Matemática Computacional

Capítulo 3

Distribuições discretas e contínuas

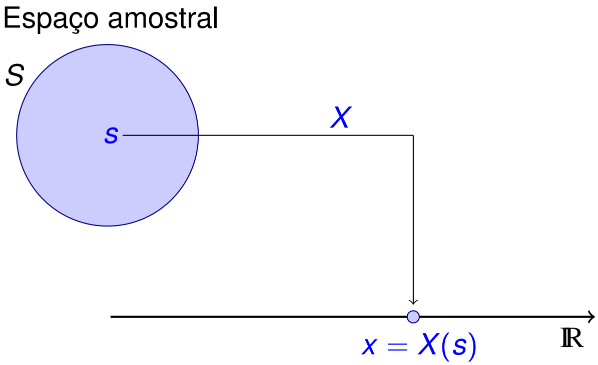
Licenciatura em Engenharia Informática
ISEP
(2024/2025)

Conteúdo

- Em análise estatística, trabalha-se com dados experimentais (dados numéricos ou não numéricos) que podem ser classificados de acordo com algum critério.
- Dados experimentais são observações de variáveis qualitativas (discretas) ou quantitativas (discretas ou contínuas) de populações finitas ou infinitas e são obtidos através da realização de experiências estatísticas.
- Experiência estatística é qualquer processo que gera um conjunto de dados que pode ser diferente de cada vez que este é executado, em iguais condições.
- Observação é qualquer registo de informação (numérica ou não) associado a uma variável.

Definição de variável aleatória

Uma variavel aleatória (v.a.) X é uma função que associa um número real x a cada resultado s do espaço amostral ou experimental S, $x = X(s)$.



- Frequentemente, o espaço experimental original já constitui a característica numérica que pretendemos estudar, e assim, $X(s) = s$ (função identidade).
- Usamos letra maiúscula para designarmos uma variavel aleatória.
- Usamos letra minúscula para designar uma das suas concretizações ou realizações.
- Uma variavel aleatória X é discreta se o conjunto de valores possíveis de X for finito ou infinito numerável, ou seja, representam dados contáveis ou numeráveis.
- Uma variavel aleatória X e contínua se tomar valores de um intervalo ou de uma colecção de intervalos, ou seja, representam dados medidos.

Exemplo 3.1: Considere-se os casais com olhos castanhos e com três filhos e a experiência estatística em que se regista a cor de olhos de cada um dos três filhos, por ordem crescente de idades. Estamos interessados no número de filhos com olhos azuis. Definir o espaço amostral e uma variável aleatória que seja apropriada.

Resolução: Considerem-se os acontecimentos,

A: "a criança tem olhos azuis".
C: "a criança tem olhos castanhos".

O espaço amostral é

$$S = \{AAA, AAC, ACA, ACC, CAA, CAC, CCA, CCC\}$$

A v.a. de interesse é

$X =$ número de filhos com olhos azuis entre os três filhos do casal.

Exemplo 3.1 (Cont.): Verifica-se:

$$X(AAA) = 3,$$

$$X(CAA) = X(ACA) = X(AAC) = 2,$$

$$X(ACC) = X(CAC) = X(CCA) = 1,$$

$$X(CCC) = 0,$$

e a correspondência entre o espaço amostral original e a característica numérica em estudo é mostrada na tabela:

s	AAA	AAC	ACA	ACC	CAA	CAC	CCA	CCC
x	3	2	2	1	2	1	1	0

Variáveis aleatórias discretas

Função de probabilidade

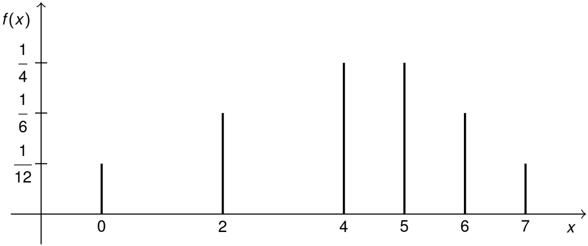
- Seja X uma variavel aleatória discreta.
- A função de probabilidade de X é uma função f que associa a cada valor possível x de X a sua probabilidade:

$$f(x) = P(X = x),$$

e tem as seguintes propriedades:

- $f(x) \geq 0$;
- $\sum_{x} f(x) = 1$.

Exemplo 3.2: A figura seguinte mostra o gráfico (diagrama de traços) de uma função de probabilidade de uma variável aleatória discreta.



A representação tabular da função de probabilidade é:

x	0	2	4	5	6	7
$f(x)$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{12}$

Função de distribuição acumulada

- Seja X uma variavel aleatória discreta com função de probabilidade f.
- A sua função de distribuição acumulada F está definida para qualquer valor real x e e dada por:

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

onde u toma todos os valores possíveis da v.a. X não superiores a x.

Propriedades da função de distribuição acumulada

- É uma função com descontinuidades e em escada;
- Se $x \leq y$, então $F(x) \leq F(y)$ (função crescente em sentido lato);
- $\lim_{x \to -\infty} F(x) = 0$ e $\lim_{x \to +\infty} F(x) = 1$.

Exemplo 3.3: Uma máquina de jogo tem 200 bolas (110 azuis, 60 brancas e 30 verdes), que são baralhadas permanentemente, à espera que alguém introduza 1€ na ranhura e carregue num botão. Nesse momento, é selecionada aleatoriamente uma das bolas e lançada para o exterior. Entretanto, essa bola é substituída automaticamente, na máquina, por outra da mesma cor. O jogador tem direito a um prémio de 1€, se a bola for branca, e de 2€, se for verde. Determine:

- A função de probabilidade do valor do prémio obtido por um jogador.
- A correspondente função de distribuição acumulada.

Resolução: Estamos interessados no prémio conseguido pelo jogador que pode ser 0€, 1€ ou 2€. Assim, verifica-se que a v.a. é:

$$X = \text{valor do prémio conseguido por um jogador.}$$

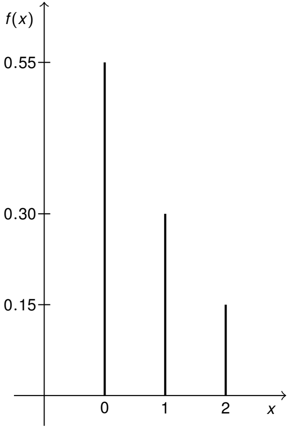
Exemplo 3.3 (Cont.):

Função de probabilidade:

$$f(0) = P(X = 0) = \frac{110}{200} = 0.55,$$

$$f(1) = P(X = 1) = \frac{60}{200} = 0.30,$$

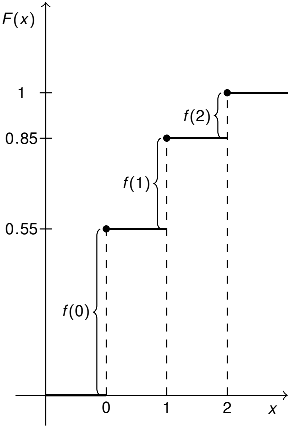
$$f(2) = P(X = 2) = \frac{30}{200} = 0.15.$$



Exemplo 3.3 (Cont.):

Função de distribuição acumulada:

$$F(x) = \begin{cases} 0, & x < 0 \\ 0.55, & 0 \leq x < 1 \\ 0.85, & 1 \leq x < 2 \\ 1, & x \geq 2 \end{cases}$$



