Estatística II

Aulas 3 & 4

Nelson Faustino [1]

[1]**Faculdade de Economia (FEUC)**
Universidade de Coimbra – Portugal

nelson@fe.uc.pt

Conteúdos

Provas de Bernoulli

Provas de Bernoulli são acontecimentos independentes e identicamente distribuídos

$n = 1$ — **Bernoulli**

Realizar n **provas de Bernoulli**

$n > 1$ — **Binomial**

$\theta = P($**'Sucesso'**$)$

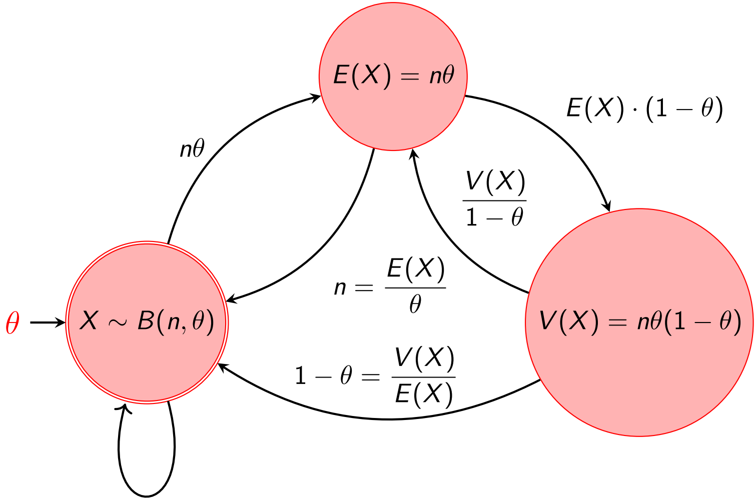
$r = 1$ — **Geométrica**

Repetir até obter r **sucessos**

$r > 1$ — **Binomial Negativa**

Distribuição Bernoulli/Binomial

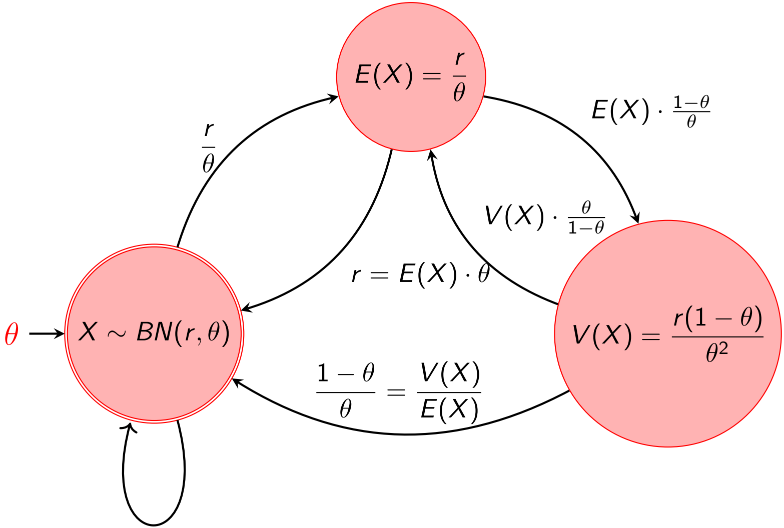
Cálculo de $E(X)$ e $V(X)$ a partir de n e θ, e vice-versa (`Semana1.pdf`)



Realizar n provas de Bernoulli

Distribuição Geométrica/Binomial Negativa

Cálculo de $E(X)$ e $V(X)$ a partir de r e θ, e vice-versa (`Semana1.pdf`)



Repetir até obter r sucessos

Teste os seus conhecimentos

Distribuições de probabilidade envolvendo provas de Bernoulli

Parâmetros n resp. r & θ [quase] conhecidos

1. Se $X \sim B(n, \theta)$ descreve uma variável aleatória que contabiliza o número de vezes que sai face 6 num dado honesto ao fim de 14 lançamentos, qual é o valor de n e θ? Adicionalmente, qual o valor de $E(X)$ e de $V(X)$?
2. Se $Y \sim BN(r, \theta)$ descreve uma variável aleatória que contabiliza o número de vezes que que precisamos de lançar um dado honesto até sair duas faces 6, qual é o valor de n e θ? Adicionalmente, qual o valor de $E(X)$ e de $V(X)$?

Valores da média e variância conhecidos

1. Se $Z \sim B(n, \theta)$ é tal que $E(Z) = 2.86$ e $V(Z) = 2.6741$, qual o valor dos parâmetros n e θ?
2. Se $W \sim BN(r, \theta)$ é tal que $E(W) = 150$ e $V(W) = 3600$, qual o valor dos parâmetros r e θ?

Distribuição de Poisson

Probabilidade de **ocorrências por unidade de tempo** (minutos,horas,dias,etc)

Função de probabilidade ($X \sim Po(\lambda)$)

$$f(x) = \begin{cases} e^{-\lambda}\dfrac{\lambda^x}{x!} & (x = 0, 1, \ldots) \\ 0 & (\text{outros } x) \end{cases}$$

Principais características

- **FGM:** $M(s) = e^{\lambda(e^s - 1)}$.
- **Média:** $E(X) = \lambda$, sendo $\lambda = M'(0)$.
- **Variância:** $V(X) = \lambda$, sendo $\lambda = M''(0) - (M'(0))^2$.

Distribuição de Poisson

Probabilidade de ocorrências em **intervalos de tempo de duração** t.

Função de probabilidade ($X(t) \sim Po(\lambda t)$)

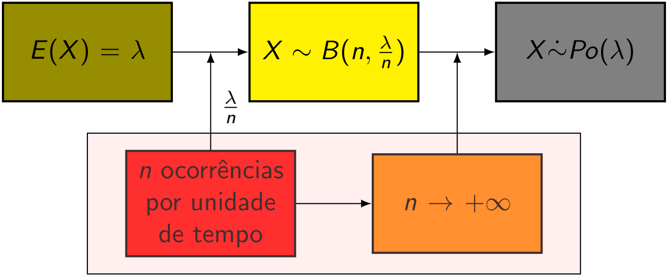
$$f(x) = \begin{cases} e^{-\lambda t} \dfrac{(\lambda t)^x}{x!} & (x = 0, 1, \ldots) \\ 0 & (\text{outros } x) \end{cases}$$

Parâmetros λ e t de $Po(\lambda t)$

- $\lambda > 0$: **Média de ocorrências** por unidade de tempo.
- $t > 0$: Intervalo de tempo de **duração** t.
- $\lambda t > 0$: **Média de ocorrências** em intervalos de tempo de **duração** t.

Processo de Poisson

Probabilidade do número ocorrências (n) que ocorrem em média (λ) por unidade de tempo



Notação: $X \dot\sim Po(\lambda)$ **lê-se** X *segue assintoticamente uma distribuição de Poisson, de parâmetro* λ.

