XÁC SUẤT - THỐNG KÊ

CHƯƠNG 3: MỘT SỐ PHÂN PHỐI XÁC SUẤT THÔNG DỤNG

TS. Phan Thị Hường

Trường Đại học Bách Khoa TP HCM
Khoa Khoa học ứng dụng, bộ môn Toán ứng dụng
Email: huongphan@hcmut.edu.vn

TP. HCM — 2020.

NỘI DUNG

1. CÁC PHÂN PHỐI RỜI RẠC

Nội dung

1. Các phân phối rời rạc
2. Các phân phối liên tục

NỘI DUNG

1. CÁC PHÂN PHỐI RỜI RẠC
2. CÁC PHÂN PHỐI LIÊN TỤC
3. ĐỊNH LÝ GIỚI HẠN TRUNG TÂM

BIẾN NGẪU NHIÊN BERNOULLI

ĐỊNH NGHĨA 1.1 (BIẾN NGẪU NHIÊN BERNOULLI)

Thực hiện một phép thử, ta quan tâm đến biến cố A*. Nếu biến cố* A *xảy ra (thành công) thì* X *nhận giá trị là 1* $(X=1)$*, ngược lại biến ngẫu nhiên* X *nhận giá trị* 0*. Phép thử này gọi là phép thử Bernoulli. Giả sử xác suất xảy ra biến cố* A *là* p*,* $0<p<1$

$$\mathbb{P}(A)=\mathbb{P}(X=1)=p$$

và

$$\mathbb{P}\left(\bar{A}\right)=\mathbb{P}(X=0)=1-p=q$$

Khi đó biến ngẫu nhiên X *được gọi là biến ngẫu nhiên có phân phối Bernoulli với tham số* p*, ký hiệu* $X\sim B(p)$*.*

Phân phối Bernoulli

Ví dụ 1.1

Các phép thử sau đây cho kết quả là một biến ngẫu nhiên Bernoulli

- *Kiểm tra ngẫu nhiên một sản phẩm trong lô hàng:* $X = 1$ *nếu gặp được sản phẩm tốt,* $X = 0$ *nếu gặp được sản phẩm kém.*

PHÂN PHỐI BERNOULLI

VÍ DỤ 1.1

Các phép thử sau đây cho kết quả là một biến ngẫu nhiên Bernoulli

- *Kiểm tra ngẫu nhiên một sản phẩm trong lô hàng:* $X = 1$ *nếu gặp được sản phẩm tốt,* $X = 0$ *nếu gặp được sản phẩm kém.*
- *Trả lời ngẫu nhiên 1 câu trắc nghiệm:* $X = 1$ *nếu trả lời đúng,* $X = 0$ *nếu trả lời sai.*

Phân phối Bernoulli

Ví dụ 1.1

Các phép thử sau đây cho kết quả là một biến ngẫu nhiên Bernoulli

- *Kiểm tra ngẫu nhiên một sản phẩm trong lô hàng:* $X = 1$ *nếu gặp được sản phẩm tốt,* $X = 0$ *nếu gặp được sản phẩm kém.*
- *Trả lời ngẫu nhiên 1 câu trắc nghiệm:* $X = 1$ *nếu trả lời đúng,* $X = 0$ *nếu trả lời sai.*
- *Kết quả xét nghiệm một loại bệnh:* $X = 1$ *nếu kết quả là dương tính,* $X = 0$ *nếu kết quả là âm tính.*

Phân phối Bernoulli

Phân phối Bernoulli

Hàm xác suất của BNN $X \sim B(p)$ có dạng:

$$f(x) = p^x(1-p)^{1-x}, \quad x = 0, 1$$

PHÂN PHỐI BERNOULLI

Hàm xác suất của BNN $X \sim B(p)$ có dạng:

$$f(x) = p^x(1-p)^{1-x}, \quad x = 0, 1$$

Bảng phân phối xác suất của BNN $X \sim B(p)$ có dạng

X	1	0
$\mathbb{P}$	p	q

với $q = 1 - p$.

PHÂN PHỐI BERNOULLI

Hàm xác suất của BNN $X \sim B(p)$ có dạng:

$$f(x) = p^x(1-p)^{1-x}, \quad x = 0, 1$$

Bảng phân phối xác suất của BNN $X \sim B(p)$ có dạng

X	1	0
$\mathbb{P}$	p	q

với $q = 1 - p$.
Dựa vào bảng phân phối xác suất của biến ngẫu nhiên X ta dễ dàng tính được

$$\mathbb{E}(X) = p$$

$$\mathbb{V}(X) = pq$$

Phân phối nhị thức

Định nghĩa 1.2 (Binomial distribution)

Thực hiện n phép thử Bernoulli độc lập với xác suất thành công trong mỗi phép thử là p. Gọi X là số lần thành công (biến cố A xảy ra) trong n phép thử thì

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

với X_i, $(i = 1, \ldots, n)$, là biến ngẫu nhiên có phân phối Bernoulli với cùng tham số p. Khi đó X được gọi là có phân phối nhị thức với các tham số n, p ký hiệu $X \sim B(n; p)$.

