

AULA 2

Breve revisão de conceitos relacionados com variáveis aleatórias.

Plano da aula 2

- 1 O conceito de variável aleatória – cálculo de probabilidades
- 2 Parâmetros de uma variável aleatória
- 3 Parâmetros em pares aleatórios
- 4 Principais modelos discretos
 - A distribuição uniforme discreta
 - A distribuição binomial
 - A distribuição binomial negativa
 - A distribuição de Poisson
- 5 Principais modelos contínuos
 - A distribuição normal ou de Gauss
 - O Teorema Limite Central
 - A distribuição uniforme contínua
 - As distribuições exponencial e gama
 - A distribuição beta
- 6 Exemplo de um exercício no 

O conceito de variável aleatória— cálculo de probabilidades

Quando se realiza uma experiência aleatória, pode associar-se a cada resultado da experiência um (ou mais) valores reais - diz-se que temos uma variável aleatória ou (um vector aleatório).

Uma variável aleatória costuma representar-se por X .

Um variável aleatória pode ser:

- **discreta** - por exemplo o número de sementes germinadas; o registo, a intervalos regulares, do número de pessoas em fila de espera na caixa de um supermercado;
- **contínua** - por exemplo o peso de um indivíduo; a largura, o comprimento de um folha.

Variável aleatória - cálculo de probabilidade

Associadas a cada variável aleatória existem:

- uma **função massa de probabilidade**, se X **discreta**.

A função massa de probabilidade é a **aplicação** que a cada valor $x_i \rightarrow p_i = P[X = x_i]$, verificando:

$$p_i \geq 0, \quad i = 1, \dots, k \quad \text{e} \quad \sum_{i=1}^k p_i = 1.$$

- ou uma **função densidade**, se X **contínua**.

Uma função f diz-se **função densidade de probabilidade** ou apenas **função densidade** se verificar as seguintes condições:

$$f(x) \geq 0 \quad \forall x \in \mathbb{R}; \quad \int_{-\infty}^{+\infty} f(x) dx = 1$$

Variável aleatória - cálculo de probabilidade

Associada a cada variável aleatória tem-se também:

- uma função real F , a que se chama **função distribuição cumulativa** assim definida:

$$F(x) = P[X \leq x]$$

Se X discreta tem-se $F(x) = P[X \leq x] = \sum_{x_i \leq x} P[X = x_i]$,
ou seja temos a probabilidade cumulativa associada à variável X calculada em qualquer $x \in \mathbb{R}$.

Se X contínua tem-se $F(x) = P[X \leq x] = \int_{-\infty}^x f(t) dt$
 $-\infty < x < \infty$ sendo f a função densidade.

Exemplos de cálculo de uma probabilidade com recurso à função distribuição cumulativa:

- 1 $P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a);$
- 2 $P(X = a) = F(a) - F(a^-)$ onde $F(a^-) = \lim_{x \rightarrow a^-} F(x)$
- 3 $P(a < X < b) = P(X < b) - P(X \leq a) = F(b^-) - F(a);$

Parâmetros de uma variável aleatória

Valor médio

Dada uma v.a. X chama-se **valor médio** ou **valor esperado** e representa-se por $E[X]$, μ_X ou simplesmente μ a

$$E[X] = \sum_{i=1}^n x_i p_i \quad X \text{ é v.a. discreta com distribuição } (x_i, p_i)$$

$$E[X] = \int_{-\infty}^{+\infty} x f(x) dx \quad X \text{ é v.a. contínua com f.d.p. } f(x)$$

Algumas propriedades

- $E[a + bX] = a + b E[X]$.
- $E[\varphi(X) + \psi(X)] = E[\varphi(X)] + E[\psi(X)]$
- $\inf(X) \leq E[X] \leq \sup(X)$

Parâmetros de uma variável aleatória

Variância

Chama-se **variância** de uma variável aleatória X e representa-se por $Var[X]$, σ_X^2 ou apenas σ^2 a

$$\sigma_X^2 = E[(X - \mu)^2]$$

$\sigma_X = \sqrt{Var[X]}$ chama-se **desvio padrão**.

Algumas propriedades

- $Var[X] = E[X^2] - (E[X])^2$
- $Var[X] \geq 0$
- $Var[a + b X] = b^2 Var[X]$.

Para o **desvio padrão** tem-se $\sigma_{(a+b X)} = |b| \sigma_X$

Parâmetros em pares aleatórios

Breve revisão de propriedades dos parâmetros em pares aleatórios

Se (X, Y) é um par aleatório, podendo ser **discreto** ou **contínuo**

Valor médio

Dado o par aleatório (X, Y) , e $g : \mathbb{R}^2 \rightarrow \mathbb{R}$, define-se

$$E[g(X, Y)] = \sum_i \sum_j g(x_i, y_j) p_{ij}, \quad \text{no caso discreto}$$

$$E[g(X, Y)] = \int \int_{\mathbb{R}^2} g(x, y) f(x, y) dx dy, \quad \text{no caso contínuo.}$$

Propriedades do Valor Médio

- $E[X \pm Y] = E[X] \pm E[Y]$
- **Desigualdade de Schwarz** Se $E[X^2]$ e $E[Y^2]$ existem então $(E[XY])^2 \leq E[X^2]E[Y^2]$.

Corolário: $(E[X])^2 \leq E[X^2]$

Nota: se $E[X^2]$ existe \implies existe $E[X]$.

