1. Conceitos gerais sobre distribuições de variáveis discretas

1.1 Função probabilidade

Função probabilidade, **f (x)** , é uma função matemática que associa a cada valor de uma variável aleatória quantitativa discreta a probabilidade de ele ocorrer (ou a percentagem de ocorrências na população).

Também é costume chamar-se a tal função a **distribuição de probabilidades** da variável X.

O seu domínio pode ser definido como o conjunto de valores possíveis para a variável, mas é habitualmente definido como sendo |R, atribuindo-se a todos os valores impossíveis a imagem 0.

$$\mathbf{f(x)}: \quad \mathbb{R} \longrightarrow [0,1]$$
$$x \longmapsto \mathbf{f(x) = P(X = x)}$$

Note que: $0 \leq f(x) \leq 1, \ \forall x_i$

e que: $\sum_{i=1}^{k} f(x_i) = 1$, em k é o número de valores possíveis para X.

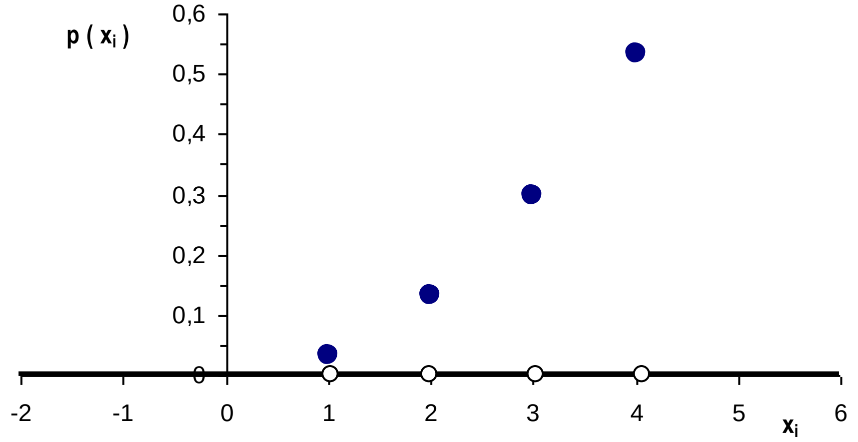
A função probabilidade pode ser definida:

- analiticamente através de uma ou várias expressões matemáticas (função definida por ramos),
- através de uma tabela,
- graficamente.

Exemplo:

$$p(x) = \begin{cases} \dfrac{x^2}{30}, & \text{se } x = 1,2,3,4 \\ 0, & \text{para outros valores de } x \end{cases}$$

xᵢ :	*1*	*2*	*3*	*4*
f (xᵢ) :	*1/30*	*2/15*	*3/10*	*8/15*



1.2 Função distribuição

A **função distribuição** , **F (x)** , é também uma função matemática que, neste caso, associa a cada valor da variável aleatória discreta a probabilidade de ele ou valores inferiores a ele ocorrerem.

O seu domínio pode ser definido como sendo |R.

$$\mathbf{F(x)}: \quad |R \longrightarrow [0,1]$$
$$x \longmapsto \mathbf{F(x) = P(X \leq x)}$$

Ou seja: $$F(x) = \sum_{x_i \leq x} f(x_i)$$

Note que: $$0 \leq F(x) \leq 1, \quad \forall x_i$$

$\sum_{i=1}^{k} F(x_i)$ não tem qualquer significado especial.

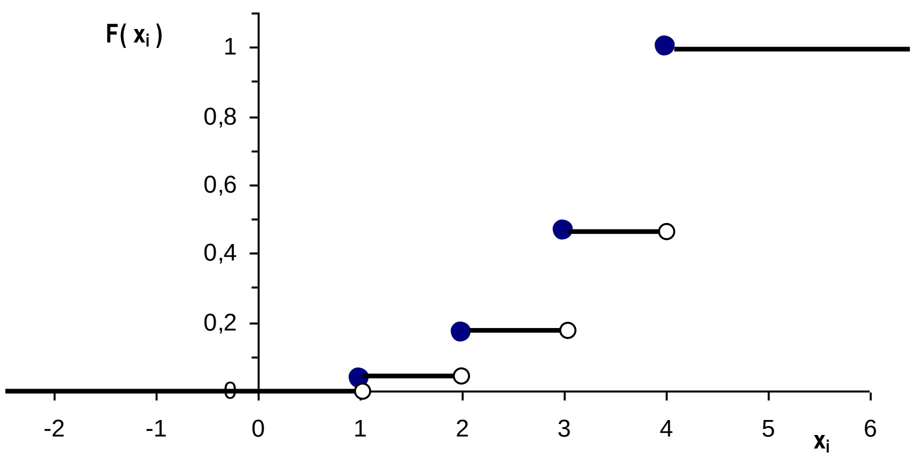
A função distribuição também pode ser definida:

- analiticamente através de uma ou várias expressões matemáticas (função definida por ramos),
- através de uma tabela,
- graficamente.

Exemplo: *Considerando a distribuição de probabilidades do exemplo anterior, tem-se que:*

$$F(x) = \begin{cases} 0\,, & \text{se } x < 1 \\ 1/30 & \text{se } 1 \leq x < 2 \\ 1/6 & \text{se } 2 \leq x < 3 \\ 7/15 & \text{se } 3 \leq x < 4 \\ 1\,, & \text{se } x \geq 4 \end{cases}$$

***x_i* :**	*1*	*2*	*3*	*4*
***F (x_i)* :**	*1/30*	*5/30 ou 1/6*	*14/30 ou 7/15*	*1*



A Função de distribuição pode tornar-se útil no cálculo de probabilidades que envolvem muitos valores da variável X, em alternativa à soma dos valores da função probabilidade:

$\mathbf{P (X = x_i) = F (x_i) - F (x_{i-1})}$

$\mathbf{P (X \leq x_i) = F (x_i)}$

$\mathbf{P (X < x_i) = F (x_{i-1})}$

$\mathbf{P (X > x_i) = 1 - F (x_i)}$

$\mathbf{P (X \geq x_i) = 1 - F (x_{i-1})}$

$\mathbf{P (x_i < X \leq x_j) = F (x_j) - F (x_i)}$ *Nota: i < j*

$\mathbf{P (x_i \leq X \leq x_j) = F (x_j) - F (x_{i-1})}$

$\mathbf{P (x_i < X < x_j) = F (x_{j-1}) - F (x_i)}$

$\mathbf{P (x_i \leq X < x_j) = F (x_{j-1}) - F (x_{i-1})}$

Exemplo: *Considerando a distribuição de probabilidades anterior, tem-se que:*

P (X < 3) = f (1) + f (2) = 1/30 + 3/15 = 1/6
Mas também: P (X < 3) = F (2) = 1/6

P (2 ≤ X ≤ 4) = f (2) + f (3) + f (4) = 3/15 + 3/10 + 8 / 15 = 29/30
Mas também: P (2 ≤ X ≤ 4) = F (4) - F (1) = 1 – 1/30 = 29/30

1.3 Média, Valor médio, Valor esperado ou Esperança matemática

Média, **Valor médio**, **Valor esperado** ou **Esperança matemática** , μ ou **E [X]** , é o valor da média populacional ; isto é, é a média de todos os valores da população que tem a distribuição de probabilidades / ocorrências f (x) e define-se do seguinte modo:

$$\boxed{\mu = E [X] = \sum_{i=1}^{k} x_i \cdot f (x_i)}$$

A média é a principal medida de localização de uma variável.

Exemplo: *Considerando a distribuição de probabilidades dos exemplos anteriores, tem-se que:*

$$\mu = E [X] = \sum_{i=1}^{4} x_i \cdot f (x_i) =$$

$$= \sum_{i=1}^{4} x_i \cdot \frac{x_i^2}{30} = \sum_{i=1}^{4} \frac{x_i^3}{30} = \frac{1}{30} + \frac{8}{30} + \frac{27}{30} + \frac{64}{30} = \frac{100}{30} = \frac{10}{3}$$

1.4 Mediana

Mediana , η , é o valor de X que divide o total da probabilidade em duas partes iguais ; idealmente, é o valor da variável tal que a probabilidade de ocorrerem valores superiores a ele ou valores inferiores a ele é igual a ½ . Na prática, tal nem sempre acontece, porque a mediana é muitas vezes ela própria um valor possível. Por isso, define-se a mediana do seguinte modo:

$$\boxed{P(X \geq \eta) \geq \tfrac{1}{2} \;\wedge\; P(X \leq \eta) \geq \tfrac{1}{2}}$$

A mediana, não sendo a principal medida de localização de uma variável, torna-se mais importante sempre que a distribuição de probabilidades é assimétrica.

