Estatística Aplicada

ano lectivo 2010/2011

A. Rita Gaio
Departamento de Matemática - FCUP
argaio@fc.up.pt

February 14, 2011

Índice Geral

Chapter 1

Variáveis aleatórias e distribuições de probabilidade

1.1. Noção de probabilidade

Experiências aleatórias: os resultados obtidos não são sempre (essencialmente) os mesmos ainda que as condições de realização das experiências se mantenham praticamente as mesmas.

Exemplos:

(a) lançar uma moeda ao ar e registar a face que fica voltada para cima

(b) lançar um dado e registar a face que fica voltada para cima

(c) seleccionar aleatoriamente uma determinada região e várias famílias dessa região com dois filhos, e registar a ordem do sexo dos filhos dessas famílias

(d) seleccionar aleatoriamente um grupo de alunos inscritos em 2009/2010 na UP e registar o peso (em Kg) desses alunos.

Espaço de probabilidade $(\Omega, \mathcal{A}, P)$

- $\Omega =$ **espaço amostral**
  $=$ conjunto de todos os resultados de uma experiência aleatória

  (a) $\Omega = \{H, T\}$ H *head*, T *toss*

  (b) $\Omega = \{1, 2, 3, 4, 5, 6\}$

  (c) $\Omega = \{♂♂,\ ♂♀,\ ♀♀,\ ♀♂\}$

  (d) $\Omega = [40, 150]$

- $\mathcal{A} =$ conjunto de partes de Ω
  $=$ conjunto dos **acontecimentos** [a] mensuráveis,

$$\emptyset \in \mathcal{A}, \qquad A \in \mathcal{A} \Longrightarrow \overline{A} = \Omega \setminus A \in \mathcal{A}$$

$$A_n \in \mathcal{A}, \quad n = 1, 2, \ldots \Longrightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$$

  (a) $\mathcal{A} = \{\{H\}, \{T\}, \{H, T\}\}$

  (b) $\mathcal{A} = \{\{1\}, \{2\}, \{3\}, \cdots, \{1, 2\}, \cdots, \{1, 2, 3, 4, 5, 6\}\}$

  (c) $\mathcal{A} = \{\{♂♂\}, \{♂♀\}, \cdots, \{♂♂,\ ♂♀\}, \{♂♂,\ ♀♂\}, \cdots$
  $\cdots, \{♂♂, ♂♀, ♀♀\}, \cdots, \{♂♂,\ ♂♀,\ ♀♀,\ ♀♂\}\}$

  Diz-se que:

  – $\emptyset$ é o **acontecimento impossível**

  – Ω é o **acontecimento certo**

  – um acontecimento que consiste de um único ponto é um **acontecimento elementar**

  – $A, B \subset \mathcal{A}$ são acontecimentos **mutuamente exclusivos** [b] se $A \cap B = \emptyset$

- $P : \mathcal{A} \to \mathbb{R}$ **função de probabilidade**

  $P(A) =$ probabilidade de se realizar o acontecimento A

$$P(A) \geq 0$$

$$P(\Omega) = 1$$

$$A_n \cap A_m = \emptyset (n \neq m) \Longrightarrow P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$$

[a] Um **acontecimento** é então um qualquer subconjunto de Ω, i.e., um qualquer conjunto de resultados possíveis de uma experiência aleatória

[b] Acontecimentos mutuamente exclusivos são também designados como **incompatíveis** ou **disjuntos**

Propriedades: Sejam A e B dois acontecimentos.

(a) $A \subseteq B \Longrightarrow P(A) \leq P(B)$,
$P(B \setminus A) = P(B) - P(A)$

(b) $0 \leq P(A) \leq 1$

(c) $P(\overline{A}) = P(\Omega \setminus A) = 1 - P(A)$

(d) $P(\emptyset) = 0$

(e) $A_1,\ A_2,\ \ldots,\ A_n$ acontecimentos mutuamente exclusivos e $A = A_1 \cup A_2 \cup \cdots \cup A_n \Longrightarrow$ $P(A) = P(A_1) + P(A_2) + \cdots + P(A_n)$

(f) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

(g) $B_1, B_2, \ldots, B_n$ acontecimentos mutuamente exclusivos e $\Omega = B_1 \cup B_2 \cup \cdots \cup B_n \Longrightarrow$ $P(A) = P(A \cap \Omega) = \sum_{i=1}^{n} P(A \cap B_i)$.

Questão:
que probabilidade atribuir a cada acontecimento?

Se o espaço amostral consiste apenas dos acontecimentos elementares $A_1,\ A_2,\ \ldots,\ A_n$ então

$$P(A_1) + P(A_2) + \cdots + P(A_n) = 1.$$

Se admitirmos probabilidades iguais para todos os acontecimentos elementares, tem-se

$$P(A_i) = \frac{1}{n}, \qquad i = 1, 2, \ldots, n.$$

Em particular, se A é um acontecimento formado por k acontecimentos elementares então

$$P(A) = k/n.$$

Exemplo 1: No lançamento de um dado não viciado,

(a) a probabilidade de sair a face i, é

$$P(i) = 1/6 \qquad (i = 1, 2, \ldots, 6)$$

dado que qualquer face tem a mesma probabilidade de sair voltada para cima (o dado é não viciado) e existem seis faces possíveis.

(b) a probabilidade de sair 2 ou 5 é

$$P(2 \cup 5) = P(2) + P(5) = 1/3.$$

dado que a saída de 2 e de 5 são acontecimentos mutuamente exclusivos.

(c) a probabilidade de não sair 6 é
$P(\text{não sair } 6) = 1 - P(6) = 1 - 1/6 = 5/6.$

(d) a probabilidade de não sair 1 nem 2 é

$$\begin{aligned} P(\text{não sair 1 nem 2}) &= 1 - P(1 \cup 2) \\ &= 1 - (P(1) + P(2)) \\ &= 1 - 1/3 \\ &= 2/3 \end{aligned}$$

pois o acontecimento que se pretende é o complementar do acontecimento "sair 1 ou sair 2".

