Naïve Bayes Classifier

Gustavo Teodoro Laureano
gustavo@inf.ufg.br

Programa de Pós-Graduação em Ciência da Computação
Instituto de Informática – Universidade Federal de Goiás

Outline

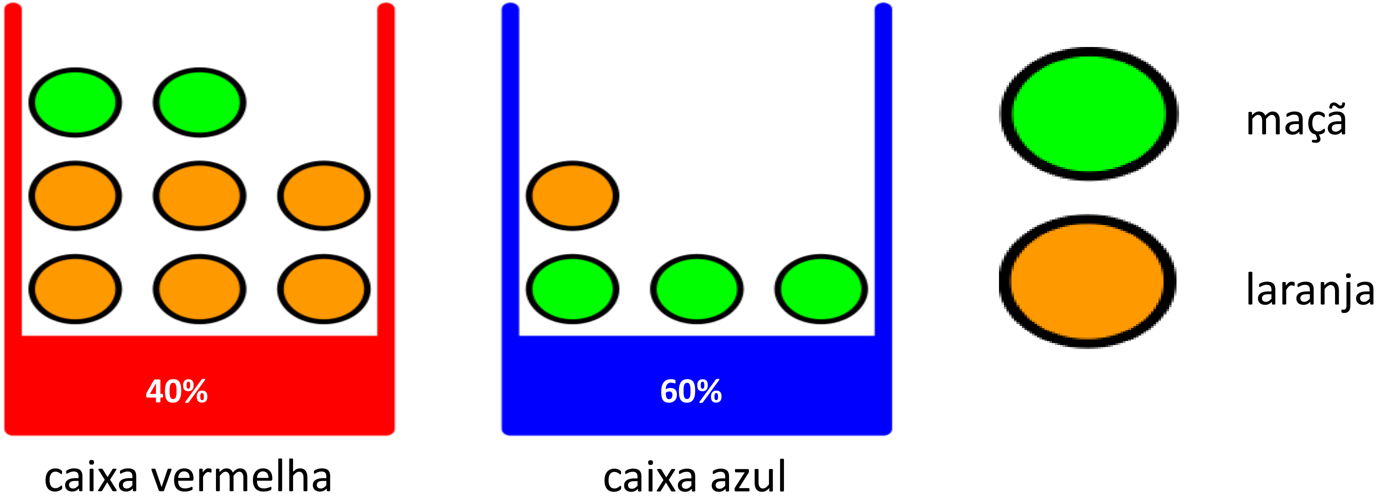
- Topics
  - Introduction
  - Fundamentals of Statistics
  - The Bayes Theorem
  -

Introduction

- Conceito chave para a modelagem de sistemas reais é a **incerteza**.
    - Devido ao **ruído** presente nas medições
    - Conjunto de **dados é de tamanho finito**.

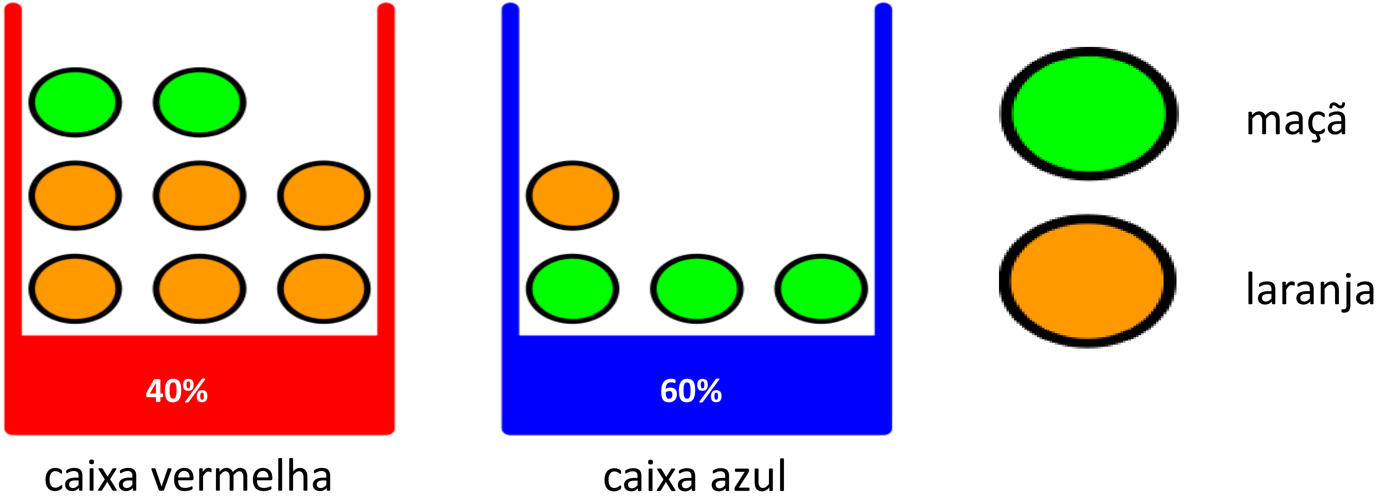
- Teoria da Probabilidade
  - Fornece meios para a **medição** e **manipulação** de **incertezas**.
  - Útil para:
    - Modelagem de sistemas reais
    - Estimação
    - Classificação
    - Reconhecimento de padrões
    - Aprendizagem de máquina

- Imagine o seguinte sistema:



  - 40% de chances de escolhermos a caixa vermelha.
  - 60% de chances de escolhermos a caixa azul.

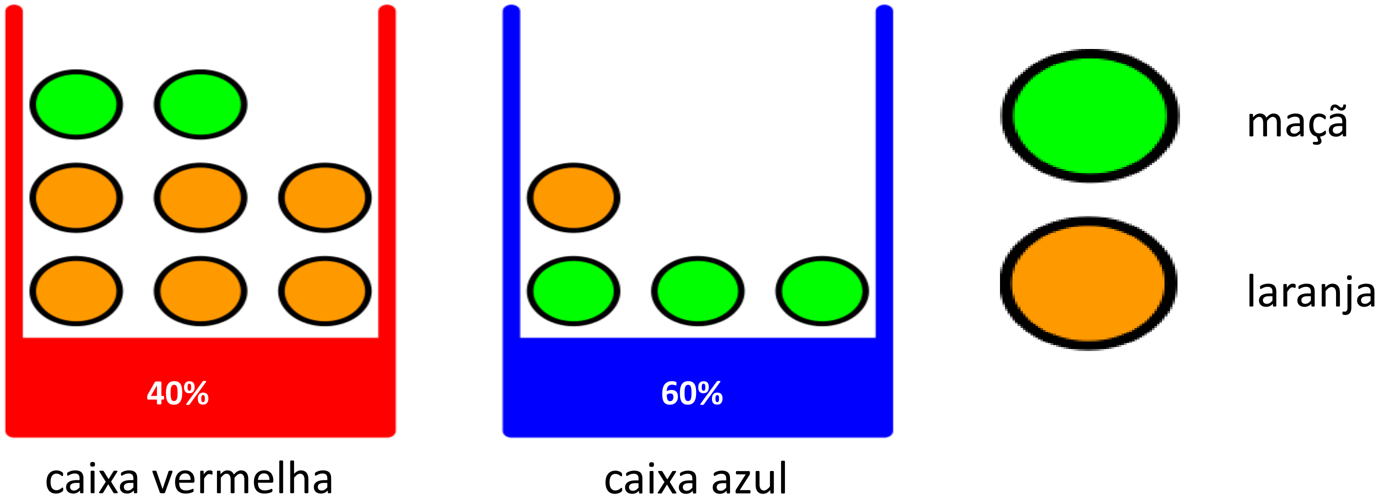
- Imagine o seguinte sistema:



  - Temos 2 variáveis aleatórias:
    - C: caixa
    - F: fruta

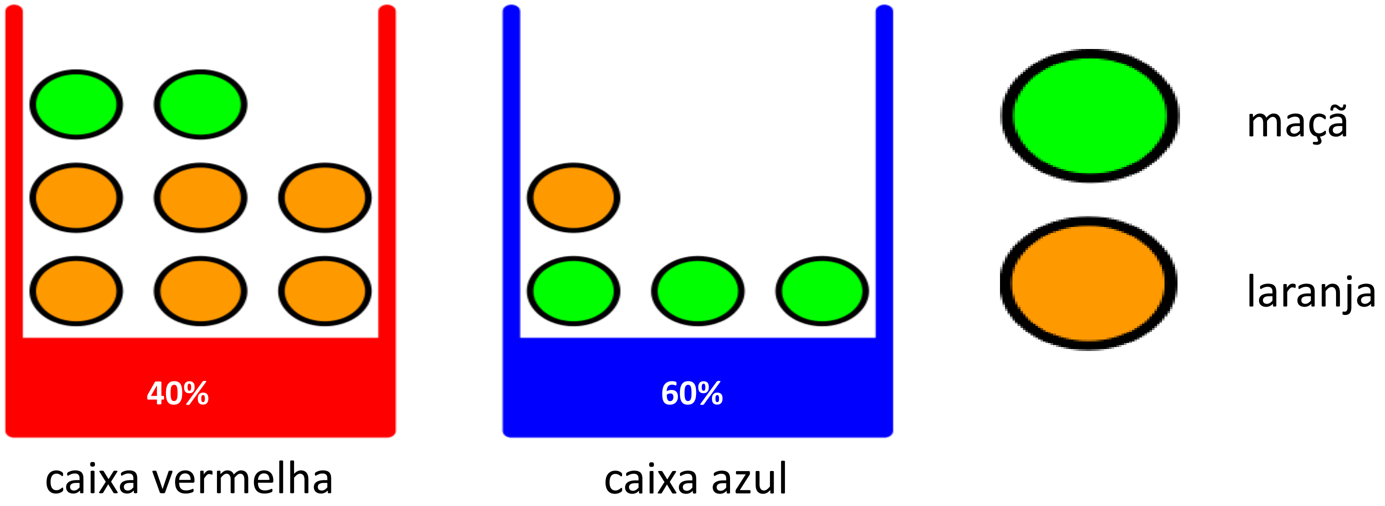
Introduction

- Imagine o seguinte sistema:



  - Temos 2 variáveis aleatórias:
    - C: caixa → 2 possibilidades: C = a e C = v
    - F: fruta → 2 possibilidades: F = l e F = m

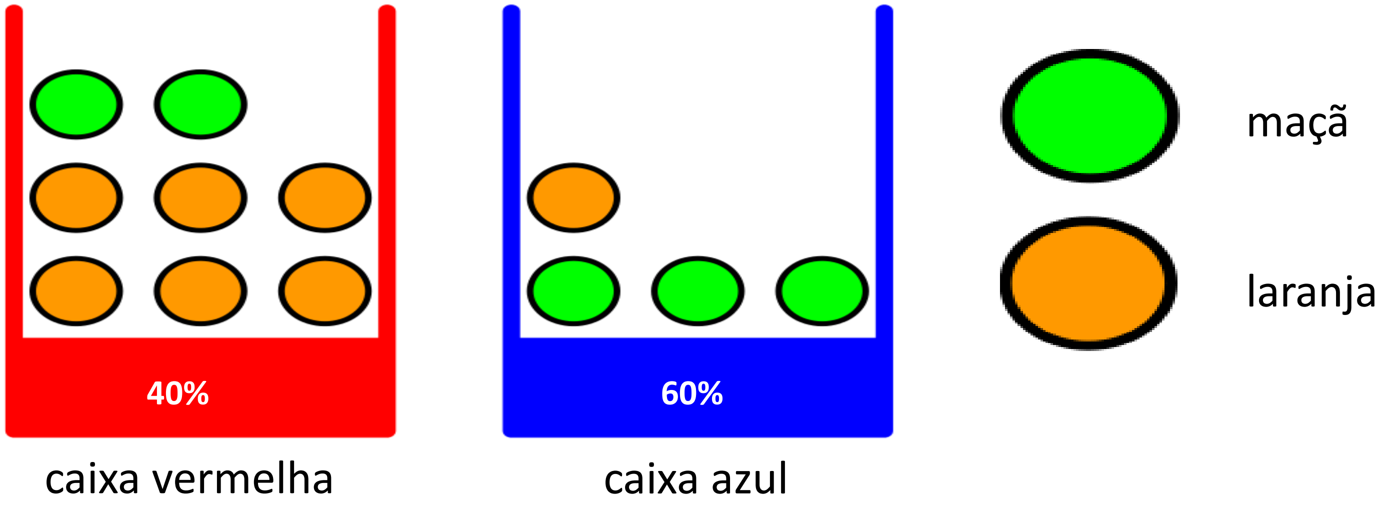
- Imagine o seguinte sistema:



  - Probabilidade de escolher a caixa vermelha: $p(C=v)=\frac{4}{10}=0.4$
  - Probabilidade de escolher a caixa azul: $p(C=a)=\frac{6}{10}=0.6$

Introduction

- Imagine o seguinte sistema:



  - *“Qual a probabilidade de escolhermos uma maçã?”*
  - *“Dado que pegamos uma laranja, qual a probabilidade dela ter vindo da caixa azul?”*

- Variáveis aleatórias:
  - Variável cujo valor depende de fatores aleatórios.