Ví dụ 1.2

Một số ví dụ về BNN nhị thức:

- *Số linh kiện điện tử bị hỏng trong 20 linh kiện độc lập với xác suất bị hỏng của một linh kiện là 5%.*
- *Số lượng người thưởng trong 20 người tham gia quay số với xác suất trúng thưởng là 2%.*
- *Số người bị sốt phản vệ với một loại vacxin biết tỷ lệ bị sốt phản vệ là 80%.*

TÍNH CHẤT 1.1

- *Hàm xác suất của BNN $X \sim B(n,p)$ có dạng*

$$f(k) = \mathbb{P}(X = k) = C_n^k p^k q^{n-k}, \quad k \in \{0, 1, \ldots, n\}$$

Tính chất 1.1

- *Hàm xác suất của BNN $X \sim B(n,p)$ có dạng*

$$f(k) = \mathbb{P}(X = k) = C_n^k p^k q^{n-k}, \quad k \in \{0, 1, \ldots, n\}$$

- *Bảng phân phối xác suất của BNN $X \sim B(n,p)$ có dạng*

X	0	1	2	...	n
$\mathbb{P}$	q^n	npq^{n-1}	$C_n^2 p^2 q^{n-2}$	...	p^n

Tính chất 1.1

- *Hàm xác suất của BNN $X \sim B(n,p)$ có dạng*

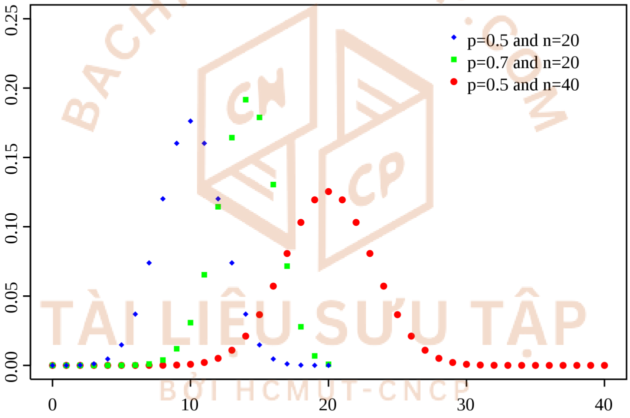
$$f(k) = \mathbb{P}(X=k) = C_n^k p^k q^{n-k}, \quad k \in \{0,1,\dots,n\}$$

- *Bảng phân phối xác suất của BNN $X \sim B(n,p)$ có dạng*

X	0	1	2	...	n
$\mathbb{P}$	q^n	npq^{n-1}	$C_n^2 p^2 q^{n-2}$	...	p^n

- $\mathbb{E}(X) = np$ *và* $\mathbb{V}(X) = npq$.

PHÂN PHỐI NHỊ THỨC - HÀM XÁC SUẤT



PHÂN PHỐI NHỊ THỨC

VÍ DỤ 1.3

Trong một nhà máy sản xuất vi mạch điện tử, biết rằng tỷ lệ vi mạch không đạt chất lượng là 5%. *Kiểm tra ngẫu nhiên 15 vi mạch. Tính xác suất*

(A) *Có đúng 7 vi mạch không đạt chất lượng.*

(B) *Có ít nhất 1 vi mạch không đạt chất lượng.*

(C) *Tính trung bình và độ lệch chuẩn cho số vi mạch bị hỏng.*

Phân phối siêu bội

Định nghĩa 1.3

Một tập hợp bao gồm N phần tử trong đó có

- *K phần tử thành công (chứa biến cố mà ta quan tâm)*
- *N − K phần tử không thành công (chứa biến cố mà ta không quan tâm)*

Một mẫu gồm n phần tử được chọn ngẫu nhiên (không hoàn lại) từ N phần tử trên ($K \leq N$ và $n \leq N$).
Biến ngẫu nhiên X là số phần tử thành công trong mẫu được chọn có phân phối siêu bội với hàm xác suất là

$$f(x) = \frac{C_K^x \, C_{N-K}^{n-x}}{C_N^n}, \quad x = \max\{0, n+K-N\}, \ldots, \max\{K, n\}.$$

PHÂN PHỐI SIÊU BỘI

VÍ DỤ 1.4

Môt thùng chứa 300 *linh kiện điện tử, trong đó có* 100 *linh kiện là hàng nội địa và* 200 *linh kiện là hàng nhập khẩu. Nếu chọn ngẫu nhiên* 4 *linh kiện (không hoàn lại). Tính*

- (A) *xác suất để cả* 4 *linh kiện đều là hàng nội địa.*
- (B) *xác suất để có* 2 *hoặc* 3 *linh kiện là hàng nội địa.*
- (C) *xác suất để có ít nhất một linh kiện từ hàng nội địa.*

PHÂN PHỐI SIÊU BỘI

CÁC ĐẶC TRUNG CỦA PHÂN PHỐI SIÊU BỘI H(N, K, N)

Nếu X là biến ngẫu nhiên siêu bội với các tham số N, K và n, thì

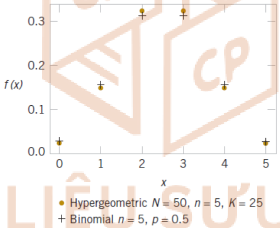
$$\mathbb{E}(X) = np, \quad \text{và} \quad \mathbb{V}ar(X) = np(1-p)\left(\frac{N-n}{N-1}\right)$$

với $p = K/N$. Ở đây, $\dfrac{N-n}{N-1}$ được gọi là hệ số hiệu chỉnh.

PHÂN PHỐI SIÊU BỘI

TÍNH CHẤT 1.2

Khi n là đủ nhỏ so với N thì phân phối siêu bội xấp xỉ phân phối nhị thức.



