PUC-RIO
INF1036 - Probabilidade Computacional

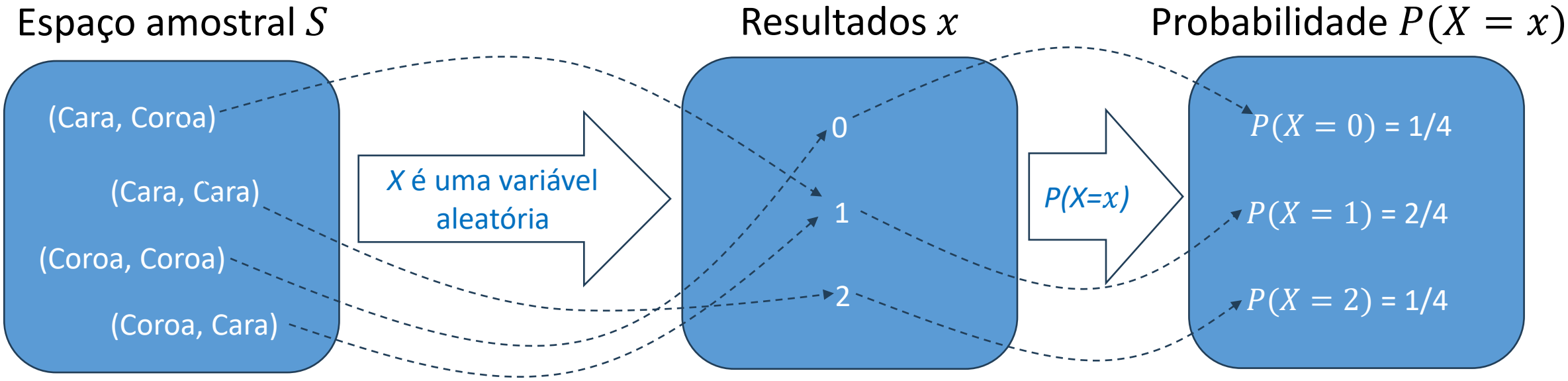
Material 6 – Variáveis Aleatórias

*Material Adaptado dos Professores Hélio Lopes / Ana Carolina Letichevsky

Experimento – Lançar uma moeda duas vezes

Uma **variável aleatória** (v.a.) é uma função que associa **cada resultado possível** de um **experimento aleatório** a **um número real**. Ela é usada em probabilidade e estatística para **quantificar incertezas**.

Considere o experimento aleatório **"Lançar uma moeda não viciada duas vezes e observar os resultados."** e a variável aleatória $\boldsymbol{X}$ **= "Número de caras que ocorre no experimento aleatório."**.



$$S = \left\{ \begin{matrix} (Cara, Coroa), (Cara, Cara), \\ (Coroa, Coroa), (Coroa\ Coroa) \end{matrix} \right\}$$

Como os 4 resultados do espaço amostral são igualmente prováveis (probabilidade 1/4 cada), podemos montar a distribuição de probabilidade de X.

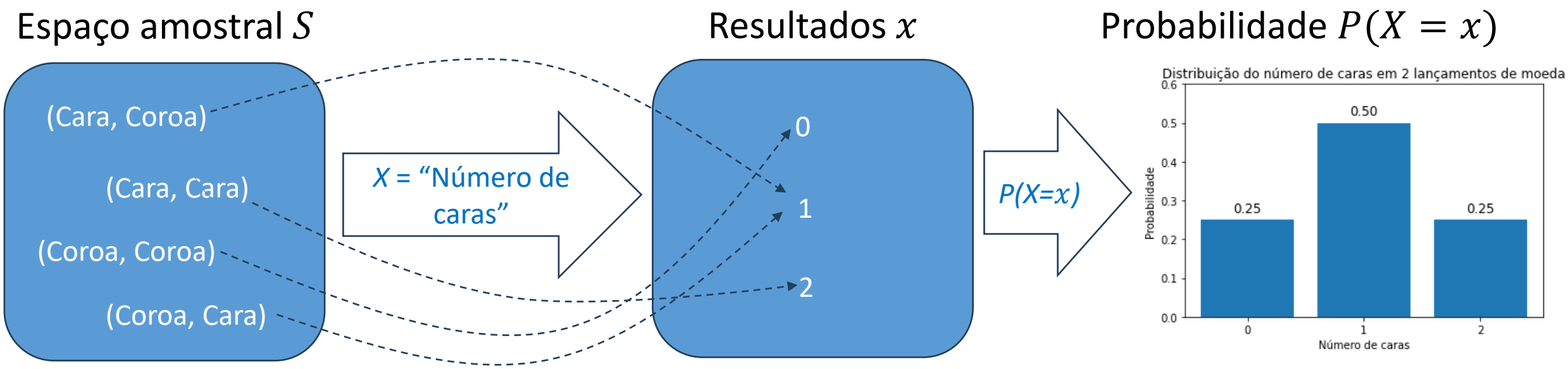
Variáveis aleatórias

Dado um experimento aleatório, descrito pelo espaço de probabilidades, uma função numérica $\boldsymbol{X}$:

$$\boldsymbol{X} : \boldsymbol{S} \rightarrow \boldsymbol{R}, \text{ onde } \boldsymbol{S} \text{ é o conjunto de resultados possíveis.}$$

É uma **variável aleatória** do experimento. Após a execução do experimento, obtém-se um **valor numérico específico** da variável aleatória $\boldsymbol{X}$, denominado **realização**.

Nos interessa estudar as **distribuições de probabilidade** das **variáveis aleatórias**.



Variáveis aleatórias

Considere o experimento aleatório **"Lançar uma moeda não viciada duas vezes e observar os resultados"** e a variável aleatória $\boldsymbol{X}$ **= "número de caras que ocorre no experimento aleatório"**.

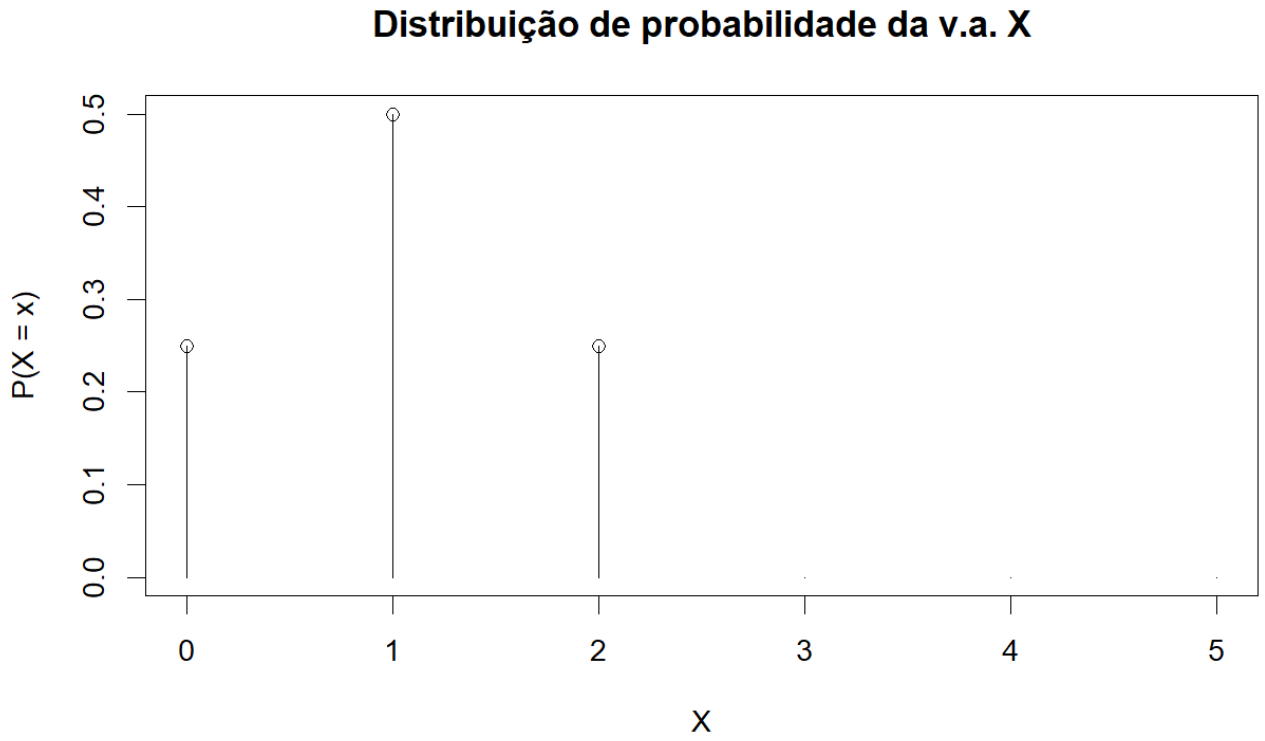
Probabilidade $P(X = x)$

$P(X = 0)$ = 1/4

$P(X = 1)$ = 2/4

$P(X = 2)$ = 1/4

$$\sum_{i=1}^{n} P(X = x_i) = 1$$



A **distribuição de probabilidade** de uma v.a. X é o conjunto de todos os pares ordenados $(x_i, P(X = x_i))$ e pode ser expressa em forma de tabela, gráfico ou fórmula. O gráfico acima apresenta a **distribuição de probabilidade** da variável aleatória X do exemplo.

Variáveis aleatórias

Uma **variável aleatória** (v.a.) é uma função que associa cada resultado possível de um experimento aleatório a um número real. Ela é usada em probabilidade e estatística para quantificar incertezas.

As **variáveis aleatórias** podem ser classificadas em:

- **Discreta:** assume um número finito ou infinito enumerável de possíveis valores;
- **Contínua:** pode assumir um número infinito de possíveis valores dentro de um intervalo contínuo;
- **Mista:** pode tanto assumir valores discretos, quanto assumir todos os valores em um determinado intervalo contínuo.

Variáveis aleatórias discretas

Uma **variável aleatória discreta** assume um número finito ou infinito enumerável de possíveis valores.

Exemplos de experimentos e variáveis aleatórias discretas:

1) E = "Observar aviões que chegam em determinado aeroporto no período de um mês".

   X = "número de aviões que chegam atrasados".

   $X = \{0, 1, 2, 3, \dots\}$

2) E = "Observar os carros de um determinado estacionamento".

   X = "número de carros que são da marca Fiat".

   $X = \{0, 1, 2, 3, \dots\}$

3) E = "Lançar uma moeda 10 vezes".

   X = "número de lançamentos até a primeira cara ser obtida".