Exemplo: *Considerando a distribuição de probabilidades dos exemplos anteriores, tem-se que:*

$$P(X \geq 4) = \frac{8}{15} \geq \frac{1}{2}$$

$$P(X \leq 4) = \frac{1}{30} + \frac{2}{15} + \frac{3}{10} + \frac{8}{15} = 1 \geq \frac{1}{2}$$

Logo, neste caso: $\eta = 4$.

1.5 Moda

Por definição, **Moda** , **Mo** , é o valor de X mais provável ; ou seja, é o valor de correspondente ao maior valor de f (x):

$$\boxed{P(X = Mo) = p(Mo) = \text{Máx}[f(x)]}$$

No entanto, esta definição pode ser estendida e podemos alargar o conceito de moda para:

$$\boxed{\text{Se } Mo = x_i\text{, então: } f(Mo) > f(x_{i-1}) \;\wedge\; f(Mo) > f(x_{i+1})}$$

Se todos os valores da variável possuem igual probabilidade de ocorrência (caso da distribuição uniforme, então diz-se que a moda não está definida.

Exemplo: *Considerando a distribuição de probabilidades dos exemplos anteriores, tem-se que:*

$Mo = 4$, porque $p(4) = 8/15$ é o valor máximo de $f(x)$.

Exemplo: *A distribuição de probabilidades seguinte tem duas modas: $Mo_1 = 2$ e $Mo_2 = 4$.*

x_i :	*1*	*2*	*3*	*4*	*5*
$f(x_i)$:	*0,05*	*0,15*	*0,10*	*0,40*	*0,30*

1.6 Variância e Desvio padrão

Variância , $\boldsymbol{\sigma^2}$ ou **V [X]** , é uma medida da dispersão da variável e é igual ao erro quadrático médio. Define-se do seguinte modo:

$$\sigma^2 = V[X] = \sum_{i=1}^{k} (x_i - \mu)^2 \cdot f(x_i)$$

Prova-se que a variância também pode ser calculada da seguinte forma, mais cómoda:

$$\sigma^2 = V[X] = E[X^2] - \mu^2 , \quad \text{com } E[X^2] = \sum_{i=1}^{k} x_i^2 \cdot f(x_i)$$

O **Desvio padrão** , $\boldsymbol{\sigma}$, é a medida da dispersão mais utilizada e é definido como sendo igual à raiz quadrada do erro quadrático médio ; ou seja, é a raiz quadrada da variância:

$$\sigma = \sqrt{V[X]} = \sqrt{\sigma^2}$$

Exemplo: *Considerando ainda a distribuição de probabilidades anterior, tem-se que:*

$\mu = 10 / 3$

$$E[X^2] = \sum_{i=1}^{k} x_i^2 \cdot f(x_i) =$$

$$= \sum_{i=1}^{4} x_i^2 \cdot \frac{x_i^2}{30} = \sum_{i=1}^{4} \frac{x_i^4}{30} = \frac{1}{30} + \frac{16}{30} + \frac{81}{30} + \frac{256}{30} = \frac{354}{30} = \frac{59}{5}$$

Logo: $$\sigma^2 = V[X] = E[X^2] - \mu^2 = \frac{59}{5} - \left(\frac{10}{3}\right)^2 = \frac{59}{5} - \frac{100}{9} = \frac{31}{45}$$

E portanto: $$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{31}{45}} \approx 0{,}830$$

2. Distribuição Uniforme

2.1 Definição. Função probabilidade.

Distribuição uniforme é um tipo de distribuição de probabilidades aplicável a todas as variáveis para as quais todos os seus **k** valores possíveis têm igual probabilidade de ocorrerem:

$$\mathbf{X \sim U(k)} \quad \Leftrightarrow \quad f(x) = \begin{cases} \dfrac{1}{k} & \text{se } x = x_1, x_2, \ldots x_k \\ 0, & \text{para outros valores de } x \end{cases}$$

A distribuição uniforme tem apenas um ***parâmetro*** a definir:

k = n.º de diferentes valores da variável X

Exemplo: *A variável X = "N.º de pontos obtidos no lançamento de um dado não viciado" segue uma distribuição uniforme, com k = 6, já que todas as faces têm igual probabilidade de saírem. Podemos pois escrever, para esta variável:*

X ~U (6)

A sua função probabilidade é a seguinte: $f(x) = \begin{cases} \dfrac{1}{6} & \text{se } x = 1, 2, 3, 4, 5, 6 \\ 0, & \text{para outros valores de } x \end{cases}$

2.2 Função distribuição

Decorre da definição de função distribuição que, para uma variável uniforme, se tem:

$$X \sim U(k) \quad \rightarrow \quad F(x_i) = \frac{i}{k}$$

Exemplo: *Para a variável X = "N.º de pontos obtidos no lançamento de um dado não viciado",*

a função distribuição é: $F(x) = \begin{cases} \dfrac{x}{6} & \text{se } x = 1, 2, 3, 4, 5, 6 \\ 0, & \text{para outros valores de } x \end{cases}$

2.3 Média, Valor médio, Valor esperado ou Esperança matemática

A média de uma variável com distribuição uniforme é igual à média aritmética dos seus valores possíveis:

$$X \sim U(k) \quad \rightarrow \quad \mu = E[X] = \frac{\sum_{i=1}^{k} x_i}{k}$$

Exemplo: *Considerando a variável X = "N.º de pontos obtidos no lançamento de um dado não viciado" a sua média é:*

$$\mu = E[X] = \frac{\sum_{x=1}^{6} x_i}{6} = \frac{1+2+3+4+5+6}{6} = 3{,}5$$

2.4 Mediana

A mediana de uma distribuição uniforme é igual à sua média.

$$\boxed{X \sim U(k) \qquad \rightarrow \qquad \eta = \mu = E[X]}$$

Exemplo: *Considerando a distribuição de probabilidades dos exemplos anteriores, tem-se que:*

$\eta = 3{,}5$

Efectivamente: $P(X \geq 3{,}5) = \frac{3}{6} = 0{,}5 \geq \frac{1}{2}$

e: $P(X \leq 3{,}5) = \frac{3}{6} = 0{,}5 \geq \frac{1}{2}$

2.5 Moda

Como todos os valores da variável uniforme possuem igual probabilidade de ocorrência, a sua moda não está definida.

2.6 Variância e Desvio padrão

Variância , $\boldsymbol{\sigma^2}$ ou **V [X]** , não possui uma definição específica no caso da distribuição uniforme e continua a ser igual ao erro quadrático médio:

$$\sigma^2 = V[X] = \sum_{i=1}^{k} (x_i - \mu)^2 . f(x_i) = E[X^2] - \mu^2 , \quad \text{com } E[X^2] = \sum_{i=1}^{k} x_i^2 . f(x_i)$$

O **Desvio padrão** , $\boldsymbol{\sigma}$, é, como sempre igual à raiz quadrada do erro quadrático médio ; ou seja, é a raiz quadrada da variância:

$$\sigma = \sqrt{V[X]} = \sqrt{\sigma^2}$$

3. Distribuição de Bernoulli

3.1 Definição. Função probabilidade.

Uma **experiência de Bernoulli** é aquela que obedece às seguintes condições:

- Apenas possui dois resultados possíveis, perfeitamente conhecidos à partida e designados por **sucesso** e **insucesso** (ou **falha**).

- A probabilidade ocorrência do resultado sucesso, chamada **probabilidade de sucesso**, **p** , é constante.

- A experiência pode sempre ser repetida em condições idênticas, sem que a probabilidade de sucesso numa experiência dependa dos resultados das experiências anteriormente realizadas.

A **variável de Bernoulli** , **X** , pode apenas assumir dois valores:

Se o resultado da experiência de Bernoulli for o ***sucesso***, **X = 1**;

Se o resultado da experiência de Bernoulli for o ***insucesso***, **X = 0**.

