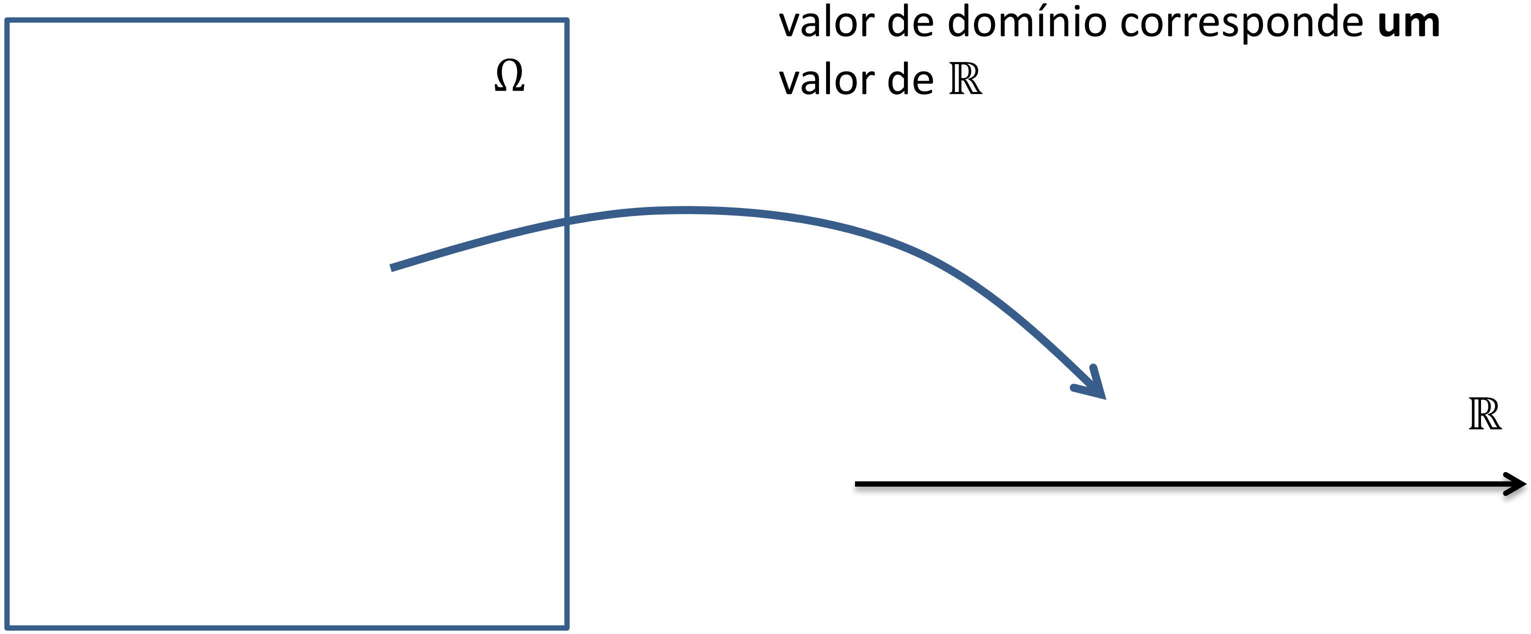
Aula 5. Variáveis Aleatórias Discretas

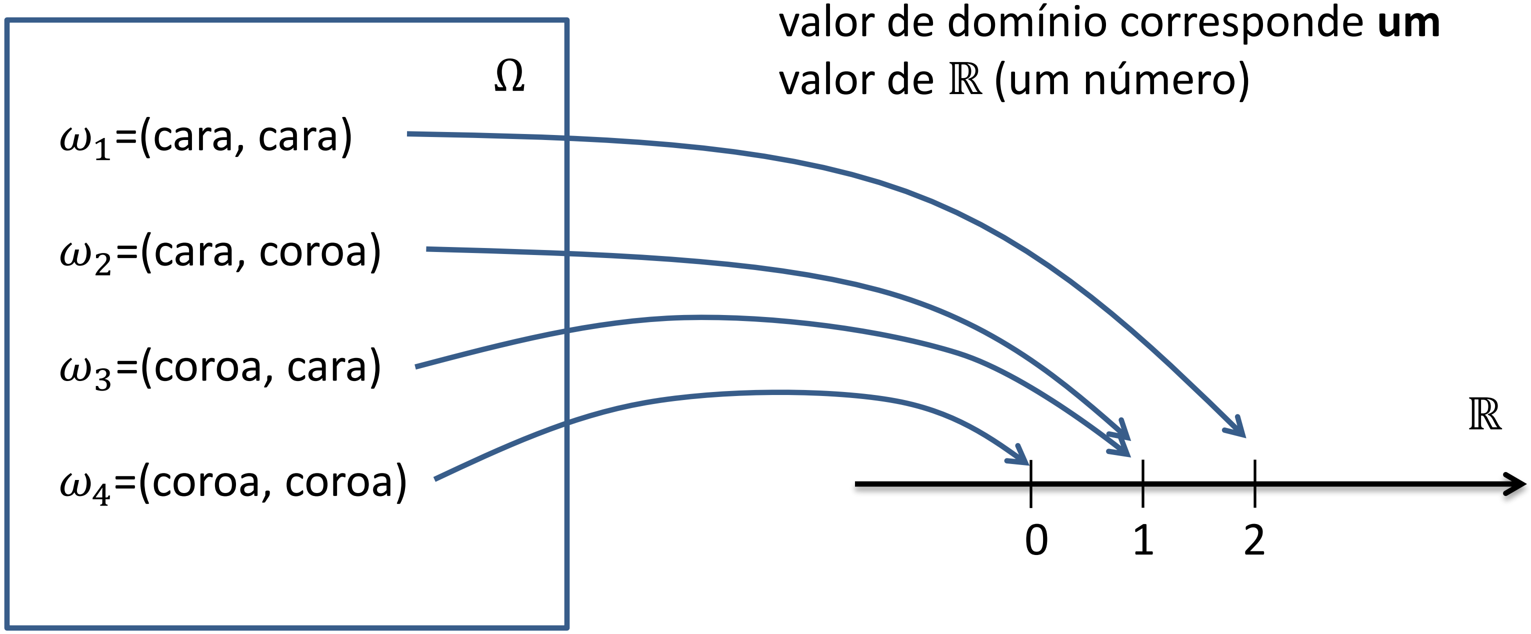
Definição "formal": Variável aleatória é qualquer *função* definida em espaço Ω.

