

## Introdução à Probabilidade

Jair Donadelli

### Sumário

§0 Generalidades sobre conjuntos	1
§1 Princípios de Análise Combinatória	4
§2 Introdução à Probabilidade	7
§3 Probabilidade Condicional e Independência	18
§4 Variáveis Aleatórias	30
§5 Esperança matemática	38
§6 Principais modelos discretos	49
§7 Principais modelos contínuos	63
§8 Distribuição Normal e Teorema Central do Limite	68
§9 Distribuição Conjunta	88
§10 Distribuição e Esperança condicionais	96
§11 Cadeias de Markov	99
§12 Passeio aleatório em grafos	131

### §0 Generalidades sobre conjuntos

$A, B, C, \dots$  denotam conjuntos quaisquer.

**pertinência**  $x \in A$  denota a proposição  $x$  é elemento de  $A$ .

$x \notin A$  denota a proposição  $x$  não é elemento de  $A$ .

**vazio**  $\emptyset$  denota o conjunto vazio.

**inclusão**  $A \subset B$  significa  $\forall a (a \in A \Rightarrow a \in B)$ .

$A = B$  equivale a  $A \subset B$  e  $B \subset A$ .

**união**  $x \in A \cup B$  significa  $(x \in A)$  ou  $(x \in B)$ .

**interseção**  $x \in A \cap B$  significa  $(x \in A)$  e  $(x \in B)$ .

**diferença**  $x \in A \setminus B$  significa  $(x \in A)$  e  $(x \notin B)$ .

**diferença simétrica**  $x \in A \Delta B$  significa  $(x \in A \cup B)$  e  $(x \notin A \cap B)$ .

$$A \Delta B = (A \cup B) \setminus (A \cap B)$$

$A$  e  $B$  são **disjuntos** se  $A \cap B = \emptyset$ .

**Teorema 1**  $\emptyset \subset C$ .

**Teorema 2 (Associatividade)**  $A \cap (B \cap C) = (A \cap B) \cap C$  e  $A \cup (B \cup C) = (A \cup B) \cup C$ .

Sejam  $A_0, A_1, A_2, \dots$  conjuntos

$x \in \bigcup_{i=0}^n A_i$  significa  $x \in A_i$  para algum  $i \in \{0, 1, \dots, n\}$ ;

$x \in \bigcup_{i \geq 0} A_i$  significa  $x \in A_i$  para algum  $i \in \mathbb{N}$ ;

$x \in \bigcap_{i=0}^n A_i$  significa  $x \in A_i$  para todo  $i \in \{0, 1, \dots, n\}$ ;

$x \in \bigcap_{i \geq 0} A_i$  significa  $x \in A_i$  para todo  $i \in \mathbb{N}$ .

## Partição

$\{A_1, A_2, \dots, A_n\}$  é uma *partição* de  $A$  se

1. as partes são subconjuntos de  $A$ , i.e.,  $A_i \subset A$  para todo  $i$ ;
2.  $A_i$  e  $A_j$  são disjuntos para quaisquer  $i \neq j$ , i.e.,  $\forall i \forall j (j \neq i \Rightarrow A_i \cap A_j = \emptyset)$
3. a união das partes resulta no todo, i.e.,  $\bigcup_{i=1}^n A_i = A$ .

## Conjunto das partes

$2^A$  denota o conjunto

$$\{B: B \subseteq A\}$$

de todos os subconjuntos de  $A$ , chamado, *conjunto das partes de  $A$* .

## Complemento

Consideremos os conjuntos  $A \subset U$  e  $B \subset U$ .

$\overline{A}^U$  denota complemento de com relação a  $U$ , i.e.,

$$\overline{A}^U = U \setminus A.$$

se não há perigo de confusão (ambiguidade) usamos simplesmente  $\overline{A}$ .

**Teorema 3** *Leis de De Morgan*

$$1. \overline{A \cup B} = \overline{A} \cap \overline{B}.$$

$$2. \overline{A \cap B} = \overline{A} \cup \overline{B}.$$

## Cardinalidade

$|A|$  denota a cardinalidade de  $A$ . Os conjuntos  $A$  e  $B$  têm a *mesma cardinalidade*, isto é  $|A| = |B|$ , se existe uma bijeção  $f: A \rightarrow B$ .

$|A| = n$ , para algum  $n$  natural, se existe uma bijeção  $f: \{1, 2, \dots, n\} \rightarrow A$ . Nesse caso, dizemos que  $A$  é **finito**. Se  $|A| > n$  para todo natural  $n$ , então  $A$  é **infinito**.

**Teorema 4** *Se  $A$  é finito então  $|2^A| = 2^{|A|}$ .*

**Teorema 5**  $|A| \neq |2^A|$ .

$A$  é *enumerável* se é finito ou se tem a mesma cardinalidade de  $\mathbb{N}$ , isto é  $|A| = |\mathbb{N}|$ . Em outras palavras,  $A$  é enumerável se existe uma função injetiva  $f: A \rightarrow \mathbb{N}$ .

**Teorema 6**  $\mathbb{N}$ ,  $\mathbb{Z}$  e  $\mathbb{Q}$  são enumeráveis.  $\mathbb{R}$  não é enumerável.

## Produto cartesiano

$A \times B$  denota o conjunto  $\{(a, b): a \in A \text{ e } b \in B\}$ . De um modo geral, se  $A_1, A_2, \dots, A_n$  são conjuntos

$$\prod_{i=1}^n A_i = A_1 \times A_2 \times \dots \times A_n = \{(a_1, a_2, \dots, a_n): a_i \in A_i (\forall i)\}. \quad (1)$$

**Teorema 7** *Para conjuntos finitos  $\left| \prod_{i=1}^n A_i \right| = |A_1| \cdot |A_2| \cdot \dots \cdot |A_n|$ .*

No caso em que os conjuntos  $A_1, A_2, \dots, A_n$  são iguais a  $A$  denotamos  $\prod_{i=1}^n A_i = A^n$ .

## §1 Princípios de Análise Combinatória

Em probabilidade muitas vezes é importante contar de quantas maneiras diferentes um determinado evento pode ocorrer. A seguir, descrevemos brevemente princípios de contagem.

**(1) Princípio Aditivo** Suponha que nós queremos comprar um computador de um dos dois fabricantes mais comuns de processadores: Intel e AMD. Suponha também fabricantes mais comuns de processadores: Intel e AMD. Suponha também que nosso orçamento nos faz ter 12 opções de modelos Intel e 18 opções de modelos AMD. Então, existem no total  $12 + 18 = 30$  opções diferentes de modelos de computador que podemos comprar.

Princípio Aditivo: Suponha que a tarefa 1 pode ser realizada de  $n$  maneiras (ou que há  $n$  desfechos possíveis para o evento 1) e a tarefa 2 de  $m$  maneiras (ou que há  $m$  desfechos possíveis para o evento 2). Se não houver nenhuma maneira de fazer as duas tarefas ao mesmo tempo (ou que ambos os eventos ocorram), então o número de maneiras de fazer uma das tarefas (ou, o número de possíveis desfechos para a ocorrência de um dos eventos) é  $n + m$ .

Formalmente, se  $T_1$  e  $T_2$  são conjuntos finitos disjuntos então  $|T_1 \cup T_2| = |T_1| + |T_2| = n + m$ . De modo geral, se  $T_1, T_2, \dots, T_k$  são conjuntos finitos disjuntos então

$$\left| \bigcup_{i=1}^k T_i \right| = \sum_{i=1}^k |T_i|.$$

**Exercício 1** Prove que se  $T_1$  e  $T_2$  são conjuntos finitos então

$$|T_1 \cup T_2| = |T_1| + |T_2| - |T_1 \cap T_2|.$$

**Exercício 2** Uma carta é escolhida de um maço de cartas de um baralho tradicional. De quantas maneiras podem ocorrer cada um dos eventos

1. escolher um rei ou uma rainha;
2. escolher uma carta de copa, ou de ouro, ou de pau;
3. escolher um número par ou uma espada;
4. escolher um rei ou uma carta preta.

**(2) Princípio Multiplicativo** Digamos que as únicas roupas limpas que você tem são duas camisetas e quatro calças. Quantas combinações diferentes de camiseta e calça você pode escolher? São duas camisetas e pra cada escolha de camiseta é possível escolher uma de quatro calças, portanto, são  $2 \times 4 = 8$  combinações possíveis.

**Princípio Multiplicativo:** Se uma tarefa é composta de duas etapas, a etapa 1 pode ser realizada de  $n$  maneiras (ou que há  $n$  desfechos possíveis para o evento 1) e a tarefa 2 de  $m$  maneiras (ou que há  $m$  desfechos possíveis para o evento 2), então o número de maneiras distintas de realizar a tarefa (ou, o número de possíveis desfechos) é  $n \cdot m$ .

Formalmente, se  $E_1$  e  $E_2$  são conjuntos finitos então  $|E_1 \times E_2| = |E_1| \cdots |E_2| = n \cdot m$ . De um modo geral, se  $E_1, \dots, E_r$  representam  $r$  etapas de uma tarefa, então o número de modos distintos de realizar a tarefa é

$$\left| \prod_{i=1}^r E_i \right| = |E_1| \cdot |E_2| \cdots |E_r|.$$

**Exercício 3** Quantos números distintos de celulares com DDD 11 podem haver, quando cada celular é identificado com um número de 8 dígitos?

**Exercício 4** Uma placa de carro é uma sequência de 3 letras seguidas por 4 dígitos. Qual é a quantidade de placas distintas que podemos obter?

**Exercício 5** Use o princípio multiplicativo para mostrar que um conjunto com  $n$  elementos tem  $2^n$  subconjuntos distintos.

**Exercício 6** De quantas maneiras distintas podemos distribuir 3 bolas distintas, digamos que uma é azul, uma verde e a outra vermelha, em 3 caixas?

**Exercício 7** De quantas maneiras distintas podemos distribuir  $r$  bolas distintas em  $n$  caixas?

**(3) Arranjos** Quantas palavras de 3 letras distintas do alfabeto latino podem ser formadas? Como o alfabeto tem 26 letras, pelo princípio multiplicativo são  $26 \cdot 25 \cdot 24 =$  palavras.

Um **arranjo** de  $r$  elementos tomados de um conjunto  $A$  de  $n$  elementos ( $r \leq n$ ) é uma sequência  $(a_1, a_2, \dots, a_r)$  de elementos não repetidos de  $A$ .

**Arranjos:** O número de arranjos de  $r$  elementos tomados de um conjunto de  $n$  elementos ( $r \leq n$ ) é

$$(n)_r \stackrel{\text{def}}{=} n(n-1)(n-2) \cdots (n-r+1) = \frac{n!}{(n-r)!}. \quad (2)$$

**Exercício 8** Quantos inteiros entre 100 e 1.000 possuem todos os dígitos distintos?

**Exercício 9** Mostre que numa população de  $n$  elementos podemos fazer uma lista de  $r$  elementos de  $n^r$  maneiras distintas caso seja permitido repetição e de  $(n)_r$  maneiras caso não haja repetição.

**Exercício 10** De quantas maneiras distintas podemos distribuir  $r$  bolas distintas em  $n$  caixas ( $r \leq n$ ) de modo que cada caixa receba no máximo 1 bola?

Alguns textos usam a notação  $A(n, r)$ .

**Permutação** Uma **permutação** de elementos de um conjunto  $A$  é uma sequência de elementos de  $A$ . Em outras palavras, é um arranjo com  $r = n$ .

Permutações: O número de permutações de elementos tomados de um conjunto de  $n$  elementos é

$$n! \stackrel{\text{def}}{=} n(n-1)(n-2) \cdots 1. \quad (3)$$

**Convenção:**  $0! = 1$ .

**Exercício 11** O número de permutações possíveis com as letras  $a, b, c, d$  e  $e$  é  $5! = 120$ . O número de permutações possíveis com as letras  $a, b, c, d, e$  e  $f$  é  $6! = 720$ . Qual é o número de permutações possíveis com as letras da palavra *banana*?

**Exercício 12** Um sinal é composto por nove bandeiras alinhadas. Quantos sinais diferentes é possível formar com há disponíveis 4 bandeiras brancas, três bandeiras vermelhas e duas bandeiras azuis? Bandeiras da mesma cor são idênticas.

**(4) Combinação** Tomemos um arranjo de  $r$  elementos escolhidos de um conjunto com  $n$  elementos. A quantidade de arranjos que têm os mesmos  $r$  elementos é  $r!$  pois a única diferença entre eles é a ordem com que se apresentam os  $r$  elementos. Quando essa ordem não importa o que temos são **combinações**. Uma combinação de  $r$  elementos escolhidos de um conjunto  $A$  com  $n$  elementos é simplesmente um subconjunto de  $A$ .

Combinações: O número de combinações de  $r$  elementos tomados de um conjunto de  $n$  elementos é

$$\binom{n}{r} \stackrel{\text{def}}{=} \frac{(n)_r}{r!} = \frac{n!}{r!(n-r)!}. \quad (4)$$

**Exercício 13** Prove que

$$(x + y)^n = \sum_{i=0}^n \binom{n}{i} x^i y^{n-i}.$$

**Exercício 14** Numa população com  $n$  elementos,  $n_1$  são azuis e  $n_2 = n - n_1$  são verdes. De quantas maneiras podemos escolher  $k$  elementos com  $r$  deles azuis? ( $0 \leq r \leq \min\{n_1, k\}$ )

**Exercício 15** De quantas maneiras distintas podemos distribuir  $r$  bolas idênticas em  $n$  caixas ( $r \leq n$ ) de modo que cada caixa receba no máximo 1 bola?

(5) **Soluções inteiras de equações** De quantas maneiras distintas podemos distribuir  $r$  bolas idênticas em  $n$  caixas? Como as bolas são indistinguíveis, só interessa quantas bolas há em cada caixa. Se a caixa  $i$  recebe  $x_i$  bolas então

$$x_1 + x_2 + \cdots + x_n = r. \quad (5)$$

Dessarte o número de modos de distribuir as bolas nas caixas é igual a quantidade de soluções inteiras e não-negativas da equação acima.

Por exemplo, para distribuir 5 bolas idênticas em 7 caixas podemos representar bolas por  $*$  e caixas pelos separadores  $|$  de modo que

$$* || | * * | * | * |$$

significa  $x_1 = 1$  (1 bola na 1ª caixa),  $x_2 = x_3 = 0$  (2ª e 3ª caixas vazias)  $x_4 = 2$ ,  $x_5 = x_6 = 1$  e  $x_7 = 0$ . A quantidade de tais padrões distintos é

$$\binom{7+5-1}{5}.$$

**Exercício 16** Mostre que a equação (5) tem

$$\binom{n+r-1}{r}.$$

*soluções inteiras e não-negativas.*

## §2 Introdução à Probabilidade

*Probabilidade*, intuitivamente, é uma forma quantitativa de expressar o conhecimento ou crença de que um evento ocorra e que está, normalmente, vinculada a interpretações como, por exemplo a **clássica** que trata os eventos elementares como igualmente prováveis; a **frequentista** que trata de probabilidades quando se lida com experimentos que são imprevisíveis, bem definidos, e podem ser repetidos, a probabilidade é a frequência relativa de ocorrência do evento; a **subjetiva** que atribui uma probabilidade a qualquer evento, mesmo quando não há aleatoriedade envolvida, a qual representa um grau de crença, quanto maior a probabilidade mais plausível é sua ocorrência.

Felizmente, para quase todas as interpretações importantes *as mesmas regras e propriedades matemáticas se aplicam*, independentemente de interpretação, embora a escolha tenha grandes implicações pelo modo em que a probabilidade é usada para modelar o mundo real. Nós estamos interessados, nessas regras que independem de interpretações.

Antes de começarmos vejamos dois exemplos, o primeiro é uma história real da época em que começou-se a estudar probabilidade. O segundo exemplo é uma aplicação moderna do dia-a-dia de pessoas que usam conexão segura na internet.

**Exemplo 1 (Chevalier de Méré)** *Antoine Gombaud, o Chevalier de Méré, viveu na França do séc. 17. Era jogador que pensou ter descoberto uma maneira de ganhar dinheiro em um jogo de dados. A sua estratégia era a seguinte*

- em 4 lançamentos de um dado o 6 ocorre pelo menos uma vez;

a idéia é que se em um lançamento o 6 ocorre com probabilidade  $1/6$  então em 4 lançamentos a chance é 4 vezes maior, ou seja,  $4/6 = 2/3$ . Assim, no longo prazo, a cada 3 apostas ele vence 2.

Essa essa estratégia o fez prosperar, até que resolveu testar outra aposta

- em 24 lançamentos de dois dados ocorre um par de 6, pelo menos uma vez;

a idéia é a mesma, a chance de um par de 6 em um lançamento é  $1/36$  e em 24 lançamentos  $24/36 = 2/3$ . Custou-lhe a falência.

Chevalier de Méré levou seu problema para o matemático Blaise Pascal, que conjuntamente com seu amigo Pierre de Fermat respondeu o problema lançando os fundamentos teoria da probabilidade.

Aprenderemos que a probabilidade de sair 6 pelo menos uma vez em 4 lançamentos é  $1 - (5/6)^4 \approx 0,51$ , mas a probabilidade de sair par de 6 pelo menos uma vez em 24 lançamentos é  $1 - (35/36)^{24} \approx 0,49$ . A razão da prosperidade no primeiro caso é a mesma que os cassinos usam em seu favor, apesar dos jogos envolverem aleatoriedade (não há dúvida de que os cassinos sempre lucram com as apostas) e a explicação probabilística é a **Lei dos Grandes Números**.

**Exemplo 2 (Comércio eletrônico)** Numa transação eletrônica entre um Consumidor (C) e o comércio eletrônico em [www.vendoenaoentrego.com.br](http://www.vendoenaoentrego.com.br) (S) há troca de vários dados, como o número do cartão de crédito, que o consumidor espera estarem protegidos sabendo que o tráfego na rede pode ser facilmente capturado. Simplificadamente, a comunicação dá como segue

1. S sorteia dois números primos  $P$  e  $Q$ , determina  $N = P \cdot Q$ ;
2. S determina um par de números, ditos chaves,  $(e, d)$  com o propriedades de que  $e \cdot d \bmod (P-1)(Q-1) = 1$ ;
3. S envia  $(N, e)$  pra C (esses números são de conhecimento público) mas os números  $P, Q$  e  $d$  devem ser mantidos em segredo;
4. se  $X$  é o número do cartão de crédito de C, então C envia  $M$  para S dado por  $M = X^e \bmod N$ ;
5. S recupera o número  $X$  da seguinte forma:  $X = M^d \bmod N$ .

Um espião, que conhece os parâmetros públicos  $N$  e  $e$ , e que captura  $M$  poderia tentar descobrir  $d$ , ou tentar fatorar  $N$ , ou tentar determinar  $\sqrt[d]{M} \bmod N$ , mas levaria mais de uma vida para conseguir fazê-lo com “força bruta”.

O que foi descrito acima é o protocolo para usar o algoritmo conhecido como RSA para criptografar e enviar  $X$  a um terceiro; mais explicações sobre esse algoritmo e porque ele funciona estão aqui. O fato que merece alguma explicação agora é onde entra probabilidade nessa história. São dois pontos: (1) por razões de segurança  $P$  e  $Q$  devem ser inteiros escolhidos aleatoriamente; esses números têm que ser grandes e qualquer pista de como eles são escolhidos pode ajudar a fatorar  $N$



*rapidamente, o que fura a criptografia; (2) os números escolhidos aleatoriamente devem ser primos e testes primalidade determinísticos são lentos, na prática são usados testes probabilísticos que podem errar com uma probabilidade muito pequena.*

*Você pode gerar um par de chaves ( $e, d$ ) usando, por exemplo, o programa `openssl` da seguinte forma: Num terminal de linha de comando do sistema operacional, o comando*

```
openssl genrsa -out private_key.pem 32
```

*para gerar um par de chaves de 32 bits (até 1024 bits não serve mais pra criptografia), tal comando devolve*

```
Generating RSA private key, 32 bit long modulus
.+++++
.+++++
e is 65537 (0x10001)
```

*cada + significa que o número sorteado passou no teste de Miller-Rabin, um teste probabilístico de primalidade. Para conhecer os parâmetros resultantes dessa operação basta digitar*

```
openssl rsa -text -in private_key.pem
```

*que devolve*

```
Private-Key: (32 bit)
modulus: 2883180901 (0xabd9d965)
publicExponent: 65537 (0x10001)
privateExponent: 386553605 (0x170a5705)
prime1: 58403 (0xe423)
prime2: 49367 (0xc0d7)
```

## Modelo probabilístico

O problema de Monty Hall é um problema que surgiu a partir de um concurso televisivo dos Estados Unidos chamado *Let's Make a Deal*, exibido na década de 1970. O jogo consiste no seguinte: Monty Hall (o apresentador) apresentava 3 portas a um concorrente, sabendo que atrás de uma delas, escolhida ao acaso, está um carro e que as outras duas têm um bode. O protocolo da brincadeira é:

1. Na 1ª etapa o concorrente escolhe uma porta (que ainda não é aberta);
2. em seguida Monty Hall abre uma das outras duas portas que o concorrente não escolheu, sabendo que ela esconde um bode;
3. em seguida, com duas portas fechadas apenas, e sabendo que o carro está atrás de uma delas, o apresentador oferece ao concorrente a oportunidade de trocar de porta.
4. O concorrente tem que se decidir se permanece com a porta que escolheu no início do jogo se muda para a outra porta que ainda está fechada;
5. feita a escolha, o apresentador abre a porta escolhida e o concorrente leva o prêmio escondido pela porta.

O problema é determinar a estratégia (trocar ou não trocar no passo 4) que maximiza a chance de ganhar o carro. Teste o jogo [aqui](#). Assista ao show [aqui](#).

Chamamos de *experimento* qualquer processo que nos fornece um resultado, particularmente, estamos interessados em experimentos idealizados e que quando repetido sob as mesmas condições não necessariamente fornece o mesmo resultado, por exemplo, o lançamento de um dado. O **espaço amostral** de um experimento, denotado por  $\Omega$ , é conjunto de *todos* os resultados possíveis do experimento.

**Exemplo 3** São experimentos com os respectivos espaços amostrais

1. lançamento de dado

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

2. lançamento de moeda

$$\Omega = \{\text{Ca}, \text{Co}\}$$

3. lançamento de moeda até sair Coroa

$$\Omega = \{(\text{Co}), (\text{Ca}, \text{Co}), (\text{Ca}, \text{Ca}, \text{Co}), (\text{Ca}, \text{Ca}, \text{Ca}, \text{Co}), \dots\}$$

4. a altura em metros de um brasileiro escolhido ao acaso

$$\Omega = \{h \in \mathbb{R} : h > 0\}$$

5. sortear um ponto no círculo de raio 1 e centro na origem do plano cartesiano

$$\Omega = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$$

Notemos que os resultados do experimento 4 estão superestimado no sentido de que, por exemplo, não é possível ocorrência do resultado 20,48m. O que é *importante* no espaço amostral é que

- para cada resultado do experimento existe um, e só um, elemento de  $\Omega$  associado a ele;
- resultados diferentes estão associados a elementos diferentes de  $\Omega$ .