    Exemplo:
    - No lançamento de um dado, qual é a variável aleatória?
    - Em uma cobrança de pênalt, qual é a variável aleatória?
  - Variáveis discretas:
  - Variáveis contínuas:
- Probabilidade:

- Distribuição de probabilidades:
  - Para variáveis discretas (Função de Distribuição de Probabilidades):

    $F(x)$

  - Variáveis contínuas (Função Densidade de Probabilidade):

    $f(x)$

- Esperânça, Espectativa ou Média:
  - Caso disctreto:

$$\mu = \frac{1}{M}\sum_{i=1}^{M} x_i \qquad\qquad E(X) = \sum_{x} xP(X = x)$$

  - Caso contínuo:

$$E(X) = \mu = \int_{-\infty}^{+\infty} xf(x)$$

- Variância
  - Medida de dispersão em relação à média:

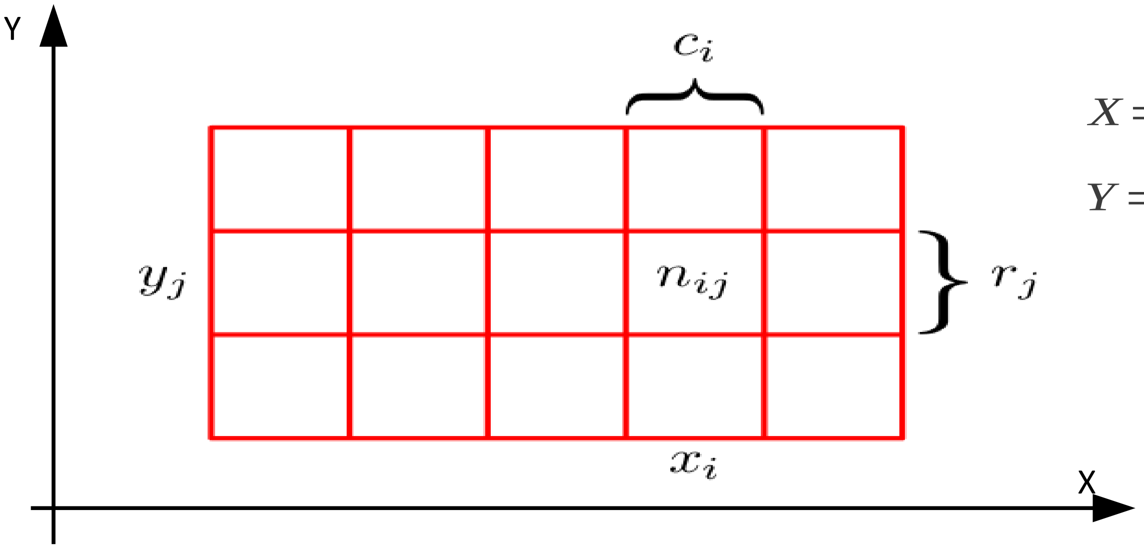
$$V(X) = E\{[X - E(X)]^2\}$$

$$\sigma^2 = \frac{1}{M}\sum_{i=1}^{M}(x_i - \mu)^2$$

- Desvio Padrão:

$$\sigma = \sqrt{\sigma^2}$$

- Considerando um sistema mais genérico:

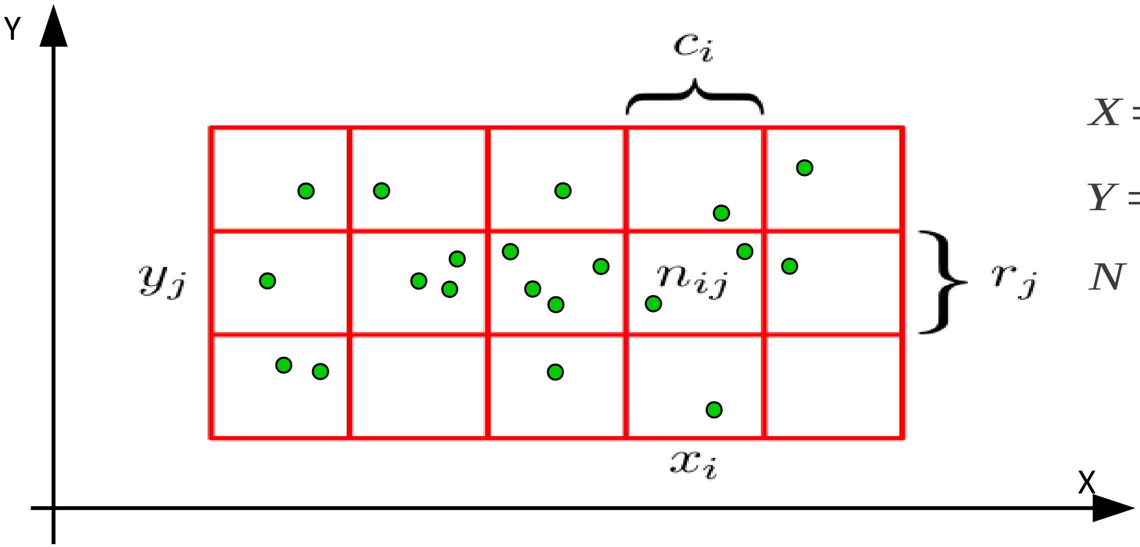


$$X = x_i, i = 1 \ldots M$$

$$Y = y_j, j = 1 \ldots L$$

  - Temos 2 variáveis aleatórias X e Y.

- Considerando um sistema mais genérico:



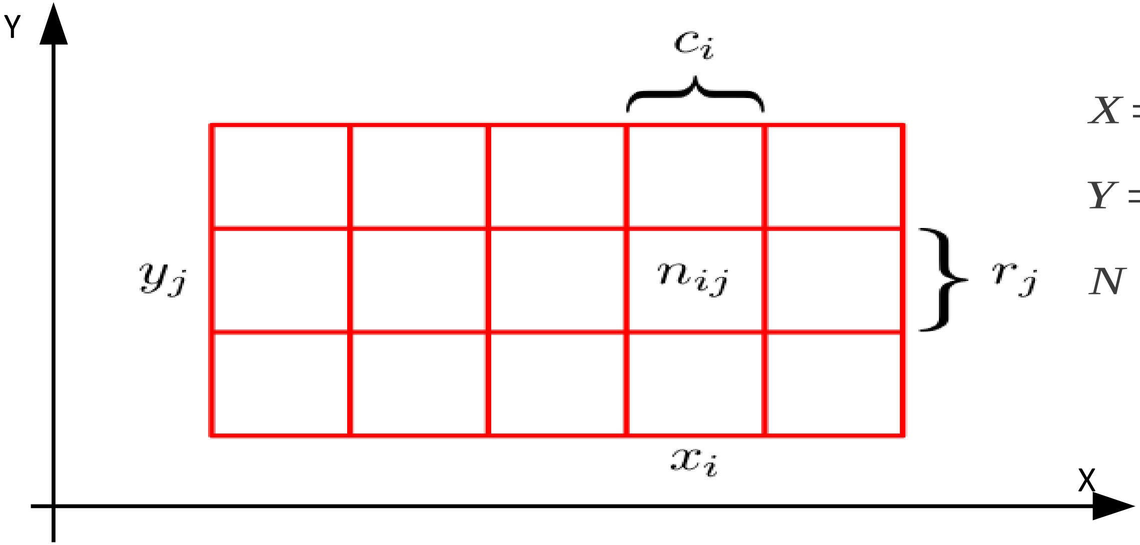
$X = x_i , i = 1 \ldots M$

$Y = y_j , j = 1 \ldots L$

N - quantidade de escolhas/tentativas/amostras

- Temos 2 variáveis aleatórias X e Y.

- Considerando um sistema mais genérico:



$X = x_i, i = 1 \ldots M$

$Y = y_j, j = 1 \ldots L$

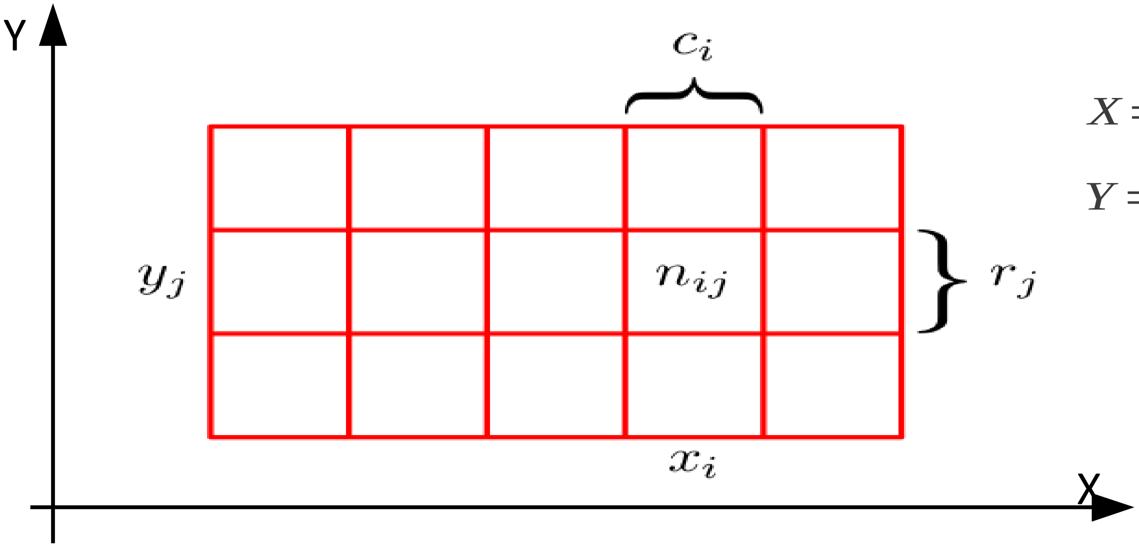
N - quantidade de escolhas/tentativas

n_{ij} - quantidade de vezes que $X = x_i$ e $Y = y_j$ são escolhidos juntos.

c_i - quantidade de vezes que $X = x_i$ independentemente da escolha de Y

r_j - quantidade de vezes que $Y = y_j$ independentemente da escolha de X

- Considerando um sistema mais genérico:

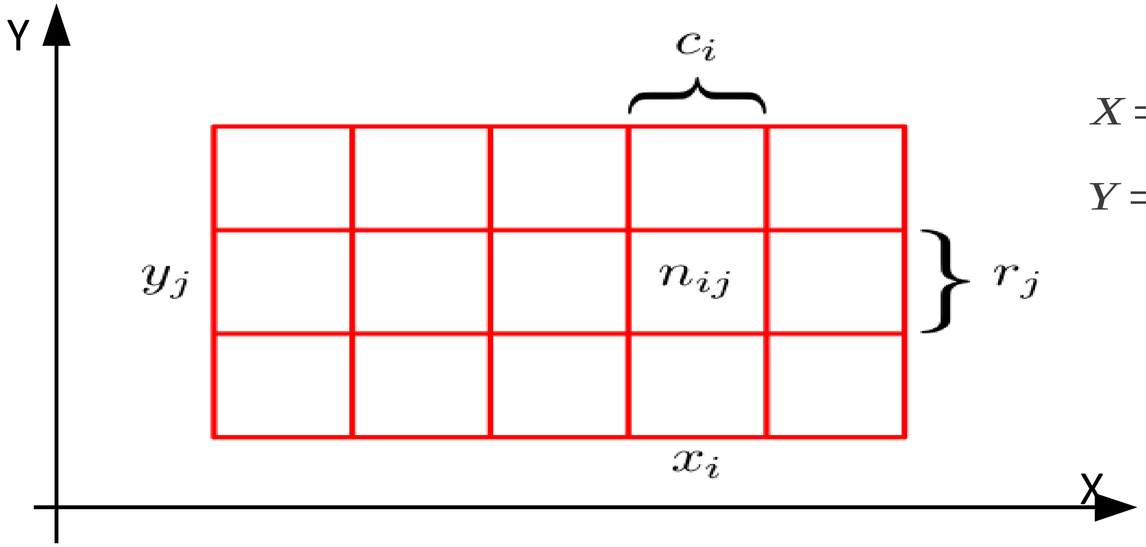


$$X = x_i, i = 1 \dots M$$

$$Y = y_j, j = 1 \dots L$$

- A probabilidade de escolhermos $X = x_i$ e $Y = y_j$ é: $p(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$
- A probabilidade de escolhermos $X = x_i$ é: $p(X = x_i) = \frac{c_i}{N}$
- A probabilidade de escolhermos $Y = y_j$ é: $p(Y = y_j) = \frac{r_i}{N}$

- Considerando um sistema mais genérico:



$X = x_i, i = 1 \ldots M$

$Y = y_j, j = 1 \ldots L$

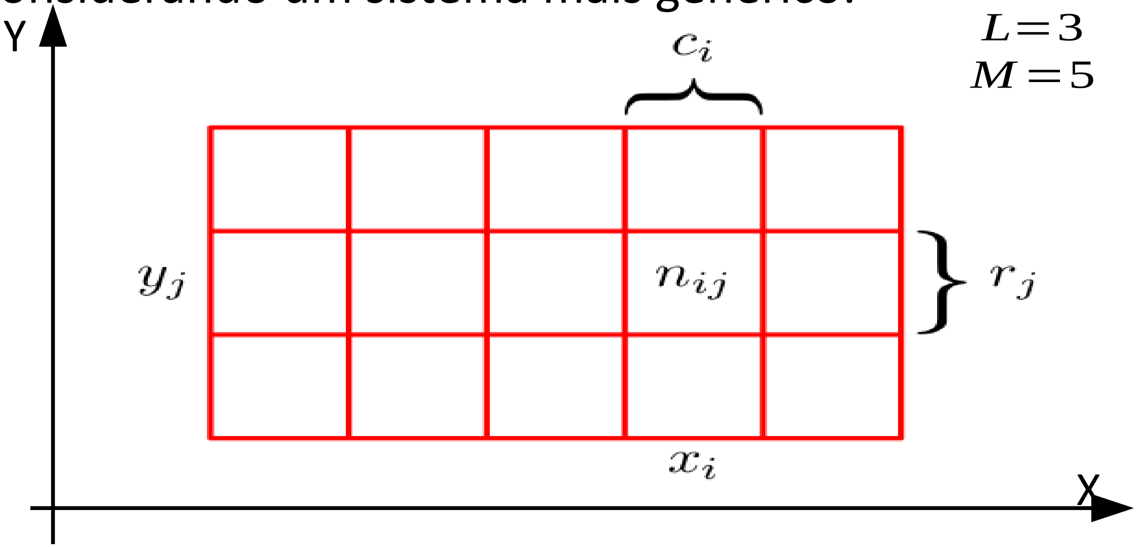
Probabilidade Conjunta

- A probabilidade de escolhermos $X = x_i$ e $Y = y_j$ é: $p(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$
- A probabilidade de escolhermos $X = x_i$ é: $p(X = x_i) = \frac{c_i}{N}$
- A probabilidade de escolhermos $Y = y_j$ é: $p(Y = y_j) = \frac{r_i}{N}$

