Ciência dos Dados

Variáveis Aleatórias Contínuas

Distribuição de probabilidades
Esperança e variância

Magalhães e Lima, 7ª. Edição. Seção 6.1 e Seção 6.2

Objetivos de Aprendizagem

Os alunos devem ser capazes de:

- Classificar se uma função $f(x)$ é dita uma função densidade de probabilidade (fdp);
- Obter e interpretar uma função de distribuição acumulada (fda);
- Calcular e interpretar esperança (média) e variância;
- Contrastar resultados teóricos e empíricos.

Acompanhe, previamente, o PLANO DE AULA no BLACKBOARD!

O que vamos aprender?

- **Conceito de v.a. Contínua**
- **Distribuição de Probabilidades de uma v.a. Contínua**
- **Função de Distribuição Acumulada**
- **Características de uma v.a. Contínua**

Variáveis Aleatórias

Variável aleatória discreta: conjunto de possibilidades é um conjunto finito ou enumerável.

Variável aleatória contínua: **são variáveis cujos possíveis valores ocorrem aleatoriamente e pertencem a um intervalo dos números reais.**

Exemplos de Variáveis Aleatórias Contínuas

- Y – Tempo de execução de uma tarefa;
- Z – Nível de colesterol no sangue;
- W – Altura de trabalhadores de uma certa região.
- T – Tempo de espera em uma fila bancária: $t \geq 0$
- U – Intenção de votos em um candidato em uma certa semana:

  $0 \leq x \leq 1$
- S – Variação do real frente ao dólar em um certo dia: $-\infty \leq y \leq +\infty$

Distribuição Contínua de Probabilidades

De forma semelhante àquela desenvolvida para as v.a.'s discretas, precisamos estabelecer para as v.a.'s contínuas a atribuição de probabilidades às suas diversas realizações.

No entanto, agora a variável pode assumir um número infinito de valores diferentes.

Caso Discreto:

X: número da face observada no lançamento de um dado, $x=1,2,\ldots,6$.

Função de probabilidade: $P(X=x) = 1/6$.

Caso Contínuo:

Y: Altura de um indivíduo

A v.a. Y pode assumir infinitos valores dentro do intervalo que representa um indivíduo (depende da precisão) .

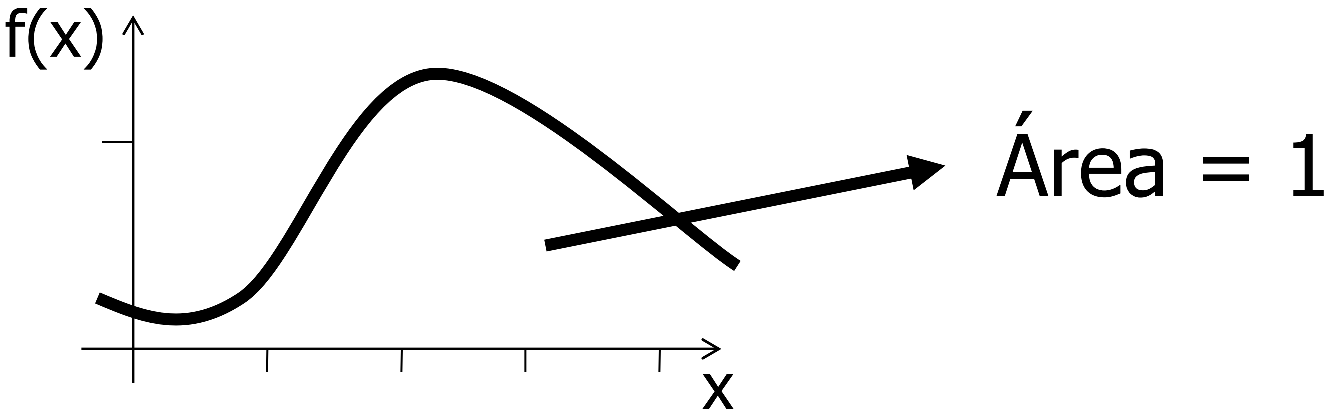
"Função de probabilidade": $P(Y=y) = 1/\infty = 0$.

?? Como calcular probabilidade para v.a.'s contínuas ??

Função de Densidade de Probabilidade (fdp)
Probability Density Function (pdf)

Definição 1.

