

THỐNG KÊ MÁY TÍNH & ỨNG DỤNG

Bài 5

MỘT SỐ PHÂN PHỐI XÁC SUẤT

Vũ Quốc Hoàng
(vqhoang@fit.hcmus.edu.vn)

FIT-HCMUS, 2018

Nội dung

- Phân phối Bernoulli
- Phân phối nhị thức
- Phân phối siêu bội
- Phân phối nhị thức âm và phân phối hình học
- Phân phối Poisson
- Phân phối đều (liên tục)
- Phân phối chuẩn
- Phân phối mũ

Phân phối Bernoulli

- B.n.n rời rạc X được gọi là có *phân phối Bernoulli* (Bernoulli distribution) với tham số p ($0 \leq p \leq 1$), kí hiệu $X \sim \text{Bernoulli}(p)$, nếu X có tập giá trị là $\{0, 1\}$ với xác suất

$$P(X = 1) = p \text{ và } P(X = 0) = 1 - p$$

- Khi đó, X có kì vọng và phương sai

$$E(X) = p \text{ và } \text{Var}(X) = p(1 - p)$$

- Ví dụ:

- Xét thí nghiệm tung một đồng xu đồng chất, gọi X là b.n.n “số lần được ngửa”, khi đó $X \sim \text{Bernoulli}(0.5)$
- Xét thí nghiệm T với biến cố A có $P(A) = p$, khi đó $I_A \sim \text{Bernoulli}(p)$

Phân phối nhị thức

- B.n.n rời rạc X được gọi là có **phân phối nhị thức** (binomial distribution) với tham số n ($n > 0$), p ($0 \leq p \leq 1$), kí hiệu $X \sim B(n, p)$, nếu X có tập giá trị là $\{0, 1, \dots, n\}$ với xác suất

$$f(k) = P(X = k) = C_n^k p^k (1 - p)^{n-k}$$

- Đặt $q = 1 - p$, ta có

$$E(X) = np \text{ và } Var(X) = np(1 - p)$$

- Nếu X_1, X_2, \dots, X_n là các b.n.n độc lập, có cùng phân phối Bernoulli(p) và $X = \sum_{i=1}^n X_i$ thì $X \sim B(n, p)$
- Nếu X_1, X_2, \dots, X_n là các b.n.n độc lập, $X_i \sim B(n_i, p)$ và $X = \sum_{i=1}^n X_i$ thì $X \sim B(\sum_{i=1}^n n_i, p)$

Phân phối nhị thức

Ví dụ

- Xét thí nghiệm T với biến cố A có $P(A) = p$. Xét thí nghiệm R “thực hiện T lặp lại n lần **độc lập**”, gọi X là b.n.n “số lần A xảy ra” thì $X \sim B(n, p)$
- Đề thi gồm 50 câu trắc nghiệm, mỗi câu trắc nghiệm chọn một trong 4 lựa chọn. Chọn đáp án ngẫu nhiên cho mỗi câu, gọi X là b.n.n “số câu đúng” thì $X \sim B(50, 1/4)$. Khi đó:

- Xác suất được 5 điểm là: $P(X = 25) = C_{50}^{25} (1/4)^{25} (3/4)^{25} = 8.45 \times 10^{-5}$

- Xác suất được điểm ≤ 2 là:

$$P(X \leq 10) = \sum_{x=0}^{10} C_{50}^x (1/4)^x (3/4)^{50-x} = 0.26$$

- Xác suất được điểm ≥ 8 là:

$$P(X \geq 40) = \sum_{x=40}^{50} C_{50}^x (1/4)^x (3/4)^{50-x} = 5.2 \times 10^{-16}$$

- Kỳ vọng của điểm đạt được là: $E\left(\frac{10}{50}X\right) = \frac{10}{50}E(X) = \frac{10}{50} \times 50 \times \frac{1}{4} = 2.5$

Phân phối nhị thức

Ví dụ

- Trong tổng thể N phần tử có K phần tử có tính chất P , xét thí nghiệm “chọn ngẫu nhiên một mẫu n phần tử **có hoàn lại** từ tổng thể”, gọi X là b.n.n “số phần tử có tính chất P trong mẫu” thì $X \sim B(n, p)$ với

$$p = \frac{K}{N}$$

- Trong hộp có 10 bi đỏ và 20 bi đen. Bốc ngẫu nhiên 10 viên có hoàn lại. Gọi X là b.n.n “số bi đỏ bốc được” thì $X \sim B(10, 1/3)$. Khi đó:

- Xác suất bốc được 5 bi đỏ là:

$$P(X = k = 5) = C_n^k p^k (1 - p)^{n-k} = C_{10}^5 \left(\frac{1}{3}\right)^5 \left(\frac{2}{3}\right)^5 = 0.1365$$

Phân phối siêu bội

- B.n.n rời rạc X được gọi là có **phân phối siêu bội** (hypergeometric distribution) với tham số n, N, K ($0 \leq K \leq N, 0 \leq n \leq N$), kí hiệu $X \sim \text{Hypergeometric}(n, N, K)$, nếu X có tập giá trị là $\{\max(0, n + K - N), \dots, \min(n, K)\}$ với xác suất

$$f(k) = P(X = k) = \frac{C_K^k C_{N-K}^{n-k}}{C_N^n}$$

- Đặt $p = \frac{K}{N}, q = 1 - p = \frac{N-K}{N}$, ta có

$$E(X) = np \text{ và } \text{Var}(X) = npq \frac{(N-n)}{(N-1)}$$

Phân phối siêu bội

Ví dụ

- Trong tổng thể N phần tử có K phần tử có tính chất P , xét thí nghiệm “chọn ngẫu nhiên một mẫu n phần tử **không hoàn lại** từ tổng thể”, gọi X là b.n.n “số phần tử có tính chất P trong mẫu” thì $X \sim \text{Hypergeometric}(n, N, K)$
- Trong hộp có 10 bi đỏ và 20 bi đen. Bốc ngẫu nhiên 10 viên không hoàn lại. Gọi X là b.n.n “số bi đỏ bốc được” thì $X \sim \text{Hypergeometric}(10, 30, 10)$. Khi đó:

- Xác suất bốc được 5 bi đỏ là:

$$P(X = k = 5) = \frac{C_K^k C_{N-K}^{n-k}}{C_N^n} = \frac{C_{10}^5 C_{30-10}^{10-5}}{C_{30}^{10}} = \frac{C_{10}^5 C_{20}^5}{C_{30}^{10}} = 0.13$$

Phân phối nhị thức âm và phân phối hình học

- B.n.n rời rạc X được gọi là có **phân phối nhị thức âm** (negative binomial distribution) với tham số r ($r > 0$), p ($0 < p < 1$), kí hiệu $X \sim NB(r, p)$, nếu X có tập giá trị là $\{0, 1, 2, \dots\}$ với xác suất

$$f(k) = P(X = k) = C_{r+k-1}^k p^r (1-p)^k$$

- Khi đó, X có kì vọng và phương sai

$$E(X) = \frac{r(1-p)}{p} \text{ và } Var(X) = \frac{r(1-p)}{p^2}$$

- Xét thí nghiệm T với biến cố A có $P(A) = p$. Xét thí nghiệm R “thực hiện T lặp lại **độc lập** cho đến khi có đúng r lần A xảy ra thì dừng”, gọi X là b.n.n “số lần A không xảy ra” thì $X \sim NB(r, p)$

Phân phối nhị thức âm và phân phối hình học

- Nếu $X \sim NB(1, p)$ thì X được gọi là có **phân phối hình học** (geometric distribution) với tham số p ($0 < p < 1$). Khi đó X có tập giá trị là $\{0, 1, 2, \dots\}$ với xác suất

$$f(k) = P(X = k) = p(1 - p)^k$$

- Khi đó, X có kì vọng và phương sai

$$E(X) = \frac{1-p}{p} \text{ và } Var(X) = \frac{1-p}{p^2}$$

- Xét thí nghiệm T với biến cố A có $P(A) = p$. Xét thí nghiệm R “thực hiện T lặp lại **độc lập** cho đến khi A xảy ra thì dừng”, gọi X là b.n.n “số lần A không xảy ra” thì $X \sim NB(1, p)$

Phân phối nhị thức âm và phân phối hình học

Ví dụ

- Xét thí nghiệm bắn đạn vào bia cho đến khi trúng thì dừng. Giả sử các lần bắn độc lập nhau với cùng xác suất trúng là $1/20$, gọi X là b.n.n “số viên đạn trật” thì $X \sim NB(1, 0.05)$. Khi đó:

- Xác suất bắn trật 5 viên đạn là: $P(X = 5) = 0.05(1 - 0.05)^5 = 0.0387$
- Xác suất dùng ≤ 5 viên đạn là:

$$P(X \leq 4) = \sum_{x=0}^4 0.05(1 - 0.05)^x = 0.2262$$

- Xác suất dung nhiều hơn 100 viên đạn là:

$$P(X > 99) = 1 - P(X \leq 99) = 1 - \sum_{x=0}^{99} 0.05(1 - 0.05)^x = 0.0059$$

- Số viên đạn trung bình: $E(X + 1) = E(X) + 1 = \frac{1-0.05}{0.05} = 20$

Phân phối Poisson

- B.n.n rời rạc X được gọi là có **phân phối Poisson** (Poisson distribution) với tham số λ ($\lambda > 0$), kí hiệu $X \sim P(\lambda)$, nếu X có tập giá trị là $\{0, 1, 2, \dots\}$ với xác suất

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

- Khi đó, X có kì vọng và phương sai

$$E(X) = \lambda \text{ và } Var(X) = \lambda$$

- Nếu X_1, X_2, \dots, X_n là các b.n.n độc lập, $X_i \sim P(\lambda_i)$ và $X = \sum_{i=1}^n X_i$ thì $X \sim P(\sum_{i=1}^n \lambda_i)$

Phân phối Poisson

- Cho $X \sim B(n, p)$, khi n lớn, p nhỏ thì phân phối của X “xấp xỉ” phân phối Poisson với tham số $\lambda = np$, tức là

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

- Ví dụ: lượng khách đến một tiệm có tỉ lệ trung bình là 4.5 khách trong một giờ. Gọi X là số lượng khách đến trong 1 giờ, xác định phân phối của X
 - Giả sử trong mỗi khoảng thời gian 1 giây có tối đa 1 khách đến với xác suất là tỉ lệ khách đến trong 1 giây, tức là $4.5/3600 = 0.00125$. Giả sử lượng khách đến trong các khoảng thời gian là độc lập thì $X \sim B(n, p)$ với $n = 3600, p = 0.00125$
 - X có phân phối Poisson với tham số $\lambda = np = 4.5$
 - Xác suất tiệm vắng khách trong một giờ là

$$P(X = 0) = e^{-\lambda} \frac{\lambda^k}{k!} = e^{-4.5} \frac{4.5^0}{0!} = 0.0111$$

Phân phối đều (liên tục)

- B.n.n liên tục X được gọi là có **phân phối đều** (uniform distribution) với tham số a, b ($a < b$), kí hiệu $X \sim U(a, b)$, nếu X có tập giá trị là $[a, b]$ với hàm mật độ xác suất

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{với } a \leq x \leq b \\ 0 & \text{khác} \end{cases}$$

- Khi đó, X có kì vọng và phương sai

$$E(X) = \frac{a+b}{2} \text{ và } Var(X) = \frac{(b-a)^2}{12}$$

- Gọi X là kết quả của thí nghiệm “chọn **ngẫu nhiên** một điểm trong khoảng $[a, b]$ ” thì $X \sim U(a, b)$