função é uma regra que para cada valor de domínio corresponde **um** valor de $\mathbb{R}$



Definição "formal": Variável aleatória é qualquer *função* definida em espaço Ω.

função é uma regra que para cada valor de domínio corresponde **um** valor de $\mathbb{R}$ (um número)



Variável aleatória X é número de "caras" em experimento de duas jogadas de uma moeda

Variável Aleatória

Uma variável aleatória pode ser classificada em:

- Variável aleatória **discreta**
- Variável aleatória **contínua**

Definição "formal": Variável aleatória é qualquer *função* definida em espaço Ω.

Exemplos de variáveis aleatórias discretas:

- número de acidentes no marginal Tietê amanha;
- número de alunos que vão passar no vestibular entre 200 inscritos;
- número de pessoas que visitam um determinado site, num certo período de tempo;
- número de futuros inadimplentes entre as pessoas que pegaram o crédito neste mês;
- número de pontos que o preço de PETR sobe/desce no fechamento do mercado em comparação de preço na abertura

Variável Aleatória

- Variável aleatória **discreta**

Uma v.a. é discreta quando o conjunto de valores possíveis que ela assume for finito ou infinito enumerável.

Exemplo:

Observa-se o sexo (característica) das crianças em famílias com três filhos (M: masculino e F: feminino).

Espaço amostral:

$$\Omega = \{\underset{\omega_1}{(MMM)}, \underset{\omega_2}{(MMF)}, \underset{\omega_3}{(MFM)}, \underset{\omega_4}{(FMM)}, \underset{\omega_5}{(MFF)}, \underset{\omega_6}{(FMF)}, \underset{\omega_7}{(FFM)}, \underset{\omega_8}{(FFF)}\}$$

Defina X: nº de crianças do sexo masculino (M).

Ω	MMM	MMF	MFM	FMM	MFF	FMF	FFM	FFF
X	3	2	2	2	1	1	1	0

$\rightarrow$ Então X assume valores no conjunto {0, 1, 2, 3}, logo é uma variável aleatória discreta.

Exemplo:

No mesmo experimento...

Espaço amostral:

$$\Omega = \{(\underset{\omega_1}{MMM}), (\underset{\omega_2}{MMF}), (\underset{\omega_3}{MFM}), (\underset{\omega_4}{FMM}), (\underset{\omega_5}{MFF}), (\underset{\omega_6}{FMF}), (\underset{\omega_7}{FFM}), (\underset{\omega_8}{FFF})\}$$

Podemos definir agora

Y: nº de crianças do sexo feminino (F).

Ω	MMM	MMF	MFM	FMM	MFF	FMF	FFM	FFF
Y	0	1	1	1	2	2	2	3

→ Então *Y* também assume valores no conjunto {0, 1, 2, 3}, porém, para outros valores de Ω.

Variável Aleatória

- Variável aleatória **contínua**

Uma v.a. é contínua quando o conjunto de valores possíveis que ela assume for não enumerável.

Exemplo:

Observa-se o tempo de vida, em horas, de lâmpadas produzidas por uma fábrica.

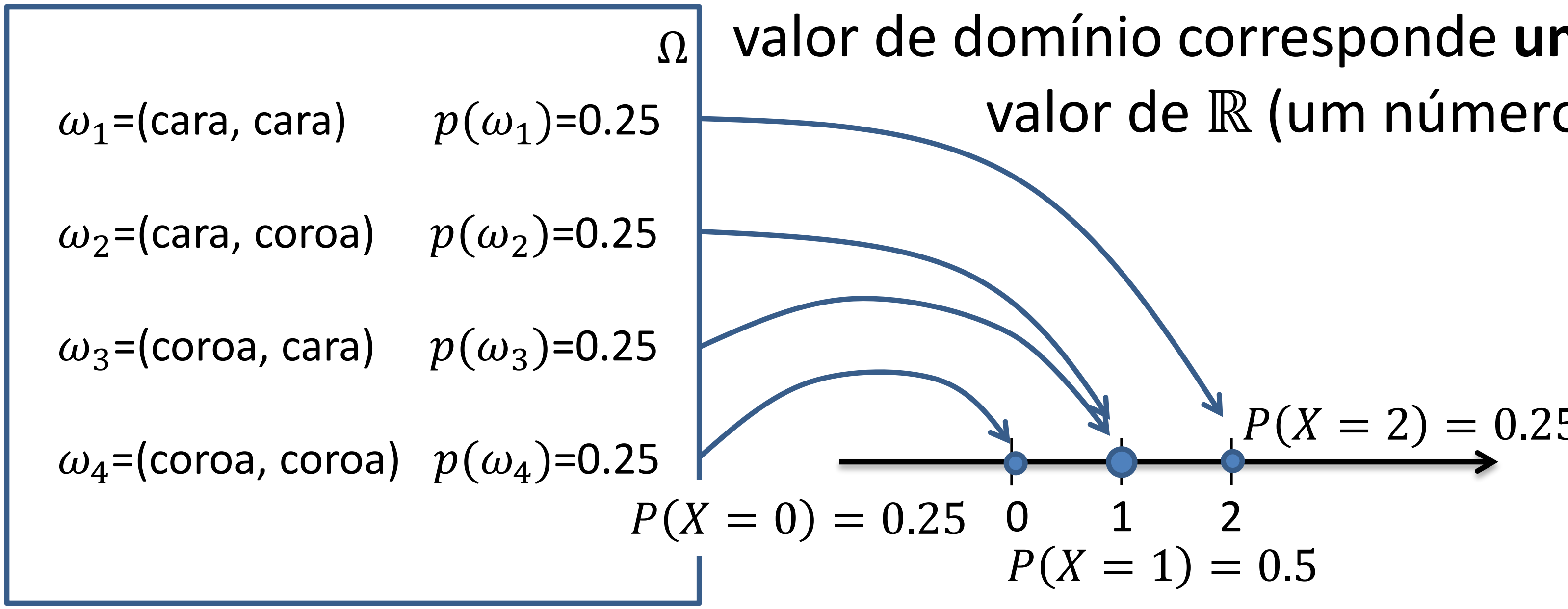
Defina ***T***: tempo de vida, em horas, da lâmpada escolhida, ao acaso, da fábrica.