	0	1	2	3	4	5
Hypergeometric probability	0.025	0.149	0.326	0.326	0.149	0.025
Binomial probability	0.031	0.156	0.312	0.312	0.156	0.031

PHÂN PHỐI POISSON

VÍ DỤ 1.5

Giả sử ta cần truyền đi n bit qua một kênh truyền tín hiệu số. Gọi X là số bit lỗi trong n bit được truyền. Khi xác suất một bit bị lỗi là hằng số và việc truyền tín hiệu độc lập, X có phân phối nhị thức. Gọi p là xác suất một bit truyền đi bị lỗi. Đặt $\lambda = np$ thì $\mathbb{E}(X) = np = \lambda$ và

$$\mathbb{P}(X = x) = C_n^x p^x (1-p)^{n-x} = C_n^x \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

Phân phối Poisson

Ví dụ 1.5

Giả sử ta cần truyền đi n bit qua một kênh truyền tín hiệu số. Gọi X là số bit lỗi trong n bit được truyền. Khi xác suất một bit bị lỗi là hằng số và việc truyền tín hiệu độc lập, X có phân phối nhị thức. Gọi p là xác suất một bit truyền đi bị lỗi. Đặt $\lambda = np$ thì $\mathbb{E}(X) = np = \lambda$ và

$$\mathbb{P}(X = x) = C_n^x p^x (1-p)^{n-x} = C_n^x \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

Giả sử số truyền đi tăng lên và xác suất một bit lỗi giảm xuống sao cho np không đổi. Nghĩa là, n và p đồng thời tăng và giảm sao cho $\mathbb{E}(X) = \lambda$ là hằng số.

Phân phối Poisson

Ví dụ 1.5

Giả sử ta cần truyền đi n bit qua một kênh truyền tín hiệu số. Gọi X là số bit lỗi trong n bit được truyền. Khi xác suất một bit bị lỗi là hằng số và việc truyền tín hiệu độc lập, X có phân phối nhị thức. Gọi p là xác suất một bit truyền đi bị lỗi. Đặt $\lambda = np$ *thì* $\mathbb{E}(X) = np = \lambda$ *và*

$$\mathbb{P}(X = x) = C_n^x p^x (1-p)^{n-x} = C_n^x \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

Giả sử số truyền đi tăng lên và xác suất một bit lỗi giảm xuống sao cho np *không đổi. Nghĩa là, n và p đồng thời tăng và giảm sao cho* $\mathbb{E}(X) = \lambda$ *là hằng số. Ta có thể chỉ ra*

$$\lim_{n\to\infty} \mathbb{P}(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, 2, \ldots$$

PHÂN PHỐI POISSON

ĐỊNH NGHĨA 1.4 (POISSSON DISTRIBUTION)

Biến ngẫu nhiên rời rạc X nhận các giá trị từ $0, 1, 2, \ldots$ gọi là có phân phối Poisson với tham số λ, ký hiệu $X \sim P(\lambda)$ nếu

$$f(x) = \mathbb{P}(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!}, x = 0, 1, 2 \ldots$$

Phân phối Poisson

Định nghĩa 1.4 (Poissson distribution)

Biến ngẫu nhiên rời rạc X nhận các giá trị từ $0, 1, 2, \ldots$ *gọi là có phân phối Poisson với tham số* λ*, ký hiệu* $X \sim P(\lambda)$ *nếu*

$$f(x) = \mathbb{P}(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, 2\ldots$$

TÍNH CHẤT 1.3

TÍNH CHẤT 1.3

- *Nếu X lần lượt bằng* $\mathbb{E}(X) = \lambda$ *và* $\mathbb{V}(X) = \lambda$.

TÍNH CHẤT 1.3

- *Nếu X lần lượt bằng* $\mathbb{E}(X) = \lambda$ *và* $\mathbb{V}(X) = \lambda$.
- *Nếu các biến ngẫu nhiên* $X_i \sim P(\lambda_i), \quad (i = 1, \ldots, n)$ *độc lập thì*

$$X_1 + X_2 + \ldots + X_n \sim P(\lambda_1 + \lambda_2 + \ldots + \lambda_n).$$

TÍNH CHẤT 1.3

- *Nếu X lần lượt bằng* $\mathbb{E}(X) = \lambda$ *và* $\mathbb{V}(X) = \lambda$.
- *Nếu các biến ngẫu nhiên* $X_i \sim P(\lambda_i), \quad (i = 1, \ldots, n)$ *độc lập thì*

$$X_1 + X_2 + \ldots + X_n \sim P(\lambda_1 + \lambda_2 + \ldots + \lambda_n).$$

- *Xét* $Z \sim P(\lambda)$. *Giả sử rằng các sự kiện (outcome) được phân loại một cách độc lập vào m nhóm và xác suất để một sự kiện được phân loại vào nhóm k tương ứng với xác suất* p_k. *Khi đó số sự kiện thuộc nhóm k, ký hiệu* Z_k, *sẽ có phân phối Poisson với trung bình* $\lambda_k = p_k\lambda$, *và* $Z_k, k = 1, \ldots, m$ *là các BNN độc lập.*

Phân phối Poisson

Phân phối Poisson thường được dùng để mô tả số sự kiện trong một đơn vị thời gian hay số cá thể trong một đơn vị không gian. Ví dụ:

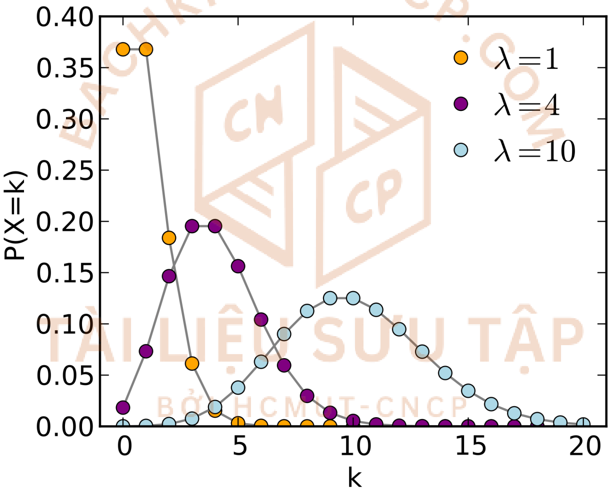
- Số côn trùng trong mỗi m^2 của một cánh đồng.