   $X = \{0, 1, 2, \dots, 10\}$

Função de probabilidade

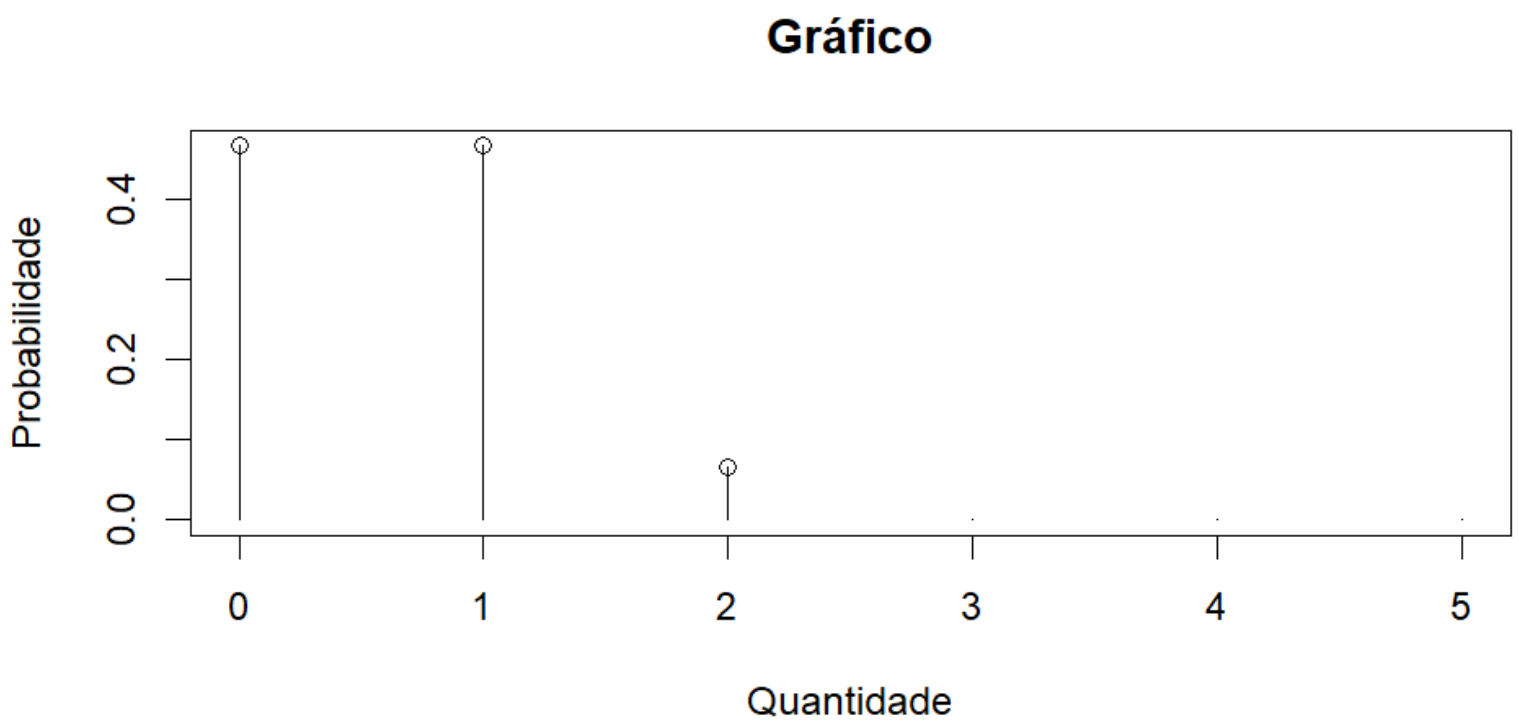
Exemplo) Uma urna possui 10 bolas sendo 3 com o número 1 e 7 com o número 2. Usando simulação, obtenha a probabilidade de se obter nenhuma, uma ou duas bolas número 1 em duas extrações sem reposição. Plote um gráfico com as probabilidades obtidas.

```
nsamples <- 10000
urna <- c(rep(1, 3), rep(2, 7))
probabilidadeSelecao <- rep(0.1, 10)
totalBolas <- rep(0, 3) # Nenhuma, uma ou duas bolas.
for (i in 1:nsamples) {
  bolas <- sample(urna, 2, replace = F, prob = probabilidadeSelecao) # Sem reposição.
  if ((bolas[1] == 1) & (bolas[2] == 1)) {
    totalBolas[3] = totalBolas[3] + 1
  } else if ((bolas[1] == 1) | (bolas[2] == 1)) {
    totalBolas[2] = totalBolas[2] + 1
  } else {
    totalBolas[1] = totalBolas[1] + 1
  }
}
sprintf("A probabilidade de obter nenhuma bola e %.2f.", totalBolas[1]/nsamples)
sprintf("A probabilidade de obter uma bola e %.2f.", totalBolas[2]/nsamples)
sprintf("A probabilidade de obter duas bolas e %.2f.", totalBolas[3]/nsamples)
```

Função de probabilidade

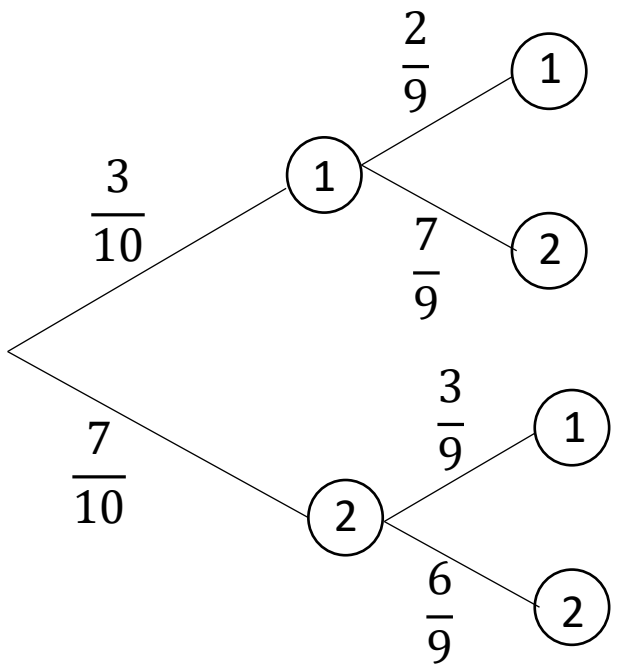
```
q <- 0:5
p <- c(totalBolas[1]/nsamples, totalBolas[2]/nsamples, totalBolas[3]/nsamples, 0, 0, 0)
plot (q, p, main = "Gráfico", xlab = "Quantidade", ylab = "Probabilidade", type = "h")
points (0, totalBolas[1]/nsamples)
points (1, totalBolas[2]/nsamples)
points (2, totalBolas[3]/nsamples)
```

Quantidade	*Probabilidade*
2	0.066491
1	0.466805
0	0.466704

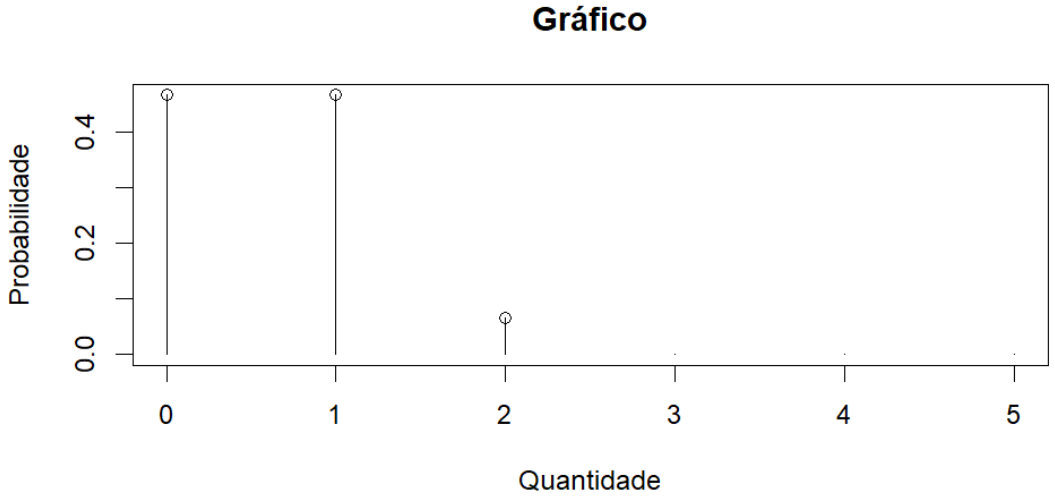


Função de probabilidade

Exemplo) Uma urna possui 10 bolas sendo 3 com o número 1 e 7 com o número 2. Definimos a variável aleatória X = **"número de bolas com o número 1 obtidas em duas extrações sem reposição"**.



Resultados Possíveis		Probabilidades	X	$P(X = x)$
1	1	6/90 = 2/30	2	2/30
1	2	21/90 = 7/30	1	14/30
2	1	21/90 = 7/30	1	
2	2	42/90 = 14/30	0	14/30



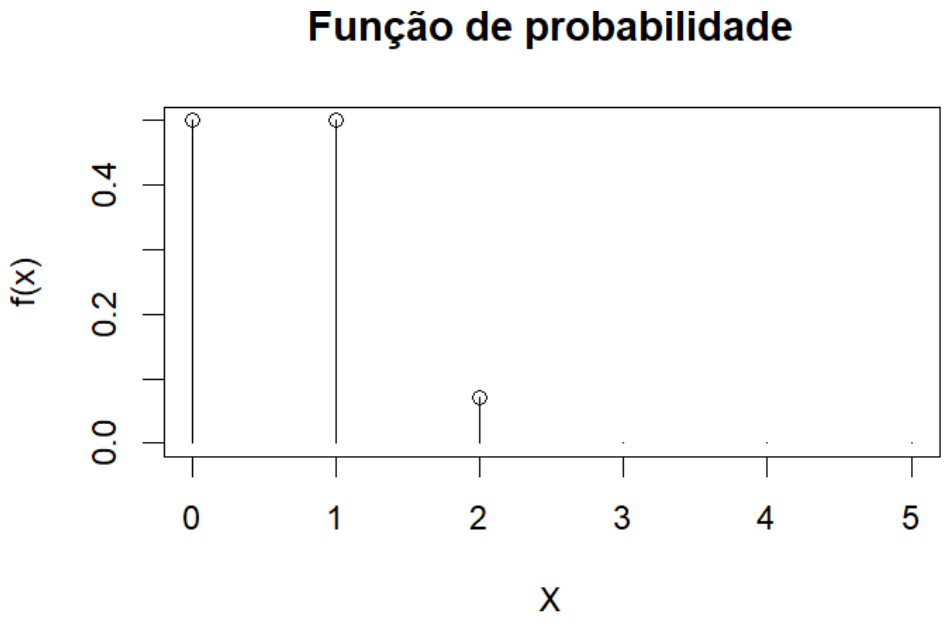
Função de probabilidade

Exemplo, continuação) Uma urna possui 10 bolas sendo 3 com o número 1 e 7 com o número 2. Definimos a variável aleatória $\boldsymbol{X}$ = **"número de bolas com o número 1 obtidas em duas extrações sem reposição"**.

A **função de probabilidade** da v.a. discreta $\boldsymbol{X}$ é a função $\boldsymbol{f(x)}$, que atribui a cada possível valor $\boldsymbol{x_i}$ de $\boldsymbol{X}$, sua respectiva probabilidade

$$f(x) = P(X = x) = \begin{cases} 2/30, se\ x = 2 \\ 14/30, se\ x = 1 \\ 14/30, se\ x = 0 \end{cases}$$

X	$f(x)$
2	2/30
1	14/30
0	14/30



$\boldsymbol{f(x)}$ apresenta as seguintes propriedades:

- $f(x_i) = P(X = x_i), \forall i = 1, 2, \dots n$
- $f(x_i) \geq 0\ ,\ \ \forall i = 1, 2, \dots n$
- $\sum_{i=1}^{n} f(x_i) = 1$

$f(2) = P(X = 2) = 2/30$

Função de probabilidade

Experimento: extração, sem reposição, de duas bolas de uma urna que possui 10 bolas sendo 3 com o número 1 e 7 com o número 2.