Ou seja, a variável de Bernoulli é por definição:

X = " N.º de sucessos, em uma experiência de Bernoulli"

A função probabilidade de tal variável é sempre a seguinte:

$$\mathbf{X \sim Be\,(\,p\,)} \quad \Leftrightarrow \quad f(x) = \begin{cases} 1-p \ , \text{ se } x=0 \\ p, \quad \text{ se } x=1 \\ 0 \ , \text{ para outros valores de } x \end{cases}$$

A distribuição de Bernoulli tem assim apenas um ***parâmetro*** a definir:

p = ***probabilidade de sucesso***

A probabilidade de falha ou insucesso, não sendo um parâmetro que é necessário definir, será desgnada por **q** e tem-se que:

$$q = 1 - p$$

Exemplo: *Considere a seguinte experiência: O João sai de casa às 8h00 para apanhar o autocarro, cujo horário de passagem é às 7h55. A probabilidade de o João apanhar o autocarro é igual a 0,30, já que este se atrasa mais de 5 minutos em 30 % dos dias.*

Seja a variável: X = "N.º de vezes que o João apanha o autocarro das 7h55 em uma manhã". Esta é uma variável de Bernoulli. Podemos portanto escrever:

X ~Be (0,30)

A sua função probabilidade é a seguinte:

$$f(x) = \begin{cases} 0{,}70 \ , \text{ se } x=0 \\ 0{,}30 \ , \text{ se } x=1 \\ 0 \ , \text{ para outros valores de } x \end{cases}$$

3.2 Função distribuição

Decorre da definição de função distribuição que, para uma variável de Bernoulli, se tem:

$$X \sim Be\,(p) \quad \rightarrow \quad F\,(x) = \begin{cases} 0 \ , \text{ se } \ x < 0 \\ 1-p \ , \quad \text{se } \ 0 \le x < 1 \\ 1 \ , \text{ se } x \ge 1 \end{cases}$$

Exemplo: *Para a variável: X = "N.º de vezes que o João apanha o autocarro das 7h55 numa manhã" a função distribuição é:*

$$F\,(x) = \begin{cases} 0 \ , \text{ se } \ x < 0 \\ 0{,}70 \ , \quad \text{se } \ 0 \le x < 1 \\ 1 \ , \text{ se } x \ge 1 \end{cases}$$

3.3 Média, Valor médio, Valor esperado ou Esperança matemática

Prova-se que a média de uma variável com distribuição de Bernoulli é igual a p:

$$X \sim Be\,(p) \quad \rightarrow \quad \mu = E\,[\,X\,] = p$$

Exemplo: *A média da variável: X = "N.º de vezes que o João apanha o autocarro das 7h55 numa manhã" é:*

$$\mu = E\,[\,X\,] = 0{,}30$$

3.4 Mediana

Para uma distribuição de Bernoulli:

- Se $p < 0{,}5 \rightarrow \eta = 0$
- Se $p > 0{,}5 \rightarrow \eta = 1$
- Se $p = 0{,}5 \rightarrow \eta = 0{,}5$

3.5 Moda

Para uma distribuição de Bernoulli:

- Se $p < 0{,}5 \rightarrow Mo = 0$
- Se $p > 0{,}5 \rightarrow Mo = 1$
- Se $p = 0{,}5 \rightarrow Mo = 0{,}5$

3.6 Variância e Desvio padrão

Prova-se que a variância , σ^2 ou V [X] , de uma distribuição de Bernoulli é igual ao produto da probabilidade de sucesso pela probabilidade de insucesso:

$$\boxed{\sigma^2 = V[X] = p \cdot (1 - p) = p \cdot q}$$

O desvio padrão , σ , é, como sempre igual à raiz quadrada do erro quadrático médio ; ou seja, é a raiz quadrada da variância:

$$\sigma = \sqrt{V[X]} = \sqrt{\sigma^2}$$

Exemplo: *A variância da variável: X = "N.º de vezes que o João apanha o autocarro das 7h55 numa manhã" é:*

$$\sigma^2 = V[X] = 0{,}30 \times 0{,}70 = 0{,}21$$

E o seu desvio padrão é: $\sigma = \sqrt{0{,}21} = 0{,}458$

4. Distribuição Binomial

4.1 Definição. Função probabilidade.

Uma **experiência binomial** consiste na repetição de um número fixo de experiências de Bernoulli ; logo, é aquela que obedece às seguintes condições:

- Consiste na repetição de **n** experiências iguais.
- Em cada experiência, só há dois resultados possíveis, perfeitamente conhecidos à partida e designados por **sucesso** e **insucesso** (ou **falha**).
- A probabilidade ocorrência do resultado sucesso, chamada **probabilidade de sucesso**, **p** , é constante, ao logo de todas as experiências.
- As várias experiências são independentes entre si ; isto é a probabilidade de sucesso em cada experiência não depende dos resultados das experiências anteriormente realizadas.

Subentendida nestas condições, está a necessidade de haver reposição entre experiências. No entanto, esta reposição é negligenciável se a população for muito grande (muito maior do que o número de experiências realizadas).

A **Variável Binomial** , **X** , define-se do seguinte modo:

X = " N.º de sucessos, em n experiências."

E os valores que pode assumir são:

$$\mathbf{X \in \{ 0 , 1 , 2 , 3 , \dots , n \}}$$

A variável é caracterizada por dois parâmetros:

- n é o número de experiências

- p é a probabilidade de sucesso em cada experiência.

De uma variável binomial, escreve-se, portanto: **X ~ Bi (n , p)**

A probabilidade de falha ou insucesso, não sendo um parâmetro que é necessário definir, será desgnada por **q** e tem-se que:

$$q = 1 - p$$

A função probabilidade de tal variável é sempre a seguinte:

$$X \sim Bi(n, p) \quad \Leftrightarrow \quad f(x) = \begin{cases} C_x^n \cdot p^x \cdot q^{n-x} \ , & \text{se } x = 0, 1, 2, \dots, n \\ 0 \ , & \text{para outros valores de } x \end{cases}$$

Os valores da função probabilidade binomial encontram-se abundantemente publicados na bibliografia sob a forma de tabela. Muitas máquinas de calcular trazem já de origem rotinas incorporadas para o cálculo destes valores.

Exemplo: *Considere a contagem de caras em 10 lançamentos de uma mesma moeda, não viciada. A variável X = " N.º de caras em 10 lançamentos" segue uma distribuição binomial (verifica todas as condições) com parâmetros n = 10 e p = 0,5. Podemos portanto escrever:*

$$X \sim Bi\,(\,10\,;\,0{,}5\,)$$

4.2 Função distribuição

A função distribuição de uma variável binomial não assume nenhuma forma especial.

5.3 Média, Valor médio, Valor esperado ou Esperança matemática

Prova-se que a média de uma variável com distribuição Binomial é igual a n . p:

$$\boxed{X \sim Bi\,(\,n\,,\,p\,) \quad \rightarrow \quad \mu = E\,[\,X\,] = n\,.\,p}$$

Exemplo: *A média da variável: X = " N.º de caras em 10 lançamentos" é:*

$\mu = E\,[\,X\,] = n\,.\,p = 10 \times 0{,}5 = 5$

4.4 Variância e Desvio padrão

Prova-se que a variância , σ^2 ou V [X] , de uma distribuição Binomial é:

$$\boxed{\sigma^2 = V\,[\,X\,] = n\,.\,p\,.\,(\,1 - p\,) = n\,.\,p\,.\,q}$$

O desvio padrão , σ , é, como sempre igual à raiz quadrada do erro quadrático médio ; ou seja, é a raiz quadrada da variância:

$$\sigma = \sqrt{V\,[\,X\,]} = \sqrt{\sigma^2}$$

Exemplo: *A variância da variável: X = " N.º de caras em 10 lançamentos" é:*

$\sigma^2 = V\,[\,X\,] = 10 \times 0{,}50 \times 0{,}50 = 2{,}5$

E o seu desvio padrão é: $\sigma = \sqrt{2{,}5} = 1{,}58$

5. Distribuição de Poisson

5.1 Definição. Função probabilidade.

Não há um conjunto de condições que nos permitam afirmar sem erro que certa variável segue uma distribuição de Poisson. Para tal é necessário realizar testes (testes de qualidade de ajuste). Podemos no entanto apontar algumas variáveis que tipicamente seguem uma distribuição de Poisson.

A **variável de Poisson** tem o seguinte formato típico:

X = " N.º de sucessos, numa quantidade fixa de meio contínuo."

