Estatística aplicada

Lino Costa

Departamento de Produção e Sistemas
Escola de Engenharia
lac@dps.uminho.pt

Ano letivo 2015/2016

Sumário

Famílias de distribuições de probabilidade

1. distribuições de probabilidade discretas
   - distribuição uniforme
   - distribuição de Bernoulli
   - distribuição binomial
   - distribuição de Poisson
   - aproximação da Poisson à binomial
2. distribuições de probabilidade contínuas
   - distribuição uniforme
   - distribuição exponencial
   - distribuição normal
   - aproximação da normal à binomial

Distribuição uniforme

Na distribuição uniforme de uma variável aleatória discreta X existe um probabilidade igual para a ocorrência de cada uma dos valores da variável.

Logo, uma variável aleatória X segue uma distribuição uniforme discreta se e só se a sua distribuição de probabilidade é dada por

$$f(x) = \frac{1}{k} \quad \text{onde} \quad x = x_1, x_2, \ldots, x_k (\text{com } x_i \neq x_j \text{ para } i \neq j)$$

A média e a variância de X são

$$\mu = \frac{1}{k}\sum_{i=1}^{k} x_i \quad \sigma^2 = \frac{1}{k}\sum_{i=1}^{k}(x_i - \mu)^2$$

Distribuição uniforme

Exemplo 1

Considere o jogo da roleta e X a variável aleatória que representa o número saído numa jogada. Qual a distribuição de probabilidade da variável X? Calcule a probabilidade de se obter o número 13?

- todos os 38 números $(1, \ldots, 36, 0, 00)$ têm igual a probabilidade de sair
- a variável X segue uma distribuição uniforme com $k = 38$
- $f(x) = \frac{1}{38}$ para $x = x_1, x_2, \ldots, x_{38}$
- $P(X = x_{13}) = \frac{1}{38}$

Exemplo 2

Qual deverá ser a distribuição para gerar uma sequência de dígitos aleatórios ?

- todos os 10 dígitos $(0, 1, 2, 3, 4, 5, 6, 7, 8, 9)$ devem ter igual probabilidade de ocorrer
- a variável X que representa o dígito gerado deve seguir uma distribuição uniforme com $k = 10$
- $f(x) = \frac{1}{10}$ para $x = x_1, x_2, \ldots, x_{10}$

Distribuição de Bernoulli

Uma tentativa de Bernoulli é uma experiência aleatória que tem apenas dois resultados possíveis: sucesso (com probabilidade p) ou insucesso (com probabilidade $1-p$).

Logo, uma variável aleatória X segue uma distribuição de Bernoulli se e só se a sua distribuição de probabilidade é dada por

$$f(x) = \begin{cases} 1-p & x=0 \\ p & x=1 \end{cases}$$

A média e a variância de X são

$$\mu = p \quad \sigma^2 = p(1-p)$$

Exemplo 3

A variável aleatória X que representa o resultado do lançamento de uma moeda equilibrada segue uma distribuição de Bernoulli com $p = 0.5$.

Distribuição binomial

Uma experiência binomial é uma experiência aleatória que consiste em realizar n tentativas que satisfazem as seguintes condições:

- as tentativas são independentes, i.e., o resultado de uma tentativa não afeta o resultado das outras
- cada tentativa tem apenas dois resultados mutuamente exclusivos: sucesso e insucesso
- a probabilidade de sucesso p em cada tentativa é constante

Uma experiência binomial consiste numa série de n tentativas de Bernoulli com uma probabilidade de sucesso constante p em cada tentativa.

Distribuição binomial

Uma variável aleatória binomial X representa o número de tentativas em que o resultado foi sucesso em n tentativas com uma probabilidade de sucesso de p.

Logo, uma variável aleatória X segue uma distribuição binomial se e só se a sua distribuição de probabilidade é dada por

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x = 0, 1, \ldots, n$$

onde $\binom{n}{x} = C_x^n = \dfrac{n!}{x!(n-x)!}$.
A média e a variância de $X \sim B(n, p)$ são

$$\mu = np \quad \sigma^2 = np(1-p)$$

Distribuição binomial

Exemplo 4

Assuma que a proporção de fumadores na população geral é de 30%. Suponha que duas pessoas são selecionadas aleatoriamente da população. Quais são os resultados possíveis da variável aleatória X, o número de pessoas que fumam? Quais as respetivas probabilidades?