Se após n experiências se observarem k ocorrências de um acontecimento A então a **probabilidade empírica** de A é

$$\widehat{P}(A) = \frac{k}{n}.$$

Exemplo 2: se em 100 jogadas de uma moeda, a face "cara" sair 47 vezes, então a probabilidade empírica do acontecimento "sair cara" é

$$\widehat{P}(\text{"sair cara"}) = \frac{47}{100} = 0.47$$

Mesmo que a moeda seja não viciada, esta probabilidade não coincide exactamente com a *probabilidade teórica* de 0.5 de sair cara!

Exemplo 3: Suponhamos que estudos realizados sobre o teste da tuberculina numa determinada região afirmam que a probabilidade deste teste ser positivo é de 0.1. Isto significa que:

- para um número **grande** de pessoas testadas, se espera uma percentagem aproximada de 10% de indivíduos com resultado positivo
- à medida que se aumentar o tamanho da amostra mais o resultado se aproximará dos 10%.
- o resultado será exactamente 10% se se conseguirem avaliar todos os indivíduos dessa região.

Exemplo 4: Um estudo afirma que a probabilidade de aparecimento de cancro da mama num período de 30 anos em mulheres com idade superior a 40 anos e que nunca tiveram esse tipo de cancro é de 1/10. Isto significa que:

- se escolhermos *à sorte* uma mulher com idade superior a 40 anos e que nunca teve cancro da mama, ela tem uma probabilidade de 0.1 de vir a desenvolver esse tipo de cancro durante os próximos 30 anos.
- se considerarmos uma amostra *aleatória* constituída por *muitas* mulheres com idade superior a 40 anos e que nunca tiveram cancro da mama, é de esperar que durante os próximos 30 anos aproximadamente 10% dessas mulheres venham a sofrer de cancro da mama. Quanto maior for a amostra considerada, melhor será a aproximação do valor em causa aos 10%.

1.2. Probabilidade condicional

Sejam A e B dois acontecimentos de um espaço de probabilidade $(\Omega, \mathcal{A}, P)$ e suponhamos que se tem $P(A) > 0$.

A **probabilidade condicional** $P(B|A)$ é a probabilidade de se realizar o acontecimento B na hipótese de A ter ocorrido. Dado que o novo espaço amostral passa a ser A, define-se

$$P(B|A) = \frac{P(B \cap A)}{P(A)}{}^{a}.$$

Exemplo 1:

Determinar a probabilidade de, no lançamento de um dado, sair um número menor do que 5 sabendo que o resultado é um número par.

Sejam A o acontecimento "número par" e B o acontecimento "menor do que 5". Pretende-se determinar $P(B|A)$. Como

$$P(A) = P(2) + P(4) + P(6) = \frac{3}{6} = \frac{1}{2},$$
$$P(A \cap B) = P(2) + P(4) = \frac{2}{6} = \frac{1}{3}$$

vem

$$P(B|A) = \frac{1/3}{1/2} = \frac{2}{3}.$$

Observe-se que $P(A \cap B)$ é a probabilidade de sair um número menor do que 5 que é par, o que é diferente da probabilidade de sair um número menor do que 5 sabendo que o resultado foi um número par.

[a] Como seria de esperar, $P(B|\Omega) = P(B)$ e $P(B|B) = 1$

Diz-se que A e B são **acontecimentos independentes** se

$$P(B|A) = P(B)$$

ou, equivalentemente,

$$P(A \cap B) = P(A)P(B).$$

Dois acontecimentos independentes são dois eventos em relação aos quais a ocorrência ou não ocorrência de um deles não altera a probabilidade de ocorrência do outro.

Exemplo 2: em dois lançamentos consecutivos de um dado não viciado, os acontecimentos

$$\begin{aligned} \text{A} &= \text{“saída de 1 no primeiro lançamento”} \\ \text{B} &= \text{“saída de 2 no segundo lançamento”} \end{aligned}$$

são independentes. Tem-se $P(A) = P(B) = 6/36 = 1/6$ e portanto a probabilidade de sair (1,2) em dois lançamentos consecutivos é

$$P(A \cap B) = \frac{1}{6}\frac{1}{6} = \frac{1}{36}.$$

Exemplo 3: Sabe-se que, em famílias de uma determinada população, a probabilidade:

- da mãe ser hipertensa é 0.2
- do pai ser hipertenso é 0.25
- de um dos pais ser hipertenso é 0.35.

Os acontecimentos "ser mãe hipertensa" (MH) e "ser pai hipertenso" (PH) são independentes?

Resolução: Os acontecimentos MH e PH serão independentes se

$$P(MH \cap PH) = P(MH)\, P(PH). \qquad (1.1)$$

Sabe-se que $P(MH) = 0.2$ e $P(PH) = 0.25$ mas desconhece-se $P(MH \cap PH)$ portanto nada se pode concluir daqui. Contudo, do enunciado resulta também que

$$P(MH \cup PH) = 0.35$$

e viu-se já que, para a reunião de dois acontecimentos, no caso $MH \cup PH$, é válida a fórmula

$$P(MH \cup PH) = P(MH) + P(PH) - P(MH \cap PH)$$

portanto

$$P(MH \cap PH) = 0.2 + 0.25 - 0.35 = 0.1.$$

Voltando à equação (1.1), como

$$0.1 \neq 0.2 \times 0.25$$

conclui-se que os acontecimentos MH e PH não são independentes.

Teorema 1.2.1 *Se A é um acontecimento qualquer e $A_1, \dots, A_n$ são acontecimentos disjuntos cuja reunião é o espaço amostral ($A_1 \cup \cdots \cup A_n = \Omega$), então*

(a) teorema da probabilidade total

$$P(A) = P(A|A_1)P(A_1) + \cdots + P(A|A_n)P(A_n)$$

(b) teorema de Bayes

$$P(A_k|A) = \frac{P(A|A_k)P(A_k)}{\sum_{j=1}^{n} P(A|A_j)P(A_j)}.$$

Exemplo 4: Todos os indivíduos adultos de uma determinada comunidade são classificados de acordo com o sexo em

feminino (F) ou masculino (M)

e de acordo com o estado civil em

solteiro (S), casado (C) ou outro (O).

