4 Famílias de distribuições

4.1 Distribuição uniforme discreta

Definição

Seja X uma variável aleatória que toma valores em $D = \{x_1, \ldots, x_n\}$. Se esses valores forem equiprováveis então diz-se que a variável tem uma **distribuição uniforme discreta** no conjunto D.

$$X \sim U(\{x_1, \ldots, x_n\}) \iff f_X(x) = \begin{cases} 1/n, & x \in D \\ 0, & x \notin D \end{cases}$$

4.2 Distribuição binomial

Definição

Uma experiência aleatória com apenas dois resultados diz-se um ensaio ou **prova de Bernoulli**.

Seja

$$X = \begin{cases} 1, & \text{se ocorreu um sucesso} \\ 0, & \text{se não ocorreu um sucesso} \end{cases}$$

A distribuição de X fica definida se conhecermos a probabilidade de "sucesso", $0 < p < 1$. Então,

$$f_X(x) = \begin{cases} p, & x = 1 \\ 1-p, & x = 0 \\ 0, & \text{c. c.} \end{cases} = \begin{cases} p^x(1-p)^{1-x}, & x \in \{0,1\} \\ 0, & \text{c. c.} \end{cases}$$

Definição

Nas condições anteriores diz-se que a variável aleatória X tem uma **distribuição de Bernoulli** ou $X \sim Ber(p)$, com $0 < p < 1$.

- $E[X] = E[X^2] = p$
- $Var[X] = p(1-p)$

Uma prova de Bernoulli isolada é um caso pouco interessante. No entanto, muitas situações de interesse prático podem ser descritas como sequências de provas desse tipo.

Consideremos uma experiência aleatória que consiste numa sequência de realizações **independentes** de uma prova de Bernoulli com probabilidade de sucesso p.

Seja $X =$ "número de sucessos em n realizações independentes da prova".

Qual a distribuição de X?

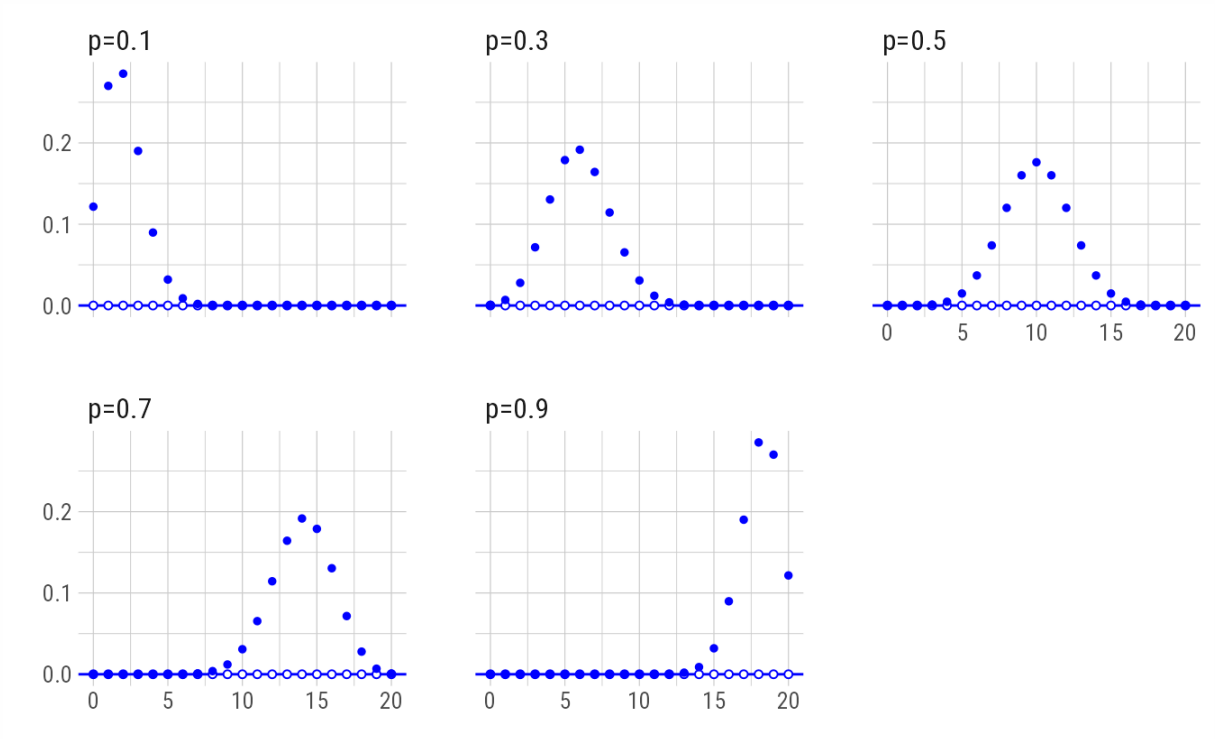
Temos então que

$$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x},\ x = 0, 1, \ldots, n$$