E os valores que pode assumir são:

$$X \in \{ 0 , 1 , 2 , 3 , \dots , +\infty \}$$

Exemplos: *São (ou poderão ser) variáveis de Poisson as seguintes:*

"N.º de clientes no período de abertura diário de 8h00."

"N.º de acidentes em 3 dias."

"N.º de terramotos em 1 ano."

"N.º de buracos por Km de estrada."

"N.º de defeitos por m^2 de tecido."

"N.º de árvores por hectare"

"N.º de erros de dactilografia por página."

"N.º de peixes por m^3 de água."

"N.º de passas por 100 g de Bolo-rei."

A afirmação de que estas variáveis seguem realmente distribuições de Poisson carece de validação.

A variável é caracterizada apenas por um parâmetro:

λ = número <u>médio</u> de sucessos na quantidadade fixada de meio contínuo

De uma variável de Poisson, escreve-se, portanto: **X ~ Po (λ)**

Note que a quantidade de meio contínuo fixada não é ela própria um parâmetro da distribuição, mas influencia o valor de λ, de acordo com uma lei de proporcionalidade directa:

$$\frac{\lambda_1}{\lambda_2} = \frac{Q_1}{Q_2}$$

Exemplo: *Vamos supor que a variável X = "N.º de clientes de uma sapataria, por hora" segue uma distribuição de Poisson com parâmetro $\lambda = 3$.*

Então, a variável Y = "N.º de clientes de uma sapataria, em 8h00" segue uma distribuição de Poisson com parâmetro $\lambda = 8 \times 3 = 24$.

A função probabilidade de tal variável é sempre a seguinte:

$$\mathbf{X \sim Po\ (\lambda)} \quad \Leftrightarrow \quad p(x) = \begin{cases} \dfrac{e^{-\lambda} \cdot \lambda^{x}}{x!}, & \text{se } x = 0, 1, 2, \ldots \\ 0, & \text{para outros valores de } x \end{cases}$$

Os valores da função probabilidade de Poisson encontram-se abundantemente publicados na bibliografia sob a forma de tabela. Muitas máquinas de calcular trazem já de origem rotinas incorporadas para o cálculo destes valores.

Exemplo: *Considere a análise do fenómeno de chegada de clientes a uma loja, sendo que em média há 4,5 clientes por hora.*
A variável X = " N.º de clientes num intervalo de tempo de 1 hora" segue uma distribuição de Poisson com parâmetro λ = 4,5. Podemos portanto escrever:

$$X \sim Po\ (4{,}5)$$

5.2 Função distribuição

A função distribuição de uma variável de Poisson não assume nenhuma forma especial.

5.3 Média, Valor médio, Valor esperado ou Esperança matemática

Como já foi dito, a média de uma variável com distribuição de Poisson é igual a λ:

$$X \sim Po\ (\lambda) \quad \rightarrow \quad \mu = E[X] = \lambda$$

Exemplo: *A média da variável: X = " N.º de clientes num intervalo de tempo de 1 hora" é:*

$\mu = E[X] = \lambda = 4{,}5$

5.4 Moda

Se λ não é um número inteiro, então a moda da variável de Poisson é igual ao valor inteiro de λ. Se λ é inteiro, então a distribuição diz-se bimodal e as duas modas são o próprio e o inteiro inferior.

$$X \sim Po\ (\lambda) \text{ e } \lambda \text{ não é inteiro} \rightarrow \quad Mo = Int(\lambda)$$

$$X \sim Po\ (\lambda) \text{ e } \lambda \text{ é inteiro} \rightarrow \quad Mo_1 = \lambda - 1 \quad Mo_2 = \lambda$$

Exemplo: *A moda de: X = " N.º de clientes num intervalo de tempo de 1 hora", com λ = 4,5 é **4**.*

5.5 Variância e Desvio padrão

Prova-se que a variância , σ^2 ou V [X] , de uma distribuição de Posson é:

$$\boxed{\sigma^2 = V[X] = \lambda}$$

O desvio padrão , σ , é, como sempre igual à raiz quadrada do erro quadrático médio ; ou seja, é a raiz quadrada da variância:

$$\sigma = \sqrt{V[X]} = \sqrt{\sigma^2}$$

Exemplo: *A variância da variável: X = " N.º de clientes num intervalo de tempo de 1 hora" é:*

$\sigma^2 = V[X] = \lambda = 4{,}5$

E o seu desvio padrão é: $\sigma = \sqrt{4{,}5} = 2{,}12$

6. Conceitos gerais sobre distribuições de variáveis contínuas

6.1 Função densidade de probabilidade

Função densidade probabilidade, **f (x)** , é uma função matemática que associa a cada valor de uma variável aleatória quantitativa discreta a densidade de probabilidade que lhe está associada.

Também é costume chamar-se a tal função a **distribuição de probabilidades** da variável X.

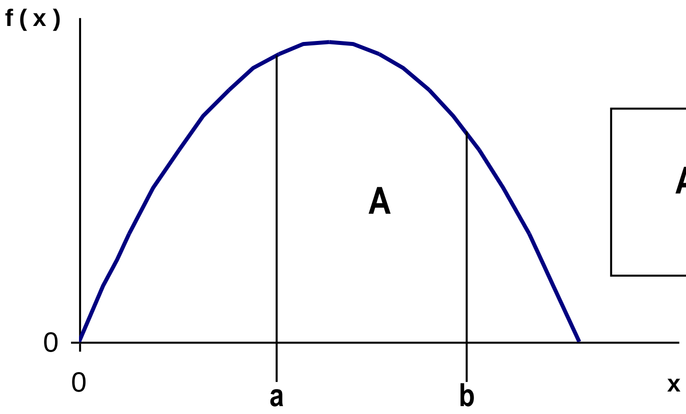
Note que densidade de probabilidade não é o mesmo que probabilidade, uma vez que para variáveis contínuas se tem sempre que:

$P (X = x) \rightarrow 0$

O seu domínio pode ser definido como o conjunto de valores possíveis para a variável, mas é habitualmente definido como sendo |R, atribuindo-se a todos os valores impossíveis a imagem 0. Dado que não se trata de uma probabilidade, pode acontecer que f (x) seja superior a 1.

$$\mathbf{f(x)}: \quad \mathbb{R} \longrightarrow [\,0\,,+\infty\,[$$
$$x \longmapsto \mathbf{f(x)}$$

O seu significado físico é o seguinte: a probabilidade de uma variável contínua X assumir valores num intervalo entre a e b (com a e b finitos ou infinitos) é matematicamente igual à área entre a representação gráfica da função e o eixo dos xx, entre a e b; ou seja, é:



$$\mathbf{A} = P (a \le X \le b) = \int_a^b f(x)\, dx$$

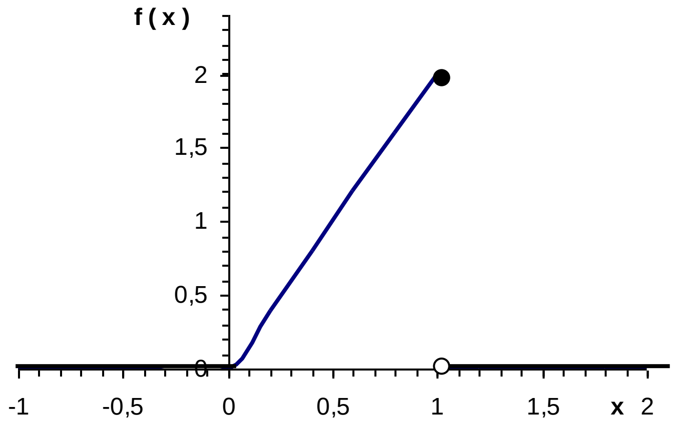
Tem-se necessariamente que:

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

A função densidade de probabilidade pode ser definida:

- analiticamente através de uma ou várias expressões matemáticas (função definida por ramos),
- graficamente.