Para essa comunidade, sabe-se que:

- 25% dos indivíduos são solteiros e 65% são casados.
- de entre os solteiros, 65% são mulheres
- de entre os casados, 40% são homens
- de entre os indivíduos com estado civil "outro", 45% são homens.

Homepage

Página de Rosto

Índice Geral

Voltar

Full Screen

Fechar

Desistir

Perguntas:

(a) Escolheu-se aleatoriamente uma mulher da população; qual é a probabilidade de ela ser solteira?

(b) Escolheu-se aleatoriamente um indivíduo da população em causa; qual é a probabilidade de ele ser do sexo masculino?

Resolução: Os dados do problema mostram que:

$$P(S) = 0.25, \quad P(C) = 0.65,$$
$$P(F|S) = 0.65, \quad P(M|C) = 0.4, \quad P(M|O) = 0.45.$$

(a) Pretende-se determinar $P(S|F)$. Pelo teorema de Bayes, tem-se

$$P(S|F) = \frac{P(F|S)P(S)}{P(F|S)P(S) + P(F|C)P(C) + P(F|O)P(O)}.$$

Agora

- $P(O) = 1 - P(S) - P(C) = 0.1$
- $P(F|C) = 1 - P(M|C) = 0.6$
- $P(F|O) = 1 - P(M|O) = 0.55$

portanto

$$\begin{aligned} P(S|F) &= \frac{0.65 \times 0.25}{0.65 \times 0.25 + 0.6 \times 0.65 + 0.55 \times 0.1} \\ &\approx 0.27\,. \end{aligned}$$

(b) Pretende-se determinar $P(M)$. Tem-se

$$\begin{aligned} &P(M) \\ &= 1 - P(F) \\ &= 1 - P(F|S)P(S) - P(F|C)P(C) - P(F|O)P(O) \\ &\approx 0.39\,. \end{aligned}$$

Homepage

Página de Rosto

Índice Geral

Página 11 de 26

Voltar

Full Screen

Fechar

Desistir

1.3. Variáveis aleatórias

Uma variável aleatória (v.a.) é uma função X que atribui um número a cada um dos resultados possíveis de uma experiência aleatória,

$$X : \Omega = \{\text{espaço amostral}\} \longrightarrow \mathbb{R}.$$

Usualmente as variáveis aleatórias são designadas por letras maiúsculas enquanto que os valores que elas podem tomar são designados por letras minúsculas. Por exemplo, $P(X = x)$ representa a probabilidade da v.a. X tomar o valor (teórico) x.

Uma v.a. pode ser:

- **discreta** (ou **categórica** ou **qualitativa**): se toma um número finito de valores ou um número infinito *numerável* de valores.
  Uma v.a. discreta pode ser:
  - **ordinal**: se existe uma ordem natural entre os diferentes valores da v.a.
  - **nominal**: caso contrário.
- **contínua** (ou **quantitativa**): se toma um número infinito *não numerável* de valores.

Exemplos de v.a. discretas:

(a) X v.a. (ordinal) que representa o resultado do lançamento de um dado;

$X : \{1, 2, 3, 4, 5, 6\} \to \mathbb{R}$, $i \mapsto i$.

(b) X v.a. (nominal) que representa o resultado do lançamento de uma moeda;

$X : \{H, T\} \to \mathbb{R}$, $X(H) = 1$ e $X(T) = 0$.

(c) X v.a. (ordinal) que representa a resposta a um determinado tratamento (excelente, boa, razoável, má);

$X : \{\text{exc}, \text{boa}, \text{raz}, \text{má}\} \to \mathbb{R}$,
$X(\text{exc}) = 3$, $X(\text{boa}) = 2$, $X(\text{raz}) = 1$, $X(\text{má}) = 0$.

(d) X v.a. (nominal) que representa o sexo dos filhos em famílias com 2 filhos;

$X : \{♂♂, ♂♀, ♀♀, ♀♂\} \to \mathbb{R}$,
$X(♂♂) = 1$, $X(♂♀) = 2$, $X(♀♀) = 3$, $X(♀♂) = 4$

(e) X v.a. (ordinal) que representa o número de episódios de otites verificados até aos 2 anos de idade.

$X : \{0, 1, 2, 3, \cdots\} \to \mathbb{R}, \quad X(x) = x$

(f) O teste de um novo fármaco passa pela prescrição do mesmo aos primeiros 4 pacientes que necessitem de medicação análoga.

X v.a. (ordinal) que representa o número de pacientes, em 4, que melhoram com o novo fármaco

$X : \{0, 1, 2, 3, 4\} \to \mathbb{R}, \quad X(x) = x$.

Exemplos de v.a. contínuas

(a) X v.a. que representa o peso (em g) dos alunos inscritos na FMUP no ano lectivo 2008/2009;

$X : [45\,000, 130\,000] \to \mathbb{R}, \quad X(x) = x$.

(b) X v.a. que representa o pH de uma solução;

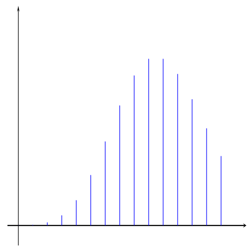
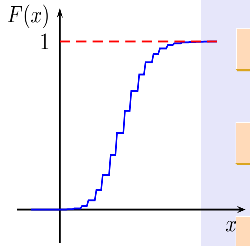
$X : [0, 14] \to \mathbb{R}, \quad X(x) = x$.

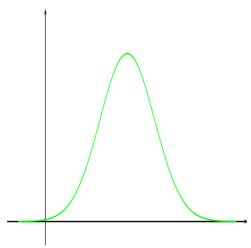
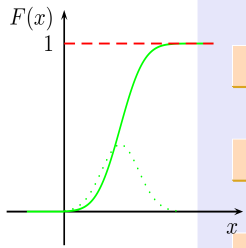
(c) X v.a. que representa a quantidade (em ml) de água ingerida por dia, por indivíduo da população do grande Porto, durante o mês de Outubro;

$X : [0, 10\,000] \to \mathbb{R}, \quad X(x) = x$.