Phân phối Poisson

Phân phối Poisson thường được dùng để mô tả số sự kiện trong một đơn vị thời gian hay số cá thể trong một đơn vị không gian. Ví dụ:

- Số côn trùng trong mỗi m^2 của một cánh đồng.
- Số tai nạn giao thông mỗi tuần tại một thành phố.
- Số bệnh nhân Covid-19 mỗi ngày ở một quốc gia.

HÀM XÁC SUẤT


0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
P(X=k)
0
5
10
15
20
k
$\lambda=1$
$\lambda=4$
$\lambda=10$

PHÂN PHỐI POISSON

VÍ DỤ 1.6

Số cuộc điện thoại gọi đến một tổng đài điện thoại trong một giờ có phân phối Poisson với $\lambda = 10$. *Tính xác suất*

(A) *Có* 5 *cuộc điện thoại gọi đến trong một giờ.*

(B) *Có nhiều nhất* 3 *cuộc điện thoại gọi đến trong một giờ.*

(C) *Có* 15 *cuộc điện thoại gọi đến trong hai giờ.*

(D) *Có* 5 *cuộc điện thoại gọi đến trong 30 phút.*

Xấp xỉ PP nhị thức bằng PP Poisson

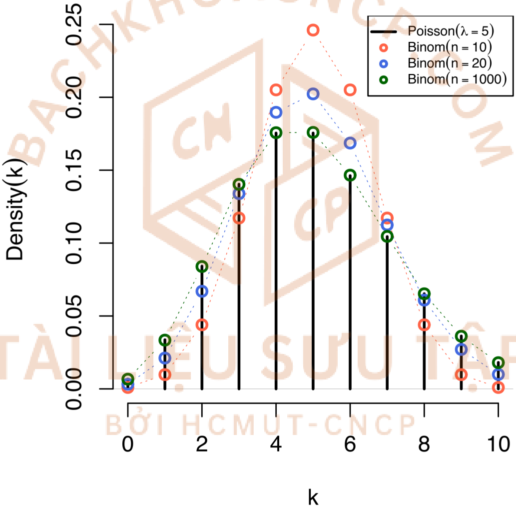
Bổ đề 1.1

Cho $X \sim B(n,p)$, nếu $n \to \infty$ và $p \to 0$ sao cho $np \to \lambda$ thì

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

Trong thực tế, phân phối Poisson sẽ xấp xỉ tốt cho phân phối nhị thức khi $n \geq 100$ và $np \leq 10$, khi đó BNN $X \sim B(n,p)$ sẽ sấp xỉ BNN $X \sim P(\lambda)$ với $\lambda = np$.

XẤP XỈ PP NHỊ THỨC BẰNG PP POISSON



XẤP XỈ PP NHỊ THỨC BẰNG PP POISSON

VÍ DỤ 1.7

Trong một đợt tiêm chủng cho trẻ em ở một khu vực, biết xác suất một trẻ bị phản ứng với thuốc sau khi tiêm là 0,001. Thực hiện tiêm cho 2000 trẻ, tính xác suất có nhiều nhất 1 trẻ bị phản ứng với thuốc sau khi tiêm.

Phân phối đều

Định nghĩa 2.1 (Uniform distribution)

Biến ngẫu nhiên liên tục X được gọi là có phân phối đều trên đoạn $[a;b]$, *ký hiệu* $X \sim U(a;b)$, *nếu hàm mật độ xác suất của X có dạng*

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{khi } x \in [a,b] \\ 0 & \text{nơi khác} \end{cases}$$

Phân phối đều

Định nghĩa 2.1 (Uniform distribution)

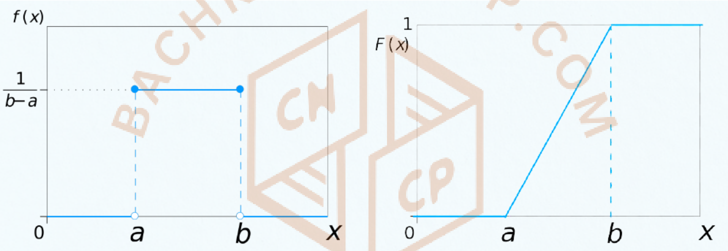
Biến ngẫu nhiên liên tục X được gọi là có phân phối đều trên đoạn $[a;b]$, *ký hiệu* $X \sim U(a;b)$, *nếu hàm mật độ xác suất của X có dạng*

$$f(x) = \begin{cases} \dfrac{1}{b-a} & khi\ x \in [a,b] \\ 0 & nơi\ khác \end{cases}$$

Từ định nghĩa trên ta có được hàm phân phối xác suất của $X \sim U(a;b)$

$$F(x) = \begin{cases} 0 & \text{khi } x < a \\ \dfrac{x-a}{b-a} & \text{khi } x \in [a,b] \\ 1 & \text{khi } x > b \end{cases}$$

PHÂN PHỐI ĐỀU - HÀM MẬT ĐỘ



VÍ DỤ 2.1

Cho $X \sim U(0,10)$ *tính*

- **A** $\mathbb{P}(X < 3)$
- **B** $\mathbb{P}(X > 6)$
- **C** $\mathbb{P}(3 < X < 8)$

Kỳ vọng và phương sai của PP đều

Bổ đề 2.1 (Các đặc trưng của biến ngẫu nhiên có phân phối đều)

Cho X là biến ngẫu nhiên có phân phối đều trên $[a,b]$ $(X \sim U(a;b))$ thì

I) *Kỳ vọng* $\mathbb{E}(X) = \dfrac{a+b}{2}$.