- Se X e Y **variáveis aleatórias independentes**



$$E[XY] = E[X]E[Y]$$

Nota: O recíproco não é verdadeiro

A covariância e propriedades da variância

Covariância entre X e Y

Dado o par aleatório (X, Y) chama-se **covariância de X e Y** a

$$\mathbf{Cov}[X, Y] \equiv \sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)]$$

Exercício

Verifique que $\mathbf{Cov}[X, Y] = E[XY] - E[X]E[Y]$

Propriedades da variância e da covariância

- Sejam X e Y variáveis aleatórias.

$$\text{Var}[X \pm Y] = \text{Var}[X] + \text{Var}[Y] \pm 2\text{Cov}[X, Y]$$

- Se X e Y são v. a.'s **independentes** $\implies \text{Cov}[X, Y] = 0$.

Nota: O recíproco não é verdadeiro.

- Se X e Y são variáveis aleatórias **independentes**

$$\text{Var}[X \pm Y] = \text{Var}[X] + \text{Var}[Y]$$

- $\text{Cov}[a + bX, c + dY] = bd \text{Cov}[X, Y]$.

- $|\text{Cov}[X, Y]| \leq \sigma_X \sigma_Y$.

- Coefficiente de correlação** define-se como:

$$\rho \equiv \rho_{X,Y} = \frac{\text{Cov}[X, Y]}{\sigma_X \sigma_Y} \quad (\sigma_X > 0; \sigma_Y > 0).$$

Os principais modelos de probabilidade discretos e contínuos.

Principais modelos discretos

A distribuição uniforme discreta

Definição Uma v.a. X diz-se ter **distribuição uniforme discreta** se toma os valores x_1, \dots, x_n com probabilidades $1/n, \dots, 1/n$, i.e.
 $P(X = x_i) = 1/n, \quad i = 1, \dots, n.$

Caso particular

$$X = \begin{cases} 1 & 2 & \dots & n \\ 1/n & 1/n & \dots & 1/n \end{cases}$$

Valor médio e variância

$$E[X] = \frac{n+1}{2} \qquad \text{Var}[X] = \frac{n^2-1}{12}$$

Instrução no  para simular `> sample(v, size, rep=TRUE)`

v vector com os valores que a variável pode tomar

A função `sample()`

A função `sample` - permite criar uma amostra aleatória a partir dos elementos de um vector, **com ou sem reposição**, com probabilidades iguais ou não.

```
>sample(1:20, 15)
```

Selecciona aleatoriamente 15 números de 1 a 20 **sem reposição**
a omissão é “sem reposição”.

Para seleccionar **com reposição** com diferentes probabilidades fazer, por exemplo:

```
>pb<-c(rep(0.1,3), .2, .3, .2);pb  
>sample(1:6,30,rep=T,prob=pb)
```

Se a probabilidade for a mesma pode omitir-se.

Nota: para gerar sempre a mesma sucessão fazer `>set.seed()`

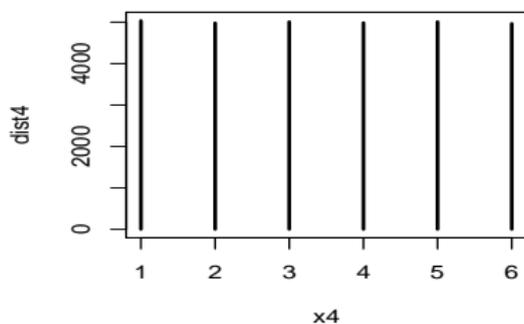
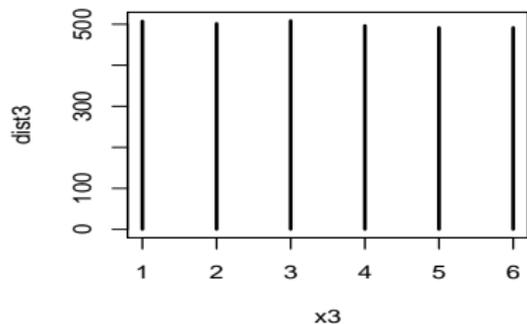
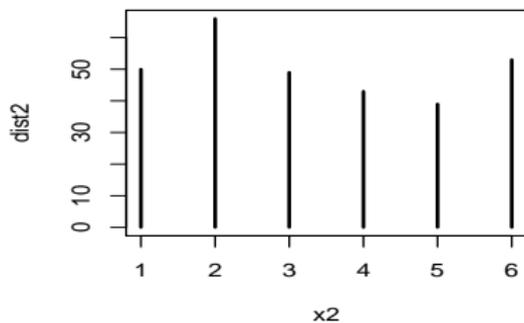
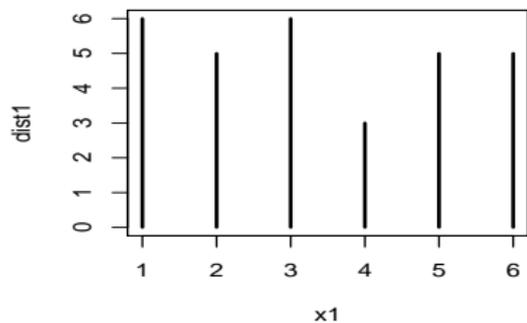
```
> par(mfrow=c(2,2))
> x1<-sample(1:6,30,rep=T);x1
> dist1<-table(x1);dist1
> plot(dist1)
```

Repetir para 300, 3000 e 30000. (ver variáveis x_2 , x_3 e x_4 nos gráficos do *slide* seguinte.