1.4. Função de probabilidade, função densidade de probabilidade e função de distribuição

$X : \Omega = \{\text{espaço amostral}\} \longrightarrow \mathbb{R}$ **variável aleatória**

Variável	Função de prob.	Função de distrib.	Gráfico de f	Gráfico de F
Discreta $X \in \{a_1, a_2, \ldots, a_N\}$ ou $\{a_1, \ldots, a_n, \ldots\}$	$f : \{\text{valores de } X\} \to \mathbb{R}$ $f(x) \geq 0$ $\sum_x f(x) = 1$ $\mathbf{f(x) = P(X = x)}$	$F : \mathbb{R} \longrightarrow [0, 1]$ $F(x) = P(X \leq x)$ $= \sum_{u \leq x} f(u)$		

Variável	Função densid. de prob.	Função de distrib.	Gráfico de f	Gráfico de F
Contínua $X \in \mathbb{R}$	$f : \mathbb{R} \longrightarrow \mathbb{R}$ $f(x) \geq 0$ $\int_{-\infty}^{+\infty} f(x)\, dx = 1$ $\mathbf{P(X = x) = 0}$!!	$F : \mathbb{R} \longrightarrow [0, 1]$ $F(x) = P(X \leq x)$ $= \int_{-\infty}^{x} f(u)\, du$		

Homepage
Página de Rosto
Índice Geral
Página 13 de 26
Voltar
Full Screen
Fechar
Desistir

Exemplo - Caso discreto:

Considere-se novamente a v.a. X que conta o número de pacientes nos quais um determinado fármaco produz o efeito desejado, de entre os primeiros 4 a necessitarem de medicação análoga.

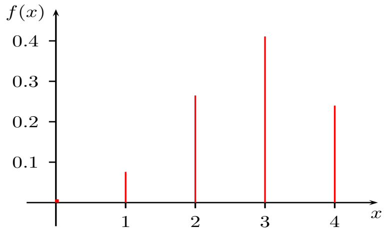
$$X : \{0, 1, 2, 3, 4\} \to \mathbb{R}, \quad X(x) = x.$$

De experiências anteriores (efectuadas em *muitos* grupos de 4 pacientes), sabe-se que:

- a probabilidade de nenhum indivíduo, em 4, melhorar com o novo fármaco é 0.008
- a probabilidade de 1 indivíduo, em 4, melhorar com o novo fármaco é 0.076
- a probabilidade de 2 indivíduos, em 4, melhorarem com o novo fármaco é 0.265
- a probabilidade de 3 indivíduos, em 4, melhorarem com o novo fármaco é 0.411
- a probabilidade de 4 indivíduos, em 4, melhorarem com o novo fármaco é 0.240

A função de probabilidade[a] de X é $f : \{0, 1, 2, 3, 4\} \to \mathbb{R}$

$$f(0) = 0.008, \quad f(1) = 0.076, \quad f(2) = 0.265,$$
$$f(3) = 0.411, \quad f(4) = 0.240.$$



[a] $f(x) = P(X = x)$; repare-se que $f(x) \geq 0$ e $\sum_x f(x) = 1$.

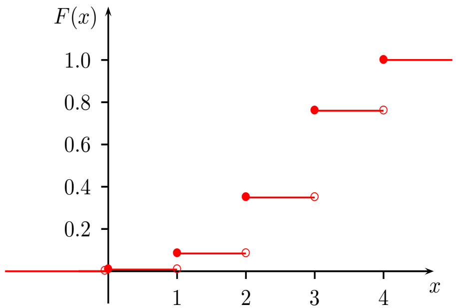
Para o cálculo da função de distribuição recorde-se que, por X ser discreta,

$$F(x) = P(X \leq x) = \sum_{u \leq x} f(u).$$

Assim $F : \mathbb{R} \to \mathbb{R}$ é a função definida por

$$F(x) = \begin{cases} 0, & x < 0 \\ 0.008, & 0 \leq x < 1 \\ 0.008 + 0.076 = 0.084, & 1 \leq x < 2 \\ 0.084 + 0.265 = 0.349, & 2 \leq x < 3 \\ 0.349 + 0.411 = 0.760, & 3 \leq x < 4 \\ 0.760 + 0.240 = 1, & 4 \leq x \end{cases}$$

que, tal como em todos os casos discretos, é uma **função em escada**:



Nota: Os valores de $f(x)$ apresentados aqui são teóricos. Isto significa que as frequências relativas obtidas por um clínico que quisesse conduzir uma experiência análoga numa amostra da população considerada não seriam, à partida, exactamente iguais aos valores de f mas tenderiam para esses valores à medida que o tamanho da amostra fosse aumentando.

Algumas cálculos e observações:

- A probabilidade de o fármaco produzir o efeito desejado em 2 indivíduos ou menos (em 4) é

$$P(X \leq 2) = F(2) = 0.349$$

- A probabilidade de o fármaco produzir o efeito desejado em menos de 2 indivíduos (em 4)

$$P(X < 2) = P(X \leq 1) = F(1) = 0.084$$

- A probabilidade de o fármaco produzir o efeito desejado em 2 ou mais indivíduos (em 4) é

$$\begin{aligned} P(X \geq 2) &= 1 - P(X < 2) \\ &= 1 - P(X \leq 1) \\ &= 1 - F(1) \\ &= 1.0.084 \\ &= 0.916 \end{aligned}$$

ou, alternativamente

$$\begin{aligned} P(X \geq 2) &= P(X = 2 \text{ ou } X = 3 \text{ ou } X = 4) \\ &= f(2) + f(3) + f(4) \\ &= 0.265 + 0.411 + 0.240 \\ &= 0.916. \end{aligned}$$

Função de probabilidade de X

$f : \{0, 1, 2, 3, 4\} \to \mathbb{R}$ definida por

$$\begin{array}{lll} f(0) = 0.008, & f(1) = 0.076, & f(2) = 0.265, \\ f(3) = 0.411, & f(4) = 0.240. & \end{array}$$

Função de distribuição de X

$F : \mathbb{R} \to \mathbb{R}$ definida por

$$F(x) = \begin{cases} 0, & x < 0 \\ 0.008, & 0 \leq x < 1 \\ 0.008 + 0.076 = 0.084, & 1 \leq x < 2 \\ 0.084 + 0.265 = 0.349, & 2 \leq x < 3 \\ 0.349 + 0.411 = 0.760, & 3 \leq x < 4 \\ 0.760 + 0.240 = 1, & 4 \leq x \end{cases}$$

Caso contínuo:

Se $f:\mathbb{R}\to\mathbb{R}$ é a função definida por

$$f(x)=\begin{cases} cx^2, & 0<x<3\\ 0, & \text{caso contrário}\end{cases} \tag{1.2}$$

determinar:

(a) o valor da constante c de forma a que f represente a função densidade de probabilidade de uma variável aleatória contínua X.