- Considerando um sistema mais genérico:

$$X = x_i, i = 1 \dots M$$

$$Y = y_j, j = 1 \dots L$$



$$L = 3$$
$$M = 5$$

$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$$

$$p(X = x_i) = \frac{c_i}{N}$$

$$p(Y = y_j) = \frac{r_i}{N}$$

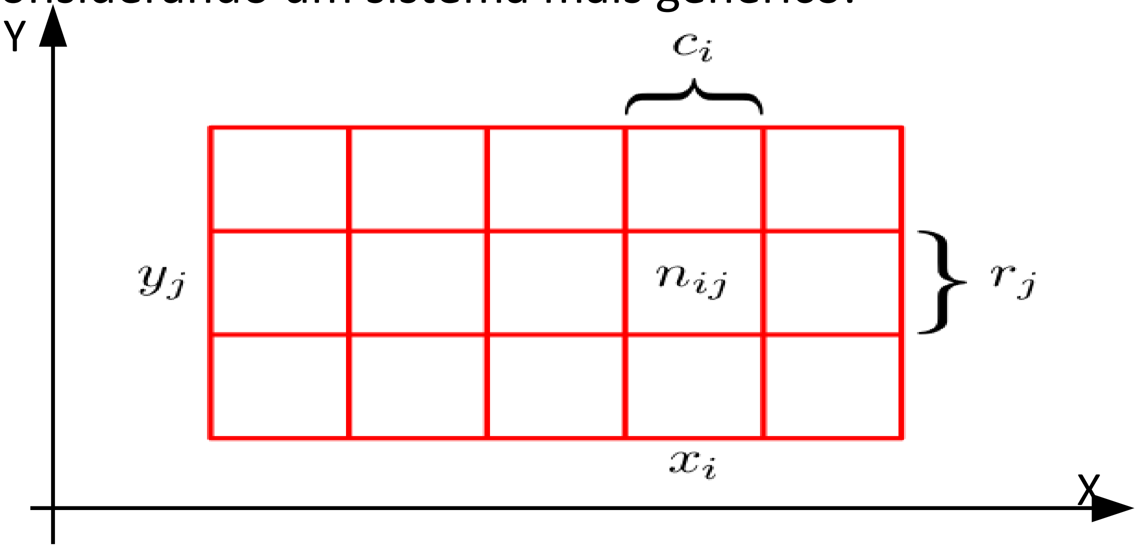
$$c_i = \sum_j^L n_{ij} \qquad r_j = \sum_i^M n_{ij}$$

- Visualmente podemos verificar que:

$$X = x_i, i = 1 \ldots M$$

$$Y = y_j, j = 1 \ldots L$$

- Considerando um sistema mais genérico:



$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$$

$$p(X = x_i) = \frac{c_i}{N}$$

$$p(Y = y_j) = \frac{r_i}{N}$$

$$c_i = \sum_j^L n_{ij} \qquad r_j = \sum_i^M n_{ij}$$

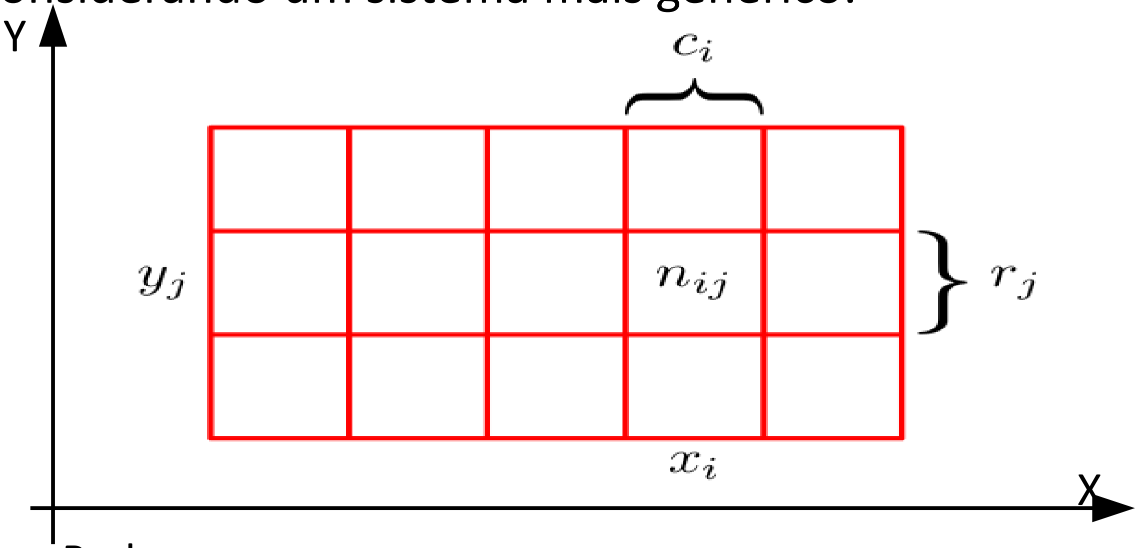
$$p(X = x_i) = \frac{c_i}{N} \longrightarrow p(X = x_i) = \frac{1}{N}\sum_j^L n_{ij} \longrightarrow p(X = x_i) = \sum_j^L p(X = x_i, Y = y_j)$$

- Podemos reescrever:

$$X = x_i, i = 1 \dots M$$

$$Y = y_j, j = 1 \dots L$$

- Considerando um sistema mais genérico:



$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$$

$$p(X = x_i) = \frac{c_i}{N}$$

$$p(Y = y_j) = \frac{r_i}{N}$$

$$c_i = \sum_{j}^{L} n_{ij} \qquad r_j = \sum_{i}^{M} n_{ij}$$

  - Podemos reescrever:

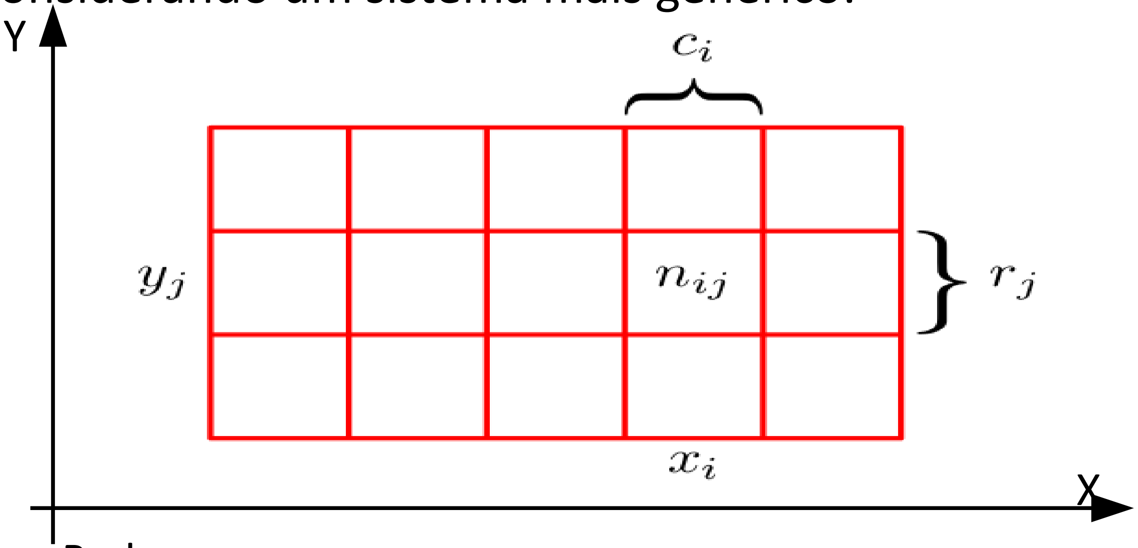
$$p(X = x_i) = \frac{c_i}{N} \longrightarrow p(X = x_i) = \frac{1}{N} \sum_{j}^{L} n_{ij} \longrightarrow p(X = x_i) = \sum_{j}^{L} p(X = x_i, Y = y_j)$$

$$p(Y = y_j) = \frac{r_j}{N} \longrightarrow p(Y = y_j) = \sum_{i}^{M} p(X = x_i, Y = y_j)$$

- Considerando um sistema mais genérico:

$$X = x_i, i = 1 \ldots M$$

$$Y = y_j, j = 1 \ldots L$$



$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$$

$$p(X = x_i) = \frac{c_i}{N}$$

$$p(Y = y_j) = \frac{r_i}{N}$$

$$c_i = \sum_j^L n_{ij} \qquad r_j = \sum_i^M n_{ij}$$

- Podemos reescrever:

$$\left.\begin{array}{l} p(X = x_i) = \dfrac{c_i}{N} \\[2ex] p(Y = y_j) = \dfrac{r_j}{N} \end{array}\right\}$$

Regra da Soma de Probabilidades

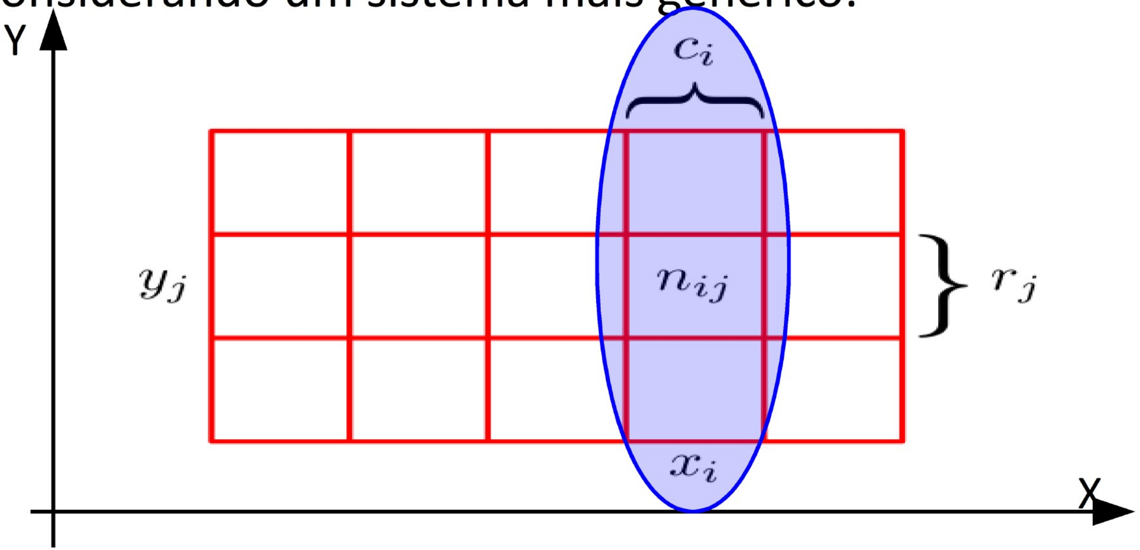
$$\longrightarrow \quad p(X = x_i) = \sum_j^L p(X = x_i, Y = y_j)$$

$$\longrightarrow \quad p(Y = y_j) = \sum_i^M p(X = x_i, Y = y_j)$$

$$X = x_i, i = 1 \ldots M$$

$$Y = y_j, j = 1 \ldots L$$

- Considerando um sistema mais genérico:



$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$$

$$p(X = x_i) = \frac{c_i}{N}$$

$$p(Y = y_j) = \frac{r_i}{N}$$

$$c_i = \sum_j^L n_{ij} \qquad r_j = \sum_i^M n_{ij}$$

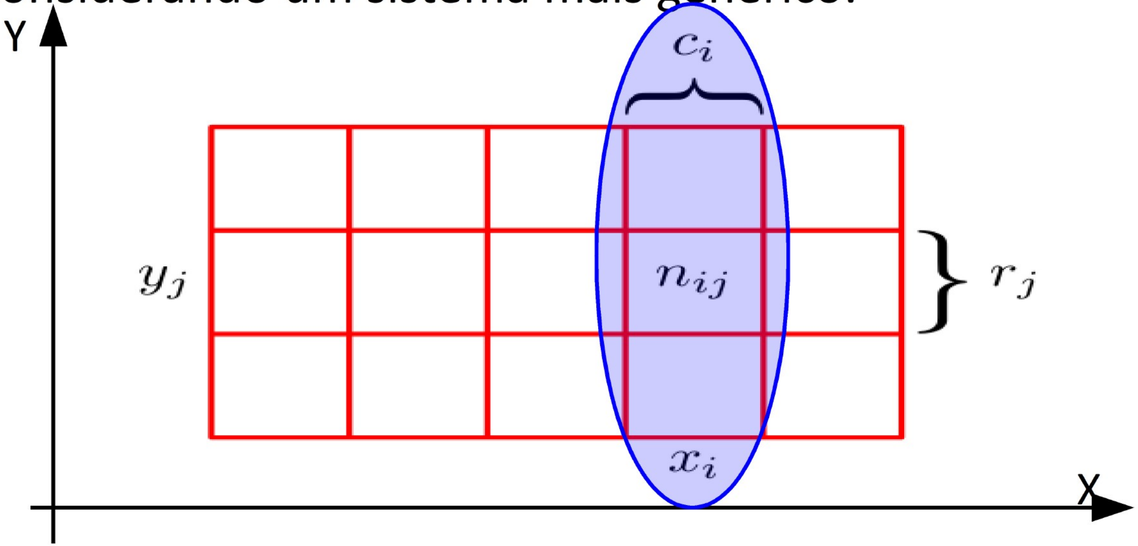
  - Considerando somente a instância de $X = x_i$
  - A fração de $Y = y_j$ dado que $X = x_i$ é chamada de **probabilidade condicional**.

$$p(Y = y_j | X = x_i)$$

$$X = x_i, i = 1 \ldots M$$

$$Y = y_j, j = 1 \ldots L$$

- Considerando um sistema mais genérico:



$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$$

$$p(X = x_i) = \frac{c_i}{N}$$

$$p(Y = y_j) = \frac{r_i}{N}$$

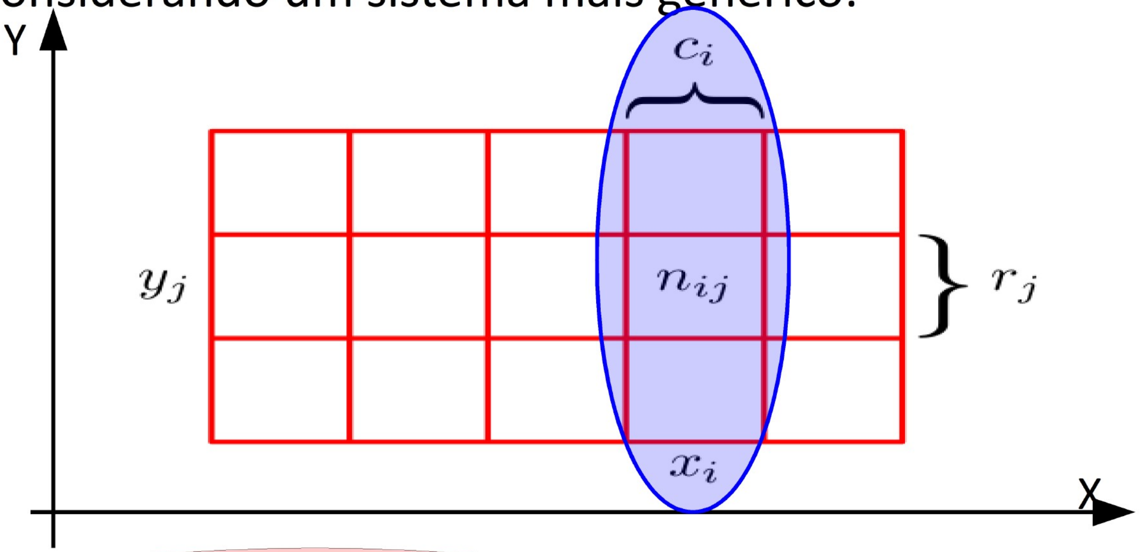
$$c_i = \sum_j^L n_{ij} \qquad r_j = \sum_i^M n_{ij}$$

$$p(Y = y_j | X = x_i) = \frac{n_{ij}}{c_i}$$

- Considerando um sistema mais genérico:

$$X = x_i, i = 1 \dots M$$

$$Y = y_j, j = 1 \dots L$$



$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$$

$$p(X = x_i) = \frac{c_i}{N}$$

$$p(Y = y_j) = \frac{r_j}{N}$$

$$c_i = \sum_j^L n_{ij} \qquad r_j = \sum_i^M n_{ij}$$

$$N \cdot p(X = x_i, Y = y_j)$$

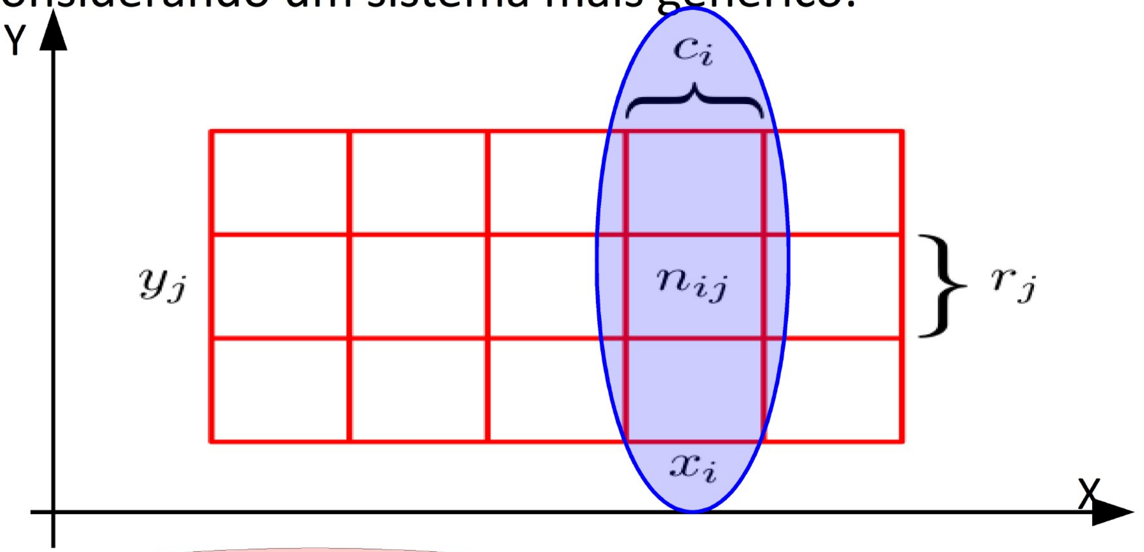
$$p(Y = y_j | X = x_i) = \frac{n_{ij}}{c_i}$$

$$N \cdot p(X = x_i)$$

$$X = x_i, i = 1 \ldots M$$

$$Y = y_j, j = 1 \ldots L$$

- Considerando um sistema mais genérico:



$$p(X=x_i, Y=y_j) = \frac{n_{ij}}{N}$$

$$p(X=x_i) = \frac{c_i}{N}$$

$$p(Y=y_j) = \frac{r_i}{N}$$

$$c_i = \sum_j^L n_{ij} \qquad r_j = \sum_i^M n_{ij}$$

$$N \cdot p(X=x_i, Y=y_j)$$

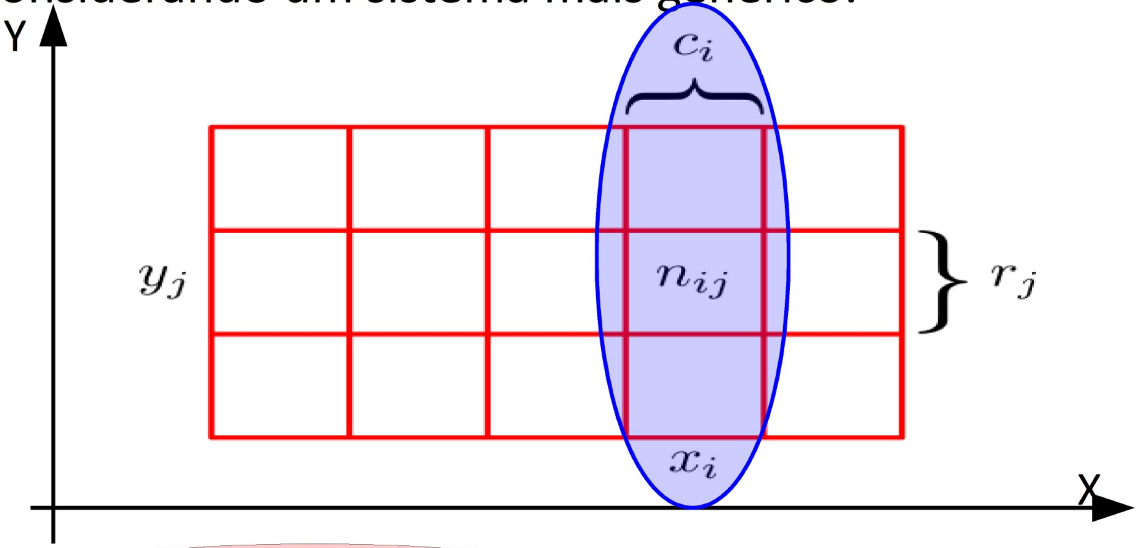
$$p(Y=y_j | X=x_i) = \frac{n_{ij}}{c_i} \Rightarrow p(Y=y_j | X=x_i) = \frac{p(X=x_i, Y=y_j)}{p(X=x_i)}$$

$$N \cdot p(X=x_i)$$

$$X = x_i, i = 1 \dots M$$

$$Y = y_j, j = 1 \dots L$$

- Considerando um sistema mais genérico:



$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$$

$$p(X = x_i) = \frac{c_i}{N}$$

$$p(Y = y_j) = \frac{r_i}{N}$$

$$c_i = \sum_j^L n_{ij} \qquad r_j = \sum_i^M n_{ij}$$

$$N \cdot p(X = x_i, Y = y_j)$$

$$p(Y = y_j | X = x_i) = \frac{n_{ij}}{c_i} \Rightarrow p(Y = y_j | X = x_i) = \frac{p(X = x_i, Y = y_j)}{p(X = x_i)}$$

$$N \,.\, p(X = x_i)$$

Regra do Produto de Probabilidades

$$\boxed{p(X = x_i, Y = y_j) = p(Y = y_j | X = x_i) \cdot p(X = x_i)}$$

- Na forma compacta:

The Rules of Probability

sum rule $$p(X) = \sum_Y p(X,Y)$$

product rule $$p(X,Y) = p(Y|X)p(X)$$

  - $p(X,Y)$ é a probabilidade de X e Y ocorrerem ao mesmo tempo.
  - $p(Y|X)$ é a probabilidade de Y ocorrer dado que X já ocorreu.
  - $p(X)$ é a probabilidade de X ocorrer.

- Na forma compacta:

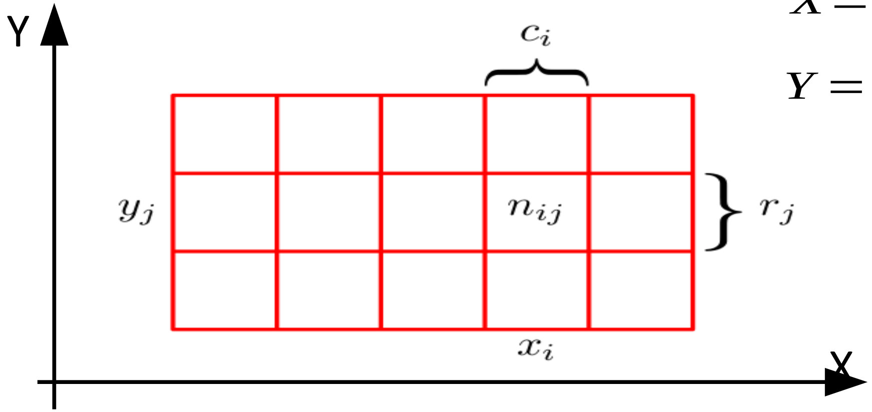
O processo de encontrar p(X) pela soma de outras probabilidades é chamado de **Marginalização.**

The Rules of Probability

sum rule $$p(X) = \sum_Y p(X, Y)$$

product rule $$p(X, Y) = p(Y|X)p(X)$$

  - $p(X, Y)$ é a probabilidade de X e Y ocorrerem ao mesmo tempo.
  - $p(Y|X)$ é a probabilidade de Y ocorrer dado que X já ocorreu.
  - $p(X)$ é a probabilidade de X ocorrer.



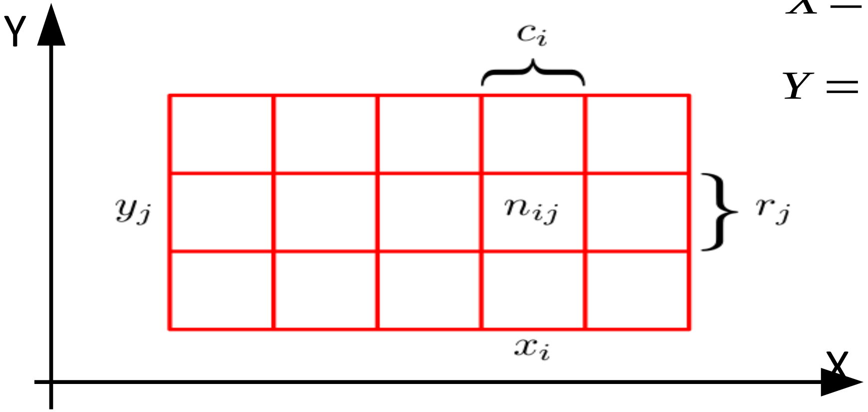
$$X = x_i, i = 1 \dots M$$

$$Y = y_j, j = 1 \dots L$$

- Por uma questão de simetria, podemos escrever: $p(X,Y) = p(Y,X)$

$$\left.\begin{aligned} p(X,Y) &= p(Y|X).\,p(X) \\ p(Y,X) &= p(X|Y).\,p(Y) \end{aligned}\right\} \Rightarrow \quad p(Y|X) = \frac{p(X|Y).\,p(Y)}{p(X)}$$

Fundamentals of Statistics



$X = x_i, i = 1 \dots M$

$Y = y_j, j = 1 \dots L$

- Por uma questão de simetria, podemos escrever: $p(X,Y) = p(Y,X)$

$$\left.\begin{aligned} p(X,Y) &= p(Y|X).\,p(X) \\ p(Y,X) &= p(X|Y).\,p(Y) \end{aligned}\right\} \Rightarrow p(Y|X) = \frac{p(X|Y).\,p(Y)}{p(X)}$$

Teorema de Bayes

The Bayes Theorem

- O Teorema de Bayes
  - Mostra a relação entre a probabilidade condicional e sua inversa;
  - Modela o ganho de informação em função de probabilidades à priori e as evidências;

$$p(Y|X) = \frac{p(X|Y).p(Y)}{p(X)}$$

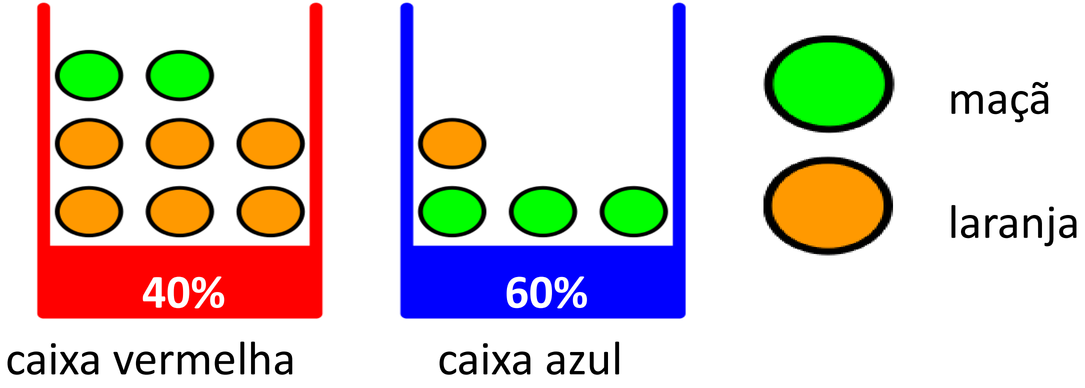
Thomas Bayes

Matemático, Estatístico e Ministro Presbiteriano

Universidade de Edinburgh

- De volta ao problema incial:

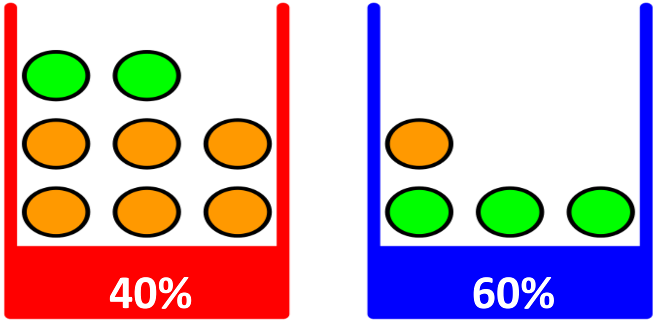
$$p(C=v)=0.4$$



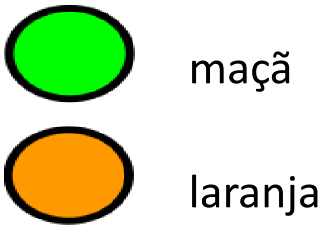
- De volta ao problema incial:

$$p(C=v)=0.4$$

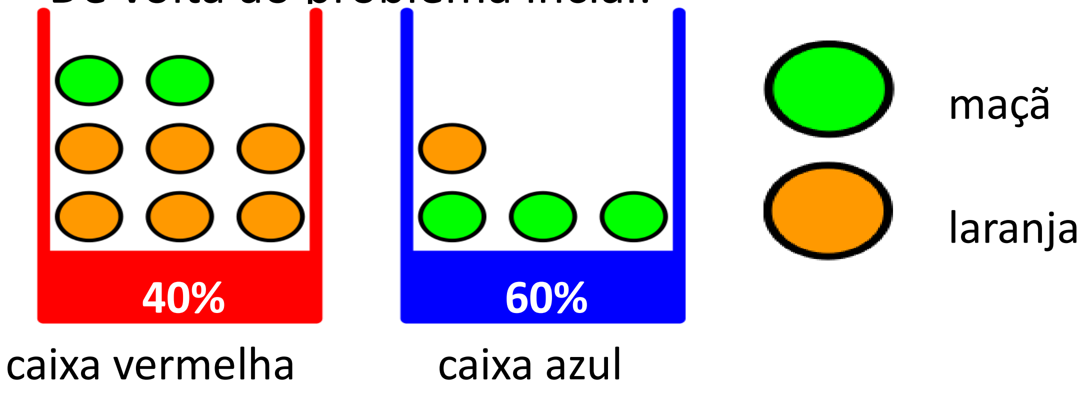
$$p(C=a)=0.6$$



caixa vermelha caixa azul



- De volta ao problema incial:

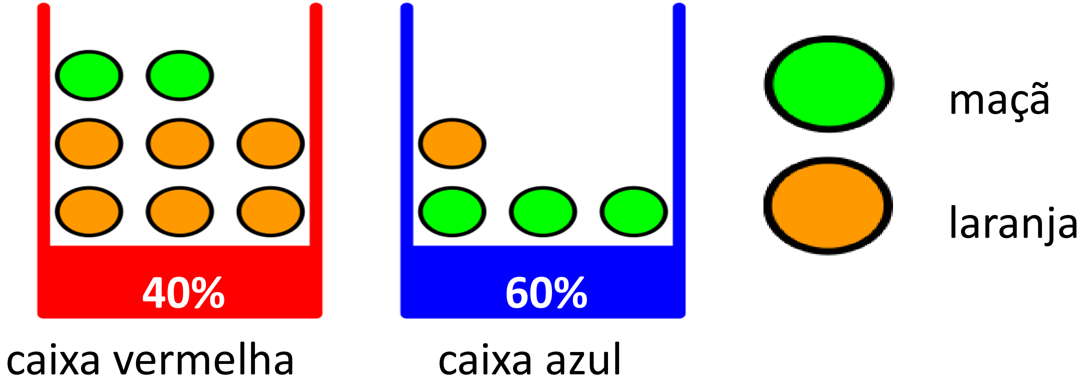


$$p(C=v)=0.4$$

$$p(C=a)=0.6$$

$$p(F=m|C=v)=\frac{2}{8}=\frac{1}{4}$$

- De volta ao problema incial:



$$p(C=v)=0.4$$

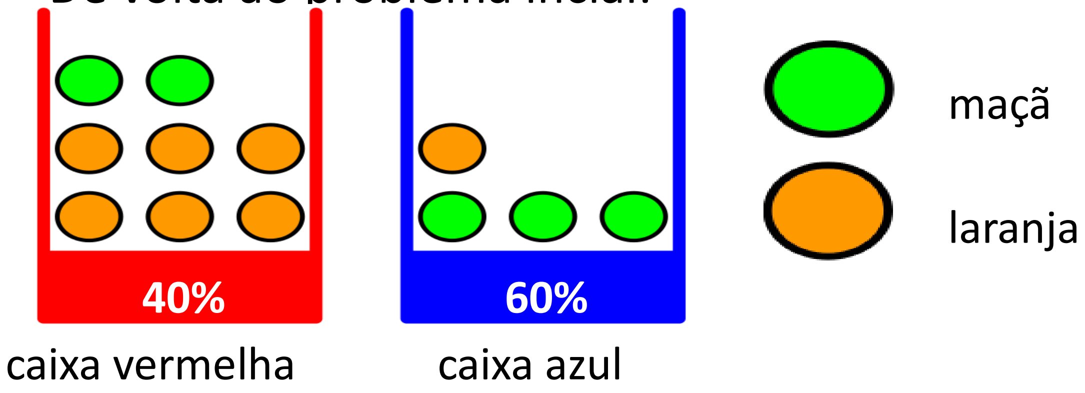
$$p(C=a)=0.6$$

$$p(F=m|C=v)=\frac{2}{8}=\frac{1}{4}$$

$$p(F=l|C=v)=\frac{6}{8}=\frac{3}{4}$$

Conceitos Básicos de Probabilidade

- De volta ao problema incial:



$$p(C=v)=0.4$$

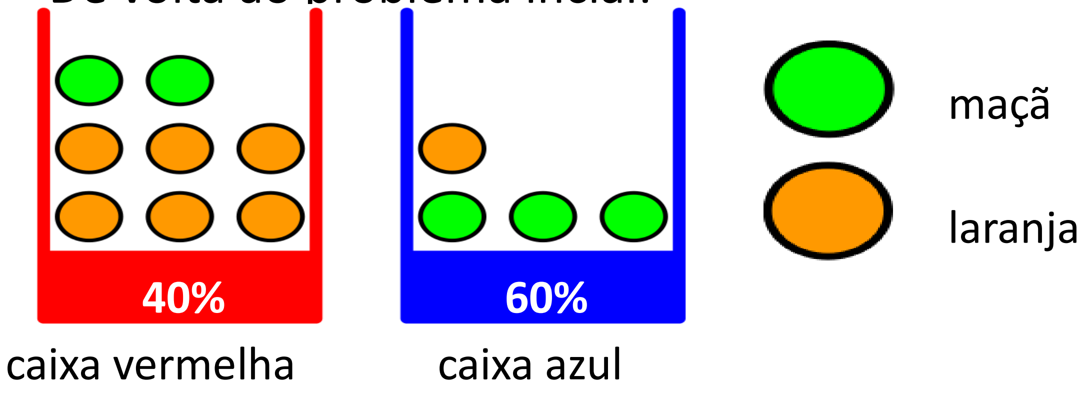
$$p(C=a)=0.6$$

$$p(F=m|C=v)=\frac{2}{8}=\frac{1}{4}$$

$$p(F=l|C=v)=\frac{6}{8}=\frac{3}{4}$$

$$p(F=m|C=a)=\frac{3}{4}$$

- De volta ao problema incial:



$$p(C=v)=0.4$$

$$p(C=a)=0.6$$

$$p(F=m|C=v)=\frac{2}{8}=\frac{1}{4}$$

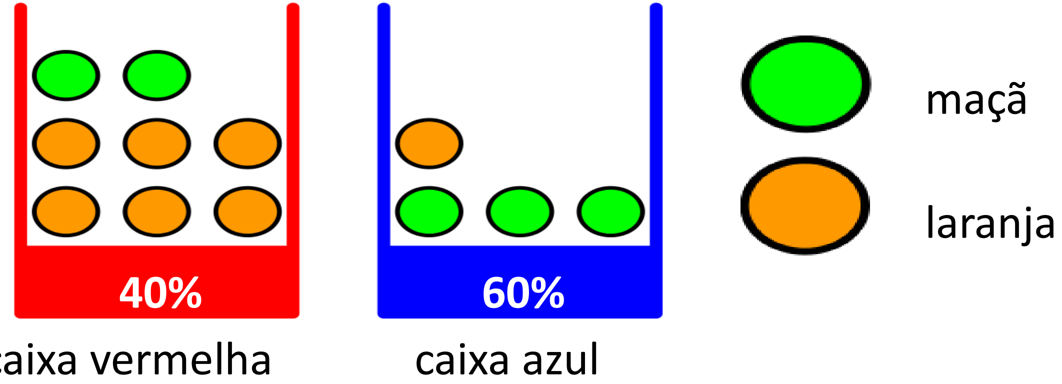
$$p(F=l|C=v)=\frac{6}{8}=\frac{3}{4}$$

$$p(F=m|C=a)=\frac{3}{4}$$

$$p(F=l|C=a)=\frac{1}{4}$$

Conceitos Básicos de Probabilidade

- De volta ao problema incial:



$$p(F=m)=?$$

$$p(C=v)=0.4$$

$$p(C=a)=0.6$$

$$p(F=m|C=v)=\frac{2}{8}=\frac{1}{4}$$

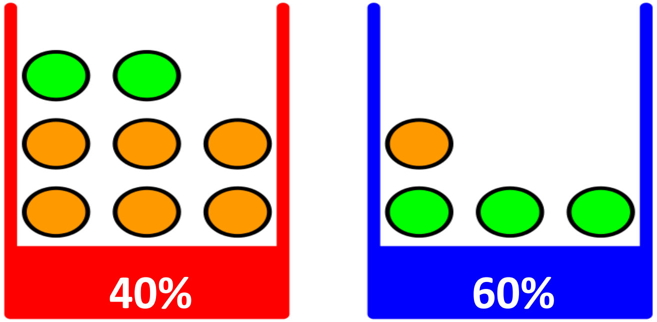
$$p(F=l|C=v)=\frac{6}{8}=\frac{3}{4}$$

$$p(F=m|C=a)=\frac{3}{4}$$

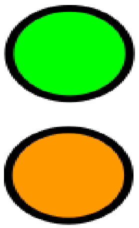
$$p(F=l|C=a)=\frac{1}{4}$$

Conceitos Básicos de Probabilidade

- De volta ao problema incial:



caixa vermelha   caixa azul



maçã
laranja

$$p(F=m)=\sum_{C} p(F=m,C)$$

$$p(C=v)=0.4$$

$$p(C=a)=0.6$$

$$p(F=m|C=v)=\frac{2}{8}=\frac{1}{4}$$

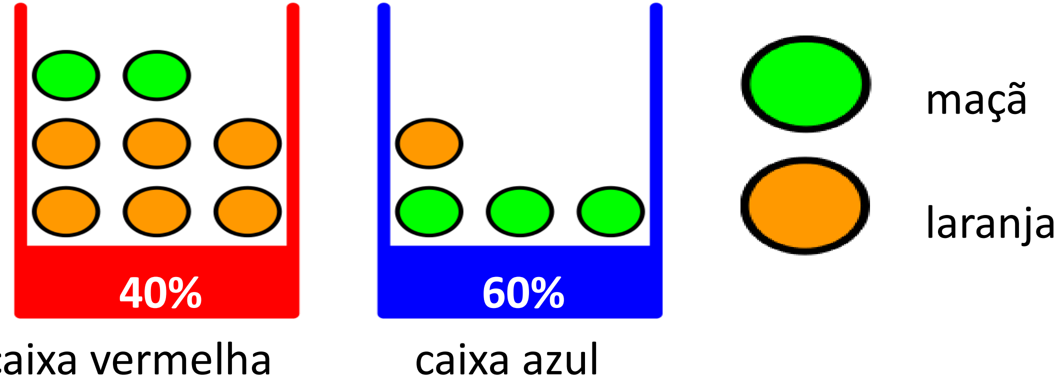
$$p(F=l|C=v)=\frac{6}{8}=\frac{3}{4}$$

$$p(F=m|C=a)=\frac{3}{4}$$

$$p(F=l|C=a)=\frac{1}{4}$$

Conceitos Básicos de Probabilidade

- De volta ao problema incial:



$$p(C=v)=0.4$$

$$p(C=a)=0.6$$

$$p(F=m|C=v)=\frac{2}{8}=\frac{1}{4}$$

$$p(F=l|C=v)=\frac{6}{8}=\frac{3}{4}$$

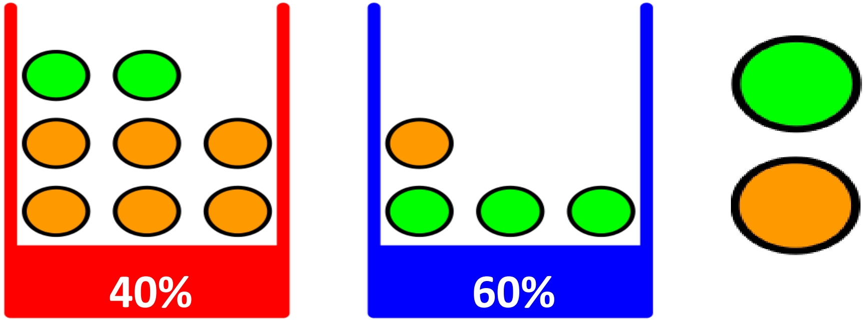
$$p(F=m|C=a)=\frac{3}{4}$$

$$p(F=l|C=a)=\frac{1}{4}$$

$$p(F=m)=\sum_{c} p(F=m,C)$$

$$p(F=m)=p(F=m,C=v)+p(F=m,C=a)$$

- De volta ao problema incial:



maçã

laranja

caixa vermelha

caixa azul

$$p(C=v)=0.4$$

$$p(C=a)=0.6$$

$$p(F=m|C=v)=\frac{2}{8}=\frac{1}{4}$$

$$p(F=l|C=v)=\frac{6}{8}=\frac{3}{4}$$

$$p(F=m|C=a)=\frac{3}{4}$$

$$p(F=l|C=a)=\frac{1}{4}$$

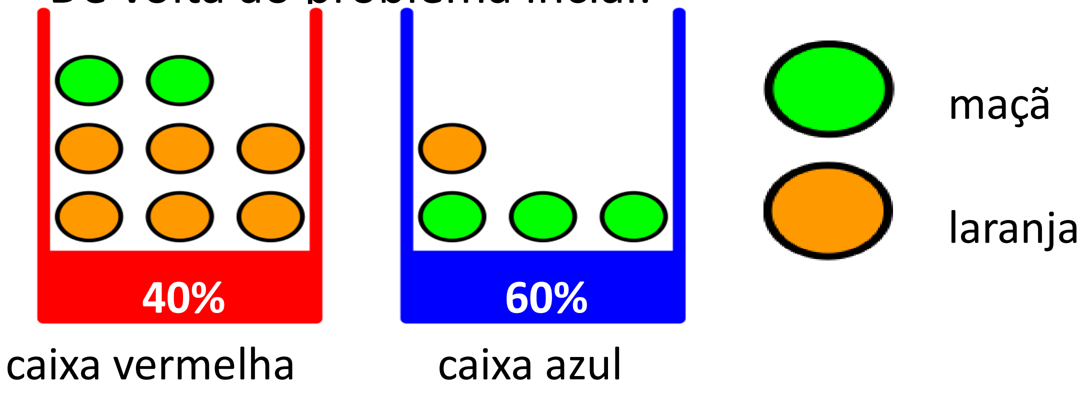
$$p(F=m)=\sum_{c} p(F=m,C)$$

$$p(F=m)=p(F=m,C=v)+p(F=m,C=a)$$

$$p(F=m)=p(F=m|C=v)\cdot p(C=v)+p(F=m|C=a)\cdot p(C=a)$$

$$p(F=m)=\frac{1}{4}\cdot\frac{4}{10}+\frac{3}{4}\cdot\frac{6}{10}=\frac{11}{20}$$

- De volta ao problema incial:



$$p(C=v)=0.4$$

$$p(C=a)=0.6$$

$$p(F=m|C=v)=\frac{2}{8}=\frac{1}{4}$$

$$p(F=l|C=v)=\frac{6}{8}=\frac{3}{4}$$

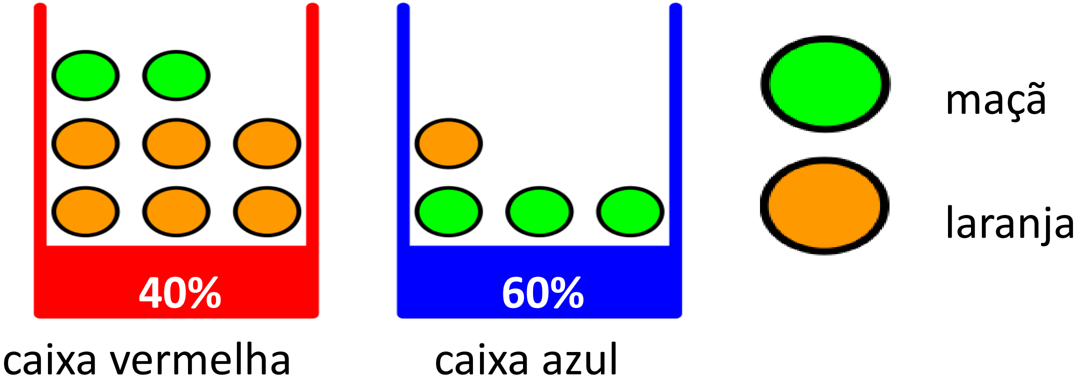
$$p(F=m|C=a)=\frac{3}{4}$$

$$p(F=l|C=a)=\frac{1}{4}$$

$$p(F=m)=\sum_c p(F=m,C)=\frac{11}{20}$$

Conceitos Básicos de Probabilidade

- De volta ao problema incial:



$$p(C=v)=0.4$$

$$p(C=a)=0.6$$

$$p(F=m|C=v)=\frac{2}{8}=\frac{1}{4}$$

$$p(F=l|C=v)=\frac{6}{8}=\frac{3}{4}$$

$$p(F=m|C=a)=\frac{3}{4}$$

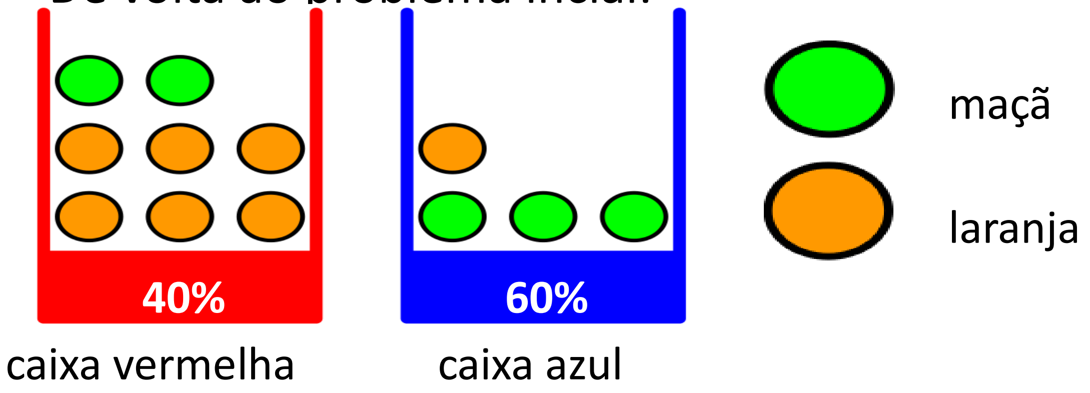
$$p(F=l|C=a)=\frac{1}{4}$$

$$p(F=m)=\sum_{c} p(F=m,C)=\frac{11}{20}$$

$$p(F=l)=?$$

Conceitos Básicos de Probabilidade

- De volta ao problema incial:



$$p(C=v)=0.4$$

$$p(C=a)=0.6$$

$$p(F=m|C=v)=\frac{2}{8}=\frac{1}{4}$$

$$p(F=l|C=v)=\frac{6}{8}=\frac{3}{4}$$

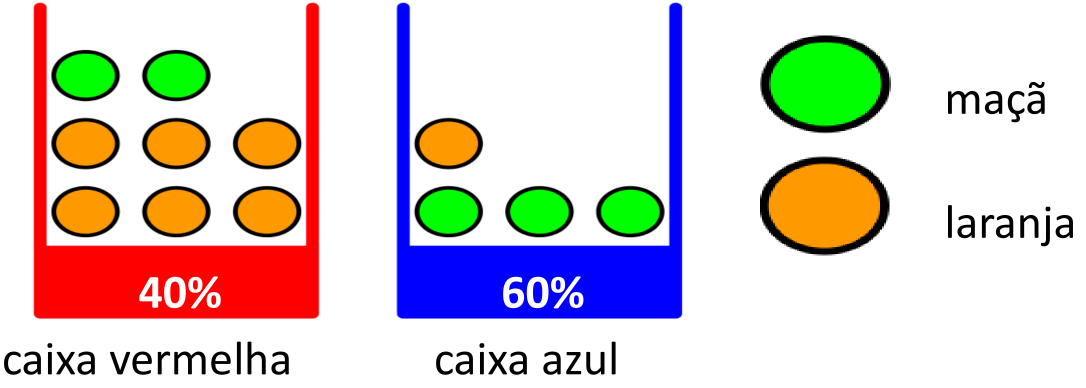
$$p(F=m|C=a)=\frac{3}{4}$$

$$p(F=l|C=a)=\frac{1}{4}$$

$$p(F=m)=\sum_{c} p(F=m,C)=\frac{11}{20}$$

$$p(F=l)=1-p(F=m)$$

- De volta ao problema incial:



$$p(F=m)=\sum_c p(F=m,C)=\frac{11}{20}$$

$$p(F=l)=1-p(F=m)=1-\frac{11}{20}=\frac{9}{20}$$

$$p(C=v)=0.4$$

$$p(C=a)=0.6$$

$$p(F=m|C=v)=\frac{2}{8}=\frac{1}{4}$$

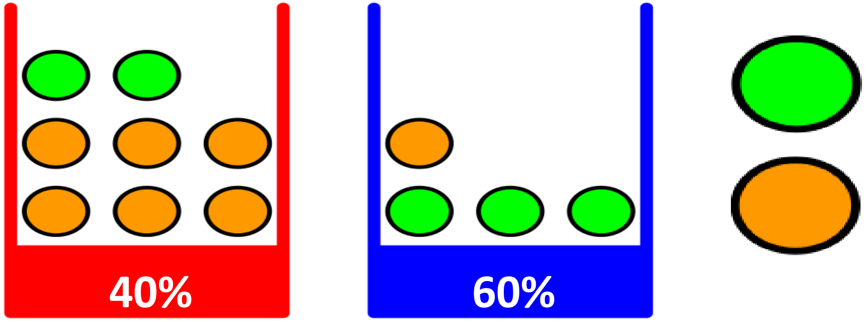
$$p(F=l|C=v)=\frac{6}{8}=\frac{3}{4}$$

$$p(F=m|C=a)=\frac{3}{4}$$

$$p(F=l|C=a)=\frac{1}{4}$$

Conceitos Básicos de Probabilidade

- De volta ao problema incial:



maçã

laranja

caixa vermelha

caixa azul

$$p(F=m)=\sum_{c} p(F=m,C)=\frac{11}{20}$$

$$p(F=l)=1-p(F=m)=\frac{9}{20}$$

$$p(C=v)=0.4$$

$$p(C=a)=0.6$$

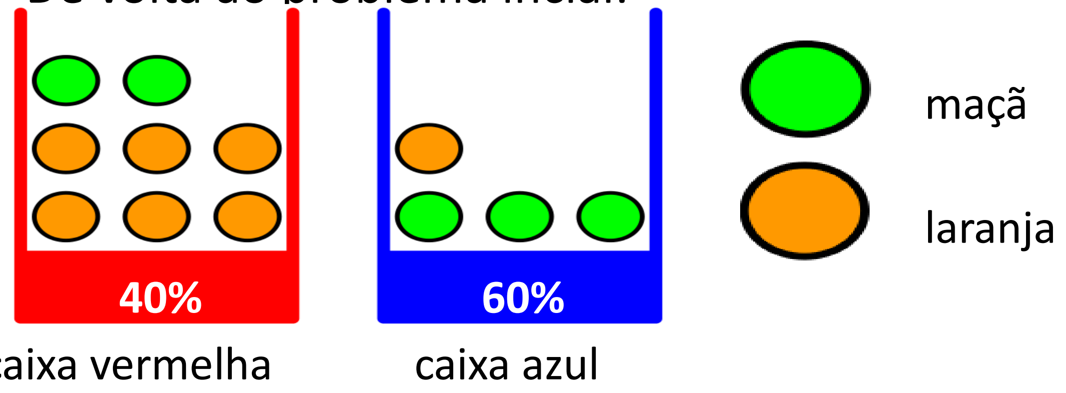
$$p(F=m|C=v)=\frac{2}{8}=\frac{1}{4}$$

$$p(F=l|C=v)=\frac{6}{8}=\frac{3}{4}$$

$$p(F=m|C=a)=\frac{3}{4}$$

$$p(F=l|C=a)=\frac{1}{4}$$

- De volta ao problema incial:



$$p(C=v)=0.4$$

$$p(C=a)=0.6$$

$$p(F=m|C=v)=\frac{2}{8}=\frac{1}{4}$$

$$p(F=l|C=v)=\frac{6}{8}=\frac{3}{4}$$

$$p(F=m|C=a)=\frac{3}{4}$$

$$p(F=l|C=a)=\frac{1}{4}$$

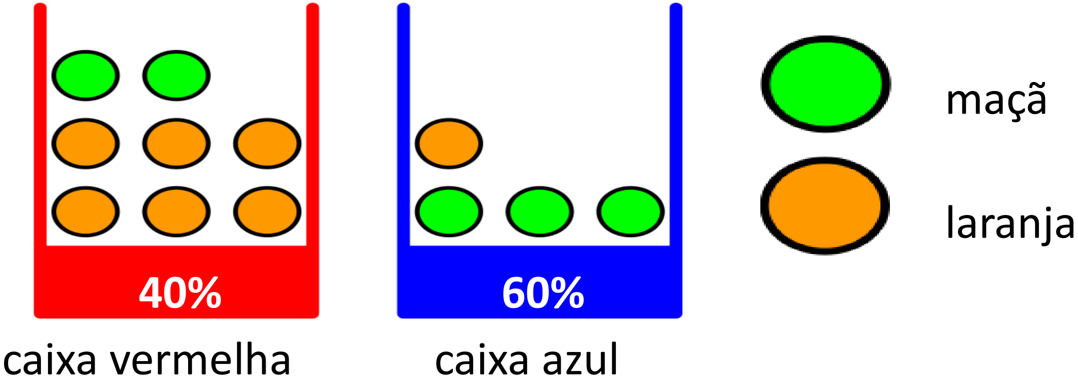
$$p(F=m)=\sum_{c} p(F=m,C)=\frac{11}{20}$$

$$p(F=l)=1-p(F=m)=\frac{9}{20}$$

$$p(C=v|F=l)=?$$

Conceitos Básicos de Probabilidade

- De volta ao problema incial:



$$p(C=v)=0.4$$

$$p(C=a)=0.6$$

$$p(F=m|C=v)=\frac{2}{8}=\frac{1}{4}$$

$$p(F=l|C=v)=\frac{6}{8}=\frac{3}{4}$$

$$p(F=m|C=a)=\frac{3}{4}$$

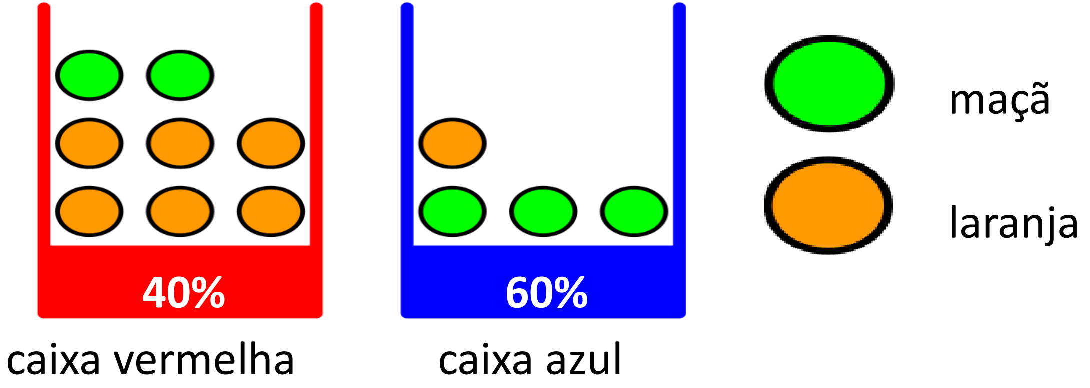
$$p(F=l|C=a)=\frac{1}{4}$$

$$p(F=m)=\sum_c p(F=m,C)=\frac{11}{20}$$

$$p(F=l)=1-p(F=m)=\frac{9}{20}$$

$$p(C=v|F=l)=\frac{p(F=l|C=v)\cdot p(C=v)}{p(F=l)}$$

- De volta ao problema incial:



$$p(C=v)=0.4$$

$$p(C=a)=0.6$$

$$p(F=m|C=v)=\frac{2}{8}=\frac{1}{4}$$

$$p(F=l|C=v)=\frac{6}{8}=\frac{3}{4}$$

$$p(F=m|C=a)=\frac{3}{4}$$

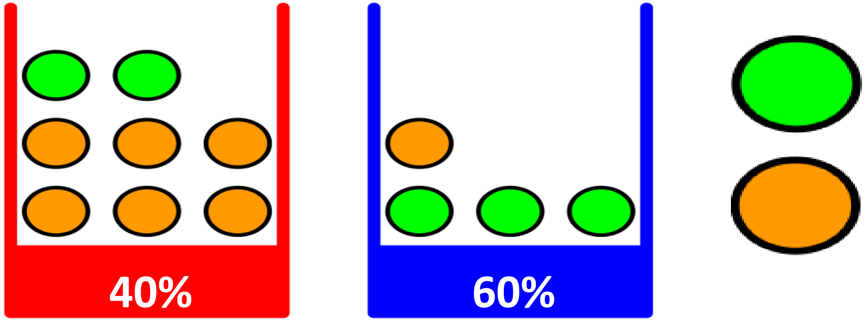
$$p(F=l|C=a)=\frac{1}{4}$$

$$p(F=m)=\sum_{c} p(F=m,C)=\frac{11}{20}$$

$$p(F=l)=1-p(F=m)=\frac{9}{20}$$

$$p(C=v|F=l)=\frac{p(F=l|C=v)\cdot p(C=v)}{p(F=l)}=\frac{3}{4}\cdot\frac{4}{10}\cdot\frac{20}{9}=\frac{2}{3}$$

- De volta ao problema incial:



caixa vermelha caixa azul

maçã

laranja

$$p(C=v)=0.4$$

$$p(C=a)=0.6$$

$$p(F=m|C=v)=\frac{2}{8}=\frac{1}{4}$$

$$p(F=l|C=v)=\frac{6}{8}=\frac{3}{4}$$

$$p(F=m|C=a)=\frac{3}{4}$$

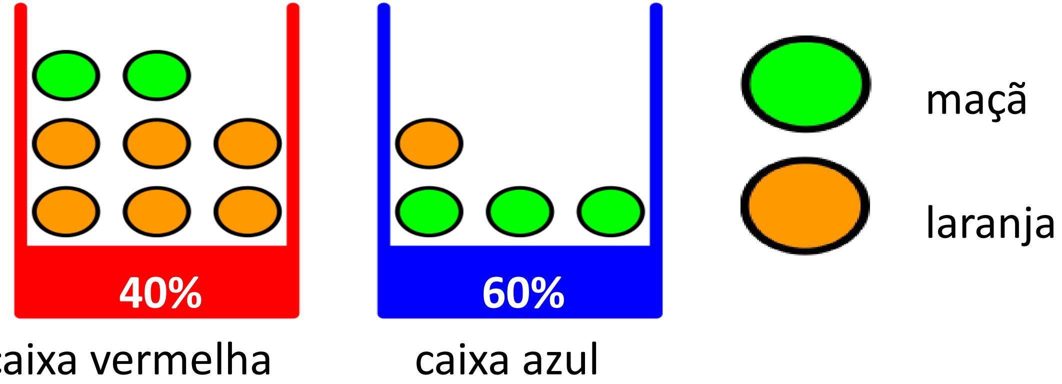
$$p(F=l|C=a)=\frac{1}{4}$$

$$p(F=m)=\sum_{c} p(F=m,C)=\frac{11}{20}$$

$$p(F=l)=1-p(F=m)=\frac{9}{20}$$

$$p(C=v|F=l)=\frac{p(F=l|C=v)\cdot p(C=v)}{p(F=l)}=\frac{2}{3}$$

- De volta ao problema incial:



$$p(C=v)=0.4$$

$$p(C=a)=0.6$$

$$p(F=m|C=v)=\frac{2}{8}=\frac{1}{4}$$

$$p(F=l|C=v)=\frac{6}{8}=\frac{3}{4}$$

$$p(F=m|C=a)=\frac{3}{4}$$

$$p(F=l|C=a)=\frac{1}{4}$$

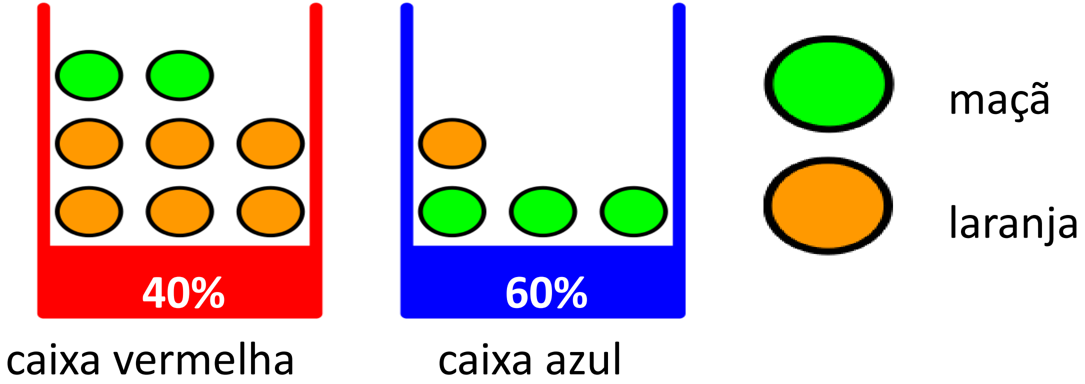
$$p(F=m)=\sum_{c} p(F=m,C)=\frac{11}{20}$$

$$p(F=l)=1-p(F=m)=\frac{9}{20}$$

$$p(C=v|F=l)=\frac{p(F=l|C=v)\cdot p(C=v)}{p(F=l)}=\frac{2}{3}$$

$$p(C=a|F=l)=?$$

- De volta ao problema incial:



$$p(C=v)=0.4$$

$$p(C=a)=0.6$$

$$p(F=m|C=v)=\frac{2}{8}=\frac{1}{4}$$

$$p(F=l|C=v)=\frac{6}{8}=\frac{3}{4}$$

$$p(F=m|C=a)=\frac{3}{4}$$

$$p(F=l|C=a)=\frac{1}{4}$$

$$p(F=m)=\sum_{c} p(F=m,C)=\frac{11}{20}$$

$$p(F=l)=1-p(F=m)=\frac{9}{20}$$

$$p(C=v|F=l)=\frac{p(F=l|C=v)\cdot p(C=v)}{p(F=l)}=\frac{2}{3}$$

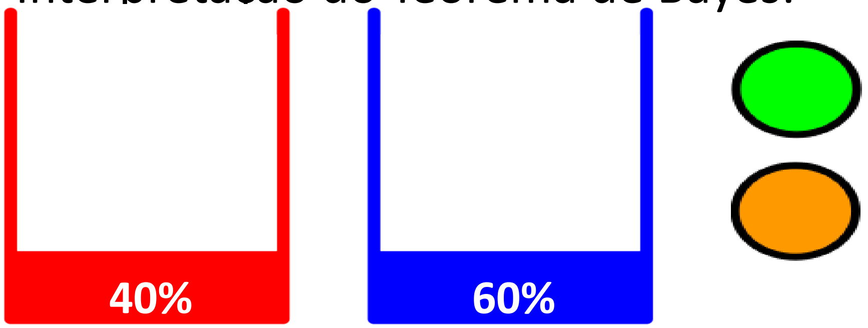
$$p(C=a|F=l)=1-\frac{2}{3}=\frac{1}{3}$$

- Interpretação do Teorema de Bayes:

  - Se queremos saber a probabilidade de ter escolhido uma caixa dado a informação do tipo da fruta.

Conceitos Básicos de Probabilidade

- Interpretação do Teorema de Bayes:



caixa vermelha caixa azul

maçã

laranja

$$\underbrace{p(C|F)}_{\text{posteriori}} = \frac{\overbrace{p(F|C)}^{\text{Likelihood}} . \overbrace{p(C)}^{\text{priori}}}{\underbrace{p(F)}_{\text{Preditor priori}}}$$

- Se queremos saber a probabilidade de ter escolhido uma caixa dado a informação do tipo da fruta.
    - Melhor suposição inicial: $p(C)$, conhecimento a priori, antes de conhecermos a identidade da fruta.

- Interpretação do Teorema de Bayes:

Likelihood

posteriori

priori

$$p(C|F)=\frac{p(F|C).\,p(C)}{p(F)}$$

Preditor priori

  - Se queremos saber a probabilidade de ter escolhido uma caixa dado a informação do tipo da fruta.
    - Melhor suposição inicial: $p(C)$, conhecimento a priori, antes de conhecermos a identidade da fruta.

Conceitos Básicos de Probabilidade

- Interpretação do Teorema de Bayes:

Likelihood

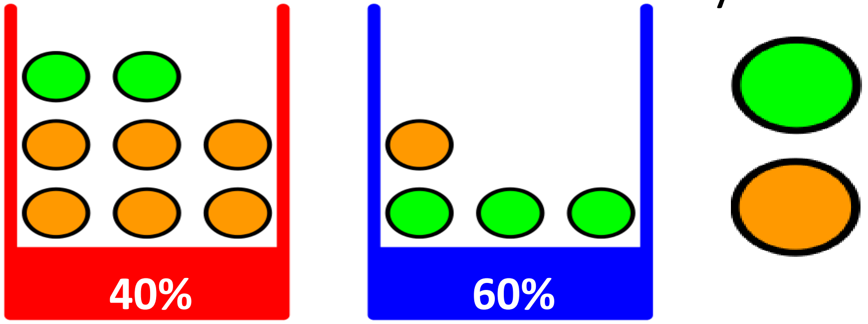
posteriori **priori**

$$p(C|F)=\frac{p(F|C).\,p(C)}{p(F)}$$

Preditor priori

  - Se queremos saber a probabilidade de ter escolhido uma caixa dado a informação do tipo da fruta.
    - Melhor suposição inicial: $p(C)$, conhecimento a priori, antes de conhecermos a identidade da fruta.
    - Com o teorema de Bayes podemos calcular $p(C|F)$ após observar a identidade da fruta.

- Interpretação do Teorema de Bayes:



maçã

laranja

caixa vermelha

caixa azul

Likelihood

posteriori

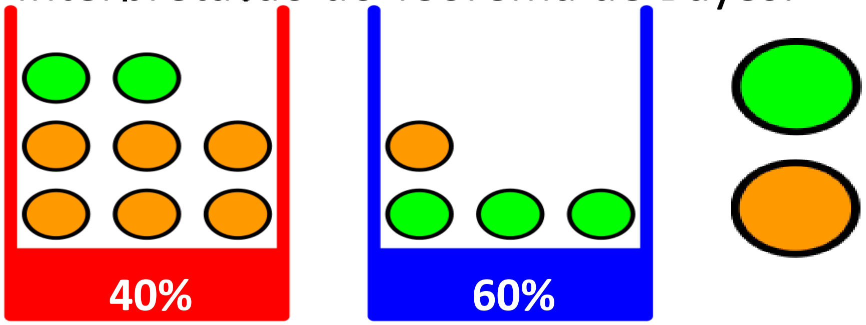
priori

$$p(C|F)=\frac{p(F|C).\,p(C)}{p(F)}$$

Preditor priori

  - Se queremos saber a probabilidade de ter escolhido uma caixa dado a informação do tipo da fruta.
    - Melhor suposição inicial: $p(C)$, conhecimento a priori, antes de conhecermos a identidade da fruta.
    - Com o teorema de Bayes podemos calcular $p(C|F)$ após observar a identidade da fruta.

- Interpretação do Teorema de Bayes:



maçã

laranja

caixa vermelha

caixa azul

Likelihood

posteriori

priori

$$p(C|F)=\frac{p(F|C).\,p(C)}{p(F)}$$

Preditor priori

  - Se a fruta for LARANJA, essa informação fornece importante evidência de que a caixa escolhida seja a VERMELHA, apesar dela ser a caixa com menor probabilidade de ser escolhida.

- Função Densidade de Probabilidade (PDF)

  - Do mesmo modo que consideramos probabilidades de eventos discretos (variáveis aleatórias discretas), podemos definir probabilidades sobre variáveis contínuas.
  - Ex:
    - Qual a probabilidade de um carro X viajar a 105.01 km/h ?