Exemplo: $f(x) = \begin{cases} 2x \ , \ \text{se } 0 \le x \le 1 \\ 0, \text{para outros valores de } x \end{cases}$



6.2 Função distribuição

A **função distribuição** , **F (x)** , é também uma função matemática que, neste caso, associa a cada valor da variável aleatória discreta a probabilidade de ele ou valores inferiores a ele ocorrerem.
O seu domínio pode ser definido como sendo |R.

$$\mathbf{F(x)}: \quad \mathbb{R} \longrightarrow [0,1]$$
$$x \mapsto \mathbf{F(x) = P(X \leq x)}$$

Ou seja:

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

Note que, enquanto f (x) não é uma probabilidade, F (x) é uma probabilidade, portanto tem-se necessariamente que:

$$\mathbf{0 \leq F(x) \leq 1}, \quad \forall x$$

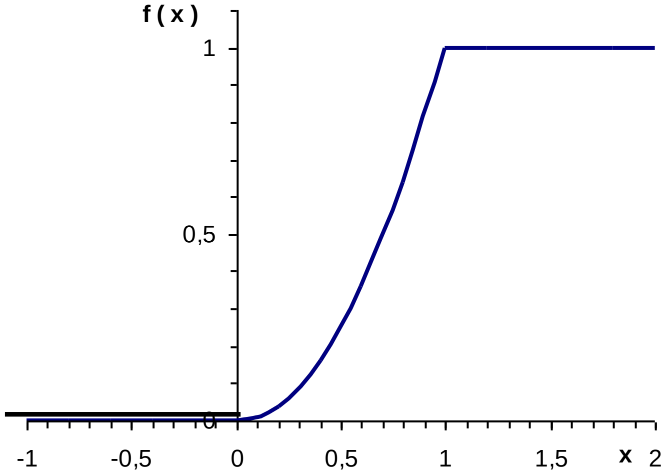
A função distribuição também pode ser definida:

- analiticamente através de uma ou várias expressões matemáticas (função definida por ramos),

- graficamente.

Exemplo: *Considerando a distribuição de probabilidades do exemplo anterior, tem-se que:*

$$F(x) = \begin{cases} 0, & \text{se } x < 0 \\ x^2 & \text{se } 0 \leq x \leq 1 \\ 1, & \text{se } x > 1 \end{cases}$$



A Função de distribuição pode tornar-se útil no cálculo de probabilidades que envolvem muitos valores da variável X, em alternativa à integração da função densidade de probabilidade:

$P(X = a) = 0$

$P(X \leq a) = P(X < a) = F(a)$

$P(X \geq x_i) = P(X > x_i) = 1 - F(a)$

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

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

$= F(b) - F(a)$ *Nota: a < b*

Exemplo: *Considerando a distribuição de probabilidades anterior, tem-se que:*

$P(X < 0{,}3) = F(0{,}3) = 0{,}3^2 = 0{,}09$

Mas também: $P(X < 0{,}3) = \int_{-\infty}^{0,3} f(x)\,dx = \int_{-\infty}^{0} 0\,dx + \int_{0}^{0,3} 2x\,dx = 0 + \left[2\frac{x^2}{2}\right]_0^{0,3} = 0{,}09$

$P(0{,}2 \leq X \leq 0{,}4) = F(0{,}4) - F(0{,}2) = 0{,}4^2 - 0{,}2^2 = 0{,}12$

Mas também: $P(0{,}2 \leq X \leq 0{,}4) = \int_{0,2}^{0,4} f(x)\,dx = \int_{0,2}^{0,4} 2x\,dx = \left[2\frac{x^2}{2}\right]_{0,2}^{0,4} = 0{,}12$

6.3 Média, Valor médio, Valor esperado ou Esperança matemática

Média, **Valor médio**, **Valor esperado** ou **Esperança matemática** , μ ou **E [X]** , é o valor da média populacional ; isto é, é a média de todos os valores da população que tem a distribuição de probabilidades f (x) e define-se do seguinte modo:

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

A média é a principal medida de localização de uma variável.

Exemplo: *Considerando a distribuição de probabilidades dos exemplos anteriores, tem-se que:*

$$\mu = E[X] = \int_{-\infty}^{+\infty} x\,f(x)\,dx = \int_{-\infty}^{0} x.0\,dx + \int_{0}^{1} x.2x\,dx + \int_{0}^{+\infty} x.0\,dx = 0 + \left[2\frac{x^3}{3}\right]_0^1 = \frac{2}{3}$$

6.4 Mediana

Mediana , η , é o valor de X que divide o total da probabilidade em duas partes iguais ; por isso, define-se a mediana do seguinte modo:

$$P(X \le \eta) = ½ \Leftrightarrow F(\eta) = ½$$

A mediana, não sendo a principal medida de localização de uma variável, torna-se mais importante sempre que a distribuição de probabilidades é assimétrica.

Exemplo: *Considerando a distribuição de probabilidades dos exemplos anteriores, tem-se que:*

$F(\eta) = ½ \Leftrightarrow \eta^2 = ½ \Leftrightarrow \eta \approx \pm 0{,}707$

Logo, neste caso: $\eta \approx 0{,}707$.

6.5 Moda

Moda , **Mo**, é o valor de correspondente ao máximo de f (x):

$$f(Mo) = Máx[f(x)]$$

Na determinação da moda deverão ter-se em atenção:

- os extremos relativos de f (x), isto é, os valores de X tais que: $f'(x) = 0$ e $f''(x) < 0$.
- os valores da função f (x) nos extremo de validade de qualquer expressão que a defina
- eventuais ponto de descontinuidade

Se todos os valores da variável possuem igual densidade de probabilidade de ocorrência (caso da distribuição uniforme, então diz-se que a moda não está definida.

Exemplo: *Considerando a distribuição de probabilidades dos exemplos anteriores, tem-se que:*

Mo = 1 , porque f (1) = 2 é o valor máximo de f (x).

6.6 Variância e Desvio padrão

Variância , σ^2 ou **V [X]** , é uma medida da dispersão da variável e é igual ao erro quadrático médio. Define-se do seguinte modo:

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

Prova-se que a variância também pode ser calculada da seguinte forma, mais cómoda:

$$\sigma^2 = V[X] = E[X^2] - \mu^2 \,, \quad \text{com } E[X^2] = \int_{-\infty}^{+\infty} x^2 \cdot f(x)\,dx$$

O **Desvio padrão** , σ , é a medida da dispersão mais utilizada e é definido como sendo igual à raiz quadrada do erro quadrático médio ; ou seja, é a raiz quadrada da variância:

$$\sigma = \sqrt{V[X]} = \sqrt{\sigma^2}$$

Exemplo: *Considerando ainda a distribuição de probabilidades anterior, tem-se que:*

$\mu = 2/3$

$$E[X^2] = \int_{-\infty}^{+\infty} x^2 \cdot f(x)\,dx = \int_{-\infty}^{0} x^2 \cdot 0\,dx + \int_{0}^{1} x^2 \cdot 2x\,dx + \int_{1}^{+\infty} x^2 \cdot 0\,dx =$$

$$= 0 + \int_{0}^{1} 2x^3\,dx + 0 = \left[2\frac{x^4}{4}\right]_0^1 = ½ - 0 = ½ = 0{,}5$$

E portanto: $\sigma = \sqrt{\sigma^2} = \sqrt{0{,}5} \approx 0{,}707$

7. Distribuição Uniforme (para variáveis contínuas)

7.1 Definição. Função densidade de probabilidade.

Distribuição uniforme é um tipo de distribuição de probabilidades aplicável a todas as variáveis para as quais todos os valores possíveis têm igual valor de função densidade de probabilidade; por outras palavras, a probabilidade de ocorrências num intervalo é directamente proporcional à largura desse intervalo:

$$\mathbf{X \sim U\,(\,a\,,\,b\,)} \quad \Leftrightarrow \quad f\,(\,x\,) = \begin{cases} \dfrac{1}{b-a} & \text{se } a \le x \le b \\ 0\ , & \text{para outros valores de } x \end{cases}$$

A distribuição uniforme tem dois ***parâmetros*** a definir:

a = limite inferior do intervalo de possíveis valores da variável X

b = limite superior do intervalo de possíveis valores da variável X

Exemplo: *A variável X, contínua, tem distribuição uniforme no intervalo [0 , 1]. Podemos pois escrever, para esta variável:*

X ~U (0 , 1)

A sua função densidade de probabilidade é pois a seguinte:

$$f\,(\,x\,) = \begin{cases} \dfrac{1}{4} & \text{se } 0 \le x \le 4 \\ 0\ , & \text{para outros valores de } x \end{cases}$$