(b) a função de distribuição F.

(c) $P(-1\leq X<2)$.

(a) A primeira condição a impor é $f(x)\geq 0$ o que se verifica para qualquer $c>0$ dado que, para qualquer $x\geq 0$, x^2 é sempre não negativo.

A segunda condição a impor é

$$\int_{-\infty}^{+\infty} f(u)du=1,$$

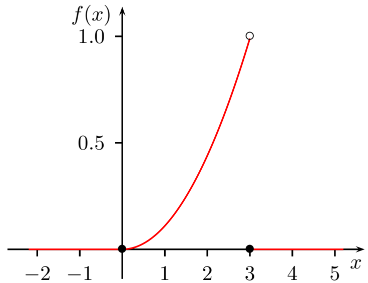
i.e., a área da região compreendida entre o gráfico da função e o eixo dos xx é 1. Ora

$$\begin{aligned} 1 &= \int_{-\infty}^{+\infty} f(u)\,du=\int_0^3 cu^2\,du\\ &= \frac{c}{3}u^3\Big|_0^3=9c\end{aligned}$$

logo $c=1/9$.

Em conclusão, de entre todas as funções do tipo (1.2) a única que pode representar uma função densidade de probabilidade de uma variável aleatória contínua é a função

$$f(x)=\begin{cases} \frac{1}{9}x^2, & 0<x<3\\ 0, & \text{caso contrário}\end{cases}$$



(b) Por definição, para qualquer $x\in\mathbb{R}$ tem-se

$$F(x)=P(X\leq x)=\int_{-\infty}^{x} f(u)\,du.$$

A determinação da função F passa pela discussão de três situações distintas:

Caso 1: $x<0$. Tem-se

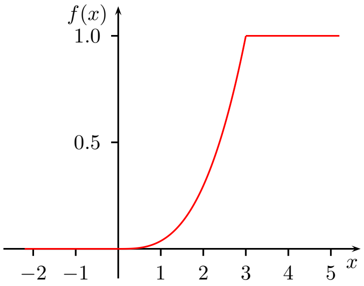
$$F(x)=\int_{-\infty}^{x} f(u)\,du=\int_{-\infty}^{x} 0\,du=0.$$

Caso 2: $0\leq x\leq 3$. Tem-se

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} f(u)\,du=\int_{-\infty}^{0} 0\,du+\int_0^x \frac{1}{9}u^2\,du\\ &= x^3/27.\end{aligned}$$

Caso 3: $x > 3$. Tem-se

$$\begin{aligned} F(x) &= \int_{-\infty}^{0} 0\,du + \int_{0}^{3} \frac{1}{9}u^3\,du + \int_{3}^{+\infty} 0\,du \\ &= \frac{1}{27}u^3\Big|_0^3 = \frac{1}{27}27 = 1. \end{aligned}$$



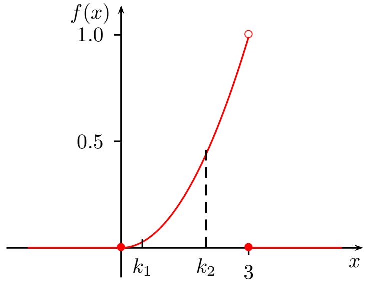
Tal como deveria acontecer, a função F é uma função crescente que toma valores entre 0 e 1.

(c) Observe-se primeiro que, no caso contínuo, $P(X = k) = 0$, para qualquer $k \in \mathbb{R}$. Para quaisquer valores $k_1, k_2 \in \mathbb{R}$ tem-se então

$$\begin{aligned} P(k_1 \leq X < k_2) &= P(k_1 < X \leq k_2) \\ &= F(k_2) - F(k_1) \\ &= \int_{k_1}^{k_2} f(u)\,du \end{aligned}$$

o que, geometricamente, corresponde à área da região compreendida entre o gráfico da função f e o eixo dos xx, entre k_1 e k_2.

- nao dei



No caso em questão,

$$\begin{aligned} P(-1 \leq X < 2) &= \int_{-1}^{2} f(u)\,du \\ &= \int_{-1}^{0} 0\,du + \int_{0}^{2} \frac{1}{9}u^2\,du \\ &= \frac{1}{27}u^3\Big|_0^2 = \frac{8}{27}. \end{aligned}$$

Homepage

Página de Rosto

Índice Geral

Página 17 de 26

Voltar

Full Screen

Fechar

Desistir

Exemplo 1: Seja X a v.a. que representa os valores (mmHg) da pressão arterial diastólica (PAD) na população[a] feminina adulta de uma determinada comunidade. Supondo que

$$P(90 < X < 95) = 0.1, \quad P(X < 90) = 0.7$$
$$P(75 < X < 100) = 0.7, \quad P(X > 100) = 0.15$$

calcule:

(a) $P(X > 90)$

(b) $P(X < 95)$

(c) $P(X < 75)$

(d) $P(75 < X < 90)$

(e) $P(90 < X < 100)$

(f) Dentro do contexto do problema, dê exemplo de 2 eventos mutuamente exclusivos e 2 não mutuamente exclusivos (para X).

Resolução: As alíneas seguintes usam o facto de $P(X = x) = 0$ para uma v.a. contínua.

(a) $P(X > 90) = 1 - P(X \leq 90) = 1 - 0.7 = 0.3$

(b) $$\begin{aligned} P(X < 95) &= P(90 < X < 95 \quad \text{ou} \quad X < 90) \\ &= P(90 < X < 95) + P(X < 90)^{b} \\ &= 0.1 + 0.7 \\ &= 0.8 \end{aligned}$$

[a] não é numa amostra.