Média ou valor esperado de uma variável aleatória discreta

- A cada distribuição de probabilidade podemos associar certas constantes que fornecem informação relevante sobre a distribuição.
- A média ou valor esperado de uma variavel aleatória discreta indica onde a distribuição de probabilidade está centrada.
- A média da variavel aleatória discreta X é a constante:

$$\mu = \mu_X = E(X) = \sum_x x\, f(x)$$

Seja X uma v.a. com função de probabilidade f. A média ou valor esperado da variavel aleatória discreta $g(X)$ é a constante:

$$E[g(X)] = \sum_x g(x) f(x).$$

Exemplo 3.4: O estabelecimento *Mais Desporto em Casa* comercializa máquinas de musculação e outros artigos desportivos. Suponha que X é a v.a. que representa o número de máquinas de musculação que esta empresa vende por semana e que tem a seguinte função de probabilidade:

x	7	8	9	10	11	12
$f(x)$	$\frac{1}{20}$	$\frac{1}{10}$	$\frac{1}{4}$	$\frac{2}{5}$	$\frac{3}{20}$	$\frac{1}{20}$

- Determine o número médio de máquinas de musculação vendidas por semana.
- Sabe-se que o lucro semanal (euros) conseguido na comercialização destas máquinas é dado por $L(X) = 60X - 50$. Determine o seu valor esperado.

Exemplo 3.4 (Cont.):

Resolução:

- O número médio de máquinas de musculação vendidas por semana é dado por:

$$\mu = E(X) = \sum_{x=7}^{12} x\, f(x) =$$

$$= 7 \times \frac{1}{20} + 8 \times \frac{1}{10} + 9 \times \frac{1}{4} + 10 \times \frac{2}{5} + 11 \times \frac{3}{20} + 12 \times \frac{1}{20} =$$

$$= \frac{193}{20} \approx 9.7 \text{ máquinas.}$$

- O valor esperado do lucro semanal (euros) conseguido na comercialização destas máquinas é dado por:

$$E[L(X)] = \sum_{x=7}^{12} L(x) f(x) = \sum_{x=7}^{12} (60x - 50) f(x) = 529 \text{ euros.}$$

Variabilidade de uma variavel aleatória discreta

- A variância da v.a. discreta X é a constante

$$\mathrm{var}(X) = \sigma^2 = \sigma_X^2 = E\left[(X-\mu)^2\right] = \sum_x (x-\mu)^2 f(x).$$

- O desvio padrão de uma v.a. X é a constante

$$\sigma = \sigma_X = \sqrt{E\left[(X-\mu)^2\right]} = \sqrt{\mathrm{var}(X)}.$$

Teorema: A variância da v.a. X é a constante

$$\mathrm{var}(X) = \sigma^2 = \sigma_X^2 = E\left[X^2\right] - \mu^2.$$

Distribuições discretas: Binomial e Poisson

Problema:

- Considere-se uma máquina que produz determinado tipo de peças de automóvel, de modo independente.
- Sabe-se que a probabilidade de a máquina produzir uma peça com defeito é $p = 0.2$.
- Considere-se que se analisou $n = 30$ peças produzidas pela máquina.
- Qual a probabilidade de haver exatamente 5 peças defeituosas nas 30 peças analisadas?

Como resolver?

- Definir a v.a. X: $X =$ Número de peças defeituosas, em 30.
- Calcular $P(X = 5)$.

Processo de Bernoulli

Um Processo de Bernoulli ou sequência de tentativas de Bernoulli é qualquer experiência estatística, com as seguintes propriedades:

- Consiste em n tentativas repetidas.
- Cada tentativa tem dois resultados possíveis: sucesso ou insucesso.
- A probabilidade de sucesso p é a mesma em qualquer tentativa.
- As tentativas repetidas são independentes e, portanto a probabilidade de sucesso não é afetada pelo possível conhecimento do resultado obtido em tentativas anteriores.

Distribuição binomial

- A distribuição binomial está associada a um processo de Bernoulli com n tentativas independentes com probabilidade p de sucesso em cada tentativa.
- Esta distribuição é muito aplicada em amostragem e em situações em que conhecemos o tamanho da amostra e sabemos a probabilidade associada ao acontecimento em estudo.
- A aplicação genérica mais importante refere-se à contagem do número de sucessos em n tentativas de Bernoulli com probabilidade p de sucesso em cada tentativa. Por exemplo, a contagem do número de artigos defeituosos num lote de tamanho n.

Distribuição binomial

- Definição da v.a.: X v.a. que representa o número de sucessos em n tentativas de Bernoulli com probabilidade p de sucesso em cada tentativa.
- Notação: $X \sim Bi(n, p)$ indica que a v.a. X tem distribuição binomial com parâmetros n e p.
- Função de probabilidade:

$$f(x) = \begin{cases} {}^nC_x\, p^x (1-p)^{n-x}, & x \in \{0, 1, ..., n\} \\ \\ 0, & x \notin \{0, 1, ..., n\} \end{cases} .$$

  em que ${}^nC_x$ representa o número de maneiras distintas de se conseguir x sucessos em n tentativas.
- Parâmetros: $n \in \mathbb{N}$ e $p \in]0, 1[$.
- Média: $E(X) = \mu_X = np$.
- Variância: $\text{var}(X) = \sigma_X^2 = np(1-p)$.

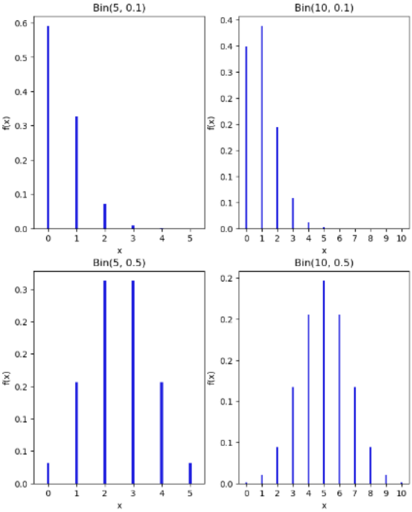
Função de distribuição acumulada de uma distribuição binomial de parâmetros n e p.

$$F(x) = \begin{cases} 0, & x < 0 \\ \sum_{i=0}^{Int[x]} p^i (1-p)^{n-i}, & 0 \leq x \leq n \\ 1, & x > n \end{cases}.$$

Distribuição Binomial

Comandos Python:

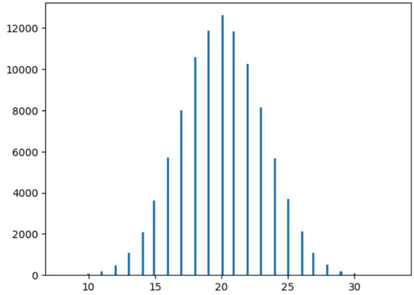
```
from scipy import stats
import seaborn as sns
(p, num) = (0.5, 5)
binomDist = stats.binom(num, p)
# gera 5 números inteiros
# sequenciais
x = np.arange(0, 6)
y = binomDist.pmf(x)
sns.barplot(x = x, y = y, width=0.1)
ax.set_xlabel('x')
ax.set_ylabel('f(x)')
ax.set_title('Bin(5, 0.5)')
```



Simulação de 100000 observações de uma distribuição Binomial, $Bi(n = 40, p = 0.5)$.

