Xác suất

Chương 2: Biến ngẫu nhiên liên tục

Phan Thanh Hồng

Ngày 21 tháng 9 năm 2022

Thang long University

Nội dung

1. Biến ngẫu nhiên liên tục và hàm mật độ xác suất
2. Hàm phân phối tích luỹ
3. Một số biến ngẫu nhiên liên tục
4. Kỳ vọng và phương sai

1. Biến ngẫu nhiên liên tục và hàm mật độ xác suất

Ví dụ: Đường dây điện thoại ngầm nối một tổng đài với một trạm dài 1 km bị đứt. Gọi X là khoảng cách từ tổng đài tới vị trí dây đứt. Ta có thể chắc chắn về giá trị của X không? X có thể nhận những giá trị nào?

Ví dụ: Đường dây điện thoại ngầm nối một tổng đài với một trạm dài 1 km bị đứt. Gọi X là khoảng cách từ tổng đài tới vị trí dây đứt. Ta có thể chắc chắn về giá trị của X không? X có thể nhận những giá trị nào?

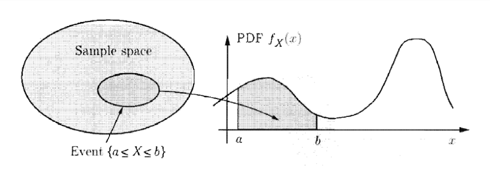
Định nghĩa 1.1

Biến ngẫu nhiên có tập giá trị không đếm được (tập giá trị của nó có thể lấp đầy một khoảng trên trục số) gọi là biến ngẫu nhiên liên tục (continuous r.v.).

Hàm mật độ xác suất của biến ngẫu nhiên liên tục

Định nghĩa 1.2

Hàm $f(x)$ là hàm mật độ xác suất (hàm PDF-probability density function) của X, nếu xác suất để X nhận giá trị trong khoảng $[a, b]$ là diện tích của hình giới hạn bởi đường mật độ đó, trục hoành và hai đường thẳng vuông góc với trục hoành tại hai điểm a, b.



Hàm mật độ xác suất của biến ngẫu nhiên liên tục

1. Xác suất để X nhận giá trị trong khoảng $[a, b]$ là

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

2. Xác suất để X nhận một giá trị x nào đó bằng 0, tức là $P(X = x) = 0$ do đó (giải thích?)

$$P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = P(a < X < b)$$

3. $P(a \leq X \leq a + \epsilon) = \int_a^{a+\epsilon} f_X(x)dx \approx f_X(a).\epsilon$ or

$$\frac{P(a \leq X \leq a + \epsilon)}{\epsilon} = f_X(a)$$

Chú ý: hàm mật độ không cho giá trị xác suất, ko cần phải nhận giá trị bé hơn 1.

Hàm mật độ xác suất của biến ngẫu nhiên liên tục

Tính chất của hàm mật độ

1. $f_X(x) \geq 0$ với mọi x
2. $\int_{-\infty}^{\infty} f_X(x)dx = P(-\infty < X < \infty) = 1$

Hàm mật độ xác suất của biến ngẫu nhiên liên tục

Tính chất của hàm mật độ

1. $f_X(x) \geq 0$ với mọi x
2. $\int_{-\infty}^{\infty} f_X(x)dx = P(-\infty < X < \infty) = 1$

Ví dụ: Đường dây điện thoại ngầm nối một tổng đài với một trạm dài 1 km bị đứt. Gọi X là b.n.n. chỉ khoảng cách từ tổng đài tới vị trí dây đứt. Giả sử xác suất X rơi vào các khoảng trong đoạn $[0, 1]$ có cùng độ dài là như nhau. Có thể mô tả hàm mật độ của X có dạng

$$f_X(x) = \begin{cases} c, & \text{if } 0 \leq x \leq 1 \\ 0, & \text{khác} \end{cases}$$

Ta có thể thấy $c = 1$.
Xác suất điểm bị đứt cách tổng đài 0.5 km là $\int_0^{1/2} 1.dx = 0.5$.

Hàm mật độ xác suất của biến ngẫu nhiên liên tục

Ví dụ: : xét trường hợp vị trí dây đứt nằm giữa hai điểm a, b với $a < b$ và nhỏ hơn khoảng cách giữa tổng đài và trạm.

Ví dụ: : xét trường hợp vị trí dây đứt nằm giữa hai điểm a, b với $a < b$ và nhỏ hơn khoảng cách giữa tổng đài và trạm.