$\rightarrow$ Então, ***T*** é uma **variável aleatória contínua** que assume qualquer valor real não negativo.

Distribuição de variável aleatória discreta.

Junto com elementos ω_i de espaço Ω vamos "transferir" para $\mathbb{R}$ a probabilidade (ou peso) de cada elemento $p(\omega_i)$

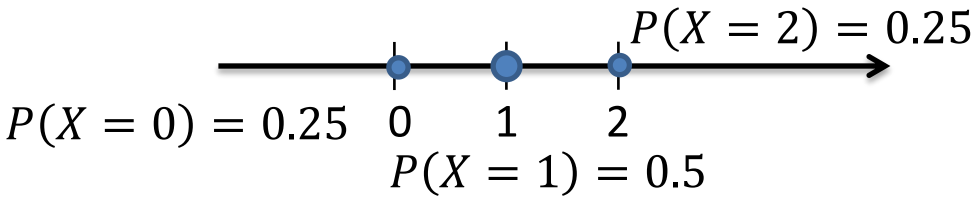
função é uma regra que para cada valor de domínio corresponde **um** valor de $\mathbb{R}$ (um número)



Variável aleatória X é número de "caras" em experimento de duas jogadas de uma moeda

Distribuição de variável aleatória discreta.

A variável aleatória, ou o resultado da "transferência"



podemos representar como a tabela seguinte

$\boldsymbol{X}$	**0**	**1**	**2**
$\boldsymbol{P}$	0.25	0.5	0.25

Obs: Toda a informação sobre a variável aleatória esta contida nesta tabela; não precisamos mais de lembrar do espaço Ω.

Obs: 0.25 + 0.5 + 0.25 = 1

Distribuição de variável aleatória discreta.

Variável aleatória discreta e a sua distribuição podem ser definidas pela sua tabela

$\boldsymbol{X}$	$\boldsymbol{x_1}$	$\boldsymbol{x_2}$	$\boldsymbol{\ldots}$	$\boldsymbol{x_n}$
$\boldsymbol{P}$	p_1	p_2	$\ldots$	p_n

onde todos os números x_i são diferentes e as probabilidades p_i de correspondentes valores satisfazem seguintes propriedades:

- $p_i \geq 0$
- $p_1 + p_2 + \cdots + p_n = 1$

Distribuição de variável aleatória discreta.

Variável aleatória X é número que sai em um experimento de jogada de um dado

$\boldsymbol{X}$	**1**	**2**	**3**	**4**	**5**	**6**
$\boldsymbol{P}$	1/6	1/6	1/6	1/6	1/6	1/6

Variável aleatória X é soma dos números que saem em um experimento de jogada de dois dados

$\boldsymbol{X}$	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
$\boldsymbol{P}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Distribuição de variável aleatória discreta.

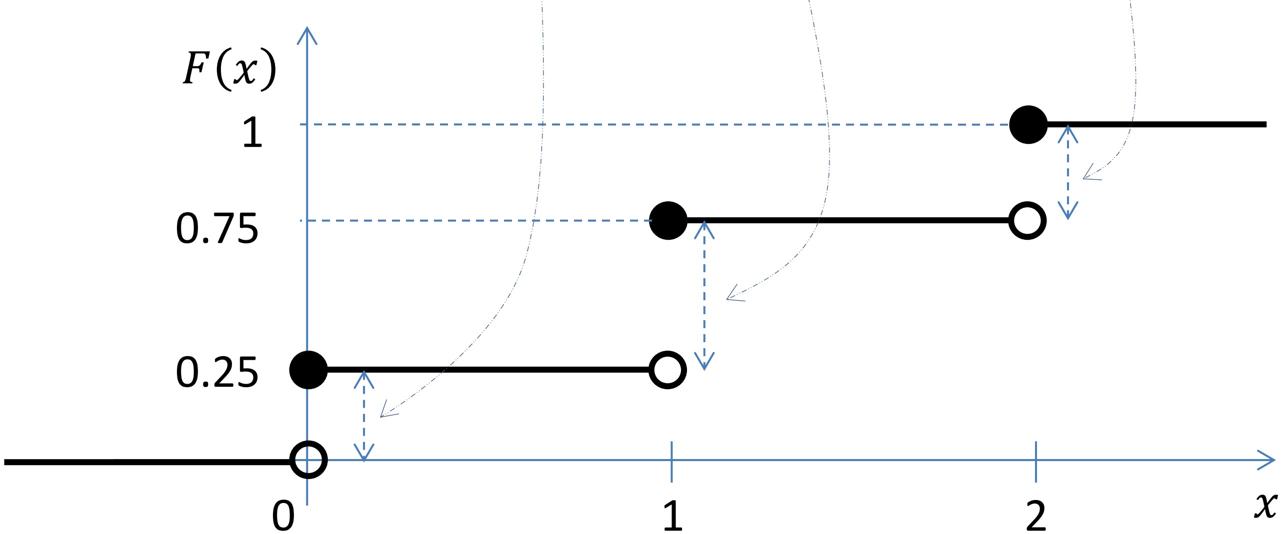
Outro jeito de apresentar uma variável aleatória discreta é função de distribuição cumulativa $F(x)$, ou, as vezes denotamos como $F_X(x)$ para destacar que uma função de variavel aleatoria X. Pela definição