Comandos Python:

```
from scipy import stats
(p, num) = (0.5, 40)
binomDist = stats.binom(num, p)
n = 100000
trials = binomDist.rvs(n)
plt.hist(trials, bins = 'auto')
```



Distribuição Binomial

Exemplo 3.5: A Sara e a Catarina são amigas que gostam muito de jogar xadrez, mas a Sara é uma grande especialista e ganha em 60% dos jogos. Assim, resolveram efetuar, no próximo fim de semana, um campeonato de 15 jogos. Qual é a probabilidade da Sara ganhar:

- exatamente em 10 jogos?
- pelo menos 10 jogos?
- entre 4 e 8 jogos?

Resolução:

Seja X uma v.a. tal que:

$$X = \text{número de jogos que a Sara vence.}$$

$$X \sim Bi(n = 15, p = 0.6).$$

Exemplo 3.5 (Cont):

- A probabilidade da Sara ganhar exatamente em 10 jogos é $P(X = 10)$:

  Comandos Python:

  ```
  from scipy import stats
  p1 = stats.binom.pmf(10, 15, 0.6)
  print(f'A probabilidade de a Sara ganhar é {p1:.3f}')
  ```

  Output: A probabilidade de a Sara ganhar é 0.186.

- A probabilidade da Sara ganhar em pelo menos 10 jogos é $P(X \geq 10) = 1 - P(X \leq 9)$:

  Comandos Python:

  ```
  from scipy import stats
  p2 = 1- stats.binom.cdf(9, 15, 0.6)
  print(f'A probabilidade de a Sara ganhar é {p2:.3f}')
  ```

  Output: A probabilidade de a Sara ganhar é 0.403.

Exemplo 3.5 (Cont):

- A probabilidade da Sara ganhar entre 4 e 8 jogos é
  $P(4 \leq X \leq 8) = P(X \leq 8) - P(X \leq 3)$:

  Comandos Python:

```
from scipy import stats
p3 = stats.binom.cdf(8, 15, 0.6) - stats.binom.cdf(3, 15, 0.6)
print(f'A probabilidade de a Sara ganhar é {p3:.3f}')
```

  Output: A probabilidade de a Sara ganhar é 0.388.

Problema:

- Considere-se uma repartição de finanças. Os clientes chegam à repartição de modo independente uns dos outros.
- Sabe-se que, em média, chegam à repartição de finanças $\lambda = 5$ pessoas em cada 15 minutos.
- Qual a probabilidade de não chegar nenhum cliente entre as 10:00h e as 10:15h?

Como resolver?

- Definir a v.a. X: $X =$ Número de clientes que chegam à repartição de fianças, num intervalo de tempo de 15 minutos.
- Calcular $P(X = 0)$.

Processo de Poisson

Um Processo de Poisson refere-se, normalmente, ao número de eventos que ocorrem num intervalo temporal ou numa região espacial. Tem as seguintes propriedades:

- Não tem memória, ou seja o número de eventos que ocorre num intervalo é independente do que ocorre noutro intervalo disjunto.
- A probabilidade de ocorrer um evento num intervalo muito pequeno é proporcional ao comprimento do intervalo.
- A probabilidade de ocorrer mais do que um evento num intervalo muito pequeno é nula.

Distribuição de Poisson

- A distribuição de Poisson baseia-se num processo de Poisson.
- Usa-se a distribuição de Poisson quando estamos interessados na contagem do número de eventos que ocorrem num intervalo temporal ou numa região espacial, quando os eventos são independentes uns ds outros.
- Por exemplo, costuma usar-se a distribuição de Poisson quando estamos interessados no:
    - número de chamadas telefónicas que chegam a uma central num intervalo de tempo;
    - número de erros que existem num livro;
    - número de peixes que existe num lago;
    - número de artigos encomendados num armazém;
    - número de artigos existentes num lote de tamanho aleatório.

Distribuição de Poisson

- Definição da v.a.: Seja $\lambda > 0$ o número médio de eventos que correm num dado intervalo de tempo (ou numa região espacial) e seja X uma v.a. que representa o número de eventos que ocorrem nesse intervalo de tempo (ou nessa região).
- Notação: $X \sim Poisson(\lambda)$ indica que a v.a. X tem distribuição de Poisson com parâmetro λ.
- Função de probabilidade:

$$f(x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^x}{x!}, & x \in \{0, 1, 2, ...\} \\ 0, & x \notin \{0, 1, 2, ...\} \end{cases}.$$

- Parâmetros: $\lambda > 0$.
- Média: $E(X) = \mu_X = \lambda$.
- Variância: $\text{var}(X) = \sigma_X^2 = \lambda$.

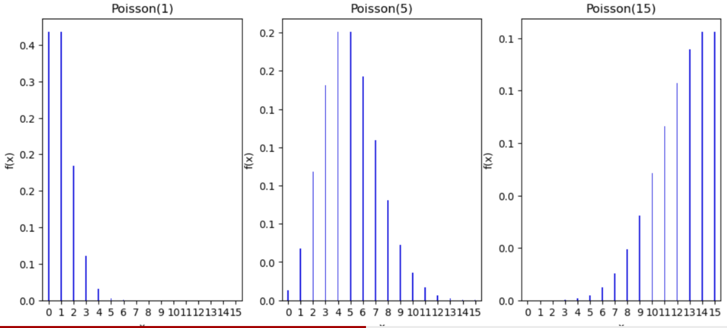
Função de distribuição acumulada de uma distribuição de Poisson de parâmetro λ.

$$F(x) = \begin{cases} 0, & x < 0 \\ \\ e^{-\lambda} \displaystyle\sum_{i=0}^{Int[x]} \frac{\lambda^i}{i!}, & x \geq 0 \end{cases}.$$

Distribuição de Poisson

Comandos Python:

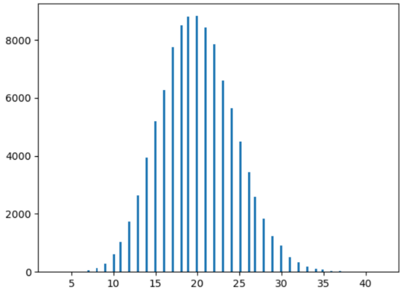
```
from scipy import stats, import seaborn as sns
poissonDist = stats.poisson(1)
x = np.arange(0, 16), y = poissonDist.pmf(x)
sns.barplot(x = x, y = y, width=0.1)
ax.set_xlabel('x'), ax.set_ylabel('f(x)'), ax.set_title('Poisson(1)')
```



Simulação de 100000 observações de uma distribuição de Poisson, *Poisson* $(\lambda = 20)$.

Comandos Python:

```
from scipy import stats
lambda = 20
poissonDist = stats.poisson(lambda)
n = 100000
trials = poissonDist.rvs(n)
plt.hist(trials, bins = 'auto')
```



Distribuição de Poisson

Exemplo 3.6: O número de chamadas que chegam à central telefónica de uma associação de defesa do consumidor é uma v. a. de Poisson com média 1.5 chamadas em cada 10 minutos.

- Considere o período entre as $9:00$h e as $09:10$h. Determine a probabilidade de a associação.
  - Não receber qualquer chamada.
  - Receber mais de duas chamadas.

- Considere o período entre as $11:00$h e as $11:30$h. Determine a probabilidade de a associação:
  - Não receber qualquer chamada.
  - Receber mais de duas chamadas.

Distribuição de Poisson

Exemplo 3.6(Cont.):

Resolução:

Considere-se $\lambda =$ número médio de chamadas que chegam à central telefónica em cada 10 minutos.