Processo de Poisson

Limite do Processo de Poisson

Limite Notável

$$\lim_{n\to+\infty}\left(1-\frac{\lambda}{n}\right)^{n-x}=e^{-\lambda}$$

Propriedades dos Limites

$$\lim_{n\to+\infty}\frac{n!}{n^x(n-x)!}=1$$

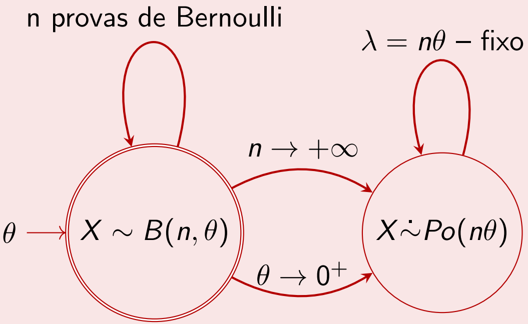
Limite do Processo de Poisson

$$\begin{aligned}
\lim_{n\to+\infty}P(X=x) &= \lim_{n\to+\infty}\binom{n}{x}\left(\frac{\lambda}{n}\right)^{x}\left(1-\frac{\lambda}{n}\right)^{n-x}\\
&= \lim_{n\to+\infty}\cancel{\left(1-\frac{\lambda}{n}\right)^{n-x}}^{\,e^{-\lambda}}\frac{\lambda^x}{x!}\,\cancel{\frac{n!}{n^x(n-x)!}}^{\,1}\\
&= e^{-\lambda}\frac{\lambda^x}{x!}
\end{aligned}$$

Poisson vs. Binomial

Aproximação de Binomial ($X \sim B(n,\theta)$) por Poisson ($X \dot{\sim} Po(n\theta)$)

Lei dos Acontecimentos Raros



Empiricamente, o resultado acima verifica-se quando:

1. Número de provas de Bernoulli é elevado (p.e. $n > 20$);
2. $\theta \approx 0$ suficientemente pequeno (p.e. $0 < \theta \leq 0.1$).

Exercícios Resolvidos

Exercício 15., página 328

Dados do Exercício 15.

- **Distribuição de Probabilidade:** $X \sim Po(4)$ descreve o número de reclamações por dia.

Resolução 15.a)

A probabilidade de haver dias sem reclamações é dada por

$$P(X=0) = e^{-4}\frac{4^0}{0!} = e^{-4}$$

Após o cálculo, conclui-se que e^{-4} é aproximadamente igual a 1.83% (2 c.d.) (**Resolução 1**).

Resolução em R

Resolução 1

```
> round(exp(-4)*100,2)
[1] 1.83
```

Resolução 2

```
> p15a<-dpois(0,4)
> p15a
[1] 0.01831564
> round(p15a*100,2)
[1] 1.83
```

Exercícios Resolvidos

Exercício 15., página 328

Probabilidade a calcular em 15.b)

A probabilidade de receber entre três (3) a seis (6) queixas é dada por $P(3 \leq X \leq 6)$. Esta pode ser determinada usando uma das seguintes propriedades:

1. $P(3 \leq X \leq 6) = P(X=3) + P(X=4) + P(X=5) + P(X=6)$ (comando **dpois**);
2. $P(3 \leq X \leq 6) = P(X \leq 6) - P(X \leq 2)$ (comando **ppois**)

Comando dpois

```
> dpois(3:6,4)
[1] 0.1953668 0.1953668
0.1562935 0.1041956
> sum(dpois(3:6,4))
[1] 0.6512227
```

Comando ppois

```
> ppois(6,4)-ppois(2,4)
[1] 0.6512227
```

Exercícios Resolvidos

Exercício 15., página 328

Probabilidade a calcular em 15.c)

A variável aleatória $X(5) \sim Po(20)$ ($\lambda t = 20$, sendo $\lambda = 4$ e $t = 5$) dá-nos o número de chamadas de reclamações numa semana de cinco (5) dias úteis. Neste caso,

$$P(X(5) = 15) = e^{-20} \frac{20^{15}}{15!}$$

corresponde ao valor exacto da probabilidade pedida.

Resolução em R

Resolução 1

```
> exp(-20)*20^(15)/factorial(15)
[1] 0.05164885
```

Resolução 2

```
> dpois(15,20)
[1] 0.05164885
```

Exercícios Resolvidos

Exercício 15., página 328

Variáveis Aleatórias em 15.d)

Para descrever cada reclamação em cada um dos dias da semana, precisamos de considerar cinco (5) variáveis aleatórias identicamente distribuídas X_1, X_2, X_3, X_4, X_5, com $X_i \sim Po(4)$ $(i = 1, 2, \ldots, 5)$.

Probabilidade pedida em 15.d)

Sendo $A_i = \{X_i \geq 1\}$ $(i = 1, 2, \ldots, 5)$ o acontecimento que descreve **haver pelo menos uma reclamação no $i-$ésimo dia da semana**, tem-se que

$$A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5 = \{X_1 \geq 1 \wedge X_2 \geq 1 \wedge X_3 \geq 1 \wedge X_4 \geq 1 \wedge X_5 \geq 1\}.$$

Portanto, $P(A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5)$ corresponde à **probabilidade pedida** em **15.d)**.

Exercícios Resolvidos

Exercício 15., página 328

Variáveis Aleatórias em 15.d) são independentes

Uma vez que o número de reclamações é independente do dia da semana considerado, retiramos que os acontecimentos $A_i = \{X_i \geq 1\}$ $(i = 1, 2, \ldots, 5)$ são mutualmente independentes (cf. [Murteira et al. (2015), Definição 2.13]). Segue então que

$$P(A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5) = P(A_1)P(A_2)P(A_3)P(A_4)P(A_5).$$

Variáveis Aleatórias em 15.d) são ainda identicamente distribuídas

Da propriedade $P(X_i \geq 1) = 1 - P(X = 0)$ $(i = 1, 2, \ldots, 4 \ \& \ X \sim Po(4))$ e da fórmula anterior concluímos, após algumas simplificações, que $(1 - e^{-4})^5$ é o **valor exacto da probabilidade pedida**.

Em R: Comandos `> (ppois(0,4,lower.tail = FALSE))^5` ou `> (1-dpois(0,4))^5`.

Exercícios Resolvidos

Exercício 30., página 332

Variáveis Aleatórias em 30.

A **variável aleatória (v.a.)** $X \sim Po(\lambda)$ **número de gralhas por página de uma revista**, e

$$Y = \begin{cases} 1 & \text{(composição feita pela Empresa 1)} \\ 2 & \text{(composição feita pela Empresa 2)} \end{cases}$$

Probabilidades condicionais

Sendo $X|Y = y$ a **v.a. condicional** que descreve a distribuição de gralhas cometidas na composição feita pela **Empresa** y **(**$y = 1, 2$**)**, tem-se

$$X|Y = 1 \sim Po(0.2) \quad \& \quad X|Y = 2 \sim Po(0.3)$$

Tem-se ainda que $P(Y = 1) + P(Y = 2) = 1$, sendo $P(Y = 1) = 0.6$ (60%) **– dado no enunciado –** e $P(Y = 2) = 0.4$ (40%) **– calculado**.

Resolução 30. a)

Aplica-se a **regra da probabilidade total** (cf. [Murteira et al. (2015), Teorema 2.1]) para calcular a percentagem (em probabilidade) de páginas da revista sem gralhas:

$$\begin{aligned} P(X=0) &= P(X=0|Y=1)P(Y=1) \\ &+ P(X=0|Y=2)P(Y=2) \\ &= e^{-0.2}\times 0.6 + e^{-0.3}\times 0.4. \end{aligned}$$

Resolução em R

```
> p1<-0.6
> p2<-1-p1
> pX1<-dpois(0,0.2)
> pX2<-dpois(0,0.3)
> pX1*p1+pX2*p2
[1] 0.7875657
```

Exercícios Resolvidos

Exercício 30., página 332

Resolução 30. b)

Sendo $P(Y=2)=0.4\ (40\%)$ a probabilidade da empresa que compõe menos páginas para a revista ($P(Y=2)<P(Y=1)$), aplica-se o **Teorema de Bayes** (cf. [Murteira et al. (2015), Teorema 2.2 da p. 98]) para calcular $P(Y=2|X=0)$ (probabilidade pedida). Em concreto:

$$P(Y=2|X=0)=\frac{P(X=0|Y=2)P(Y=2)}{P(X=0)},$$

sendo $P(X=0)$ a probabilidade calculada em **30. a)** usando a **regra da probabilidade total**.