$\boldsymbol{X}$: número de bolas com o número 1 obtidas.

Espaço amostral

1	1
1	2
2	1
2	2

X
2
1
0

X	$f(x) = P(X = x)$
2	2/30
1	14/30
0	14/30

$\boldsymbol{X}$ é uma variável aleatória, ou seja, é uma variável numérica obtida a partir de um experimento aleatório.

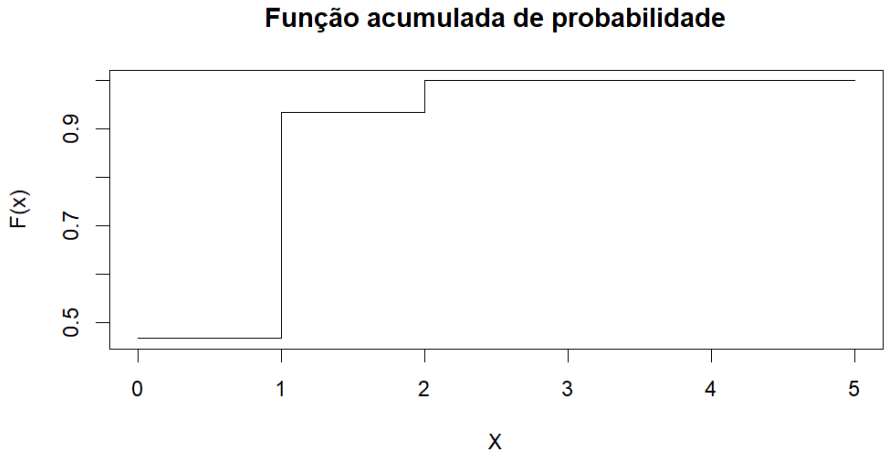
Função acumulada de probabilidade

Exemplo, continuação) Uma urna possui 10 bolas sendo 3 com o número 1 e 7 com o número 2. Definimos a variável aleatória $\boldsymbol{X}$ = **"número de bolas com o número 1 obtidas em duas extrações sem reposição"**.

A **função acumulada de probabilidade** de uma v.a. $\boldsymbol{X}$ discreta é dada pela expressão $\boldsymbol{F(x) = P(X \leq x)}$ e fornece a probabilidade acumulada até um certo valor $\boldsymbol{x}$.

X	$f(x)$
2	2/30
1	14/30
0	14/30

$$F(x) = P(X \leq x) = \begin{cases} 30/30, & se\ x \geq 2 \\ 28/30, & se\ 1 \leq x < 2 \\ 14/30, & se\ 0 \leq x < 1 \\ 0, & se\ x < 0 \end{cases}$$




```
x <- -1:5
p <- c (0, 14/30, 28/30, 30/30, 30/30, 30/30, 30/30)
plot (x, p, main = "Função acumulada de probabilidade", xlab = "X", ylab = "F(x)", type = "s")
```

Função acumulada de probabilidade

Experimento : extração, sem reposição, de duas bolas de uma urna que possui 10 bolas sendo 3 com o número 1 e 7 com o número 2.

$\boldsymbol{X}$: número de bolas com o número 1 obtidas.

Espaço amostral

1	1
1	2
2	1
2	2

X
2
1
0

X	$f(x) = P(X = x)$	$F(x) = P(X \leq x)$
2	2/30	30/30
1	14/30	28/30
0	14/30	14/30

$\boldsymbol{X}$ é uma variável aleatória, ou seja, é uma variável numérica obtida a partir de um experimento aleatório.

Função acumulada de probabilidade

Exercício) Uma população de 1000 crianças foi analisada para determinar a efetividade de um remédio contra um tipo de micose. As crianças receberam uma dose do remédio e, após um mês, passavam por uma nova bateria de exames. Caso ainda tivessem alguma micose, recebiam outra dose de remédio. Ao fim de cinco doses todas as crianças foram consideradas como não apresentando mais micose. A tabela abaixo apresenta os resultados obtidos. Elabore a função de probabilidade e a função acumulada de probabilidade. Para uma criança sorteada ao acaso, qual a probabilidade dela ter recebido duas doses? E até duas doses?

Doses	Números de crianças que receberem o remédio
1	240
2	270
3	260
4	180
5	50

Função acumulada de probabilidade

Exercício) Uma população de 1000 crianças foi analisada para determinar a efetividade de um remédio contra um tipo de micose. As crianças receberam uma dose do remédio e, após um mês, passavam por uma nova bateria de exames. Caso ainda tivessem alguma micose, recebiam outra dose de remédio. Ao fim de cinco doses todas as crianças foram consideradas como não apresentando mais micose. A tabela abaixo apresenta os resultados obtidos. Elabore a função de probabilidade e a função acumulada de probabilidade. Para uma criança sorteada ao acaso, qual a probabilidade dela ter recebido duas doses? E até duas doses?

Doses	Números de crianças que receberem o remédio	$f(x)$	$F(x)$
1	240	0,24	0,24
2	270	0,27	0,51
3	260	0,26	0,77
4	180	0,18	0,95
5	50	0,05	1,00

Distribuições

Diversas situações reais muitas vezes se aproximam de certas **distribuições estocásticas** definidas por algumas hipóteses, daí a importância de se conhecer e manipular algumas destas distribuições tão presentes em nosso cotidiano.

As distribuições de probabilidade podem ser **contínuas** ou **discretas**, dependendo se elas definem probabilidades para **variáveis contínuas** ou **discretas**.

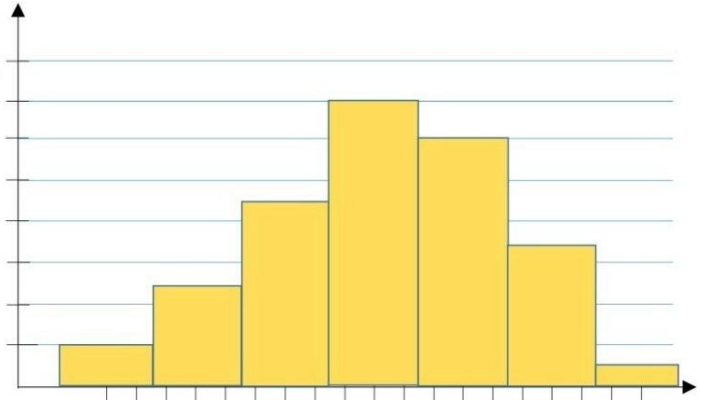
Com uma **distribuição de probabilidade discreta**, cada valor possível da variável aleatória discreta pode ser associado a uma probabilidade diferente de zero. Deste modo, uma distribuição de probabilidade discreta é, por vezes, apresentada em forma de tabela.

As **probabilidades de variáveis aleatórias contínuas** são definidas como a área sob a curva da sua distribuição. Assim, apenas as faixas de valores podem ter uma probabilidade diferente de zero. A probabilidade de que uma variável aleatória contínua seja igual a algum valor é sempre zero.

Distribuições Discretas e Contínuas

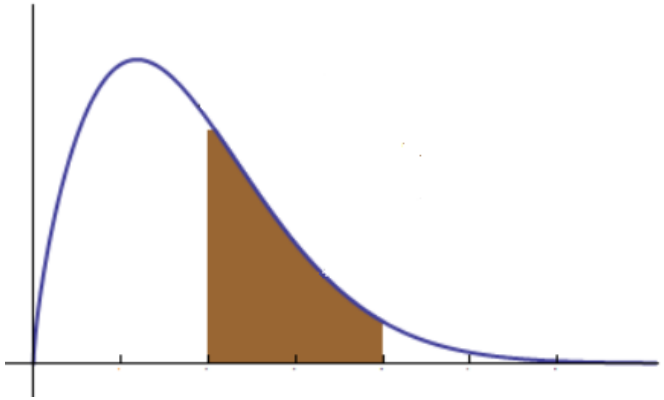
Variável aleatória discreta

Distribuição de probabilidade discreta



Variável aleatória contínua

Distribuição de probabilidade contínua



Distribuição Uniforme Discreta

A v.a. discreta X, que assume os valores x_1, x_2,..., x_n, tem **Distribuição Uniforme Discreta** se:

$$f(x_i) = P(X = x_i) = \frac{1}{n}, \qquad \forall i = 1, 2, \dots, n$$

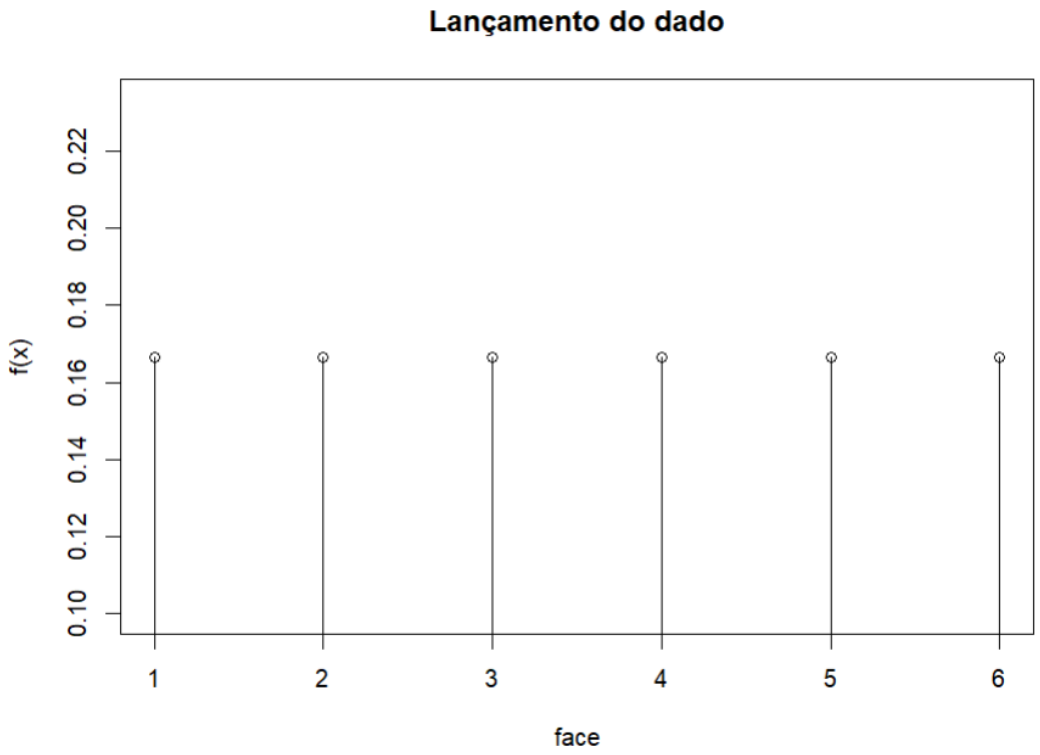
Note que, em uma **Distribuição Uniforme Discreta**, todos os valores são igualmente prováveis.

Além disso, para que uma v.a. X tenha **Distribuição Uniforme Discreta**, é necessário que X assuma um número finito de valores, já que $\sum_{i=1}^{n} f(x_i) = 1$.