As v.a. de interesse são:

- $X_{10} =$ número chamadas que chegam à central telefónica no intervalo de tempo de 10 minutos.

$$X_{10} \sim Poisson(\lambda = 1.5)$$

- $X_{30} =$ número chamadas que chegam à central telefónica no intervalo de tempo de 30 minutos.

$$X_{30} \sim Poisson(3 \times \lambda = 4.5)$$

Exemplo 3.6 (Cont): No período entre as $09:00$h e as $09:10$h:

- A probabilidade de a associação não receber qualquer chamada é $P(X_{10} = 0)$:

  Comandos Python:

```
from scipy import stats
p1 = stats.poisson.pmf(0, 1.5)
print(f'A prob. de não receber qualquer chamada é {p1:.3f}')
```

  Output: A prob. de não receber qualquer chamada é 0.223.

- A probabilidade de a associação receber mais de duas chamadas é $P(X_{10} > 2) = 1 - P(X_{10} \leq 2)$:

  Comandos Python:

```
from scipy import stats
p2 = 1- stats.poisson.cdf(2, 1.5)
print(f'A prob. de receber mais de duas chamadas um é {p2:.3f}')
```

  Output: A prob. de receber mais de duas chamadas um é 0.191.

Exemplo 3.6 (Cont): No período entre as $11:00$h e as $11:30$h:

- A probabilidade de a associação não receber qualquer chamada é $P(X_{30} = 0)$:

  Comandos Python:

  ```
  from scipy import stats
  p3 = stats.poisson.pmf(0, 4.5)
  print(f'A prob. de não receber qualquer chamada é {p3:.3f}')
  ```

  Output: A prob. de não receber qualquer chamada é 0.011.

- A probabilidade de a associação receber mais de duas chamadas é $P(X_{30} > 2) = 1 - P(X_{30} \leq 2)$:

  Comandos Python:

  ```
  from scipy import stats
  p4 = 1 - stats.poisson.cdf(2, 4.5)
  print(f'A prob. de receber mais de duas chamadas um é {p4:.3f}')
  ```

  Output: A prob. de receber mais de duas chamadas um é 0.826.

Variáveis aleatórias contínuas

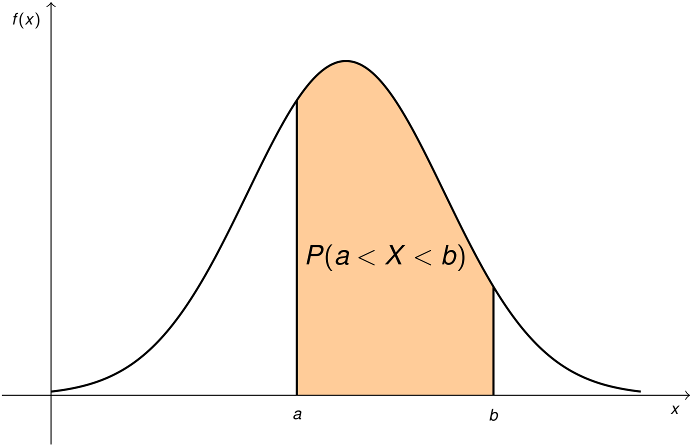
Função densidade de probabilidade (f.d.p.)

- Seja X uma variavel aleatória contínua.
- A função densidade de probabilidade de X é uma função f tal que:

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx, \forall a, b \in \mathbb{R} : a \leq b,$$

  e tem as seguintes propriedades:
  - $f(x) \geq 0, \forall x \in \mathbb{R}$;
  - $\displaystyle\int_{-\infty}^{+\infty} f(x)\,dx = 1.$
- $P(X = x) = 0, \forall x \in \mathbb{R}$, ou seja, uma v.a. contínua tem probabilidade nula de tomar exatamente qualquer um dos seus valores.
- $P(a \leq X \leq b) = P(a < X < b) + P(X = a) + P(X = b) =$
  $= P(a < X < b)$.

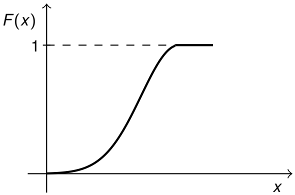
Função densidade de probabilidade (f.d.p.)

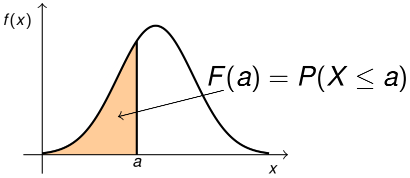


Função de distribuição acumulada

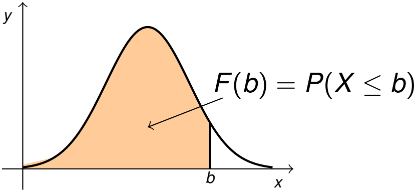
- Seja X uma variavel aleatória contínua com função densidade de probabilidade f.
- A sua função de distribuição acumulada F está definida para qualquer valor real x e é dada por:

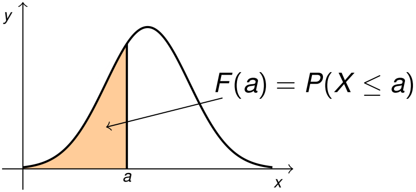
$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)\, du.$$

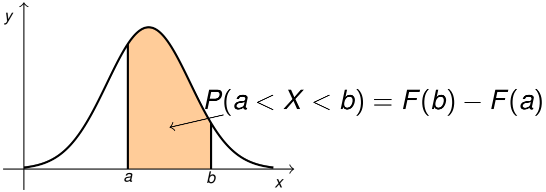




Função de distribuição acumulada







Proprieades da função de distribuição acumulada F da variável aleatória contínua X

- É uma função contínua;
- Se $x \leq y$, então $F(x) \leq F(y)$ (função crescente em sentido lato);
- $\lim_{x \to -\infty} F(x) = 0$;
- $\lim_{x \to +\infty} F(x) = 1$;
- $f(x) = \dfrac{dF(x)}{dx}$, se a função de distribuição acumulada for derivável;
- $P(a \leq X \leq b) = F(b) - F(a)$.

Exemplo 3.7: Numa experiência laboratorial, mede-se a temperatura (^oC) de uma reação química. O erro dessa medição é uma variável aleatória contínua X com a seguinte função densidade de probabilidade:

$$f(x) = \begin{cases} \frac{3}{4}(1 - x^2), & x \in [-1, 1] \\ 0, & x \notin [-1, 1] \end{cases} .$$

- Verifique que $\displaystyle\int_{-\infty}^{+\infty} f(x)\, dx = 1$.
- Determine a função de distribuição acumulada de X.
- Determine $P(0.5 < X \leq 1)$.

Exemplo 3.7 (Cont.):

Resolução:

- $$\int_{-\infty}^{+\infty} f(x)\,dx = \int_{-1}^{1} \frac{3}{4}(1 - x^2)\,dx = \frac{3}{4}\left[x - \frac{x^3}{3}\right]_{-1}^{1} = 1.$$

Comandos Python:

```
import scipy.integrate as integrate
result = integrate.quad(lambda x: (3/4)*(1-x**2), -1, 1)
print(f'O resultado do integral é result[0]:.1f.')
```

Output: O resultado do integral é 1.0.