(função densidade de probabilidade – f.d.p)



Função de Densidade de Probabilidade (fdp)
Probability Density Function (pdf)

Definição 1. (cont.)

(função densidade de probabilidade – f.d.p.)

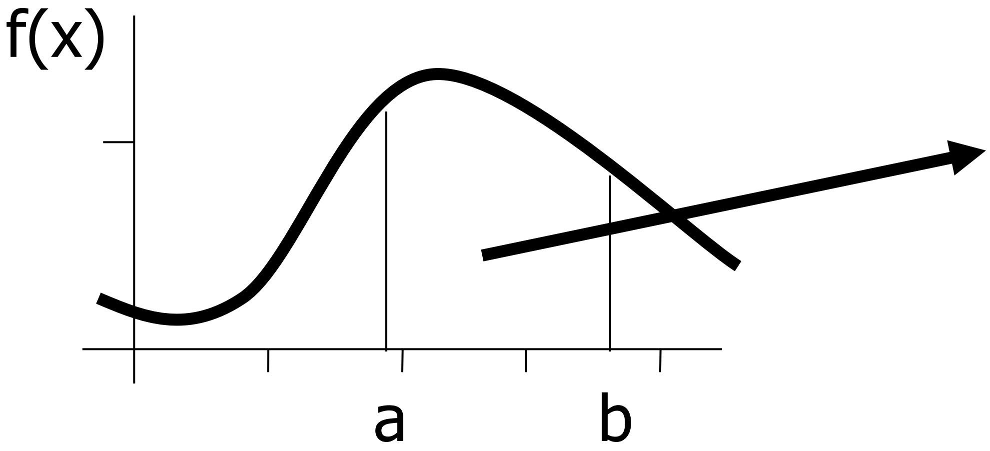
Dizemos que *f*(*x*) é uma f.d.p. para uma v.a. contínua *X*, se satisfaz a duas condições:

$$\text{a) } f(x) \geq 0, \ \mathrm{x} \in (-\infty;+\infty)$$

$$\text{b) } \int_{-\infty}^{+\infty} f(x)\, dx = 1$$

Função de Densidade de Probabilidade (fdp)
Probability Density Function (pdf)

Definição 2.



$P(a \leq X \leq b) = ?$

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx$$

Função de Densidade de Probabilidade (fdp)
Probability Density Function (pdf)

Definição 2. (cont.)

Para calcular probabilidades, temos que

$$P(a \le X \le b) = \int_a^b f(x)dx;$$

para $a \le b$.

A integral acima retorna a área sob a função $f(\cdot)$ definida pelo intervalo $[a;b]$.

Exemplo 1

O tempo de execução de um determinado loop (uso de um "for", por exemplo) pode demorar de (0,4] segundos. Logo, esse tempo é uma variável aleatória, aqui representada por X.

A distribuição de probabilidades de X é dada pela seguinte função:

$$f(x) = \begin{cases} cx(4-x), & \text{se} \quad 0 < x \leq 4 \\ 0, & \text{caso contrário} \end{cases}$$

a) **Encontre c para que $f(x)$ seja considerada uma fdp.**

b) **Qual a probabilidade de um loop demorar até 2,5 segundos?**

c) **Qual a probabilidade de um loop demorar até 0,5 segundo?**

d) **E menor do que um valor *k*, sendo *k* qualquer valor entre 0 e 4?**

e) **Qual a probabilidade de um loop demorar entre 0,5 e 2,5?**

f) **Qual o tempo média de execução e a variância do tempo de execução.**

Solução do Exemplo 1

Definição 1:

$i)\ f(x) \geq 0,\ x \in (-\infty; +\infty)$ $\qquad ii)\ \int_{-\infty}^{+\infty} f(x)\,dx = 1$

a)

$$\int_0^4 cx(4-x)dx = 1 \Rightarrow c\int_0^4 x(4-x)dx = 1 \Rightarrow c\int_0^4 (4x - x^2)dx = 1 \Rightarrow$$