Definição

Nas condições anteriores diz-se que a variável aleatória X tem uma **distribuição binomial** ou $X \sim Bi(n, p)$, com $n \in \mathbb{N}$ e $0 < p < 1$.

Funções de probabilidade binomiais – Bi(20,p)



Notas

1. $Bi(1,p) \equiv Ber(p)$

2. $X = \sum_{i=1}^{n} X_i$ onde $X_i \sim Ber(p)$ indica o resultado da i-ésima realização

3. $E[X] = np$ e $Var[X] = np(1-p)$

4. Se $X \sim Bi(n,p)$ representar o número de sucessos numa experiência aleatória do tipo referido, qual a distribuição do número de insucessos na mesma experiência, Y ? E qual a relação entre as variáveis aleatórias X e Y?

Exemplo – Tiro ao alvo

Um teste de escolha múltipla é formado por 10 questões com 4 alíneas das quais apenas uma está certa. Considere que alguém responde a todas as questões ao acaso:

1. Qual a probabilidade de responder acertadamente a pelo menos metade das questões?

2. Qual é o número mais provável de respostas certas?

4.3 Distribuição geométrica

Consideremos de novo a última experiência aleatória considerada: uma sequência de realizações **independentes** de uma prova de Bernoulli com probabilidade de sucesso p.

Seja $X =$ "número de realizações da prova até ao primeiro sucesso".

Qual a distribuição de X?