Exemplo 3.7 (Cont.):

- Para $-1 \leq x \leq 1$, temos

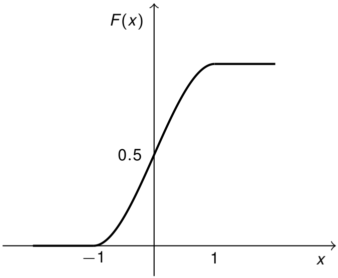
$$F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{-1}^{x} \frac{3}{4}(1 - t^2)\,dt =$$

$$= \frac{3}{4}\left[t - \frac{t^3}{3}\right]_{-1}^{x} = \frac{2 + 3x - x^3}{4}.$$

Então,

$$F(x) = \begin{cases} 0, & x < -1 \\ \dfrac{2 + 3x - x^3}{4}, & -1 \leq x \leq 1 \\ 1, & x > 1 \end{cases}.$$

Exemplo 3.7 (Cont.):



- Usando o resultado da alínea anterior:

$$P(0.5 < X \leq 1) = F(1) - F(0.5) = \frac{5}{32}.$$

Média ou valor esperado de uma variável aleatória contínua

- A cada distribuição de probabilidade podemos associar certas constantes que fornecem informação relevante sobre a distribuição.
- A média ou valor esperado de uma variavel aleatória contínua indica onde a distribuição de probabilidade está centrada.
- A média da variavel aleatória contínua X é a constante:

$$\mu = \mu_X = E(X) = \int_{-\infty}^{+\infty} x\, f(x)\, dx.$$

Seja X uma v.a. com função densidade de probabilidade f. A média ou valor esperado da variavel aleatória contínua $g(X)$ é a constante:

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

Variabilidade de uma variavel aleatória contínua

- A variância da v.a. contínua X é a constante

$$\text{var}(X) = \sigma^2 = \sigma_X^2 = E\left[(X-\mu)^2\right] = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)\, dx.$$

- O desvio padrão de uma v.a. X é a constante

$$\sigma = \sigma_X = \sqrt{E\left[(X-\mu)^2\right]} = \sqrt{\text{var}(X)}.$$

Teorema: A variância da v.a. X é a constante

$$\text{var}(X) = \sigma^2 = \sigma_X^2 = E\left[X^2\right] - \mu^2.$$

- A média e o desvio padrão são medidas que dependem das unidades das observações, dificultando a comparação entre distribuições de duas ou mais v.a. com unidades diferentes e não relacionadas. Por exemplo, quando uma das v.a. representa uma distância e a outra um tempo.

- O procedimento analítico que nos permite comparar tais distribuições é a estandardização das v.a..

Estandardização de variáveis aleatórias

Seja X uma v.a. com média μ_X e desvio padrão σ_X. Então,

$$Z = \frac{X - \mu_X}{\sigma_X}$$

é uma v.a. estandardizada, tal que $E(Z) = 0$ e $\mathrm{var}(Z) = 1$.

Desigualdade de Chebyshev

Seja X uma v.a. com média μ e variância σ^2 finitas e $k \in \mathbb{R}^+$. Então,

$$P(|X - \mu| \leq k\sigma) = P(\mu - k\sigma \leq X \leq \mu + k\sigma) \geq \frac{1}{k^2}$$

ou, de forma equivalente,

$$P\left(-k \leq \frac{X - \mu}{\sigma} \leq k\right) \geq \frac{1}{k^2}$$

Assim, a desigualdade de Chebyshev dá uma estimativa da probabilidade de uma v.a. tomar um valor distanciado da sua média menos de k desvios padrão. Este resultado é muito útil porque nos permite estimar tais probabilidades mesmo que não se conheça a distribuição da variável aleatória.

Distribuições contínuas: Uniforme, Exponencial e Normal

Distribuição Uniforme

- A distribuição uniforme, $U(a, b)$, é usada para representar uma quantidade que varia aleatoriamente no intervalo $[a, b]$ e cuja probabilidade de tomar valores num subintervalo de $[a, b]$ é proporcional ao seu comprimento.

- Tal é o caso dos números aleatórios, que são concretizações de uma v.a. $U(0, 1)$.

- Esta distribuição também é usada como um primeiro modelo para uma quantidade que se julga varia aleatoriamnete entre a e b, mas acerca da qual pouco mais é sabido.

Distribuição uniforme

- Definição da v.a.: X v.a. tal que a sua probabilidade de tomar um valor num subintervalo de $[a, b]$ é proporcional ao comprimento desse subintervalo.
- Notação: $X \sim U(a, b)$ indica que a v.a. X tem distribuição uniforme em $[a, b]$.
- Função densidade de probabilidade:

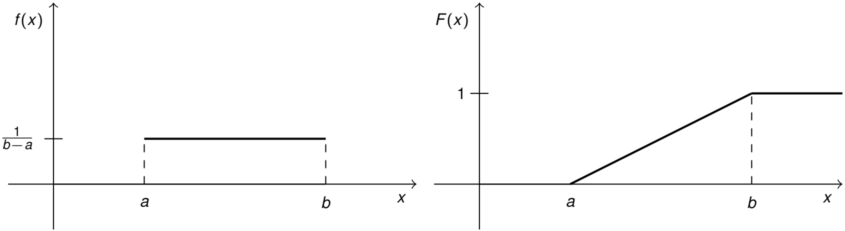
$$f(x) = \begin{cases} \dfrac{1}{b-a}, & x \in [a, b] \\ 0, & x \notin [a, b] \end{cases} .$$

- Parâmetros: a, b.
- Média: $E(X) = \mu_X = \dfrac{a+b}{2}$.
- Variância: $\mathrm{var}(X) = \dfrac{(b-a)^2}{12}$.

Função de distribuição acumulada de uma distribuição uniforme em $[a, b]$

$$F(x) = \begin{cases} 0, & x < a \\ \dfrac{x-a}{b-a}, & x \in [a, b] \\ 1, & x > b \end{cases}.$$

Função de densidade de probabilidade e distribuição acumulada de uma distribuição uniforme em $[a, b]$



Exemplo 3.8: Considere o processo de selecionar números reais, ao acaso no intervalo [a, b].

- Suponha que se seleciona um número no intervalo $[5, 7]$. Determine a probabilidade do número estar compreendido entre:
  - 6.30 e 6.95.
  - 6.50 e 7.50.

- Suponha que se selecionam dois números reais no intervalo $[0, 1]$. Determine a probabilidade de:
  - Ambos serem maiores do que 0.75.
  - Um ser maior do que 0.8 e o outro menor do que 0.5.

Exemplo 3.8 (Cont): Seja X a v.a. contínua tal que $X \sim U(5,7)$.

- A probabilidade do número estar compreendido entre 6.30 e 6.95 é $P(6.30 \leq X \leq 6.95)$.

  Analiticamente podemos obter o valor:

  $$P(6.30 \leq X \leq 6.95) = F(6.95) - F(6.3) =$$

  $$= \frac{6.95 - 5}{7 - 5} - \frac{6.3 - 5}{7 - 5} = 0.325.$$

  Comandos Python:

  ```
  from scipy import stats
  p1 = stats.uniform.cdf(6.95, 5, 2) - stats.uniform.cdf(6.30, 5, 2)
  print(f'A probabilidade de do número estar compreendido entre 6.30 e 6.95 é {p1:.3f}')
  ```

  Output: A probabilidade de do número estar compreendido entre 6.30 e 6.95 é 0.325.

Exemplo 3.8 (Cont): Seja X a v.a. contínua tal que $X \sim U(5, 7)$.