Nota: Podia definir-se uma função, por exemplo:

```
> dado<-function(n) sample(1:6,n,replace=T)
> d1<-dado(30);table(d1)
> table(dado(30)) # Haverá alguma diferença?
> dado(300);dado(3000)
```

Gráficos de vários lançamentos



A distribuição binomial

Quando se realizam n provas de Bernoulli **independentes**, a variável que conta o número de sucessos que ocorrem diz-se ter **distribuição binomial** e representa-se por $X \sim \text{Binom}(n, p)$, sendo p a probabilidade de sucesso. A probabilidade de insucesso, $1 - p$, é costume representar por q .

X toma os valores $x = 0, 1, 2, \dots, n$ com probabilidades

$$P[X = x] = \binom{n}{x} p^x (1 - p)^{n-x}$$

Valor médio e variância

$$E[X] = np \qquad \text{Var}[X] = np(1 - p) = npq$$

Para determinar o valor daquelas probabilidades, quantis ou a função distribuição cumulativa o  possui **funções já definidas** para os modelos implementados.

Funções no R para os modelos existentes

- **d**função (x, \dots) - permite obter a função massa de probabilidade (modelo discreto) ou a função densidade (modelo contínuo) em x ;
- **p**função(q, \dots) - permite obter a função de distribuição cumulativa, i.e., devolve a probabilidade de a variável ser menor ou igual a q ;
- **q**função (p, \dots) - permite calcular o quantil associado à probabilidade p ;
- **r**função (n, \dots) - permite gerar uma amostra de n números pseudo-aleatórios do modelo especificado.

Significado:

density, **p**robability, **q**uantile, **r**andom

Exercícios

Exercício Vamos experimentar a utilização das funções `d`, `p`, `q`, `r`. Considere-se uma *Binomial* ($n = 10, p = 0.2$).

```
> x<- 0:10
> dbinom(x,size=10,prob=0.2)
> pbinom(3,size=10,prob=0.2,lower.tail = TRUE) #dá P[X<=3]
> qbinom(0.75, size=10, prob=0.2, lower.tail = TRUE)
+       #dá o quantil de probabilidade 0.75
> rbinom(5, size=10, prob=0.2)
> pbinom(3, size=10, prob=0.2, lower.tail = F) #dá P[X>3]
```

O **quantil** é definido como o **menor valor** χ_p tal que $F(\chi_p) \geq p$, sendo F a função distribuição cumulativa.

```
> par(mfrow=c(1,2))
> plot(x,dbinom(x,size=10,prob=0.2),type="h")
> plot(x,dbinom(x,size=10,prob=0.4),type="h")
```

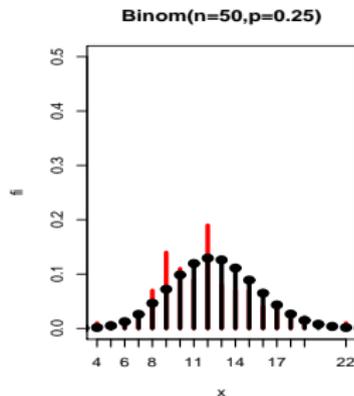
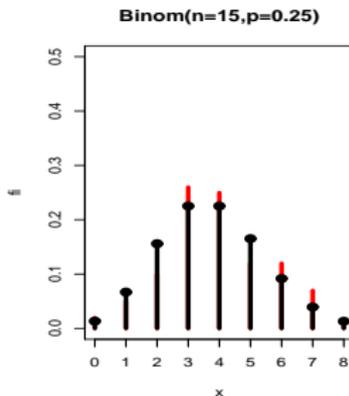
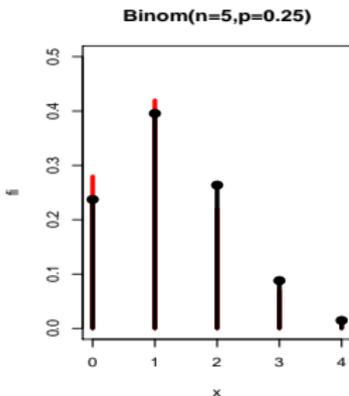
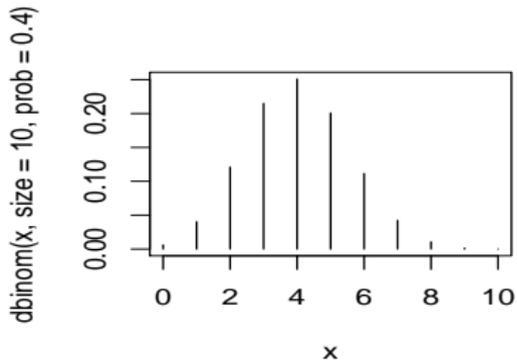
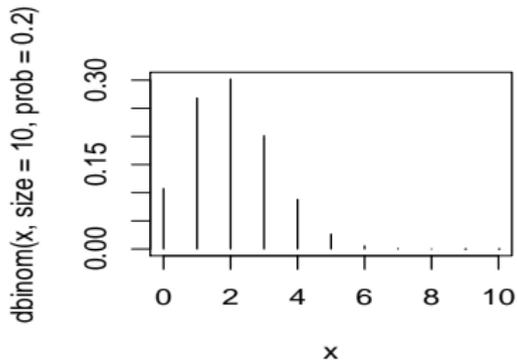
Exercícios (continuação)

Para exemplificar a **distribuição binomial teórica** e a **simulada** (geração de números pseudo-aleatórios)

```
> par(mfrow=c(1,3))
> n<-5;p<-0.25
> x<-rbinom(100,n,p) # 100 random numbers
> ni<-table(x);ni
> fi<-ni/sum(ni);fi
> dbinom(0:n,size=5,prob=0.25)
> plot(fi,type = "h", col = "red",lwd=3,
+      main="Binom(n=5,p=0.25)",ylim=c(0,.5))
> xvals<-0:n;points(xvals,dbinom(xvals,n,p),type="h",lwd=3)
> points(xvals,dbinom(xvals,n,p),type="p",lwd=3)
```

... Repetir com $n=15$, $n=50$.