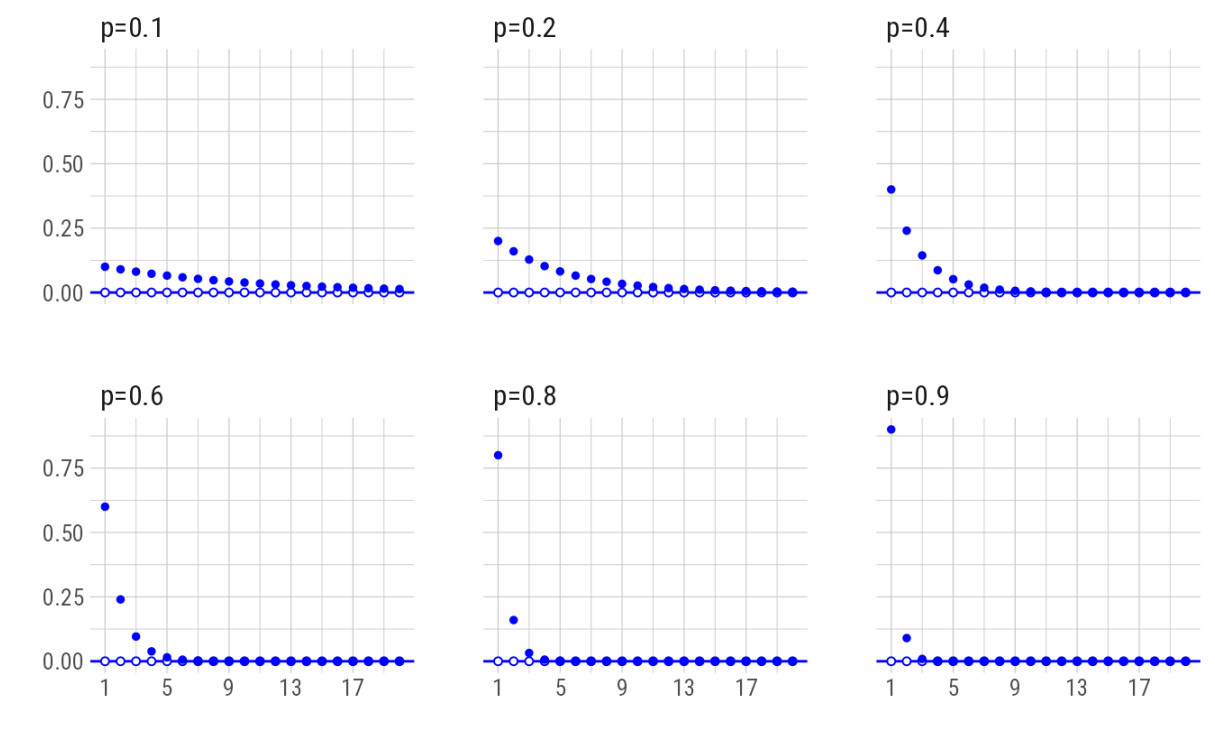
Temos então que

$$f_X(x) = (1-p)^{x-1}p,\ x = 1, 2, \ldots$$

Definição

Nas condições anteriores diz-se que a variável aleatória X tem uma **distribuição geométrica** ou $X \sim Geo(p)$, com $0 < p < 1$.

Funções de probabilidade geométricas



$$f_X(x) = (1-p)^{x-1}p,\ x = 1, 2, \ldots$$

Notas

1. $f_X(x)$ é sempre decrescente

2. $E[X] = \frac{1}{p}$ e $Var[X] = \frac{1-p}{p^2}$

3. $F_X(x) = \begin{cases} 0, & x < 1 \\ 1-(1-p)^k, & k \leq x < k+1,\ k \in \mathbb{N} \end{cases}$

Amnésia da distribuição geométrica

$X \sim Geo(p) \Rightarrow P(X > i+j \mid X > i) = P(X > j),$

$\forall i, j = 1, 2, \ldots$

4.4 Distribuição de Poisson

Suponhamos que se está interessado em contar as ocorrências de um dado fenómeno ao longo do tempo.

Seja $N(t) =$"número de ocorrências em $[0, t]$", para $t > 0$, e admitamos que:

1. $N(0) = 0$,

2. $E[N(t)] = \lambda t$, para algum $\lambda \in \mathbb{R}^+$.

Admitamos também que $\exists n \in \mathbb{N}$ tal que, sendo

X_i="número de ocorrências em $\left]\frac{(i-1)}{n}t, \frac{i}{n}t\right]$", $i = 1, \ldots, n$,

se tem:

1. $X_i \sim Ber\,(p)$;

2. X_i independente de X_j, $\forall i \neq j$.

Note-se que $N(t) = \sum_{i=1}^{n} X_i$ e, consequentemente.

$$N(t) \sim Bi\,(n, p)$$

Então, para que $E[N(t)] = \lambda t$ deve-se ter:

- $p = \frac{\lambda t}{n}$ e

- $n > \lambda t$

Esta versão com n finito não acrescentaria nada, mas

$$\lim_{n \to +\infty} f_{N(t)}(x) = e^{-\lambda t} \frac{(\lambda t)^x}{x!}, \ x \in \mathbb{N}_0.$$

Notas

1. $e^{-\lambda t} \frac{(\lambda t)^x}{x!} > 0, \ \forall x \in \mathbb{N}_0, \forall \lambda \in \mathbb{R}^+, \forall t \in \mathbb{R}^+;$

2. $\sum\limits_{x=0}^{+\infty} e^{-\lambda t} \frac{(\lambda t)^x}{x!} = 1, \forall \lambda \in \mathbb{R}^+, \forall t \in \mathbb{R}^+;$