Resolução em R

```
> p30a<-pX1*p1+pX2*p2
> (pX2*p2)/p30a
[1] 0.3762572
```

Exercícios Resolvidos

Exercício 30., página 332

Variável aleatória em 30. c)

A variável aleatória $Z \sim B(20, \theta)$ descreve o número de páginas da revista com gralhas ($n = 20$ é o **número de provas de Bernoulli**), sendo θ a probabilidade de uma página não conter gralha, i.e.

'Sucesso'='Página sem gralhas' $\Longrightarrow \theta = P(X = 0)$ (**30.a)**)

Cálculo da probabilidade pedida usando o R

A probabilidade $P(Z = 20) = (P(X = 0))^{20}$ (20 páginas sem gralhas) pode ser calculada de dois modos distintos:

Resolução 1:

```
> (p30a)^(20)
[1] 0.008428229
```

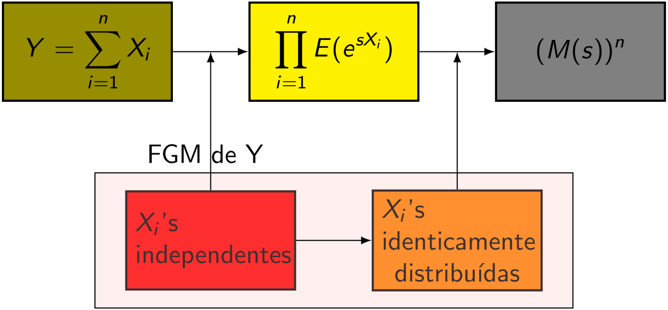
Resolução 2:

```
> dbinom(20,20,p30a)
[1] 0.008428229
```

Conteúdos

Independência de Variáveis Aleatórias (v.a.'s)

Relação com a função geradora de momentos (FGM)



- $M_{X_i}(s) = E(e^{sX_i})$ FGM de X_i & $M_Y(s) = E(e^{sY})$ FGM de Y;
- **X_i's são independentes:** $M_Y(s) = \prod_{i=1}^{n} E(e^{sX_i})$;
- **X_i's são i.i.d's:** $M_Y(s) = (M(s))^n$, sendo $E(e^{sX_i}) = M(s)$.

Independência de Variáveis Aleatórias (v.a.'s)

Distribuições Binomial, Binomial Negativa e de Poisson

Binomial – cf. [Murteira et al. (2015), Teorema 5.1]

$$X_i \sim B(n_i, \theta) \text{ independentes} \Longrightarrow \sum_{i=1}^{n} X_i \sim B(n, \theta), \quad \text{com } n = \sum_{i=1}^{n} n_i.$$

Binomial Negativa – cf. [Murteira et al. (2015), Teorema 5.2]

$$X_i \sim BN(r_i, \theta) \text{ independentes} \Longrightarrow \sum_{i=1}^{n} X_i \sim B(r, \theta), \quad \text{com } r = \sum_{i=1}^{n} r_i.$$

Poisson – cf. [Murteira et al. (2015), Teorema 5.3]

$$X_i \sim Po(\lambda_i) \Longrightarrow \sum_{i=1}^{n} X_i \sim Po(\lambda), \quad \text{com } \lambda = \sum_{i=1}^{n} \lambda_i.$$

Provas de Bernoulli

Provas de Bernoulli Independentes e Identicamente Distribuídas (i.i.d's)

$$\theta = P(\textbf{'Sucesso'})$$

$$\theta = P(\textbf{'Sucesso'}) \longrightarrow X \sim B(1,\theta)$$

$$X \sim B(1,\theta) \longrightarrow n = 1 \text{ — } X_1 \sim B(1,\theta)$$

$$X \sim B(1,\theta) \longrightarrow n > 1 \text{ - } \sum_{i=1}^{n} X_i \sim B(n,\theta)$$

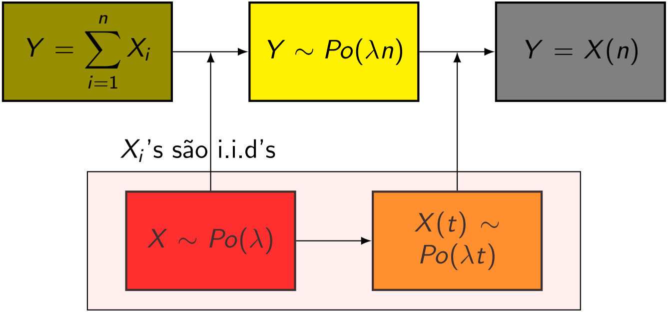
$$\theta = P(\textbf{'Sucesso'}) \longrightarrow X \sim BN(1,\theta)$$

$$X \sim BN(1,\theta) \longrightarrow r = 1 \text{ — } X_1 \sim BN(1,\theta)$$

$$X \sim BN(1,\theta) \longrightarrow r > 1 \text{ - } \sum_{i=1}^{r} X_i \sim BN(r,\theta)$$

Ocorrências por Unidade de Tempo

Probabilidade de ocorrências em n intervalos de tempo unitário ($t = 1$)

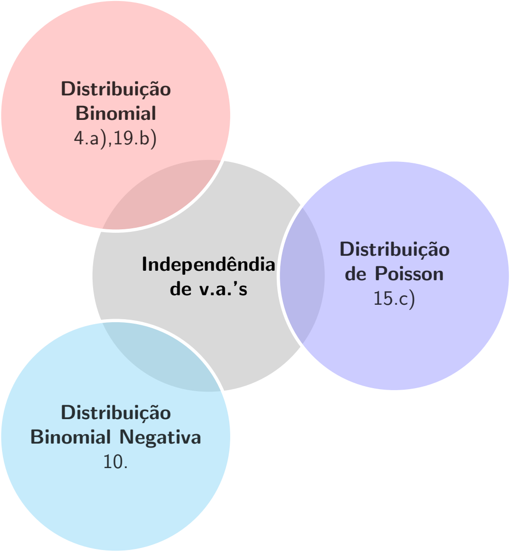


Variáveis Aleatórias:

- $\sum_{i=1}^{n} X_i$ – nº de ocorrências em n unidades de tempo unitário.
- $X(t)$ – nº de ocorrências em intervalos de tempo de duração t.

Independência de Variáveis Aleatórias (v.a.'s)

Alguns exercícios do Capítulo 5 em que pode aplicar os resultados do slide anterior



FEUC FACULDADE DE ECONOMIA
UNIVERSIDADE DE COIMBRA

Conteúdos

Distribuição Uniforme Contínua

Não confundir com distribuição uniforme discreta tratada na Aula 2

Função densidade de probabilidade ($X \sim U(a, b)$)

$$f(x) = \begin{cases} \dfrac{1}{b-a} & (a < x < b) \\ 0 & (\text{outros } x) \end{cases}$$

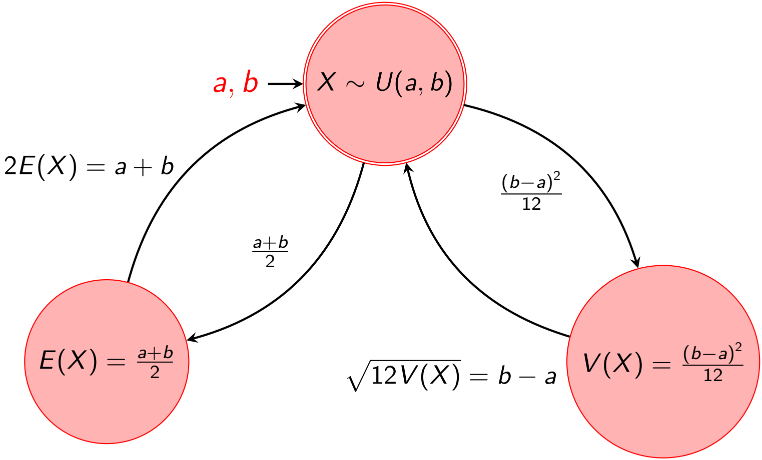
Quantidades de interesse

- **Parâmetros** a, b **(**$a < b$**):** Intervalo de ocorrências $]a, b[$.
- **Média:** $E(X) = \frac{a+b}{2}$ (ponto médio do intervalo $]a, b[$).
- **Variância:** $V(X) = \frac{(b-a)^2}{12}$.