Exemplos (continuação)



Mais modelos de probabilidade

No \mathbb{R} o modelo Binomial Negativo surge associado a contagens do número de insucessos que se observam até obter k "sucessos", num contexto de provas de Bernoulli independentes.

A variável X , nº de contagens nas condições acima referidas diz-se ter **distribuição Binomial Negativa** e é costume representar-se por $X \sim BN(k, p)$

p é a probabilidade constante de "sucesso" de prova para prova
 k é o número de "sucessos" que se pretende obter.

A distribuição binomial negativa

Caracterização da v.a. $X \sim BN(k, p)$:

Valores $x = 0, 1, 2, \dots$

Probabilidades $P[X = x] = \binom{x+k-1}{x} p^k q^x$

$0 < p < 1, \quad q = 1 - p$

Valor médio e variância de $X \sim BN(k, p)$

$$E[X] = \frac{kq}{p}$$

$$\text{Var}[X] = \frac{kq}{p^2}$$

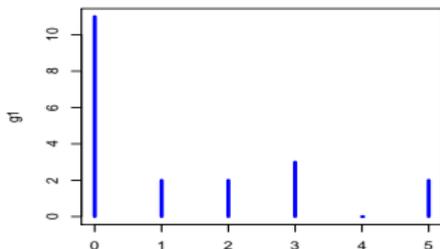
Exemplo em 

```
> x <- 0:15 #vector de valores da variável
> dnbinom(x,size=6, prob= 0.4) #probabilidade de se
+ # verificarem 0 a 15 insucessos até haver 6 sucessos
+ #outra parametrização que usa o valor médio indicado acima
> dnbinom(x, mu = 9, size = 6)
```

A distribuição Geométrica

Se $k = 1$, isto é, se pretendemos determinar o número de insucessos (falhas) até obter o 1º sucesso, a variável X diz-se ter **distribuição Geométrica**, $X \sim \text{Geo}(p)$

```
> Ni <- rgeom(20, prob = 1/4)
> g1 <- table(factor(Ni, 0:max(Ni)))
> plot(g1)
```



A distribuição de Poisson

Definição

A v.a X que conta o número de sucessos que ocorrem num dado intervalo de tempo ou domínio (independentemente do número que ocorre em qualquer outro intervalo ou domínio disjunto) diz-se ter **distribuição de Poisson**.

Depende apenas do parâmetro $\lambda \rightarrow$ **número médio de sucessos** que ocorrem no intervalo de tempo (ou na região especificada).

Representa-se por $X \sim P(\lambda)$ e a lei de probabilidade é:

$$P[X = x] = \frac{e^{-\lambda} \lambda^x}{x!}, \quad x = 0, 1, 2, \dots$$

A distribuição de Poisson

Valor médio e variância

$$E[X] = \lambda \quad \text{Var}[X] = \lambda.$$

Uso do

```
> diff(ppois(c(47, 50), lambda = 50)) # P[47 < X <=50]
> ppois(50,50)-ppois(47,50) # verificar que é o mesmo
```

A distribuição normal ou de Gauss

Tem um papel fulcral nas Probabilidades e Estatística, porque:

- muitas variáveis biométricas têm uma distribuição muito próxima da normal;
- por vezes uma variável que não é normal pode ser transformada de um modo simples numa outra com distribuição normal;
- a parte central de muitos modelos não normais é por vezes razoavelmente bem aproximada por uma distribuição normal.

A distribuição normal ou de Gauss

Uma v.a. contínua X diz-se ter **distribuição normal** com parâmetros μ e σ e representa-se por $X \sim \mathcal{N}(\mu, \sigma)$ se a sua f.d.p. é da forma:

$$f(x) = \frac{1}{\sqrt{2\pi} \sigma} \exp \left[-\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2 \right]$$

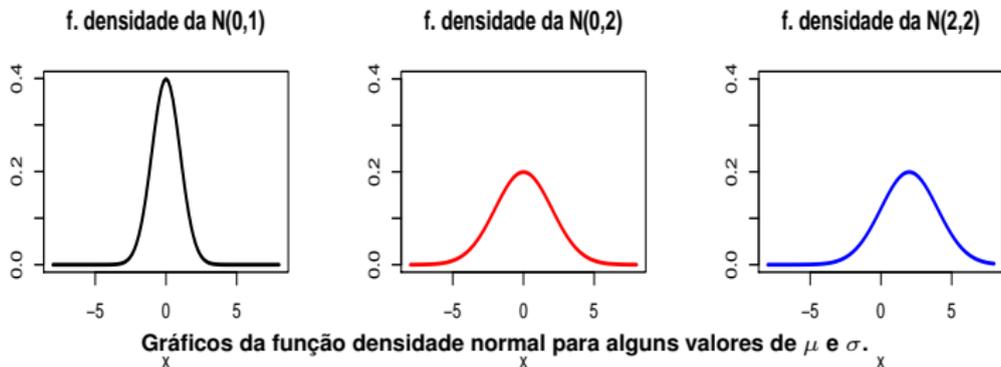
$$-\infty < x < +\infty, \quad -\infty < \mu < +\infty, \quad 0 < \sigma < +\infty$$

A distribuição normal ou de Gauss

Propriedades da curva densidade da distribuição normal

1. É simétrica relativamente a μ .
2. É uma curva unimodal, a moda é μ .
3. Tem pontos de inflexão em $\mu + \sigma$ e $\mu - \sigma$.