3. O processo de contagem atrás esboçado chama-se um **processo de Poisson**.

Definição

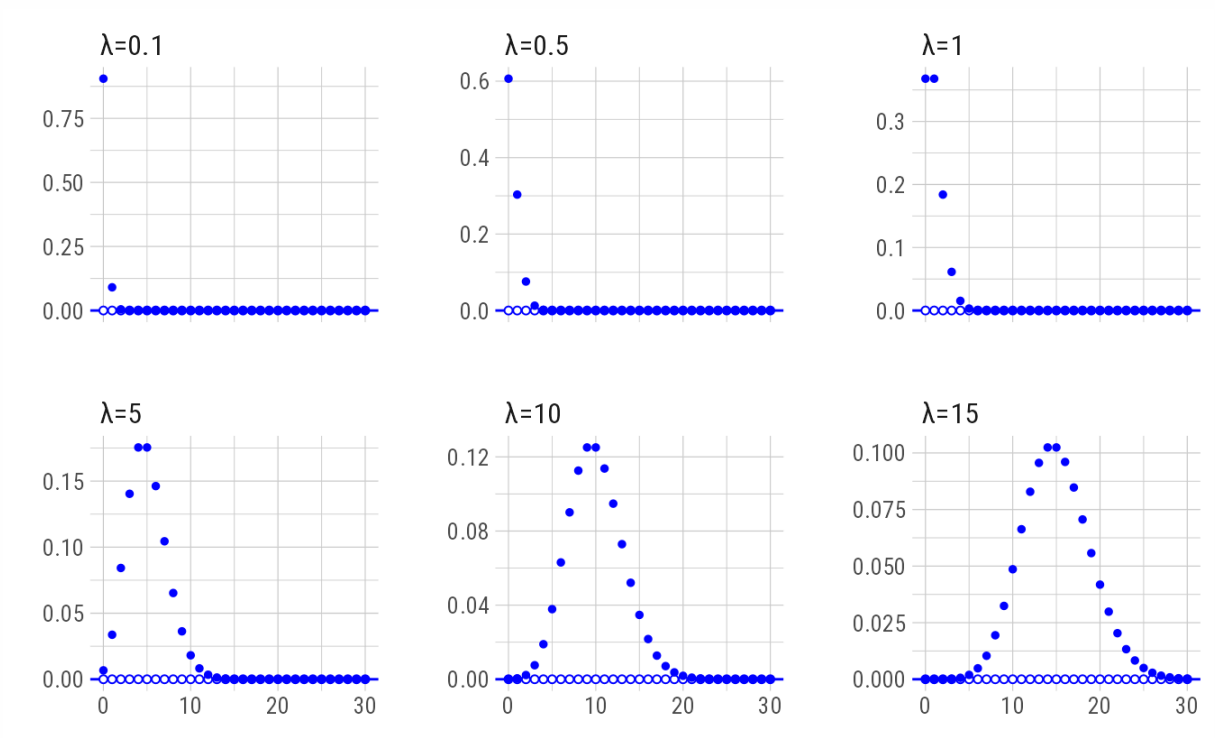
Seja X a variável aleatória que representa o número de ocorrências de um fenómeno por unidade de tempo (comprimento, área, ...). Diz-se que X tem uma **distribuição de Poisson** ou $X \sim Poi(\lambda)$, com $\lambda \in \mathbb{R}^+$ quando

$$f_X(x) = e^{-\lambda} \frac{\lambda^x}{x!}, \ x = 0, 1, \ldots$$

em que λ é a taxa média de ocorrências por unidade de tempo.

- $E[X] = Var[X] = \lambda$

Funções de probabilidade de Poisson



4.5 Distribuição uniforme contínua

Definição

Se

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{caso contrário} \end{cases}$$

então diz-se que X tem uma **distribuição uniforme contínua** no intervalo $[a, b]$ ou $X \sim U(a, b)$, com $a < b \in \mathbb{R}$.

Notas

1. $E[X] = \int_a^b x \frac{1}{b-a}\, dx = \frac{a+b}{2}$

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

4.6 Distribuição exponencial

Definição

Se

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

então diz-se que X tem uma **distribuição exponencial** ou $X \sim Exp(\lambda)$, com $\lambda > 0$.

Notas

1. $F_X(x) = \begin{cases} 0, & x < 0 \\ \int_0^x \lambda e^{-\lambda t}\, dt = 1 - e^{-\lambda x}, & x \geq 0 \end{cases}$

2. $E[X] = \frac{1}{\lambda}$ e $Var[X] = \frac{1}{\lambda^2}$

Amnésia da distribuição exponencial

$X \sim Exp(\lambda) \Rightarrow$

$\Rightarrow P(X > s + t \mid X > t) = P(X > s),\ \forall s, t \geq 0$

Teorema

Seja X uma variável aleatória que representa o número de ocorrências por unidade de tempo (comprimento, área, etc.) de um qualquer fenómeno e Y uma outra variável aleatória que representa o tempo entre ocorrências sucessivas.

Se $X \sim Poi(\lambda)$ então $Y \sim Exp(\lambda)$.

Nota

O teorema anterior também se aplica se a variável aleatória Y representar o tempo até à primeira ocorrência do fenómeno.

Exemplo – A ver navios . . .

O tempo em horas entre chegadas sucessivas de veleiros a uma marina é uma variável aleatória com distribuição exponencial de valor esperado 6.

a. Sabendo que o último veleiro chegou há mais de 2 horas, calcule a probabilidade de se passar um período de mais de 8 horas sem qualquer nova chegada.

b. Calcule a probabilidade de chegarem 2 ou mais veleiros num período de 8 horas.

4.7 Distribuição normal

Definição

Se

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

então diz-se que X tem uma **distribuição normal** ou gaussiana ou $X \sim N(\mu, \sigma^2)$, com $\mu \in \mathbb{R}$ e $\sigma^2 > 0$.