Para mais detalhes: Vide [Murteira et al. (2015), pp. 268–271].

Distribuição Uniforme Contínua

Cálculo de $E(X)$ e $V(X)$ a partir de a e b, e vice-versa



Distribuição Normal

Usada com frequência para aproximação assintótica de outras distribuições de probabilidade

Função densidade de probabilidade ($X \sim N(\mu, \sigma^2)$)

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (-\infty < x < +\infty)$$

Principais características

- **FGM:** $M(s) = e^{\mu s + \frac{\sigma^2 s^2}{2}}$
- **Média:** $E(X) = \mu$, sendo $\mu = M'(0)$.
- **Variância:** $V(X) = \sigma^2$, sendo $\sigma^2 = M''(0) - (M'(0))^2$.

Distribuição Normal $N(\mu, \sigma^2)$

Relação com a distribuição normal estandartizada, $N(0, 1)$

$\Phi(z) = P(Z \leq z)$, $Z \sim N(0, 1)$

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}} dt$$

Mudança de variável

$$t = \frac{s-\mu}{\sigma} \Rightarrow dt = \frac{1}{\sigma} ds$$

$$s \leq x \Rightarrow t \leq \frac{x-\mu}{\sigma}$$

Função de distribuição $X \sim N(\mu, \sigma^2)$

$$\begin{aligned}
P(X \leq x) &= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(s-\mu)^2}{2\sigma^2}} ds \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{\left(\frac{s-\mu}{\sigma}\right)^2}{2}} \frac{1}{\sigma} ds \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-\mu}{\sigma}} e^{-\frac{t^2}{2}} dt \\
&= \Phi\left(\frac{x-\mu}{\sigma}\right)
\end{aligned}$$

Distribuição Normal $N(\mu, \sigma^2)$

Relação com a distribuição normal estandartizada, $N(0,1)$

Estandartização da v.a. $X \sim N(\mu, \sigma^2)$

$$\mu, \sigma^2 \longrightarrow X \sim N(\mu, \sigma^2) \xrightarrow[\sigma]{\mu} \frac{X-\mu}{\sigma} \sim N(0,1)$$

Obs: Sempre que X é uma v.a. de uma distribuição normal, segue que

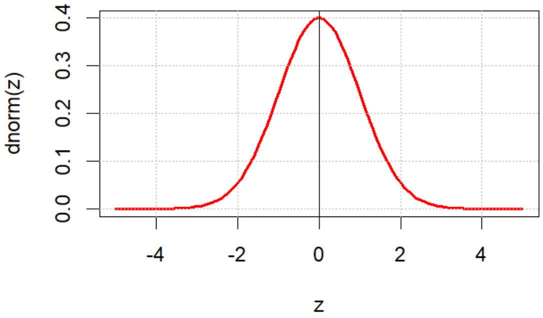
$$Z = \frac{X - E(X)}{\sqrt{V(X)}} \sim N(0,1).$$

Distribuição Normal Estandartizada

Representação Gráfica

Função Densidade de Probabilidade (f.d.p.) $Z \sim N(0,1)$

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad (-\infty < z < +\infty)$$



Distribuição Normal Estandartizada

Simetria da função de distribuição em relação ao eixo Oy

Função de distribuição $Z \sim N(0,1)$

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}}\, dt$$

f.d.p. $Z \sim N(0,1)$

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}}\, dt = 1$$

Propriedade $\Phi(-z) = 1 - \Phi(z)$

$$\begin{aligned}
\Phi(z) + \Phi(-z) &= \lim_{R\to+\infty} \frac{1}{\sqrt{2\pi}} \int_{-R}^{z} e^{-\frac{t^2}{2}}\, dt \\
&+ \lim_{R\to+\infty} \frac{1}{\sqrt{2\pi}} \int_{-R}^{-z} e^{-\frac{t^2}{2}}\, dt \quad (t = -u) \\
&= \frac{1}{\sqrt{2\pi}} \lim_{R\to+\infty} \left(\int_{-R}^{z} e^{-\frac{t^2}{2}}\, dt + \int_{z}^{R} e^{-\frac{u^2}{2}}\, du \right) \\
&= 1 \quad \textbf{(após simplificações)}.
\end{aligned}$$

Função **pnorm**

Síntaxe

Comandos R para calcular função distribuição da normal:

- **pnorm(z)** – Calcula, por defeito, $P(Z \leq z) = \Phi(z)$ $(Z \sim N(0,1))$;
- **pnorm(z,lower.tail = FALSE)** – Calcula, por defeito, $P(Z > z) = 1 - \Phi(z)$ $(Z \sim N(0,1))$;
- **pnorm(x,mean,sd)** – Calcula $P(X \leq x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$ ($\mu =$**mean**;$\sigma =$**sd**) para $X \sim N(\mu, \sigma^2)$.
- **pnorm(x,mean,sd,lower.tail = FALSE)** – Calcula $P(X > x) = 1 - \Phi\left(\frac{x-\mu}{\sigma}\right)$ ($\mu =$**mean**;$\sigma =$**sd**) para $X \sim N(\mu, \sigma^2)$.

Função **qnorm**

Síntaxe

Comandos R para calcular os quantis da função distribuição da normal:

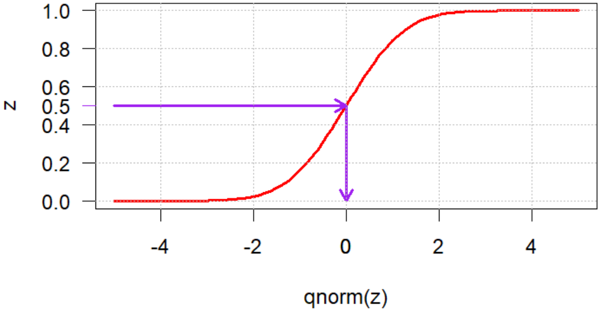
- **qnorm(q)** – Calcula, por defeito, $\Phi^{-1}(q)$;
- **qnorm(q,lower.tail = FALSE)** – Calcula, por defeito, o valor de z tal que $P(Z > z) = q$ (valor $\Phi^{-1}(1-q)$);
- **qnorm(q,mean,sd)** – Calcula o valor de x tal que $P(X \leq x) = q$ ($\mu =$**mean**;$\sigma =$**sd**) para $X \sim N(\mu, \sigma^2)$.
- **qnorm(q,mean,sd,lower.tail = FALSE)** – Calcula o valor de x tal que $P(X > x) = q$ ($\mu =$**mean**;$\sigma =$**sd**) para $X \sim N(\mu, \sigma^2)$.