- A probabilidade do número estar compreendido entre 6.50 e 7.50 é $P(6.50 \leq X \leq 7.50)$.

  Analiticamente podemos obter o valor:

  $$P(6.50 \leq X \leq 7.50) = F(7.50) - F(6.50) =$$

  $$= 1 - \frac{6.50 - 5}{7 - 5} = 0.250.$$

  Comandos Python:

```
from scipy import stats
p2 = stats.uniform.cdf(7.50, 5, 2) - stats.uniform.cdf(6.50, 5, 2)
print(f'A probabilidade de do número estar compreendido entre
6.50 e 7.50 é {p2:.3f}')
```

  Output: A probabilidade de do número estar compreendido entre 6.50 e 7.50 é 0.250.

Exemplo 3.8 (Cont): Sejam X_i, $i = 1, 2$ v.a. contínuas tais que $X_i \sim U(0, 1)$.

- A probabilidade de ambas serem maiores do que 0.75 é $P(X_1 > 0.75 \text{ e } X_2 > 0.75)$.

  Analiticamente podemos obter o valor:

$$P(X_1 > 0.75 \text{ e } X_2 > 0.75) = P(X_1 > 0.75)P(X_2 > 0.75) =$$

$$= (1 - P(X_1 \leq 0.75))(1 - P(X_2 \leq 0.75)) = (0.25)^2 = 0.0625.$$

  Comandos Python:

```
from scipy import stats
p3 = (1 - stats.uniform.cdf(0.75, 0, 1))**2
print(f'A probabilidade de ambos serem maiores
do que 0.75 é {p3:.4f}')
```

  Output: A probabilidade de ambos serem maiores do que 0.75 é 0.0625.

Exemplo 3.8 (Cont): Sejam X_i, $i = 1, 2$ v.a. contínuas tais que $X_i \sim U(0, 1)$.

- A probabilidade de um ser maior do que 0.80 e o outro menor do que 0.50 é
  $P((X_1 > 0.80 \text{ e } X_2 < 0.50) \text{ ou } (X_1 < 0.50 \text{ e } X_2 > 0.80))$.

  Analiticamente podemos obter o valor:

$$P((X_1 > 0.80 \text{ e } X_2 < 0.50) \text{ ou } (X_1 < 0.50 \text{ e } X_2 > 0.80)) =$$

$$= 2 \times P((X_1 > 0.80 \text{ e } X_2 < 0.50) =$$

$$= 2 \times (1 - P((X_1 < 0.80))P(X_2 < 0.50) =$$

$$= 2 \times 0.20 \times 0.50 = 0.20.$$

Distribuição exponencial

- A distribuição exponencial é muito usada no estudo de filas de espera e de fiabilidade de sistemas complexos (fiabilidade no instante *t* é a probabilidade do sistema ainda funcionar nesse instante).
- Aparece, frequentemente, associada a um processo de Poisson.
- Enquanto a distribuição de Poisson serve como modelo probabilístico da contagem do número de eventos num intervalo de tempo ou região espacial, a distribuição exponencial usa-se como modelo probabilístico para representar o intervalo de tempo entre dois eventos independentes. Por exemplo, o tempo entre a chegada de clientes a um estabelecimento, ou o tempo entre ocorrência de avarias de uma máquina.

Distribuição exponencial

- Definição da v.a.: X v.a. representa o intervalo de tempo entre dois eventos independentes.
- Notação: $X \sim Exp(\beta)$ indica que a v.a. X tem distribuição exponencial de parâmetro β ($\frac{1}{\beta}$ representa o número médio de eventos que ocorrem por unidade de tempo ou região espacial).
- Função densidade de probabilidade:

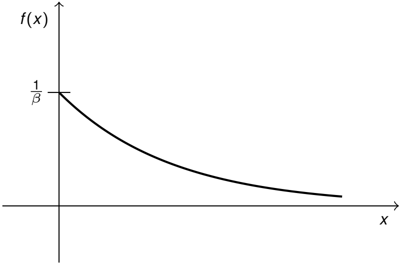
$$f(x) = \begin{cases} 0, & x < 0 \\ \\ \frac{1}{\beta} e^{-\frac{x}{\beta}}, & x \geq 0 \end{cases}.$$

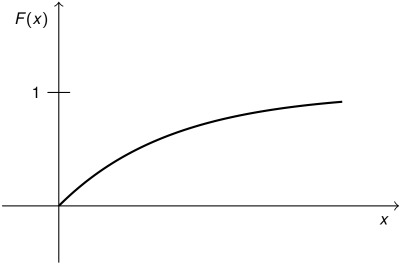
- Parâmetros: $\beta > 0$.
- Média: $E(X) = \mu_X = \beta$.
- Variância: $\text{var}(X) = \beta^2$.

Função de distribuição acumulada de uma distribuição exponencial de parâmetro β

$$F(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-\frac{x}{\beta}} & x \geq 0 \end{cases}.$$

Função de densidade de probabilidade e distribuição acumulada de uma distribuição exponencial de parâmetro β



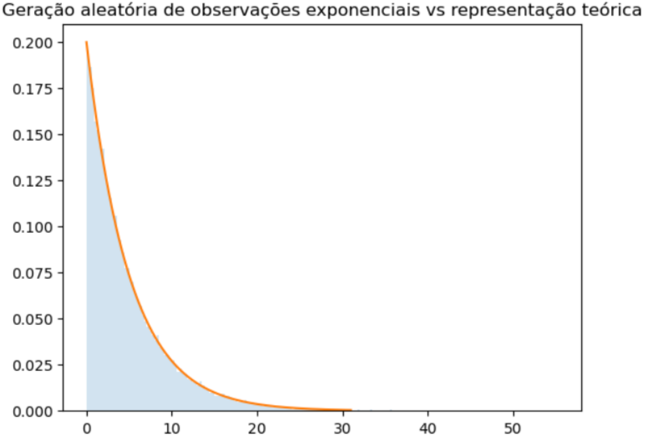


Geração aleatória de 100000 observações exponenciais vs representação teórica de uma distribuição exponencial, *Exp*$(\beta = 5)$.

Comandos Python:

```
from scipy import stats
myBeta = 5, n = 100000
# Se β = 1, usamos stats.expon.rvs(t)
# Se β ≠ 1, usamos stats.gamma.rvs(t, a = 1, scale = beta)
trials = stats.gamma.rvs(a = 1, scale = myBeta, size = n)
plt.hist(data, density=True, bins='auto', histtype='stepfilled', alpha=0.2)
# Gráfico da distribuição teórica
t = np.arange(0,31,0.02)
# Se β = 1, usamos stats.expon.pdf(t),
# Se β ≠ 1, usamos stats.gamma.pdf(t, a = 1, scale = beta)
y = stats.gamma.pdf(t, a = 1, scale = myBeta)
plt.plot(t,y)
plt.title('Geração aleatória de observações exponenciais vs
representação teórica')
```

Geração aleatória de 100000 observações exponenciais vs representação teórica de uma distribuição exponencial, *Exp*($\beta = 5$).



Exemplo 3.8: Os acidentes que ocorrem num estaleiro de construção naval são estatisticamente independentes e o número de acidentes tem uma distribuição de Poisson com média de 6 por mês. Considere a v.a. T que representa o intervalo de tempo entre acidentes consecutivos. Suponha que o estaleiro está em laboração contínua.