Se o espaço amostral é enumerável então ele é chamado de **espaço amostral discreto**. Caso contrário, o espaço é chamado de **espaço amostral contínuo**.

**Exemplo 4** São espaços discretos os espaços amostrais dos experimentos 1, 2 e 3 dados no exemplo 3 acima. Os experimentos 4 e 5 têm espaços contínuos.

Um subconjunto de  $\Omega$  é chamado de **evento**, e em especial

- $\emptyset$  é o evento *impossível*.

- $\Omega$  é o evento *certo*.
- $\{\omega\}$  é um evento *elementar* para cada elemento  $\omega \in \Omega$ .
- o *complemento* do evento  $A$  é o evento *não-A* dado por

$$\bar{A} = \Omega \setminus A = \{\omega \in \Omega: \omega \notin A\}.$$

Na realização de um experimento o evento  $A$  *ocorre* se o resultado é um elemento de  $A$ , por conseguinte, o evento complementar de  $A$  é o evento em que  $A$  *não* ocorre.

**Exemplo 5** No lançamento de dados,  $\Omega = \{1, 2, 3, 4, 5, 6\}$ , são eventos

- $A = \{2, 4, 6\}$ , i.e.,  $A$  é o evento “ocorre número par”;
- $\bar{A} = \{1, 3, 5\}$ , i.e.,  $\bar{A}$  é o evento “não ocorre número par”;
- $B = \{4, 5, 6\}$ , i.e.,  $B$  é o evento “ocorre um resultado  $> 3$ ”;
- $C = \{4\}$ , i.e.,  $C$  é o evento “ocorre o resultado 4”;

também são eventos

- $A \cap \bar{A} = \emptyset$ , i.e.,  $A \cap \bar{A}$  é o evento “ocorre um número par e ocorre um número ímpar”, que é o evento impossível;
- $A \cup \bar{A} = \Omega$ , i.e.,  $A \cup \bar{A}$  é o evento “ocorre par **ou** ocorre ímpar”, que é o evento certo;
- $B \cap C = \{4\}$ , i.e.,  $B \cap C$  é o evento “ocorre um número  $> 3$  e ocorre 4”;
- $B \cap A = \{4, 6\}$ , i.e.,  $B \cap A$  é o evento “ocorre número  $> 3$  e ocorre número par”.

Notação	Conjunto	Evento
$\Omega$	universo	espaço amostral, evento certo
$\emptyset$	conjunto vazio	evento impossível
$\omega$	elemento	evento elementar, resultado
$A$	subconjunto	evento $A$ ocorre
$\bar{A}$	complemento com relação a $\Omega$	evento $A$ não ocorre
$A \cap B$	intersecção	ocorre $A$ e $B$
$A \cup B$	união	ocorre $A$ ou $B$
$A \Delta B$	diferença simétrica	ocorre $A$ ou $B$ , não ambos
$A \subset B$	inclusão	se ocorre $A$ , então ocorre $B$

**Exercício 17** Sejam  $A$ ,  $B$  e  $C$  eventos. Determine expressões para

1. somente  $A$  ocorre;
2.  $A$  e  $B$  mas não  $C$  ocorrem;

3. os três eventos ocorrem;
4. pelo menos um evento ocorre;
5. pelo menos dois eventos ocorrem;
6. exatamente um evento ocorre;
7. exatamente dois eventos ocorrem;
8. nenhum evento ocorre;
9. não mais que dois eventos ocorrem.

$A$  e  $B$  são eventos **disjuntos** ou **mutuamente exclusivos** quando não têm elementos em comum, isto é,  $A \cap B = \emptyset$ . Uma sequência de eventos é dita de eventos **mutuamente exclusivos** se os eventos são mutuamente exclusivos quando tomados dois-a-dois.

Uma **medida de probabilidade** é uma função que atribui para os eventos de um espaço amostral fixo,  $A \subset \Omega$ , um real  $\mathbb{P}(A)$  satisfazendo

$$\mathbf{A1} \quad 0 \leq \mathbb{P}(A) \leq 1;$$

$$\mathbf{A2} \quad \mathbb{P}(\Omega) = 1;$$

$$\mathbf{A3} \quad \mathbb{P}\left(\bigcup_{i \geq 1} A_i\right) = \sum_{i \geq 1} \mathbb{P}(A_i) \text{ se } \{A_i\}_{i \geq 1} \text{ é qualquer sequência de eventos mutuamente exclusivos.}$$

**Exemplo 6 (probabilidade do evento impossível)** A probabilidade de  $\emptyset$  é 0 pois fazendo  $A_1 = \Omega$  e  $A_i = \emptyset$  para todo  $i \geq 2$  temos por **A3** que

$$\mathbb{P}(\Omega) = \mathbb{P}(\Omega \cup \emptyset \cup \emptyset \cup \dots \cup \emptyset \cup \dots) = \mathbb{P}(\Omega) + \sum_{i \geq 0} \mathbb{P}(\emptyset)$$

portanto, por **A1** resta que  $\mathbb{P}(\emptyset) = 0$ .

**Exemplo 7 (probabilidade de uma união finita de eventos disjuntos)** Se  $A_1, A_2, \dots, A_n$  é qualquer sequência de eventos mutuamente exclusivos então

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mathbb{P}(A_i) \tag{6}$$

pois podemos fazer  $A_i = \emptyset$  para todo  $i > n$  e por **A3**

$$\begin{aligned} \mathbb{P}\left(\bigcup_{i=1}^n A_i\right) &= \mathbb{P}\left(\bigcup_{i \geq 1} A_i\right) = \sum_{i \geq 1} \mathbb{P}(A_i) = \sum_{i=1}^n \mathbb{P}(A_i) + \sum_{i > n} \mathbb{P}(\emptyset) \\ &= \sum_{i=1}^n \mathbb{P}(A_i) \end{aligned}$$

Decorrem dos axiomas acima as seguintes propriedades

**(P1)**  $\mathbb{P}(E) + \mathbb{P}(\bar{E}) = 1$

**(P2)** se  $A \subseteq B$  então  $\mathbb{P}(A) \leq \mathbb{P}(B)$ ;

**(P3) regra da adição**  $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$ .

**Exercício 18** *Prove que valem tais propriedades dadas acima.*

### Modelo probabilístico discreto

No caso de espaço amostral discreto, todo experimento tem seu modelo probabilístico especificado quando estabelecemos

- O espaço amostral  $\Omega = \{\omega_1, \omega_2, \dots\}$
- A probabilidade

$$\mathbb{P}(\omega) \stackrel{\text{def}}{=} \mathbb{P}(\{\omega\})$$

para cada evento elementar de modo que:

**(A1')**  $0 \leq \mathbb{P}(\omega_i) \leq 1$ , para todo  $i$ , e

**(A2')**  $\mathbb{P}(\Omega) = \mathbb{P}(\{\omega_1, \omega_2, \dots\}) = \sum_{i \geq 1} \mathbb{P}(\omega_i) = 1$ .

Se  $A$  é um evento, então

$$\mathbb{P}(A) \stackrel{\text{def}}{=} \sum_{\omega_i \in A} \mathbb{P}(\omega_i) \quad (7)$$

**Exemplo 8** *No caso do lançamento de um dado é usual atribuírmos a probabilidade 1/6 a cada uma das faces, o que é interpretado como todas as faces serem equiprováveis. A partir daí é natural definir para qualquer evento  $A$  a probabilidade*

$$\mathbb{P}(A) = \frac{|A|}{6}$$

*que é o modo clássico de interpretar probabilidade no caso finito.*

**Exercício 19 (eventos equiprováveis em espaços finitos)** *Prove que se  $\Omega$  é finito e todos eventos elementares têm a mesma probabilidade de ocorrer então para todo  $A \subset \Omega$  vale que  $\mathbb{P}(A) = \frac{|A|}{|\Omega|}$ .*

**Exemplo 9** *Escolhemos um inteiro positivo ao acaso, a probabilidade de escolher  $i$  é  $(\frac{1}{2})^i$ . Estendamos a probabilidade a qualquer evento  $A$  pondo*

$$\mathbb{P}(A) = \sum_{a \in A} \mathbb{P}(\{a\}).$$

*A probabilidade de escolher um número par é*

$$\sum_{\substack{a \in A \\ a \text{ par}}} \mathbb{P}(\{a\}) = \sum_{k \geq 1} \left(\frac{1}{2}\right)^{2k} = \sum_{k \geq 1} \left(\frac{1}{4}\right)^k = \frac{1}{3}.$$

**Exemplo 10** Um casal é escolhido ao acaso e  $(i, j) \in \mathbb{N} \times \mathbb{N}$  representa o número de filhos ( $i$ ) e o número de filhas ( $j$ ) do casal. Admitamos que

$$\mathbb{P}(\{(i, j)\}) = \frac{1}{2^{i+j+2}}$$

qual é a probabilidade de um casal não ter filho? O evento é dado por  $A = \{(0, j) : j \in \mathbb{N}\}$  e

$$\mathbb{P}(A) = \sum_{j \geq 0} \frac{1}{2^{j+2}} = \frac{1}{2}.$$

Você acha isso um modelo realista?

**Exercício 20** Verifique que  $\mathbb{P}$  dos exemplos anteriores é uma medida de probabilidade. Também, prove que  $\mathbb{P}$  como definida em (7), de acordo com **A1'** e **A2'**, é uma medida de probabilidade.

**Exemplo 11 (Lançamento(s) de uma moeda equilibrada)** Uma moeda equilibrada é lançada

- $\Omega = \{\text{Ca}, \text{Co}\};$
- $\mathbb{P}(\text{Ca}) = \mathbb{P}(\text{Co}) = 1/2;$

Uma moeda equilibrada é lançada até sair coroa

- $\Omega = \{\omega_1, \omega_2, \dots\}$ , onde  $\omega_i = c_1 c_2 \dots c_i$  com  $c_j = \begin{cases} \text{Co} & \text{se } j = i \\ \text{Ca} & \text{se } 1 \leq j < i \end{cases}$
- $\mathbb{P}(\omega_i) = \left(\frac{1}{2}\right)^i.$

Notemos que

$$\sum_{\omega_i \in \Omega} \mathbb{P}(\omega_i) = \sum_{i \geq 1} 2^{-i} = 1$$

além disso,  $0 < \mathbb{P}(\omega_i) < 1$  para todo  $\omega_i$ .

### Modelo probabilístico contínuo

No caso contínuo temos um pouco mais de trabalho, como mostra o seguinte exemplo.

**Exemplo 12** Consideremos o experimento 5 do exemplo 3 e interpretemos a probabilidade de um evento  $A \subset \Omega$  como a área de  $A$  proporcionalmente a de  $\Omega$ , i.e.,

$$\mathbb{P}(A) = \frac{\text{Área}(A)}{\pi}.$$

Como não é possível definir área para todo subconjunto do plano, alguns eventos não tem uma probabilidade associada.

No caso geral, exigimos que  $\mathbb{P}$  esteja definida sobre uma família  $\mathcal{A}$  de todos os eventos aos quais são atribuídos uma probabilidade, chamados de **eventos mensuráveis**, que deve satisfazer

1.  $\Omega \in \mathcal{A}$ ;
2. se  $A \in \mathcal{A}$  então  $\bar{A} \in \mathcal{A}$ ;
3. se  $A_i \in \mathcal{A}$  para todo  $i \geq 1$ , então  $\bigcup_{i \geq 1} A_i \in \mathcal{A}$ ;

uma família de subconjuntos como acima é dita  $\sigma$ -álgebra de subconjuntos de  $\Omega$ .

**Exercício 21** Mostre que se  $\Omega$  é finito então  $2^\Omega$  é uma  $\sigma$ -álgebra.

### Modelo probabilístico

Sabendo que nem sempre é possível definir probabilidade para todo evento estabelecemos que um **Modelo Probabilístico** para um experimento consiste de

1. um **espaço amostral**  $\Omega$ ;
2. uma  $\sigma$ -álgebra  $\mathcal{A}$  de eventos mensuráveis;
3. uma **medida de probabilidade**  $\mathbb{P}: \mathcal{A} \rightarrow [0, 1]$ .

**Observação 1 (sobre o item 2 de Modelo Probabilístico)** *Mais forte que o fato novo descrito no exemplo 12 acima é o de que não há medida de probabilidade que possa ser definida para todo evento daquele exemplo, esse é um resultado difícil para mostrar aqui. Isso ocorre em espaços amostrais não-enumeráveis. Há casos em que não é possível definir  $\mathbb{P}(E)$  para todo  $E \subset \Omega$  quando  $\Omega$  é infinito não-enumerável.*

*Decorre daí que família  $\mathcal{A}$  é uma necessidade técnica e sua compreensão vai muito além do que precisamos e podemos compreender nesse momento. Nós não vamos nos preocupar com esse fato nessa disciplina pois para os eventos de nosso interesse a medida  $\mathbb{P}$  estará sempre definida.*

**Exemplo 13 (Modelo probabilístico — Monty-Hall)** *No caso do Monty Hall, consideremos o experimento que consiste das seguintes três etapas*

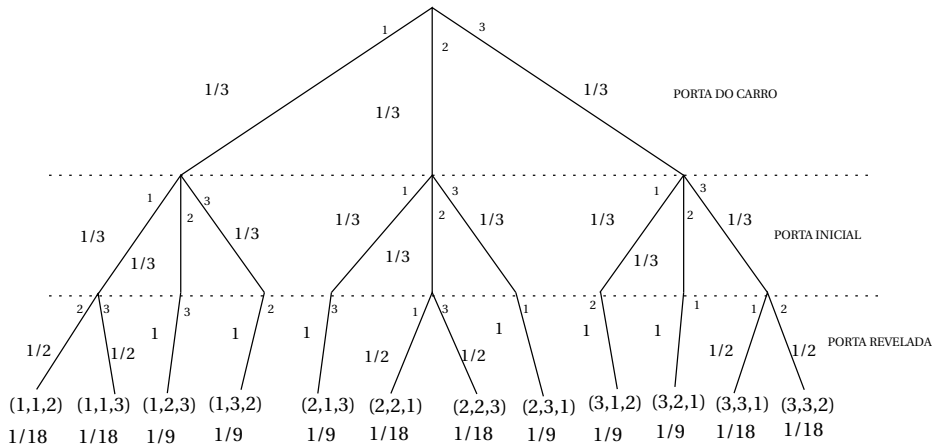
1. uma porta dentre 1,2,3 é escolhida ao acaso e o carro é colocado atrás dessa porta;
2. uma porta, dentre 1,2,3, é escolhida pelo jogador;
3. Monty revela uma porta, dentre 1,2,3, que não esconde o carro.

*O espaço amostral é*

$$\Omega = \left\{ \begin{array}{llll} (1, 1, 2), & (1, 1, 3), & (1, 3, 2), & (1, 2, 3), \\ (2, 2, 1), & (2, 2, 3), & (2, 3, 1), & (2, 1, 3), \\ (3, 3, 1), & (3, 3, 2), & (3, 1, 2), & (3, 2, 1) \end{array} \right\}$$

em que cada terna significa (porta do carro, escolha inicial, porta revelada).

Definimos uma medida de probabilidade de acordo com o diagrama a seguir



Os eventos de interesse são

$A \equiv$  "o jogador vence trocando de porta"

$A = \{(1, 2, 3), (1, 3, 2), (2, 1, 3), (2, 3, 1), (3, 1, 2), (3, 2, 1)\}$

$B \equiv$  "o carro está na porta escolhida inicialmente"

$B = \{(1, 1, 2), (1, 1, 3), (2, 2, 1), (2, 2, 3), (3, 3, 1), (3, 3, 2)\}$

$$\mathbb{P}(B) = \mathbb{P}(\{(1, 1, 2), (1, 1, 3), (2, 2, 1), (2, 2, 3), (3, 3, 1), (3, 3, 2)\}) = \frac{1}{3}.$$

**Exercício 22** *Pra pensar: No meu bolso há 3 cartas. Uma é verde em ambas as faces, outra é laranja em ambos os lados, a última tem uma face de cada uma dessas cores.*

*Retiro uma das três cartas escolhida ao acaso e um dos lados é exibido; é uma face laranja.*

*Qual é a probabilidade de que o outro lado, que está oculto, seja laranja?*

*Um argumento comum diz: "Não pode ser a carta verde-verde. Se for verde-laranja, então o outro lado é verde, enquanto se for no cartão de laranja-laranja, o outro lado é laranja. Uma vez que essas possibilidades são igualmente prováveis, o outro lado é igualmente provável de ser verde ou laranja." Essa conclusão está errada, comente.*

**Exercício 23** *Sejam  $A_1, A_2, \dots, A_n$  eventos. Prove que*

$$\begin{aligned} \mathbb{P}\left(\bigcup_{i=1}^n A_i\right) &= \sum_{i=1}^n \mathbb{P}(A_i) - \sum_{1 \leq i < j \leq n} \mathbb{P}(A_i \cap A_j) + \dots + \\ &+ (-1)^{k+1} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} \mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) + \dots + \\ &+ (-1)^{n+1} \mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_n) \end{aligned}$$



A soma

$$(-1)^{k+1} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} \mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k})$$

é feita ao longo dos  $\binom{n}{k}$  subconjuntos de  $k$  de elementos de  $\{1, 2, \dots, n\}$ .

**Exercício 24** Suponha que  $N$  pessoas chegam de chapéu em uma festa e deixam seu chapéu na entrada. Ao ir embora da festa, cada pessoa pega um chapéu ao acaso. Qual é a probabilidade que nenhuma pessoa pegue o próprio chapéu? (dica: considere o evento “o  $i$ -ésimo a sair pega o próprio chapéu” e determine a probabilidade que pelo menos uma pessoa pegue o próprio chapéu)

**Exercício 25 (continuidade de  $\mathbb{P}$ )** Seja  $A_n$ ,  $n \geq 1$ , uma sequência de eventos. Essa sequência é

- *crescente* se  $A_1 \subset A_2 \subset \dots \subset A_n \subset A_{n+1} \subset \dots$ . Nesse caso, definimos

$$\lim_{n \rightarrow \infty} A_n = \bigcup_{n \geq 1} A_n.$$

- *decrescente* se  $A_1 \supset A_2 \supset \dots \supset A_n \supset A_{n+1} \supset \dots$ . Nesse caso, definimos

$$\lim_{n \rightarrow \infty} A_n = \bigcap_{n \geq 1} A_n.$$

Prove que em ambos os casos

$$\lim_{n \rightarrow \infty} \mathbb{P}(A_n) = \mathbb{P}(\lim_{n \rightarrow \infty} A_n).$$

**Exercício 26** Uma moeda equilibrada é lançada repetidamente. Mostre que o resultado CARA ocorre com probabilidade 1.

**Problema:** Uma urna contém bolas  $k$  pretas e uma única bola vermelha. Pedro e Paula sorteiam bolas sem reposição alternadamente até que a bola vermelha é extraída. O jogo é ganho pelo jogador que jogar por último. Pedro é um cavalheiro oferece a Paula a opção de escolher se ela quer começar ou não. Paula tem um palpite de que ela tem mais chance se ela começar, afinal, ela pode ter sucesso no primeiro sorteio. Por outro lado, se o seu primeiro sorteio produz uma bola preta, então a chance de Pedro ganhar em seu primeiro sorteio é aumentada, já que a urna atem uma bola preta a menos. Como Paula deve decidir a fim de maximizar sua probabilidade de ganhar?

## Espaços de probabilidades

Probabilidade pode ser estudada do ponto de vista formal/abstrato sem se referir a experimentos e sem que os números associados aos eventos tenham qualquer interpretação. De fato, a Probabilidade é uma disciplina matemática que é estudada independentemente da natureza dos números que definem a função.

Um *espaço de probabilidade* é uma tripla  $(\Omega, \mathcal{E}, \mathbb{P})$  tal que

1.  $\Omega$  é um conjunto, chamado **espaço amostral**;
2.  $\mathcal{E}$  é uma  $\sigma$ -álgebra de subconjuntos de  $\Omega$  ditos **eventos**;
3.  $\mathbb{P} : \mathcal{E} \rightarrow \mathbb{R}$  **medida de probabilidade**.

Todo modelo probabilístico corresponde a um espaço de probabilidades e todo espaço de probabilidades corresponde a um modelo probabilístico de um experimento.

### §3 Probabilidade Condicional e Independência

Consideremos a escolha de uma família com dois filhos. Suponha que ambos os sexos para o filho mais velho e para o filho mais novo são igualmente prováveis. O espaço amostral é

$$\Omega = \{(\sigma, \sigma), (\sigma, \varphi), (\varphi, \sigma), (\varphi, \varphi)\}$$

e a probabilidade de cada evento elementar é  $\frac{1}{4}$ . Qual é a probabilidade de ambos serem meninos sabendo-se que pelo menos um deles é menino?

A **probabilidade condicional** do evento  $A$  com respeito ao evento  $E$ , em que  $\mathbb{P}(E) > 0$ , é definida por

$$\mathbb{P}(A | E) = \frac{\mathbb{P}(A \cap E)}{\mathbb{P}(E)}$$

e  $\mathbb{P}(A | E)$  é lido como a **probabilidade de  $E$  dado  $A$** .

**Exemplo 14** Se  $\Omega$  é finito e os eventos elementares equiprováveis então

$$\mathbb{P}(A | E) = \frac{\mathbb{P}(A \cap E)}{\mathbb{P}(E)} = \frac{|A \cap E|/|\Omega|}{|E|/|\Omega|} = \frac{|A \cap E|}{|E|}$$

logo, no enunciado acima, probabilidade de ambos serem meninos sabendo-se que pelo menos um deles é menino é  $\frac{1}{3}$ . Notemos que  $\mathbb{P}(A) = \frac{|A \cap \Omega|}{|\Omega|}$ .

**(P4) regra do produto** Se  $A$  e  $B$  são eventos com  $\mathbb{P}(A) > 0$  então

$$\mathbb{P}(A \cap B) = \mathbb{P}(B | A)\mathbb{P}(A).$$

Analogamente,  $\mathbb{P}(A \cap B) = \mathbb{P}(A | B)\mathbb{P}(B)$ , quando  $\mathbb{P}(B) > 0$ .

**Exemplo 15** Dois dados, um azul e outro verde, são lançados e cada uma das seis faces são equiprováveis nos dois dados. Qual é a probabilidade do dado verde ter resultado 6 dado que a soma dos resultados foi 8?