II) *Phương sai* $\mathbb{V}(X) = \dfrac{(b-a)^2}{12}$.

Phân phối đều

Ví dụ 2.2

Lịch xuất bến của một trạm xe buýt như sau: chiếc xe đầu tiên trong ngày sẽ khởi hành vào lúc 7 giờ, sau 15 phút sẽ có một xe khác đến trạm. Giả sử một hành khách đến trạm trong khoảng thời gian từ 7 giờ - 7 giờ 30. Tìm xác suất để hành khách này chờ

(A) *ít hơn 5 phút.*

(B) *ít nhất 12 phút.*

(C) *Tính trung bình và phương sai thời gian mà hành khách sẽ đến trạm xe buýt tính từ thời điểm 7 giờ.*

Phân phối mũ

Định nghĩa 2.2 (Exponential distribution)

Biến ngẫu nhiên T ($t > 0$) gọi là có phân phối mũ, ký hiệu $T \sim E(\lambda)$, nếu nó có hàm mật độ xác suất

$$f(t) = \lambda e^{-\lambda t}, \qquad t > 0 \tag{1}$$

trong đó

- *λ: số biến cố trung bình xảy ra trong một đơn vị thời gian*
- *t: số đơn vị thời gian cho đến biến cố kế tiếp*

Nhận xét 2.1

Nếu số sự kiện xảy ra trong mỗi đơn vị thời gian là một BNN $X \sim P(\lambda)$, thì khoảng cách (đơn vị thời gian) giữa hai sự kiên liên tiếp là BNN $T \sim E(\lambda)$.

CÁC ĐẶC TRƯNG CỦA PHÂN PHỐI MŨ

Hàm phân phối của T:

$$F(t) = \mathbb{P}(T \leq t) = 1 - e^{-\lambda t}, \quad t > 0 \tag{2}$$

CÁC ĐẶC TRƯNG CỦA PHÂN PHỐI MŨ

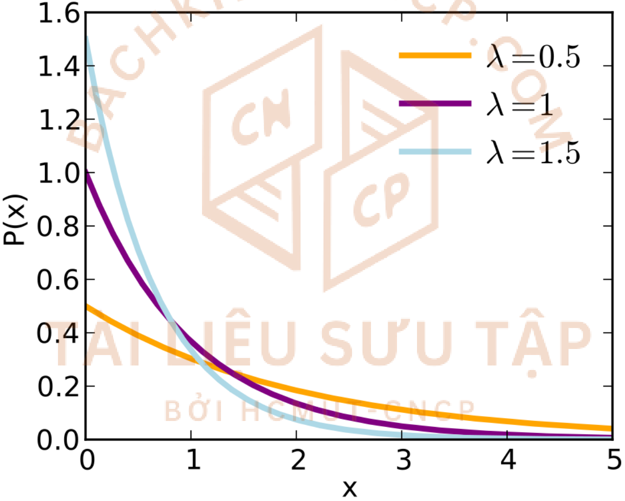
Hàm phân phối của T:

$$F(t) = \mathbb{P}(T \le t) = 1 - e^{-\lambda t}, \quad t > 0 \tag{2}$$

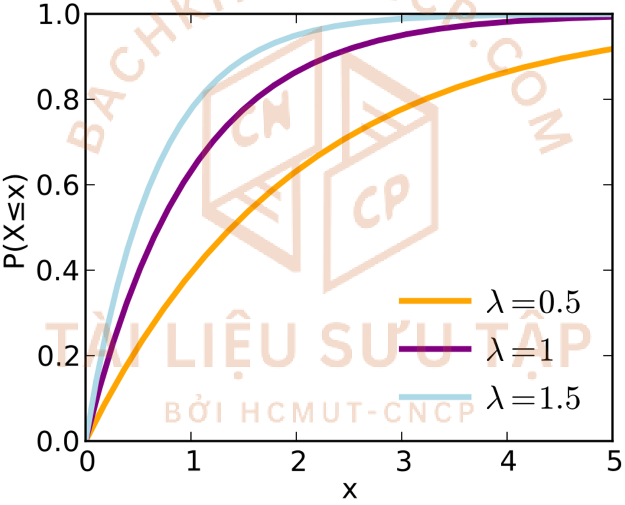
BỔ ĐỀ 2.2

Nếu $T \sim Exp(\lambda)$ thì kỳ vọng và phương sai của T lần lượng bằng $\mathbb{E}(T) = \frac{1}{\lambda}$ *và* $\mathbb{V}ar(T) = \frac{1}{\lambda^2}$

PHÂN PHỐI MŨ - HÀM MẬT ĐỘ



PHÂN PHỐI MŨ - HÀM PHÂN PHỐI



Phân phối mũ

Ví dụ 2.3

Trong một mạng máy tính ở một công ty, biết rằng số người dùng đăng nhập vào mạng trong một giờ có phân phối Poisson với trung bình bằng 25.

(A) *Tính xác suất không có người dùng nào đăng nhập trong khoảng thời gian 6 phút.*

(B) *Tính xác suất lần đăng nhập kế tiếp cách lần đăng nhập đầu từ 2 đến 3 phút.*

Ví dụ 2.4

Giả sử X *là thời gian một máy đếm Geiger dò được một nguồn phóng xạ có phân phối mũ với kỳ vọng* 1.4 *phút. Nếu ta bật máy đếm Geiger va chờ* 3 *phút trước khi tiến hành dò tìm phóng xạ thì xác suất nó phát hiện được nguồn phóng xạ trong* 30 *giây kế tiếp là bao nhiêu?*

PHÂN PHỐI CHUẨN

ĐỊNH NGHĨA 2.3 (NORMAL DISTRIBUTION)

Biến ngẫu nhiên liên tục X nhận giá trị trong khoảng $(-\infty, +\infty)$ được gọi là có phân phối chuẩn tham số μ, σ nếu hàm mật độ xác suất có dạng

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad -\infty < x < +\infty \tag{3}$$

trong đó μ, σ là hằng số và $\sigma > 0$, $-\infty < \mu < +\infty$, ký hiệu $X \sim \mathcal{N}(\mu; \sigma^2)$.