Phân phối chuẩn

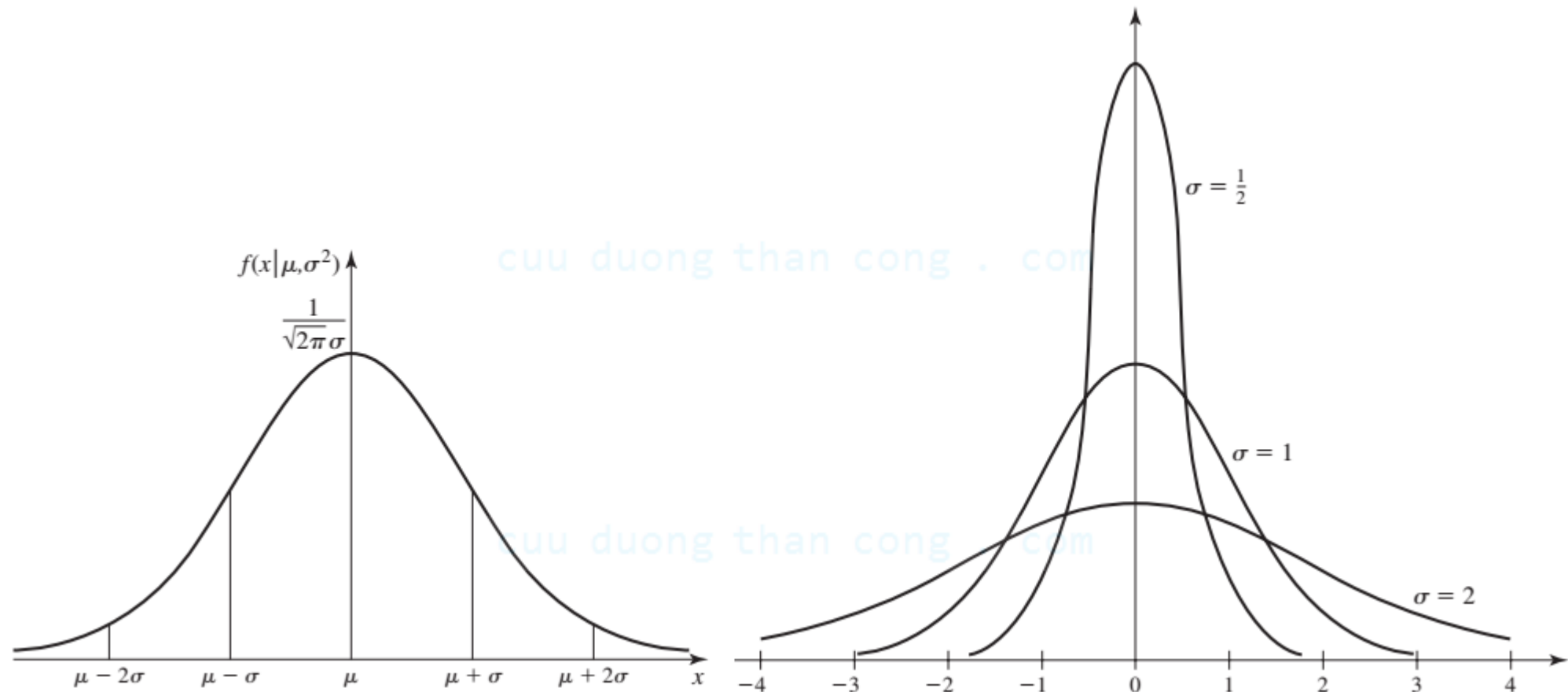
- B.n.n liên tục X được gọi là có *phân phối chuẩn* (normal distribution) với trung bình μ và phương sai σ^2 ($\sigma > 0$), kí hiệu $X \sim N(\mu, \sigma^2)$ nếu X có hàm mật độ xác suất:

$$f(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in \mathbb{R}$$

- Trường hợp $Z \sim N(0, 1)$ thì Z được gọi là có *phân phối chuẩn tắc* (standard normal distribution). Khi đó Z có hàm mật độ xác suất:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, z \in \mathbb{R}$$

Phân phối chuẩn Hàm mật độ



Phân phối chuẩn

Các tính chất

- $X \sim N(\mu, \sigma^2)$ thì $E(X) = \mu$ và $Var(X) = \sigma^2$
- $Z \sim N(0, 1)$ thì $E(Z) = 0$ và $Var(Z) = 1$
- Nếu $X \sim N(\mu, \sigma^2)$ và $Y = aX + b$ ($a \neq 0$) thì $Y \sim N(a\mu + b, a^2\sigma^2)$
- Nếu $X \sim N(\mu, \sigma^2)$ và $Z = \frac{X-\mu}{\sigma}$ thì $Z \sim N(0, 1)$
- Nếu $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$ và X_1, X_2 độc lập thì $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

Phân phối chuẩn

Hàm phân phối tích lũy và hàm phân vị

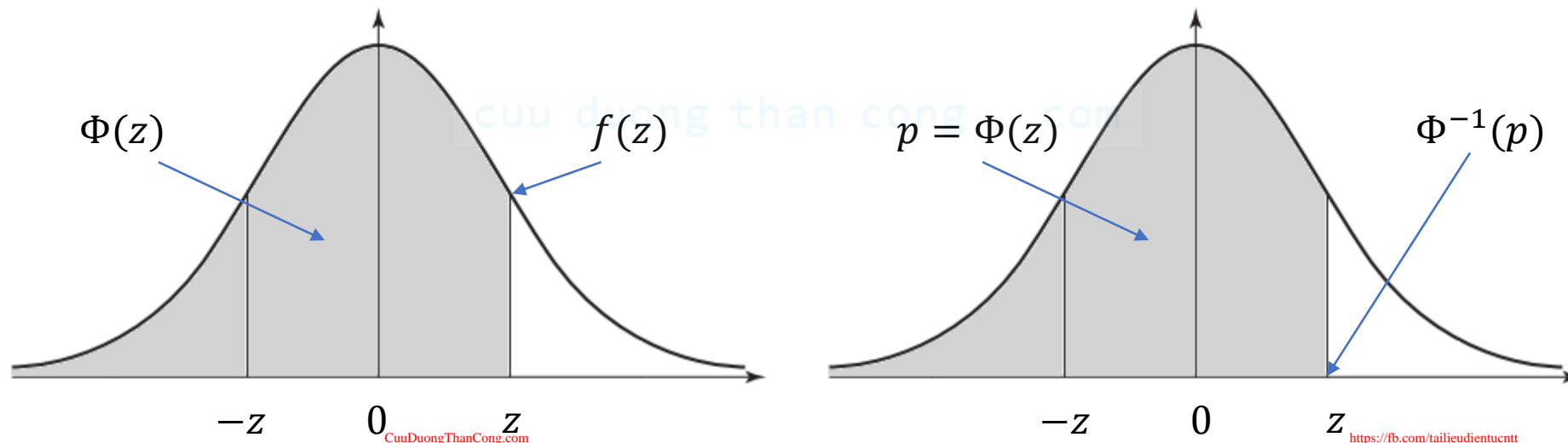
- Hàm phân phối tích lũy của b.n.n chuẩn tắc $Z \sim N(0, 1)$