O espaço amostral tem 36 elementos equiprováveis, formado pelos pares ordenados dos resultados possíveis. Os eventos de interesse são

$$E \equiv \text{"a soma é 8"} \quad E = \{(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)\}$$

$$A \equiv \text{"no verde saiu 6"} \quad A = \{(x, 6) : 1 \leq x \leq 6, x \in \mathbb{N}\}$$

$$\mathbb{P}(A | E) = \frac{\mathbb{P}(A \cap E)}{\mathbb{P}(E)} = \frac{\mathbb{P}(\{(2, 6)\})}{\mathbb{P}(E)} = \frac{1/36}{5/36} = \frac{1}{5}.$$

**Exemplo 16** Numa cômoda há três gavetas e em cada gaveta um par de meias. Na primeira gaveta há um par de meias brancas, na segunda um par de meias pretas e na terceira gaveta um par com um pé de cada cor, preta e branca. Uma gaveta é escolhida aleatoriamente e, sem olhar para o interior da gaveta, um pé de meia é escolhido aleatoriamente e a gaveta é fechada. O pé de meia retirado é branco. Qual a probabilidade de o outro pé que ficou sozinho na gaveta ser preto?

Consideremos o espaço amostral formado pelos pares (gaveta, cor da meia retirada),

$$\Omega = \{ (1, \text{branca}), (2, \text{preta}), \\ (3, \text{preta}), (3, \text{branca}) \}$$

com os eventos elementares da primeira linha 2 vezes mais prováveis (por quê?) que os da segunda. Se  $E$  é o evento “retirou uma meia branca”, e  $A$  o evento “ficou uma meia preta”, então

$$\mathbb{P}(A | E) = \frac{\mathbb{P}(A \cap E)}{\mathbb{P}(E)} = \frac{\mathbb{P}((3, \text{branca}))}{\mathbb{P}(\{(1, \text{branca}), (3, \text{branca})\})} = \frac{1}{3}.$$

**Exemplo 17** No problema do exercício 22 os eventos que interessam são  $E \equiv$  “a face exibida é laranja” e  $A \equiv$  “a face não-exibida é laranja”. Como metade das faces são laranjas,  $\mathbb{P}(E) = 1/2$ . Só uma das cartas tem as duas faces laranjas logo  $\mathbb{P}(A \cap E) = 1/3$ , assim a probabilidade que interessa é

$$\mathbb{P}(A | E) = \frac{\mathbb{P}(A \cap E)}{\mathbb{P}(E)} = \frac{1/3}{1/2} = \frac{2}{3}.$$

Descreva detalhadamente o espaço amostral e os eventos desse experimento.

**Exercício 27 (Lewis Carroll)** Leia a seguinte “prova” de que uma urna não pode ter duas bolas da mesma cor:

Uma urna contém um par de bolas que podem ser da cor branca ( $B$ ) ou preta ( $P$ ), de modo que  $\mathbb{P}(BP) = \mathbb{P}(PB) = \mathbb{P}(PP) = \mathbb{P}(BB) = \frac{1}{4}$ . Adicionamos uma bola preta de modo que  $\mathbb{P}(PBP) = \mathbb{P}(PPB) = \mathbb{P}(PPP) = \mathbb{P}(PBB) = \frac{1}{4}$ .

Sorteamos uma bola e a probabilidade que essa bola é preta é, usando probabilidade condicional

$$1 \cdot \frac{1}{4} + \frac{2}{3} \cdot \frac{1}{4} + \frac{2}{3} \cdot \frac{1}{4} + \frac{1}{3} \cdot \frac{1}{4} = \frac{2}{3}.$$

Se a probabilidade de sortear uma bola preta é  $\frac{2}{3}$  então o par original de bolas não pode ser  $PP$  nem  $BB$ , ou seja, não podem ter a mesma cor. Faça uma análise criteriosa dessa dedução.

**Exercício 28** Mostre que se  $\mathbb{P}$  é uma medida de probabilidade e  $E$  um evento com probabilidade positiva, então a função  $\mathbb{Q}$  dada por  $\mathbb{Q}(A) = \mathbb{P}(A | E)$  é uma medida de probabilidade para os mesmos eventos mensuráveis.

Observemos que se  $E$  e  $A$  são eventos,  $0 < \mathbb{P}(E) < 1$ , então  $A$  ocorre se e só se

$A$  e  $E$  ocorre

ou

$A$  e  $\bar{E}$  ocorre

e esses eventos são disjuntos, ou seja,  $\{A \cap E, A \cap \bar{E}\}$  é uma partição de  $A$ , e de

$$A = (A \cap E) \cup (A \cap \bar{E})$$

deduzimos

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}\left((A \cap E) \cup (A \cap \bar{E})\right) \\ &= \mathbb{P}(A \cap E) + \mathbb{P}(A \cap \bar{E}) \\ &= \mathbb{P}(A | E)\mathbb{P}(E) + \mathbb{P}(A | \bar{E})\mathbb{P}(\bar{E}). \end{aligned}$$

**Lei da probabilidade total — caso geral.** Seja  $\{E_1, E_2, \dots, E_n\}$  uma partição de  $\Omega$  com  $0 < \mathbb{P}(E_i) < 1$  para todo  $i$ . Para qualquer evento  $A$

$$\mathbb{P}(A) = \sum_{i=1}^n \mathbb{P}(E_i \cap A) = \sum_{i=1}^n \mathbb{P}(A | E_i)\mathbb{P}(E_i). \quad (8)$$

**Exemplo 18 (Monty Hall revisitado)** Se o jogador não troca de porta, ele acerta o carro com probabilidade  $1/3$ .

Se o jogador troca de porta, então consideremos os eventos  $A \equiv$  “ganha o carro” e  $E \equiv$  “acerta na escolha inicial”. Então  $\mathbb{P}(E) = 1/3$ ,  $\mathbb{P}(A | E) = 0$  e  $\mathbb{P}(A | \bar{E}) = 1$ . Pela lei da probabilidade total

$$\mathbb{P}(A) = \mathbb{P}(A | E)\mathbb{P}(E) + \mathbb{P}(A | \bar{E})\mathbb{P}(\bar{E}) = \frac{2}{3}$$

portanto, é melhor trocar de porta.

**Exemplo 19** Uma fábrica de sorvete compra galões de leite de 3 fazendas diferentes.

A fazenda F1 entrega todo dia 20% do leite usado e uma inspeção revelou que 20% desse leite estava adulterado.

A fazenda F2 entrega 30% do leite usado e a inspeção revelou que 5% estava adulterado.

A fazenda F3 entrega 50% do leite usado, 2% adulterado.

Os galões são indistinguíveis. Um é escolhido ao acaso. Qual a probabilidade do evento

$A \equiv$  “ter leite adulterado”?

O espaço amostral  $\Omega$  representa os galões de leite e pode ser particionado como  $\Omega = F_1 \cup F_2 \cup F_3$ . As inspeções nos fornecem as probabilidades

$$\mathbb{P}(A | F_1) = 0,2, \quad \mathbb{P}(A | F_2) = 0,05 \quad e \quad \mathbb{P}(A | F_3) = 0,02$$

usando (8) determinamos

$$\mathbb{P}(A) = \mathbb{P}(A | F_1)\mathbb{P}(F_1) + \mathbb{P}(A | F_2)\mathbb{P}(F_2) + \mathbb{P}(A | F_3)\mathbb{P}(F_3) = 0,06504065.$$

Agora, dado que a amostra de leite está adulterada, qual é a probabilidade de ter sido escolhida da fazenda F1?

$$\begin{aligned} \mathbb{P}(F_1 | A) &= \frac{\mathbb{P}(F_1 \cap A)}{\mathbb{P}(A)} && \text{por definição} \\ &= \frac{\mathbb{P}(A | F_1)\mathbb{P}(F_1)}{\mathbb{P}(A)} && \text{pela regra do produto} \\ &= \frac{0,2 \cdot 0,2}{0,06504065} = 0,615. \end{aligned}$$

## Teorema de Bayes

Suponha que *probabilite* é uma doença rara que afeta apenas 0,1% da população, e que existe um teste para detectar *probabilite* e que não é perfeito: há 3% de falsos positivos e 2% de falsos negativos. Dado que o teste para um determinado paciente deu positivo, qual é a probabilidade que ele tenha de fato a doença?

A lenda diz que a maioria dos médicos respondem 97% sem pestanejar, já que há 3% de falsos positivos. Essa justificativa está errada.

Da definição de probabilidade condicional, se  $A$  e  $B$  são eventos com probabilidade positiva então

$$\mathbb{P}(A | B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B | A)\mathbb{P}(A)}{\mathbb{P}(B)}.$$

Usando a lei da probabilidade total para  $B$  com a partição  $\{A, \bar{A}\}$  temos

$$\mathbb{P}(A | B) = \frac{\mathbb{P}(B | A)\mathbb{P}(A)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B | A)\mathbb{P}(A)}{\mathbb{P}(B | A)\mathbb{P}(A) + \mathbb{P}(B | \bar{A})\mathbb{P}(\bar{A})}.$$

Essa dedução pode ser facilmente generalizada para provar o seguinte.

**Teorema 8 (Teorema de Bayes)** Sejam  $A_1, A_2, \dots, A_m$  uma **partição** de  $\Omega$ ,  $B \subset \Omega$ , todos eventos com probabilidade positiva. Para todo  $1 \leq j \leq m$

$$\mathbb{P}(A_j | B) = \frac{\mathbb{P}(B | A_j)\mathbb{P}(A_j)}{\sum_{i=1}^m \mathbb{P}(B | A_i)\mathbb{P}(A_i)}. \quad (9)$$

**Exemplo 20** Temos 3 moedas e apenas uma delas, não sabemos qual, é desbalanceada. Essa moeda desbalanceada resulta cara com probabilidade  $2/3$ . Um lançamento de cada uma delas resulta em Ca, Ca, Co. Consideremos os eventos

$E_i \equiv$  “a  $i$ -ésima moeda é a desbalanceada”;  $\mathbb{P}(E_i) = 1/3$ .

$B \equiv$  “o resultado dos lançamentos é Ca, Ca, Co”

$$\mathbb{P}(B | E_1) = \frac{2}{3} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{6}$$

$$\mathbb{P}(B | E_2) = \frac{1}{2} \times \frac{2}{3} \times \frac{1}{2} = \frac{1}{6}$$

$$\mathbb{P}(B | E_3) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{3} = \frac{1}{12}$$

Usando o Teorema de Bayes

$$\mathbb{P}(E_1 | B) = \frac{\mathbb{P}(B | E_1)\mathbb{P}(E_1)}{\mathbb{P}(B | E_1)\mathbb{P}(E_1) + \mathbb{P}(B | E_2)\mathbb{P}(E_2) + \mathbb{P}(B | E_3)\mathbb{P}(E_3)} = \frac{2}{5}$$

lembramos que  $\mathbb{P}(E_1) = 1/3$ , logo “aprendemos mais” sobre as moedas depois do lançamento. Também,  $\mathbb{P}(E_2 | B) = 2/5$  mas  $\mathbb{P}(E_3 | B) = 1/5$ .

**Exemplo 21** A probabilidade afeta 0,1% da população; há 3% de falsos positivos e 2% de falsos negativos. Dado que o teste para um determinado paciente deu positivo, qual é a probabilidade que ele tenha de fato a doença?

Os eventos de interesse são

$B \equiv$  “o exame deu positivo” e  $A \equiv$  “tem probabilidade”

Por (9)

$$\mathbb{P}(A | B) = \frac{\mathbb{P}(B | A)\mathbb{P}(A)}{\mathbb{P}(B | A)\mathbb{P}(A) + \mathbb{P}(B | \bar{A})\mathbb{P}(\bar{A})} = \frac{0,98 \cdot 0,001}{0,98 \cdot 0,001 + 0,03 \cdot 0,999}$$

a chance de ter a doença dado que o teste foi positivo é ligeiramente menor que 3,2%. Teste inútil.

**Exemplo 22** Um teste para H1N1 tem 97% de chance de responder certo. Se uma pessoa tem resultado positivo (teste+), então ela pode estar doente ou não – doente; o que podemos concluir se 60% da população que fica doente?

Usando a mesma idéia do exemplo acima

$$\begin{aligned} \mathbb{P}(\text{doente} | \text{teste}+) &= \frac{\mathbb{P}(\text{teste}+ | \text{doente})\mathbb{P}(\text{doente})}{\mathbb{P}(\text{teste}+)} = \\ &= \frac{\mathbb{P}(\text{teste}+ | \text{doente})\mathbb{P}(\text{doente})}{\mathbb{P}(\text{teste}+ | \text{doente})\mathbb{P}(\text{doente}) + \mathbb{P}(\text{teste}+ | \text{não – doente})\mathbb{P}(\text{não – doente})} = \\ &= \frac{0,97 \cdot 0,6}{0,97 \cdot 0,6 + 0,03 \cdot 0,4} \end{aligned}$$

logo a probabilidade é 0,979798.

**Exercício 29** Um país extremamente democrático mantém três prisioneiros sem qualquer acusação na prisão de Guantánamo, Said, Osama e Ayman. O carcereiro diz a eles que um será libertado e os outros serão executados, mas não lhe foi permitido dizer o que acontecerá a cada um. Said avalia que ele tem chance de  $1/3$  de escapar e pensou em perguntar em segredo ao carcereiro qual dos outros prisioneiros seria executado; em seguida avaliou que se o carcereiro respondesse o Osama, por exemplo, então a chance dele escapar passaria a ser  $1/2$  pois seria Ele e Osama, ou ele e Ayman. Essa avaliação está correta?

## Independência

Vimos nos exemplos acima, como o exemplo 20, que o conhecimento da ocorrência de um evento  $E$  afeta a probabilidade de ocorrência de  $A$ , a probabilidade  $\mathbb{P}(A | E)$  não é, em geral, igual a  $\mathbb{P}(A)$ . Quando esse não é o caso, dizemos que os eventos são independentes.

Os eventos  $A$  e  $B$  são ditos **independentes** se

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

Uma coleção de eventos  $E_1, E_2, \dots, E_n, \dots$  é dita **independente** se para toda subcoleção finita  $E_{i_1}, E_{i_2}, \dots, E_{i_k}$  vale que

$$\mathbb{P}\left(\bigcap_{\ell=1}^k E_{i_\ell}\right) = \mathbb{P}(E_{i_1})\mathbb{P}(E_{i_2}) \cdots \mathbb{P}(E_{i_k})$$

qualquer que seja  $k \in \{2, 3, \dots, n\}$ .

**Exemplo 23** Consideremos o espaço amostral com 9 elementos dado pelas 6 permutações das letras  $a, b, c$  mais as 3 ternas  $(a, a, a)$ ,  $(b, b, b)$  e  $(c, c, c)$ , cada uma com probabilidade  $1/9$ . Seja  $E_i$  o evento “a coordenada  $i$  é  $a$ ”. Então  $\mathbb{P}(E_i) = 1/3$  e  $\mathbb{P}(E_i \cap E_j) = 1/9$  para  $i \neq j$ , portanto os eventos são 2-a-2 independentes, mas não são independentes pois, por exemplo, a ocorrência de  $E_1$  e  $E_3$  implica na ocorrência de  $E_2$ ; de fato,  $\mathbb{P}(E_1 \cap E_2 \cap E_3) = 1/9$ .

**Exemplo 24** Consideremos três lançamentos de uma moeda equilibrada e os eventos

$E_{12} \equiv$  “o resultado do primeiro e do segundo coincidem”

$E_{13} \equiv$  “o resultado do primeira e do terceiro coincidem” e

$E_{23} \equiv$  “o resultado do segundo e do terceiro coincidem”.

$$\mathbb{P}(E_{12}) = \mathbb{P}(E_{13}) = \mathbb{P}(E_{23}) = \frac{1}{2}$$

Os eventos são 2-a-2 independentes,

$$E_{12} \cap E_{13} = \{(Ca, Ca, Ca), (Co, Co, Co)\}$$

com probabilidade  $1/4$ . Analogamente, para  $E_{12} \cap E_{23}$  e  $E_{13} \cap E_{23}$ .

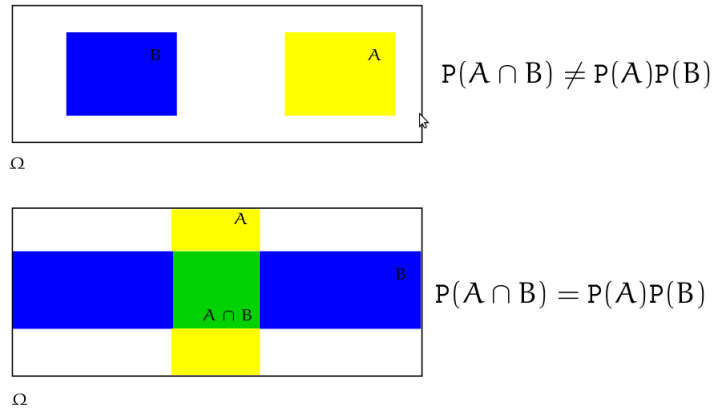
Esses eventos não são independentes pois

$$\mathbb{P}(E_{12} \cap E_{13} \cap E_{23}) = \frac{1}{4} \text{ enquanto que } \mathbb{P}(E_{12})\mathbb{P}(E_{13})\mathbb{P}(E_{23}) = \frac{1}{8}$$

**Exercício 30** Mostre que se  $A$  e  $B$  são eventos independentes, então (a)  $\bar{A}$  e  $\bar{B}$  são eventos independentes; (b)  $A$  e  $\bar{B}$  são eventos independentes; (c)  $\bar{A}$  e  $B$  são eventos independentes.

**Independência × exclusão:** Não confunda esses termos.

**Exemplo 25** Ao jogar um dado  $A \equiv$  “ocorreu par” e  $\bar{A} \equiv$  “não ocorreu par” não são evento independentes.



## Espaços produto

Se temos dois experimentos com modelos  $(\Omega_1, \mathcal{A}_1, \mathbb{P}_1)$  e  $(\Omega_2, \mathcal{A}_2, \mathbb{P}_2)$  e que ocorrem de modo independente então podemos definir um modelo probabilístico para a realização conjunta desses experimentos da seguinte forma.

O espaço amostral da realização conjunta desses experimentos é dada pelos pares  $(\omega_1, \omega_2)$ , i.e., o espaço amostral é  $\Omega_1 \times \Omega_2$ .

Os eventos mensuráveis são difíceis de definir, não é simplesmente  $\mathcal{A}_1 \times \mathcal{A}_2$ , entretanto pode ser feito e ignoraremos como isso se dá apenas chamaremos de  $\mathcal{A}$  a  $\sigma$ -álgebra conveniente para esse caso. No caso discreto, se a probabilidade está definida em todos os subconjuntos então basta tomar  $2^{\Omega_1} \times 2^{\Omega_2}$ .

Definimos a medida de probabilidade

$$\mathbb{P}(A_1 \times A_2) \stackrel{\text{def}}{=} \mathbb{P}_1(A_1)\mathbb{P}_2(A_2).$$

O espaço de probabilidade  $(\Omega_1 \times \Omega_2, \mathcal{A}, \mathbb{P})$  é chamada de **espaço produto**.

Espaços produto podem ser usado para os modelos probabilísticos dos experimentos nos exemplos 11, 13 e 16.

Isso pode ser estendido para o produto de vários espaços, por exemplo, no caso do exemplo 11 e no exemplo 26 a seguir, em que  $\Omega = \{\text{sucesso}, \text{fracasso}\}$ ,  $\mathbb{P}(\text{sucesso}) = p$ , e o modelo para tal exemplo é

$$\prod_{i \geq 1} (\Omega, 2^\Omega, \mathbb{P}).$$



**Exemplo 26** Consideremos uma sequência infinita de realizações de um experimento, cada realização independente das outras, e cada realização resulta em sucesso com probabilidade  $p$  e fracasso com probabilidade  $(1 - p)$ . Se  $E_i$  é o evento, “fracasso na  $i$ -ésima tentativa”, então a probabilidade de não haver sucesso nas  $n$  primeiras realizações é

$$\mathbb{P}(E_1)\mathbb{P}(E_2)\cdots\mathbb{P}(E_n) = (1 - p)^n$$

por causa da independência das realizações. Portanto, pelo menos 1 sucesso nas  $n$  primeiras realizações ocorre com probabilidade  $1 - (1 - p)^n$ . A ocorrência de exatamente  $k$  sucessos ocorre com probabilidade

$$\binom{n}{k} p^k (1 - p)^{n-k}.$$

**Exercício 31** No exemplo anterior, qual a probabilidade de todas as realizações resultarem sucesso?

**Exercício 32** Seja  $\Omega = \{1, 2, \dots, p\}$  para  $p$  primo e  $\mathbb{P}(A) = |A|/p$ , para todo  $A \subset \Omega$ . Prove que se  $A$  e  $B$  são independentes então pelo menos um desses eventos deve ser  $\emptyset$  ou  $\Omega$ .

**Exercício 33** Mostre que para eventos  $A$ ,  $B$  e  $C$  vale

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B | A)\mathbb{P}(C | A \cap B). \quad (10)$$

Também vale que, se  $D$  é evento então

$$\mathbb{P}(A \cap B \cap C \cap D) = \mathbb{P}(A)\mathbb{P}(B | A)\mathbb{P}(C | A \cap B)\mathbb{P}(D | A \cap B \cap C).$$

Assuma que todos os eventos envolvidos tenha probabilidade positiva para que as probabilidades condicionais estejam definidas.

**Exercício 34** Use (10) para determinar qual é a probabilidade de ocorrer três reis ao se retirar três cartas ao acaso de um baralho.

**Exercício 35 (Regra generalizada do produto)** Sejam  $A_1, A_2, \dots, A_n$  eventos tais que  $\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_n) > 0$ . Escreva uma versão geral, com relação ao exercício anterior, para

$$\mathbb{P}\left(\bigcap_{i=1}^n A_i\right).$$

**Exercício 36** Seja  $(\Omega, \mathcal{A}, \mathbb{P})$  um espaço de probabilidades e  $C \subset \Omega$  um evento com probabilidade positiva. De acordo com o exercício 28 acima a função  $\mathbb{P}_C: \mathcal{A} \rightarrow \mathbb{R}$ , definida por  $\mathbb{P}_C(A) = \mathbb{P}(A | C)$  é uma medida de probabilidade dos eventos mensuráveis de  $\Omega$ ; portanto deve satisfazer as propriedades que uma medida de probabilidade satisfaz.

Verifique que  $\mathbb{P}(\cdot | C)$  satisfaz as propriedades **P1**, **P2**, **P3**, **P4**, por exemplo, mostre que vale que se  $A \subset B$  então  $\mathbb{P}(A | C) \leq \mathbb{P}(B | C)$ .

