Variáveis aleatórias Contínuas

Variáveis aletórias CONTÍNUAS:

suporte contínuo

A probabilidade pontual não faz sentido! $P[X = x]$ = 0 sempre

Função densidade de probabilidade, $f(x)$

"taxa instantânea de acumulação de probabilidade"

Propriedades de $f(x)$

Função não negativa	$f(x) \geq 0$
O integral em todo o domínio é 1	$\int_{-\infty}^{+\infty} f(x)\,dx = 1$

Variáveis contínuas – Função de distribuição

A f.d.p. é a derivada da função distribuição

$$F(x) = P[X \le x] = \int_{-\infty}^{x} f(u)\,du$$

$$P[x_1 \le X \le x_2] = P[x_1 < X \le x_2] = \int_{x_1}^{x_2} f(u)\,du$$

$$E[X] = \int_{-\infty}^{+\infty} x\, f(x)\,dx$$

$$Var[X] = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x)\,dx$$

Variáveis contínuas – Parâmetros

Valor Esperado de X $E[X] = \int_{-\infty}^{+\infty} x f(x) dx$

Variância de X $Var[X] = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx$

Momentos não centrados de ordem k $E[X^k] = \int_{-\infty}^{+\infty} x^k f(x) dx$

Momentos centrados de ordem k

$$E[(X - \mu)^k] = \int_{-\infty}^{+\infty} (x - \mu)^k f(x) dx$$

Tem-se em geral $E[g(X)] = \int_{-\infty}^{+\infty} g(x) f(x) dx$

Variáveis contínuas – exercício em R

A procura diária em centenas de horas dos serviços prestados por uma empresa de limpeza é uma v.a. com a seguinte função densidade de probabilidade

$$f(x) = \begin{cases} 2x - 1 & 0{,}5 \leq x \leq 1{,}5 \\ 0 & outros\ valores \end{cases}$$

a) Crie o gráfico da função densidade de probabilidade em causa.

b) Qual a probabilidade da procura exceder 100 horas?

c) Calcule a receita diária esperada, sabendo que a empresa cobra 10€ por hora de serviço prestado

Distribuição uniforme contínua, $X \sim U(a, b)$

- A densidade de probabilidade é constante ao longo de um intervalo

$$f(x) = \frac{1}{b-a} \quad , a \leq x \leq b$$

- A função de distribuição é dada por

$$F(x) = \begin{cases} 0 & x < a \\ \dfrac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$$

$$E[X] = \frac{a+b}{2}$$

$$Var[X] = \frac{(b-a)^2}{12}$$

Funções em R associadas a variáveis contínuas

De forma análoga ao caso das variáveis discretas (ver slide 7 do EstComp-VarAleat-Parte2.pptx):

runif (,,,) → geração de observações aleatórias com base no tipo de variável aleatória especificada.

dunif (,,) → função densidade de probabilidade (f.d.p.) do tipo de variável aleatória especificada.

punif (,,,) → função DISTRIBUIÇÃO i.e. P(X<=x)

qunif(,,,) → "Inversa" da Função Distribuição: dada uma probabilidade, determina o (primeiro) x em que esse acumulado é atingido (quantil)

Tal como nas variáveis discretas, basta mudar o nome, por exemplo unif para norm, para mudarmos o tipo de variável aleatória em causa…

Uniforme contínua – exercício em R

O verdadeiro peso em mg por comprimido de certa marca de analgésicos é uma variável aleatória com função densidade de probabilidade dada por

$$f(x) = \begin{cases} 10 & 9{,}95 \leq x \leq 10{,}05 \\ 0 & outros\ valores \end{cases}$$

1) Represente graficamente a f.d.p. e a Função de Distribuição

2) Qual a probabilidade de um determinado comprimido pesar mais de 9,98mg

3) Calcule uma aproximação à probabilidade anterior, tendo por base uma simulação (amostra) de 10 000 comprimidos.

Distribuição Normal

- É uma das distribuições mais utilizadas (talvez mesmo a mais utilizada)
- Caracteriza-se pela sua função densidade apresentar uma forma em sino, simétrica em torno da respectiva média
- Esta curva é conhecida pela *curva de Gauss*, sendo a distribuição Normal conhecida também como a Gaussiana
- Muitas características físicas se distribuem desta forma (pesos, alturas, etc) – daqui resultam até as tabelas de percentis utilizadas nos boletins médicos infantis...
- Também descreve bem a distribuição de erros de medição...

Distribuição Normal

- Esta distribuição depende de dois parâmetros caracterizadores: a média, μ, ($\mu \in |R$) e a variância, σ^2.
- A sua função densidade é

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

- Tem a particularidade de não ser possível determinar explicitamente a sua primitiva. Assim, não dispomos de uma expressão algébrica para a Função Distribuição

Distribuição NORMAL

- Sendo X uma v.a com distribuição Normal, qualquer sua transformação linear, aX+b, tem ainda distribuição Normal.

Assim, se $X \sim N(\mu, \sigma)$ e $Y = aX + b$, onde a e b são duas constantes reais, sendo *a* não nula, então,

$$Y \sim N(a\mu+b; a\sigma)$$

já que

$$E[Y] = E[aX+b] = a\,E[X] + b = a\mu+b$$

e

$$Var[Y] = Var[aX+b] = a^2\,Var[X] = a^2\,\sigma^2$$

- Em particular (processo de estandardização)

Se $X \sim N(\mu, \sigma)$ então, $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$. (z-score)

Normal – exercício em R – Número 1

Uma certa função pode ser desempenhada com um de dois dispositivos, D1 ou D2, indiferentemente. A duração em horas de D1 e de D2 pode em ambos os casos ser modelizada por uma distribuição Normal, de média 43h e desvio-padrão 6h, para o primeiro, sendo os respetivos valores para o segundo, 45h e 3h.

1) Se a função tiver de ser executada por um período superior a 48 horas, qual dos dois dispositivos aconselharia usar?

2) De seguida, crie uma amostra de 10000 dispositivos, para as distribuições em causa, representando-as graficamente (histograma). No mesmo gráfico, inclua as f.d.p. teóricas para as duas distribuições em causa. Compare as durações simuladas, com as durações da f.d.p. criadas.

3) Estime a probabilidade pedida em 1) usando uma simulação de 100 000 dispositivos.

Normal – exercício em R – Número 2

Uma certa universidade escolhe os seus alunos através da realização de uma prova. O número de candidatos é usualmente muito elevado. Do histórico sabe-se que as pontuações obtidas nas provas podem ser modelizadas por uma distribuição Normal, de média 55 pontos e variância 25 pontos^2.

Em certo ano houve 3500 candidatos para 700 vagas.

1) Qual terá sido a nota de corte? Represente graficamente a zona da f.d.p. em causa.

2) Quantos candidatos deverão ter tido pontuação inferior a 50 pontos?

3) Quais deverão ter sido as pontuações extremas do grupo intermédio constituído por 50% dos candidatos?

4) Gere 3500 observações de uma normal como a especificada e compare os valores teóricos dos quartis (obtidos na alínea 3) com os obtidos empiricamente.

5) Calcule o Z-score da alínea 2). Represente graficamente todos os z-score da amostra da alínea 4 e os z-score teóricos da distribuição em causa.

Outras distribuições importantes para inferência

t-Student

F-Snedecor

Qui-Quadrado