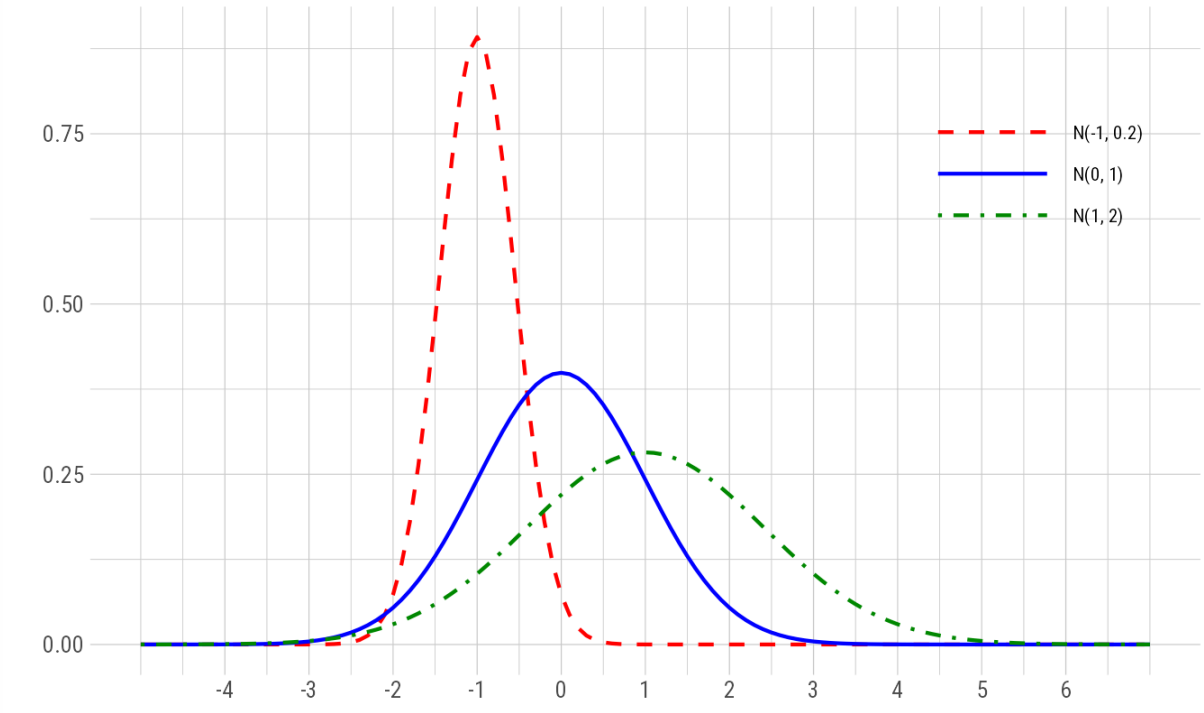
Notas

1. $f_X(\mu - x) = f_X(\mu + x), \forall x > 0$

2. $E[X] = \mu$ e $Var[X] = \sigma^2$

3. Moda $=$ Mediana $= \mu$

Funções densidade de probabilidade gaussianas



Teorema

Se $X \sim N(\mu, \sigma^2)$ e $Y = aX + b$, com $a \neq 0$, então

$$Y \sim N\left(a\mu + b, a^2\sigma^2\right)$$

Aplicação

Sejam $a = \dfrac{1}{\sigma}$ e $b = -\dfrac{\mu}{\sigma}$, isto é, $Y = \dfrac{X - \mu}{\sigma}$.

Então $Y \sim N(0, 1) \equiv$ **distribuição normal reduzida** ou *standard*.

⭐ As funções $\Phi(y) = F_Y(y)$ e $\Phi^{-1}(y)$ encontram-se tabeladas.

4.8 Distribuições no R

No *R* básico temos acesso a 17 das mais comuns distribuições univariadas e a mais outras duas menos comuns. Todas as funções tem as seguintes formas:

Função	Descrição
pnome(. . .)	função de distribuição
dnome(. . .)	função de probabilidade ou densidade de probabilidade
qnome(. . .)	inversa da função de distribuição
rnome(. . .)	geração de números aleatórios

em que **nome** é uma abreviatura do nome usual da distribuição (binom, geom, pois, unif, exp, norm, . . .).

Muitas outras distribuições são disponibilizadas por diversos pacotes (extraDistr, . . .).