**Problema:** Uma urna contém seis bolas, digamos que três são vermelhas e três azuis. Uma dessas bolas, chamemos-a bola  $A$ , é selecionada ao acaso removida permanentemente da urna sem que a sua cor seja revelada para um observador. Esse observador pode agora retirar bolas sucessivamente, uma de cada vez, de forma aleatória e com reposição um número pré-determinado de vezes. O objetivo dessa experimentação é a formar uma idéia aproximada da proporção de bolas vermelhas e azuis que permaneceram na urna após a retirada da bola  $A$ .

Pedro, na condição de observador, retirou uma bola seis vezes, e todos resultados foram bolas vermelhas. Paula retirou uma bola da urna 600 vezes, 303 vezes foram bolas vermelhas e 297 vezes bolas azuis. Claramente, ambos tendem a prever que a bola  $A$  é, provavelmente, azul. Qual deles, se é que algum, tem a evidência empírica mais forte de sua previsão?

### Mais exemplos

**Exemplo 27** Suponhamos que o @alunodaufabc está em dúvida entre matricular-se em Física Quântica e Introdução à Probabilidade e à Estatística. Ele estima que a probabilidade de passar com  $A$  em FQ é  $1/2$  e a de passar com  $A$  em IPE é  $2/3$ . Se ele usar uma moeda para escolher a disciplina, qual é a probabilidade de passar com  $A$  em IPE?

Se  $A$  representa “passar com  $A$ ” e  $P$  representa “cursar IPE”

$$\mathbb{P}(P \cap A) = \mathbb{P}(A | P)\mathbb{P}(P) = \frac{1}{2} \cdot \frac{2}{3} = \frac{1}{3}.$$

**Exemplo 28** Um investigador de polícia está 60% convencido da culpa de um determinado suspeito. Uma nova evidência na cena do crime prova que o criminoso tinha uma determinada característica. Sabe-se que 20% da população tem tal característica. Dado que o suspeito tem essa característica, qual é o grau de convencimento do investigador?

Representemos por  $H$  o evento “suspeito é culpado” e por  $E$  “tem a característica”, então

$$\mathbb{P}(H | E) = \frac{\mathbb{P}(E | H)\mathbb{P}(H)}{\mathbb{P}(E | H)\mathbb{P}(H) + \mathbb{P}(E | \bar{H})\mathbb{P}(\bar{H})} = \frac{1 \cdot 0,6}{1 \cdot 0,6 + 0,2 \cdot 0,4}$$

ou seja, 88,2%.

**Exemplo 29** No contexto do exemplo anterior, se  $H$  é uma hipótese e  $E$  uma evidência então

$$\mathbb{P}(H | E) = \frac{\mathbb{P}(E | H)\mathbb{P}(H)}{\mathbb{P}(E)} \quad \text{e} \quad \mathbb{P}(\bar{H} | E) = \frac{\mathbb{P}(E | \bar{H})\mathbb{P}(\bar{H})}{\mathbb{P}(E)}$$

portanto

$$\frac{\mathbb{P}(H | E)}{\mathbb{P}(\bar{H} | E)} = \frac{\mathbb{P}(H)}{\mathbb{P}(\bar{H})} \cdot \frac{\mathbb{P}(E | H)}{\mathbb{P}(E | \bar{H})} \quad (11)$$

onde  $\frac{\mathbb{P}(H)}{\mathbb{P}(\bar{H})}$  quantifica quanto provável é a ocorrência de  $H$  com relação a não ocorrência de  $H$ , portanto  $\frac{\mathbb{P}(H | E)}{\mathbb{P}(\bar{H} | E)}$  é quão mais provável é a ocorrência de  $H$  com relação a não ocorrência após a introdução da evidência  $E$ .

Portanto, se a hipótese é culpa e a evidência é uma característica do criminoso, no exemplo anterior temos

$$\frac{\mathbb{P}(E | H)}{\mathbb{P}(E | \overline{H})} = \frac{1}{0,2} = 5$$

ou seja, após a nova prova a culpa é 5 vezes mais provável.

**Exemplo 30 (Falácia do promotor)** Essa falácia geralmente resulta em assumir que a probabilidade prévia de que uma evidência condenaria um membro escolhido aleatoriamente da população é igual à probabilidade de que ela condenaria o réu. Por exemplo, você ganha na loteria, mas é acusado de fraude. No julgamento o promotor afirma que ganhar na loteria sem trapacear é extremamente improvável portanto é igualmente extremamente improvável que você seja inocente. Isso desconsidera que a probabilidade de qualquer outra pessoa ganhe na loteria é igualmente pequena.

Uma confusão típica é, usando a linguagem dos tribunais, que

se  $\mathbb{P}(\text{evidência comprovada} | \text{inocência do suspeito})$   
é pequena, então  $\mathbb{P}(\text{inocência do suspeito} | \text{evidência comprovada})$   
é pequena.

De fato, essas probabilidade podem ser muito diferentes como mostramos a seguir. No caso de teste de DNA para identificar criminosos, dependendo do número de *locus* considerado na análise, é possível que a frequência de determinado padrão seja da ordem de 1 em 1 milhão [Technology in Forensic Science]. O erro comum é a confusão resultante de tomar por aproximadamente igual

1. a probabilidade que a amostra de DNA de uma outra pessoa (não o criminoso) case com o padrão do DNA descoberto na cena do crime, dado que a pessoa é inocente;
2. a probabilidade que a pessoa é inocente dado que seu DNA casa com o padrão do DNA descoberto na cena do crime.

A primeira é a probabilidade atestada por um perito em exame de DNA; quando um teste de DNA atesta que “a probabilidade é de 1 em 100 milhões que o DNA da cena do crime case como DNA do suspeito” isso significa que “a chance do casamento de padrões entre o DNA na cena do crime e do DNA do acusado é de 1 em 100 milhões, se o acusado é inocente e não é parente do criminoso” ( $\mathbb{P}(\text{evidência} | \text{inocência})$  é pequena). A segunda probabilidade, descrita acima, é a que interessa ao promotor para poder acusar sem erro o suspeito e uma vez que o resultado positivo é certo ( $\mathbb{P}(\text{inocência} | \text{evidência})$ ), se o acusado deixou a amostra de DNA na cena do crime então

$$\frac{\mathbb{P}(\text{evidência} | \text{culpa})}{\mathbb{P}(\text{evidência} | \text{inocência})} = 100 \text{ milhões}$$

ou seja, independentemente da chance a priori de culpa, a evidência do DNA torna a probabilidade de culpa 100 milhões de vezes maior que a de inocência, de acordo com a equação (11) acima

$$\frac{\mathbb{P}(\text{culpa} \mid \text{evidência})}{\mathbb{P}(\text{inocência} \mid \text{evidência})} = 100.000.000 \frac{\mathbb{P}(\text{culpa})}{\mathbb{P}(\text{inocência})}$$

**Exemplo 31** O seguinte relato foi extraído de uma reportagem da revista *TIME* de sexta-feira, 8 de janeiro de 1965. “Por volta do meio-dia, em um dia de Junho último, uma mulher idosa foi assaltada em um beco em San Pedro, na Califórnia. Pouco depois, uma testemunha viu uma garota loira com rabo de cavalo saindo do beco e entrando em um carro amarelo conduzido por um negro com barba e bigode. A polícia prendeu Janet e Malcolm Collins, um casal que que encaixava na descrição física dada pela testemunha e que possuía um Lincoln amarelo.”

Ao apresentar o caso o promotor, que teve a assessoria de um professor de matemática local, salientou a baixa probabilidade de haver casais como os Collins em San Pedro, em um carro amarelo pois

- probabilidade de carro amarelo = 1 / 10;
- probabilidade de homem com bigode = 1 / 4;
- probabilidade de mulher com rabo de cavalo = 1 / 10;
- probabilidade mulher tem cabelo loiro = 1 / 3;
- probabilidade de homem negro com barba = 1 / 10;
- probabilidade de casal inter-racial = 1 / 1000;

Multiplicando todos esses valores leva a um valor de 1 em 12 milhões para a probabilidade de haver um casal com as mesmas características de Janet e Malcolm Collins. O casal acabou sendo condenado. Qual é o problema desse argumento?

Veja [este link](#) se quiser mais detalhes sobre as considerações legais do processo e o desdobramento. Há vários casos reais de uso incorreto de probabilidade num julgamento, veja os casos *Lucia de Berk*, *Kevin Sweeney*, *Sally Clark*.

**Exemplo 32 (Filtros bayesiano anti-spam)** A abordagem bayesiana para filtrar mensagens (uma das ferramentas do *spamassassin*, por exemplo) foi idéia de *Paul Graham* que descreve aqui como eles funcionam. Abaixo daremos uma explicação resumida e formal de como usar o teorema de Bayes na classificação de mensagens que são spam.

O espaço amostral é dado por  $\{0, 1\}^n$  de modo que cada coordenada indica se uma mensagem tem (representado por 1) ou não tem (representado por 0) uma determinada característica e a primeira coordenada de uma  $n$ -nupla, especificamente, é 1 se a mensagem é spam e 0 caso, contrário.

Assim o evento  $S$  dado por  $\{1\} \times \{0, 1\}^{n-1}$  é o evento é spam e  $\bar{S} = \{0\} \times \{0, 1\}^{n-1}$  é o evento não é spam.

Denotamos por  $C_i$ ,  $i \geq 2$ , o evento “tem a característica  $i$ ” que corresponde a todos os vetores cuja  $i$ -ésima coordenada é 1.

Vamos assumir que as características  $2, 3, \dots, n$  são independentes condicionalmente

$$\mathbb{P}\left(\bigcap_{i \in I} C_i | S\right) = \prod_{i \in I} \Pr(C_i | S)$$

$$\mathbb{P}\left(\bigcap_{i \in I} C_i | \bar{S}\right) = \prod_{i \in I} \Pr(C_i | \bar{S})$$

para todo  $I \subseteq \{2, 3, \dots, n\}$  (pode não ser uma hipótese muito realista. Por exemplo, se a característica 2 é conter a palavra *watch* e a característica 3 é conter a palavra *replica* então um email que contém *replica* tem muito mais chance de conter *watch*; o meu endereço de email recebe várias dessas mensagens que anunciam réplicas de relógios caros e *watch* só aparece nesses spams; esse pode não ser o caso de um relojoeiro num país de língua inglesa).

Pelo teorema de Bayes

$$\mathbb{P}\left(S \mid \bigcap_{i \in I} C_i\right) = \frac{\prod_{i \in I} \Pr(S | C_i)}{\prod_{i \in I} \Pr(S | C_i) + (\Pr(S) / \Pr(\bar{S}))^{|I|-1} \prod_{i \in I} \Pr(\bar{S} | C_i)}. \quad (12)$$

(Dica: use o teorema de Bayes com  $S$ ,  $\bar{S}$  e  $\bigcap_i C_i$ .)

Agora, suponha que temos 100 mensagens, 50 spams e 50 não spams. O trabalho inicial é definir características das mensagens que são spams das que não são. Por exemplo, nas minhas mensagens muitos dos spams têm a palavra *watch* enquanto que muitos dos não spams têm a palavra *reunião*; a maioria das mensagens têm a palavra *a*, tantos spams quanto não spams, logo *a* não deve ser uma característica levada em consideração.

Para a característica  $i$ , seja  $k_i$  a quantidade de spams que têm a característica  $i$  e seja  $\ell_i$  a quantidade de não spams que têm a característica  $i$ ; definimos  $\Pr(C_i | S) = \frac{k_i}{50}$  e  $\Pr(C_i | \bar{S}) = \frac{\ell_i}{50}$  e pelo teorema de Bayes,

$$\Pr(S | C_i) = \frac{\Pr(C_i | S) \Pr(S)}{\Pr(C_i | S) \Pr(S) + \Pr(C_i | \bar{S}) \Pr(\bar{S})} = \frac{k_i \Pr(S)}{k_i \Pr(S) + \ell_i \Pr(\bar{S})}.$$

Recebida uma mensagem, filtramo-la para determinar suas características, digamos  $C_i$  com  $i \in I$ , e (12) determina a probabilidade dessa mensagem ser spam.

**Exemplo 33** As seguradoras de automóveis classificam motoristas em propensos a acidentes e não propensos a acidentes; estimam que os propensos são 30% da população. As estatísticas mostram que os propensos se envolvem em acidente no período de um ano com probabilidade 0,4 e os não propensos com probabilidade 0,2. Seja  $A$  o evento definido pelos motoristas propensos. Então, a probabilidade de um novo segurado se envolver em acidente em um ano é

$$\mathbb{P}(A_1) = \mathbb{P}(A_1 | A) \mathbb{P}(A) + \mathbb{P}(A_1 | \bar{A}) \mathbb{P}(\bar{A}) = 0,4 \cdot 0,3 + 0,2 \cdot 0,7 = 0,26$$

se um novo segurado se envolve em acidente nesse prazo, a probabilidade dele ser propenso é

$$\mathbb{P}(A | A_1) = \frac{\mathbb{P}(A_1 | A) \mathbb{P}(A)}{\mathbb{P}(A_1)} = \frac{0,3 \cdot 0,4}{0,26} = \frac{6}{13}.$$

Qual a probabilidade de ocorrer um acidente no 2º ano dado que tenha acidentado no 1º ano de contrato? Definimos a medida de probabilidade, sobre os mesmos evento,  $\mathbb{Q}(\cdot) = \mathbb{P}(\cdot | A_1)$  (veja os exercícios 28 e 36). Dessa forma, pela Lei de Probabilidade Total

$$\mathbb{Q}(A_2) = \mathbb{Q}(A_2 | A)\mathbb{Q}(A) + \mathbb{Q}(A_2 | \bar{A})\mathbb{Q}(\bar{A})$$

**Exercício 37** Mostre que  $\mathbb{Q}(A_2 | A) = \mathbb{P}(A_2 | A \cap A_1)$ .

Assumiremos que  $A_1$  e  $A_2$  são condicionalmente independentes dado  $A$ , ou seja,

$$\mathbb{P}(A_2 | A \cap A_1) = \mathbb{P}(A_2 | A)$$

que equivale a  $\mathbb{P}(A_2 \cap A_1 | A) = \mathbb{P}(A_1 | A)\mathbb{P}(A_2 | A)$ .

Portanto,  $\mathbb{Q}(A_2 | A) = 0,4$  e temos que

$$\begin{aligned}\mathbb{Q}(A_2) &= \mathbb{Q}(A_2 | A)\mathbb{Q}(A) + \mathbb{Q}(A_2 | \bar{A})\mathbb{Q}(\bar{A}) \\ \mathbb{P}(A_2 | A_1) &= \mathbb{P}(A_2 | A)\mathbb{P}(A | A_1) + \mathbb{P}(A_2 | \bar{A})\mathbb{P}(\bar{A} | A_1) \\ \mathbb{P}(A_2 | A_1) &= 0,4 \cdot \frac{6}{13} + 0,2 \cdot \frac{1}{13} = 0,29.\end{aligned}$$

## §4 Variáveis Aleatórias

Se uma moeda é lançada 3 vezes, supondo os resultados independentes, qual é o número de caras ocorridas? Qual é a probabilidade de termos 2 caras?

resultado 1	resultado 2	resultado 3	Nº de caras
Ca	Ca	Ca	3
Ca	Ca	Co	2
Ca	Co	Ca	2
Ca	Co	Co	1
Co	Ca	Ca	2
Co	Ca	Co	1
Co	Co	Ca	1
Co	Co	Co	0

A probabilidade de ocorrerem 2 caras é  $\frac{3}{8}$ .

Muitas vezes estamos mais interessados numa característica numérica de um evento, mais do que no evento propriamente dito.

**Exemplo 34** Vejamos os seguintes exemplos de experimentos e o característico numérico de interesse.

1. Considera-se o número de chamadas telefônicas que chegam a uma central em um intervalo de tempo.
2. Escolher um ponto no círculo unitário e determinar a distância até a origem.

3. Mede-se a altura de um cidadão escolhido ao acaso.
4. Retira-se uma lâmpada da linha de produção, acende-a e observa-se a mesma até que se queime.

Uma **variável aleatória** (v.a.) é uma função que associa a cada elemento de  $\Omega$  um número real

$$X: \Omega \rightarrow \mathbb{R}.$$

**Exemplo 35** Se  $X$  é o número de caras em 3 lançamentos de uma moeda então

$$X((Ca, Ca, Ca)) = 3$$

$$X((Ca, Ca, Co)) = 2$$

$$X((Ca, Co, Ca)) = 2$$

$$X((Ca, Co, Co)) = 1$$

$$X((Co, Ca, Ca)) = 2$$

$$X((Co, Ca, Co)) = 1$$

$$X((Co, Co, Ca)) = 1$$

$$X((Co, Co, Co)) = 0$$

Para uma v.a.  $X: \Omega \rightarrow \mathbb{R}$  usaremos as seguintes notações

$$[X = a] \text{ denota o evento } \{\omega \in \Omega: X(\omega) = a\} \quad (\forall a \in \mathbb{R})$$

$$[X \leq a] \text{ denota o evento } \{\omega \in \Omega: X(\omega) \leq a\} \quad (\forall a \in \mathbb{R})$$

e com essa notação usamos, por exemplo

1.  $\mathbb{P}([X = 3])$ , ou  $\mathbb{P}(X = 3)$ , com o significado de  $\mathbb{P}(\{\omega \in \Omega: X(\omega) = 3\})$
2.  $[X \leq 3] \cap [X \geq 3] = \{\omega \in \Omega: X(\omega) \leq 3\} \cap \{\omega \in \Omega: X(\omega) \geq 3\} = [X = 3];$
3.  $[X < 3] \cap [X > 3] = \emptyset$

e valem todas as definições e operações com evento, por exemplo, se  $X, Y: \Omega \rightarrow \mathbb{R}$  são duas v.a.

1. podemos dizer  $[X < a]$  e  $[Y > b]$  são independentes;
2. usando a Lei de Probabilidade Total podemos escrever

$$\mathbb{P}(X = 3) = \mathbb{P}(X = 3 \mid Y < 1)\mathbb{P}(Y < 1) + \mathbb{P}(X = 3 \mid Y \geq 1)\mathbb{P}(Y \geq 1).$$

**Exemplo 36** Se  $X$  é o número de caras em 3 lançamentos de uma moeda então

$$[X = 2] = \{(Ca, Ca, Co), (Ca, Co, Ca), (Co, Ca, Ca)\}$$

$$[X \geq 2] = \{(Ca, Ca, Co), (Ca, Co, Ca), (Co, Ca, Ca), (Ca, Ca, Ca)\}$$

e  $\mathbb{P}([X = 2]) = 3/8$  e  $\mathbb{P}([X \geq 2]) = 1/2$ .

O evento complementar a  $[X \geq 2]$  é o evento

$$\overline{[X \geq 2]} = [X < 2] = \{(Ca, Co, Co), (Co, Ca, Co), (Co, Co, Ca), (Co, Co, Co)\}$$

que ocorre com probabilidade  $\mathbb{P}(\overline{[X \geq 2]}) = 1 - \mathbb{P}(X \geq 2) = \frac{1}{2}$ .

Se  $A$  é o evento “o primeiro lançamento foi Ca” então

$$\begin{aligned} \mathbb{P}(X = 2 \mid A) &= \frac{\mathbb{P}([X = 2] \cap A)}{\mathbb{P}(A)} \\ &= \frac{\mathbb{P}(\{(Ca, Ca, Co), (Ca, Co, Ca)\})}{\mathbb{P}(\{(Ca, Ca, Ca), (Ca, Ca, Co), (Ca, Co, Ca), (Ca, Co, Co)\})} \\ &= \frac{1}{2}. \end{aligned}$$

$$\begin{aligned} \mathbb{P}(X = 2 \mid X \geq 2) &= \frac{\mathbb{P}([X = 2] \cap [X \geq 2])}{\mathbb{P}(X \geq 2)} \\ &= \frac{\mathbb{P}(\{(Ca, Ca, Co), (Ca, Co, Ca), (Co, Ca, Ca)\})}{\mathbb{P}(\{(Ca, Ca, Ca), (Ca, Ca, Co), (Ca, Co, Ca), (Co, Ca, Ca)\})} \\ &= \frac{3}{4}. \end{aligned}$$

**Observação 2** A rigor,  $X$  é uma variável aleatória só se  $[X \leq a] \subset \Omega$  é um evento mensurável para todo  $a \in \mathbb{R}$  mas, como antes, não nos preocuparemos com essa questão pois para nós as funções encontradas na prática satisfazem esse requisito.

## Função de distribuição de uma v.a.

A [função de distribuição \(acumulada/cumulativa\)](#) da v.a.  $X$  é a função  $F_X: \mathbb{R} \rightarrow \mathbb{R}$  dada por

$$F_X(x) = \mathbb{P}(X \leq x)$$

Observemos que  $F(x) = F_X(x)$  está definida para todo real  $x$  e valem:

**(fda1)**  $0 \leq F(x) \leq 1$ ;  $\lim_{x \rightarrow +\infty} F(x) = 1$ ;  $\lim_{x \rightarrow -\infty} F(x) = 0$ ;

**(fda2)**  $F$  é não decrescente:  $x < y \implies F(x) \leq F(y)$ ;

**(fda3)**  $F$  é contínua à direita:  $\lim_{x \rightarrow a^+} F(x) = F(a)$ , para todo  $a \in \mathbb{R}$ .

A prova do item 2 é fácil, se  $x < y$  então  $[X \leq x] \subset [X \leq y]$ , portanto,  $\mathbb{P}(X \leq x) \leq \mathbb{P}(X \leq y)$  por **(P2)**. As outras propriedades precisam do exercício 25, página 17. As 3 propriedades acima *caracterizam* uma função de distribuição (i.e., qualquer  $F$  que satisfaz essas condições é a f.d.a. de alguma v.a.). Além delas, pode-se provar que

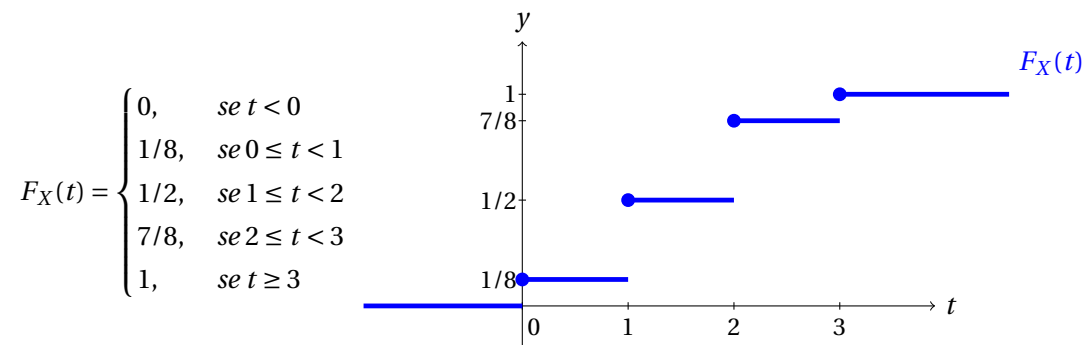


- $1 - F(a) = \mathbb{P}(X > a)$ , para todo  $a \in \mathbb{R}$ ;
- para quaisquer  $a < b$  temos  $F(b) - F(a) = \mathbb{P}(a < X \leq b)$ ;
- $\mathbb{P}(X = x) = F(x) - F(x-)$ , em que  $F(x-) \stackrel{\text{def}}{=} \lim_{y \rightarrow x^-} F(y)$ .