- Função Densidade de Probabilidade (PDF)

  – Do mesmo modo que consideramos probabilidades de eventos discretos (variáveis aleatórias discretas), podemos definir probabilidades sobre variáveis contínuas.

  – Ex:

    - Qual a probabilidade de um carro X viajar a 105.01 km/h ?

  – Se a probabilidade de uma variável aleatória contínua x estar no intervalo $(x, x+\delta x)$ é dada por $p(x)\cdot\delta x$ quando $\delta x \rightarrow 0$, então $p(x)$ é chamda de **função densidade probabilidade** (pdf) sobre x .

- Função Densidade de Probabilidade (PDF)
  - Propriedades de uma pdf:

$$p(x) \geq 0$$

$$\int_{-\infty}^{+\infty} p(x)\,dx = 1$$

  - A probabilidade de x estar dentro do intervalo (a, b) é dada por:

$$p(a \leq x \leq b) = \int_{a}^{b} p(x)\,dx$$

- Função Densidade de Probabilidade (PDF)
  - Propriedades de uma pdf:

$$p(x) \geq 0$$

$$\int_{-\infty}^{+\infty} p(x)\,dx = 1$$

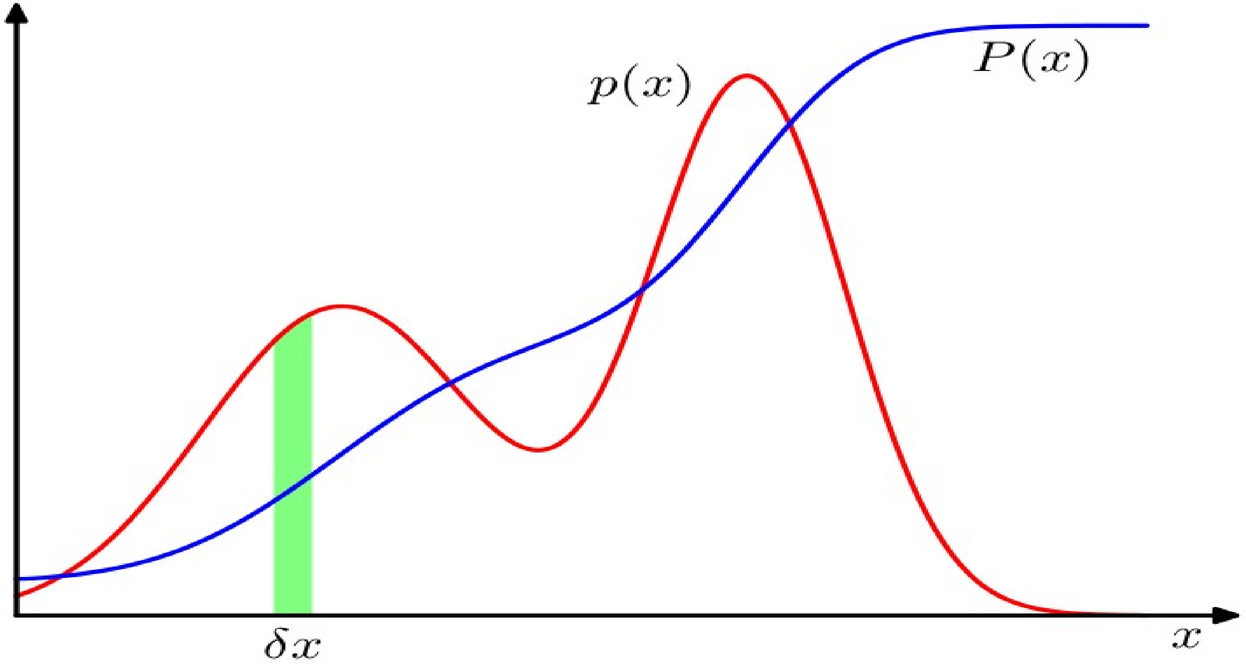
  - A probabilidade de x estar dentro do intervalo (a, b) é dada por:

$$p(a \leq x \leq b) = \int_a^b p(x)\,dx$$

  - A probabilidade de estar no intervalo $(-\infty, z)$ é dada pela **função distribuição acumulativa** (CDF) definida por:

$$P(z) = \int_{-\infty}^{z} p(x)\,dx$$

- Função Densidade de Probabilidade (PDF) e Função Densidade Acumulativa (CDF)



- Função Densidade de Probabilidade (PDF)

  - As regras da soma e do produto, bem como o teorema de Bayes, se aplicam igualmente para o caso de densidades de probabilidades.

  - Sendo x e y variáveis contínuas, temos:

    - Regra da soma: $p(x)=\int p(x,y)dy$

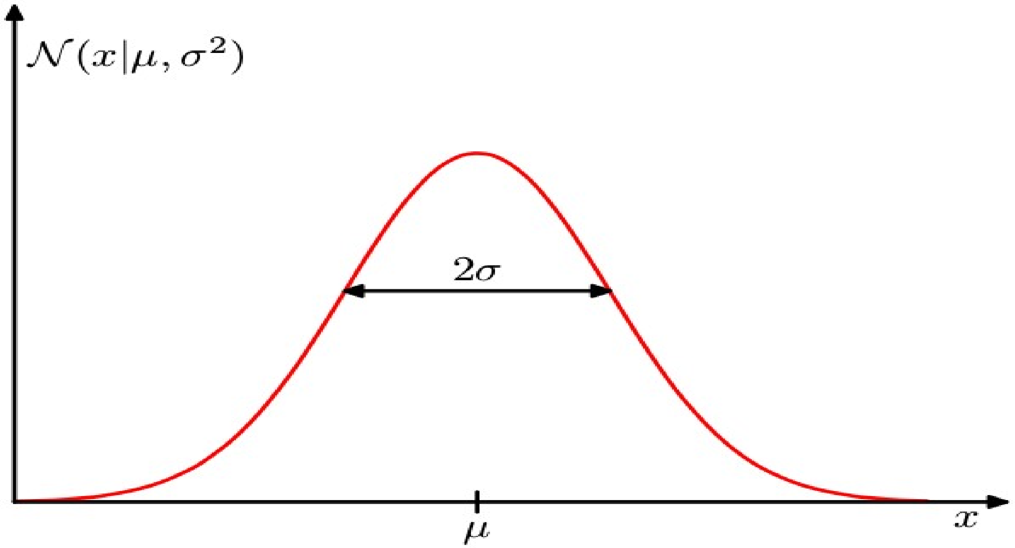
    - Regra do produto: $p(x,y)=p(y|x)\cdot p(x)$

- Função Gaussiana
  - É um das funções de distribuição de probabilidades mais conhecidas;
  - Muitos fenômenos seguem essa distribuição;
  - Também conhecida como Normal ou Distribuição Gaussiana.
  - Para o caso de uma variável:

$$\mathcal{N}(x|\bar{x},\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}$$

onde: $\bar{x} = E[X] = \frac{1}{M}\sum_{i=1}^{M} x^{(i)}$ e $\sigma^2 = V[X] = E[(X-E[X])^2] = \frac{1}{M}\sum_{i=1}^{M}(x^{(i)}-\sigma)^2$

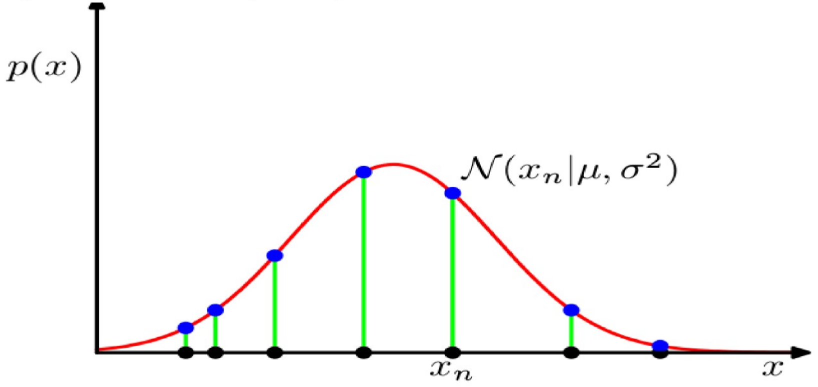
- Função Gaussiana



$$\int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu,\sigma^2\right)\,\mathrm{d}x = 1.$$

- Função Gaussiana
  - Supondo que tenhamos um conjunto de dados $\boldsymbol{X} = [x_{1,} x_{2,} \ldots, x_n]^T$ representando N observações da variável x;
  - Supondo que as observações são dispostas em um gráfico independentemente da distribuição gaussiana que os representa e sem conhecermos a média e a variância dessa distribuição;
  - E se os pontos são ditos: *independent and identically distributed (i.i.d.)*
  - Os parâmetros μ e σ^2 são aqueles que maximizam o produtório acima.

$$p(\mathbf{x}|\mu, \sigma^2) = \prod_{n=1}^{N} \mathcal{N}\left(x_n|\mu, \sigma^2\right)$$

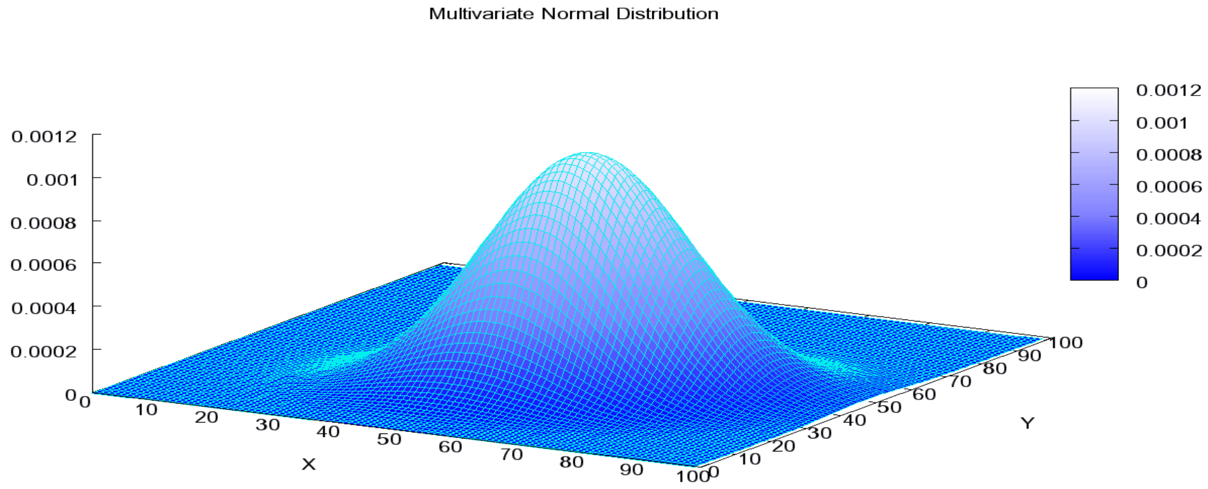


- Função Gaussiana Multivariada
  - Supondo $\boldsymbol{x}=[x_{1,}x_{2,}\ldots,x_n]^T$ uma amostra formada por n características.
  - A probabilidade da amostra x é dada pela probabilidade conjunta de suas características.

- Função Gaussiana Multivariada

Supondo $\boldsymbol{x} = [x_1, x_2, \ldots, x_n]^T$ uma amostra formada por n variáveis.

$$\mathcal{N}(\mathbf{x}|\bar{\mathbf{x}}, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x}-\bar{\mathbf{x}})^\top \Sigma^{-1}(\mathbf{x}-\bar{\mathbf{x}})}$$



Naïve Bayes Classifier

- The Naïve Bayes Classifier is a conditional probability model where, given an instance to be classified:

$$\mathbf{x} = [x_1, x_2, \cdots, x_n]^\intercal, \mathbf{x} \in \mathbb{R}^n$$

  - What the probability of a Class C_k , given the instance?

$$P(C_k|\mathbf{x}) = P(C_k|x_1, x_2, \cdots, x_n)$$

- Using Bayes's Theorem:

$$P(C_k|\mathbf{x}) = \frac{P(C_k)P(\mathbf{x}|C_k)}{P(\mathbf{x})}$$

- The Naïve Bayes Classifier is a conditional probability model where, given an instance to be classified:

$$\mathbf{x} = [x_1, x_2, \cdots, x_n]^\intercal, \mathbf{x} \in \mathbb{R}^n$$

  - What the probability of a Class C_k , given the instance?

$$P(C_k|\mathbf{x}) = P(C_k|x_1, x_2, \cdots, x_n)$$

- Using Bayes's Theorem:

$$P(C_k|\mathbf{x}) = \frac{P(C_k)P(\mathbf{x}|C_k)}{P(\mathbf{x})}$$

Naïve Bayes Classifier

- Using conditional probability:

$$
\begin{aligned}
p(C_k|x_1, x_2, \cdots, x_n) &= p(C_k)p(x_1, x_2, \cdots, x_n|C_k) \\
&= p(C_k)p(x_1|C_k)p(x_2, \cdots, x_n|C_k, x_1) \\
&= p(C_k)p(x_1|C_k)p(x_2|C_k, x_1)p(x_3, \cdots, x_n|C_k, x_1,_x 2) \\
&= p(C_k)p(x_1|C_k)p(x_2|C_k, x_1) \cdots p(x_n, \cdots, x_n|C_k, x_1,_x 2 \cdots x_{n-1})
\end{aligned}
$$

- Considering Independence of features:

$$p(x_i|C_k, x_j) = p(x_i|C_k)$$

$$
\begin{aligned}
p(C_k|x_1, x_2, \cdots, x_n) &= p(C_k)p(x_1|C_k)p(x_2|C_k) \cdots p(x_n|C_k) \\
&= p(C_k) \prod_{i=1}^{n} p(x_i|C_k)
\end{aligned}
$$

Naïve Bayes Classifier

- The classifier:

$$\hat{y} = \underset{k \in \{1, \cdots, K\}}{\operatorname{argmax}} \; p(C_k) \prod_{i=1}^{n} p(x_i | C_k)$$