Distribuição Normal Estandartizada

Representação Gráfica dos Quantis de ordem α, q_α $(\Phi(q_\alpha) = \alpha)$

Mediana em R

```
> qnorm(0.5)
[1] 0
```



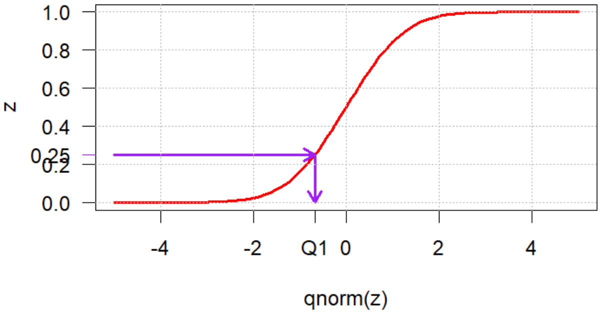
Distribuição Normal Estandartizada

Representação Gráfica dos Quantis de ordem α, q_α $(\Phi(q_\alpha) = \alpha)$

Estatística II
N. Faustino
Distribuições discretas
Binomial & Binomial Negativa
Poisson
Poisson vs. Binomial
Exercícios Resolvidos
Independência de v.a.'s
Função Geradora de Momentos
Resultados para v.a.'s discretas
Distribuições Contínuas
Uniforme Contínua
Normal
Exercícios Resolvidos
Exercícios Extra-Aula
Implementação em R

1^o quartil em R

```
> qnorm(0.25)
[1] -0.6744898
```

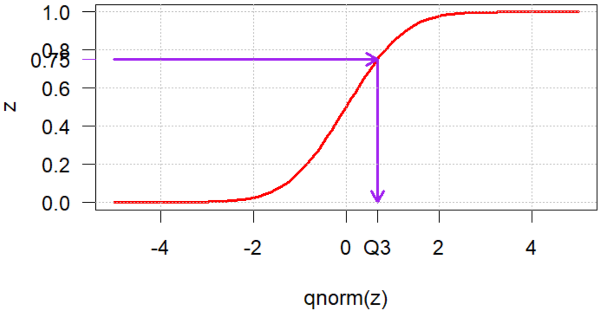


Distribuição Normal Estandartizada

Representação Gráfica dos Quantis de ordem α, q_α $(\Phi(q_\alpha) = \alpha)$

3^o quartil em R

```
> qnorm(0.75)
[1] 0.6744898
```



Exercícios Resolvidos

Exercício 34., página 333

Função distribuição a calcular 34.a) – vide [Murteira et al. (2015), p. 269]

$$F(x)=\int_{-\infty}^{x} f(u)du, \text{ sendo } f(x)=\tfrac{1}{7} \quad (5<x<12)$$

Caso $x<5$: $F(x)=\int_{-\infty}^{x} \overbrace{f(u)}^{=0} du = 0.$

Caso $5\leq x<12$:

$$F(x)=\int_{-\infty}^{5} f(u)du+\int_{5}^{x} f(u)du=\overbrace{F(5)}^{=0}+\int_{5}^{x}\frac{1}{7}du=\frac{x-5}{7}.$$

Caso $x\geq 12$: $F(x)=\int_{-\infty}^{12} f(u)du+\int_{12}^{x}\overbrace{f(u)}^{=0}du=\overbrace{F(12)}^{=\frac{12-5}{7}}=1.$

Solução 34.a)

$$F(x) = \begin{cases} 0 & (x < 5) \\ \dfrac{x-5}{7} & (5 \leq x < 12) \\ 1 & (x \geq 12) \end{cases}$$

Solução 34.a) em R – representação gráfica

Introduza os comandos abaixo para representar a função de distribuição de $X \sim U(5, 12)$.

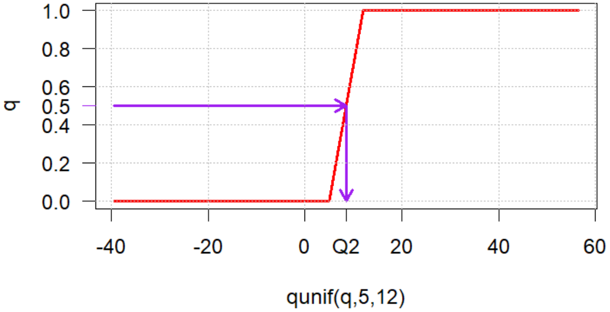
```
> a<-5
> b<-12
> m<-(a+b)/2
> Uab<-function(x){punif(x,a,b)}
> curve(Uab,col="red",lwd=2,from = m-4*b, to = m+4*b,
xlab ="x",ylab = "F(x)")
> grid(col="grey")
```

Distribuição Uniforme $U(5, 12)$

Representação gráfica dos quantis de $F(x)$ calculada em 34.a)

Mediana em R

```
> qunif(0.5,5,12)
[1] 8.5
```

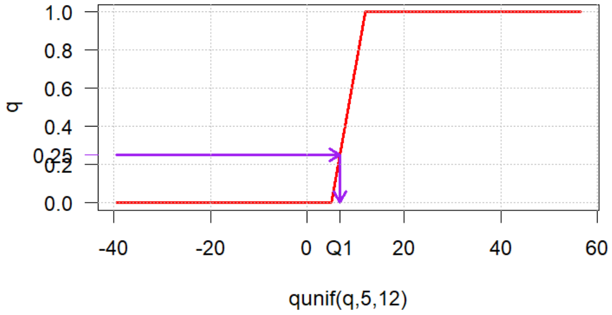


Distribuição Uniforme $U(5, 12)$

Representação gráfica dos quantis de $F(x)$ calculada em 34.a)

1^o quartil em R

```
> qunif(0.25,5,12)
[1] 6.75
```

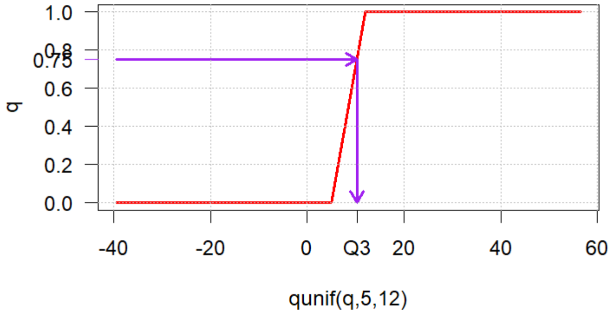


Distribuição Uniforme $U(5, 12)$

Representação gráfica dos quantis de $F(x)$ calculada em 34.a)

3^o quartil em R

```
> qunif(0.75,5,12)
[1] 10.25
```



Exercícios Resolvidos

Exercício 34., página 333

Resolução 34.b) – usando $F(x)$ calculada em 34.a)

A probabilidade de um pequeno anúncio ter a duração superior a 7 segundos corresponde é dada por $P(X > 7) = 1 - F(7) = 1 - \frac{2}{7} = \frac{5}{7}$.

34.b) em R > `punif(q,min,max)`

```
> punif(7,5,12)
[1] 0.2857143
> 2/7
[1] 0.2857143
> 1-punif(7,5,12)
[1] 0.7142857
```

34.b) em R > `punif(q,min,max,lower.tail = FALSE)`

```
> punif(7,5,12,lower.tail = FALSE)
[1] 0.7142857
```

Resolução 34.c) – análoga à resolução de 27.b)

Usar [Murteira et al. (2015), Definição 2.10], propriedades da função distribuição e a solução obtida em **34.a)**.

34.c) em R > `punif(q,min,max)`

```
> (punif(10,5,12)-punif(6,5,12))/punif(10,5,12)
[1] 0.8
> 4/5
[1] 0.8
```

Resolução 34.d)

Usar fórmulas $E(X) = \frac{a+b}{2}$ e $V(X) = \frac{(b-a)^2}{12}$ associadas a $X \sim U(a,b)$.