- Identifique a distribuição T.
- Qual a probabilidade de não ocorrerem acidentes na próxima semana?
- Qual a probabilidade de ocorrerem acidentes em, pelo menos, 4 das próximas 5 semanas?

Exemplo 3.8 (Cont.):

- Identifique a distribuição T.

  $T =$ tempo entre acidentes consecutivos (dias).

  $\lambda = \dfrac{6}{30} = 0.2$ é o número médio de acidentes por dia.

  $\beta = \dfrac{1}{\lambda} = 5$ é o tempo médio entre acidentes (dias).

  Então:

$$T \sim Exp(\beta = 5)$$

Exemplo 3.8 (Cont.):

- A probabilidade de não ocorrerem acidentes na próxima semana é $P(T > 7) = 1 - P(T \leq 7)$.

  Podemos obter o valor da probabilidade analiticamente:

  $$P(T > 7) = 1 - P(T \leq 7) = 1 - F(7) =$$

  $$= 1 - (1 - e^{-\frac{7}{5}}) = e^{-\frac{7}{5}} = 0.247.$$

  Comandos Python:

```
from scipy import stats
p = 1 - stats.gamma.cdf(7, a = 1, scale = 5)
print(f'A probabilidade de de não ocorrerem acidentes na
próxima semana é {p:.3f}')
```

  Output: A probabilidade de de não ocorrerem acidentes na próxima semana é 0.247.

Exemplo 3.8 (Cont.):

- Qual a probabilidade de ocorrerem acidentes em, pelo menos, 4 das próximas 5 semanas?

  $Y =$ número de semanas, em 5, em que ocorrem acidentes.
  $p =$ probabilidade de ocorrerem acidentes numa semana.
  $p = P(T \leq 7) = 1 - e^{-\frac{7}{5}} = 0.753.$
  $Y \sim Bi(n = 5, p = 0.735).$

  Logo a probabilidade pedida é $P(Y \geq 4) = P(Y = 4) + P(Y = 5)$.

  Comandos Python:

```
from scipy import stats
p1 = stats.binom.pmf(4, 5, p) + stats.binom.pmf(5, 5, p)
print(f'A probabilidade de ocorrerem acidentes em,
pelo menos, 4 das próximas 5 semanas é {p1:.3f}')
```

  Output: A probabilidade de ocorrerem acidentes em, pelo menos, 4 das próximas 5 semanas é 0.640.

Distribuição Normal

- A Distribuição normal é a mais importante das distribuições contínuas, nomeadamente por:
  - ser um modelo adequado para representar muitos dos fenómenos do mundo real, nomeadamente características relacionadas com medições;
  - ser muito usada na inferencia estatística;
  - muitas técnicas desenvolvidas em Estatística são exatas no caso de distribuições normais;
  - algumas variaveis aleatórias (como por exemplo, a binomial e a de Poisson) podem ser aproximadas por uma v.a. normal.
- Usa-se a distribuição normal como modelo para representar
  - quantidades que sejam a soma de um grande número de outras quantidades;
  - características da população relacionadas com medições ou os respectivos erros.

Distribuição Normal

- Notação: $X \sim N(\mu, \sigma^2)$ indica que a v.a. X tem distribuição normal com média μ e variância σ^2.
- Função densidade de probabilidade:

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

- Parâmetros: Média μ e variância σ^2.
- Média: $E(X) = \mu$.
- Variância: $\text{var}(X) = \sigma^2$.

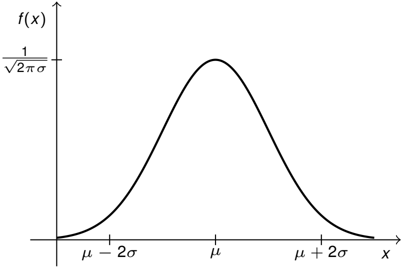
Distribuição normal estandardizada

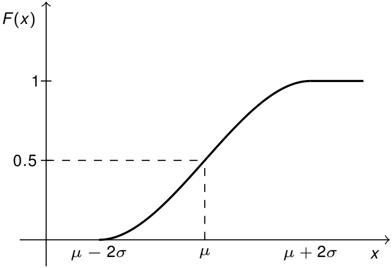
$$X \sim N(\mu, \sigma^2) \Rightarrow Z = \frac{X - \mu}{\sigma} \sim N(0, 1).$$

Função de distribuição acumulada de uma distribuição normal de média e variância σ^2

$$F(x) = P(X \leq x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dt$$

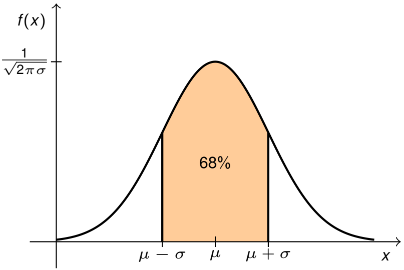
Função densiddae de probabilidade e função de distribuição acumulada de uma distribuição normal de média e variância σ^2





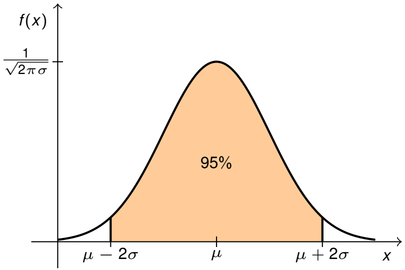
- Na distribuição normal proximadamente 68% da população difere da média menos de 1 desvio padrão:

$$P(\mu - \sigma < X < \mu + \sigma) \approx 0.68.$$



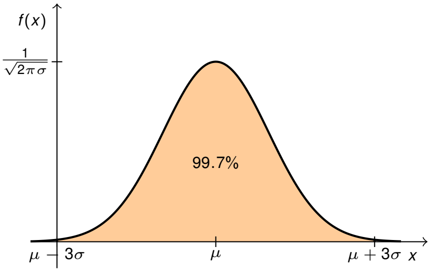
- Na distribuição normal proximadamente 95% da população difere da média menos de 2 desvios padrão:

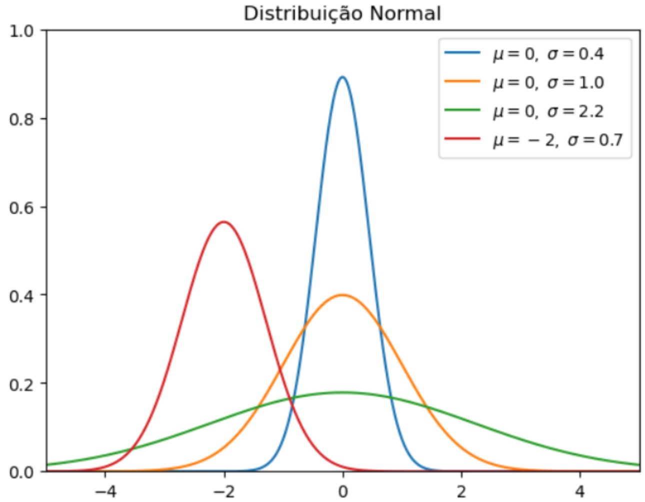
$$P(\mu - 2\sigma < X < \mu + 2\sigma) \approx 0.95.$$