7.2 Função distribuição

Decorre da definição de função distribuição para variáveis contínuas que, para uma variável uniforme, se tem:

$$\mathbf{X \sim U\,(\,a\,,\,b\,)} \quad \Leftrightarrow \quad F\,(\,x\,) = \begin{cases} 0\ , & \text{se } x < a \\ \dfrac{x-a}{b-a}\ , & \text{se } a \le x \le b \\ 1\ , & \text{se } x > b \end{cases}$$

Exemplo: *Para a variável anterior, a função distribuição é:*

$$F\,(\,x\,) = \begin{cases} 0\ , & \text{se } x < 0 \\ \dfrac{x}{4}\ , & \text{se } 0 \le x \le 4 \\ 1\ , & \text{se } x > 4 \end{cases}$$

7.3 Média, Valor médio, Valor esperado ou Esperança matemática

A média de uma variável com distribuição uniforme é igual à média dos seus extremos:

$$X \sim U(a, b) \quad \rightarrow \quad \mu = E[X] = \frac{a+b}{2}$$

Exemplo: *Considerando ainda a variável anterior, a sua média é:*

$$\mu = E[X] = \frac{0+4}{2} = 2$$

7.4 Mediana

A mediana de uma distribuição uniforme é igual à sua média.

$$X \sim U(a, b) \quad \rightarrow \quad \eta = \mu = E[X]$$

7.5 Moda

Como todos os valores da variável uniforme possuem igual função densidade de probabilidade, a sua moda não está definida.

7.6 Variância e Desvio padrão

Prova-se que a variância , $\boldsymbol{\sigma^2}$ ou **V [X]** , de uma variável contínua uniforme é igual a:

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

O **Desvio padrão** , $\boldsymbol{\sigma}$, é, como sempre, igual à raiz quadrada da variância:

$$\sigma = \sqrt{V[X]} = \sqrt{\sigma^2}$$

Exemplo: *Ainda em relação ao exemplo anterior, a sua variância e o seu desvio padrão são:*

$$\sigma^2 = V[X] = \frac{(4-0)^2}{12} = 4/3 \qquad \text{e} \qquad \sigma = \sqrt{4/3} = 2\sqrt{3}/3$$

8. Distribuição Exponencial

8.1 Definição. Função densidade de probabilidade.

A **Distribuição Exponencial** está intimamente ligada à distribuição de Poisson e podem definir-se ambas sobre um mesmo fenómeno. Enquanto que a variável de Poisson é, por exemplo, Y = " N.º de acontecimento num período de tempo fixo", a variável exponencial será X = " Tempo decorrido entre dois acontecimentos consecutivos" ou X = " Tempo sem acontecimentos".

A sua função densidade de probabilidade é:

$$\mathbf{X \sim Exp\ (\lambda)} \quad \Leftrightarrow \quad f(x) = \begin{cases} 0 \ \text{ se } \ x < 0 \\ \lambda \, . \, e^{-\lambda \, . \, x} \ , \ \text{se } x \geq 0 \end{cases}$$

A distribuição uniforme tem um ***parâmetro*** a definir:

λ = média da distribuição de Poisson / Δt correspondente

Ou seja, enquanto que na distribuição de Poisson o parâmetro λ é sempre proporcional ao intervalo de tempo fixado, aqui, onde não se fixa um intervalo de tempo (antes pelo contrário, ele é a variável !), o valor de λ corresponde ao número médio de acontecimentos por unidade de tempo. Claro qestá que esta unidade de tempo deverá ser a mesma da variável.

Tal como acontecia na distribuição de Poisson, esta distribuição pode ser considerada, quando o que está em causa é uma quantidade de meio contínuo, não apenas tempo. Assim também é aplicável a comprimento, área, etc…

Exemplo: *A variável Y = "N.o de clientes durante 2 h" tem distribuição de Poisson comparâmetro igual a 5; isto é: Y ~ Po ($\lambda = 5$).*
Então a variável X = " Tempo em que a loja está sem clientes, em horas." segue uma distribuição exponencial com parâmetro λ = 5 / 2 = 2,5: X ~ Exp (λ = 2,5) e a sua função densidade de probabilidade é:

$$X \sim Exp\ (2{,}5) \quad \Leftrightarrow \quad f(x) = \begin{cases} 0 \ \text{ se } \ x < 0 \\ 2{,}5 \, . \, e^{-2{,}5 \, . \, x} \ , \ \text{se } x \geq 0 \end{cases}$$

8.2 Função distribuição

Decorre da definição de função distribuição para variáveis contínuas que, para uma variável exponencial, se tem:

$$\mathbf{X \sim Exp\ (\lambda)} \quad \Leftrightarrow \quad F(x) = \begin{cases} 0 \ , \quad \text{se } x < 0 \\ 1 - e^{-\lambda \, . \, x} \ , \ \text{ se } \ x \geq 0 \end{cases}$$

Também interessante é a fórmula de cálculo da probabilidade acumulada superior. Juntando as duas, podemos pois dizer:

$$P(X \le x) = P(X < x) = 1 - e^{-\lambda \cdot x}$$

$$P(X \ge x) = P(X > x) = e^{-\lambda \cdot x}$$

Exemplo: *Na sequência do exemplo anterior, a probabilidade de a loja estar menos de 1 hora sem clientes é igual a:*

$$P(X \le 1) = P(X < x) = 1 - e^{-2,5 \cdot 1} = 0,9179$$

Já a probabilidade de a loja estar 2 horas sem clientes (ou seja, pelo menos 2 horas sem clientes) é igual a:

$$P(X \ge x) = P(X > x) = e^{-2,5 \cdot 2} = 0,0821$$

8.3 Média, Valor médio, Valor esperado ou Esperança matemática

A média de uma variável com distribuição exponencial é igual ao inverso do seu parâmetro:

$$X \sim Exp(\lambda) \rightarrow \quad \mu = E[X] = \frac{1}{\lambda}$$

Exemplo: *Considerando ainda a variável anterior, a sua média é:* $\mu = E[X] = \frac{1}{2,5} = 0,4$ horas

8.4 Mediana

A mediana de uma distribuição exponencial deduz-se directamente da definição de mediana de uma variável contínua:

$$F(x) = 0,5 \Leftrightarrow 1 - e^{-\lambda \cdot x} = 0,5 \Leftrightarrow = e^{-\lambda \cdot x}\ 0,5 \Leftrightarrow -\lambda . x = \ln(0,5) \Leftrightarrow x = -\frac{\ln(0,5)}{\lambda}$$

$$X \sim Exp(\lambda) \quad \rightarrow \quad \eta = -\frac{\ln(0,5)}{\lambda}$$

Exemplo: *Considerando ainda a variável anterior, a sua mediana é:* $\eta = -\frac{\ln(0,5)}{2,5} = 0,2773$

8.5 Moda

A mediana de uma distribuição exponencial é sempre igual a 0.

$$X \sim Exp(\lambda) \quad \rightarrow \quad Mo = 0$$

8.6 Variância e Desvio padrão

Prova-se que a variância , $\boldsymbol{\sigma^2}$ ou **V [X]** , de uma variável exponencial é igual a:

$$\boxed{\sigma^2 = V[X] = \frac{1}{\lambda^2}}$$

O **Desvio padrão** , $\boldsymbol{\sigma}$, é, como sempre, igual à raiz quadrada da variância:

$$\boxed{\sigma = \sqrt{V[X]} = \sqrt{\sigma^2} = \frac{1}{\lambda}}$$

Exemplo: Ainda em relação ao exemplo anterior, a sua variância e o seu desvio padrão são:

$$\sigma^2 = V[X] = \frac{1}{2{,}5^2} = 0{,}16 \qquad \text{e} \qquad \sigma = \frac{1}{2{,}5} = 0{,}4$$

9. Distribuição Normal

9.1 Definição. Função densidade de probabilidade.

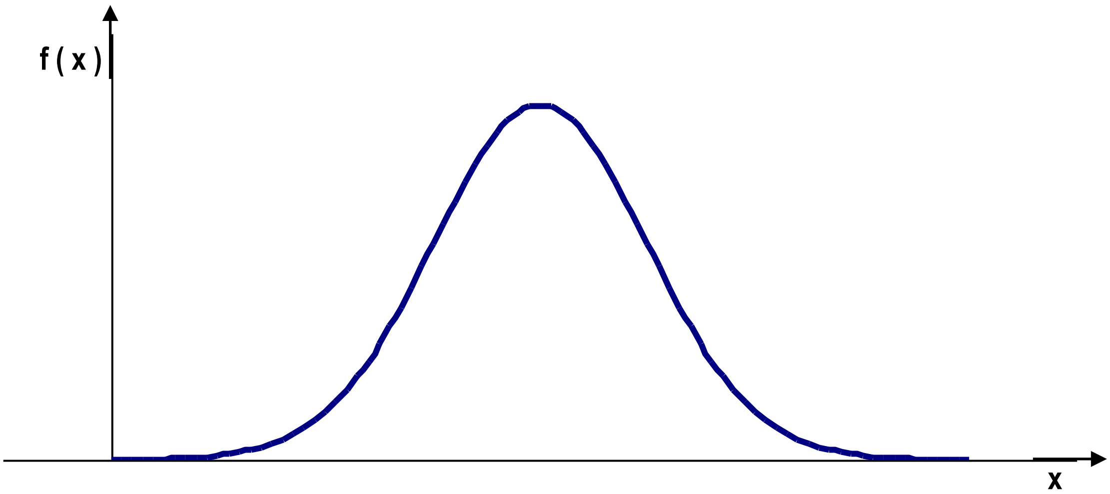
População normal – População cujos valores se encontram distribuídos segundo uma Lei de distribuição normal.