PHÂN PHỐI CHUẨN

ĐỊNH NGHĨA 2.3 (NORMAL DISTRIBUTION)

Biến ngẫu nhiên liên tục X nhận giá trị trong khoảng $(-\infty,+\infty)$ được gọi là có phân phối chuẩn tham số μ, σ nếu hàm mật độ xác suất có dạng

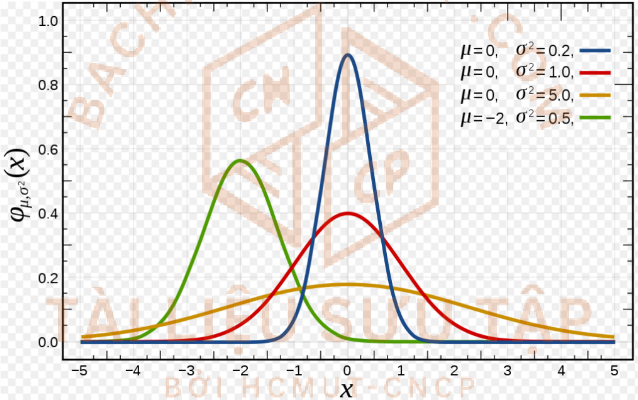
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad -\infty < x < +\infty \tag{3}$$

trong đó μ, σ là hằng số và $\sigma > 0$, $-\infty < \mu < +\infty$, ký hiệu $X \sim \mathcal{N}(\mu;\sigma^2)$.

Nếu $X \sim \mathcal{N}(\mu,\sigma^2)$

$$\mathbb{E}(X) = \mu$$
$$\mathbb{V}ar(X) = \sigma^2$$

PHÂN PHỐI CHUẨN



Phân phối chuẩn

Nhờ vào định lý sau, nên nếu biến ngẫu nhiên X có phân phối chuẩn thì biến đổi tuyến tính của X cũng có phân phối chuẩn.

Bổ đề 2.3 (Tính "tuyến tính" của phân phối chuẩn)

Nếu biến ngẫu nhiên X có phân phối chuẩn với kỳ vọng μ, phương sai σ^2 và nếu $Y = aX + b$, (a, b là hằng số và $a \neq 0$), thì Y có phân phối chuẩn với kỳ vọng $a\mu + b$ và phương sai $a^2\sigma^2$.

Phân phối chuẩn

Nhờ vào định lý sau, nên nếu biến ngẫu nhiên X có phân phối chuẩn thì biến đổi tuyến tính của X cũng có phân phối chuẩn.

Bổ đề 2.3 (Tính "tuyến tính" của phân phối chuẩn)

Nếu biến ngẫu nhiên X có phân phối chuẩn với kỳ vọng μ, phương sai σ^2 và nếu $Y = aX + b$, (a, b là hằng số và $a \neq 0$), thì Y có phân phối chuẩn với kỳ vọng $a\mu + b$ và phương sai $a^2\sigma^2$.

Bổ đề 2.4

Nếu các biến ngẫu nhiên $X_1, \ldots, X_n$ là độc lập và nếu X_i có phân phối chuẩn với kỳ vọng μ_i và phương sai σ_i^2, $(i = 1, 2, \ldots, n)$, thì tổng $X_1 + \cdots + X_n$ có phân phối chuẩn với kỳ vọng là $\mu_1 + \cdots + \mu_n$ và phương sai là $\sigma_1^2 + \cdots + \sigma_n^2$.

Phân phối chuẩn

Tính chất 2.1

Nếu các biến ngẫu nhiên $X_1, \ldots, X_n$ là độc lập và X_i có phân phối chuẩn với kỳ vọng μ_i và phương sai σ_i^2, $(i = 1, \ldots, n)$. $a_i, \ldots, a_n$ là các hằng số sao cho có ít nhất một $a_i \neq 0$, thì biến ngẫu nhiên $a_1X_1 + \cdots + a_nX_n$ có phân phối chuẩn với kỳ vọng $a_1\mu_1 + \cdots + a_n\mu_n$ và phương sai $a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2$.

Ví dụ 2.5

Nếu $X_1 \sim \mathcal{N}(1,2)$, $X_2 \sim \mathcal{N}(-1,1/2)$, và $X_3 \sim \mathcal{N}(0,2)$, Tìm phân phối của $X_1 + 2X_2 + X_3$.

PHÂN PHỐI CHUẨN HÓA

ĐỊNH NGHĨA 2.4 (STANDARD NORMAL DISTRIBUTION)

Biến ngẫu nhiên Z được gọi là có phân phối chuẩn hóa nếu nó có phân phối chuẩn với tham số $\mu = 0$ và $\sigma^2 = 1$, ký hiệu $Z \sim \mathcal{N}(0;1)$.

Phân phối chuẩn hóa

Định nghĩa 2.4 (Standard normal distribution)

Biến ngẫu nhiên Z được gọi là có phân phối chuẩn hóa nếu nó có phân phối chuẩn với tham số $\mu = 0$ và $\sigma^2 = 1$, ký hiệu $Z \sim \mathcal{N}(0;1)$.