- Seja X a variável que representa o número de pessoas que fumam nas duas pessoas selecionadas
-

Pessoa 1	Pessoa 2	X	probabilidade
NF	NF	0	$(1-p)(1-p) = 0.49$
NF	F	1	$(1-p)p = 0.21$
F	NF	1	$p(1-p) = 0.21$
F	F	2	$pp = 0.09$

-

X	$f(x)$
0	$\binom{2}{0}p^0(1-p)^{2-0} = 0.49$
1	$\binom{2}{1}p^1(1-p)^{2-1} = 0.42$
2	$\binom{2}{2}p^2(1-p)^{2-2} = 0.09$

- $X \sim B(n, p)$ com $n = 2$ e $p = 0.3$

Distribuição binomial

Exemplo 5

Um dado é lançado 10 vezes. Qual a probabilidade de obter exatamente 3 senas? E pelo menos 5 senas?

- Seja X a variável que representa o número de senas que se obtém nos 10 lançamentos do dado.
- $X \sim B(n, p)$ com $n = 10$ e $p = 1/6$
- $P(X = 3) = f(3) = 0.155$
- $P(X \geq 5) = 1 - P(X < 5) = 1 - P(X \leq 4) = 1 - F(4) = 1 - 0.985 = 0.015$

Exemplo 6

Uma moeda é lançada 4 vezes. Qual a probabilidade de obter exatamente uma cara? E de obter no máximo duas caras? E de obter pelo menos 3 caras?

- Seja X a variável que representa o número de caras que se obtém nos 4 lançamentos da moeda
- $X \sim B(n, p)$ com $n = 4$ e $p = 0.5$
- $P(X = 1) = f(1) = 0.25$ (Tabela 1)
- $P(X \leq 2) = F(2) = 0.6875$ (Tabela 3)
- $P(X \geq 3) = 1 - P(X < 3) = 1 - P(X \leq 2) = 1 - F(2) = 1 - 0.6875 = 0.3125$ (Tabela 3)

Distribuição de Poisson

Suponha que a ocorrência de um acontecimento num intervalo (de tempo, comprimentos, área, espaço, etc) é contável e o intervalo pode ser dividido em subintervalos.
Uma experiência é uma experiência aleatória de Poisson se:

- a probabilidade do acontecimento ocorrer mais do que uma vez num subintervalo é infinitesimal (próxima de zero)
- as ocorrências de acontecimentos em subintervalos distintos são independentes
- a probabilidade de um ocorrência de um acontecimento num subintervalo é a mesma em todos os subintervalos e proporcional ao comprimento do subintervalo

Distribuição de Poisson

Uma variável aleatória de Poisson X representa o número de ocorrências de um acontecimento num intervalo (de tempo, espaço, etc). A variável X pode assumir qualquer valor entre zero e infinito e λ representa o número médio de ocorrências do acontecimento num intervalo.

Logo, uma variável aleatória X segue uma distribuição de Poisson se e só se a sua distribuição de probabilidade é dada por

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

A média e a variância de $X \sim P(\lambda)$ são

$$\mu = \lambda \quad \sigma^2 = \lambda$$

Distribuição de Poisson

Exemplo 7

Em média chegam 5 clientes cada 10 minutos a um restaurante. Assumindo que o número de clientes que chega ao restaurante segue uma distribuição de Poisson. Calcule a probabilidade de chegarem exatamente 3 clientes em 10 minutos e a probabilidade de chegarem mais de 5 clientes em 10 minutos. Qual seria a probabilidade de chegarem mais de 10 clientes em meia hora?

- Seja X a variável que representa o número de clientes que chegam cada 10 minutos
- $X \sim P(\lambda)$ com $\lambda = 5$ (5 clientes/10 minutos)
- $P(X = 3) = f(3) = 0.1404$ (Tabela 2)
- $P(X > 5) = 1 - P(X \leq 5) = 1 - F(5) = 1 - 0.616 = 0.384$ (Tabela 4)
- Seja X' a variável que representa o número de clientes que chegam cada 30 minutos
- $X' \sim P(\lambda')$ com $\lambda' = 15$ (15 clientes/30 minutos)
- $P(X' > 10) = 1 - P(X' \leq 10) = 1 - F(10) = 1 - 0.1185 = 0.8815$ (Tabela 4)

Distribuição de Poisson

Uma experiência de Poisson pode ser vista como uma experiência binomial com um número de tentativas infinito.