Variável normal – Variável quantitativa cujos possíveis valores estão associados a probabilidades distribuídas segundo uma Lei de distribuição normal.

Lei da Distribuição normal – Distribuição de probabilidades na seguinte forma:

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

cuja representação gráfica é uma **Curva de Gauss**:



A distribuição normal tem dois ***parâmetros*** a definir:

μ = média populacional

σ = desvio padrão populacional [1]

A Curva de Gauss varia consoante a média populacional (deslocando-se mais para a direita ou mais para a esquerda, consoante a média seja superior ou inferior, respetivamente) e também com o desvio padrão populacional (quando o desvio padrão é menor, a curva apresenta-se menos alargada).

Assim, o aspeto particular de cada curva varia apenas com os valores de μ e de σ, pelo que é costume escrever-se em relação a uma variável normal:

[1] Alguns autores definem como segundo parâmetro a variância, σ^2 , e não o desvio padrão.

X ~ N (μ , σ) → *A variável X <u>está distribuída de acordo com</u> ou <u>segue</u> uma distribuição normal com média μ e desvio padrão σ.*

Exemplo: *X = "Erro de enchimento de uma garrafa de água de 2 L, em mL", em certa empresa, segue uma distribuição normal, com média 0,5 mL e desvio padrão 3,1 mL. Podemos assim escrever:*

X ~ N (0,5 ; 3,1)

Muito embora a distribuição normal de probabilidades tenha sido desenvolvida para ser aplicada a variáveis contínuas, como, por exemplo o erro experimental aleatório na medição de um valor, ela também se aplica por aproximação a algumas variáveis discretas.

9.2 Características da Distribuição normal

A curva de Gauss apresenta algumas <u>propriedades</u> a saber:

- A curva é perfeitamente simétrica em torno da média, cujo valor é o valor de x correspondente ao máximo / centro da curva (ver figura).

- A área compreendida entre a curva e o eixo dos xx, num dado intervalo de x, digamos de **a** a **b**, é direitamente proporcional à probabilidade / ocorrência de valores entre **a** e **b** (ver figura).

- A área total vale, obviamente, 100%.

- A probabilidade / ocorrência de valores em intervalos $\mu \pm p\ \sigma$ é sempre igual em todas as populações normais e independente dos valores de μ e de σ (ver tabela e figuras).

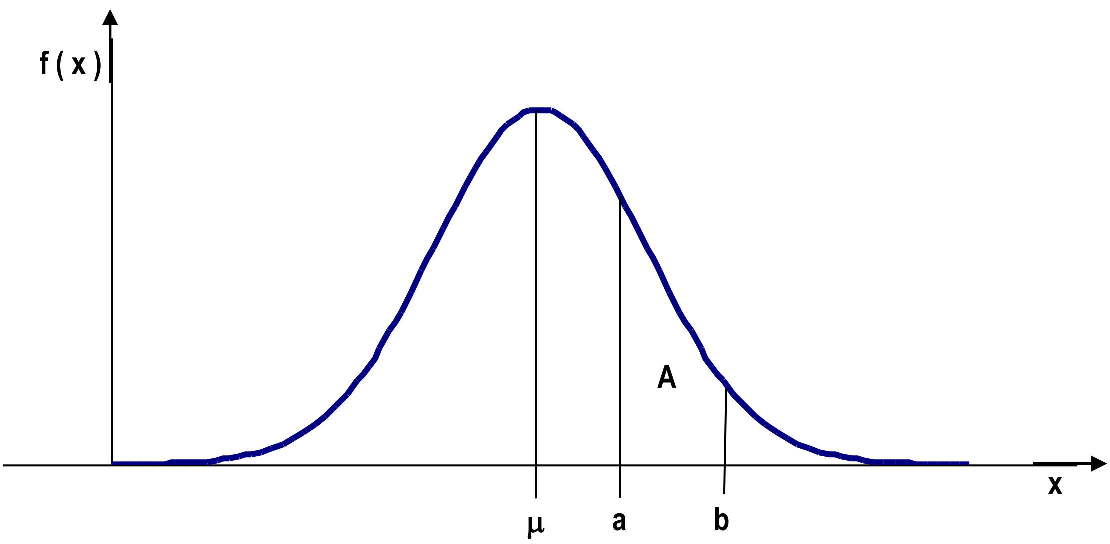


Figura – Simetria e significado da distribuição normal de probabilidades (Curva de Gauss).
- A curva é simétrica em torno de μ .
- A ocorrência de valores entre **a** e **b** é direitamente proporcional à área **A**.

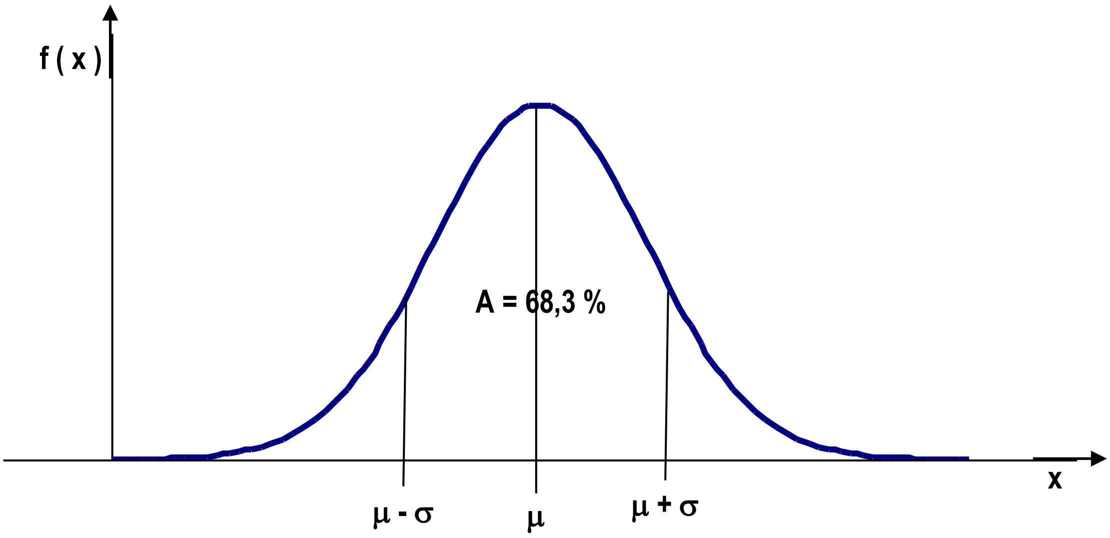


Figura – Exemplo: Na distribuição normal, a ocorrência / probabilidade de valores compreendidos entre $\mu - \sigma$ e $\mu + \sigma$ é sempre igual a 68,3 %.

Tabela – Alguns valores típicos relacionadas com uma distribuição normal.