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

Por exemplo, consideramos v.a. X dada pela tabela

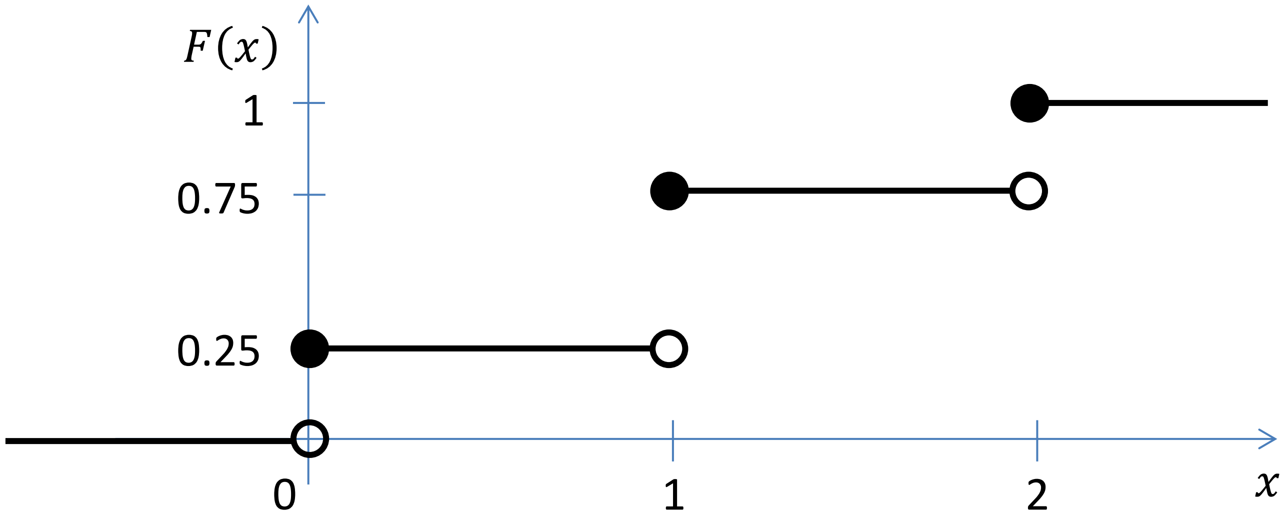
$\boldsymbol{X}$	**0**	**1**	**2**
$\boldsymbol{P}$	0.25	0.5	0.25

Desenhamos gráfico de $F(x)$:



Distribuição de variável aleatória discreta.

X	**0**	**1**	**2**
P	0.25	0.5	0.25



Distribuição de variável aleatória discreta. Exemplos.

Distribuição Bernoulli.

Supomos um simples modelo de alteração de preço de uma ação. Seja s_1 o preço no instante “agora”. No próximo instante (um tick, próxima negociação, próximo dia etc.) o preço aumentou com probabilidade p ou diminuiu em um ponto com probabilidade $q = 1 - p$. Se o evento “preço aumentou” vou codificar como “1” e o evento “preço diminuiu” como “0”, então tenho uma variável Bernoulli

$\boldsymbol{X}$	**0**	**1**
$\boldsymbol{P}$	q	p

Caso quero a distribuição de incremento do preço posso considerar

$\boldsymbol{X}$	**-1**	**1**
$\boldsymbol{P}$	q	p

Situações com alternativas *dicotômicas* podem ser representadas, genericamente, por respostas do tipo sucesso-fracasso.

Esses experimentos recebem o nome de ***Ensaios de Bernoulli*** e originam uma v.a. com *distribuição de Bernoulli.*

Variável aleatória de Bernoulli: É uma v.a. que assume apenas dois valores:

- ***1*** se ocorrer ***sucesso***,
- ***0*** se ocorrer ***fracasso***.

Geralmente, a probabilidade de sucesso é representada por ***p***, **0 < *p* < 1**.

"X ~ B(p)" indica uma v.a. X tem a ***distribuição de Bernoulli*** com parâmetro ***p***, isto é,

$$X = \begin{cases} 1, \text{ se ocorrer "sucesso"} \\ 0, \text{ se ocorrer "fracasso"} \end{cases}$$

e sua função de probabilidade pode ser representada pela tabela

X	1	0
$P(X=x)$	p	$1 - p$

Distribuição de variável aleatória discreta. Valor esperado, a média da variável.

Como em caso de medidas de posição gostaria de achar algum tipo de “centro” de uma variável aleatória. “Centro” é um número que resuma toda v.a. (*eu espero a temperatura de amanha “em média” 19 graus*). “Centro” deveria possuir um valor que não pode ser menor de valor minimal e maior de um valor maximal de v.a. Vejamos a tabela de uma v.a.