**Exercício 38** Prove que  $1 - F(a) = \mathbb{P}(X > a)$ , para todo  $a \in \mathbb{R}$ .

**Exercício 39** Prove que  $F(b) - F(a) = \mathbb{P}(a < X \leq b)$  para quaisquer  $a < b$ .

**Exemplo 37** No caso dos 3 lançamentos de uma moeda



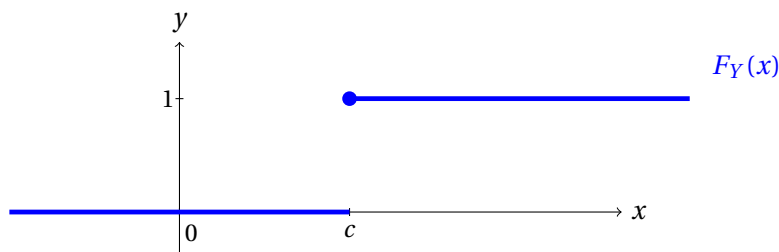
Usando as propriedades de uma função de distribuição temos, por exemplo

- $\mathbb{P}(X > 2) = 1 - F(2) = 1 - 7/8 = 3/8$ ;
- $\mathbb{P}(X > 3) = 1 - F(3) = 0$ ;
- $\mathbb{P}(0,5 < X \leq 2,5) = F(2,5) - F(0,5) = 7/8 - 1/8 = 3/4$ ;
- $\mathbb{P}(X = 1) = F(1) - F(1-) = 1/2 - 1/8 = 3/8$ ;
- $\mathbb{P}(X = 1.8) = F(2) - F(2-) = 7/8 - 1/2 = 3/8$ ;
- $\mathbb{P}(X = -1) = F(-1) - F(-1-) = 0 - 0 = 0$ ;
- $\mathbb{P}(X = 7) = F(7) - F(7-) = 1 - 1 = 0$ .

As descontinuidades das funções de distribuição são do tipo salto. Se  $F$  é descontínua então o salto em  $a$  é de  $\mathbb{P}(X = a) = F(a) - F(a-)$ . No exemplo acima, o salto em  $x = 1.8$  é  $7/8 - 1/2 = 3/8 = \mathbb{P}(X = 1.8)$ . Notemos que a soma dos saltos de tamanho  $\geq 1/n$  não deve ser maior que 1, portanto, há no máximo  $n$  desses saltos; desse fato podemos concluir que há um número enumerável de pontos de descontinuidade.

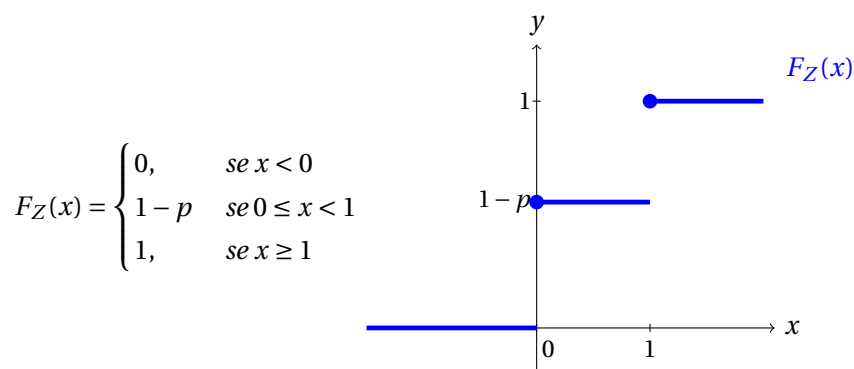
**Exemplo 38** Um caso bem simples de v.a. é  $Y(\omega) = c$  para algum  $c \in \mathbb{R}$  e todo  $\omega \in \Omega$ .

$$F_Y(x) = \begin{cases} 0, & \text{se } x < c \\ 1, & \text{se } x \geq c \end{cases}$$



**Exemplo 39** Consideremos uma moeda com probabilidade  $p$  de sair Ca. Seja  $Z: \Omega \rightarrow \mathbb{R}$  dada por

$$Z(\text{Ca}) = 1, Z(\text{Co}) = 0$$



Estudaremos dois tipos de variáveis aleatórias:

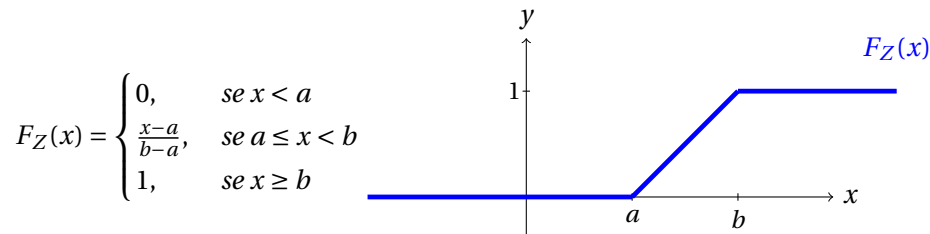
- variável aleatória discreta Uma v.a.  $X$  é *discreta* se assume valores em um subconjunto enumerável de  $\mathbb{R}$ ; a função de massa de probabilidade para  $X$  é uma função  $f: \mathbb{R} \rightarrow [0, 1]$  que atribui a cada valor da v.a. sua probabilidade,  $f(a) = \mathbb{P}(X = a)$ ; se  $X$  assume valores reais  $x_1, x_2, x_3, \dots, x_n, \dots$  então  $f(x_i) = \mathbb{P}(X = x_i)$  e  $f(x) = 0$  se  $x \neq x_i$ , para todo  $i$ .
- variável aleatória contínua Uma v.a.  $X$  é *contínua* se

$$F_X(a) = \int_{-\infty}^a f(u) du \quad (\forall a \in \mathbb{R})$$

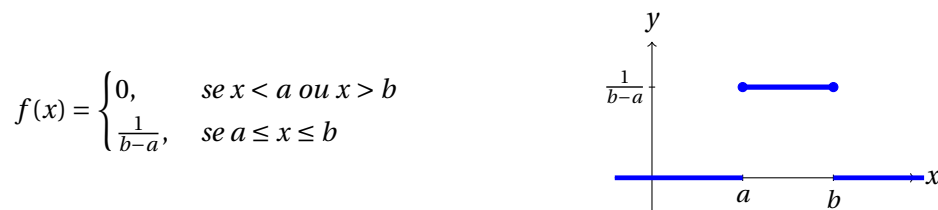
para alguma função integrável  $f: \mathbb{R} \rightarrow [0, +\infty)$  chamada função de densidade de probabilidade de  $X$ .

Para nós, sempre valerá que  $X$  tem uma função de densidade se a distribuição  $F_X$  é contínua e é derivável em todo ponto da reta, exceto por um número finito deles ou exceto pelos inteiros. Uma função integrável  $f: \mathbb{R} \rightarrow [0, +\infty)$  é densidade de alguma v.a. se, e somente se,  $\int_{-\infty}^{+\infty} f(x) dx = 1$ , pois neste caso  $F(x) = \int_{-\infty}^x f(u) du$  satisfaz as três propriedades citadas acima que caracterizam uma função de distribuição.

**Exemplo 40** Se  $Z: \Omega \rightarrow \mathbb{R}$  tem função de distribuição dada por



então  $F_Z$  é contínua e tem derivada em todo ponto, exceto  $a$  e  $b$ , a função de densidade é a derivada, i.e.

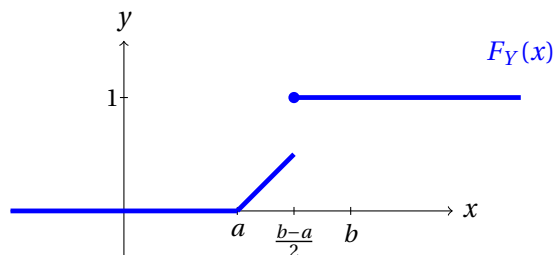


Notemos que  $f(a)$  e  $f(b)$  é arbitrário, pois quaisquer que sejam esses valores, a integral  $\int_{-\infty}^x f(u) du$  ainda vale  $F_Z(x)$  (veja o exemplo 57 nessa página).

**Exemplo 41** Defina a v.a.  $Y$  por

$$Y(\omega) = \min\{Z(\omega), (b-a)/2\}$$

em que  $Z$  é a v.a. do exemplo anterior. Pela definição de  $Y$ ,  $\mathbb{P}(Y \leq (b-a)/2) = 1$ ; para  $x < (b-a)/2$ ,  $Y(\omega) \leq x$  se e só se  $Z(\omega) \leq x$ , logo  $F_Y(x) = F_Z(x)$ .



Nesse caso

$$\mathbb{P}\left(Y = \frac{b-a}{2}\right) = F_Y\left(\frac{b-a}{2}\right) - F_Y\left(\frac{b-a}{2} -\right) = \frac{b-a}{2} - \frac{b-3a}{2(b-a)}.$$

Se  $Y$  fosse uma v.a. contínua esse valor deveria ser 0 (por quê?). Se  $Y$  fosse discreta sua imagem deveria ser enumerável. Essa v.a. não é discreta nem contínua.

Notemos que se  $X$  é v.a. contínua com distribuição  $F$  e densidade  $f$  então

$$P(a \leq X \leq b) = F(b) - F(a) = \int_{-\infty}^b f(u) du - \int_{-\infty}^a f(u) du = \int_a^b f(u) du$$

e  $\mathbb{P}(X = a) = 0$  para qualquer  $a$ .

**Exercício 40** Seja  $X$  uma v.a. e  $a, b \in \mathbb{R}$ . Defina a v.a.  $Y$  por  $Y(\omega) = aX(\omega) + b$ , para todo  $\omega \in \Omega$ . Se  $F$  é função de distribuição de  $X$ , determine a função de distribuição de  $Y$ .

**Exercício 41** Seja  $g: \mathbb{R} \rightarrow \mathbb{R}$  uma função contínua e crescente (estritamente). A função  $Y = g(X)$  é uma variável aleatória?

**Exercício 42** Prove que se  $X$  é v.a. discreta com função de massa de probabilidade  $f$  então  $\mathbb{P}(a \leq X \leq b) = \sum_x f(x)$  onde a soma é sobre todo  $a \leq x \leq b$  tal que  $f(x) > 0$ .

**Exercício 43** Prove que se  $a < b$  e  $Y$  é v.a. contínua com função de densidade de probabilidade  $f$  então  $\mathbb{P}(a < X < b) = \mathbb{P}(a \leq X \leq b) = \int_a^b f(x) dx$ .

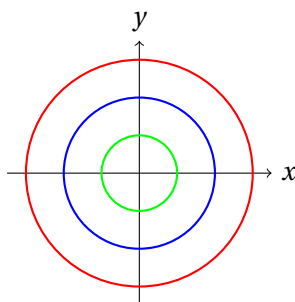
Notemos que do exercício acima deduzimos que para  $\varepsilon > 0$  pequeno temos

$$\mathbb{P}\left(a - \frac{\varepsilon}{2} < X < a + \frac{\varepsilon}{2}\right) = \int_{a-\frac{\varepsilon}{2}}^{a+\frac{\varepsilon}{2}} f(x) dx \approx \varepsilon f(a)$$

ou seja,  $X$  assume valor na vizinhança de  $a$  com probabilidade próxima a  $\varepsilon f(a)$ .

### Exemplo 42

Num jogo de dardos o alvo é composto de 3 círculos concêntricos de raios 1, 2 e 3. Consideremos o centro desses círculos no ponto  $(0, 0)$  do plano cartesiano



Assuma que o dardo acerte qualquer ponto dentro do círculo vermelho, i.e.,

$$\Omega = \left\{ (x, y) : \sqrt{x^2 + y^2} < 3 \right\}$$

e que a probabilidade de atingir qualquer ponto de uma região  $A$  é proporcional a área de  $A$ ,

$$\mathbb{P}(A) = \frac{\text{Área}(A)}{9\pi^2}. \quad (13)$$

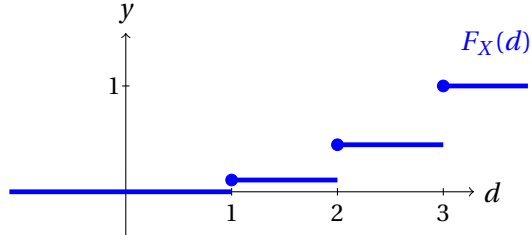
Para cada  $k = 1, 2, 3$ , considere as regiões

$$A_k = \left\{ (x, y) : k-1 \leq \sqrt{x^2 + y^2} < k \right\}.$$

Se num lançamento o jogador acerta a região  $A_k$  então ele marca  $k$  pontos. Se  $X(\omega) = k$  quando  $\omega \in A_k$  então  $\mathbb{P}(X = k) = \mathbb{P}(A_k) = (2k - 1)/9$  e  $X$  tem distribuição acumulada

$$F_X(d) = \begin{cases} 0, & \text{se } d < 1 \\ \frac{\lfloor d \rfloor^2}{9}, & \text{se } 1 \leq d < 3 \\ 1, & \text{se } d \geq 3 \end{cases}$$

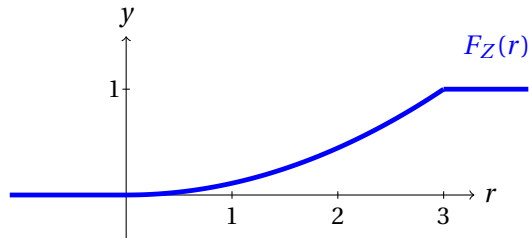
onde  $\lfloor d \rfloor$  é o maior inteiro menor ou igual a  $d$ .



Se  $\omega = (x, y) \in \Omega$  e  $Y(\omega) = \sqrt{x^2 + y^2}$  é a distância do ponto atingido ao centro. Denotemos por  $C_r$  o disco

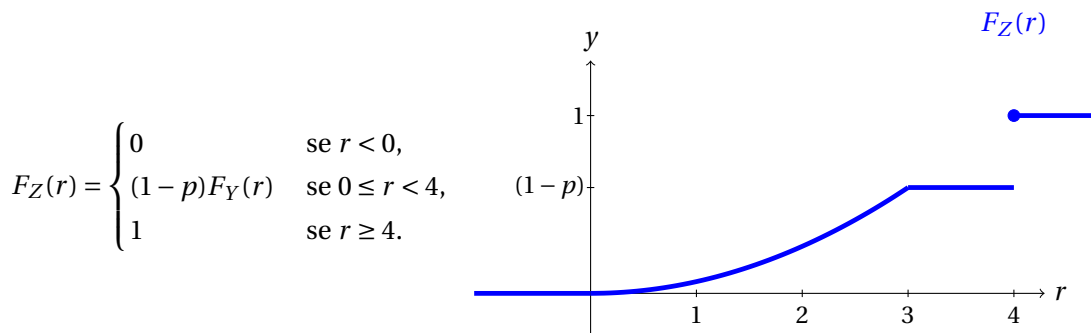
$$C_r = \{(x, y) : x^2 + y^2 \leq r^2\}.$$

A função de distribuição acumulada de  $Y$  é  $F_Y(r) = \mathbb{P}(C_r) = r^2/9$  se  $0 \leq r \leq 3$



Agora, suponha que o jogador erre o alvo com probabilidade  $p$ , para algum  $p > 0$  fixo, caso acerte então vale a equação (13). A pontuação, caso acerte, é a distância ao centro e, caso erre o alvo, é 4. Seja  $Z$  a v.a. da pontuação de um lance. Então

$$F_Z(r) = \mathbb{P}(Z \leq r \mid \text{acertou})\mathbb{P}(\text{acertou}) + \mathbb{P}(Z \leq r \mid \text{errou})\mathbb{P}(\text{errou})$$



**Problema** Num jogo de apostas, se o ganho é  $x$  e a perda é  $y$  em cada rodada, então o ganho médio é  $x \cdot \mathbb{P}(\text{ocorrendas favoráveis}) + y \cdot \mathbb{P}(\text{ocorrendas desfavoráveis})$ . Uma v.a.  $U$  não-negativa em função de distribuição acumulada  $F$  e densidade  $f = F'$ . Um jogo lhe é oferecido da seguinte forma: você pode escolher um número não negativo  $c$ , se  $U > c$  então você ganha a quantidade  $c$ , caso contrário, você não ganha nada. Como exemplo, suponha que  $U$  é a altura (medida em cm) da próxima pessoa entrando em uma estação ferroviária pública específica. Se você escolher  $c = 100$ , então você quase certamente ganha essa quantia. Um valor de  $c = 200$  dobraria a quantia se você ganhar, mas reduz drasticamente a sua probabilidade de ganhar.

Encontrar uma equação para caracterizar o valor de  $c$  que maximiza o ganho médio.

## §5 Esperança matemática

Se duas moedas são lançadas 16 vezes e o número de caras por lançamento é 0 em 4 deles, é 1 em 7 deles e é dois em 5 deles, qual é o número médio de caras por lançamento?

$$0 \frac{4}{16} + 1 \frac{7}{16} + 2 \frac{5}{16}$$

Se  $X$  é uma v.a. então **valor médio (ou valor esperado, ou esperança)** da v.a.  $X$  é dado por

- se  $X$  é discreta e assume valores  $x_1, x_2, \dots$  e tem função de massa de probabilidade  $f$  então

$$\mathbb{E}(X) \stackrel{\text{def}}{=} \sum_{i \geq 1} x_i f(x_i) \quad (14)$$

no caso infinito exigimos **convergência absoluta**, na prática significa que se alterarmos o ordem dos fatores o limite não muda, o que nos permite escrever (verifique a igualdade)

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\omega); \quad (15)$$

- se  $X$  é contínua e tem função de densidade de probabilidade  $f$  então

$$\mathbb{E}(X) \stackrel{\text{def}}{=} \int_{-\infty}^{+\infty} x f(x) dx. \quad (16)$$

**Exemplo 43 (variável aleatória indicadora)** *Seja  $I_A$  variável aleatória indicadora da ocorrência do evento  $A$ , i.e.,  $I_A = 1$  se  $A$  ocorre e  $I_A = 0$  se  $\bar{A}$  ocorre, ou ainda, para todo  $\omega \in \Omega$*

$$I_A(\omega) = \begin{cases} 1 & \text{se } \omega \in A; \\ 0 & \text{se } \omega \notin A. \end{cases} \quad (17)$$

Então  $\mathbb{E}(I_A) = \mathbb{P}(A)$ .

**Observação 3 (uma justificativa informal para (16))** *A definição de valor médio no caso discreto é intuitiva. No caso contínuo podemos justificar, ingenuamente, da seguinte maneira: Sejam  $I_n = (y_n, y_{n+1}]$ , para todo  $n \in \mathbb{Z}$ , uma coleção de intervalos centrados em  $x_i$  que particiona a reta e que, por simplicidade, supomos de mesmo comprimento  $\varepsilon$ . Definimos a v.a. discreta  $Y$  sobre o mesmo espaço amostral dada por*

$$Y = \sum_n x_n I_{[X \in I_n]}$$

*que assume os valores  $x_n$  ( $n \in \mathbb{Z}$ ). Assim  $[Y = x_n] = [X \in I_n]$  e a esperança de  $Y$  é*

$$\mathbb{E}(Y) = \sum_n x_n \mathbb{P}(Y = x_n) = \sum_n x_n \mathbb{P}(X \in I_n).$$

Notemos que se  $\omega \in [X \in I_n]$ , então  $X(\omega) \in I_n$  e  $Y(\omega) = x_n$ , logo

$$|Y(\omega) - X(\omega)| \leq \frac{|y_{n+1} - y_n|}{2} = \frac{\varepsilon}{2} \quad (\forall \omega \in \Omega).$$

Portanto, a definição de esperança para a variável  $X$  deve satisfazer  $|\mathbb{E}(X) - \mathbb{E}(Y)| \leq \frac{\varepsilon}{2}$ , logo  $\mathbb{E}(X) = \lim \mathbb{E}(Y)$  quando  $\varepsilon \rightarrow 0$ , i.e.

$$\begin{aligned} \mathbb{E}(X) &= \lim_{\varepsilon \rightarrow 0} \sum_n x_n \mathbb{P}(y_n < X \leq y_{n+1}) \\ &= \lim_{\varepsilon \rightarrow 0} \sum_n x_n (F_X(y_{n+1}) - F_X(y_n)) \\ &= \int_{-\infty}^{+\infty} x \, dF_X \\ &= \int_{-\infty}^{+\infty} x f(x) \, dx \end{aligned}$$

pois  $dF_X = f(x) \, dx$ ; (lembramos que  $F'_X = f$ ).

**Exemplo 44** Seja  $X$  o resultado de um lançamento de um dado,

$$\mathbb{E}X = 1 \frac{1}{6} + 2 \frac{1}{6} + 3 \frac{1}{6} + 4 \frac{1}{6} + 5 \frac{1}{6} + 6 \frac{1}{6} = \frac{7}{2}.$$

Qual a probabilidade  $\mathbb{P}(X = 7/2)$ ?

**Exemplo 45** Num jogo de azar você ganha \$1.000.000 com probabilidade  $p$  e \$10 com probabilidade  $1 - p$ . O ganho médio de uma aposta é

$$\mathbb{E}(Y) = 10^6 p + 10(1 - p).$$

No caso de  $p = 1/2$ , temos  $\mathbb{E}(Y) = 500.005$ , qual é a probabilidade de você ganhar \$500.000 numa aposta?

**Exemplo 46** Num jogo com 3 moedas, você ganha \$5 se ocorrerem três caras ou 3 coroas, você perde \$3 se ocorrer uma ou duas caras, se  $Y$  é o ganho então o ganho médio é

$$\mathbb{E}(Y) = 5 \frac{1}{4} - 3 \frac{3}{4} = -1.$$

**Exemplo 47** Seja  $T$  o tempo de vida útil de um equipamento eletrônico.  $T$  tem f.d.p.

$$f(t) = \begin{cases} \frac{20.000}{t^3} & \text{se } t > 100 \\ 0 & \text{caso contrário.} \end{cases}$$

O tempo médio de vida é

$$\mathbb{E}(T) = \int_{-\infty}^{+\infty} x \frac{20.000}{x^3} \, dx = \int_{100}^{+\infty} \frac{20.000}{x^2} \, dx = 200 \text{ horas.}$$

O seguinte resultado é muito útil. A prova do caso discreto é bem fácil e fica como exercício, a prova do caso contínuo é difícil e pulamos.