[b] $P(90 < X < 95) + P(X < 90) - P(90 < X < 95 \text{ e } X < 90)$

(c) $$\begin{aligned} & P(X < 75) \\ &= 1 - P(X > 75) \\ &= 1 - (P(75 < X < 100) + P(X > 100)) \\ &= 1 - (0.70 + 0.15) \\ &= 0.15 \end{aligned}$$

(d) $$\begin{aligned} & P(75 < X < 90) \\ &= P(75 < X < 100 \text{ e } X < 90) \\ &= P(75 < X < 100) + P(X < 90) - P(X < 100)^{a} \\ &= 0.7 + 0.7 - 0.85 \\ &= 0.55 \end{aligned}$$

(e) $$\begin{aligned} & P(90 < X < 100) \\ &= 1 - [P(X < 90 \text{ ou } X > 100)] \\ &= 1 - \left[P(X < 90) + P(X > 100)^{b}\right] \\ &= 1 - (0.7 + 0.15) \\ &= 0.15 \end{aligned}$$

(f) 2 eventos mutuamente exclusivos:

$X < 100$ e $X > 100$ (por exemplo)

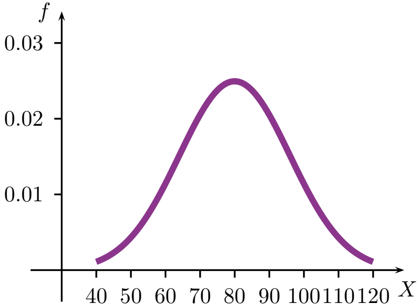
2 eventos que não são mutuamente exclusivos:

$75 < X < 100$ e $X < 90$ (por exemplo).

[a] $P(75 < X < 100 \text{ ou } X < 90) = P(X < 100)$

[b] $P(X < 90 \text{ e } X > 100) = 0$

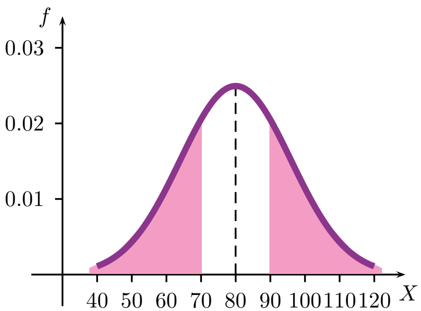
Exemplo 2: Seja X a v.a. que representa os valores (mmHg) da pressão arterial diastólica (PAD) numa população de homens com idades compreendidas entre os 33 e 44 anos e suponha que a f.d.p. de X é a apresentada abaixo.



Várias observações podem ser feitas:

- o gráfico da f.d.p. é em forma de sino
- a f.d.p. é **simétrica** relativamente a $X = 80$:

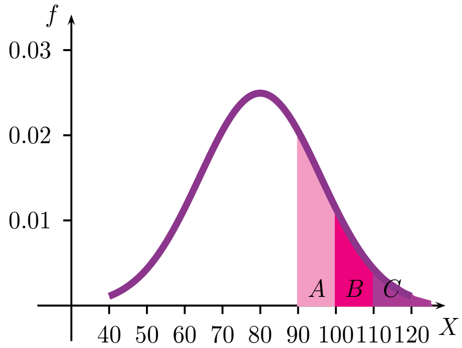
  $P(X > 80+y) = P(X < 80-y)$ para qualquer $y \geq 0$.



Em particular, $P(X \geq 80) = P(X \leq 80) = 0.5$.

- Os valores mais prováveis de PAD para um indivíduo da população em causa são os que rondam os 80 mmHg sendo que a probabilidade de um indivíduo ter PAD cada vez mais diferente de 80 é cada vez menor.

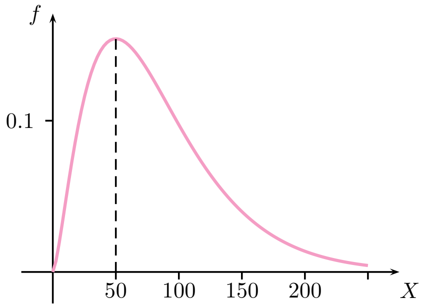
- A probabilidade de um indivíduo da população em causa ser:
  - ligeiramente hipertenso (90¡PAD¡100) é $P(90 < X < 100) = \text{área}(A)$
  - moderadamente hipertenso (100¡PAD¡110) é $P(100 < X < 110) = \text{área}(B)$
  - severamente hipertenso (PAD¿110) é $P(X > 110) = \text{área}(C)$



- A f.d.p. apresentada (e as conclusões anteriores) dizem respeito à <u>população</u> de indivíduos do sexo masculino com idades compreendidas entre os 33 e os 44 anos, e não a uma amostra de indivíduos com essas características.

Homepage
Página de Rosto
Índice Geral
Página 19 de 26
Voltar
Full Screen
Fechar
Desistir

Exemplo 3: Seja X a v.a. que representa os níveis de triglicéridos no sangue (mg/dL) de uma determinada população. Suponha que estudos realizados anteriormente conduziram à seguinte f.d.p. para X:



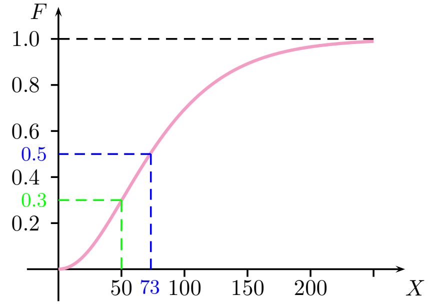
Algumas observações:

- os valores com maior probabilidade de ocorrência rondam os 50 mg/dL
- a distribuição não é simétrica (em relação a nenhum valor). Por exemplo, para $X = 50$, tem-se

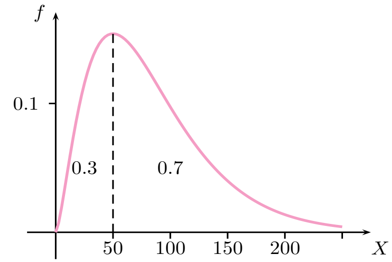
$$P(X \geq 50) > P(X \leq 50)$$

- a distribuição é assimétrica à direita
- a probabilidade de encontrar um indivíduo da população com um nível de triglicéridos superior a 200 mg/mL é muito baixa
- dado um qualquer indivíduo da população, a probabilidade do seu nível de triglicéridos ser inferior a 50 mg/dL é superior à probabilidade desse nível ser superior a 150 mg/dL