$\boldsymbol{X}$	$\boldsymbol{x_1}$	$\boldsymbol{x_2}$	**…**	$\boldsymbol{x_n}$
$\boldsymbol{P}$	p_1	p_2	…	p_n

ou mehor, vejamos a distribuição de v.a. Bernoulli

$\boldsymbol{X}$	**0**	**1**
$\boldsymbol{P}$	q	p

ou mehor, vejamos um caso particular da distribuição de v.a. Bernoulli

$\boldsymbol{X}$	**0**	**1**
$\boldsymbol{P}$	0.1	0.9

Como resumir essa variável em um número?

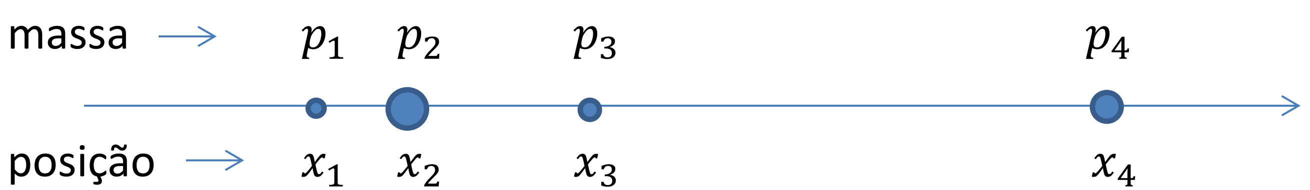
Distribuição de variável aleatória discreta. Valor esperado, a média da variável.

$\boldsymbol{X}$	**0**	**1**
$\boldsymbol{P}$	0.1	0.9

vou representar essa distribuição em um sistema físico:

em que se-reduz esse sistema ?

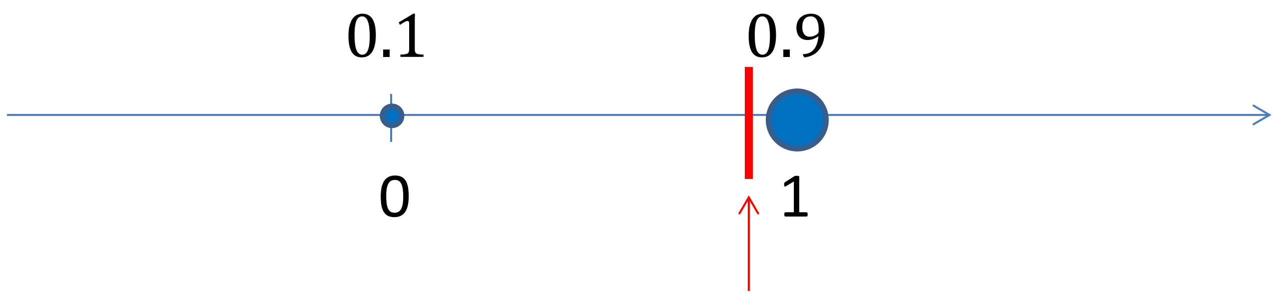
os físicos reduzem qualquer sistema físico em um ponto – ponto material, um ponto que representa a posição do sistema (medida de posição), esse ponto calcula-se como **centro de massas**



centro de massas $= \sum_i x_i p_i$ = valor esperado

Distribuição de variável aleatória discreta. Valor esperado, a média da variável.

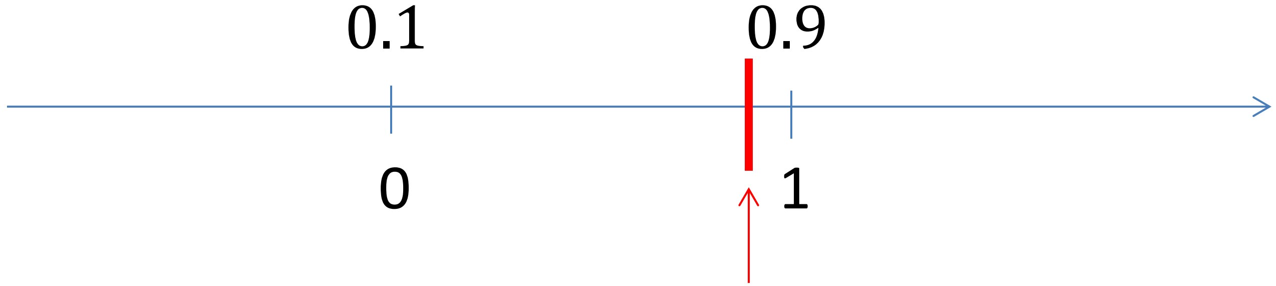
X	**0**	**1**
P	0.1	0.9



valor esperado $= 0 \cdot 0.1 + 1 \cdot 0.9 = 0.9$

Distribuição de variável aleatória discreta. Valor esperado, a média da variável.

X	**0**	**1**
P	0.1	0.9



valor esperado $= 0 \cdot 0.1 + 1 \cdot 0.9 = 0.9$

Distribuição de variável aleatória discreta. Valor esperado, a média da variável.

$\boldsymbol{X}$	$\boldsymbol{x_1}$	$\boldsymbol{x_2}$	**…**	$\boldsymbol{x_n}$
$\boldsymbol{P}$	p_1	p_2	…	p_n

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

Distribuição de variável aleatória discreta. Valor esperado, a média da variável.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i \frac{1}{n}$$

supondo que todos os valores observados x_i são diferentes podemos representar os valores observados como os valores de uma v.a. com esses valores $x_i\ i = 1, \dots, n$ e com probabilidades $p_i = \frac{1}{n}, \quad i = 1, \dots, n$ iguais.

usa-se na simulação *bootstrap*, previsão dos valores futuros baseando-se em valores históricos.

Distribuição de variável aleatória discreta. Valor de dispersão, variância.