Distribuição Uniforme Discreta

Exemplo 1) Suponha o lançamento de um dado não viciado de 6 faces. Qual a probabilidade de se obter a face 5?

$$f(x_i) = P(X = x_i) = \frac{1}{6}, \qquad \forall i = 1, 2, \dots, 6$$



Ensaio de Bernoulli

Um ensaio ou experimento de **Bernoulli** é um experimento aleatório com apenas dois resultados possíveis, por convenção, um deles é chamado “sucesso” e o outro, “fracasso”.

Um ensaio de **Bernoulli** é, portanto, definido por:

- Em cada ensaio considera-se somente a ocorrência ou não ocorrência de um certo evento que será denominado “sucesso” e cuja não ocorrência será denominada “fracasso”.
- Os ensaios são independentes.
- A probabilidade de “sucesso”, p, é a mesma para cada ensaio.
- A probabilidade de “fracasso”, $1 - p$, é a mesma para cada ensaio.

Distribuição de Bernoulli

Seja a v.a. X, definida como o número de sucessos num ensaio de **Bernoulli**. Então:

$$X = \begin{cases} 1, & se\ ocorre\ sucesso \\ 0, & se\ ocorre\ fracasso \end{cases}$$

A v.a. X assim definida chama-se variável aleatória de **Bernoulli**.

Chamando de p a probabilidade de sucesso $(0 < p < 1)$, a **Distribuição de Bernoulli** é dada por:

$$f(x) = \begin{cases} p, & se\ x = 1 \\ 1 - p, & se\ x = 0 \end{cases}$$

Distribuição de Bernoulli

Exemplo 1) Considere o lançamento de uma moeda viciada e a seguinte v.a. X associada a esse experimento:

$$X = \begin{cases} 1, & se\ ocorre\ cara \\ 0, & se\ ocorre\ coroa \end{cases}$$

Seja $p = 0{,}4$ a probabilidade de cara. Então, X tem **Distribuição de Bernoulli** com parâmetro p dada por:

$$f(x) = \begin{cases} 0{,}4, & se\ x = 1 \\ 0{,}6, & se\ x = 0 \end{cases}$$

Note que, nesse caso, a **Distribuição Bernoulli** com parâmetro $p = 0{,}5$ é equivalente à **Distribuição Uniforme**.

Distribuição Binomial

A **Distribuição Binomial** descreve o **número de sucessos** que ocorrem em um determinado número de experimentos de **Bernoulli**. Recordando, um ensaio de Bernoulli é definido por:

- Em cada ensaio considera-se somente a ocorrência ou não ocorrência de um certo evento que será denominado "sucesso" e cuja não ocorrência será denominada "fracasso".
- Os ensaios são independentes.
- A probabilidade de "sucesso", p, é a mesma para cada ensaio.
- A probabilidade de "falha", $1 - p$, é a mesma para cada ensaio.

Para um experimento que consiste na realização de n ensaios independentes de Bernoulli, a probabilidade de um ponto amostral com sucessos nos k primeiros ensaios e fracassos nos $n - k$ ensaios seguintes é $p^k(1-p)^{n-k}$.

Distribuição Binomial

O número de pontos do espaço amostral que satisfaz essa condição é igual ao número de maneiras com que podemos escolher k ensaios para a ocorrência de sucesso dentre o total de n ensaios, pois nos $n - k$ restantes deverão ocorrer fracassos. Este número é igual ao número de combinações de n elementos tomados k a k, ou seja:

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$

Seja X o número de sucessos obtidos na realização de n ensaios de Bernoulli independentes. Se X tem Distribuição Binomial com parâmetros n e p, em que p é a probabilidade de sucesso em cada ensaio, então sua função de probabilidade é:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} = \left(\frac{n!}{k!(n-k)!}\right) p^k (1-p)^{n-k}$$

Distribuição Binomial

A **função acumulada** da **Distribuição Binomial** (n, p) é dada por:

$$P(X \leq k) = \sum_{x=0}^{k} \binom{n}{x} p^x (1-p)^{n-x}$$

A **função acumulada** $F(x)$ da **Distribuição Binomial** é o somatório das probabilidades individuais de cada ponto da $f(x)$ até que um determinado valor k seja obtido.

Distribuição Binomial

Exemplo 1) Suponha que, para uma máquina que produz peças automotivas, a probabilidade de se obter uma peça defeituosa (sucesso) é $p = 0{,}1$. Das peças produzidas por esta máquina, seleciona-se uma amostra de 10 peças para análise. Qual a probabilidade de se obter:

a) Uma peça defeituosa?

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} = \left(\frac{10!}{1!(10-1)!}\right) 0{,}1^1 (1-0{,}1)^{10-1} = 0{,}387$$

b) Nenhuma peça defeituosa?

$$P(X = 0) = \binom{10}{0} 0{,}1^0 (1-0{,}1)^{10-0} = 0{,}349$$

c) Duas peças defeituosas?

$$P(X = 2) = \binom{10}{2} 0{,}1^2 (1-0{,}1)^{10-2} = 0{,}194$$

Distribuição Binomial

Exemplo 1) Suponha que, para uma máquina que produz peças automotivas, a probabilidade de se obter uma peça defeituosa (sucesso) é $p = 0{,}1$. Das peças produzidas por esta máquina, seleciona-se uma amostra de 10 peças para análise. Qual a probabilidade de se obter:

d) No mínimo duas peças defeituosas?

$$P(X \geq 2) = P(X = 2) + \cdots + P(X = 10) = 1 - [P(X = 0) + P(X = 1)] = 0{,}264$$

e) No máximo duas peças defeituosas?

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = 0{,}930$$

Distribuições no R

No R, o trabalho com uma **variável aleatória X** está baseado em 4 procedimentos identificados pelas letras $\boldsymbol{p}$, $\boldsymbol{d}$, $\boldsymbol{q}$ e $\boldsymbol{r}$ que, adicionadas previamente ao **código das distribuições,** permitem o seu uso com **diferentes propósitos**:

- **rdistrib**: Gera **números aleatórios** baseados na distribuição definida. Necessita de argumentos especificando o tamanho da amostra, além dos parâmetros requeridos pela distribuição de interesse;
- **pdistrib**: Calcula a **função acumulada de probabilidade $\boldsymbol{F(x)}$;**
- **ddistrib**: Calcula a **função de probabilidade $\boldsymbol{f(x)}$**;
- **qdistrib**: Calcula o **quantil** correspondente a uma dada probabilidade.

Distribuições no Python

No Python, o trabalho com uma **variável aleatória X** pode ser realizado através do módulo stats da biblioteca SciPy. Ele apresenta métodos comuns, como rvs, pdf, cdf, ppf e stats, às distribuições que permitem o seu uso com **diferentes propósitos**:

- **rvs**: Gera **números aleatórios** baseados na distribuição definida;
- **cdf**: Calcula a **função acumulada de probabilidade $F(x)$;**
- **pmf**: Calcula a **função de probabilidade $f(x)$**;
- **ppf**: Percent Point Function (inverse of cdf — percentiles).
- **stats**: Return mean, variance, skew, and/or kurtosis.

Distribuição Binomial no R

Distribuição	Sintaxe	Opções
Binomial (**binom**)	**dbinom** (x, n, prob, opções)	**x**: número de sucessos em n ensaios de Bernoulli (pode ser um vetor). **p**: probabilidade de sucesso (pode ser um vetor). **q**: número máximo de sucessos em ensaios de Bernoulli (pode ser um vetor). **prob**: probabilidade de sucesso em cada ensaio de Bernoulli. **obs**: número de observações. **n**: número de ensaios de Bernoulli.
	pbinom (q, n, prob, opções)	
	qbinom (p, n, prob, opções)	
	rbinom (obs, n, p)	

Distribuição Binomial

Exemplo 2) Suponha que, para uma máquina que produz peças automotivas, a probabilidade de se obter uma peça defeituosa (sucesso) é $p = 0{,}1$. Das peças produzidas por esta máquina, seleciona-se uma amostra de 10 peças para análise. Qual a probabilidade de se obter:

a) Uma peça defeituosa?

```
dbinom(1, 10, 0.1)
0.3874205
```

b) Nenhuma peça defeituosa?

```
dbinom(0, 10, 0.1)
0.3486784
```

c) Duas peças defeituosas?

```
dbinom(2, 10, 0.1)
0.1937102
```

Distribuição Binomial

Exemplo 2) Suponha que, para uma máquina que produz peças automotivas, a probabilidade de se obter uma peça defeituosa (sucesso) é $p = 0{,}1$. Das peças produzidas por esta máquina, seleciona-se uma amostra de 10 peças para análise. Qual a probabilidade de se obter:

d) No mínimo duas peças defeituosas?

```
1 - pbinom(1, 10, 0.1)
0.2639011
```

e) No máximo duas peças defeituosas?

```
pbinom(2, 10, 0.1)
0.9298092
```

Distribuição Binomial

```
nsamples <- 1000000
namostra <- 10; cont.defeituosa <- 0; npeca.sem.defeito <- 0; npeca.defeituosa <- rep(0, 10)
for (i in 1:nsamples) {
  cont.defeituosa <- 0
  for (j in 1:namostra) { # seleciono 10 peças
    peca.selecionada.defeituosa <- sample(c(T, F), 1, prob = c(0.1, 0.9))
    if (peca.selecionada.defeituosa) {
      cont.defeituosa <- cont.defeituosa + 1
    }
  }
  if (cont.defeituosa > 0) { # npeca.defeituosa[i] contém o total de peças i defeituosas
    npeca.defeituosa[cont.defeituosa] <- npeca.defeituosa[cont.defeituosa] + 1
  } else {
    npeca.sem.defeito <- npeca.sem.defeito + 1
  }
}
prob.nenhuma.defeituosa <- npeca.sem.defeito/nsamples
print(prob.nenhuma.defeituosa)
prob.uma.defeituosa <- npeca.defeituosa[1]/nsamples
print(prob.uma.defeituosa)
prob.duas.defeituosas <- npeca.defeituosa[2]/nsamples
print(prob.duas.defeituosas)
prob.minimo.duas.defeituosas <- sum(npeca.defeituosa[2:10])/nsamples # filtro no vetor
print(prob.minimo.duas.defeituosas)
prob.maximo.duas.defeituosas <- (npeca.sem.defeito + sum(npeca.defeituosa[1:2]))/nsamples
print(prob.maximo.duas.defeituosas)
```

Distribuição Binomial

```
nsamples <- 1000000
namostra <- 10; cont.defeituosa <- 0; npeca.sem.defeito <- 0; npeca.defeituosa <- rep(0, 10)
for (i in 1:nsamples) {
  # sorteio e obtenho o total de peças defeituosas
  cont.defeituosa <- sum(sample(c(T, F), 10, prob = c(0.1, 0.9), replace = T) == T)
  if (cont.defeituosa > 0) { # npeca.defeituosa[i] contém o total de peças i defeituosas
    npeca.defeituosa[cont.defeituosa] <- npeca.defeituosa[cont.defeituosa] + 1
  } else {
    npeca.sem.defeito <- npeca.sem.defeito + 1
  }
}
prob.nenhuma.defeituosa <- npeca.sem.defeito/nsamples
print(prob.nenhuma.defeituosa)
prob.uma.defeituosa <- npeca.defeituosa[1]/nsamples
print(prob.uma.defeituosa)
prob.duas.defeituosas <- npeca.defeituosa[2]/nsamples
print(prob.duas.defeituosas)
prob.minimo.duas.defeituosas <- sum(npeca.defeituosa[2:10])/nsamples # filtro no vetor
print(prob.minimo.duas.defeituosas)
prob.maximo.duas.defeituosas <- (npeca.sem.defeito + sum(npeca.defeituosa[1:2]))/nsamples
print(prob.maximo.duas.defeituosas)
```