Se $\mu = 0$ e $\sigma = 1$ a variável aleatória com distribuição $\mathcal{N}(0, 1)$ chama-se **normal reduzida** e costuma representar-se por Z , $Z \sim \mathcal{N}(0, 1)$

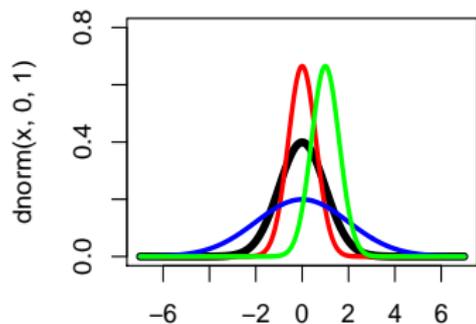


Exercícios com a distribuição normal no R

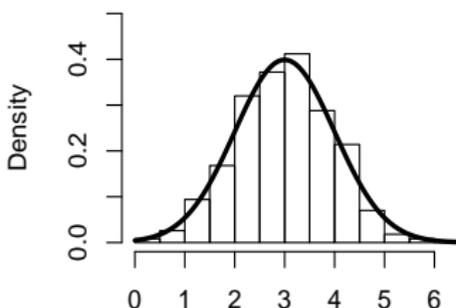
```
#cálculos e gráficos com a normal
> pnorm(1.96)
> pnorm(-1.96)
> pnorm(3,mean=5,sd=2)
> qnorm(0.75,mean=5,sd=1)
> qnorm(0.75,mean=5,sd=1,lower.tail=T)
> qnorm(0.25,mean=5,sd=1,lower.tail=F)
+           #graficos
> par(mfrow=c(1,2))
> x<-seq(-7,7,.01)
> plot(x,dnorm(x,0,1),type="l",ylim=c(0,.8),lwd=5)
> lines(x,dnorm(x,0,.6),col="red",lwd=3)
> lines(x,dnorm(x,0,2),col="blue",lwd=3)
> lines(x,dnorm(x,1,.6),col="blue",lwd=3)
```

A distribuição normal (gráficos)

```
# gerar valores (continuação do exercício)
> y<-rnorm(1000,mean=3,sd=1)
> hist(y,freq=F,ylim=c(0,0.5),
+ main="valores gerados+curva",col=gray(.9))
> curve(dnorm(x,mean=3,sd=1),add=T,lwd=3)
```



valores gerados+curva



Resultados importantes com a distribuição normal

- Seja $X \sim \mathcal{N}(\mu, \sigma)$ então a v.a. $\frac{X - \mu}{\sigma}$ tem distribuição normal reduzida, i.e., $Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$.
- Sejam X_i n v.a. normais independentes e semelhantes, i.e., tendo todas o mesmo valor médio μ e a mesma variância σ^2 .

As variáveis aleatórias **soma** e **média**, definidas respectivamente como $S_n = \sum_{i=1}^n X_i$ e $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$

têm distribuição normal assim definida

$$S_n \sim \mathcal{N}(n\mu, \sigma\sqrt{n}) \quad \text{e} \quad \bar{X}_n \sim \mathcal{N}(\mu, \sigma/\sqrt{n}).$$

O Teorema Limite Central

Vimos que a soma de normais independentes é ainda uma normal. Mas a **distribuição aproximada da soma de n variáveis aleatórias independentes e identicamente distribuídas**, sob certas condições é também **normal**.

O Teorema limite central

Sejam X_1, \dots, X_n variáveis aleatórias independentes e identicamente distribuídas, com valor médio μ e variância σ^2 (finita). Se n 'grande' a v.a. $S_n = \sum_{i=1}^n X_i$, verifica:

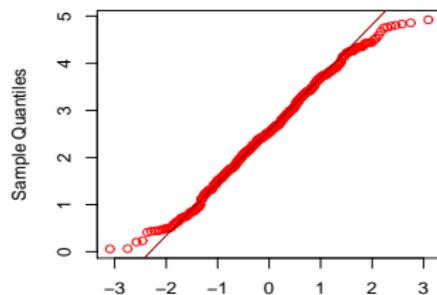
$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \sim \mathcal{N}(0, 1) \quad \text{e também se tem} \quad \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1).$$

O Teorema Limite Central...exercício

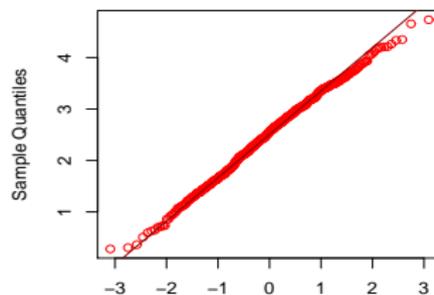
```
> # População Uniforme(0,5)
> par(mfrow=c(2,2))
> am<-500
> vec.med<-c(rep(0,am))
> n<-c(2,3,10,30)
> for(j in 1:4)
+ {for(i in 1:am)
+ {x<-runif(n[j],0,5)
+ vec.med[i]<-mean(x)}
+ qqnorm(vec.med,main=paste("Q-QPlot Normal, n =",n[j],
+ "n","Médias Pop. U(0,5),"),xlab=" ",
+ col="red")
+ qqline(vec.med,col="darkred")}
```