Ver https://cran.r-project.org/web/views/Distributions.html

Atenção!

Verificar sempre a definição e a parametrização de qualquer distribuição disponível no *R*.

Simulação

A simulação de sistemas sujeitos a variações aleatórias é fundamental em muitas áreas científicas.

Em qualquer simulação é central a geração de **números pseudo-aleatórios**:

1. números gerados por algum algoritmo determinista, e que,
2. uma vez gerados, escapam aos melhores esforços para se detetarem padrões.

Cada réplica de uma experiência de simulação nas mesmas condições iniciais produz resultados diferentes.

Para controlar uma simulação e torná-la repetível é necessário fixar a **semente do gerador**:

```
set.seed(integer)
```

Tiragens

O comando `sample` permite simular tiragens de um qualquer conjunto, com ou sem reposição.

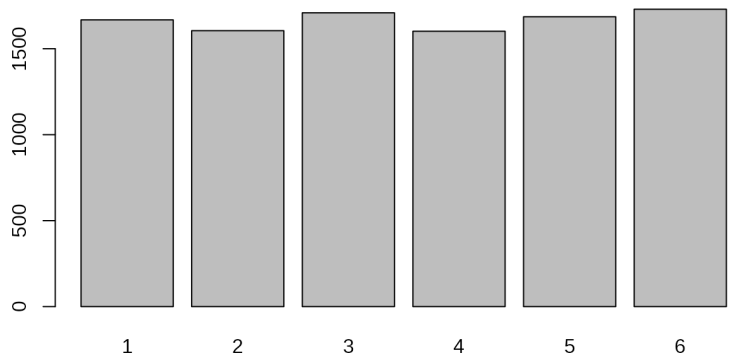
Exemplo – 1 lançamento de um dado cúbico equilibrado

```
# sample(c(1, 2, 3, 4, 5, 6), size = 1)
# sample(1:6, 1)
sample(6, 1)
```

```
[1] 3
```

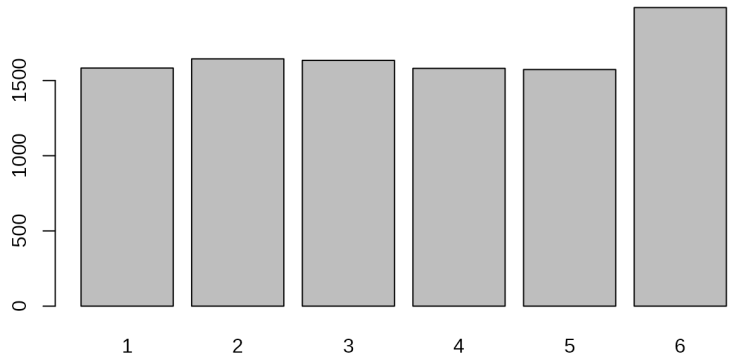
Exemplo – 10000 lançamentos de um dado cúbico equilibrado

```
res <- sample(6, 10000, replace = TRUE)
barplot(table(res))
```



Exemplo – 10000 lançamentos de um dado cúbico viciado

```
res <- sample(6, 10000, replace = TRUE, prob = c(1, 1, 1, 1, 1, 1.25))
barplot(table(res))
```