Định nghĩa 1.3
Cho $a < b$, biến ngẫu nhiên X có hàm mật độ

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & \textit{if } a \leq x \leq b \\ 0, & \textit{khác} \end{cases}$$

được gọi là b.n.n. có phân phối đều (uniform r.v.) trên $[a, b]$. Ký hiệu $X \sim U(a, b)$

Ví dụ

Tuổi thọ một thiết bị (đơn vị: năm) là biến ngẫu nhiên X có mật độ

$$f(x) = \begin{cases} \frac{k}{x^3} & \text{for} \quad x \geq 1 \\ 0 & \text{for} \quad x < 1 \end{cases}$$

1. Tìm k.
2. Tìm hàm CDF $F_X(x)$.
3. Tính xác suất tuổi thọ của thiết bị trên 5 năm

Hàm phân phối tích luỹ

Nhắc lại: Cho X là một biến ngẫu nhiên (liên tục, rời rạc), hàm phân phối tích luỹ của X, CDF- cumulative distribution function, kí hiệu $F_X(x)$ được định nghĩa là

$$F_X(x) = P(X \leq x), \quad x \in R$$

Như vậy nếu X là bnn liên tục

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt$$

Ví dụ: Viết, vẽ đồ thị cho CDF của biến ngẫu nhiên có hàm mật độ

$$f(x) = \begin{cases} 1 & \text{với} \quad 0 \leq x \leq 1 \\ 0 & \text{trong những trường hợp còn lại} \end{cases}$$

Tính chất của hàm phân phối tích luỹ

1. F_X đơn điệu tăng
2. $\lim_{x \to -\infty} F_X(x) = 0$, $\lim_{x \to \infty} F_X(x) = 1$
3. Nếu X liên tục thì

$$f_X(x) = F'_X(x)$$

2. Kỳ vọng và phương sai

Kỳ vọng và phương sai

Giá trị kỳ vọng (trung bình) của bnn liên tục X là

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx$$

Kỳ vọng và phương sai

Giá trị kỳ vọng (trung bình) của bnn liên tục X là

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx$$

Phương sai của X là

$$V(X) = E\left[(X - E(X))^2\right] = \int_{-\infty}^{\infty} x^2 f_X(x) dx - [E(X)]^2$$

Ví dụ: Cho bnn X có hàm mật độ

$$f_X(x) = \begin{cases} cx^2 & 0 \leq x \leq 9 \\ 0 & \text{otherwise} \end{cases}$$

Tìm $c, CDF, E(X), V(X)$.

Kỳ vọng và phương sai

Tương tự như đối với b.n.n rời, các tính chất sau cũng đúng cho b.n.n. liên tục

1. $E(aX + b) = aE(X) + b$
2. $V(X) = E(X^2) - (E(X))^2 \geq 0$
3. $V(aX + b) = a^2 V(X)$

Kỳ vọng và phương sai

Tương tự như đối với b.n.n rời, các tính chất sau cũng đúng cho b.n.n. liên tục

1. $E(aX + b) = aE(X) + b$
2. $V(X) = E(X^2) - (E(X))^2 \geq 0$
3. $V(aX + b) = a^2V(X)$

3. Một số biến ngẫu nhiên liên tục

Biến ngẫu nhiên phân phối đều (Uniform r.v)

Cho a, b là hai số thực với $a < b$, b.n.n. liên tục X có hàm mật độ dạng

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{với} \quad a \le x \le b \\ 0 & \text{trong những trường hợp còn lại} \end{cases}$$

thì X được gọi là có phân phối đều trên đoạn $[a, b]$, ký hiệu $X \sim U(a, b)$.

Biến ngẫu nhiên phân phối đều (Uniform r.v)

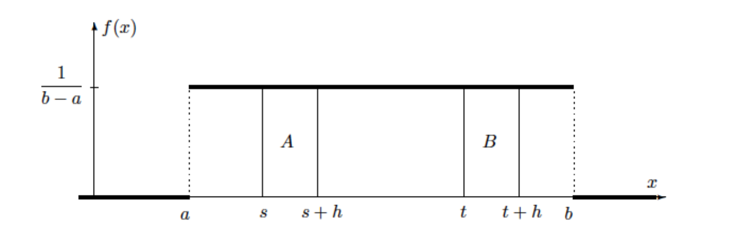
Cho a, b là hai số thực với $a < b$, b.n.n. liên tục X có hàm mật độ dạng

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{với} \quad a \le x \le b \\ 0 & \text{trong những trường hợp còn lại} \end{cases}$$

thì X được gọi là có phân phối đều trên đoạn $[a, b]$, ký hiệu $X \sim U(a, b)$.

Câu hỏi: Tính kì vọng và phương sai của $X \sim U(a, b)$.