O Teorema Limite Central...exercício

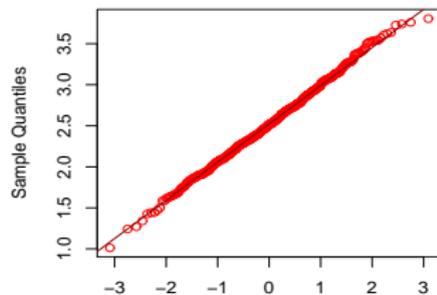
Q-QPlot Normal, $n = 2$
Médias Pop. $U(0,5)$,



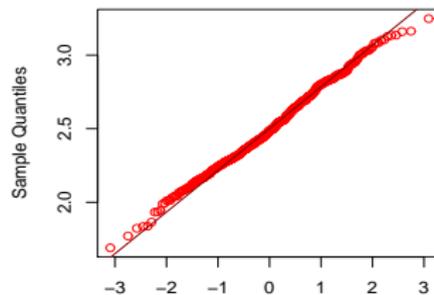
Q-QPlot Normal, $n = 3$
Médias Pop. $U(0,5)$,



Q-QPlot Normal, $n = 10$
Médias Pop. $U(0,5)$,



Q-QPlot Normal, $n = 30$
Médias Pop. $U(0,5)$,



Aplicações do Teorema Limite Central

Seja X uma v.a. com distribuição binomial com valor médio $\mu = np$ e variância $\sigma^2 = npq$.

$X \sim \mathcal{B}(n, p)$, portanto $\mu = np$ e $\sigma^2 = npq$

$$\frac{X - np}{\sqrt{npq}} \sim \mathcal{N}(0, 1) \quad \text{se} \quad n \rightarrow \infty$$

Regra prática Se na distribuição binomial, $np > 5$ e $nq > 5 \implies$ a aproximação pela distribuição normal é boa.

$$X \sim P(\lambda)$$

Se $\lambda \rightarrow \infty$ então $\frac{X - \lambda}{\sqrt{\lambda}} \sim \mathcal{N}(0, 1)$.

Mais uma convergência

Se na distribuição binomial $n \rightarrow \infty$ e p pequeno (digamos $p < 0.05$ e $n > 20$) $X \sim B(n, p) \sim P(np)$

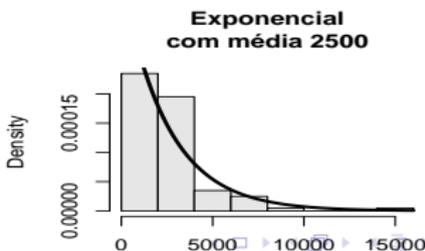
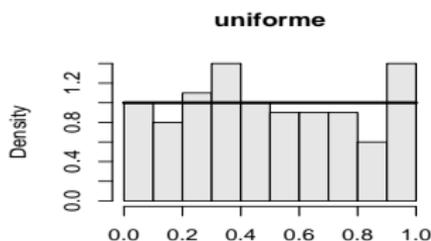
Outras distribuições contínuas

Distribuição uniforme contínua (já referida atrás)

```
> u<-runif(100)
> hist(u,freq=F,col=gray(.9),main="uniforme")
> curve(dunif(x),add=T,lwd=3)
```

... distribuição exponencial de valor médio 2500

```
> x<-rexp(100,1/2500)
> hist(x,probability=TRUE,col=gray(.9),main="Exponencial
+ com média 2500")
> curve(dexp(x,1/2500),add=T)
```



A distribuição gama

Em muitas áreas há ainda muitas situações em que a lei de Gauss não serve para modelar o fenómeno.

Começemos por referir brevemente a **distribuição gama** que deve o seu nome à **função gama**, estudada em muitas áreas da matemática, assim definida:

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx \quad \text{para } \alpha > 0$$

Algumas propriedades da função gama:

- $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$ (fórmula de recorrência)
- No caso de $\alpha = n$ inteiro, é fácil verificar que se tem

$$\Gamma(n) = (n - 1)(n - 2)\dots\Gamma(1) = (n - 1)!$$

A distribuição gama

Mais algumas propriedades da função gama:

- $\Gamma(1/2) = \sqrt{\pi}$
- As derivadas da função gama são assim definidas:

$$\Gamma^{(k)}(\alpha) = \int_0^{\infty} x^{\alpha-1} (\log x)^k e^{-x} dx$$

Alguns valores particulares das derivadas úteis em muitas aplicações são

$\Gamma'(1) = \gamma = 0.57722\dots$ valor é designado por **constante de Euler**

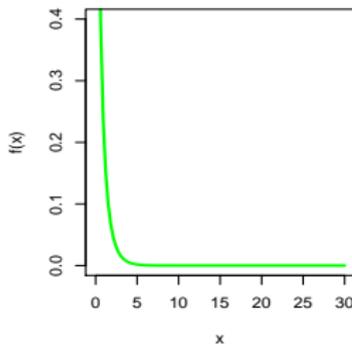
$\Gamma''(1) = \gamma^2 + \pi^2/6 = 1.97811\dots$