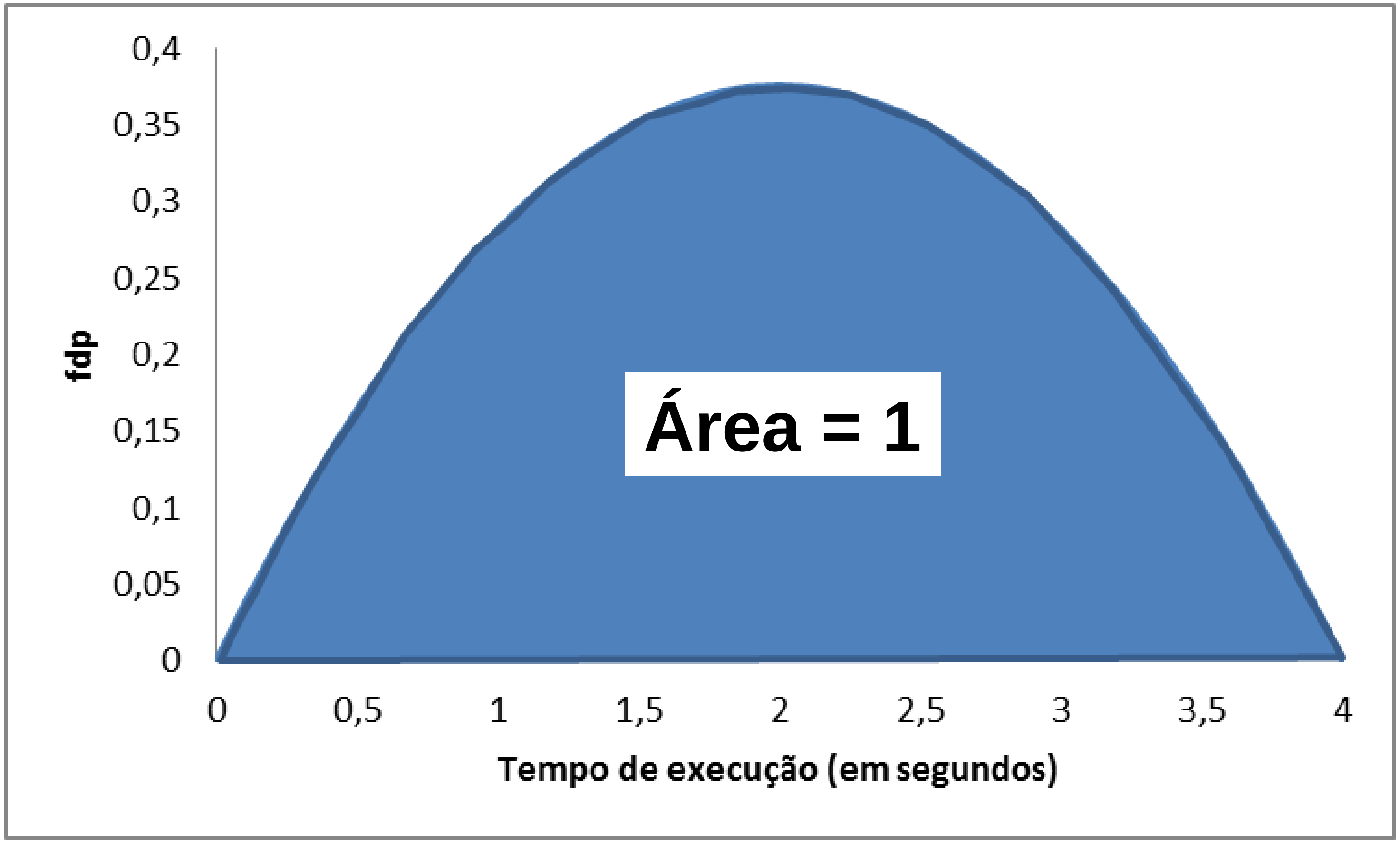
$$\Rightarrow c\left[\int_0^4 4x\,dx - \int_0^4 x^2 dx\right] = 1 \Rightarrow c\left[2x^2\Big|_0^4 - \frac{x^3}{3}\Big|_0^4\right] = 1 \Rightarrow$$

$$\Rightarrow c\left[32 - \frac{64}{3}\right] = 1 \Rightarrow c\left[\frac{96-64}{3}\right] = 1 \Rightarrow c\left[\frac{32}{3}\right] = 1 \Rightarrow c = \frac{3}{32}$$

Solução do Exemplo 1 (cont.)

a)

$$f(x)=\begin{cases}\dfrac{3}{32}x(4-x), & \text{se } 0\le x\le 4\\ 0, & \text{caso contrário}\end{cases}$$



Exemplo 1

O tempo de execução de um determinado loop (uso de um "for", por exemplo) pode demorar de (0,4] segundos. Logo, esse tempo é uma variável aleatória, aqui representada por X.

