XÁC SUẤT THỐNG KÊ
Chương 2: Biến ngẫu nhiên

Phạm Việt Nga

Bm Toán, Khoa CNTT, Học viện Nông nghiệp VN
01/2019

Nội dung

Nội dung

Một số ví dụ

Ví dụ 1

Từ một nhóm gồm 10 sinh viên (6 nam, 4 nữ) chọn ngẫu nhiên 2 bạn.

$\rightarrow$

Sk có thể có	2 nam	2 nữ	1 nam,1 nữ
Số nữ SV trong 2 bạn được chọn (X)	0	2	1

Ví dụ 2

Một người ném bóng vào rổ cho tới khi có bóng trúng rổ thì dừng. Các lần ném độc lập, xác suất ném trúng ở mỗi lần đều là 0,7.

Sk có thể có	lần 1 trúng	lần 2 mới trúng	lần 3 mới trúng	$\cdots$
Số bóng dùng (Y)	1	2	3	$\cdots$

Ví dụ 3

Xem cân nặng Z của 1 em bé lúc mới sinh.
$\rightarrow$ Z có thể nhận 1 giá trị thực lớn hơn 0.

Định nghĩa biến ngẫu nhiên

Định nghĩa

BNN là đại lượng nhận giá trị thực phụ thuộc kết quả của phép thử ngẫu nhiên.

- Ký hiệu BNN bởi $X, Y, Z, ..., X_1, X_2, ...$.
- Ký hiệu giá trị mà BNN có thể nhận bởi $x, y, ..., x_1, x_2, ...$
  $\rightarrow$ sk $(X = x_1), (X = x_2), ...$
- Phân loại: BNN rời rạc / BNN liên tục

Bảng phân phối xác suất của biến ngẫu nhiên rời rạc

Bảng ppxs của BNN rời rạc

Bảng gồm 2 dòng: dòng trên ghi các giá trị mà BNN có thể nhận, dòng dưới ghi các xác suất tương ứng

X	x_1	x_2	$\cdots$	x_k	$\cdots$
P	p_1	p_2	$\cdots$	p_k	$\cdots$

Chú ý: $p_k = P(X = x_k)$ và $\sum p_k = 1$

Ví dụ: Lập bảng ppxs cho BNN X, Y trong VD1, VD2 ở trên.

Hàm phân phối xác suất

Ví dụ: Xem xét điểm thi (X) môn Toán học kỳ 1 của học sinh khối 7 tại một trường THCS ở Hà Nội. Hỏi tỷ lệ học sinh có điểm thi hk môn Toán dưới 5 điểm, dưới 8 điểm?

Định nghĩa

Hàm ppxs của BNN X là hàm số $F : \mathbb{R} \to [0; 1]$ xác định bởi:

$$F(x) = P(X < x), \quad x \in \mathbb{R}$$

Ví dụ:

Một số tính chất của hàm ppxs:

1. $0 \leq F(x) \leq 1$
2. $\lim_{x \to -\infty} F(x) = 0, \ \lim_{x \to +\infty} F(x) = 1$
3. $\forall a, b \in \mathbb{R}, a < b$, ta có $P(a \leq X < b) = F(b) - F(a)$.
4. Hàm ppxs là hàm không giảm.
5. Hàm ppxs là hàm liên tục trái.

Hàm mật độ xác suất

Định nghĩa

Nếu hàm ppxs của BNN X có đạo hàm tại mọi $x \in \mathbb{R}$ thì hàm đạo hàm F' được gọi là hàm mật độ xác suất của X, ký hiệu f.

$$f(x) = F'(x)$$

Chú ý: BNN rời rạc không có hàm mật độ xác suất.

Một số tính chất của hàm mật độ xác suất:

1. $f(x) \geq 0, \forall x \in \mathbb{R}$.
2. $F(x) = \int\limits_{-\infty}^{x} f(t)dt$.
3. $P(a \leq X < b) = \int\limits_{a}^{b} f(t)dt$
4. Nếu BNN X có hàm mật độ xác suất f thì $P(X = a) = 0$, $\forall a \in \mathbb{R}$.
5. $\int_{-\infty}^{+\infty} f(t)dt = 1$

Nội dung

Kỳ vọng

Định nghĩa

Kỳ vọng (hay trung bình lý thuyết) của BNN X, ký hiệu EX hoặc MX là số thực xác định bởi:

1. Nếu X là BNN rời rạc nhận các giá trị $x_1, x_2, \cdots$ với $P(X = x_k) = p_k$ thì
$$EX = \sum x_k p_k$$
Nếu X nhận vô hạn giá trị thì cần điều kiện $\sum |x_k| p_k$ hội tụ.
2. Nếu X là BNN liên tục có hàm mật độ xác suất f thì
$$EX = \int_{-\infty}^{+\infty} x f(x) dx$$
với điều kiện $\int_{-\infty}^{+\infty} |x| f(x) dx$ hội tụ.

Ví dụ:

Kỳ vọng

Một số tính chất của kỳ vọng:

1. $EC = C$, với C là hằng số.
2. $E(aX) = a.EX$, với a là số thực.
3. $E(X + Y) = EX + EY$.
4. $E(X - Y) = EX - EY$.
5. Nếu X, Y là hai BNN độc lập thì $E(XY) = (EX).(EY)$.

Phương sai

Định nghĩa

Nếu BNN X có kỳ vọng EX thì phương sai của X là số thực DX xác định bởi

$$DX = E[X - EX]^2$$

Suy ra $DX = E(X^2) - [EX]^2$

Ví dụ:

Một số tính chất của phương sai:

1. $DX \geq 0, \forall X$.
   $DC = 0$, với C là hằng số.
2. $D(aX) = a^2.DX$, với a là số thực.
3. Nếu X, Y là hai BNN độc lập thì $D(X + Y) = DX + DY$;
   $D(X - Y) = DX + DY$.

