb{P}(X = 25) = C_{100}^{25} \cdot 0.2^{25} \cdot 0.8^{75}$$

sẽ khó hơn.

2 Một số phân bố liên tục quan trọng

2.1 Phân bố đều

2.2 Phân bố mũ

2.3 Phân bố chuẩn

2.4 Phân bố χ^2