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2\frac{1}{n} = E(X - E(X))^2$$

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

$$DP(X) = \sqrt{Var(X)}$$

OBS: A volatilidade pode ser definida como uma medida de dispersão

Distribuição de variável aleatória discreta. Valor de dispersão, variância.

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

$$= E(X^2) - \big(E(X)\big)^2$$

$$= \sum_{i=1}^{n} x_i^2 p_i - \left(\sum_{i=1}^{n} x_i p_i\right)^2$$

Variância: É o valor esperado da v.a. $(X - E(X))^2$, ou seja, se X assume os valores $x_1, x_2, \ldots, x_n$, então

$$\mathrm{Var}(X) = \sum_{i=1}^{n} [x_i - E(X)]^2 \times P(X = x_i)$$

Notação: $\sigma^2 = \mathrm{Var}(X)$.

Da relação acima, segue que

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2.$$

Desvio Padrão: É definido como a raiz quadrada positiva da variância, isto é,

$$\mathrm{DP}(X) = \sqrt{\mathrm{Var}(X)}.$$

Notação: $\sigma = \mathrm{DP}(X)$.

Distribuição de variável aleatória discreta. Exemplos.

voltaremos para primeiro exemplo de v.a. - Bernoulli

$\boldsymbol{X}$	**0**	**1**
$\boldsymbol{P}$	q	p

esperança $E(X) = 0 \cdot q + 1 \cdot p = p$

variância $$\mathrm{Var}(X) = E(X^2) - \big(E(X)\big)^2 = E(X) - \big(E(X)\big)^2$$
$$= E(X)(1 - E(X)) = p(1-p) = pq$$

$\boldsymbol{Y}$	**-1**	**1**
$\boldsymbol{P}$	q	p

notamos que $Y = 2X - 1$

esperança $E(Y) = 2E(X) - 1 = 2p - 1$

variância $\mathrm{Var}(Y) = \mathrm{Var}(2X - 1) = \mathrm{Var}(2X) = 4\mathrm{Var}(X) = 4pq$

Distribuição de variável aleatória discreta. Propriedades Esperança e Variância.

$E(X + a) = E(X) + a$
$E(aX) = aE(X)$
$E(a) = a$

$Var(X + a) = Var(X)$
$Var(aX) = a^2 Var(X)$
$Var(a) = 0$

Para duas v.a. quaisquer X, Y
$E(X + Y) = E(X) + E(Y)$

Para duas v.a. quaisquer X, Y
e independentes
$Var(X + Y) = Var(X) + Var(Y)$

Observação: Seja $Y = f(X)$
em geral $E(Y) \neq f(E(X))$, mas isso é verdade, caso f é uma função linear

Distribuição de variável aleatória discreta. Exemplos.

Distribuição Binomial

$\rightarrow$ Repetições independentes de um ensaio de Bernoulli, com a mesma probabilidade de ocorrência de "sucesso", dão origem ao ***modelo de probabilidade binomial***.

consideramos caso quando o preço em cada "tick" aumenta ou diminua em um ponto independentemente de valor de preço. Quantos aumentos teremos depois de n ticks consecutivos? em média?

número de aumento é v.a. que tem o nome: Binomial. Veremos porque

denotamos número de aumentos S_n que é a soma de aumentos consecutivos: $S_n = X_1 + \cdots + X_n$, em que $X_i \sim B(p)$

$\boldsymbol{X_i}$	**0**	**1**
$\boldsymbol{P}$	q	p

$$P(S_n = k) = \binom{n}{k} p^k q^{n-k}, \qquad k = 0,1, \dots, n$$

Modelo Binomial

Exemplo: Um dado equilibrado é lançado 3 vezes.
Qual é a probabilidade de se obter a *face 5* duas vezes?

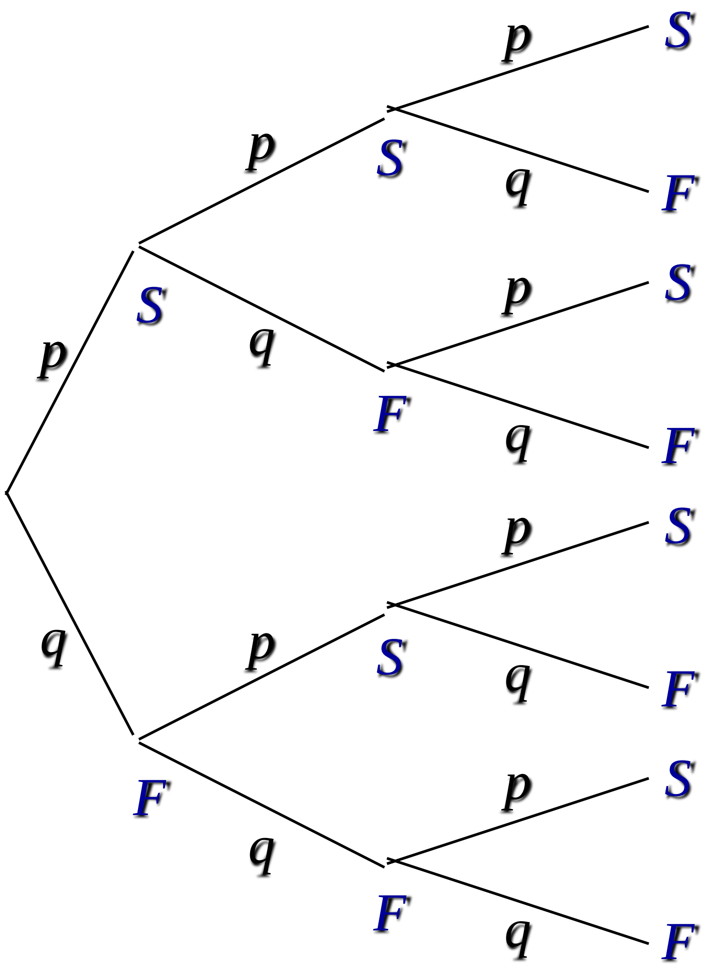
Denotamos,

S: "sucesso", ocorrer face 5;
F: "fracasso", não ocorrer face 5.