Obs: O desvio padrão corresponde a $\sqrt{V(X)} = \frac{b-a}{\sqrt{12}}$ $(a < b)$.

Distribuição Normal Estandartizada

$$Z \sim N(0,1): \quad \Phi(z) = P(Z \leq z) \;\&\; \Phi(-z) = 1 - \Phi(z)$$

38.a)

$$P(0 < Z \leq 2.05) = \\ = \Phi(2.05) - \Phi(0)$$

38.a) > pnorm(q)

```
> pnorm(2.05)-pnorm(0)
[1] 0.4798178
```

38.b)

$$P(-1.22 < Z \leq 1.05) = \\ = \Phi(1.05) - \Phi(-1.22) \\ = \Phi(1.05) - (1 - \Phi(1.22))$$

38.(b) – > pnorm(q)

```
> pnorm(1.05)-pnorm(-1.22)
[1] 0.7419085
> pnorm(1.05)-(1-pnorm(1.22))
[1] 0.7419085
```

Exercícios Resolvidos

Exercício 38, página 334

Distribuição Normal Estandartizada

$$Z \sim N(0,1): \quad \Phi(z) = P(Z \leq z) \ \& \ \Phi(-z) = 1 - \Phi(z)$$

38.c)

$$P(Z \geq -2.05) = 1 - \Phi(-2.05) = \Phi(2.05)$$

38.c) > pnorm(q)

```
> 1-pnorm(-2.05)
[1] 0.9798178
> pnorm(2.05)
[1] 0.9798178
```

38.c) > pnorm(q,lower.tail = FALSE)

```
> pnorm(-2.05,lower.tail = FALSE)
[1] 0.9798178
```

38.d)

Quantil k de $N(0,1)$: $\quad P(Z \leq k) = w \Longleftrightarrow w = \Phi^{-1}(k)$

$$
\begin{aligned}
\overbrace{P(|Z| > k)}^{=1-P(|Z|\leq k)} = 0.05 \quad &\Longleftrightarrow \quad P(|Z| \leq k) = 0.95 \\
&\Longleftrightarrow \quad P(-k \leq Z \leq k) = 0.95 \\
&\Longleftrightarrow \quad \Phi(k) - \underbrace{\Phi(-k)}_{=1-\Phi(k)} = 0.95 \\
&\Longleftrightarrow \quad \Phi(k) = \frac{1.95}{2} \Leftrightarrow k = \Phi^{-1}\left(\frac{1.95}{2}\right)
\end{aligned}
$$

38.d) > qnorm(p)

```
> qnorm(1.95/2)
[1] 1.959964
```

38.(e)

Quantil k de $N(0,1)$: $\quad P(Z \leq k) = w \iff w = \Phi^{-1}(k)$

$$
\begin{aligned}
P(|Z| < k) = 0.90 \quad &\iff \quad P(-k \leq Z \leq k) = 0.90 \\
&\iff \quad \Phi(k) - \underbrace{\Phi(-k)}_{=1-\Phi(k)} = 0.90 \\
&\iff \quad \Phi(k) = \frac{1.90}{2} (= 0.95) \\
&\iff \quad k \approx 1.64 \quad (2 \text{ c.d.})
\end{aligned}
$$

38.(e) > qnorm(p)

```
> qnorm(1.90/2)
[1] 1.644854
> round(qnorm(0.95),2)
[1] 1.64
```

Exercícios Resolvidos

Exercício 39, página 334

Distribuição Normal de média $\mu = 6$ e variância $\sigma^2 = 25$

$$X \sim N(6,25) \iff Z = \frac{X-6}{5} \sim N(0,1)$$

$$P(X \leq x) = \Phi\left(\frac{x-6}{5}\right)$$

39.a)

$$P(6 < X \leq 12) =$$
$$= \underbrace{P(X \leq 12)}_{=\Phi\left(\frac{12-6}{5}\right)} - \underbrace{P(X \leq 6)}_{=\Phi\left(\frac{6-6}{5}\right)}$$
$$= \Phi\left(\frac{6}{5}\right) - \Phi(0)$$
$$= \Phi(1.2) - \Phi(0)$$

39.a) > pnorm(q,mean,sd)

```
> pnorm(12,6,5)-pnorm(6,6,5)
[1] 0.3849303
```

39.a) > pnorm(q)

```
> pnorm(1.2)-pnorm(0)
[1] 0.3849303
```

39.d) – via acontecimento contrário

$$\begin{aligned} P(|X-6|>10) &= 1-P(|X-6|\leq 10) \\ &= 1-P(-10\leq X-6\leq 10) \\ &= 1-(\Phi(2)-\Phi(-2)) \quad \text{[Porquê ?]} \\ &= 2(1-\Phi(2)) \quad \text{[Porquê ?]} \end{aligned}$$

Obs: $P(|X-6|\leq 10) = P(-4\leq X\leq 16)$

39.d) `> pnorm(q)` ou `> pnorm(q,mean,sd)`

```
> 2*(1-pnorm(2))
[1] 0.04550026
> 1-(pnorm(2)-pnorm(-2))
[1] 0.04550026
> 1-(pnorm(16,6,5)-pnorm(-4,6,5))
[1] 0.04550026
```

Exercícios Resolvidos

Exercício 39, página 334

39.d) – via acontecimentos mutualmente exclusivos

$$\begin{aligned} P(|X-6|>10) &= P(X-6>10 \vee X-6<-10) \\ &= \underbrace{P(X-6>10)}_{=P(X>16)} + \underbrace{P(X-6<-10)}_{=P(X<-4)} \\ &= (1-\Phi(2)) + \Phi(-2) \\ &= 2(1-\Phi(2)) \quad [\text{Porquê ?}] \end{aligned}$$

39.d) `> pnorm(q,mean,sd)` e/ou `> pnorm(q,mean,sd)`

```
> pnorm(16,6,5,lower.tail = FALSE)+pnorm(-4,6,5)
[1] 0.04550026
> (1-pnorm(16,6,5))+pnorm(-4,6,5)
[1] 0.04550026
> 1-pnorm(2)+pnorm(-2)
[1] 0.04550026
```

Cálculo de quantis q: $P(X \leq q) = p \Longleftrightarrow P(X > q) = 1 - q$

39.e)

$$
\begin{aligned}
P(X > k) = 0.90 &\Longleftrightarrow \overbrace{P(X \leq k)}^{=\Phi\left(\frac{k-6}{5}\right)} = 0.10 \\
&\Longleftrightarrow \Phi\left(\frac{6-k}{5}\right) = 0.90 \quad \text{[Porquê ?]} \\
&\Longleftrightarrow k = 6 - 5 \times \Phi^{-1}(0.90) \quad \text{[Porquê ?]}
\end{aligned}
$$

39.(e) `> pnorm(q,mean,sd,lower.tail = FALSE)`

```
> qnorm(0.90,6,5,lower.tail = FALSE)
[1] -0.4077578
> 6-5*qnorm(0.90)
[1] -0.4077578
```

Exercícios Resolvidos

Exercício 42., página 335

Dados do Exercício 42.a)

Distribuição de Probabilidade: $X \sim N(330, 16)$ sendo $\mu = 330$ ml a média e $\sigma^2 = 16$ ml^2 a variância ($\sigma = 4$ ml).