**Teorema 9** *Seja  $X$  uma variável aleatória com função de probabilidade  $f$ . Seja  $g$  uma função real. Então  $g(X)$  é uma variável aleatória cuja média é*

$$\mathbb{E}(g(X)) = \sum_x g(x)f(x)$$

*onde a soma é sobre todo real  $x$  tal que  $f(x) > 0$ ; caso  $X$  seja uma v.a. discreta, e cuja média é*

$$\mathbb{E}(g(X)) = \int_{-\infty}^{+\infty} g(x)f(x) dx$$

*caso  $X$  seja uma v.a. contínua.*  $\square$

**Exemplo 48** *Seja  $X$  o resultado do lançamento de um dado, como no exemplo 44.*

$$\mathbb{E}(X^2) = \frac{1 + 4 + 9 + 16 + 25 + 36}{6} = \frac{91}{6}.$$

*Notemos que, com o resultado do exemplo 44 podemos concluir que  $\mathbb{E}(X \cdot X) \neq \mathbb{E}(X)\mathbb{E}(X)$ ; em geral não vale  $\mathbb{E}(X \cdot Y) = \mathbb{E}(X)\mathbb{E}(Y)$  para  $X, Y$  v.a.'s quaisquer.*

**Exemplo 49** *Seja  $X$  o número de carros lavados num lava-rápido em 1 hora*

$x$	4	5	6	7	8	9
$f(x)$	1/12	1/12	1/4	1/4	1/6	1/6

*Para  $x$  carros lavados o atendente recebe  $2x - 1$  reais do gerente. O ganho médio é*

$$\mathbb{E}(2X - 1) = \sum_{x=4}^9 (2x - 1)f(x) = 12,67.$$

**Exemplo 50** *Se  $Z$  é uma v.a. com f.d.p*

$$f(x) = \begin{cases} \frac{x^2}{3}, & \text{se } -1 < x < 2 \\ 0 & \text{caso contrário.} \end{cases}$$

*e  $g: \mathbb{R} \rightarrow \mathbb{R}$  é dada por  $g(x) = 4x + 1$  então*

$$\mathbb{E}(g(Z)) = \mathbb{E}(4Z + 1) = \int_{-1}^2 (4x + 1) \frac{x^2}{3} dx = 8$$

**Corolário 4** *Para quaisquer  $a, b \in \mathbb{R}$  e v.a.  $X$*

$$\mathbb{E}(aX + b) = a\mathbb{E}(X) + b.$$



DEMONSTRAÇÃO: Usando o teorema anterior com  $g(x) = ax + b$  temos, no caso discreto, somando-se sobre todo  $x \in \mathbb{R}$  tal que  $f(x) > 0$

$$\begin{aligned}\mathbb{E}(aX + b) &= \sum_x (ax + b)f(x) = \sum_x axf(x) + \sum_x bf(x) \\ &= a \sum_x xf(x) + b \sum_x f(x) = a\mathbb{E}(X) + b\end{aligned}$$

e no caso contínuo

$$\begin{aligned}\mathbb{E}(aX + b) &= \int_{-\infty}^{+\infty} (ax + b)f(x) dx = \int_{-\infty}^{+\infty} axf(x) dx + \int_{-\infty}^{+\infty} bf(x) dx \\ &= a \int_{-\infty}^{+\infty} xf(x) dx + b \int_{-\infty}^{+\infty} f(x) dx = a\mathbb{E}(X) + b.\end{aligned}$$

□

**Exemplo 51** Reconsiderando o exemplo 49, podemos usar o corolário e calcular

$$\mathbb{E}(2X - 1) = 2\mathbb{E}(X) - 1 = 2 \sum_x xf(x) - 1 = 2 \frac{41}{6} - 1 = 12,67.$$

No caso do exemplo 50

$$\mathbb{E}(4Z + 3) = 4\mathbb{E}(Z) + 3 = 4 \int_{-1}^2 \frac{x^2}{3} dx + 3 = 4x^3 \Big|_{-1}^2 + 3 = 8.$$

**Exemplo 52** Numa urna estão 1 bola branca e 1 bola preta; uma bola é escolhida ao acaso, se for preta ela é devolvida e mais uma bola preta é colocada na urna e o sorteio é repetido, se sair bola branca o experimento termina. Se  $X$  é o número de rodadas até terminar então  $\mathbb{P}(X = k) = \frac{1}{k(k+1)}$  e média é  $\sum_{k \geq 1} k \frac{1}{k(k+1)} = \sum_{k > 1} \frac{1}{k} = \infty$  (essa é a série harmônica).

**Exemplo 53** Seja  $X$  uma variável com f.d.p.

$$f(x) = \begin{cases} \frac{10}{x^2} & \text{se } x > 10 \\ 0 & \text{caso contrário.} \end{cases}$$

$X$  tem esperança

$$\mathbb{E}(X) = \int_{10}^{\infty} x \frac{10}{x^2} dx = \int_{10}^{\infty} \frac{10}{x} dx = 10 \ln(x) \Big|_{10}^{\infty} = \infty$$

**Observação 5** No caso dos dois últimos exemplos dizemos que a variável aleatória não tem esperança finita.

**Exercício 44** Prove o caso discreto do teorema acima, i.e.,  $\mathbb{E}(g(X)) = \sum_x g(x)f(x)$ .

## Variância

A **variância** da v.a.  $X$  é uma medida de quão dispersos estão os valores que a variável assume com relação ao valor médio, é dada pelo valor esperado da v.a.  $g(X) = (X - \mathbb{E}(X))^2$

$$\text{Var}(X) \stackrel{\text{def}}{=} \mathbb{E}(X - \mathbb{E}(X))^2 = \begin{cases} \sum_x (x - \mathbb{E}(X))^2 f(x) & \text{no caso discreto,} \\ \int_{-\infty}^{+\infty} (x - \mathbb{E}(X))^2 f(x) dx & \text{no caso contínuo.} \end{cases}$$

O seguinte exercício fornece um modo, em geral, mais fácil para computar a variância.

**Exercício 45** *Prove que*

$$\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2.$$

O **desvio padrão** é definido como a raiz quadrada positiva da variância

$$\sigma_X = \sqrt{\text{Var}(X)}.$$

Suponhamos que  $X$  é, por exemplo, a quantidade de refrigerante engarrafada por uma máquina de uma fábrica em  $ml$  (mililitros). Então  $\text{Var}(X)$  é a dispersão dos valores de  $X$  com respeito a média em  $ml^2$ , o desvio padrão é uma medida de dispersão em  $ml$ .

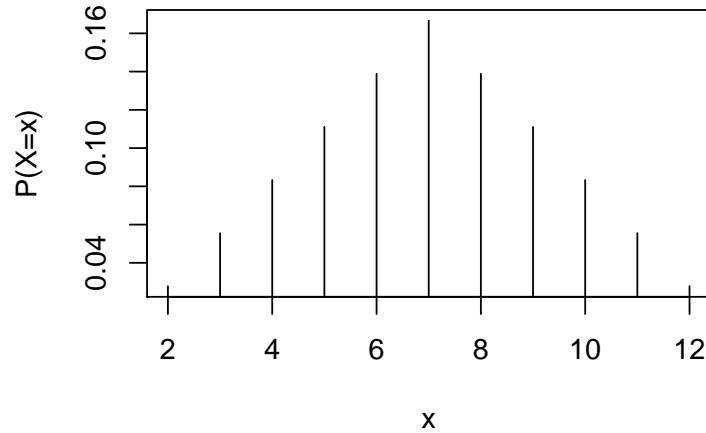
**Exemplo 54** *A variância no valor resultante de um lançamento de dado é (veja os exemplos 44 e 48)  $91/6 - 49/4 = 35/12$  ( $\approx 2,91$ ). O desvio padrão é  $\approx 1,7$ .*

**Exemplo 55** *Qual é o valor médio da soma dos pontos no lançamento de dois dados? O espaço amostral é composto por 36 eventos elementares igualmente prováveis*

$$\Omega = \left\{ (1,1), (1,2), (1,3), (1,4), (1,5), (1,6), \right. \\ (2,1), (2,2), (2,3), (2,4), (2,5), (2,6), \\ (3,1), (3,2), (3,3), (3,4), (3,5), (3,6), \\ (4,1), (4,2), (4,3), (4,4), (4,5), (4,6), \\ (5,1), (5,2), (5,3), (5,4), (5,5), (5,6), \\ \left. (6,1), (6,2), (6,3), (6,4), (6,5), (6,6) \right\}$$

*Se  $X$  é o resultado da soma dos lançamentos, então sua função de massa de probabilidade é*

$x$	$f(x)$
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36



O valor esperado da soma é

$$\mathbb{E}(X) = 2(1/36) + 3(2/36) + \cdots + 11(2/36) + 12(1/36) = \frac{252}{36} = 7$$

e o da variância é

$$\begin{aligned} \text{Var}(X) &= (2-7)^2(1/36) + (3-7)^2(2/36) + \cdots \\ &\quad + (11-7)^2(2/36) + (12-7)^2(1/36) = \frac{210}{36} = 5,83. \end{aligned}$$

o desvio padrão vale 2,41, aproximadamente. Notemos que no intervalo  $(\mathbb{E}(X) - \sigma_X, \mathbb{E}(X) + \sigma) = \{5, 6, 7, 8, 9\}$  está concentrado 2/3 da massa de probabilidade.

**Exemplo 56** Um canal digital transmite informação em pacotes de 4 bits. Os bit podem ser recebidos com erro e  $X$  denota o número de bits errados num pacote, com função de distribuição acumulada

$$F(x) = \begin{cases} 0 & x < 0 \\ 0,6561 & 0 \leq x < 1 \\ 0,9477 & 1 \leq x < 2 \\ 0,9963 & 2 \leq x < 3 \\ 0,9999 & 3 \leq x < 4 \\ 1 & 4 \leq x \end{cases}$$

assim,

- $\mathbb{P}(X = 0) = f(0) = 0,6561 - 0 = 0,6561$
- $\mathbb{P}(X = 1) = f(1) = 0,9477 - 0,6561 = 0,2916$

- $\mathbb{P}(X = 2) = f(2) = 0,9963 - 0,9477 = 0,0486$
- $\mathbb{P}(X = 3) = f(3) = 0,9999 - 0,9963 = 0,0036$
- $\mathbb{P}(X = 4) = f(4) = 1 - 0,9999 = 0,0001$

e o valor médio do número de bits errados é

$$\begin{aligned}\mathbb{E}(X) &= 0 \cdot f(0) + 1 \cdot f(1) + 2 \cdot f(2) + 3 \cdot f(3) + 4 \cdot f(4) \\ &= 0 \cdot 0,6561 + 1 \cdot 0,2916 + 2 \cdot 0,0486 + 3 \cdot 0,0036 + 4 \cdot 0,0001 \\ &= 0,4\end{aligned}$$

e a variância

$$\begin{aligned}\text{Var}(X) &= \sum_{x=0}^4 (x - \mathbb{E}(X))^2 f(x) \\ &= 0,16 \cdot 0,6561 + 0,36 \cdot 0,2916 + 2,56 \cdot 0,0486 + 6,76 \cdot 0,0036 \\ &\quad + 12,96 \cdot 0,0001 \\ &= 0,36\end{aligned}$$

portanto o desvio padrão é 0,6.

## Propriedades da esperança

A esperança de uma variável aleatória satisfaz as seguintes propriedades enunciadas como exercícios. Convém ressaltar que o caso contínuo tem, em geral, maior dificuldade, mas que o caso discreto é simples na maioria dos casos.

**Exercício 46** Para  $c \in \mathbb{R}$  fixado e  $X(\omega) = c$  ( $\forall \omega \in \Omega$ ), vale  $\mathbb{E}(X) = c$  e  $\text{Var}(X) = 0$ .

**Exercício 47** Prove que se  $\mathbb{P}(a \leq Z \leq b) = 1$  então  $a \leq \mathbb{E}(Z) \leq b$ .

**Exercício 48** Prove que se  $X \leq Y$  (i.e.,  $X(\omega) \leq Y(\omega)$ ,  $\forall \omega \in \Omega$ ) então  $\mathbb{E}(X) \leq \mathbb{E}(Y)$ .

**Exercício 49** Se  $Y$  é uma v.a. contínua que assume valores não-negativos então  $\mathbb{E}(Y) = \int_0^\infty \mathbb{P}(Y > x) dx$ . Se além de não-negativa é inteira (portanto discreta), então  $\mathbb{E}(Y) = \sum_{i \geq 1} \mathbb{P}(Y \geq i)$ .

**Exercício 50 (soma de variáveis aleatórias)** Sejam  $X$  e  $Y$  v.a.'s sobre o mesmo espaço amostral e a soma delas é a função  $X + Y$  dada por  $X + Y(\omega) = X(\omega) + Y(\omega)$ . Prove que

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y).$$

Prove que se  $\mathbb{E}(X \cdot Y) = \mathbb{E}(X)\mathbb{E}(Y)$  então

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).$$

**Exercício 51** Para  $a, b \in \mathbb{R}$ , vale  $\text{Var}(aX + b) = a^2 \text{Var}(X)$ .

**Exercício 52** Se  $\lambda$  é real positivo e  $Z$  uma v.a. que assume valores não negativos, então

$$\mathbb{P}(Z \geq \lambda) \leq \frac{\mathbb{E}(Z)}{\lambda}.$$

(Dica: Defina  $Y$  por  $Y(w) = \lambda$  se  $Z(w) \geq \lambda$ , caso contrário  $Y(w) = 0$ . Determine  $\mathbb{E}(Y)$  e use o exercício 48)

**Exercício 53 (desigualdade de Chebyshev)** Se  $\lambda$  é real positivo e  $X$  uma v.a. de esperança finita então

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \lambda) \leq \frac{\text{Var}(X)}{\lambda^2}. \quad (18)$$

(Dica: use o exercício anterior com  $Z = X - \mathbb{E}(X)$ .)

Se fizermos  $\lambda = k\sigma_X$  em (18) obtemos a probabilidade de  $X$  desviar de  $\mathbb{E}(X)$  por pelo menos  $k$  desvios padrão

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq k\sigma_X) \leq \frac{1}{k^2}. \quad (19)$$

**Exercício 54 (desigualdade de Markov)** Se  $\lambda$  é real positivo e  $X$  uma v.a. então

$$\mathbb{P}(|X| \geq \lambda) \leq \frac{\mathbb{E}(|X|^t)}{\lambda^t} \quad (20)$$

para todo  $t > 0$ .

**Exercício 55** Se  $Z \geq 0$  e  $\mathbb{E}(Z) = 0$  então  $\mathbb{P}(Z = 0) = 1$ .

**Exemplo 57** Consideremos  $n$  lançamentos de uma moeda com os resultados independentes. Seja  $X_i$  a variável indicadora (exemplo 43) do evento “ocorre cara”. A v.a.

$$S_n = X_1 + \cdots + X_n$$

é a quantidade de ocorrência de cara e  $\frac{S_n}{n}$  é o número médio de ocorrência de cara.

A esperança de  $S_n$  é, usando o exercício 50

$$\mathbb{E}(S_n) = \mathbb{E}\left(\sum_{i=1}^n X_i\right) = \left(\sum_{i=1}^n \mathbb{E}(X_i)\right) = \frac{n}{2}.$$

e a variância é  $\text{Var}(S_n) = \mathbb{E}(S_n^2) - \mathbb{E}(S_n)^2$ , assim precisamos calcular  $\mathbb{E}(S_n^2)$ .

$$\mathbb{E}(S_n^2) = \mathbb{E}\left(\sum_{i=1}^n \sum_{j=1}^n X_i X_j\right) = \sum_{i=1}^n \sum_{j=1}^n \mathbb{E}(X_i X_j),$$

se  $i \neq j$  então

$$\mathbb{E}(X_i \cdot X_j) = \mathbb{P}(X_i \cdot X_j = 1) = \mathbb{P}([X_i = 1] \cap [X_j = 1]) = \frac{1}{2} \cdot \frac{1}{2}$$

e se  $i = j$  então

$$\mathbb{E}(X_i \cdot X_j) = \mathbb{E}(X_i^2) = P(X_i = 1) = \frac{1}{2}$$

portanto

$$\mathbb{E}(S_n^2) = \sum_{i=1}^n \sum_{j=1}^n \mathbb{E}(X_i X_j) = \sum_{i=1}^n \mathbb{E}(X_i^2) + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \mathbb{E}(X_i X_j) = \frac{n}{2} + \frac{n(n-1)}{4}.$$

De volta à variância de  $S_n$  temos

$$\text{Var}(S_n) = \mathbb{E}(S_n^2) - \mathbb{E}(S_n)^2 = \frac{n}{2} + \frac{n(n-1)}{4} - \left(\frac{n}{2}\right)^2 = \frac{n}{4}.$$

A esperança de  $S_n/n$  é, usando linearidade da esperança (corolário 4)

$$\mathbb{E}\left(\frac{S_n}{n}\right) = \frac{1}{n} \mathbb{E}(S_n) = \frac{1}{2}. \quad (21)$$

e a variância é, pelo exercício 51,

$$\text{Var}\left(\frac{S_n}{n}\right) = \frac{1}{n^2} \text{Var}(S_n) = \frac{1}{n^2} \frac{n}{4} = \frac{1}{4n}.$$

Usando a desigualdade de Chebyshev, eq. (18), para qualquer  $\lambda > 0$

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \frac{1}{2}\right| \geq \lambda\right) \leq \frac{1}{4n\lambda^2}$$

portanto

$$\mathbb{P}\left(\frac{1}{2} - \lambda \leq \frac{S_n}{n} \leq \frac{1}{2} + \lambda\right) > 1 - \frac{1}{4n\lambda^2}.$$

Por exemplo, fazendo  $\lambda = n^{-1/2}$  temos que o número médio de caras está no intervalo  $(\frac{1}{2} - \frac{1}{\sqrt{n}}, \frac{1}{2} + \frac{1}{\sqrt{n}})$  com probabilidade maior que  $3/4$ .

Mais que isso, a probabilidade tende a 1 quando  $n \rightarrow \infty$ . Esse resultado foi provado pela primeira vez em 1713, por Jacob Bernoulli, sem usar a desigualdade de Chebyshev, desconhecida na época.

**Teorema 10 (Lei Fraca dos Grandes Números)** Sejam  $X_1, X_2, \dots, X_n$  v.a.'s independentes e identicamente distribuídas, cada uma com média  $\mathbb{E}(X_i) = \mu$  finita e variâncias finita. Então para todo  $\lambda > 0$

$$\mathbb{P}\left(\left|\frac{X_1 + \dots + X_n}{n} - \mu\right| \geq \lambda\right) \rightarrow 0 \quad (22)$$

quando  $n \rightarrow \infty$ .

A Lei Forte dos Grandes Números estabelece que, sob as mesmas hipóteses da lei fraca,

$$\frac{X_1 + \dots + X_n}{n} \rightarrow \mu \text{ quando } n \rightarrow \infty. \quad (23)$$

A Lei dos Grandes Números mostra que o modelo probabilístico é consistente com a interpretação frequentista de probabilidade.

**Exercício 56** Mostre que se  $\varepsilon > 0$  e  $X$  assume os valores  $-\varepsilon$  e  $\varepsilon$  com probabilidade  $1/2$ , cada, então vale a igualdade na desigualdade de Chebyshev.

**Exercício 57** Uma moeda honesta é lançada 100 vezes. O número esperado de caras é 50, e o desvio padrão para o número de caras é 5. O que a desigualdade de Chebyshev nos diz sobre a probabilidade de que o número de caras que ocorrem desvia do número esperado por três ou mais desvios-padrão (ou seja, em pelo menos 15)?

**Exercício 58** Uma aposta de 1 real tem ganho esperado de  $-0,0141$ . O que a Lei dos Grandes Números nos diz sobre os seus ganhos se você fizer um grande número apostas de 1 real? Será que ela lhe assegura que suas perdas serão pequenas? Será que ela lhe assegura que, se  $n$  for muito grande você vai perder?

**Demonstração do teorema 9.** Vamos provar a parte referente a v.a. contínua. Primeiro, provaremos que se  $X$  assume valores não-negativos então

$$\mathbb{E}(X) = \int_0^{+\infty} \mathbb{P}(X > x) dx = \int_0^{+\infty} (1 - F_X(x)) dx.$$

De fato,

$$\int_0^{+\infty} \mathbb{P}(X > x) dx = \int_0^{+\infty} \int_x^{+\infty} f(u) du dx = \int_0^{+\infty} \int_0^u dx f(u) du = \int_0^{+\infty} u f(u) du$$

na segunda igualdade trocamos a ordem de integração.

Segundo, assumiremos que  $g(x) \geq 0$  para todo  $x$ . Então,

$$\mathbb{E}(g(X)) = \int_0^{+\infty} \mathbb{P}(g(X) > x) dx = \int_0^{+\infty} \int_B f(u) du dx$$

em que  $B = \{u \in \mathbb{R} : g(u) > x\}$

$$\mathbb{E}(g(X)) = \int_0^{+\infty} \int_B f(u) du dx = \int_0^{+\infty} \int_0^{g(u)} dx f(u) du = \int_0^{+\infty} g(u) f(u) du$$

que é o estabelecido no enunciado para  $g$  não negativa.