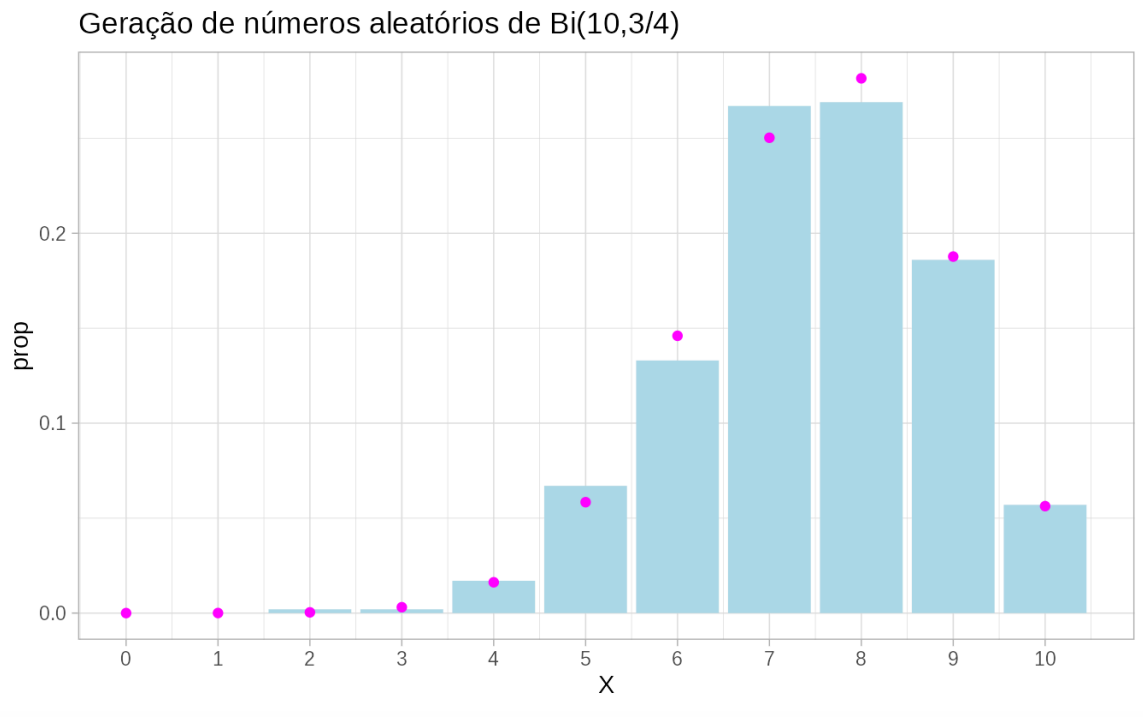
Exemplo – Geração de números aleatórios

Gerar conjuntos de valores de dimensão 1000 de

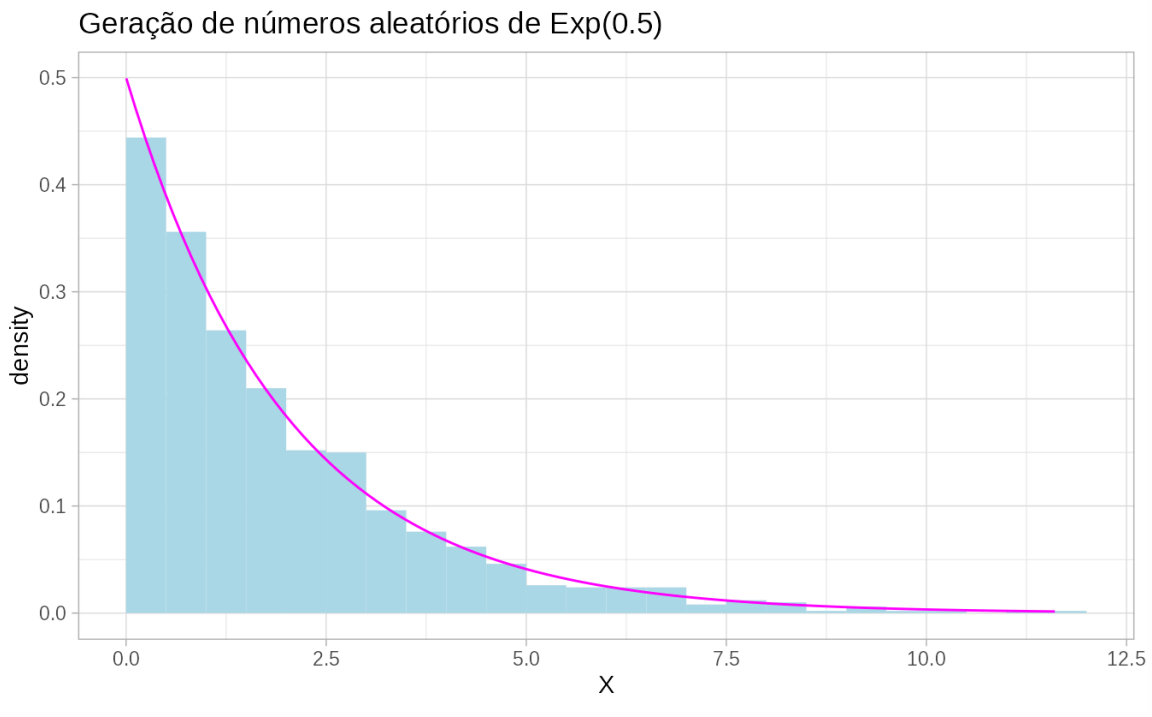
1. $X \sim Bi(10, 3/4)$ e
2. $X \sim Exp(0.5)$,

e comparar a distribuição empírica dos valores gerados com as respetivas distribuições teóricas.

```
n <- 1000
dados <- data.frame(X = rbinom(n, size = 10, prob = 3/4))

teorico <- data.frame(x = 0:10, y = dbinom(0:10, 10, 3/4))

ggplot(dados) +
  geom_bar(aes(x = X, y = after_stat(prop)), fill = "lightblue") +
  geom_point(data = teorico, aes(x, y), color = "magenta") +
  scale_x_continuous(breaks = 0:10) +
  labs(title = "Geração de números aleatórios de Bi(10,3/4)") +
  theme_light()
```