Resolução 42.a)

A percentagem pedida é $P(|X-330|>6)\times 100\%$ (percentagem de garrafas rejeitadas cujo valor absoluto difere de $\mu = 330$), onde a probabilidade associada pode ser calculada do seguinte modo:

$$
\begin{aligned}
P(|X-330|>6) &= 1 - P(|X-330|\leq 6) \quad \text{[Porquê ?]}\\
&= 1 - P(-6 \leq X-330 \leq 6)\\
&= 1 - P\left(-\frac{6}{4} \leq \frac{X-330}{4} \leq \frac{6}{4}\right)\\
&= 1 - (\Phi(1.5) - \Phi(-1.5)) \quad \text{[Porquê ?]}\\
&= 2(1-\Phi(1.5)). \quad \text{[Porquê ?]}
\end{aligned}
$$

Exercícios Resolvidos

Exercício 42., página 335

Percentagem de 42.a) calculado em R (arred. 2 c.d.)

```
> prob42a<-2*(1-pnorm(1.5))
> round(prob42a*100,2)
[1] 13.36
```

42a.) via > pnorm(q,mean,sd)

Cálculo $P(|X-330|>6)\times 100\%$ **baseado na seq. de igualdades**

$$\begin{aligned} P(|X-330|>6) &= 1-P(324<X\leq 336) \\ &= 1-(P(X\leq 336)-P(X\leq 324)) \end{aligned}$$

```
> mean<-330
> sd<-4
> prob42aV2<-1-(pnorm(336,mean,sd)-pnorm(324,mean,sd))
> round(prob42aV2*100,2))
[1] 13.36
```

Exercícios Resolvidos

Exercício 42., página 335

42a.) via > pnorm(q,mean,sd,lower.tail=FALSE)

Baseado na identidade

$$\begin{aligned} P(|X-330|>6) &= P(X-330>6 \ \vee \ X-300<-6) \\ &= P(X>336)+P(X<324) \end{aligned}$$

```
> mean<-330
> sd<-4
> prob42aV3<-pnorm(336,mean,sd,lower.tail =
FALSE)+pnorm(324,mean,sd)
> round(prob42aV3*100,2))
[1] 13.36
```

Dados do Exercício 42.b)

Variável aleatória: Precisamos de considerar a soma de 100 v.a.'s independentes e identicamente distribuídas (i.i.d.'s)

$$Y = \sum_{i=1}^{100} X_i$$

para descrever a quantidade de líquido colocado em cada garrafa, com $X_i \sim N(330, 16)$ ($i-$ésima garrafa).

Resolução 42.b) – vide [Murteira et al. (2015), Corolário 5.1]

$$X_i \sim N(330, 16) \; (i = 1, 2, \ldots, 100) \quad \Longrightarrow \quad \underbrace{\sum_{i=1}^{100} X_i}_{=Y} \sim N(33000, 1600)$$

42b.) via `> pnorm(q,mean,sd,lower.tail=FALSE)`

A probabilidade pedida $P(Y > 33100)$ (probabilidade de a quantidade de refrigerante superar a capacidade de armazenamento das 100 garrafas) calcula-se do seguinte modo em R:

```
> mean<-330
> sd<-4
> mean42b<-mean*100
> sd42b<-sqrt(100*sd^2)
> pnorm(33100,mean42b,sd42b,lower.tail = FALSE)
[1] 0.006209665
```

Dados do Exercício 43.a)

Distribuição de Probabilidade: $X \sim N(2, \sigma^2)$ de média conhecida ($\mu = 2$ horas) e desvio padrão desconhecido ($\sigma =?$ horas).
Sabe-se que $P(X > 3) = 0.025$ ($= 2.5\%$) corresponde à probabilidade dos visitantes permanecerem mais de 3 horas na feira do livro.

Resolução 43.a)

$$
\begin{aligned}
0.025 = P(X > 3) &= P\left(\frac{X-2}{\sigma} > \frac{3-2}{\sigma}\right) \\
&= 1 - \Phi\left(\frac{1}{\sigma}\right) \\
\Longleftrightarrow &\quad \sigma = \frac{1}{\Phi^{-1}(0.975)} \quad \text{[Porquê ?]}
\end{aligned}
$$

Exercícios Resolvidos

Exercício 43., página 335

43.a) via `> qnorm(q,mean,sd)`

```
> prob43a<-0.025
> quantil43a<-qnorm(1-prob43a)
> sigma<-1/quantil43a
> sigma
[1] 0.5102135
```

43.a) via `> qnorm(q,mean,sd,lower.tail=FALSE)`

```
> prob43a<-0.025
> 1/qnorm(prob43a,lower.tail = FALSE)
[1] 0.5102135
```

Dados do Exercício 43.b)

Sabemos que o visitante já já chegou há mais de uma hora ($X > 1$ hora). Pretendemos saber qual a probabilidade de permanecer na feira mais 30 minutos ($X \leq 1.5$ horas). Para tal, precisamos de calcular a probabilidade condicional $P(X \leq 1.5|X > 1)$:

$$P(X \leq 1.5|X > 1) = \frac{P(1 < X \leq 1.5)}{P(X > 1)}$$

43.b) via `> pnorm(q,mean,sd)`

```
> F1h<-pnorm(1,2,sigma)
> F1h30m<-pnorm(1.5,2,sigma)
> prob43b<-(F1h30m-F1h)/(1-F1h)
> prob43b
[1] 0.1421
```

Ideia de resolução do Exercício 43.c)

Mediana – quantil $q_{0.50}$ determinado a partir de $P(X \leq q_{0.50})$ – coincide com o valor esperado de X, $E(X)$, dado a distribuição normal ser simétrica. Portanto $E(X) = 2$ é o valor da mediana. Para calcularmos o intervalo interquantil de X, $]q_{0.25}, q_{0.75}[$, precisamos de usar o comando **qnorm** para determinar os valores de $q_\alpha = \alpha$ para $\alpha = 0.25, 0.75$, tais que $P(X \leq q_\alpha) = \alpha$.

43.c) via `> qnorm(x,mean,sd)`

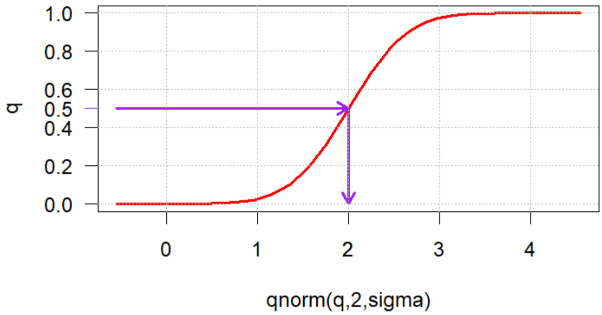
```
> qnorm(0.5,2,sigma)
[1] 2
> qnorm(0.25,2,sigma)
[1] 1.655866
> qnorm(0.75,2,sigma)
[1] 2.344134
```

Distribuição Normal $N(2, \sigma^2)$

Resolução Gráfica do 43.c)

Mediana em R

```
> qnorm(0.5,2,sigma)
[1] 2
```

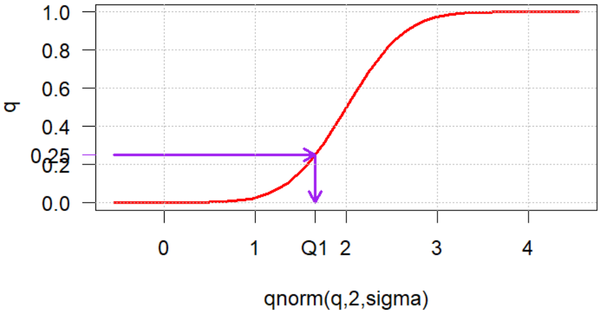


Distribuição Normal $N(2,\sigma^2)$

Resolução Gráfica do 43.c)

1^o quartil em R

```
> qnorm(0.25,2,sigma)
[1] 1.655866
```

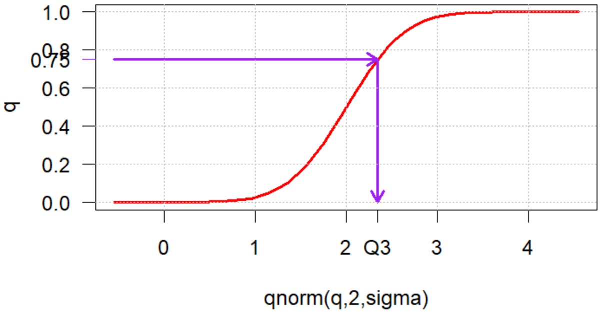


Distribuição Normal $N(2, \sigma^2)$

Resolução Gráfica do 43.c)

3^o quartil em R

```
> qnorm(0.75,2,sigma)
[1] 2.344134
```



Ideia de resolução do Exercício 43.d)

O número de visitantes seleccionados ao acaso pode ser modelado em termos de provas de Bernoulli, considerando

'Sucesso'='Permanecer mais de três horas'.