A distribuição de probabilidades de X é dada pela seguinte função:

$$f(x) = \begin{cases} \dfrac{3}{32}x(4-x), & \text{se} \quad 0 < x \le 4 \\ 0, & \text{caso contrário} \end{cases}$$

a) Encontre *c* para que $f(x)$ seja considerada uma fdp.

b) Qual a probabilidade de um loop demorar até 2,5 segundos?

c) Qual a probabilidade de um loop demorar até 0,5 segundo?

d) E menor do que um valor *k*, sendo *k* qualquer valor entre 0 e 4?

e) Qual a probabilidade de um loop demorar entre 0,5 e 2,5?

f) Qual o tempo média de execução e a variância do tempo de execução.

Função de Distribuição Acumulada (fda)
Cumulative Distribution Function (cdf)

Definição 3.

Dada uma v.a. *X* com f.d.p. *f*(*x*), podemos definir a sua função de distribuição acumulada, *F(x)*, por

$$F(x) = P(X \leq x), -\infty \leq x \leq \infty.$$

Assim, segue-se que

$$F(x) = \int_{-\infty}^{x} f(t)\,dt,$$

para todo real *x*.

Insper

Função de Distribuição Acumulada (fda)
Cumulative Distribution Function (cdf)

É fácil ver que 0 ≤ *F(x)* ≤ 1, para todo *x* real; além disso, *F(x)* é não-decrescente e possui as duas seguintes propriedades:

$$(i) \quad \lim_{x \to -\infty} F(x) = 0,$$

$$(ii) \quad \lim_{x \to \infty} F(x) = 1.$$

Insper

Função de Distribuição Acumulada (fda)
Cumulative Distribution Function (cdf)

Proposição.

Para todos os valores de *x* para os quais *F*(*x*) é derivável temos

$$F'(x) = \frac{dF(x)}{dx} = f(x)$$

Solução do Exemplo 1 (cont.)

d)

$$F(k) = P(X \leq k) = \int_0^k \frac{3}{32} x(4-x)dt = \frac{3}{32}\int_0^k (4x - x^2)dx =$$

$$= \frac{3}{32}\left[2x^2\Big|_0^k - \frac{x^3}{3}\Bigg|_0^k \right] = \frac{3}{32}\left[2k^2 - \frac{k^3}{3} \right] = \frac{k^2}{32}[6-k]$$

Exemplo 1

O tempo de execução de um determinado loop (uso de um "for", por exemplo) pode demorar de (0,4] segundos. Logo, esse tempo é uma variável aleatória, aqui representada por X.

A distribuição de probabilidades de X é dada pela seguinte função:

$$f(x)=\begin{cases}\dfrac{3}{32}x(4-x), & \text{se} \quad 0<x\le 4\\ 0, & \text{caso contrário}\end{cases}$$

a) Encontre *c* para que $f(x)$ seja considerada uma fdp.

b) Qual a probabilidade de um loop demorar até 2,5 segundos?

c) Qual a probabilidade de um loop demorar até 0,5 segundo?

d) E menor do que um valor *k*, sendo *k* qualquer valor entre 0 e 4?

e) Qual a probabilidade de um loop demorar entre 0,5 e 2,5?

f) Qual o tempo média de execução e a variância do tempo de execução.

Função de Distribuição Acumulada (fda)
Cumulative Distribution Function (cdf)

Resultado:

$$P(a \leq X \leq b) = P(X \leq b) - P(X \leq a) =$$

$$= F(b) - F(a),$$

para todo $a \leq b$.

Exemplo 1

O tempo de execução de um determinado loop (uso de um “for”, por exemplo) pode demorar de (0,4] segundos. Logo, esse tempo é uma variável aleatória, aqui representada por X.

A distribuição de probabilidades de X é dada pela seguinte função:

$$f(x)=\begin{cases}\dfrac{3}{32}x(4-x), & \text{se} \quad 0<x\leq 4\\ 0, & \text{caso contrário}\end{cases}$$

a) Encontre *c* para que $f(x)$ seja considerada uma fdp.

b) Qual a probabilidade de um loop demorar até 2,5 segundos?

c) Qual a probabilidade de um loop demorar até 0,5 segundo?

d) E menor do que um valor *k*, sendo *k* qualquer valor entre 0 e 4?

e) Qual a probabilidade de um loop demorar entre 0,5 e 2,5?

f) Qual o tempo média de execução e a variância do tempo de execução.