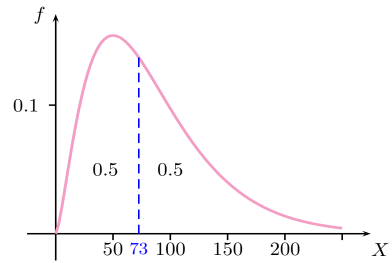
A função de distribuição de X é



- $P(X \leq 50) = F(50) = 0.3 \quad P(X \geq 50) = 0.7$



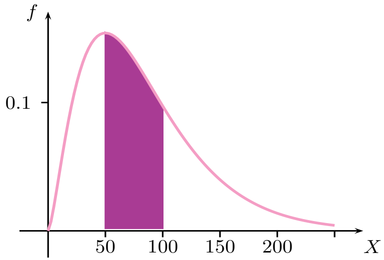
- o ponto que divide a área do gráfico de f em duas partes iguais é $x = 73$



FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO
Homepage
Página de Rosto
Índice Geral
Página 20 de 26
Voltar
Full Screen
Fechar
Desistir

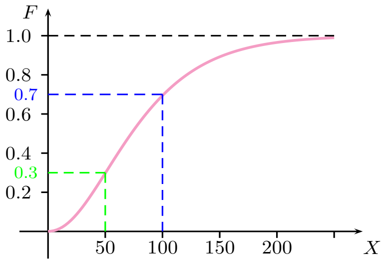
- existe uma probabilidade superior a 0.9 de encontrar na população um indivíduo com um nível de triglicéridos inferior a 200 mg/dL

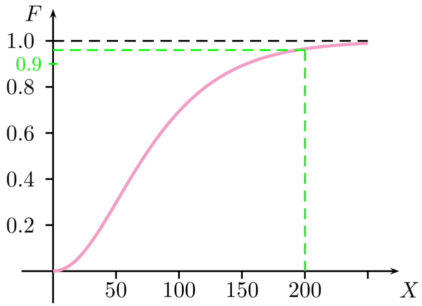
- a probabilidade de encontrar na população um indivíduo com um nível de triglicéridos compreendido entre 50 e 100 mg/dL



pode ser determinada através dos valores da função de distribuição. Tem-se

$P(50 < X < 100) = F(100) - F(50) = 0.7 - 0.3 = 0.4$





1.6. Medidas de localização e escala de uma variável aleatória

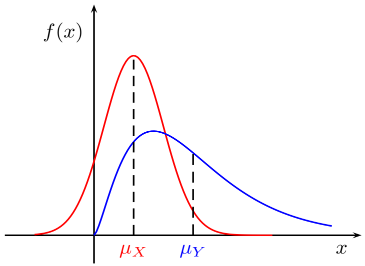
Espaço de Probabilidade $(\Omega = \mathbb{R}, \mathcal{A}, P)$

	X v.a. discreta	X v.a. contínua
Média (ou **Valor Esperado**) $\mu = \mu_X = E(X)$ *Mean* (ou *Expected Value*)	$\mu = \sum_k x_k f(x_k)$	$\mu = \int_{-\infty}^{+\infty} x f(x) dx$
Variância $\sigma^2 = \sigma_X^2 = \mathrm{Var}(X) = E[(X-\mu)^2]$ *Variance*	$\sigma^2 = \sum_k (x_k - \mu)^2 f(x_k)$	$\sigma^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x) dx$
Desvio Padrão *Standard Deviation*	$\sigma = \sqrt{\text{variância}}$	$\sigma = \sqrt{\text{variância}}$
Moda *Mode*	$x \in \mathbb{R} : f(x)$ é máximo (x que ocorre com maior freq.)	$x \in \mathbb{R} : f(x)$ é máximo (x que ocorre com maior freq.)
p-**Quantil** *p-Quantile*		$q \in \mathbb{R}$ tal que $P(X \leq q) = p$
Mediana $m = \mathrm{med}(X) = q_{0.5}$ *Median*		$m \in \mathbb{R}$ tal que $P(X \leq m) = P(X \geq m) = 0.5$

Nota: O conceito de quantil (e de mediana) pode ser generalizado ao caso discreto mas são os quantis de v.a. contínuas que têm mais interesse. No caso contínuo, a mediana corresponde ao 0.5-quantil.

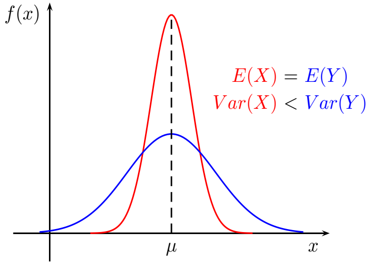
De entre as medidas anteriores:

- a média, a moda e a mediana são medidas de localização (ou tendência central) da distribuição



- a determinação dos p-quantis (e, em particular, da mediana) passa pela determinação da função de distribuição F

- a variância (ou equivalentemente, o desvio-padrão) é uma medida de escala (ou dispersão) dos valores da variável em torno das medidas de localização



Pode-se mostrar que

$$\mathrm{Var}(X) = E\left(X^2\right) - (E(X))^2$$

onde, no caso discreto,

$$E\left(X^2\right) = \sum_k x_k^2\, f(x_k)$$

e, no caso contínuo,

$$E\left(X^2\right) = \int_{-\infty}^{+\infty} x^2\, f(x)\, dx.$$

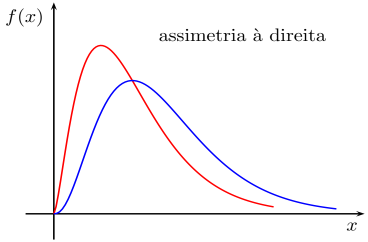
- Existe uma medida - **coeficiente de assimetria** γ_1 - que serve para avaliar a simetria da distribuição (relativamente à sua média). No caso discreto

$$\gamma_1 \quad = \quad \frac{1}{\sigma^3} \sum_k (x_k - \mu)^3 f(x_k)$$

e no caso contínuo

$$\gamma_1 = \frac{1}{\sigma^3} \int_{-\infty}^{+\infty} (x-\mu)^3 f(x) dx.$$

Se a cauda mais longa se situa à direita, a distribuição diz-se **assimétrica à direita** ou **assimétrica positiva**; caso contrário diz-se **assimétrica à esquerda** ou **assimétrica negativa**.