Assim, a v.a. $Y \sim B(20, P(X>3))$ dá-nos o número de visitantes que permanece mais de 3 horas ($\theta = P(X>3)$) em 20 visitantes seleccionados ao acaso ($n=20$), e $P(Y \leq 1)$ a probabilidade de não mais do que um (1) visitante **permanecer mais de 3 horas.**

43.d) em R – Valor de $P(X>3) = 0.025$ é dado em 43.

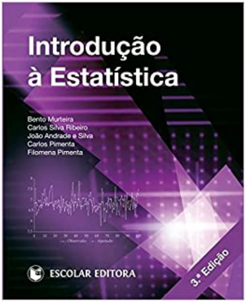
```
> pbinom(1,20,0.025)
[1] 0.9117583
> theta<-pnorm(3,2,sigma,lower.tail = FALSE)
> pbinom(1,20,theta)
[1] 0.9117583
```

Conteúdos

Conceitos de Probabilidade

Probabilidade Condicionada. Teorema de Bayes (Páginas 92 – 107)

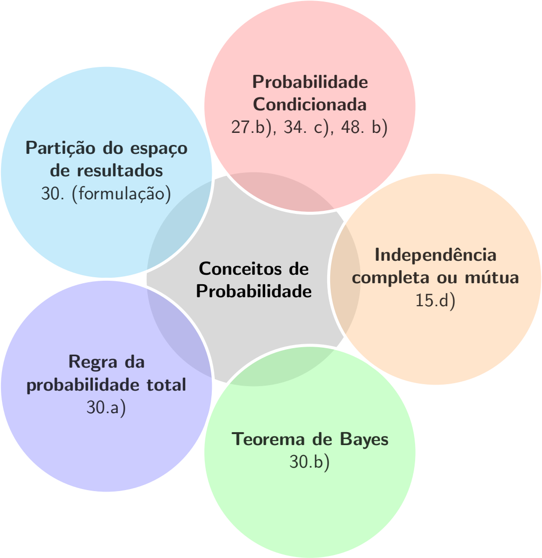
- Probabilidade condicionada: Definição 2.10;
- Partição do espaço de resultados: Definição 2.11;
- Regra da probabilidade total: Teorema 2.1;
- Teorema de Bayes: Teorema 2.2;
- Independência completa ou mútua: Definição 2.13.



Murteira, B., C. Silva Ribeiro, J. Andrade e Silva, C. Pimenta, F. Pimenta (2015), **Introdução à Estatística, 3ª Edição**', Escolar Editora.

Conceitos de Probabilidade Utilizados

Dos exercícios resolvidos até agora do Capítulo 5



Exercícios Propostos

Complemento da Aula 3

Revisão Distribuição Binomial

- **págs. 331-332:** 25., 28.

Distribuição de Poisson

- **págs. 328– 329:** 14., 16., 17., 18., 20.,21., 22., 23.

Distribuições de Poisson e Binomial

- **págs. 329:** 17., 19., 20., 21., 24., 29.

Exercícios Propostos

Complemento da Aula 4

Distribuição Uniforme Contínua

- **pág. 333:** 33., 35., 36.

Distribuição Normal

- **pág. 334:** 39.b), 39.c)., 40., 41., 44., 45., 46., 47.

Conteúdos

Lei dos Acontecimentos Raros

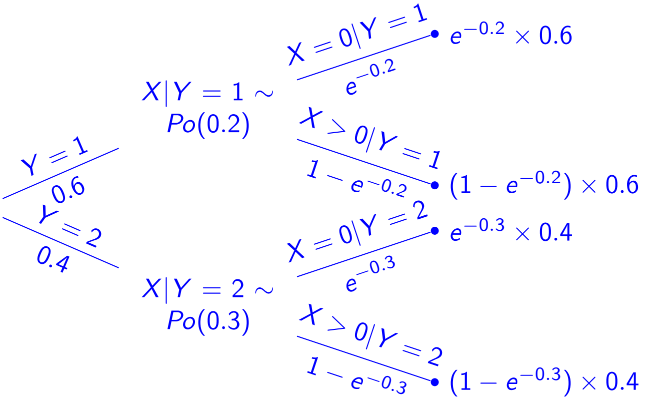
Sejam $X \sim B(510, 0.005)$, $Y \sim B(590, 0.005)$ e $Z \sim Po(2.55)$.

1. Calcule $P(X > 1)$ usando o comando **pbinom(q, size, prob)**.
2. Calcule $P(Y > 1)$ usando o comando R **pbinom(q, size, prob,lower.tail=FALSE)**.
3. Após comparar os valores $P(X > 1)$ e $P(Y > 1)$, diga, justificando, porque **lei dos acontecimentos raros não se aplica neste caso**.
4. Após calcular o valor de $P(Z > 1)$ com recurso com à instrução `> ppois(1,2.55,lower.tail = FALSE)`, verifique se $P(Z > 1) \approx P(X > 1)$.
5. Procure determinar empiricamente o **menor número natural** k para o qual o output de `> 1-pbinom(1,510*k,0.005/k)` nos dá uma aproximação de três (3) casas decimais de $P(Z > 1)$.

Desafios Computacionais

Para implementar em R

O diagrama de árvore abaixo corresponde a uma possível representação pictória do **Teorema de Bayes** com base nos dados do **Exercício 30. (p. 332 do livro)** **[resolvido anteriormente]**.



Desafios Computacionais

Para implementar em *R*

Extensão do Exercício 30.

Resolva os seguintes itens, tendo como referência as variâveis aleatórias definidas anteriormente, assim como o diagrama de árvore do slide anterior:

1. Qual a probabilidade de haver páginas da revista com gralhas?
2. Perante uma página com uma (1) ou mais gralhas, qual a probabilidade de ter sido composta pela empresa que maior número de páginas compõe?
3. Se a empresa decidir parar com a impressão assim que detectar uma (1) página com gralhas, qual a probabilidade que isso aconteça ao se imprimir a página 6.
4. Se, ao invés, a empresa decidir apenas parar com a impressão assim que detectar duas (2) páginas com gralhas, qual a probabilidade que isso aconteça ao se imprimir a página 15.

Desafios Computacionais

Para implementar em *R*

Distribuição Uniforme aplicada a tempos de espera

Suponha que os autocarros da SMTUC chegam numa determinada paragem em intervalos de 20 minutos, começando às 7:00. Isto é, **eles chegam às 7:00, 7:20, 7:40, 8:00, e assim por diante**. Se um passageiro chegar na paragem em um instante de tempo que é **uniformemente distribuído entre as 7:00 e as 8:20**, use comandos do R que lhe permitam determinar tempos de espera:

1. **inferiores a** 7 **minutos;**
2. **superiores a** 12 **minutos.**