$$\Phi(z) = F_Z(z) = P(Z \leq z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

- Do tính đối xứng của f_z nên với mọi z và mọi $0 < p < 1$ ta có:

$$\Phi(-z) = 1 - \Phi(z) \text{ và } \Phi^{-1}(p) = -\Phi^{-1}(1 - p)$$

- $\Phi(0) = 0.5, \Phi^{-1}(0.5) = 0$. Với mọi $a \geq 0$ ta có $P(-a \leq Z \leq a) = 2\Phi(a) - 1$



Phân phối chuẩn

Ví dụ

- Cho b.n.n $Z \sim N(0, 1)$, ta có:
 - $P(Z \leq 0.6) = \Phi(0.6) = 0.7257$
 - $P(Z \leq -0.6) = 1 - \Phi(0.6) = 0.2743$
 - $P(0.6 \leq Z \leq 1.2)$
 $= P(Z \leq 1.2) - P(Z < 0.6)$
 $= \Phi(1.2) - \Phi(0.6) = 0.8849 - 0.7257$
 - Xác định z để $P(Z \leq z) = 0.56$
 $z = \Phi^{-1}(0.56) = 0.15$
 - Xác định z để $P(Z \leq z) = 0.44$
 $z = -\Phi^{-1}(1 - 0.44) = -\Phi^{-1}(0.56)$
 $= -0.15$

Table of the Standard Normal Distribution

$$\Phi(x) = \int_{-\infty}^x \frac{1}{(2\pi)^{1/2}} \exp\left(-\frac{1}{2}u^2\right) du$$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0.00	0.5000	0.60	0.7257	1.20	0.8849
0.01	0.5040	0.61	0.7291	1.21	0.8869
0.02	0.5080	0.62	0.7324	1.22	0.8888
0.03	0.5120	0.63	0.7357	1.23	0.8907
0.04	0.5160	0.64	0.7389	1.24	0.8925
0.05	0.5199	0.65	0.7422	1.25	0.8944
0.06	0.5239	0.66	0.7454	1.26	0.8962
0.07	0.5279	0.67	0.7486	1.27	0.8980
0.08	0.5319	0.68	0.7517	1.28	0.8997
0.09	0.5359	0.69	0.7549	1.29	0.9015
0.10	0.5398	0.70	0.7580	1.30	0.9032
0.11	0.5438	0.71	0.7611	1.31	0.9049
0.12	0.5478	0.72	0.7642	1.32	0.9066
0.13	0.5517	0.73	0.7673	1.33	0.9082
0.14	0.5557	0.74	0.7704	1.34	0.9099
0.15	0.5596	0.75	0.7734	1.35	0.9115
0.16	0.5636	0.76	0.7764	1.36	0.9131
0.17	0.5675	0.77	0.7794	1.37	0.9147
0.18	0.5714	0.78	0.7823	1.38	0.9162
0.19	0.5753	0.79	0.7852	1.39	0.9177
0.20	0.5793	0.80	0.7881	1.40	0.9192
0.21	0.5832	0.81	0.7910	1.41	0.9207
0.22	0.5871	0.82	0.7939	1.42	0.9222
0.23	0.5910	0.83	0.7967	1.43	0.9236
0.24	0.5948	0.84	0.7995	1.44	0.9251
0.25	0.5987	0.85	0.8023	1.45	0.9265