Độ lệch chuẩn

Độ lệch chuẩn của BNN X ký hiệu σX, là số thực xác định bởi

$$\sigma X = \sqrt{DX}$$

Mod

1. Nếu X là BNN rời rạc nhận các giá trị $x_1, x_2, \cdots$ thì $Mod(X)$ là số thực x_k sao cho $P(X = x_k)$ là xác suất lớn nhất.
2. Nếu X là BNN liên tục có hàm mật độ xác suất f thì $Mod(X)$ là điểm mà tại đó hàm f đạt giá trị lớn nhất.

Ví dụ:

Nội dung

Phân phối nhị thức

Ví dụ: Gieo 5 hạt đậu, xác suất nảy mầm của mỗi hạt đậu là 0,8. Gọi X là số hạt đậu nảy mầm trong 5 hạt được gieo. Lập bảng phân phối xác suất của X.

Định nghĩa

BNN rời rạc X được gọi là tuân theo quy luật phân phối nhị thức với hai tham số là n và p ($n \in \mathbb{N}$ và $0 < p < 1$) nếu X có bảng ppxs

X	0	1	$\cdots$	k	$\cdots$	n
P	$C_n^0 p^0 q^n$	$C_n^1 p^1 q^{n-1}$	$\cdots$	$C_n^k p^k q^{n-k}$	$\cdots$	$C_n^n p^n q^0$

với $q = 1 - p$.
Ký hiệu: $X \sim B(n; p)$

Chú ý

Nếu $X \sim B(n; p)$ thì $EX = np$ và $DX = npq$.

Phân phối Poisson

Định nghĩa

BNN rời rạc X được gọi là có phân phối Poisson với tham số thực $\lambda > 0$ nếu X có bảng ppxs

X	0	1	$\cdots$	k	$\cdots$	n	$\cdots$
P	$e^{-\lambda}$	$e^{-\lambda}\dfrac{\lambda^1}{1!}$	$\cdots$	$e^{-\lambda}\dfrac{\lambda^k}{k!}$	$\cdots$	$e^{-\lambda}\dfrac{\lambda^n}{n!}$	$\cdots$

Ký hiệu: $X \sim P_\lambda$.

Chú ý

Nếu $X \sim P_\lambda$ thì $EX = \lambda$ và $DX = \lambda$.

Phân phối chuẩn

Định nghĩa

BNN liên tục X được gọi là có quy luật phân phối chuẩn với hai tham số μ và σ^2 $(\sigma > 0)$ nếu X có hàm mật độ xác suất

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Ký hiệu $X \sim N(\mu, \sigma^2)$

Chú ý

Nếu $X \sim N(\mu, \sigma^2)$ thì $EX = \mu$ và $DX = \sigma^2$.

Hàm mật độ và hàm phân phối của phân phối chuẩn

PP chuẩn $N(\mu, \sigma^2)$

Hàm mật độ xác suất: $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

Hàm phân phối xác suất: $F(x) = \int\limits_{-\infty}^{x} f(t)dt$

PP chuẩn tắc $N(0, 1)$ ($\mu = 0$ và $\sigma = 1$)

Hàm mật độ xác suất: $\varphi(x) = \dfrac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$ (hàm Gauss)

Hàm phân phối xác suất: $\Phi(x) = \int\limits_{-\infty}^{x} \phi(t)dt$ (hàm Laplace)

Một số tính chất của hàm ppxs và hàm mđxs của biến chuẩn

1. $\varphi(-x) = \varphi(x), \quad \forall x \in (0; +\infty)$
2. $\varphi(x) \approx 0, \quad \forall x \geq 4$
3. $\Phi(-x) = 1 - \Phi(x), \quad \forall x$
4. $\Phi(x) \approx 1, \quad \forall x \geq 4$

Áp dụng tính xác suất với pp chuẩn $N(\mu, \sigma^2)$

Gỉa sử $X \sim N(\mu, \sigma^2)$ và $a, b \in \mathbb{R}, a < b$,

$$P(a \leq X \leq b) = \Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})$$

Ví dụ: G/s $X \sim N(3, 4)$. Tính $P(2 \leq X \leq 4,5)$ và $P(X \leq 4)$.

Xấp xỉ phân phối nhị thức bởi pp chuẩn

Cho $X \sim B(n;p)$ với n lớn ($n > 30$), p không gần 0, không gần 1
$\Rightarrow$ xấp xỉ $X \sim N(\mu, \sigma^2)$ với $\mu = np, \sigma^2 = npq$.
Khi đó với $k, k_1, k_2 \in \mathbb{N}^*$,

$$P(X = k) \approx \frac{1}{\sigma}\varphi(\frac{k-\mu}{\sigma}); \quad P(k_1 \leq X \leq k_2) \approx \Phi(\frac{k_2-\mu}{\sigma}) - \Phi(\frac{k_1-\mu}{\sigma})$$

Xấp xỉ phân phối nhị thức bởi pp Poisson

Xét 1 dãy gồm vô hạn phép thử độc lập liên quan đến sk A, có xác suất xuất hiện sk A ở phép thử thứ n là p_n thoả mãn $\lim\limits_{n\to\infty} np_n = \lambda$
$\Rightarrow$ xấp xỉ $X \sim P_\lambda$.
Áp dụng: Nếu $X \sim B(n;p)$ với n rất lớn, p rất nhỏ gần 0 thì có thể xấp xỉ $X \sim P_\lambda$ với $\lambda = np$. Khi đó, với $k \in \mathbb{N}$,

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