Biến ngẫu nhiên phân phối đều (Uniform r.v)

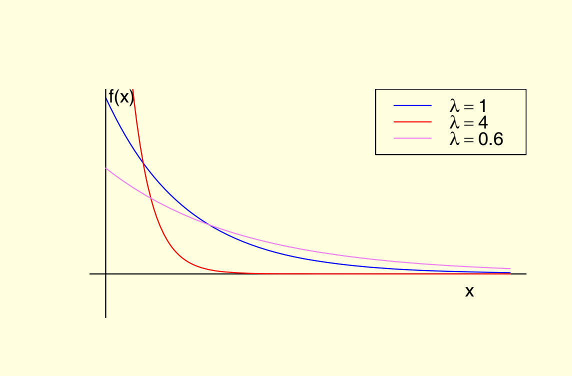


Biến ngẫu nhiên phân phối mũ (Exponential r.v)

Cho λ là một số dương, bnn X có hàm mật độ dạng

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{với} \quad x \geq 0 \\ 0 & \text{trong những trường hợp còn lại} \end{cases}$$

được gọi là bnn có phân phối mũ, ký hiệu $X \sim E(\lambda)$.



BNN phân phối mũ

Tính chất

1. X có trung bình và độ lệch chuẩn (chỉ rõ)

$$E(X) = \frac{1}{\lambda}, \quad \sigma_X = \frac{1}{\lambda}$$

BNN phân phối mũ

Tính chất

1. X có trung bình và độ lệch chuẩn (chỉ rõ)

$$E(X) = \frac{1}{\lambda}, \quad \sigma_X = \frac{1}{\lambda}$$

2. $P(X > a) = e^{-\lambda a}$

BNN phân phối mũ

Tính chất

1. X có trung bình và độ lệch chuẩn (chỉ rõ)

$$E(X) = \frac{1}{\lambda}, \quad \sigma_X = \frac{1}{\lambda}$$

2. $P(X > a) = e^{-\lambda a}$

3. Liên hệ với phân phối Poisson: Giả sử số lần xảy ra 1 sự kiện trong 1 đơn vị thời gian tuân theo phân phối Poisson với tham số λ (tức là tính trung bình có λ lần xảy ra sự kiện trên mỗi đơn vị thời gian) thì khoảng thời gian giữa hai sự kiện liên tiếp tuân theo phân phối mũ với tham số λ.

4. Tính chất không có trí nhớ: $X \sim E(\lambda)$,

$$\boldsymbol{P}\{T > t + x \mid T > t\} = \boldsymbol{P}\{T > x\} \quad \text{for } t, x > 0$$

(Google: Memoryless property of exponential distribution)

BNN phân phối mũ

Ví dụ: Giả sử số lệnh in gửi đến một máy in tuân theo phân phối Poisson với trung bình 3 lệnh in mỗi giờ.

1. Thời gian trung bình giữa hai lệnh in liên tiếp là bao nhiêu?
2. Máy vừa nhận 1 lệnh in, xác suất để lệnh in kế tiếp có trong vòng 5 phút tới là bao nhiêu?

BNN phân phối chuẩn (Normal r.v.)

Định nghĩa 3.1

Bnn liên tục X được gọi là có phân phối chuẩn nếu hàm mật độ có dạng

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

Trong đó μ và σ là hai tham số, ký hiệu $X \sim N(\mu, \sigma^2)$.

BNN phân phối chuẩn (Normal r.v.)

Định nghĩa 3.1
Bnn liên tục X được gọi là có phân phối chuẩn nếu hàm mật độ có dạng

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

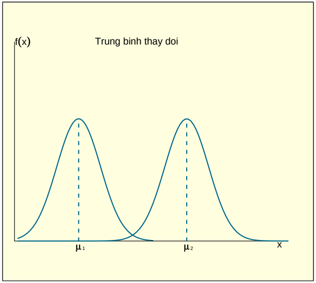
Trong đó μ và σ là hai tham số, ký hiệu $X \sim N(\mu, \sigma^2)$.

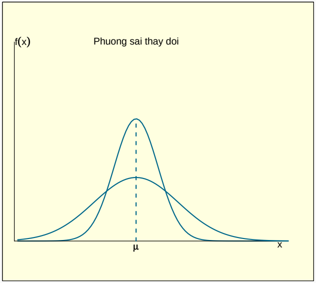
Nhận xét
1. X có trung bình và độ lệch chuẩn (chỉ rõ)

$$E(X) = \mu, \quad \sigma_X = \sigma$$

2. Nếu X phân phối chuẩn thì $Y = aX + b$ cũng có phân phối chuẩn. Câu hỏi: Tìm kì vọng và độ lệch chuẩn của Y.

BNN phân phối chuẩn





BNN phân phối chuẩn tắc

$Z \sim N(0,1)$ gọi là bnn chuẩn tắc. ? Hàm mật độ của bnn chuẩn tắc Nếu $X \sim N(\mu, \sigma^2)$ thì $\frac{X-\mu}{\sigma} \sim N(0,1)$

Đọc thêm: Bảng tra phân phối chuẩn.