Pra finalizar, se  $g$  assume valores reais então definimos as v.a.'s não negativas

$$g^+(x) = \max\{g(x), 0\} \quad \text{e} \quad g^-(x) = \max\{-g(x), 0\}$$

e temos que  $g(x) = g^+(x) - g^-(x)$ , portanto,

$$\mathbb{E}(g(X)) = \mathbb{E}(g^+(X)) - \mathbb{E}(g^-(X)) = \int_0^{+\infty} g^+(u) f(u) du - \int_0^{+\infty} g^-(u) f(u) du = \int_0^{+\infty} g(u) f(u) du.$$

**Prova da desigualdade de Chebyshev** Para demonstrar a desigualdade de Chebyshev, primeiro mostramos a seguinte desigualdade de Markov: Se  $X$  é uma variável aleatória não-negativa e  $t > 0$  então

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}X}{t}. \quad (24)$$

DEMONSTRAÇÃO: Dados  $X$  e  $t$  como no enunciado defina  $Z(\omega)$  ( $\forall \omega \in \Omega$ ) por

$$Z(\omega) = \begin{cases} t, & \text{se } X(\omega) \geq t, \\ 0, & \text{se } X(\omega) < t. \end{cases}$$

Então  $X \geq Z$  e  $\mathbb{E}X \geq \mathbb{E}(Z) = 0\mathbb{P}(X < t) + t\mathbb{P}(X \geq t)$  portanto

$$\mathbb{E}(X) \geq t\mathbb{P}(X \geq t)$$

donde concluímos (24).  $\square$

Agora, se  $X$  é uma variável aleatória qualquer com  $\mathbb{E}(X), \text{Var}(X) < \infty$  então  $0 \leq (X - \mathbb{E}X)^2 < \infty$  logo para todo  $\lambda > 0$  temos por (24) com  $t = \lambda$

$$\mathbb{P}\left((X - \mathbb{E}X)^2 \geq \lambda^2\right) \leq \frac{\mathbb{E}(X - \mathbb{E}X)^2}{\lambda^2}$$

e de  $\{(X - \mathbb{E}X)^2 \geq \lambda^2\} = \{|X - \mathbb{E}X| \geq \lambda\}$  temos a seguinte desigualdade de Chebyshev

$$\mathbb{P}(|X - \mathbb{E}X| \geq \lambda) \leq \frac{\text{Var}(X)}{\lambda^2}$$

para todo  $\lambda > 0$ .  $\square$

**Demonstração da Lei Fraca dos Grande Números** Sejam  $X_1, X_2, \dots$  uma seqüência de variáveis aleatórias independentes, identicamente distribuídas e todas com valor esperado  $\mu$  e variância  $\sigma^2$  finitos.

$$M_n = \frac{X_1 + X_2 + \dots + X_n}{n}.$$

e temos

$$\mathbb{E}(M_n) = \mu \text{ e } \text{Var}(M_n) = \frac{\sigma^2}{n}.$$

Usando a desigualdade de Chebyshev (18), concluímos para todo  $\lambda > 0$

$$\lim_{n \rightarrow \infty} \mathbb{P}(|M_n - \mu| \geq \lambda) \leq \lim_{n \rightarrow \infty} \frac{\sigma^2}{n\lambda^2} = 0. \quad (25)$$

**Problema:** Pedro e Paula ambos querem cortar um pedaço de papel retangular. Como ambos são probabilistas eles determinam a forma exata do retângulo utilizando realizações de uma v.a. positiva, digamos  $U$ , como se segue. Pedro é preguiçoso e gera apenas uma única realização dessa v.a.; então ele corta um quadrado que tem comprimento e largura igual a esse valor. Paula gosta de diversidade e gera duas realizações independentes de  $U$ . Ela, então, corta um retângulo com largura igual a primeira realização e comprimento igual ao da segunda realização. (a) Serão as áreas cortadas por Pedro e Paula diferentes em média? (b) se forem, Pedro ou Paula deverá ter um retângulo com área maior?



## §6 Principais modelos discretos

Sejam  $X$  uma v.a. discreta e  $f: \mathbb{R} \rightarrow \mathbb{R}$  uma função de massa de probabilidade (f.m.p.) para  $X$ , então

1.  $f(x) > 0$  para um subconjunto enumerável de pontos de  $\mathbb{R}$ ;
2.  $\sum_x f(x) = 1$ , em que a soma é sobre todo  $x$  tal que  $f(x) > 0$ ;
3. a função de distribuição acumulada (f.d.a) é  $F(x) = \mathbb{P}(X \leq x) = \sum_{y \leq x} f(y)$ , em que a soma é sobre todo  $y$  tal que  $f(y) > 0$ ;
4.  $f(x) = P(X = x) = F(x) - F(x-)$ .

A f.m.p. caracteriza a distribuição de probabilidades da variável aleatória.

O valor esperado de  $X$  e a variância de  $X$  são dados, respectivamente, por

$$\mathbb{E}(X) = \sum_x x f(x) \quad \text{e} \quad \sum_x (x - \mathbb{E}(X))^2 f(x)$$

onde as somas são sobre todo  $x$  tal que  $f(x) > 0$ .

### Distribuição de Bernoulli

Na prática, ocorrem muitas situações com experimentos que admitem apenas dois resultados, por exemplo

1. uma peça é classificada como *boa* ou *defeituosa*;
2. o resultado de um exame médico é *positivo* ou *negativo*;
3. um paciente submetido a um tratamento é *curado* ou *não* da doença;
4. um entrevistado *concorda* ou *não concorda* com a afirmação feita;
5. no lançamento de um dado *ocorre* ou *não ocorre* a face “5”.

Nessas situações podemos representar, genericamente, os resultados do experimento com o espaço amostral

$$\Omega = \{\text{sucesso}, \text{fracasso}\}$$

e o modelo probabilístico fica determinado dado  $p = \mathbb{P}(\text{sucesso})$ . Esses experimentos recebem o nome de *Ensaio de Bernoulli* e a v.a. indicadora do evento “sucesso” é uma *variável aleatória de Bernoulli* com parâmetro  $p$ .

A notação  $X \sim \text{Bernoulli}(p)$  indica que  $X$  é uma v.a. de Bernoulli com parâmetro  $p$ ; ela assume dois valores

- 1 se ocorre sucesso,

- 0 se ocorre fracasso;

e sua f.m.p. é

$$\text{be}_p(x) = p^x(1-p)^{1-x} \quad \forall x \in \{0, 1\}$$

e vale 0 para outros valores de  $x$ .

A média e a variância de uma v.a.  $X \sim \text{Bernoulli}(p)$  são dadas por

$$\mathbb{E}(X) = p \quad \text{e} \quad \text{Var}(X) = p(1-p).$$

**Exercício 59** Verifique as fórmulas acima para média e variância de uma v.a. de Bernoulli.

## Distribuição Binomial

Consideremos  $n$  repetições independentes de um Ensaio de Bernoulli. Seja  $X$  o número de sucessos nas repetições.

**Exemplo 58** Um dado equilibrado é lançado 3 vezes. Qual é a probabilidade de se obter a face 5 duas vezes? Se  $S$  denota sucesso, i.e., “ocorre face 5” e  $F$  denota fracasso, “não ocorrer face 5” então podemos representar o espaço amostral por

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

e  $p = \mathbb{P}(\text{sucesso}) = 1/6$  e  $1-p = \mathbb{P}(\text{fracasso}) = 5/6$ . Como podemos determinar, por exemplo,  $\mathbb{P}(X=2)$ ? A função de massa de probabilidade é

$x$	$f(x)$
0	$(1-p)^3$
1	$3p(1-p)^2$
2	$3p^2(1-p)$
3	$p^3$

e podemos escrever essa função como  $f(x) = \binom{3}{x} p^x (1-p)^{3-x}$  para todo  $x \in \{0, 1, 2, 3\}$ . Assim,  $\mathbb{P}(X=2) = \binom{3}{2} p^2 (1-p)^{3-2} = 0,0694$ .

Uma *variável aleatória binomial* de parâmetros  $n \in \mathbb{N}$  e  $p \in (0, 1)$  é uma v.a. com f.m.p.

$$\text{Bi}_{n,p}(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad \forall x \in \{0, 1, 2, \dots, n\}$$

que pode ser vista como o número de sucessos em  $n$  ensaios independentes de Bernoulli e com mesma probabilidade  $p$  de sucesso.

De fato, se  $Y_1, Y_2, \dots, Y_n$  são as v.a. indicadoras de sucesso em cada um dos ensaios, então  $X = \sum_{i=1}^n Y_i$  é a quantidade de sucessos ocorridos.

A notação  $X \sim \text{Binomial}(n, p)$  indica que  $X$  é uma v.a. com distribuição binomial com parâmetros  $n$  e  $p$ . A média e a variância de  $X$  são dadas por

$$\mathbb{E}(X) = np \quad \text{e} \quad \text{Var}(X) = np(1 - p).$$

Essas duas igualdades podem ser deduzidas da identidade  $\mathbb{E}(X^k) = np\mathbb{E}((Y + 1)^{k-1})$  em que  $Y \sim \text{Binomial}(n - 1, p)$ . De fato, usando que  $i\binom{n}{i} = n\binom{n-1}{i-1}$  na segunda identidade abaixo

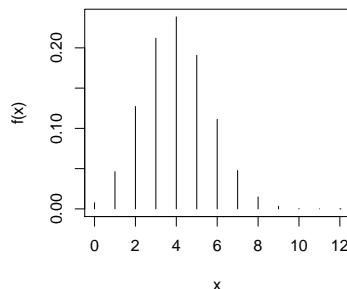
$$\begin{aligned} \mathbb{E}(X^k) &= \sum_{i=0}^n i^k \binom{n}{i} p^i (1-p)^{n-i} \\ &= \sum_{i=1}^n i^{k-1} i \binom{n}{i} p^i (1-p)^{n-i} \\ &= \sum_{i=1}^n i^{k-1} n \binom{n-1}{i-1} p^i (1-p)^{n-i} \\ &= np \sum_{i=1}^n i^{k-1} \binom{n-1}{i-1} p^{i-1} (1-p)^{n-i} \\ &= np \sum_{j=0}^{n-1} (j+1)^{k-1} \binom{n-1}{j} p^j (1-p)^{n-1-j} \\ &= np \mathbb{E}((Y + 1)^{k-1}). \end{aligned}$$

**Exercício 60** Verifique as fórmulas para média e variância de uma v.a. binomial (use a identidade dada acima).

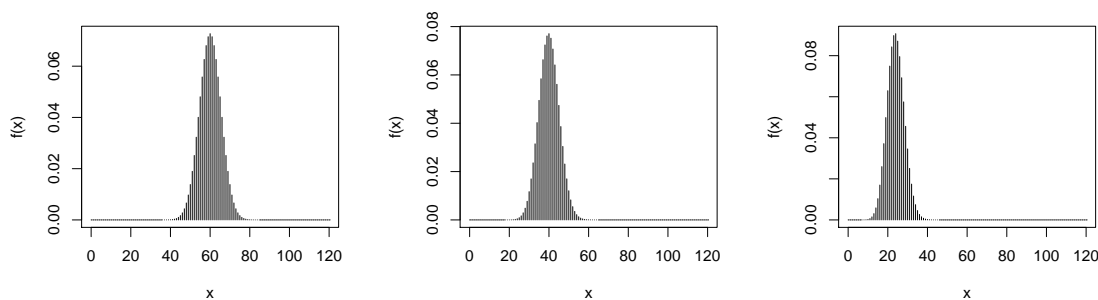
**Exemplo 59** Numa prova com 12 questões de múltipla escolha, com 3 alternativas, se todas as respostas forem chute, então a função de massa de probabilidade

$$f(x) = \text{Bi}_{12, \frac{1}{3}}(x) = \binom{12}{x} \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{12-x}$$

tem o seguinte gráfico de barras



A seguir, respectivamente, os gráficos mostram a distribuição nos casos de uma prova com 120 questões e 2 alternativas, uma prova com 120 questões e 3 alternativas e uma prova com 120 questões e 5 alternativas



Usando R:

o primeiro dos gráficos pode ser obtido com o R fazendo

```
x<-0:12
p=1/3
n=12
plot(x,dbinom(x,n,p),type="h")
```

A primeira linha diz os valores de  $x$  pros quais queremos  $Bi_{n,p}(x)$ . A segunda e terceira linhas atribuem os parâmetros da distribuição binomial, o parâmetro  $p$  é 0,5 e  $n$  é 12. Podemos usar o único comando

```
plot(0:12,dbinom(0:12,12,1/3),type="h")
```

`type="h"` é para o tipo de gráfico com barras. Os outros 3 gráficos acima são dados, respectivamente, por

```
plot(0:120,dbinom(0:120,120,1/2),type="h")
```

**Exemplo 60** Um equipamento resiste a um teste de choque com probabilidade  $3/4$ . Qual é probabilidade de que em 4 testes 2 equipamentos sobrevivam ao choque?

$$Bi_{4,3/4}(2) = \binom{4}{2} \left(\frac{3}{4}\right)^2 \left(\frac{1}{4}\right)^2 = \frac{27}{128}.$$

**Exemplo 61** Uma febre atinge 25% dos rebanhos. Se  $X$  animais não ficam doente então

$$\mathbb{P}(X = x) = Bi_{n,p}(x) = \binom{n}{x} (0,75)^x (0,25)^{n-x}$$

Os indivíduos de rebanhos de vários tamanhos recebem vacina.

população	proporção de sadios	análise
$n = x = 10$	$\mathbb{P}(X = x) = 0,056$	se nenhum animal ficar doente, pode ser que a vacina seja eficaz
$n = x = 12$	$\mathbb{P}(X = x) = 0,032$	se nenhum animal ficar doente, pode ser que a vacina seja eficaz
$n = 17, x = 16$	$\mathbb{P}(X \geq x) = 0,0501$	se $\leq 1$ ficar doente, evidência mais forte de eficácia que 0 doentes em 10
$n = 23, x = 21$	$\mathbb{P}(X \geq x) = 0,0492$	$\leq 2$ doentes em 23 é evidência mais forte que 1 em 17 e 0 em 10 de eficácia

Usando R:

Para  $X \sim \text{Binom}(n, p)$ , podemos calcular  $\text{Bi}_{n,p}(x) = \mathbb{P}(X \leq x)$  em R usando

`pbinom(x, n, p)`

no exemplo acima, para cada linha, respectivamente, os valores são obtidos com

```
pbinom(10, 10, .75) - pbinom(9, 10, .75)
pbinom(12, 12, .75) - pbinom(11, 12, .75)
1 - pbinom(15, 17, .75)
1 - pbinom(20, 23, .75)
```

A primeira linha da tabela também é obtida com `dbinom(10, 10, .75)` e a segunda com `dbinom(12, 12, .75)`.

**Exemplo 62** Um fabricante garante que seu produto tem uma taxa de itens defeituosos de 3%.

Numa seleção de 20 itens a serem inspecionados, qual é a probabilidade de ao menos um ser defeituoso?

Se  $X$  é a quantidade de itens defeituosos

$$\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X = 0) = 1 - \binom{20}{0} (0,03)^0 (0,97)^{20} = 0,4562.$$

Se 10 carregamentos por mês deixam a fábrica e de cada carregamento 20 itens são inspecionados, com que probabilidade 3 carregamentos tem pelo menos um item defeituoso?

$$\binom{10}{3} (0,4562)^3 (1 - 0,4562)^7 = 0,1602.$$

Seja  $X \sim \text{Binomial}(n, p)$ . Para  $n$  e  $p$  fixos, quando  $x$  varia de 0 a  $n$  o valor de  $\text{Bi}_{n,p}(x)$  cresce monotonicamente e depois decresce monotonicamente. De fato,

$$\frac{\text{Bi}_{n,p}(x)}{\text{Bi}_{n,p}(x-1)} = \frac{(n-x+1)p}{(1-p)x} \quad (26)$$

que  $\geq 1$  se e só se  $(n-x+1)p \geq (1-p)x$  ou, equivalentemente,  $x \leq (n+1)p$ .

**Proposição 6** Seja  $X \sim \text{Binomial}(n, p)$ . Para  $n$  e  $p$  fixos, quando  $x$  varia de 0 a  $n$  o valor de  $\text{Bi}_{n,p}(x)$  cresce monotonicamente e depois decresce monotonicamente atingindo o máximo quando  $x = (n+1)p$ .  $\square$

Notemos que o ponto de máximo é aproximadamente o valor médio, o que sugere que os maiores valores de  $\text{Bi}_{n,p}$  estão em torno da média (veja os gráficos do exemplo 59).

**Exemplo 63** Numa doença rara, um paciente se recupera com probabilidade 0,4. Em 15 pacientes doentes, seja  $X$  o número de sobreviventes.

1. pelo menos 10 sobrevivem.

$$\begin{aligned}\mathbb{P}(X \geq 10) &= 1 - \mathbb{P}(X < 10) \\ &= 1 - \sum_{x=0}^9 \binom{15}{x} (0,4)^x (0,6)^{15-x} = 0,0338.\end{aligned}$$

2. de 3 a 8 pacientes sobrevivam

$$\begin{aligned}\mathbb{P}(3 \leq X \leq 8) &= \sum_{x=3}^8 \binom{15}{x} (0,4)^x (0,6)^{15-x} \\ &= 0,8779\end{aligned}$$

3.  $X \in (\mu - 2\sigma, \mu + 2\sigma)$ . Usando a desigualdade de Chebyshev

$$\mathbb{P}(|X - \mu| \geq 2\sigma) \leq \frac{1}{4}$$

portanto  $X \in (\mu - 2\sigma, \mu + 2\sigma)$  com probabilidade  $3/4$ . A v.a.  $X$  tem média e variância dadas, respectivamente, por  $\mu = 6$  e  $\sigma^2 = 3,6$ , portanto com probabilidade  $3/4$   $X$  está no intervalo  $(2,20, 9,89)$ .

Usando R:

1-pbinom(9, 15, 0.4) para o item 1  
pbinom(8, 15, 0.4)-pbinom(2, 15, 0.4) para o item 2

Notemos que, no exemplo acima, não foi preciso usar os valores de  $\mu$  e  $\sigma$  para derivar o limitante  $3/4$  para a probabilidade. Isso se deve ao fato de Chebyshev valer pra qualquer distribuição e, por isso, a estimativa não é muito precisa.

**Exemplo 64** Numa eleição, seja  $p$  a fração (desconhecida) da população que vota no candidato  $D$ . Para simplificar, assumimos que um voto em  $D$  é ensaio de Bernoulli com parâmetro  $p$ .

Suponha que são realizadas  $n$  entrevistas:  $V_i \sim \text{Bernoulli}(p)$  é a v.a. indicadora do  $i$ -ésimo voto ser para  $D$ , para  $1 \leq i \leq n$ . Então

$$S_n = \sum_{i=1}^n V_i \sim \text{Binomial}(n, p)$$

é o total de entrevistados a favor de D. Ademais, a razão  $\frac{S_n}{n}$  é uma estimativa para  $p$ .

Queremos determinar  $n$  para obtermos uma estimativa para  $p$  com erro de 4 pontos percentuais com pelo menos 95% de certeza (grau de confiança). i.e.,

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \geq 0,04\right) < 0,05. \quad (27)$$

Usando a desigualdade de Chebyshev (18),

$$\mathbb{P}(|S_n - np| \geq 0,04n) \leq \frac{np(1-p)}{0,0016n^2} \leq \frac{1}{4 \cdot 0,0016 \cdot n}$$

e fazendo o lado direito igual a 0,05 e resolvendo para  $n$  obtemos  $n = 3125$  entrevistas para que  $S_n/n$  estime  $p$  dentro dos parâmetros de exigência. Acontece, que Chebyshev é uma desigualdade fraca devido a sua generalidade com respeito às v.a.'s, ela não usa nenhuma particularidade da distribuição binomial que, pela proposição anterior, parece estar bem concentrada em torno da média.

A eq. (27) equivale a

$$\mathbb{P}\left(\frac{S_n}{n} \leq p - 0,04\right) + \mathbb{P}\left(\frac{S_n}{n} \geq p + 0,04\right) < 0,05$$

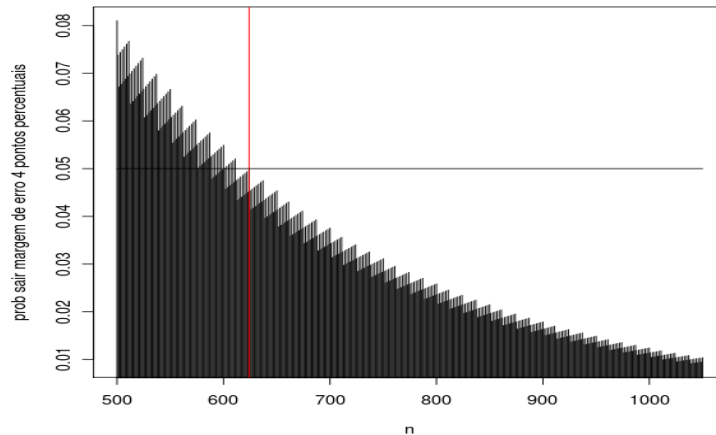
e a soma a esquerda da desigualdade é

$$F_{S_n}((p - 0,04)n) + F_{S_n}(((1 - p) - 0,04)n)$$

que é máxima para  $p = 0,5$  para todo  $n$  fixo, portanto

$$\begin{aligned} \mathbb{P}\left(\frac{S_n}{n} \leq p - 0,04\right) + \mathbb{P}\left(\frac{S_n}{n} \geq p + 0,04\right) &\leq F_{S_n}((0,5 - 0,04)n) + F_{S_n}((0,5 - 0,04)n) \\ &= 2F_{S_n}(0,46n) \end{aligned}$$

que é menor que 0,05 para  $n \geq 624$ .



Usando R:

```
n <- 500:1000
plot(n, 2 * pbinom(.46 * n, n, 1/2), type = "h")
```

Essa estimativa é bem próxima a de dados reais. Informações obtidas no sítio do IBOPE referentes ao 1º turno das eleições municipais de 2008 exibem a seguinte quantidade de entrevistados de acordo como o erro e o grau de confiança:

cidade	no. de eleitores	no. de entrevistas	proporção eleitores entrevistados	margem de erro	grau de confiança
São Paulo	8.198.282	1.204	0,01468%	3%	95%
Rio de Janeiro	4.579.282	1.204	0,02629%	3%	95%
Belo Horizonte	1.772.227	1.204	0,06793%	3%	95%
Santo André	533.428	504	0,09448%	4%	95%
Diadema	301.229	504	0,1673%	4%	95%
São Carlos	154.572	504	0,326%	4%	95%
Cubatão	91.693	504	0,5496%	4%	95%
Registro	41.001	504	1,2292%	4%	95%
Campinas	724.143	2.000	0,2761%	3%	99%
São José dos Campos	414.353	2.000	0,4826%	3%	99%
Ribeirão Preto	388.690	2.000	0,5145%	3%	99%

## Distribuição de Poisson

A notação  $X \sim \text{Poisson}(\lambda)$  indica que  $X$  é uma *v.a. de Poisson* com parâmetro  $\lambda > 0$ , a qual assume valores no conjunto dos naturais com f.m.p. dada por

$$\text{Po}_{\lambda}(x) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad \forall x \in \mathbb{N}.$$

A média e a variância de  $X \sim \text{Poisson}(\lambda)$  é

$$\mathbb{E}(X) = \text{Var}(X) = \lambda.$$

**Exercício 61** Use que  $\sum_{j \geq 0} \frac{\lambda^j}{j!} = e^{\lambda}$  para provar que  $\mathbb{E}(X) = \lambda$ . Prove que  $\mathbb{E}(X^2) = \lambda(\lambda + 1)$ . Conclua que  $\text{Var}(X) = \lambda$ .

Ocasões em que aparecem uma variável aleatória de Poisson são aquelas em que, geralmente, estamos interessados no número de eventos que ocorrem em um intervalo fixo de tempo (ou numa região de área fixa, ou de volume fixo) se esses eventos ocorrem com uma *taxa média conhecida* e independentemente do tempo desde a última ocorrência. Por exemplo,

1. o número de erros de impressão numa página de livro;



2. o número de chamadas que chega a uma central telefônica;
3. o número de partículas  $\alpha$  descarregadas por um material radioativo em um período fixo de tempo;

o que é justificado pelo seguinte. Suponha que em  $n$  ensaios independentes de Bernoulli com probabilidade de sucesso  $p = p(n)$  a taxa média de sucesso é  $\lambda = np$ . A probabilidade de  $x$  sucessos é  $\text{Bi}_{n,p}(x) = \mathbb{P}(X = x)$ , para  $X \sim \text{Binomial}(n, p)$ .