Theo quy ước, hàm phân phối của biến ngẫu nhiên chuẩn hóa được ký hiệu là $\Phi(z)$, tức

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{x^2}{2}}\, dx$$

Tính chất 2.2

$\Phi(-z) = 1 - \Phi(z)$.

Phân phối chuẩn hóa

Theo định lý về tính tuyến tính của phân phối chuẩn, nếu $X \sim \mathcal{N}(\mu;\sigma^2)$ thì $\dfrac{X-\mu}{\sigma}$ có phân phối chuẩn hóa hay

$$\frac{X-\mu}{\sigma} \sim N(0;1)$$

Phân phối chuẩn hóa

Theo định lý về tính tuyến tính của phân phối chuẩn, nếu $X \sim \mathcal{N}\left(\mu;\sigma^2\right)$ thì $\dfrac{X-\mu}{\sigma}$ có phân phối chuẩn hóa hay

$$\frac{X-\mu}{\sigma} \sim N(0;1)$$

Dựa vào tính chất này ta có thể tính xác suất của biến ngẫu nhiên $X \sim \mathcal{N}\left(\mu;\sigma^2\right)$.

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

Phân phối chuẩn hóa

Theo định lý về tính tuyến tính của phân phối chuẩn, nếu $X \sim \mathcal{N}\left(\mu;\sigma^2\right)$ thì $\dfrac{X-\mu}{\sigma}$ có phân phối chuẩn hóa hay

$$\frac{X-\mu}{\sigma} \sim N(0;1)$$

Dựa vào tính chất này ta có thể tính xác suất của biến ngẫu nhiên $X \sim \mathcal{N}\left(\mu;\sigma^2\right)$.

$$\mathbb{P}(X \le b) = \mathbb{P}\left(\frac{X-\mu}{\sigma} \le \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right)$$

Tương tự, với $a \le b$ thì

$$\mathbb{P}(a < X \le b) = \mathbb{P}(X \le b) - \mathbb{P}(X \le a) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

Phân phối chuẩn hóa

Nếu $X \sim \mathcal{N}(\mu;\sigma^2)$ thì

$$\mathbb{P}(|X-\mu| \le k\sigma) = \mathbb{P}\left(-k \le \frac{X-\mu}{\sigma} \le k\right) = 2\Phi(k) - 1$$

người ta hay gọi đẳng thức trên là "Quy tắc k-sigma $(k\sigma)$".

Phân phối chuẩn hóa

Nếu $X \sim \mathcal{N}\left(\mu; \sigma^2\right)$ thì

$$\begin{aligned}\mathbb{P}\left(|X-\mu| \leq k\sigma\right) &= \mathbb{P}\left(-k \leq \frac{X-\mu}{\sigma} \leq k\right) \\ &= 2\Phi(k) - 1\end{aligned}$$

người ta hay gọi đẳng thức trên là "Quy tắc k-sigma $(k\sigma)$". Với $k = 3$ ta có quy tắc 3-sigma:

$$\begin{aligned}\mathbb{P}\left(|X-\mu| \leq 3\sigma\right) &= \mathbb{P}\left(-k \leq \frac{X-\mu}{\sigma} \leq k\right) \\ &= 2\Phi(3) - 1 \approx 0.9973\end{aligned}$$

"Sai số giữa X và μ không quá 3 σ là gần chắc chắn (xác suất gần bằng 1)."

PHÂN VỊ CHUẨN HÓA

ĐỊNH NGHĨA 2.5 (PHÂN VỊ CHUẨN HÓA, NORMAL QUARTILE)

Cho biến ngẫu nhiên $X \sim \mathcal{N}(\mu;\sigma^2)$, *phân vị chuẩn hóa mức* α, *ký hiệu* x_α, *là giá trị của biến ngẫu nhiên* X *thỏa mãn điều kiện* $\mathbb{P}(X \leq x_\alpha) = \alpha$

Ví dụ 2.6

Đường kính của một chi tiết máy do một máy tiện sản xuất có phân phối chuẩn với kỳ vọng 20 mm, phương sai 0.2 mm^2. Lấy ngẫu nhiên một chi tiết, tính các xác suất sau.

A) *Chi tiết được chọn có đường kính trong khoảng 19.9 mm đến 20.3 mm.*

B) *Chi tiết được chọn có đường kính sai khác với kỳ vọng không quá 0.3 mm.*

C) *Tìm giá trị đường kính để có 90% các chi tiết máy có đường kính nhỏ hơn giá trị này.*

ĐỊNH LÝ GIỚI HẠN TRUNG TÂM

MỆNH ĐỀ 3.1

Xét $X_1, \ldots, X_n$ là n BNN độc lập, có cùng phân phối với kỳ vọng μ và phương sai hữu hạn σ^2. Xét $\overline{X} = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$ là trung bình mẫu.

ĐỊNH LÝ GIỚI HẠN TRUNG TÂM

MỆNH ĐỀ 3.1

Xét $X_1, \dots, X_n$ là n BNN độc lập, có cùng phân phối với kỳ vọng μ và phương sai hữu hạn σ^2. Xét $\overline{X} = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$ là trung bình mẫu. Khi đó, BNN

$$Z_n = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$

hội tụ theo phân phối về BNN chuẩn hóa khi n lớn, tức là

$$\lim_{n\to\infty} \mathbb{P}(Z_n \le x) = \Phi(x),\ \forall x \in \mathbb{R},$$

trong đó $\Phi(x)$ là hàm phân phối xác suất của BNN chuẩn hóa.