A distribuição gama

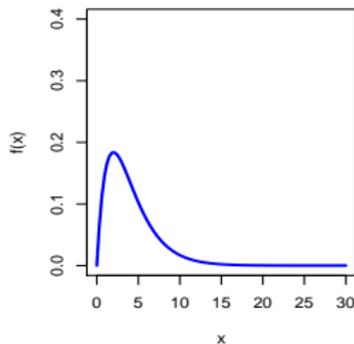
Uma v.a. diz-se ter **distribuição gama** de parâmetros α e β , ($\alpha > 0$, $\beta > 0$) e escreve-se $X \sim G(\alpha, \beta)$ se (α – parâmetro de forma e β – parâmetro de escala) função densidade é:

$$f(x) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

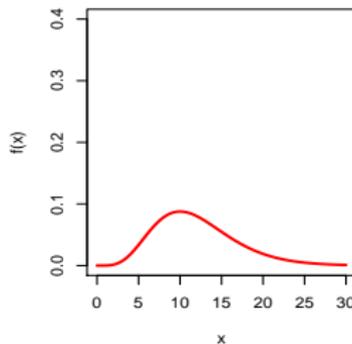
f. dens. da Gamma(0.5,1)



f. dens. da Gamma(2,2)



f. dens. da Gamma(6,2)



Gráficos da função densidade de uma v.a. com distribuição $G(1/2, 1)$, $G(2, 2)$ e $G(6, 2)$, da esquerda para a direita, respectivamente.

A distribuição gama

Valor médio e variância de $X \sim G(\alpha, \beta)$

$$E[X] = \alpha \beta \quad \text{Var}[X] = \alpha \beta^2$$

Um caso particular muito importante é o que se obtém fazendo $\alpha = 1$. A v.a. resultante diz-se ter **distribuição exponencial**, foi falada atrás, representa-se por $X \sim \text{Exp}(\beta)$ e a função densidade é assim definida,

$$f(x) = \begin{cases} \frac{1}{\beta} e^{-x/\beta} & x > 0 \quad \beta > 0 \\ 0 & x \leq 0 \end{cases}$$

A distribuição exponencial

Valor médio e variância

$$E[X] = \beta$$

$$\text{Var}[X] = \beta^2$$

A distribuição exponencial tem sido **largamente usada como um modelo de problemas relativos a duração de vida, teoria da fiabilidade, tempos de espera**, etc.

Propriedade

Se $X_i, i = 1, \dots, n$ são variáveis aleatórias independentes e semelhantes, com distribuição $\text{Exp}(\beta)$, então

$$\sum_{i=1}^n X_i \sim G(n, \beta).$$

A distribuição exponencial

Observações:

- Existe uma relação muito importante entre a distribuição exponencial e a distribuição de Poisson, que surge muitas vezes na prática. Estando a observar a **ocorrência de certos acontecimentos em intervalos de tempo**, pretendemos caracterizar T o tempo ao fim do qual se verifica a primeira ocorrência.

Teorema

Seja X uma v.a. de Poisson de parâmetro λ . Seja T a v.a. que designa o tempo de espera pela ocorrência do primeiro acontecimento, então T tem distribuição exponencial, $T \sim \text{Exp}[\beta]$, de parâmetro $\beta = 1/\lambda$.

A distribuição beta

Uma v.a. X contínua diz-se ter **distribuição beta** de parâmetros (m, n) e escreve-se $X \sim Be(m, n)$ se a sua função densidade é da forma

$$f(x) = \begin{cases} \frac{1}{B(m,n)} x^{m-1} (1-x)^{n-1} & 0 < x < 1 \quad m > 0, n > 0 \\ 0 & \text{outros valores de } x \end{cases}$$

onde $B(m, n)$ é a **função beta** assim definida

$$B(m, n) = \frac{\Gamma(m)\Gamma(n)}{\Gamma(m+n)} = \int_0^1 x^{m-1} (1-x)^{n-1} dx$$