Distribuição Binomial

Exemplo 3) Considere que a probabilidade de certa peça artesanal ser produzida com perfeição por uma pessoa que está em treinamento é igual a 0,5. Considere que a pessoa produza 6 peças por vez.

a) Obter a distribuição de probabilidades do número de peças perfeitas produzidas pela pessoa em 6 peças.

```
intervalo <- 0:6
# 0 1 2 3 4 5 6

vetor.probabilidade <- dbinom (intervalo, 6, 0.5)
# 0.015625 0.093750 0.234375 0.312500 0.234375 0.093750 0.015625
```

Distribuição Binomial

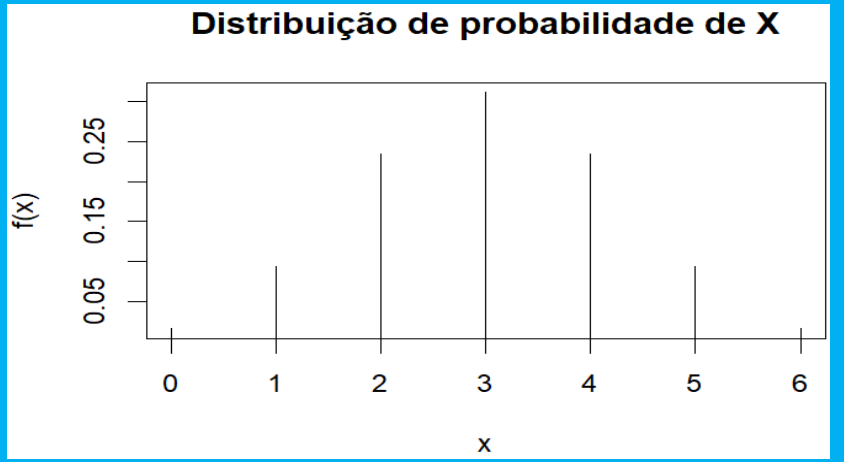
Exemplo 3) Considere que a probabilidade de certa peça artesanal ser produzida com perfeição por uma pessoa que está em treinamento é igual a 0,5. Considere que a pessoa produza 6 peças por vez.

b) Construir o gráfico de distribuição de probabilidade obtida no item a.

```
intervalo <- 0:6
# 0 1 2 3 4 5 6

vetor.probabilidade <- dbinom (intervalo, 6, 0.5)
# 0.015625 0.093750 0.234375 0.312500 0.234375 0.093750 0.015625


plot(intervalo, #intervalo desejado
     vetor.probabilidade, #valores de probabilidade
     type = "h", #traço do eixo x ao ponto
     xlab = 'x',
     ylab = 'f(x)',
     main = 'Distribuição de probabilidade de X')
```



Distribuição Binomial

```
nsamples <- 1000000
namostra <- 6; cont.defeituosa <- 0; npeca.sem.defeito <- 0; npeca.defeituosa <- rep(0, 6)
for (i in 1:nsamples) {
  cont.defeituosa <- 0
  for (j in 1:namostra) { # são produzidas 6 peças
    peca.selecionada.defeituosa <- sample(c(T, F), 1, prob = c(0.5, 0.5))
    if (peca.selecionada.defeituosa) {
      cont.defeituosa <- cont.defeituosa + 1
    }
  }
  if (cont.defeituosa > 0) { # npeca.defeituosa[i] contém o total de peças i defeituosas
    npeca.defeituosa[cont.defeituosa] <- npeca.defeituosa[cont.defeituosa] + 1
  } else {
    npeca.sem.defeito <- npeca.sem.defeito + 1
  }
}
vetor.probabilidade <- c(npeca.sem.defeito, npeca.defeituosa)
vetor.probabilidade <- vetor.probabilidade / nsamples
intervalo <- 0:6
plot(intervalo,
     vetor.probabilidade, #valores de probabilidade
     type = 'h', #traço do eixo ao ponto
     xlab = 'x',
     ylab = 'f(x)',
     main = 'Distribuição de probabilidade de X')
```

Distribuição Poisson

Frequentemente nos deparamos com situações em que o número de ensaios n é grande ($n \to \infty$) e p é pequeno ($p \to 0$), no cálculo da função **Binomial**, o que nos leva a algumas dificuldades. Nestes casos, trabalhamos com a distribuição **Poisson**.

A probabilidade de que existam exatamente k ocorrências (k sendo um inteiro não negativo, $k = 0, 1, 2, \ldots$) é:

$$f(k;\lambda) = P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

λ é um número real, que indica a taxa de ocorrência por unidade de medida.

Distribuição Poisson

Exemplo 1) A análise das estradas feita por uma empresa aponta que ocorrem, em média, 4 buracos por Km. A análise aponta, ainda, que a quantidade de buracos segue uma Distribuição Poisson. Qual a probabilidade de que em um Km qualquer contenha:

a) Um buraco?

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!} = \frac{e^{-4}4^1}{1!} = 0{,}073$$

b) Dois buracos?

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!} = \frac{e^{-4}4^2}{2!} = 0{,}147$$

c) Nenhum buraco?

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!} = \frac{e^{-4}4^0}{0!} = 0{,}018$$

d) Pelo menos um buraco?

$$P(X \geq k) = 1 - P(X < k) = 1 - P(X < 1) = 1 - P(X = 0) = 1 - \frac{e^{-\lambda}\lambda^0}{0!} = 0{,}982$$

Distribuições no R

No R, o trabalho com uma **variável aleatória** $\boldsymbol{X}$ está baseado em 4 procedimentos identificados pelas letras $\boldsymbol{p}$, $\boldsymbol{d}$, $\boldsymbol{q}$ e $\boldsymbol{r}$ que, adicionadas previamente ao código das **distribuições,** permitem o seu uso com **diferentes propósitos**:

- **rdistrib**: Gera **números aleatórios** baseados na distribuição definida. Necessita de argumentos especificando o tamanho da amostra, além dos parâmetros requeridos pela distribuição de interesse;
- **pdistrib**: Calcula a **função acumulada de probabilidade** $\boldsymbol{F(x)}$**;**
- **ddistrib**: Calcula a **função de probabilidade** $\boldsymbol{f(x)}$;
- **qdistrib**: Calcula o **quantil** correspondente a uma dada probabilidade.

Distribuição Poisson no R

Distribuição	Sintaxe	Opções
Poisson (**pois**)	**dpois** (x, lambda, opções)	**x**: número de ocorrências (pode ser um vetor). **lambda**: número médio de eventos ocorrendo no intervalo considerado. **q**: vetor contendo os quantis. **prob**: vetor contendo as probabilidades. **obs**: número de observações.
	ppois (q, lambda, opções)	
	qpois (prob, lambda, opções)	
	rpois (obs, lambda)	

Distribuição Poisson

Exemplo 2) A análise das estradas feita por uma empresa aponta que ocorrem, em média, 4 buracos por Km. A análise aponta, ainda, que a quantidade de buracos segue uma distribuição Poisson. Qual a probabilidade de que um Km qualquer contenha:

a) Um buraco?

```
dpois (1, 4)

0.07326256
```

b) Dois buracos?

```
dpois (2, 4)

0.1465251
```

c) Nenhum buraco?

```
dpois (0, 4)

0.01831564
```

d) Pelo menos um buraco?

```
1 - dpois (0, 4)

0.9816844
```

```
intervalo <- 0:20

vetor.probabilidade <- dpois (intervalo, 4)

plot(intervalo, #intervalo desejado
     vetor.probabilidade, #valores de probabilidade
     type = "h", #traço do eixo x ao ponto
     xlab = 'x',
     ylab = 'f(x)',
     main = 'Distribuição de probabilidade de X')
```

Distribuição Poisson

Exemplo 2) A análise das estradas feita por uma empresa aponta que ocorrem, em média, 4 buracos por Km. A análise aponta, ainda, que a quantidade de buracos segue uma distribuição Poisson. Qual a probabilidade de que um Km qualquer contenha:

```
from scipy.stats import poisson
```

a) Um buraco?

```
poisson.pmf(1, 4)

0.07326256
```

b) Dois buracos?

```
poisson.pmf(2, 4)

0.1465251
```

c) Nenhum buraco?

```
poisson.pmf(0, 4)

0.01831564
```

d) Pelo menos um buraco?

```
1 - poisson.pmf(0, 4)

0.9816844
```

Distribuição Poisson

Exemplo 2, continuação) A análise das estradas feita por uma empresa aponta que ocorrem, em média, 4 buracos por Km. A análise aponta, ainda, que a quantidade de buracos segue uma distribuição Poisson. Qual a probabilidade de que um Km qualquer contenha:

a) Um buraco?

```
nsamples <- 1000000; nkm.sem.buraco <- 0; nkm.com.buraco <- rep(0, 30)
for (i in 1:nsamples) {
  buraco <- rpois(1, 4) # uso de rpois com lambda = 4 e pegando uma observação
  if (buraco > 0) {
    nkm.com.buraco[buraco] <- nkm.com.buraco[buraco] + 1
  } else {
    nkm.sem.buraco <- nkm.sem.buraco + 1
  }
}
prob.um.buraco <- nkm.com.buraco[1] / nsamples
print(prob.um.buraco)
```

Distribuição Poisson

Exemplo 2, continuação) A análise das estradas feita por uma empresa aponta que ocorrem, em média, 4 buracos por Km. A análise aponta, ainda, que a quantidade de buracos segue uma distribuição Poisson. Qual a probabilidade de que um Km qualquer contenha:

b) Dois buracos?

```
nsamples <- 1000000; nkm.sem.buraco <- 0; nkm.com.buraco <- rep(0, 30)
for (i in 1:nsamples) {
  buraco <- rpois(1, 4) # uso de rpois com lambda = 4 e pegando uma observação
  if (buraco > 0) {
    nkm.com.buraco[buraco] <- nkm.com.buraco[buraco] + 1
  } else {
    nkm.sem.buraco <- nkm.sem.buraco + 1
  }
}
prob.dois.buracos <- nkm.com.buraco[2] / nsamples
print(prob.dois.buracos)
```

Distribuição Poisson

Exemplo 2, continuação) A análise das estradas feita por uma empresa aponta que ocorrem, em média, 4 buracos por Km. A análise aponta, ainda, que a quantidade de buracos segue uma distribuição Poisson. Qual a probabilidade de que um Km qualquer contenha:

c) Nenhum buraco?

```
nsamples <- 1000000; nkm.sem.buraco <- 0; nkm.com.buraco <- rep(0, 30)
for (i in 1:nsamples) {
  buraco <- rpois(1, 4) # uso de rpois com lambda = 4 e pegando uma observação
  if (buraco > 0) {
    nkm.com.buraco[buraco] <- nkm.com.buraco[buraco] + 1
  } else {
    nkm.sem.buraco <- nkm.sem.buraco + 1
  }
}
prob.nenhum.buraco <- nkm.sem.buraco / nsamples
print(prob.nenhum.buraco)
```

Distribuição Poisson

Exemplo 2, continuação) A análise das estradas feita por uma empresa aponta que ocorrem, em média, 4 buracos por Km. A análise aponta, ainda, que a quantidade de buracos segue uma distribuição Poisson. Qual a probabilidade de que um Km qualquer contenha:

d) Pelo menos um buraco?

```
nsamples <- 1000000; nkm.sem.buraco <- 0; nkm.com.buraco <- rep(0, 30)
for (i in 1:nsamples) {
  buraco <- rpois(1, 4) # uso de rpois com lambda = 4 e pegando uma observação
  if (buraco > 0) {
    nkm.com.buraco[buraco] <- nkm.com.buraco[buraco] + 1
  } else {
    nkm.sem.buraco <- nkm.sem.buraco + 1
  }
}
prob.pelo.menos.um.buraco <- sum(nkm.com.buraco) / nsamples
print(prob.pelo.menos.um.buraco)
```

Distribuição Geométrica

Descreve a quantidade de experimentos de **Bernoulli** que precisam ser realizados até a ocorrência do primeiro sucesso.