É fácil ver que $p = P(\text{sucesso}) = 1/6$ e
$$q = 1 - p = P(\text{fracasso}) = 5/6$$

$$\Omega = \{SSS, SSF, SFS, FSS, SFF, FSF, FFS, FFF\}$$

Estamos interessados no número total de sucessos que, no caso, é o _número de vezes que a face 5 é observada_ nos 3 lançamentos do dado.



Ω	Prob	X
(SSS)	p^3	3
(SSF)	p^2q	2
(SFS)	p^2q	2
(SFF)	pq^2	1
(FSS)	p^2q	2
(FSF)	pq^2	1
(FFS)	pq^2	1
(FFF)	q^3	0

A função de probabilidade de X é dada por:

Probabilidades binomiais para n = 3 e $P(S) = p$		
nº. de sucessos	**probabilidades**	**p = 1/6**
0	q^3	125/216=0,5787
1	$3pq^2$	75/216=0,3472
2	$3p^2q$	15/216=0,0694
3	p^3	1/216=0,0046

$\Rightarrow$

Podemos escrever essa função como

$$P(X = k) = \binom{3}{k} p^k \, q^{3-k} \quad , \quad k = 0, 1, 2, 3.$$

No exemplo, para n = 3 e p = 1/6, $P\,(X = 2) = 0{,}0694.$

Distribuição binomial:

A v.a. X correspondente ao **número de sucessos em n ensaios de Bernoulli independentes e com mesma probabilidade p de sucesso** tem *distribuição binomial com parâmetros n e p.*

Sua função de probabilidade é dada por

$$\boxed{P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \quad , \quad k = 0, 1, \ldots, n.}$$

Notação: $X \sim B(n; p)$.

Resultado: Se $X \sim B(n; p)$, então

$$\boxed{\begin{array}{l} \text{média: } \mu = \mathrm{E}(X) = np \\ \text{variância: } \sigma^2 = \mathrm{Var}(X) = np(1-p) = npq \end{array}}$$

Exemplo utilizando o R:

Considere uma prova com 12 questões, cada uma com 4 alternativas. Suponha que o aluno escolha as respostas ao acaso. Qual é a probabilidade de que ele ***acerte pelo menos 6 questões***?

X: nº de questões que o aluno acertará

X pode assumir valores no conjunto {0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 ou 12}.

$X \sim B(12; 0{,}25)$

$$\boxed{P(X = x) = \binom{12}{x} 0{,}25^{x} (1 - 0{,}25)^{12-x}}$$

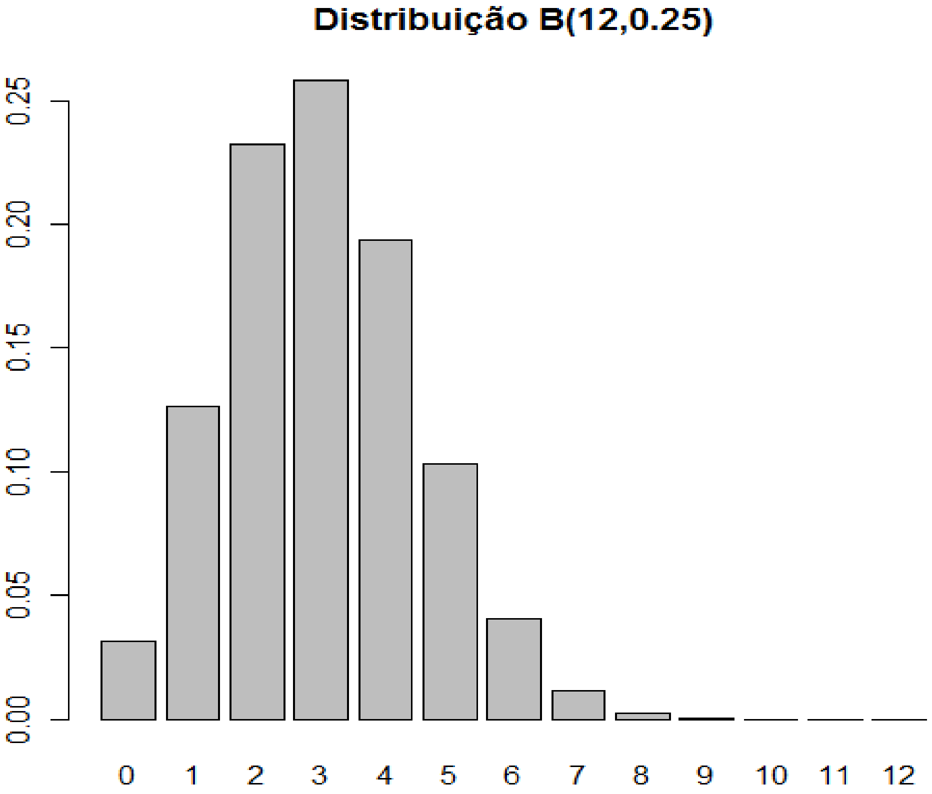
$\Downarrow$

Uso do R ou Excel para os cálculos!