ĐỊNH LÝ GIỚI HẠN TRUNG TÂM

MỆNH ĐỀ 3.1

Xét $X_1, \ldots, X_n$ là n BNN độc lập, có cùng phân phối với kỳ vọng μ và phương sai hữu hạn σ^2. Xét $\overline{X} = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$ là trung bình mẫu. Khi đó, BNN

$$Z_n = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$

hội tụ theo phân phối về BNN chuẩn hóa khi n lớn, tức là

$$\lim_{n\to\infty} \mathbb{P}(Z_n \le x) = \Phi(x),\ \forall x \in \mathbb{R},$$

trong đó $\Phi(x)$ là hàm phân phối xác suất của BNN chuẩn hóa.
Tương tự, gọi $S_n = X_1 + \ldots + X_n$ là tổng mẫu, khi đó $S_n \approx N\left(n\mu; n\sigma^2\right)$.

VÍ DỤ 3.1 (PHÂN PHỐI NHỊ THỨC)

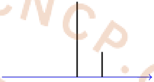
Assumptions:

- $X_1, X_2 \ldots$ are iid Bernoulli(p).
- $Z_n = \dfrac{X_1 + X_2 + \ldots + X_n - np}{\sqrt{np(1-p)}}$.

We choose $p = \frac{1}{3}$.

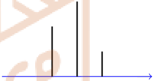
$$Z_1 = \frac{X_1 - p}{\sqrt{p(1-p)}}$$

PMF of Z_1



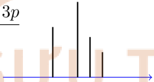
$$Z_2 = \frac{X_1 + X_2 - 2p}{\sqrt{2p(1-p)}}$$

PMF of Z_2



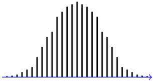
$$Z_3 = \frac{X_1 + X_2 + X_3 - 3p}{\sqrt{3p(1-p)}}$$

PMF of Z_3



$$Z_{30} = \frac{\sum_{i=1}^{30} X_i - 30p}{\sqrt{30p(1-p)}}$$

PMF of Z_{30}



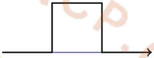
VÍ DỤ 3.2 (PHÂN PHỐI ĐỀU)

Assumptions:

- $X_1, X_2 \dots$ are iid Uniform(0,1).
- $Z_n = \dfrac{X_1 + X_2 + \ldots + X_n - \frac{n}{2}}{\sqrt{\frac{n}{12}}}$.

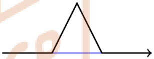
$$Z_1 = \frac{X_1 - \frac{1}{2}}{\sqrt{\frac{1}{12}}}$$

PDF of Z_1



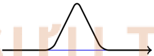
$$Z_2 = \frac{X_1 + X_2 - 1}{\sqrt{\frac{2}{12}}}$$

PDF of Z_2



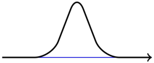
$$Z_3 = \frac{X_1 + X_2 + X_3 - \frac{3}{2}}{\sqrt{\frac{3}{12}}}$$

PDF of Z_3



$$Z_{30} = \frac{\sum_{i=1}^{30} X_i - \frac{30}{2}}{\sqrt{\frac{30}{12}}}$$

PDF of Z_{30}



NHẬN XÉT:

- Định lý giá trị trung tâm có thể áp dụng cho mọi BNN X_i có phân phối bất kì: phân phối rời rạc, phân phối liên tục, BNN với phân phối hỗn hợp.

NHẬN XÉT:

- Định lý giá trị trung tâm có thể áp dụng cho mọi BNN X_i có phân phối bất kì: phân phối rời rạc, phân phối liên tục, BNN với phân phối hỗn hợp.
- Định lý giá trị trung tâm có nhiều ứng dụng khi mà BNN mà ta quan tâm là tổng của nhiều BNN.

ĐỊNH LÝ GIÁ TRỊ TRUNG TÂM

VÍ DỤ 3.3

Nhân viên giao dịch ngân hàng phục vụ từng khách hàng đang xếp hàng. Giả sử rằng thời gian phục vụ cho khách hàng thứ i là X_i với trung bình là $\mathbb{E}(X_i) = 2$ (phút) và phương sai $\mathbb{V}(X_i) = 1$. Giả sử rằng thời gian phục vụ cho các khách hàng là độc lập với nhau. Gọi Y là tổng thời gian giao dịch viên phải phục vụ cho 50 khách hàng. Tính $P(90 < Y < 110)$.

Ví dụ 3.4

Một công ty sản xuất kẹo trẻ em đã triển khai chương trình tặng quà ngẫu nhiên trong dịp kỷ niệm 50 thành lập của công ty. 50% các túi kẹo trong chương trình này sẽ có trúng thưởng một món đồ chơi trẻ em. Một bạn nhỏ đã mua 20 gói kẹo trong đợt này. Hãy tính xác suất để bạn này đạt 8 đến 10 gói có thưởng.

Hiệu chỉnh liên tục cho BNN rời rạc

Ghi chú:

Xét $X_1, X_2, \ldots, X_n$ là các BNN độc lập cùng phân phối rời rạc. Gọi

$$Y = X_1 + X_2 + \cdots + X_n.$$

Giả sử chúng ta cần tính $\mathbb{P}(A) = \mathbb{P}(l \leq Y \leq u)$ bằng cách sử dụng định giới hạn trung tâm, trong đó l và u là các số nguyên và Y là BNN nhận giá nguyên. Thay vào đó chúng ta có thể tính

$$\mathbb{P}(A) = \mathbb{P}(l - 0.5 \leq Y \leq u + 0.5)$$

để ước lượng cho $P(A)$ khi áp dụng định lý giá trị trung tâm. Cách làm này gọi là **Hiệu chỉnh liên tục cho BNN rời rạc**, nó đặc biệt hữu ích khi áp dụng cho X_i là các BNN Bernoulli (Y là BNN nhị thức).