Para introduzir o modelo de variável geométrica, vamos retomar uma sequência de ensaios de **Bernoulli**, que é definida por:

- Em cada ensaio, considera-se somente a ocorrência ou não ocorrência de um certo evento que será denominado sucesso e cuja não ocorrência será denominada falha.
- Os ensaios são independentes.
- A probabilidade de sucesso, p, é a mesma para cada ensaio.
- A probabilidade de falha, $1 - p$, é a mesma para cada ensaio.

Seja X a variável aleatória que fornece o **número de repetições necessárias** para se obter o primeiro sucesso, ela tem **Distribuição Geométrica** com parâmetro p, $0 < p < 1$, se sua função de probabilidade é dada por:

$$P(X = k) = (1 - p)^{k-1}p$$

Distribuição Geométrica

Exemplo 1) Suponha que a probabilidade de um equipamento eletrônico ser defeituoso é de 0,2. Realizaram-se testes na produção de um determinado lote. Determine a probabilidade do primeiro defeito encontrado:

$$P(X = k) = (1 - p)^{k-1}p$$

a) Ocorrer no sétimo equipamento do teste:

$$P(7) = (1 - 0{,}2)^{7-1}(0{,}2) = 0{,}0524 = 5{,}24\%$$

b) Não ocorrer no quarto equipamento testado:

$$1 - P(4) = 1 - [(1 - 0{,}2)^{4-1}(0{,}2)]$$

$$= 1 - 0{,}1024 = 0{,}8976 = 89{,}76\%$$

Distribuições no R

No R, o trabalho com uma **variável aleatória** $\boldsymbol{X}$ está baseado em 4 procedimentos identificados pelas letras $\boldsymbol{p}$, $\boldsymbol{d}$, $\boldsymbol{q}$ e $\boldsymbol{r}$ que, adicionadas previamente ao código das **distribuições,** permitem o seu uso com **diferentes propósitos**:

- **rdistrib**: Gera **números aleatórios** baseados na distribuição definida. Necessita de argumentos especificando o tamanho da amostra, além dos parâmetros requeridos pela distribuição de interesse;
- **pdistrib**: Calcula a **função acumulada de probabilidade** $\boldsymbol{F(x)}$**;**
- **ddistrib**: Calcula a **função de probabilidade** $\boldsymbol{f(x)}$;
- **qdistrib**: Calcula o **quantil** correspondente a uma dada probabilidade.

Distribuição Geométrica no R

Distribuição	Sintaxe	Opções
Geométrica (**geom**)	**dgeom**(x, prob, opções)	**x**: **número de falhas ocorridas em uma sequência de Bernoulli antes do sucesso** (pode ser um vetor). **prob**: probabilidade de sucesso em cada tentativa. **q**: número de falhas (pode se um vetor). **p**: vetor contendo as probabilidades . **n**: número de observações.
	pgeom(q, prob, opções)	
	qgeom(p, prob, opções)	
	rgeom(n, prob)	

Distribuição Geométrica

Exemplo 2) Suponha que a probabilidade de um equipamento eletrônico ser defeituoso é de 0,2. Realizaram-se testes na produção de um determinado lote. Determine a probabilidade do primeiro defeito encontrado:

a) Ocorrer no sétimo equipamento do teste:

```
dgeom(6, 0.2)
0.0524288
```

Seis falhas devem ocorrer para termos sucesso na sétima.

b) Não ocorrer no quarto equipamento testado:

```
1 - dgeom(3, 0.2)
0.8976
```

Distribuição Geométrica

Exemplo 2, continuação) Suponha que a probabilidade de um equipamento eletrônico ser defeituoso é de 0,2. Realizaram-se testes na produção de um determinado lote. Determine a probabilidade do primeiro defeito encontrado:

a) Ocorrer no sétimo equipamento do teste:

```
nsamples <- 10000000
n.sem.defeito <- 0; n.com.defeito <- rep(0, 100)
for (i in 1:nsamples) {
  defeito <- rgeom(1, 0.2) # uso de rgeom com prob = 0.2 e pegando uma observação
                           # rgeon retorna o número de falhas antes do primeiro sucesso
  if (defeito > 0) {
    n.com.defeito[defeito] <- n.com.defeito[defeito] + 1
  } else {
    n.sem.defeito <- n.sem.defeito + 1
  }
}
prob.ocorrer.setimo <- n.com.defeito[6] / nsamples # seis sem defeito e o sétimo com defeito
print(prob.ocorrer.setimo)
```

Distribuição Geométrica

Exemplo 2, continuação) Suponha que a probabilidade de um equipamento eletrônico ser defeituoso é de 0,2. Realizaram-se testes na produção de um determinado lote. Determine a probabilidade do primeiro defeito encontrado:

b) Não ocorrer no quarto equipamento testado:

```
nsamples <- 10000000
n.sem.defeito <- 0; n.com.defeito <- rep(0, 100)
for (i in 1:nsamples) {
  defeito <- rgeom(1, 0.2) # uso de rgeom com prob = 0.2 e pegando uma observação
  if (defeito > 0) {
    n.com.defeito[defeito] <- n.com.defeito[defeito] + 1
  } else {
    n.sem.defeito <- n.sem.defeito + 1
  }
}

# Para não ocorrer no quarto, não há três falhas seguida de um sucesso.
# também poderia ser 1 - n.com.defeito[3]/nsamples
prob.nao.ocorrer.quarto <- (n.sem.defeito + sum(n.com.defeito) - n.com.defeito[3]) / nsamples
print(prob.nao.ocorrer.quarto)
```

Distribuição Geométrica

Exemplo 3) A probabilidade de uma regulagem bem sucedida na montagem de um determinado tipo de produto é de 0.80. Assuma que as tentativas são independentes.

a) Qual é a probabilidade de que a primeira regulagem bem sucedida requeira exatamente quatro tentativas?

```
dgeom (3, 0.8)
0.0064
```

Na função dgeom(x, p) x representa o número de fracassos, sendo neste caso $x = 4 - 1 = 3$.

b) Qual é a probabilidade de que a primeira regulagem bem sucedida requeira no máximo quatro tentativas ?

```
dgeom (0, 0.8) + dgeom (1, 0.8) + dgeom (2, 0.8) + dgeom (3, 0.8)
0.9984
```

c) Qual é a probabilidade de que a primeira regulagem bem sucedida requeira ao menos quatro tentativas ?

```
1 - (dgeom (0, 0.8) + dgeom (1, 0.8) + dgeom (2, 0.8))
0.008
```

Distribuição Geométrica

Exemplo 3, continuação) A probabilidade de uma regulagem bem sucedida na montagem de um determinado tipo de produto é de 0.80. Assuma que as tentativas são independentes.

a) Qual é a probabilidade de que a primeira regulagem bem sucedida requeira exatamente quatro tentativas?

```
nsamples <- 10000000
n.bem.sucedida <- 0; n.mal.sucedida <- rep(0, 100)
for (i in 1:nsamples) {
  regulagem <- rgeom(1, 0.8) # uso de rgeom com prob = 0.8 e pegando uma observação
  if (regulagem > 0) {
    n.mal.sucedida[regulagem] <- n.mal.sucedida[regulagem] + 1
  } else {
    n.bem.sucedida <- n.bem.sucedida + 1
  }
}
prob.quatro.tentativas <- n.mal.sucedida[3] / nsamples
print(prob.quatro.tentativas)
```

Distribuição Geométrica

Exemplo 3, continuação) A probabilidade de uma regulagem bem sucedida na montagem de um determinado tipo de produto é de 0.80. Assuma que as tentativas são independentes.

b) Qual é a probabilidade de que a primeira regulagem bem sucedida requeira no máximo quatro tentativas ?

```
nsamples <- 10000000
n.bem.sucedida <- 0; n.mal.sucedida <- rep(0, 100)
for (i in 1:nsamples) {
  regulagem <- rgeom(1, 0.8) # uso de rgeom com prob = 0.8 e pegando uma observação
  if (regulagem > 0) {
    n.mal.sucedida[regulagem] <- n.mal.sucedida[regulagem] + 1
  } else {
    n.bem.sucedida <- n.bem.sucedida + 1
  }
}
prob.no.maximo.quatro <- (n.bem.sucedida + sum(n.mal.sucedida[1:3])) / nsamples
print(prob.no.maximo.quatro)
```

Distribuição Geométrica

Exemplo 3, continuação) A probabilidade de uma regulagem bem sucedida na montagem de um determinado tipo de produto é de 0.80. Assuma que as tentativas são independentes.

c) Qual é a probabilidade de que a primeira regulagem bem sucedida requeira ao menos quatro tentativas ?

```
nsamples <- 10000000
n.bem.sucedida <- 0; n.mal.sucedida <- rep(0, 100)
for (i in 1:nsamples) {
  regulagem <- rgeom(1, 0.8) # uso de rgeom com prob = 0.8 e pegando uma observação
  if (regulagem > 0) {
    n.mal.sucedida[regulagem] <- n.mal.sucedida[regulagem] + 1
  } else {
    n.bem.sucedida <- n.bem.sucedida + 1
  }
}
prob.ao.menos.quatro <- (1 - (n.bem.sucedida+sum(n.mal.sucedida[1:2])) / nsamples)
print(prob.ao.menos.quatro)
```

Distribuição Hipergeométrica

É utilizada quando se consideram extrações aleatórias, sem reposição, de uma população dividida segundo dois atributos.