Aproximação da Binomial pela Poisson

Suponha uma variável aleatória binomial X representa o número de tentativas em que o resultado foi sucesso em n tentativas com uma probabilidade de sucesso de p.
Se n for grande ($n \to \infty$) e p muito pequeno ($p \to 0$) pode-se aproximar a distribuição binomial ($X \sim B(n, p)$) pela distribuição de Poisson ($X \sim P(\lambda)$) fazendo $\lambda = np$.

Aproximação da Binomial pela Poisson

Exemplo 8

Seja X uma variável aleatória que representa o número de pessoas portadoras da doença de Gaucher na população portuguesa. A probabilidade de um indivíduo ser portador é 0.00042. Suponha que pretende saber, numa determinada região com 10000 indivíduos, qual a probabilidade de haver pelo menos 6 indivíduos portadores da doença de Gaucher.

- Seja X a variável que representa o número de indivíduos em Portugal portadores da doença de Gaucher
- $X \sim B(n, p)$ com $n = 10000$ e $p = 0.00042$
- como n é grande e p muito pequeno vai-se aproximar a binomial pela Poisson
- $X \sim P(\lambda)$ com $\lambda = 10000 \times 0.00042 = 4.2$
- $P(X \geq 6) = 1 - P(X < 6) = 1 - P(X \leq 5) = 1 - F(5) = 1 - 0.7531 = 0.2469$ (Tabela 4)

Distribuição uniforme

Uma variável aleatória uniforme contínua X tem uma função densidade de probabilidade constante na gama de valores da variável X.

Logo, uma variável aleatória X segue uma distribuição uniforme contínua se e só se a sua função densidade de probabilidade é dada por

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha} & \alpha \leq x \leq \beta \\ 0 & \text{caso contrário} \end{cases}$$

A média e a variância de X são

$$\mu = \frac{\beta + \alpha}{2} \quad \sigma^2 = \frac{(\beta - \alpha)^2}{12}$$

Distribuição uniforme

Exemplo 9

Qual deverá ser a distribuição para gerar um número real entre 0 e 1?

- a densidade de probabilidade na gama de valores de X entre 0 e 1 deve ser constante
- a variável X segue uma distribuição uniforme com $\alpha = 0$ e $\beta = 1$, i.e., $X \sim U(0, 1)$
- $f(x) = \begin{cases} \frac{1}{1-0} = 1 & 0 \leq x \leq 1 \\ 0 & \text{caso contrário} \end{cases}$

Exemplo 2

Uma fábrica produz folhas de cartão com uma espessura uniforme entre 0.8 e 1.2 cm. Qual a percentagem de folhas abaixo de 1 cm?

- a variável X segue uma distribuição uniforme com $\alpha = 0.8$ e $\beta = 1.2$, i.e., $X \sim U(0.8, 1.2)$
- $P(X < 1) = \int_{0.8}^{1} \frac{1}{1.2-0.8} dx = [\frac{1}{0.4} x]_{0.8}^{1} = \frac{0.2}{0.4} = 0.5$

Distribuição exponencial

Uma variável contínua exponencial representa o comprimento de um intervalo (em tempo, espaço,...) de um certo ponto até que ocorra o próximo sucesso numa experiência aleatória de Poisson.

Logo, uma variável aleatória contínua X segue uma distribuição exponencial com parâmetro θ se e só se a sua função densidade de probabilidade é dada por

$$f(x) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}} & x \geq 0, \theta \geq 0 \\ 0 & \text{caso contrário} \end{cases}$$