Phân phối chuẩn

Chuẩn tắc hóa

- Ta đã biết nếu $X \sim N(\mu, \sigma^2)$ thì bằng cách đặt $Z = \frac{X - \mu}{\sigma}$ ta chuẩn tắc hóa được X thành Z

$$Z \sim N(0, 1)$$

- Khi đó với mọi số thực a ta có:

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

- Tương tự với mọi $a \leq b$ ta có:

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

- Lưu ý, vì X là b.n.n liên tục nên $P(X = a) = 0, P(X < a) = P(X \leq a)$

Phân phối chuẩn

Ví dụ

- Xét thí nghiệm bắt cá trong hồ, gọi X là chiều dài cá bắt được. Giả sử X có phân phối chuẩn với trung bình $\mu = 16\text{cm}$ và độ lệch chuẩn $\sigma = 4\text{cm}$.

Chuẩn tắc hóa $Z = \frac{X - \mu}{\sigma} = \frac{X - 16}{4}$

- Xác suất bắt được cá nhỏ hơn 10cm

$$P(X < 10) = P\left(Z < \frac{10 - 16}{4}\right) = \Phi(-1.5) = 1 - \Phi(1.5) = 0.0668$$

- Xác suất bắt được cá lớn hơn 24cm

$$P(X > 24) = 1 - P(X \leq 24) = 1 - P\left(Z \leq \frac{24 - 16}{4}\right) = 1 - \Phi(2) = 0.0228$$

- 10% cá nhỏ nhất trong hồ có chiều dài $\leq a$ là bao nhiêu?

$$P(X \leq a) = P\left(Z \leq \frac{a - 16}{4}\right) = 0.1$$

$$\Rightarrow \frac{a - 16}{4} = \Phi^{-1}(0.1) = -\Phi^{-1}(0.9) = -1.2816 \Rightarrow a = 10.87\text{cm}$$

Phân phối mũ

- B.n.n liên tục X được gọi là có **phân phối mũ** (exponential distribution) với tham số λ ($\lambda > 0$), kí hiệu $X \sim \text{Exponential}(\lambda)$ nếu X có hàm mật độ xác suất:

$$f(x; \lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

- Khi đó, X có kì vọng và phương sai

$$E(X) = \frac{1}{\lambda} \text{ và } \text{Var}(X) = \frac{1}{\lambda^2}$$

- X có hàm phân phối tích lũy

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(u; \lambda) du = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Phân phối mũ

- Đặt $N(t)$ là số lượng xảy ra một sự kiện quan tâm nào đó trong khoảng thời gian từ thời điểm 0 đến thời điểm t ($t \geq 0$). Ta nói $\{N(t), t \geq 0\}$ là **quá trình Poisson** (Poisson process) với tỉ lệ λ trên một đơn vị thời gian nếu nó thỏa mãn 2 tính chất:
 - Số lượng sự kiện xảy ra trong mọi khoảng thời gian độ dài Δt có phân phối Poisson với kì vọng $\lambda \Delta t$
 - Số lượng sự kiện xảy ra trong các khoảng thời gian rời nhau là độc lập nhau
- Nếu $\{N(t), t \geq 0\}$ là quá trình Poisson với tỉ lệ λ thì khoảng thời gian giữa các lần xảy ra sự kiện có phân phối mũ với tham số λ

Phân phối mũ

Ví dụ

- Giả sử lượng khách đến một tiệm là quá trình Poisson với tỉ lệ $\lambda = 4.5$ khách trong một giờ
- Đặt X là lượng khách đến trong một giờ thì $X \sim P(4.5)$
 - Xác suất có nhiều hơn 2 khách trong một giờ là:
$$P(X > 2) = 1 - P(X \leq 2) = 1 - P(X = 0) - P(X = 1) - P(X = 2)$$
$$= 1 - e^{-4.5} \frac{4.5^0}{0!} - e^{-4.5} \frac{4.5^1}{1!} - e^{-4.5} \frac{4.5^2}{2!} = 0.9271$$
 - Lượng khách trung bình trong một giờ là: $E(X) = 4.5$ khách/giờ
- Đặt Y là thời gian chờ (giữa hai lần khách đến) thì $Y \sim \text{Exponential}(4.5)$
 - Xác suất chờ không quá 15 phút (1/4 giờ) là:
$$P(Y \leq 0.25) = 1 - e^{-4.5 \times 0.25} = 0.6753$$
 - Thời gian chờ trung bình là: $E(Y) = \frac{1}{4.5} = 0.2222$ giờ = 13 phút