Considere uma população com N objetos nos quais M são classificados como do tipo A e $N - M$ são classificados como do tipo B. Seja X a variável aleatória que conta o número de objetos classificados como do tipo A em uma amostra de tamanho n. Então a distribuição de probabilidade de X é dada por:

$$P(X = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$$

Distribuição Hipergeométrica

Exemplo 1) Um banco de investimentos possui 20 clientes (N) com mais de $R\$\ 100\ milhões$ aplicados. Destes, 5 (M) moram fora do Brasil:

a) Ao retirar 3 (n) ao acaso, qual a probabilidade de que 2 (k) deles morem fora do Brasil?

$$P(X = 2) = \frac{\binom{5}{2}\binom{20-5}{3-2}}{\binom{20}{3}} = \frac{10\ x\ 15}{1140} = 0{,}1315 = 13{,}15\%$$

b) Ao retirar 5 ao acaso, qual a probabilidade de que todos morem no Brasil?

$$P(X = 0) = \frac{\binom{5}{0}\binom{20-5}{5}}{\binom{20}{5}} = 0{,}1937 = 19{,}37\%$$

$$P(X = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$$

Distribuições no R

No R, o trabalho com uma **variável aleatória** $\boldsymbol{X}$ está baseado em 4 procedimentos identificados pelas letras $\boldsymbol{p}$, $\boldsymbol{d}$, $\boldsymbol{q}$ e $\boldsymbol{r}$ que, adicionadas previamente ao código das **distribuições,** permitem o seu uso com **diferentes propósitos**:

- **rdistrib**: Gera **números aleatórios** baseados na distribuição definida. Necessita de argumentos especificando o tamanho da amostra, além dos parâmetros requeridos pela distribuição de interesse;
- **pdistrib**: Calcula a **função acumulada de probabilidade** $\boldsymbol{F(x)}$**;**
- **ddistrib**: Calcula a **função de probabilidade** $\boldsymbol{f(x)}$;
- **qdistrib**: Calcula o **quantil** correspondente a uma dada probabilidade.

Distribuição Hipergeométrica no R

Distribuição	Sintaxe	Opções
Hipergeométrica (**hyper**)	**dhyper**(x, m, n, k, opções)	**x**: número de elementos com a característica A extraídos sem reposição de uma urna que contém elementos com características A e B (pode ser um vetor). **m**: número de elementos com a característica A. **n**: número de elementos com a característica B. **k**: número de elementos extraídos da urna. **q**: vetor contendo os quantis. **p**: vetor contendo as probabilidades. **nn**: número de observações.
	phyper(q, m, n, k, opções)	
	qhyper(p, m, n, k, opções)	
	rhyper(nn, m, n, k)	

Distribuição Hipergeométrica

Exemplo 1, continuação) Um banco de investimentos possui 20 clientes com mais de $R\$\ 100\ milhões$ aplicados. Destes, 5 moram fora do Brasil:

a) Ao retirar 3 ao acaso, qual a probabilidade de que 2 deles morem fora do Brasil?

```
dhyper(2, 5, 15, 3)
0.1315789
```

b) Ao retirar 5 ao acaso, qual a probabilidade de que todos morem no Brasil?

```
dhyper(0, 5, 15, 5)
0.193692
```

Distribuição Hipergeométrica

Exemplo 1, continuação) Um banco de investimentos possui 20 clientes com mais de $R\$\ 100\ milhões$ aplicados. Destes, 5 moram fora do Brasil:

a) Ao retirar 3 ao acaso, qual a probabilidade de que 2 deles morem fora do Brasil?

```
nsamples <- 10000000
nobservacao <- 20
n.obs.fora <- 5
n.obs.Brasil <- 15
n.extraido <- 3
n.mora.fora <- rep(0, n.extraido)
n.mora.brasil <- 0
for (i in 1:nsamples) {
  cliente <- rhyper(1, n.obs.fora, n.obs.Brasil, n.extraido) # uma observação
  if (cliente > 0) { # 0 indica mora no Brasil
    n.mora.fora[cliente] <- n.mora.fora[cliente] + 1
  } else {
    n.mora.brasil <- n.mora.brasil + 1
  }
}
prob.dois.mora.fora <- n.mora.fora[2] / nsamples
print(prob.dois.mora.fora)
```

Distribuição Hipergeométrica

Exemplo 1, continuação) Um banco de investimentos possui 20 clientes com mais de $R\$\ 100\ milhões$ aplicados. Destes, 5 moram fora do Brasil:

b) Ao retirar 5 ao acaso, qual a probabilidade de que todos morem no Brasil?

```
nsamples <- 10000000
nobservacao <- 20
n.obs.fora <- 5
n.obs.Brasil <- 15
n.extraido <- 5
n.mora.fora <- rep(0, n.extraido)
n.mora.brasil <- 0
for (i in 1:nsamples) {
  cliente <- rhyper(1, n.obs.fora, n.obs.Brasil, n.extraido) # uma observação
  if (cliente > 0) { # 0 indica mora no Brasil
    n.mora.fora[cliente] <- n.mora.fora[cliente] + 1
  } else {
    n.mora.brasil <- n.mora.brasil + 1
  }
}
prob.todos.mora.Brasil <- n.mora.brasil / nsamples
print(prob.todos.mora.Brasil)
```

Distribuição Hipergeométrica

Exemplo 2) Uma urna contém 15 bolas, sendo 10 brancas e 5 vermelhas. Uma amostra de extensão 4, não ordenada e sem reposição, é retirada ao acaso da urna. Determine qual é a probabilidade de que três bolas sejam vermelhas.

```
dhyper (3, 5, 10, 4)
0.07326007
```

Distribuição Hipergeométrica

Exemplo 2, continuação) Uma urna contém 15 bolas, sendo 10 brancas e 5 vermelhas. Uma amostra de extensão 4, não ordenada e sem reposição, é retirada ao acaso da urna. Determine qual é a probabilidade de que três bolas sejam vermelhas.

```
nsamples <- 10000000
nobservacao <- 15
n.obs.branca <- 10
n.obs.vermelha <- 5
n.extraido <- 4
n.bola.vermelha <- rep(0, n.extraido)
n.bola.branca <- 0
for (i in 1:nsamples) {
  bola <- rhyper(1, n.obs.vermelha, n.obs.branca, n.extraido) # uma observação
  if (bola > 0) {
    n.bola.vermelha[bola] <- n.bola.vermelha[bola] + 1
  } else {
    n.bola.branca <- n.bola.branca + 1
  }
}
prob.tres.bola.vermelha <- n.bola.vermelha[3] / nsamples
print(prob.tres.bola.vermelha)
```

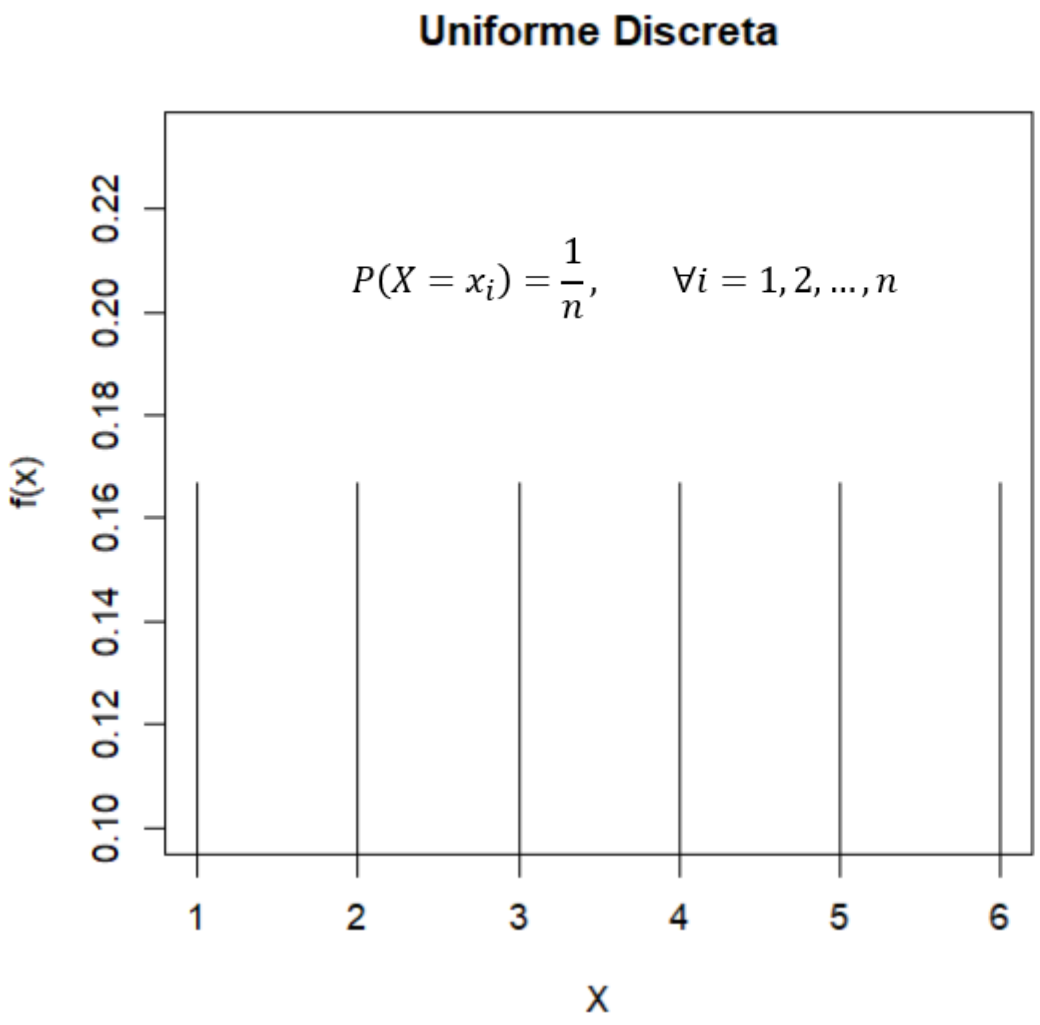
Distribuições no Python

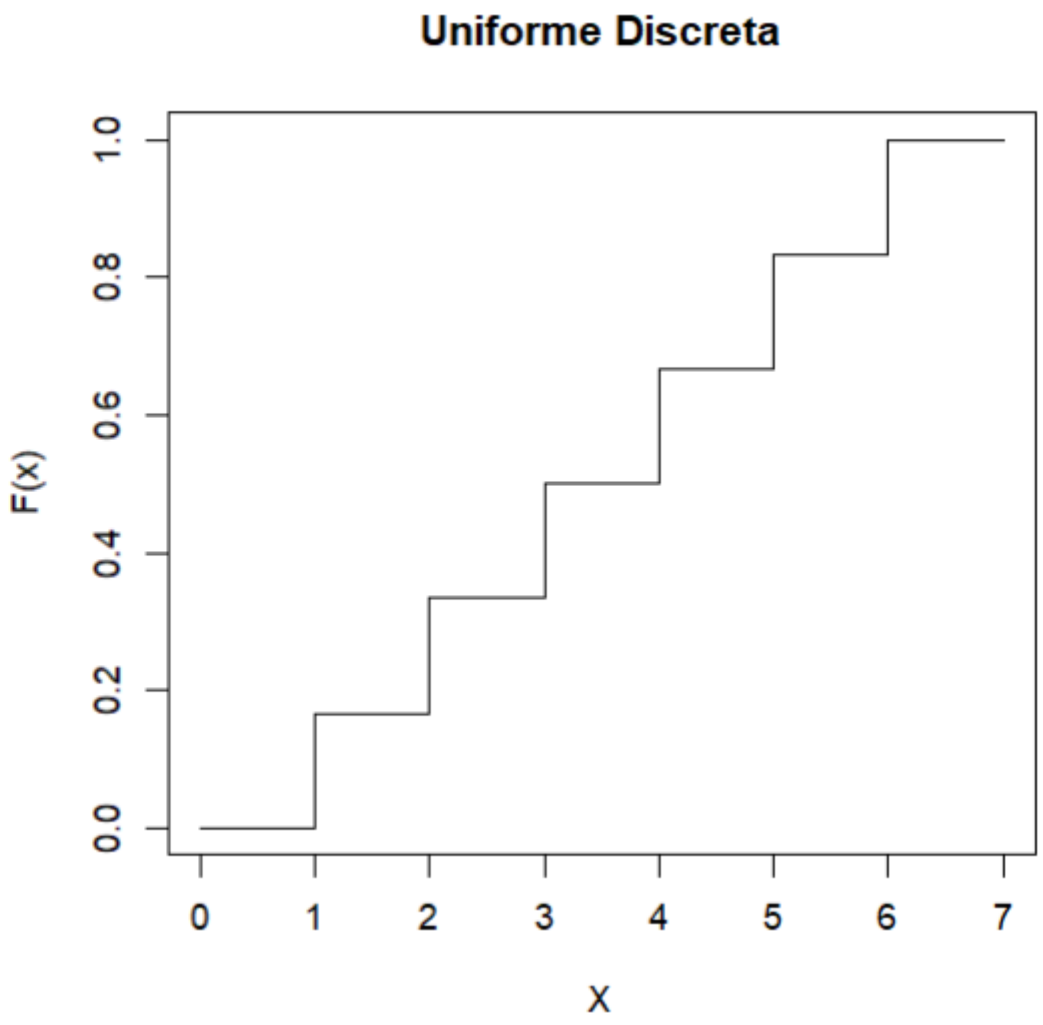
Utilizando o Python (módulo stats da biblioteca SciPy), refaça todos os exemplos deste material que foram resolvidos em R, inclusive os que foram resolvidos por simulação.