Insper

Momentos de uma v.a. Esperança e Variância

Definição 4.

(Valor médio de uma v.a. contínua):

$$E(X)=\int_{-\infty}^{\infty} xf(x)dx$$

Definição 5.

(Variância para uma v.a. contínua):

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

Momentos de uma v.a. Esperança e Variância

Propriedade:

$$\mathrm{Var}(X) = E[(X - E(X)]^2 =$$

$$= E(X^2) - [E(X)]^2$$

No caso discreto

$$E(X^2) = \sum_x x^2 P(X = x)$$

No caso contínuo

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

Solução do Exemplo 1 (cont.)

f)

$$f(x) = \begin{cases} \frac{3}{32} x(4 - x), & \text{se } \ 0 < x \leq 4 \\ 0, & \text{caso contrário} \end{cases}$$

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

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

Solução do Exemplo 1 (cont.)

f)

$$\mathrm{E(X)} = \int_0^4 x\,f(x)dx = \int_0^4 x\frac{3}{32}x(4-x)dx = \frac{3}{32}\int_0^4 x^2(4-x)dx = 2$$

$$\mathrm{Var(X)} = \int_0^4 (x-2)^2 f(x)dx = \int_0^4 (x-2)^2 \frac{3}{32}x(4-x)dx =$$

$$= \frac{3}{32}\int_0^4 (x^2-4x+4)x(4-x)dx =$$

$$= \frac{3}{32}\int_0^4 (4x^3 - x^4 - 16x^2 + 4x^3 + 16x - 4x)dx = \ldots = 0{,}80$$

Insper

Exercícios

Exercício 1

A f.d.p. abaixo modela o tempo, em horas, até a execução de uma tarefa.

$$f(x) = \begin{cases} \dfrac{2}{25}x, & 0 \leq x \leq 5 \\ 0, & \text{caso contrário} \end{cases}$$

Determine a probabilidade de

a) uma pessoa levar mais de 1 hora.

b) uma pessoa levar menos de ½ hora, uma vez que se sabe que levou menos de 1 hora.

c) uma pessoa levar exatamente 1 hora.

d) Encontre a função de distribuição acumulada (cdf).

e) Qual o tempo médio e desvio padrão de execução da tarefa?

f) Qual o tempo mediano de execução da tarefa?

Exercício 2

Seja *X* a variável aleatória que denota a sentença de prisão, em anos, para pessoas condenadas por furtos num particular estado brasileiro. Suponha que a f.d.p. de *X* seja dada por

$$f(x) = \begin{cases} c \cdot x^2, & 0 < x < 4 \\ 0, & \textit{caso contrário} \end{cases}$$

Exercício 2 (cont.)

É correto afirmar que:

(1) O valor da constante *c* para que a função descrita anteriormente represente uma função de probabilidade deve ser igual a $^{3}\!/_{64}$ **.**

(2) Em média, a sentença de prisão para pessoas condenadas por furtos neste particular estado brasileiro é de 3 anos.

(3) A probabilidade da pena de uma pessoa ser superior a 18 meses, devido a este tipo de delito, é inferior a 90%.

(4) O desvio padrão das sentenças é igual a 3,1 anos.

Exercício 3

Seja *X* uma v.a. com f.d.p. dada por

$$f(x) = \begin{cases} cx^2, & se \quad -1 < x < 1 \\ 0, & caso \;\; contrário \end{cases}$$

Verifique se as seguintes afirmações são verdadeiras ou falsas:

(1) O valor de *c*, para que *f*(*x*) seja uma função densidade de probabilidade, é igual a 2/3.

(2) O valor *k* tal que $F_x(k)$ = 1/2 é igual a 0. Logo, do exposto, podemos interpretar *k* como sendo a mediana da distribuição de *X*.

Exercício 4

O tempo entre a realização de duas vendas pode ser representado por uma v.a. com f.d.p. dada por

$$f(x) = 0{,}2e^{-0{,}2x}, \quad x > 0$$

a. **Encontre a função distribuição acumulada.**

b. **Determine P(2 < X < 4).**