- Na distribuição normal proximadamente 99.7% da população difere da média menos de 3 desvios padrão:

$$P(\mu - 3\sigma < X < \mu + 3\sigma) \approx 0.997.$$




Distribuição Normal
1.0
0.8
0.6
0.4
0.2
0.0
−4
−2
0
2
4
$\mu = 0,\ \sigma = 0.4$
$\mu = 0,\ \sigma = 1.0$
$\mu = 0,\ \sigma = 2.2$
$\mu = -2,\ \sigma = 0.7$

Aditividade da distribuição normal

Sejam X_1 e X_2 duas variáveis aleatórias independentes que seguem distribuições normais tais que:

$$X_1 \sim N(\mu_1, \sigma_1^2) \qquad \text{e} \qquad X_2 \sim N(\mu_2, \sigma_2^2).$$

Então,

$$X_1 \pm X_2 \sim N(\mu_1 \pm \mu_2, \sigma_1^2 + \sigma_2^2).$$

No caso de n variáveis aleatórias normais independentes , tais que $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, ..., n$, verifica-se:

$$X = \sum_{i=1}^{n} X_i \sim N\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right).$$

Exemplo 3.10: A duração T de um relógio de determinado modelo, antes de avariar, é uma v.a. aproximadamente normal, com média de 11 anos e desvio padrão de 1 ano. O fabricante pretende oferecer um período de garantia, dentro do qual os relógios avariados são substituídos por relógios novos. Esse período deverá ser tão grande quanto possível, mas sem que os custos se tornem incomportáveis.

- Qual a probabilidade de um relógio durar mais de 11 anos?
- Qual a probabilidade de um relógio durar menos de 10 anos?
- Qual deverá ser o período de garantia, se a fábrica não pretender substituir mais do que 5% dos relógios?
- Considere um cliente que comprou 5 relógios.
  - Qual a probabilidade de pelo menos dois deles durarem mais de 12 anos?
  - Qual a probabilidade de pelo menos um deles durar menos de 10 anos?

Exemplo 3.10 (Cont.): $T \sim N(11, 1)$.

- A probabilidade de um relógio durar mais de 11 anos é $P(T > 11) = 1 - P(T \leq 11)$.

  Comandos Python:

```
from scipy import stats
p1 = 1- stats.norm.cdf(11, 11, 1)
print(f'A prob. de um relógio durar mais de 11 anos é {p1:.3f}')
```

  Output: A prob. de um relógio durar mais de 11 anos é 0.500.

- A probabilidade de um relógio durar menos de 10 anos é $P(T < 10)$.

  Comandos Python:

```
from scipy import stats
p2 = stats.norm.cdf(10, 11, 1)
print(f'A prob. de um relógio durar menos de 10 anos é {p2:.3f}')
```

  Output: A prob. de um relógio durar menos de 10 anos é 0.159.

Exemplo 3.10 (Cont.):

- Qual deverá ser o período de garantia, se a fábrica não pretender substituir mais do que 5% dos relógios?

  Pretendemos descobrir o tempo t (em anos), tal que $P(T < t) = 0.05$.

  Comandos Python:

```
from scipy import stats
t = stats.norm.ppf(0.05, 11, 1)
print(f'O período de garantia, se a fábrica não pretender
substituir mais do que 5% dos relógios é t:.3f anos')
```

  Output: O período de garantia, se a fábrica não pretender substituir mais do que 5% dos relógios é 9.355 anos.

Exemplo 3.10 (Cont.): Considere um cliente que comprou 5 relógios.

- Qual a probabilidade de pelo menos dois deles durarem mais de 12 anos? Sejam,

  $X =$ número de relógios que duram mais de 12 anos

  $n = 5$, número de relógios comprados

  $p = P(T > 12) = 1 - P(T \leq 12) = 0.159$

  $X \sim Bi(n = 5, p = 0.159)$.

  A probabilidade pedida é dada por $P(X \geq 2) = 1 - P(X \leq 1)$.

  Comandos Python:

```
from scipy import stats
p = 1− stats.norm.cdf(12, 11, 1), p3 = 1 - stats.binom.cdf(1, 5, p)
print(f'A prob. de pelo menos 2 deles durarem mais de 12 anos
é {p3:.3f}')
```

  Output: A prob. de pelo menos 2 deles durarem mais de 12 anos é 0.181.

Exemplo 3.10 (Cont.): Considere um cliente que comprou 5 relógios.

- Qual a probabilidade de pelo menos um deles durar menos de 10 anos? Sejam,

  $X =$ número de relógios que duram menos de 10 anos

  $n = 5$, número de relógios comprados

  $p = P(T < 10) = 0.159$

  $X \sim Bi(n = 5, p = 0.159)$.

  A probabilidade pedida é dada por $P(X \geq 1) = 1 - P(X = 0)$.

  Comandos Python:

```
from scipy import stats
p = stats.norm.cdf(10, 11, 1), p4 = 1 - stats.binom.pmf(0, 5, p)
print(f'A prob. de pelo menos 1 deles durar menos de 10 anos
é {p4:.3f}')
```

  Output: A prob. de pelo menos 1 deles durar menos de 10 anos é 0.578.

Valor esperado de uma combinação linear de variáveis aleatórias

Sejam X_i variáveis aleatórias (discretas ou contínuas) e a_i constantes reais, para $i = 1, 2, ..., n$. Então:

$$E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E\left(X_i\right).$$

Em particular,

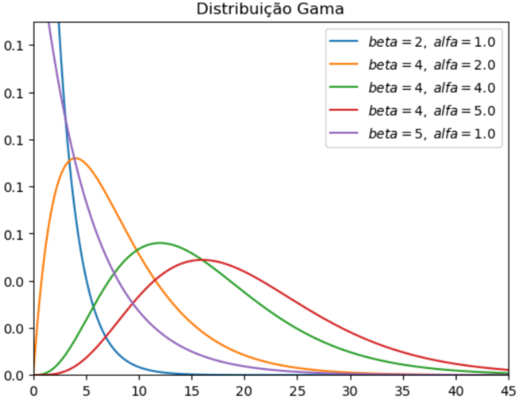
$$E\left(a_1 X_1 + a_2 X_2\right) = a_1 E(X_1) + a_2 E(X_2).$$

Breve referência às distribuições contínuas: Gama, Qui-quadrado e t-Student

Distribuição Gama, *Gama*(α, β)

A distribuição exponencial é um caso particular da distribuição gama.

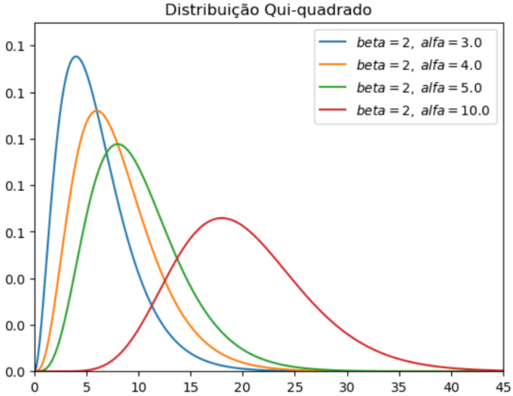
$$Gama(1, \beta) = Exp(\beta)$$



Distribuição Qui-quadrado, $\chi^2(\nu)$

A distribuição qui-quadrado é um caso particular da distribuição gama.

$$\chi^2(\nu) = Gama\left(\alpha = \frac{\nu}{2}, \beta = 2\right)$$



Distribuição t-Student, $T(\nu)$

A distribuição t-Student de parâmetro $\nu \in \mathbb{N}$, tem média zero para $\nu > 1$ e só tem variância para $\nu > 2$.