Utilize métodos como rvs, pdf, cdf, ppf e stats relacionados às distribuições.

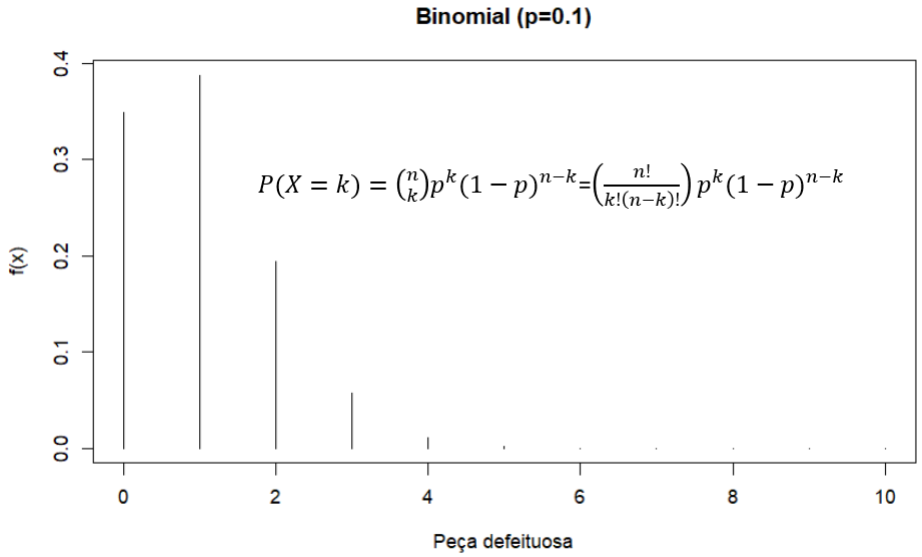
- **rvs**: Gera **números aleatórios** baseados na distribuição definida;
- **cdf**: Calcula a **função acumulada de probabilidade** $\boldsymbol{F(x)}$**;**
- **pdf**: Calcula a **função de probabilidade** $\boldsymbol{f(x)}$;
- **ppf**: Percent Point Function (inverse of cdf — percentiles).
- **stats**: Return mean, variance, skew, and/or kurtosis.

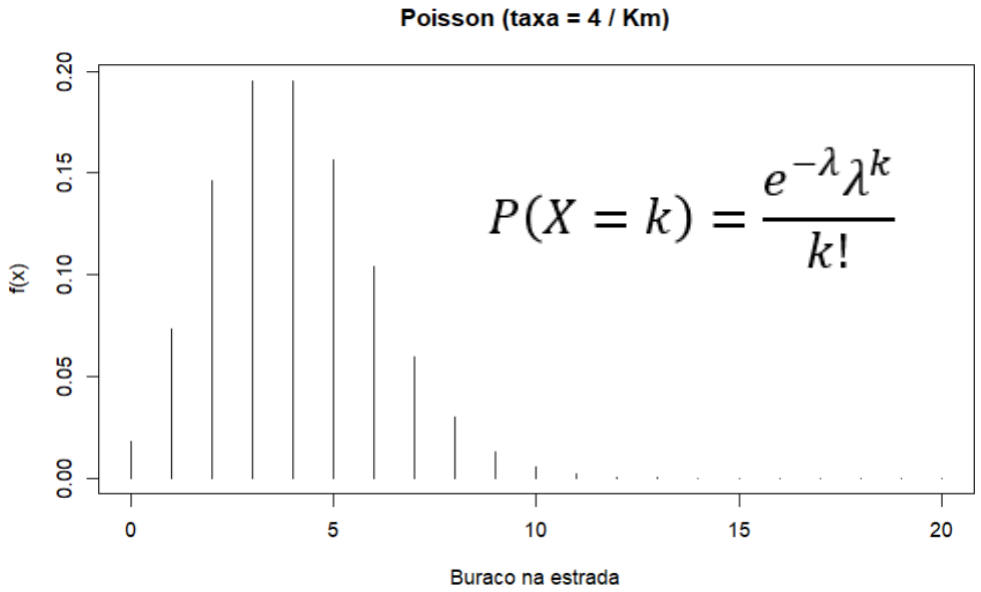
Distribuição Uniforme Discreta

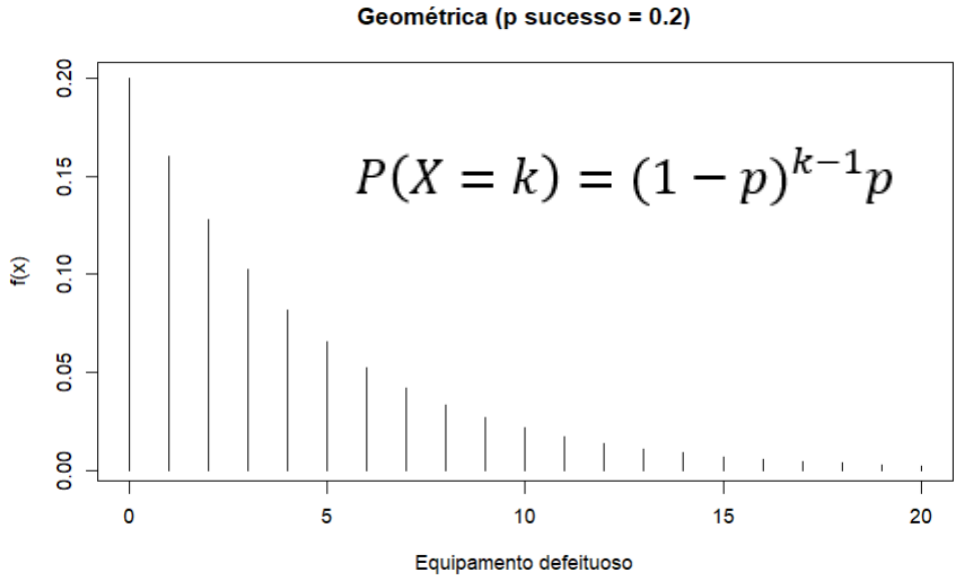


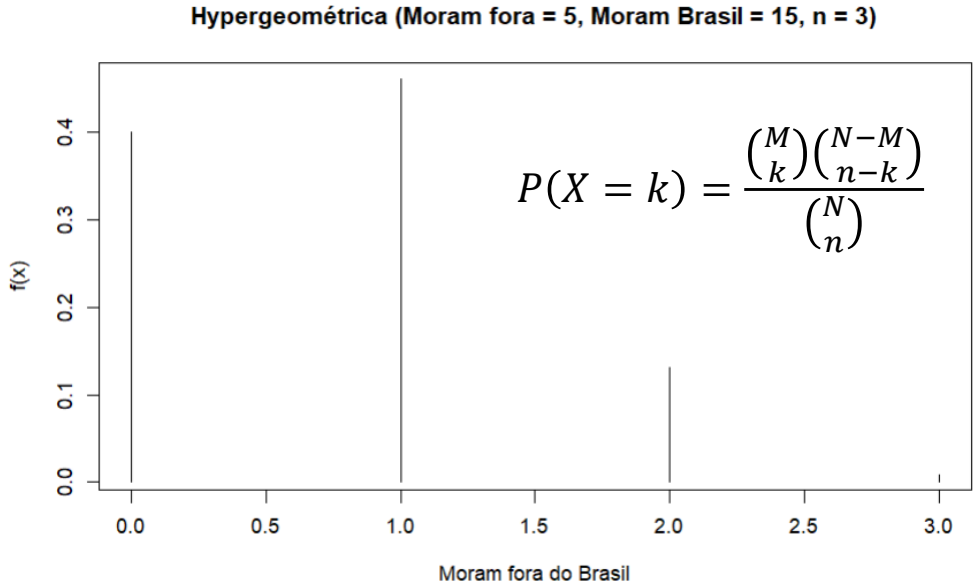


Distribuições Discretas









Valor Esperado

O **valor esperado** de uma variável aleatória representa o valor médio "esperado" de uma experiência se ela for repetida diversas vezes.

$$\mu = E(X) = \sum_{i=1}^{n} x_i P(X = x_i)$$

Propriedades:

$$E(k) = k, se\ k\ \text{é}\ uma\ constante.$$

$$E(kX) = kE(X), se\ k\ \text{é}\ uma\ constante.$$

$$E(X + Y) = E(X) + E(Y)$$

$$E(X - Y) = E(X) - E(Y)$$

$$E(X.Y) = E(X).E(Y), se\ X\ e\ Y\ \text{são}\ indenpendentes.$$

Variância

A **variância** de uma variável aleatória é uma medida da sua dispersão, indicando "o quão longe" em geral os seus valores se encontram do valor esperado.

$$\sigma^2 = Var(X) = \sum_{i=1}^{n} (x_i - \mu)^2 P(X = x_i) = E(X^2) - [E(X)]^2 \, , sendo:$$

$$E(X) = \sum_{i=1}^{n} x_i P(X = x_i)$$

$$E(X^2) = \sum_{i=1}^{n} x_i^2 P(X = x_i)$$

Variância

Exemplo 1) Suponha que X seja uma variável aleatória discreta com a seguinte distribuição: $P(X = -1) = 0{,}2$; $P(X = 0) = 0{,}5$; e $P(X = 2) = 0{,}3$. Calcule o valor esperado e a variância de X.

$$E(X) = \sum_{i=1}^{n} x_i P(X = x_i) = -1 * 0{,}2 + 0 * 0{,}5 + 2 * 0{,}3 = 0{,}4$$

$$E(X^2) = \sum_{i=1}^{n} x_i^2 P(X = x_i) = (-1)^2 * 0.2 + (0)^2 * 0{,}5 + (2)^2 * 0{,}3 = 1{,}4$$

$$Var(X) = \sum_{i=1}^{n} (x_i - \mu)^2 P(X = x_i) = E(X^2) - [E(X)]^2 = 1{,}4 - (0{,}4)^2 = 1{,}24$$