A distribuição beta

Propriedades

1. $B(m, n) = B(n, m)$

2. $B(1, 1) = 1$

3. $B\left(\frac{1}{2}, \frac{1}{2}\right) = \pi$

4. $B(m, n) = \int_0^{+\infty} \frac{x^{m-1}}{(1+x)^{m+n}} dx$

Valor médio e variância da distribuição beta

$$E[X] = \frac{m}{m+n}$$

$$\text{var}[X] = \frac{mn}{(m+n)^2(m+n+1)}$$

A distribuição beta

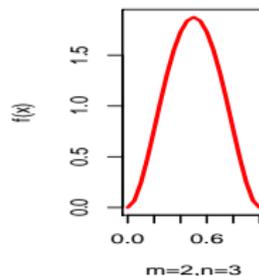
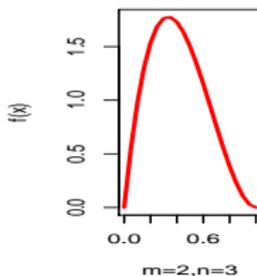
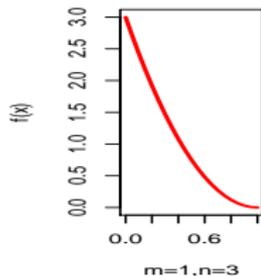
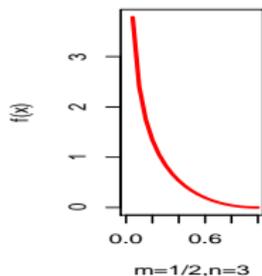
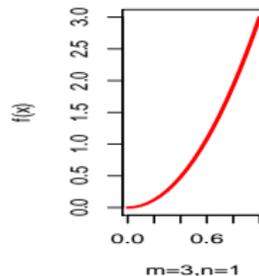
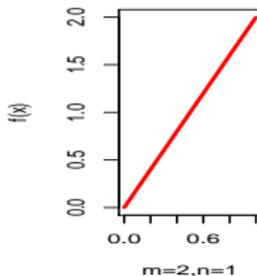
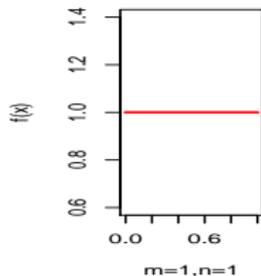
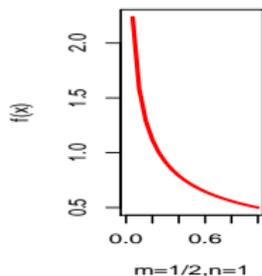
A função densidade da v.a. com distribuição beta apresenta, como dissemos, uma grande variabilidade de formas.

Assim podemos caracterizar o aspecto da densidade em função do valor dos parâmetros.

- se $m > 1, n > 1 \Rightarrow$ existe uma única moda em $x = (m - 1)/(m + n - 2)$
- se $m < 1, n < 1 \Rightarrow$ existe uma antimoda (forma de U)
- se $(m - 1)(n - 1) \leq 0 \Rightarrow$ forma de J
- se $m = n \Rightarrow$ simetria relativamente a 0.5.

A distribuição beta

Nas figuras seguintes, vejam-se alguns desses aspectos:



A distribuição beta

Sejam X e Y variáveis aleatórias independentes tais que $X \sim G(a_1, b_1)$ e $Y \sim G(a_2, b_2)$, então

$$X|(X + Y) \sim Be(a_1, a_2).$$

A distribuição beta, acabada de estudar diz-se estar na forma estandardizada e é de facto a forma mais usada. A sua forma mais geral apresenta quatro parâmetros (a, b, m, n) e a função densidade é

$$f(x) = \begin{cases} \frac{1}{B(m,n)} \frac{(x-a)^{m-1}(b-x)^{n-1}}{(b-a)^{m+n-1}} & a < x < b \quad m > 0, n > 0 \\ 0 & \text{outros valores de } x \end{cases}$$

RESUMO de algumas distribuições no R

| Nome da distribuição no  | Função | Argumentos |
|---|---------|-----------------|
| Beta | beta | shape1, shape2 |
| Binomial | binom | size, prob |
| Cauchy | cauchy | location, scale |
| Chisquare | chisq | df |
| Exponential | exp | rate |
| FDist | f | df1, df2 |
| GammaDist | gamma | shape, scale |
| Geometric | geom | prob |
| Hypergeometric | hyper | m, n, k |
| Lognormal | lnorm | meanlog, sdlog |
| Logistic | logis | location, scale |
| NegBinomial | nbinom | size, prob |
| Normal | norm | mean, sd |
| Poisson | pois | lambda |
| TDist | t | df |
| Uniform | unif | min,max |
| Weibull | weibull | shape, scale |

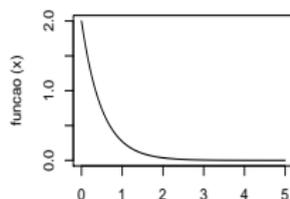
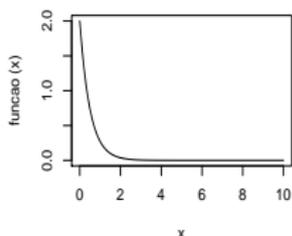
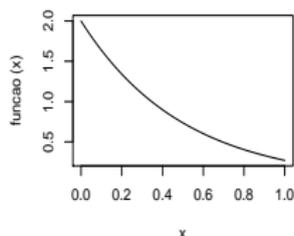
Exemplo de um exercício no R

Definição e gráfico de uma função. Cálculo de integrais.

Considera-se a seguinte função $f(x) = \begin{cases} 2 e^{-2x} & x > 0 \\ 0 & x \leq 0 \end{cases}$

Vamos ver que f é uma função densidade e calcular $P[X > 1]$ e $P[0.2 < X < 0.8]$

```
>funcao<-function(x) {  
+   fx<-ifelse(x<0,0,2*exp(-2*x))  
+   return(fx)}  
>par(mfrow=c(1,3))  
>plot(funcao);plot(funcao,0,10);plot(funcao,0,5)
```



Exemplo de um exercício no R

```
>integrate(funcao,0,Inf)
>integrate(funcao,1,Inf)
>res<-integrate(funcao,0,1);res;str(res)
>1-res$value
```

```
1 with absolute error < 5e-07
```

```
0.1353353 with absolute error < 2.1e-05
```

```
0.8646647 with absolute error < 9.6e-15
```

```
List of 5
```

```
$ value      : num 0.865
```

```
$ abs.error  : num 9.6e-15
```

```
$ subdivisions: int 1
```

```
$ message    : chr "OK"
```

```
$ call       : language integrate(f =funcao,lower = 0,upper = 1)
  attr(*, "class")= chr "integrate"
```

```
[1] 0.1353353
```