Exemplo 1:
Considere-se novamente a v.a. X que conta o número de pacientes nos quais um determinado fármaco produz o efeito desejado, de entre os primeiros 4 a necessitarem de medicação análoga, e a sua função de probabilidade $f : \{0, 1, 2, 3, 4\} \to \mathbb{R}$ dada por

$$f(0) = 0.008, \quad f(1) = 0.076, \quad f(2) = 0.265,$$
$$f(3) = 0.411, \quad f(4) = 0.240.$$

Em média, o número de pacientes (em 4) nos quais se espera que o fármaco produza efeito é

$$\begin{aligned} E(X) &= \sum_k x_k\, f(x_k) \\ &= 0(0.008) + 1(0.076) + 2(0.265) \\ &\quad +3(0.411) + 4(0.240) \\ &= 2.80 \end{aligned}$$

Quanto à variância de X, tem-se

$$\begin{aligned} &\mathrm{Var}(X) \\ &= \sum_k (x_k - \mu)^2\, f(x_k) \\ &= (0-2.80)^2 0.008 + (1-2.80)^2 0.076 \\ &\quad +(2-2.80)^2 0.265 + (3-2.80)^2 0.411 \\ &\quad +(4-2.80)^2 0.240 \\ &\approx 0.841. \end{aligned}$$

(Também se podia ter usado a fórmula $\mathrm{Var}(X) = E(X^2) - (E(X))^2$.) Em particular o desvio padrão é

$$\sigma_X = \sqrt{0.841} \approx 0.917.$$

Nota: Para uma amostra *muito grande* de grupos de 4 indivíduos, espera-se que o número médio de indivíduos nos quais o fármaco produz efeito esteja *muito próximo* de 2.8 e a variância *muito próxima* de 0.841 (sendo a aproximação tanto melhor quanto maior for o tamanho da amostra, pois isso implica que a amostra se está a aproximar da população). Veremos mais tarde, no capítulo sobre estatísticas amostrais, os fundamentos teóricos que sustentam esta afirmação.

Exemplo 2:
Considere-se um jogo feito com um dado não viciado em que um jogador ganha 20 euros se sair um 2, 40 euros se sair um 4, perde 30 euros se sair um 6 e não ganha nem perde se aparecer qualquer outra face do dado. Qual é o ganho esperado por um jogador numa jogada?

Seja X a variável aleatória que representa a quantia ganha pelo jogador em qualquer jogada, mais precisamente, seja $X(i)$ o ganho do jogador originado pela saída da face i do dado numa jogada. Então X toma os valores -30, 0, 20 e 40 e a função de probabilidade é

$$f(-30) = \frac{1}{6}, \quad f(0) = \frac{3}{6}, \quad f(20) = \frac{1}{6}, \quad f(40) = \frac{1}{6}.$$

Com estes valores,

$$E(X) = (-30)\left(\frac{1}{6}\right) + 0\left(\frac{3}{6}\right) + 20\left(\frac{1}{6}\right) + 40\left(\frac{1}{6}\right) = 5.$$

Isto significa que numa qualquer jogada um jogador pode esperar ganhar 5 euros.

Observe-se ainda que a dispersão dos valores de X em torno da média é relativamente grande

$$\begin{aligned}\mathrm{Var}(X) &= (-30-5)^2\left(\frac{1}{6}\right)+(0-5)^2\left(\frac{3}{6}\right)\\ &\quad +(20-5)^2\left(\frac{1}{6}\right)+(40-5)^2\left(\frac{1}{6}\right)\\ &= 458.3\end{aligned}$$

Outros cálculos e observações:

- A função de distribuição é $F:\mathbb{R}\to\mathbb{R}$ definida por

$$\begin{array}{rl} x<-30 & F(x)=0\\ -30\le x<0 & F(x)=1/6\\ 0\le x<20 & F(x)=4/6\\ 20\le x<40 & F(x)=5/6\\ 40\le x & F(x)=1.\end{array}$$

- A probabilidade de um indivíduo não perder dinheiro numa qualquer jogada é

$$P(X\ge 0)=1-P(X<0)=1-P(X=-30)=5/6$$

- A probabilidade de um indivíduo numa qualquer jogada ter um ganho superior a 10 euros é

$$P(X>10)=1-P(X\le 10)=1-F(10)=1-4/6=2/6$$

- A moda de X é 0 (pois f atinge o seu valor máximo em $f(0)$).

Exemplo 3:
Considere-se uma variável aleatória contínua com função densidade de probabilidade f dada por

$$f(x) = \begin{cases} \frac{1}{2}x & 0 < x < 2 \\ 0 & \text{caso contrário} \end{cases}$$

O valor esperado de X é, por definição,

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x f(x)\, dx = \int_0^2 x \frac{1}{2} x\, dx \\ &= \frac{x^3}{6}\Big|_0^2 = \frac{4}{3} \end{aligned}$$

A variância é dada por

$$\begin{aligned} \text{Var}(X) &= E\left[\left(X - \frac{4}{3}\right)^2\right] \\ &= \int_{-\infty}^{+\infty} \left(x - \frac{4}{3}\right)^2 f(x)\, dx \\ &= \int_0^2 \left(x - \frac{4}{3}\right)^2 \frac{1}{2} x\, dx \\ &= \frac{2}{9} \end{aligned}$$

sendo portanto o desvio padrão igual a

$$\sigma_X = \sqrt{\text{Var}(X)} = \frac{\sqrt{2}}{3}.$$

Homepage

Página de Rosto

Índice Geral

Página 26 de 26

Voltar

Full Screen

Fechar

Desistir