A média e a variância de X são

$$\mu = \theta \quad \sigma^2 = \theta^2$$

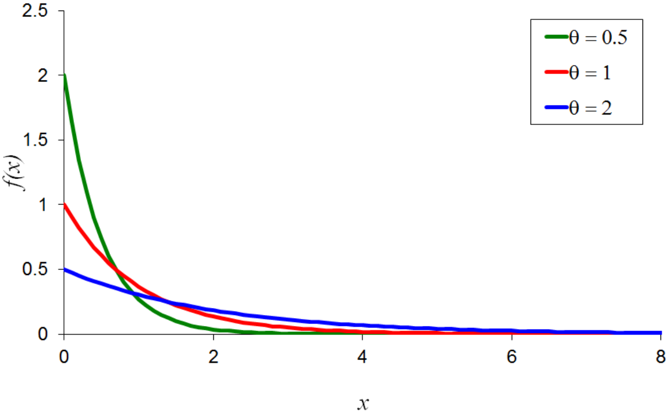
A função de distribuição acumulada é dada por

$$F(x) = 1 - e^{-\frac{x}{\theta}} \quad x \geq 0, \theta \geq 0$$

Nota: $\theta = \frac{1}{\lambda}$, i.e., θ é o inverso do número médio de sucessos por intervalo λ de uma distribuição de Poisson.

Distribuição exponencial

A curva da distribuição exponencial é assimétrica à direita.
Quanto maior o valor de θ, menor é a assimetria à direita.



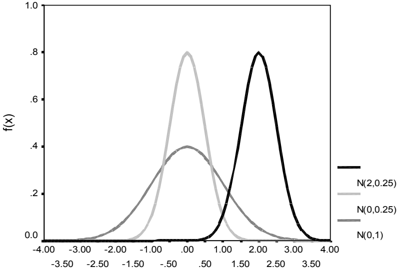
Distribuição exponencial

Exemplo 10

Um componente eletrónico requer, em média, uma reparação de 2 em 2 anos. Qual a probabilidade de que funcione por pelo menos 3 anos? Sabendo que o componente dura há já dois anos, qual a probabilidade de funcionar durante pelo menos mais um ano?

- a variável X segue uma distribuição exponencial com $\theta = 2$, i.e., $X \sim EN(1/2)$
- $F(x) = 1 - e^{-\frac{x}{2}} \quad x \geq 0$
- $P(X > 3) = 1 - P(X < 3) = 1 - \int_0^3 \frac{1}{2} e^{-\frac{x}{2}} dx = 1 - [-e^{-\frac{x}{2}}]_0^3 = e^{-\frac{3}{2}} = 0.2231$
- $P(X > 3 | X > 2) = \frac{P(X>3 \cap X>2)}{P(X>2)} = \frac{P(X>3)}{P(X>2)} = \frac{e^{-\frac{3}{2}}}{e^{-\frac{2}{2}}} = e^{-\frac{1}{2}} = 0.6065$

Distribuição normal



- uma distribuição normal com média μ e variância σ^2 é simétrica em relação a μ e em forma de sino
- a simetria da distribuição implica que $P(\mu < X) = P(X > \mu) = 0.5$
- os parâmetros μ e σ^2 determinam o centro a forma da curva normal:
  - quanto maior o valor de μ, mais à direita está o centro
  - maiores valores de σ^2 correspondem a uma curva normal mais “achatada”

Distribuição normal

Uma variável aleatória normal X com média μ e variância σ^2, $X \sim N(\mu, \sigma^2)$, tem a seguinte função densidade de probabilidade

$$f\left(x;\mu,\sigma\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad \begin{array}{l} -\infty < x < \infty \\ \sigma > 0 \end{array}$$

Variável normal padrão (ou standard)

Uma variável aleatória normal padrão Z é uma variável normal com média $\mu = 0$ e variância $\sigma^2 = 1$.

Padronização (standardização)

Uma qualquer variável aleatória normal X com média μ e variância σ^2 pode ser transformada numa variável Z normal padrão, utilizando a seguinte relação

$$Z = \frac{X-\mu}{\sigma} \sim N(0,1)$$

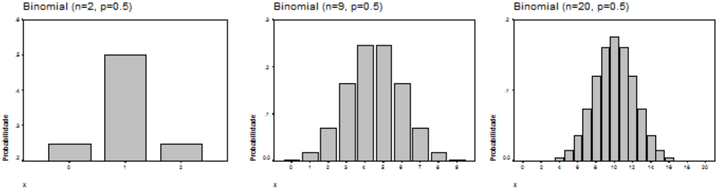
Distribuição normal

Exemplo 11

As classificações de um exame de admissão a um colégio seguem uma distribuição normal de média 500 e desvio padrão 100. Determine a probabilidade de um estudante ter classificação: i) superior a 650; ii) inferior a 250; iii) entre 325 e 675.