Intervalo de valores de x	*Probabilidade / Ocorrência*
$\mu - \sigma$ **a** $\mu + \sigma$	68,27 %
$\mu - 2\sigma$ **a** $\mu + 2\sigma$	95,45 %
$\mu - 3\sigma$ **a** $\mu + 3\sigma$	99,73 %
$\mu - 4\sigma$ **a** $\mu + 4\sigma$	99,9937 %

Note que, para uma população com distribuição normal, podemos esperar encontrar a totalidade ou quase totalidade dos valores entre $\mu - 3\,\sigma$ e $\mu + 3\,\sigma$:

Exemplo: *Numa população normal com média igual a 10 e desvio padrão igual a 2:*

- 68,27 % dos valores estão compreendidos entre 8 e 12
(Note que: 10 – 2 = 8 ; 10 + 2 = 12)

- 95,45 % dos valores estão compreendidos entre 6 e 14

- 99,73 % dos valores estão compreendidos entre 4 e 16

Como a área total é 100%, teremos 100 % - 68,27 % = 31,73 % de valores abaixo de 8 e acima de 12.

Ainda dada a simetria da distribuição normal, isso significa que 31,73 % /2 ≈ 15,87 % dos valores da população são inferiores a 8 e igualmente que 15,87% dos valores são superiores a 12.

Etc.

9.3 Distribuição normal padrão

A **distribuição normal padrão** (ou **distribuição normal reduzida**) é uma distribuição normal (como outra qualquer), com as mesmas propriedades, mas com a particularidade de ser caracterizada por ter <u>média igual a 0 e desvio-padrão igual a 1</u>. Habitualmente, reserva-se a letra **Z** para designar a variável que segue uma distribuição normal padrão. Podemos portanto escrever:

$$\mathbf{Z \sim N\,(\,\mu = 0\ ,\ \sigma = 1\,)}$$

Esta distribuição assumirá um papel muito importante nos capítulos subsequentes.

Qualquer variável normalmente distribuída é convertível na variável normal padrão pela fórmula:

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

9.4 Cálculo de probabilidades

A função densidade de probabilidade característica de uma distribuição normal não é analiticamente integrável. Isso significa que não podemos ter uma expressão analítica exacta de cálculo da função distribuição.

É possível integrar numericamente a função densidade de probabilidade de uma distribuição normal e assim calcular funções cumulativas (distribuição ou outras), como rigor desejado. Existem na bibliografia muitas <u>tabelas</u> com vários valores de probabilidades da distribuição normal padrão. Como qualquer variável normal é convertível na variável normal padrão, podemos a partir de tais tabelas obter probabilidades relacionadas com qualquer distribuição normal.

Hoje em dia, já existem várias <u>ferramentas de software</u> (por exemplo, o Excel) e <u>rotinas em máquinas calculadoras</u> que calculam direitamente probabilidades para qualquer distribuição normal (consulte o manual da sua máquina calculadora).

Exemplo: *Uma variável X está distribuída de acordo com uma distribuição normal de média igual a 14,0 e desvio padrão igual a 2,5.*

A probabilidade de tal variável assumir valores superiores a 16,0 é:

$$P\,(\,X > 16{,}0\,) = P\,(\,Z > \frac{16{,}0 - 14{,}0}{2{,}5}\,) \;=\; P\,(\,Z > 0{,}80\,) =$$

$$= 1 - F\,(\,0{,}80\,) = 1 - 0{,}7881 = 0{,}2119$$

A probabilidade desta variável assumir valores entre 11,0 e 15,0 é:

$$P\,(\,11{,}0 < X < 15{,}0\,) = P\,(\frac{11{,}0 - 14{,}0}{2{,}5} < Z < \frac{15{,}0 - 14{,}0}{2{,}5}) = P\,(\,-1{,}20 < Z < 0{,}40\,) =$$

$$= F\,(\,0{,}40\,) - F\,(\,-1{,}20\,) = F\,(\,0{,}40\,) - [\,1 - F\,(\,1{,}20\,)\,] = 0{,}6554 - 1 + 0{,}8849 = 0{,}5403$$

9.5 Cálculo de probabilidades com Excel

O Excel devolve sempre a **função cumulativa** de uma distribuição normal, **F (x)** , ou seja, a área sob a curva de Gauss à esquerda de um valor x, ou ainda, a probabilidade de a variável assumir valores inferiores (ou inferiores ou iguais) a esse valor x.

Se pretendermos calcular outro tipo de probabilidades, devemos relacioná-las com a função cumulativa:

$$P(X < x) = P(X \le x) = F(x)$$

$$P(X > x) = P(X \ge x) = 1 - F(x)$$

$$P(x_1 < X < x_2) = P(x_1 < X \le x_2) = P(x_1 \le X < x_2) = P(x_1 \le X \le x_2) = F(x_2) - F(x_1)$$

Procedimento para cálculo no Excel do valor da função cumulativa, F (x):

1 – Na célula onde pretende obter o resultado, introduza **=**

2 – Seleccione a função **DIST.NORM** do sub-menú *Estatística*.

3 – Introduza os valores (ou as células contendo os valores) de *x*, da *média populacional* e do *desvio-padrão populacional*.

4 – Em Cumulativo, introduza ***1***.

5 – Seleccione OK.

Também podemos utilizar o Excel para resolver o problema inverso, isto é, conhecendo o valor da função cumulativa, descobrir o valor de x que lhe corresponde.

Procedimento para cálculo no Excel de um valor de X, a partir do valor da função cumulativa, F (x):

1 – Na célula onde pretende obter o resultado, introduza **=**

2 – Seleccione a função **INV.NORM** do sub-menú *Estatística*.

3 – Introduza os valores (ou as células contendo os valores) da função cumulativa, F (x), da *média populacional* e do *desvio-padrão populacional*.

4 – Seleccione OK.

Para a distribuição normal padrão, existem funções específicas, que dispensam a indicação dos valores da média e do desvio-padrão (uma vez que estes são sempre iguais a 0 e 1, respectivamente).

Procedimento para cálculo no Excel de um valor de X, a partir do valor da função cumulativa, F (x):

Siga os procedimentos atrás apresentados, mas substitua:

- a função DIST.NORM pela função DIST.NORMP *(NOTA: Não necessita de introduzir **1**)*
- a função INV.NORM pela função INV.NORMP

9.6 Teste simples para verificação da normalidade

Um teste simples para verificação da natureza normal de uma população a partir de uma amostra de dimensão razoável consiste em verificar se as percentagens de ocorrência de valores nos intervalos referidos na tabela 2.1 coincidem razoavelmente com as percentagens que se sabe existirem na distribuição normal. Para definir os intervalos, utilizam-se os valores da média amostral e do desvio padrão amostral.

Exemplo: *Verifique, de forma aproximada, o pressuposto da normalidade, na seguinte amostra de 50 valores:*

10,1	*11,9*	*12,1*	*12,2*	*14,8*	*15,2*	*16,0*	***16,8***	***17,3***	***17,4***	***17,6***
17,7	***18,1***	***18,3***	***18,4***	***18,7***	***18,9***	***19,2***	***19,3***	***19,4***	***19,4***	***19,7***
19,7	***19,9***	***19,9***	***20,5***	***20,7***	***20,8***	***20,8***	***21,2***	***21,3***	***21,4***	***21,6***
21,6	***21,8***	***22,1***	***22,5***	***22,6***	***22,7***	***23,0***	***23,6***	***23,6***	*24,8*	*25,5*
25,6	*26,7*	*26,9*	*27,1*	*27,3*	*27,6*					

A média e o desvio padrão amostrais já foram calculados, sendo:
$\bar{x} = 20{,}23$ *e* $s = 4{,}08$

$\bar{x} \pm s = 20{,}23 \pm 4{,}08 = 16{,}15 \ ; \ 24{,}31$

$\bar{x} \pm 2\,s = 20{,}23 \pm 2 \times 4{,}08 = 12{,}07 \ ; \ 28{,}39$

$\bar{x} \pm 3\,s = 20{,}23 \pm 3 \times 4{,}08 = 7{,}99 \ ; \ 32{,}47$

Percentagem de valores entre $\bar{x} \pm s = 35 / 50 \times 100\ \% = 70{,}0\ \% \approx 68{,}27\ \%$

Percentagem de valores entre $\bar{x} \pm 2\,s = 48 / 50 \times 100\ \% = 96{,}0\ \% \approx 95{,}45\ \%$

Percentagem de valores entre $\bar{x} \pm 3\,s = 50 / 50 \times 100\ \% = 100{,}0\ \% \approx 99{,}73\ \%$

Embora esta conclusão careça ainda de validade estatística, podemos afirmar sem grande erro que esta amostra foi retirada de uma população normal.

- FIM -