Ví dụ

Một tín hiệu nhị phân có thể nhận hai giá trị $s = 1, -1$ được truyền đi, tuy nhiên trong quá trình truyền tin, tín hiệu bị nhiễu với nhiễu cộng tính chuẩn có trung bình 0 và phương sai σ^2. Thiết bị nhận tín hiệu nhận giá trị âm sẽ kết luận tín hiệu truyền đi là -1, ngược lại nhận giá trị dương thì kết luận tín hiệu được truyền đi là 1. Xác suất kết luận nhầm tr0ng mỗi trường hợp của s là bao nhiêu?

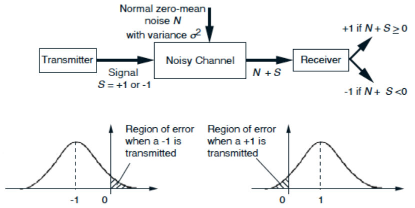


Figure 3.11: The signal detection scheme of Example 3.9. The area of the shaded region gives the probability of error in the two cases where -1 and $+1$ is transmitted.

4. Định lý giới hạn trung tâm

Định lý giới hạn trung tâm

Cho $X_1, X_2, ...$ là dãy bnn độc lập cùng phân phối với trung bình μ và phương sai σ^2. Đặt

$$Z_n := \frac{\frac{X_1+X_2+\cdots X_n}{n} - \mu}{\sigma/\sqrt{n}} = \frac{X_1 + X_2 + \cdots X_n - n\mu}{\sigma\sqrt{n}}$$

Bài tập: tính $E(Z_n), V(Z_n)$

Định lí 4.1
Hàm pp tích lũy CDF của Z_n, $F_n(z) = P(Z_n \leq z)$ hội tụ tới hàm pp tích luỹ của pp chuẩn tắc. Tức là với mỗi z

$$\lim_{n\to\infty} F_n(z) = \int_{-\infty}^{z} e^{-x^2/2} dx =: \Phi(z)$$

Định lý giới hạn trung tâm

Cho $X_1, X_2, ...$ là dãy bnn độc lập cùng phân phối với trung bình μ và phương sai σ^2. Đặt

$$Z_n := \frac{\frac{X_1+X_2+\cdots X_n}{n} - \mu}{\sigma/\sqrt{n}} = \frac{X_1 + X_2 + \cdots X_n - n\mu}{\sigma\sqrt{n}}$$

Bài tập: tính $E(Z_n), V(Z_n)$

Định lí 4.1

Hàm pp tích lũy CDF của Z_n, $F_n(z) = P(Z_n \leq z)$ hội tụ tới hàm pp tích luỹ của pp chuẩn tắc. Tức là với mỗi z

$$\lim_{n\to\infty} F_n(z) = \int_{-\infty}^{z} e^{-x^2/2} dx =: \Phi(z)$$

Nhận xét: Mặc dù không biết phân phối của các Z_n ta vẫn có thể tính được xấp xỉ $F_n(z)$ (với n đủ lớn)

Định lý giới hạn trung tâm

Ví dụ: Một kiện hàng trên máy bay gồm 100 gói hàng nhỏ được giả định có trọng lượng (đơn vị kg) từng gói là những biến ngẫu nhiên độc lập cùng phân phối đều trên đoạn $[5, 50]$. Tính xác suất trọng lượng toàn bộ kiện hàng vượt quá 3000 kg.

Ví dụ: Một kiện hàng trên máy bay gồm 100 gói hàng nhỏ được giả định có trọng lượng (đơn vị kg) từng gói là những biến ngẫu nhiên độc lập cùng phân phối đều trên đoạn $[5, 50]$. Tính xác suất trọng lượng toàn bộ kiện hàng vượt quá 3000 kg.

Lời giải: Ta cần tính $P(S_{100} \geq 3000)$

Định lý giới hạn trung tâm

Ví dụ: Một chiếc đĩa còn 330 MB dung lượng trống. Liệu có khả năng cao có thể ghi vào đĩa 300 bức ảnh độc lập nếu dung lượng một bức ảnh là biến ngẫu nhiên có kỳ vọng 1 MB và độ lệch chuẩn 0.5MB?

Ví dụ: Một chiếc đĩa còn 330 MB dung lượng trống. Liệu có khả năng cao có thể ghi vào đĩa 300 bức ảnh độc lập nếu dung lượng một bức ảnh là biến ngẫu nhiên có kỳ vọng 1 MB và độ lệch chuẩn 0.5MB?

Lời giải: Ta cần tính $P(S_{300} \geq 330)$

5. Ứng dụng

Thuật toán đếm số phần tử phân biệt

Đọc: Section 9.5, trang 332, tài liệu
Probability & Statistics with Applications to Computing.

Tổng kết chương

1. Biến ngẫu nhiên liên tục và hàm mật độ xác suất
2. Một số biến ngẫu nhiên
3. Hàm phân phối tích luỹ
4. Kì vọng và phương sai