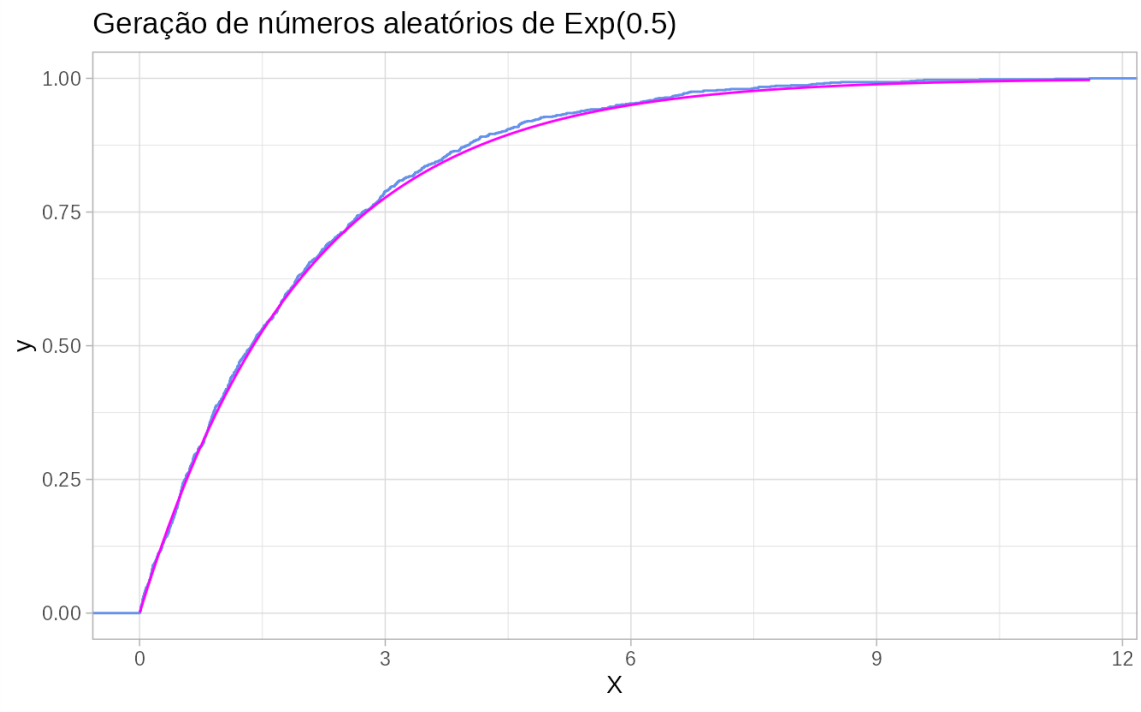

```
n <- 1000
dados <- data.frame(X = rexp(n, 0.5))

func <- function(x) dexp(x, 0.5)

ggplot(dados) +
  geom_histogram(aes(x = X, y = after_stat(density)), binwidth = 0.5,
                 fill = "lightblue", boundary = 0) +
  geom_function(fun = func, color = "magenta") +
  labs(title = "Geração de números aleatórios de Exp(0.5)") +
  theme_light()
```



Exemplo – Função de distribuição empírica

$$F_n(x) = \frac{\#\{x_i : x_i \leq x\}}{n}$$

```
ggplot(dados) +
  geom_function(fun = pexp, args = list(rate = 0.5),
                color = "magenta") +
  stat_ecdf(aes(X), color = "cornflowerblue") +
  labs(title = "Geração de números aleatórios de Exp(0.5)") +
  theme_light()
```



Exemplo – Geração de números aleatórios

Seja $X \sim N(0, 4)$. Calcular um valor aproximado de $E\left[\cos^2 X\right]$.

```
n <- 50000
x <- rnorm(n, 0, 2)
y <- cos(x)^2
mean(y)
```

```
[1] 0.503
```

Valor exato: $1/2\left(1 + e^{-8}\right) \approx 0.500168$