No R, probabilidades

```
> dbinom(0:12,12,0.25)
 [1] 3.167635e-02 1.267054e-01 2.322932e-01 2.581036e-01 1.935777e-01
 [6] 1.032414e-01 4.014945e-02 1.147127e-02 2.389848e-03 3.540516e-04
[11] 3.540516e-05 2.145767e-06 5.960464e-08
```

```
> cbind(0:12,dbinom(0:12,12,0.25))
      [,1]         [,2]
 [1,]    0 3.167635e-02
 [2,]    1 1.267054e-01
 [3,]    2 2.322932e-01
 [4,]    3 2.581036e-01
 [5,]    4 1.935777e-01
 [6,]    5 1.032414e-01
 [7,]    6 4.014945e-02
 [8,]    7 1.147127e-02
 [9,]    8 2.389848e-03
[10,]    9 3.540516e-04
[11,]   10 3.540516e-05
[12,]   11 2.145767e-06
[13,]   12 5.960464e-08
```

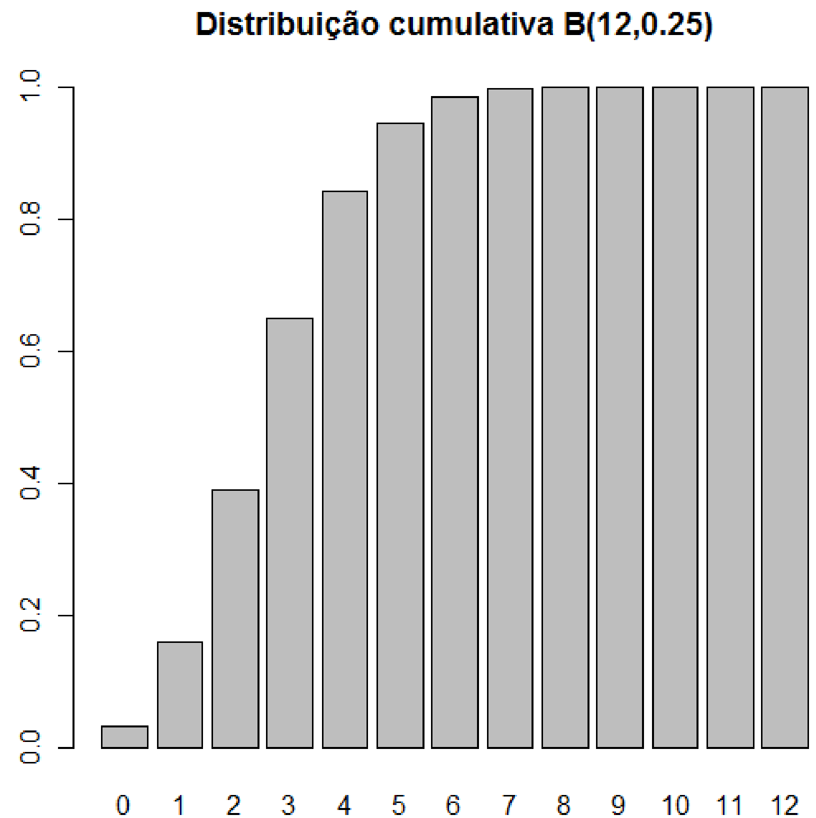



```
> barplot(dbinom(0:12,12,0.25),names.arg=0:12,main="Distribuição B(12,0.25)")
```

No R, distribuição cumulativa $P(X \leq x)$

```
> pbinom(0:12,12,0.25)
 [1] 0.03167635 0.15838176 0.39067501 0.64877862 0.84235632 0.94559777
 [7] 0.98574722 0.99721849 0.99960834 0.99996239 0.99999779 0.99999994
[13] 1.00000000
```

```
> cbind(0:12,pbinom(0:12,12,0.25))
      [,1]       [,2]
 [1,]    0 0.03167635
 [2,]    1 0.15838176
 [3,]    2 0.39067501
 [4,]    3 0.64877862
 [5,]    4 0.84235632
 [6,]    5 0.94559777
 [7,]    6 0.98574722
 [8,]    7 0.99721849
 [9,]    8 0.99960834
[10,]    9 0.99996239
[11,]   10 0.99999779
[12,]   11 0.99999994
[13,]   12 1.00000000
```




```
> barplot(pbinom(0:12,12,0.25),names.arg=0:12,main="Distribuição cumulativa B(12,0.25)")
```

No R, calcularemos $P(X \geq 6)$

```
> 1-pbinom(5,12,0.25)
[1] 0.05440223
```

$1 - P(X \leq 5) = P(X \geq 6)$

```
> pbinom(5,12,0.25,lower.tail=FALSE)
[1] 0.05440223
```

calcula a probabilidade $P(X > 5) = P(X \geq 6)$

A média é

$E(X) = n \times p = 12 \times 0{,}25 = 3$,

ou seja, *em média*, o aluno que responder ao acaso todas as questões acertará 3.

Distribuição de variável aleatória discreta. Exemplos.

Distribuição Geométrica

$X \sim Geom(p)$ ou $X \sim G(p)$

X é número de insaios de Benoulli ate o primeiro "sucesso", que ocorre com a probabilidade p:

$$P(X = k) = (1-p)^{k-1}p, \qquad k = 0,1,\ldots$$

É um exemplo de v.a. com o número infinito de valores

Distribuição de variável aleatória discreta. Exemplos.

Distribuição de Poisson

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

É um outro exemplo de v.a. com o número infinito de valores

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

$$E(X) = Var(X) = \lambda$$

Na prática, muitas situações nas quais interessa o número de observações de uma variável em um intervalo contínuo (tempo ou espaço) podem ser convenientemente explicadas pela distribuição de Poisson. Exemplos:

– chamadas telefônicas por minuto,

– mensagens que chegam a um servidor por segundo

– acidentes por dia, número de terremotos com certa magnitude

– defeitos por m^2, etc...