- Seja X a variável que representa as classificações do exame de admissão ao colégio
- $X \sim N(\mu, \sigma^2)$ com $\mu = 500$ e $\sigma^2 = 100^2$
- $Z = \dfrac{X - \mu}{\sigma} = \dfrac{X - 500}{100} \sim N(0, 1)$
- $P\,(X > 650) = 1 - P(X \leq 650) = 1 - P(Z \leq \frac{650-500}{100})$
  $= 1 - P(Z \leq 1.5) = 1 - 0.9332 = 0.0668$ (tabela 5)
- $P\,(X < 250) = P(Z < \frac{250-500}{100}) = P(Z < -2.5) = 0.0062$ (tabela 5)
- $P\,(325 < X < 675) = P(X < 675) - P(X \leq 325)$
  $= P(X < \frac{675-500}{100}) - P(X \leq \frac{325-500}{100})$
  $= P(Z < 1.75) - P(Z \leq -1.75) = 0.9599 - 0.0401 = 0.9198$ (tabela 5)

Distribuição normal



Aproximação da binomial pela normal

Uma variável aleatória binomial $X \sim B(n, p)$ pode ser aproximada pela distribuição normal com $\mu = np$ e $\sigma^2 = np(1-p)$ se $np > 5$ e $n(1-p) > 5$.

Correção de Yates (continuidade)

Para aproximar a distribuição discreta $X \sim B(n, p)$ pela distribuição contínua $X' \sim N(\mu, \sigma^2)$ com $\mu = np$ e $\sigma^2 = np(1-p)$ deve introduzir-se uma correção de continuidade de acordo com

$$P(X \leq x) \approx P\left(X' < x + 0.5\right) \text{ ou } P(X \geq x) \approx P\left(X' > x - 0.5\right)$$

Aproximação da Binomial pela Normal

Exemplo 12

Sabe-se que 30% dos estudantes de uma determinada universidade frequentaram colégios particulares. Considere uma amostra aleatória de 50 estudantes.
i) Qual a probabilidade de exactamente 10 dos estudantes seleccionados terem frequentado um colégio particular?
ii) Qual a probabilidade de 20 ou mais dos estudantes seleccionados terem frequentado um colégio particular?
iii) Qual a probabilidade de o número de estudantes provenientes de colégios particulares estar entre 10 e 20 inclusive?

- Seja X a variável que representa o número de estudantes da universidade que frequentaram colégios particulares
- $X \sim B(n, p)$ com $n = 50$ e $p = 0.3$
- $np = 50 \times 0.3 = 15 > 5$ e $n(1-p) = 50 \times (1-0.3) = 35 > 5$
- a aproximação de $X \sim B(n, p)$ por $X' \sim N(\mu, \sigma^2)$ com $\mu = np = 15$ e $\sigma^2 = np(1-p) = 10.5$ é adequada

Aproximação da Binomial pela Normal

Exemplo 13

- $X' \sim N(15, 10.5)$ e $Z = \dfrac{X' - 15}{\sqrt{10.5}} \sim N(0,1)$
- $P(X = 10) = P(X \leq 10) - P(X < 10) = P(X \leq 10) - P(X \leq 9)$
  $\approx P(X' \leq 10.5) - P(X' \leq 9.5) = P(Z \leq -1.39) - P(Z \leq -1.70)$
  $= 0.0823 - 0.0446 = 0.0377$ (tabela 5)
- $P(X \geq 20) = 1 - P(X < 20) = 1 - P(X \leq 19) \approx 1 - P(X' \leq 19.5)$
  $= 1 - P(Z \leq 1.39) = 1 - 0.9177 = 0.0823$ (tabela 5)
- $P(10 < X \leq 20) = P(X \leq 20) - P(X \leq 10)$
  $\approx P(X' \leq 20.5) - P(X' \leq 10.5) = P(Z \leq 1.70) - P(Z \leq -1.39)$
  $= 0.9544 - 0.0823 = 0.8721$ (tabela 5)