Para  $x = 0$

$$\text{Bi}_{n,p}(0) = (1 - p)^n = \left(1 - \frac{\lambda}{n}\right)^n$$

portanto, para  $n$  suficientemente grande  $\text{Bi}_{n,p}(0) \approx e^{-\lambda}$  (pela definição de  $e$  como limite de uma sequência). Agora, usando (26) podemos concluir que

$$\begin{aligned}\text{Bi}_{n,p}(1) &\approx \text{Bi}_{n,p}(0)\lambda \approx \lambda e^{-\lambda} \\ \text{Bi}_{n,p}(2) &\approx \text{Bi}_{n,p}(1)\lambda/2 \approx (\lambda^2/2!)e^{-\lambda} \\ \text{Bi}_{n,p}(3) &\approx \text{Bi}_{n,p}(2)\lambda/6 \approx (\lambda^3/3!)e^{-\lambda}\end{aligned}$$

o que podemos estender usando indução para

$$\text{Bi}_{n,p}(x) \approx \frac{\lambda^x e^{-\lambda}}{x!} = \text{Po}_{\lambda}(x)$$

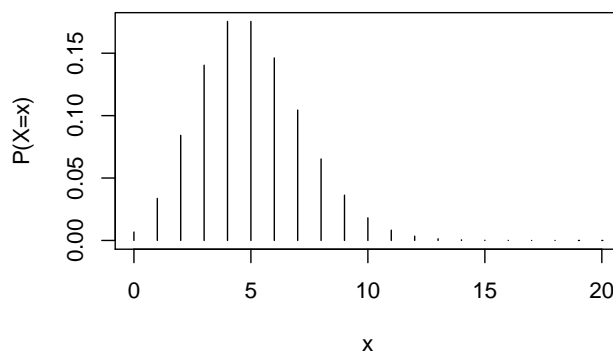
que é conhecido como a *aproximação de Poisson para a distribuição binomial*; em resumo fixado  $\lambda$  e fixado  $x$ , se  $X_n \sim \text{Binomial}(n, \lambda/n)$  e  $Y \sim \text{Poisson}(\lambda)$ , então

$$\lim_{n \rightarrow \infty} \mathbb{P}(X_n = x) = \mathbb{P}(Y = x)$$

e uma prova pode ser vista [aqui](#).

**Exemplo 65** Um telefone recebe em média 5 chamadas por minuto. Supondo que a distribuição de Poisson seja um modelo adequado para essa situação, qual a probabilidade com que o telefone não receba chamadas durante um intervalo de 1 minuto?

$$\text{Po}_5(0) = \frac{5^0 e^{-5}}{0!} = 0,0067$$



Usando R:

```
plot(0:20,dpois(0:20,5),type="h") calcula  $Po_5(0)$   
dpois(0,5) desenha o gráfico acima.
```

**Exemplo 66** O número de partículas que contaminam a superfície de um CD no processo de fabricação tem distribuição de Poisson. O número médio de partículas é  $0,1$  partículas/cm<sup>2</sup> e a área de um CD é  $100\text{cm}^2$ . Seja  $X$  o número de partículas num CD;  $X \sim \text{Poisson}(10)$ .

1. a probabilidade de ter 12 partículas é

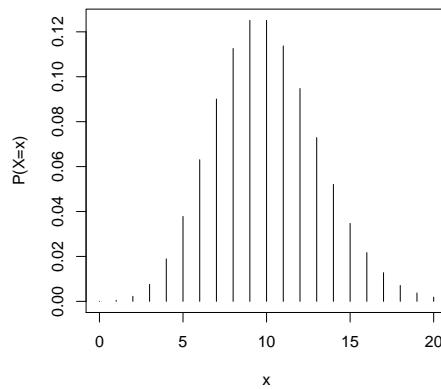
$$\mathbb{P}(X = 12) = \frac{e^{-10} 10^{12}}{12!} = 0,095$$

2. a probabilidade de ter 0 partículas é

$$\mathbb{P}(X = 0) = \frac{e^{-10} 10^0}{0!} = 4,54 \times 10^{-5}$$

3. a probabilidade de ter  $\leq 12$  partículas é

$$\mathbb{P}(X \leq 12) = \sum_{x \leq 12} \frac{e^{-10} 10^x}{x!} = 0,792$$



Usando R:

```
dpois(12,10) ou ppois(12,10) - ppois(11,10) para item 1  
ppois(12,10) para o item 3  
plot(0:20,dpois(0:20,10),type="h") para o gráfico?
```

**Exemplo 67** Suponha que essas notas tenham erros tipográficos por página que segue uma distribuição de Poisson com  $\lambda = 1/2$ . Qual é a probabilidade de haver pelo menos um erro nessa página? Se  $X$  é o número de erros por página

$$\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X = 0) = 1 - e^{-1/2} \approx 0,393.$$

**Exemplo 68** Um analisador conta um número médio de 4 partículas por milésimo de segundo. Com que probabilidade um dado milésimo de segundo o contador contará 6 partículas?

$$\text{Po}_4(6) = \frac{e^{-4} 4^6}{6!} = 0,1042.$$

Usando a desigualdade de Chebyshev (18) a probabilidade do evento  $X \in (\mu - 2\sigma, \mu + 2\sigma)$ , em que  $X$  é o número de partículas num determinado milésimo de segundo, é

$$\mathbb{P}(0 \leq X \leq 4) \geq \frac{3}{4}.$$

**Exercício 62** Prove que a f.m.p. de  $X \sim \text{Poisson}(\lambda)$  satisfaz  $\text{Po}_\lambda(x+1) = \frac{\lambda}{x+1} \text{Po}_\lambda(x)$ .

## Distribuição Hipergeométrica

Uma v.a. hipergeométrica com parâmetros  $n, a, s$  tem f.m.p.

$$H_{n,a,s}(x) = \frac{\binom{a}{x} \binom{n-a}{s-x}}{\binom{n}{s}}, \quad \forall x \in \left\{ \max\{0, s - (n - a)\}, \dots, \min\{a, n\} \right\}.$$

A média e a variância de uma v.a. hipergeométrica são dadas por

$$\mathbb{E}(X) = sp \quad \text{e} \quad \text{Var}(X) = sp(1-p) \frac{n-s}{n-1}$$

em que  $p = \frac{a}{n}$ . A prova dessas fórmulas é similar à prova para v.a. binomial, segue da identidade  $\mathbb{E}(X^k) = \frac{sa}{n} \mathbb{E}((Y+1)^k)$  em que  $Y$  é uma v.a. hipergeométrica com parâmetros  $n-1, a-1, s-1$ .

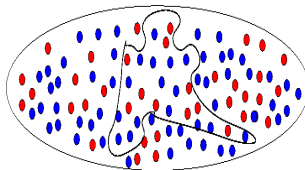
**Exercício 63** Mostre as identidades

$$i \binom{a}{i} = a \binom{a-1}{i-1} \quad \text{e} \quad s \binom{n}{s} = n \binom{n-1}{s-1}$$

e obtenha delas a identidade  $\mathbb{E}(X^k) = \frac{as}{n} \mathbb{E}((Y+1)^k)$ . Em seguida, deduza as igualdade para média e variância.

Uma coleção de  $n$  objetos contém

1.  $a$  objetos azuis,
2.  $n - a$  objetos vermelhos.



Uma amostra com  $s$  elementos é selecionada. Qual a probabilidade da amostra conter  $x$  ( $x \leq a$ ) bolas azuis? O número de bolas azuis é uma v.a. hipergeométrica.

**Exemplo 69** Qual a probabilidade de acertar 4 dos 6 números sorteados na mega-sena? Os parâmetros são  $a = s = 6$  e  $n = 60$ , logo

$$H_{60,6,6}(4) = \frac{\binom{6}{4} \binom{54}{2}}{\binom{60}{6}} = 0,0004287524.$$

Usando R:

$H_{n,a,s}(x)$  é calculado com

`dhyper(x, a, n - a, s)`

Nesse exemplo `dhyper(4, 6, 54, 6)`.

**Exemplo 70** Um comprador de componentes elétricos compra os componentes em lote de 10. A política de controle de qualidade é inspecionar 3 componentes escolhidos aleatoriamente e comprar o lote somente se os 3 não apresentarem defeitos. Se 30% dos lotes têm 4 componentes com defeito e 70% apenas 1, qual é a proporção de lotes aceitos?

Consideremos os eventos:  $A \equiv$  “aceita um lote” e  $B \equiv$  “lote com 4 peças com defeito”.

$$\mathbb{P}(A) = \mathbb{P}(A | B)\mathbb{P}(B) + \mathbb{P}(A | \bar{B})\mathbb{P}(\bar{B}) = \frac{\binom{4}{0} \binom{6}{3}}{\binom{10}{3}} \frac{3}{10} + \frac{\binom{1}{0} \binom{9}{3}}{\binom{10}{3}} \frac{7}{10} = \frac{54}{100}.$$

**Exemplo 71** Um produto é distribuído em lotes de 40 unidades. Um lote é inaceitável se três ou mais itens apresentam defeito. O departamento de controle de qualidade de um comprador adotou o plano de selecionar 5 itens e rejeitar o lote se um item inspecionado for defeituoso.

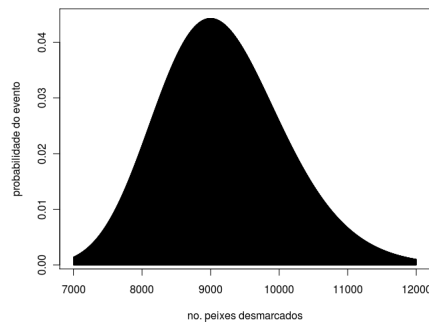
Num lote com 3 itens defeituoso, a probabilidade de haver um defeituoso numa amostra de 5 é

$$H_{40,3,5}(1) = \frac{\binom{3}{1} \binom{37}{4}}{\binom{40}{5}} = 0,3017$$

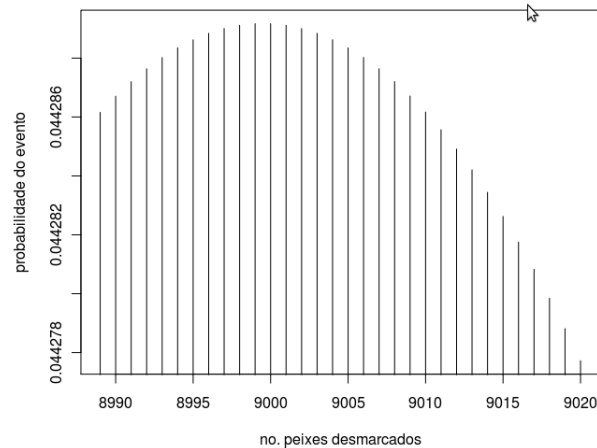
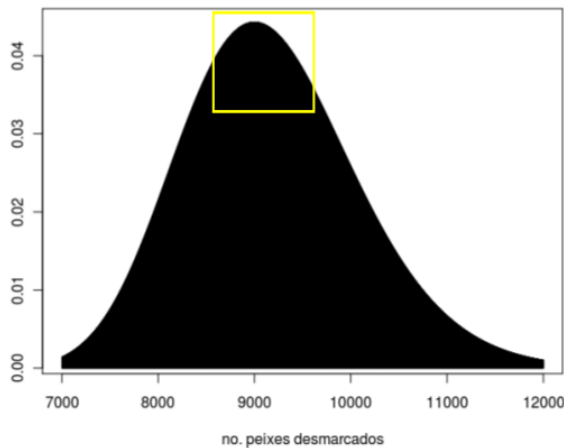
ou seja, detecta um lote ruim em apenas 30% dos casos. Como já vimos, se  $X$  é o número de itens defeituosos na amostra, então  $X \in (\mu - 2\sigma, \mu + 2\sigma)$  com probabilidade pelo menos  $3/4$ , pela desigualdade de Chebyshev. Nesse caso  $\mu = 0,375$  e  $\sigma = 0,558$  e temos então que menos que 2 itens em 5 são defeituosos com probabilidade  $\geq 3/4$ , ou seja, em 75% dos casos.

**Exemplo 72 (Estimativa de máxima verossimilhança)** Num lago 1000 peixes foram capturados, marcados e devolvidos. Uma nova captura de 1000 peixes é feita 100 deles estão marcados. O que pode ser dito a respeito do tamanho da população de peixes no lago?

A probabilidade do evento em função do número (desconhecido) de peixes desmarcados tem gráfico de barras



*olhando mais de perto*



*sugere uma população de aproximadamente 9.000 + 1.000 peixes (máxima verossimilhança — estimativa que maximiza a probabilidade do evento ocorrido, aqui essa estimativa por ser feita de modo análogo a (26)).*

## Outras distribuições

### Distribuição Geométrica

Uma v.a. tem *distribuição geométrica* com parâmetro  $p$  se tem f.m.p. dada por

$$G_p(x) = (1 - p)^{x-1} p, \quad \forall x \in \{1, 2, \dots\}$$

que correspondente ao número de ensaios de Bernoulli independentes com parâmetro  $p$  até ocorrer um sucesso. A notação  $X \sim \text{Geometrica}(p)$  indica que  $X$  é uma v.a. geométrica com parâmetro  $p$ . A média e a variância de  $X$  são

$$\mathbb{E}(X) = \frac{1}{p} \text{ e } \text{Var}(X) = \frac{(1 - p)}{p^2}.$$

Usando R:

$G_p(x)$  é calculado com

`dgeom(x,p)`

$F_X(x)$  para  $X \sim \text{Geometrica}(p)$  é calculado com

`pgeom(x,p)`

### Distribuição binomial negativa

Uma *v.a. binomial negativa* com parâmetros  $p$  e  $r$  tem f.m.p. dada por

$$\text{NBi}_{p,r}(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, \quad \forall x \in \{r+1, r+2, \dots\}$$

que correspondente ao número de ensaios de Bernoulli independentes com parâmetro  $p$  até ocorrer um total de  $r$  sucessos.

**Exercício 64** Mostre que se  $X$  é *v.a. binomial negativa* com parâmetros  $p$  e  $r$  e  $Y$  é *v.a. binomial negativa* com parâmetros  $p$  e  $r+1$ , então

$$\mathbb{E}(X^k) = \frac{r}{p} \mathbb{E}((Y-1)^{k-1}).$$

Deduza daí a média e a variância de uma *v.a. binomial negativa* com parâmetros  $p$  e  $r$ .

Usando R:

$\text{NBi}_{p,r}(x)$  é calculado com

`dnbinom(x,r,p)`

A f.d.a.  $F(x)$  é calculado com

`dnbinom(x,r,p)`

### Distribuição zeta (ou Zipf)

Uma *v.a. binomial zeta* com parâmetros  $\alpha$  tem f.m.p. dada por

$$f(x) = \frac{C}{x^{\alpha+1}}, \quad \forall x \in \{1, 2, \dots\}.$$

A constante é

$$C = \left( \sum_{k \geq 1} \left( \frac{1}{x} \right)^{\alpha+1} \right)^{-1}.$$

A distribuição deve seu nome a função zeta de Riemann e deve-se a George K. Zipf sua popularização.

## Distribuição de Uniforme

$X \sim \text{Uniforme}(k)$  se  $X$  assume valores  $x_1, x_2, \dots, x_k$  com a mesma probabilidade, a saber  $1/k$ .

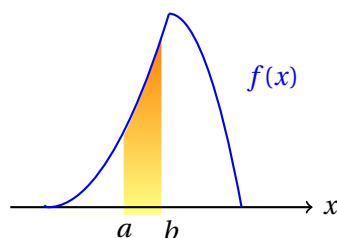
## §7 Principais modelos contínuos

Uma variável aleatória  $X$  é contínua se sua função de distribuição acumulada  $F$  pode ser escrita como

$$F(x) = \int_{-\infty}^x f(t) dt.$$

para alguma função integrável  $f$  chamada função de densidade de probabilidade (f.d.p.) de  $X$  e valem

1.  $f(x) \geq 0$  para todo  $x \in \mathbb{R}$ ;
2.  $\int_{-\infty}^{+\infty} f(x) dx = 1$ ;
3.  $\mathbb{P}(a \leq X \leq b) = \int_a^b f(x) dx = F(b) - F(a)$  para todo  $a \leq b$ ;



$\mathbb{P}(a \leq X \leq b) = \text{área da região delimitada pelo gráfico e pelo eixo } x \text{ no intervalo } [a, b]$

4.  $\mathbb{P}(X = a) = 0$ , para qualquer  $a \in \mathbb{R}$ .

Note-se o aparente paradoxo criado pelo item 4. A v.a. assume o valor associado ao resultado do experimento, mas a probabilidade de assumir qualquer valor particular é zero.

O valor esperado de  $X$  e a variância de  $X$  são dados, respectivamente, por

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x f(x) dx$$

$$\text{Var}(X) = \int_{-\infty}^{+\infty} (x - \mathbb{E}(X))^2 f(x) dx = \mathbb{E}(X^2) - \mathbb{E}(X)^2.$$

## Distribuição uniforme contínua

Uma v.a. contínua  $X$  é *uniforme* no intervalo  $[a, b]$ , para  $a < b$ , denotado  $X \sim \text{Uniforme}([a, b])$  se sua f.d.p. é

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{se } x \in [a, b] \\ 0 & \text{caso contrário} \end{cases}$$

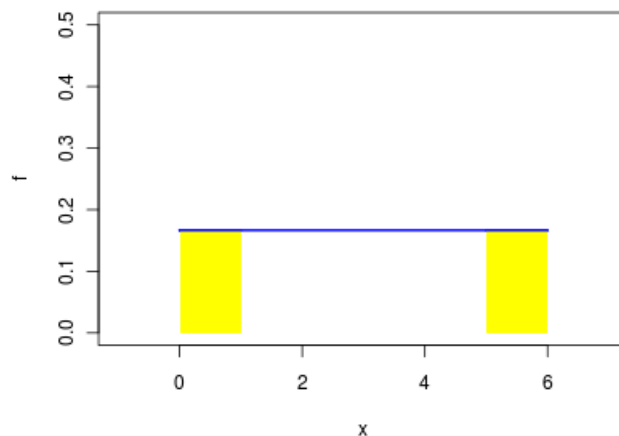
e denotamos esse fato por  $X \sim \text{Uniforme}([a, b])$ .

A média e a variância são

$$\mathbb{E}(X) = \int_a^b x \frac{1}{b-a} dx = \frac{a+b}{2}$$

$$\text{Var}(X) = \int_a^b x^2 \frac{1}{b-a} dx - \left( \frac{a+b}{2} \right)^2 = \frac{(b-a)^2}{12}$$

**Exemplo 73** Num teste, tubos de PVC de 6m são submetidos a grande pressão d'água até que o primeiro vazamento ocorra. A distância do início do tubo até o vazamento é anotada.  $X \sim \text{Uniforme}([0, 6])$  denota a distância anotada para um tubo escolhido ao acaso. Qual a probabilidade de que o vazamento esteja a  $a \leq 1$ m das extremidades?



$$\mathbb{P}(\{0 \leq X \leq 1\} \cup \{5 \leq X \leq 6\}) = \mathbb{P}(0 \leq X \leq 1) + \mathbb{P}(5 \leq X \leq 6) = \int_0^1 \frac{1}{6} dx + \int_5^6 \frac{1}{6} dx = \frac{1}{3}$$

Usando R:

$\mathbb{P}(0 \leq X \leq 1)$  é calculado com

`punif(1,min = 0,max = 6)`

e  $\mathbb{P}(5 \leq X \leq 6)$  é calculado com

`punif(6,min = 0,max = 6) - punif(5,min = 0,max = 6).`



**Exemplo 74** Seja  $g: [0, 1] \rightarrow [0, 1]$  uma função contínua. Se  $X \sim \text{Uniforme}([0, 1])$  com f.d.p.  $f$ , então

$$\mu = \mathbb{E}(g(X)) = \int_0^1 g(x)f(x)dx = \int_0^1 g(x)dx$$

e

$$\sigma^2 = \mathbb{E}(Y - \mu)^2 = \int_0^1 (g(x) - \mu)^2 dx < 1$$

pois  $|g(x) - \mu| \leq 1$ .

Agora, sejam  $X_1, \dots, X_n$  valores escolhidos aleatoriamente em  $[0, 1]$  de acordo com a distribuição uniforme. Uma estimativa para a integral de  $g$  no intervalo  $[0, 1]$  é obtida tomando a média

$$A_n = \frac{g(X_1) + \dots + g(X_n)}{n}.$$

Pela desigualdade de Chebyshev

$$\mathbb{P}(|A_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} < \frac{1}{n\varepsilon^2}$$

para qualquer erro  $\varepsilon > 0$ . Por exemplo, para  $\varepsilon = 0,001$ , se  $n = 100.000.000$  então

$$|A_n - \mu| = \left| A_n - \int_0^1 g(x)dx \right| < 0,001$$

com probabilidade  $> 0,99$ .

**Exercício 65** Determine a f.d.a. de  $X \sim \text{Uniforme}([a, b])$  e esboce o seu gráfico.

## Distribuição Exponencial

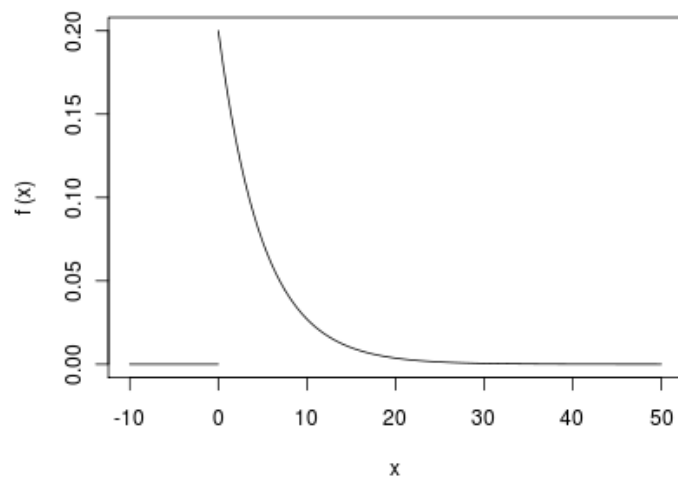
Uma v.a. contínua  $X$  é *exponencial* com parâmetro  $\alpha > 0$ , e denotamos esse fato por  $X \sim \text{Exp}(\alpha)$ , se sua f.d.p. é

$$f(x) = \begin{cases} \alpha e^{-\alpha x}, & \text{se } x \geq 0 \\ 0 & \text{caso contrário.} \end{cases}$$

A média e a variância de  $X \sim \text{Exp}(\alpha)$  é

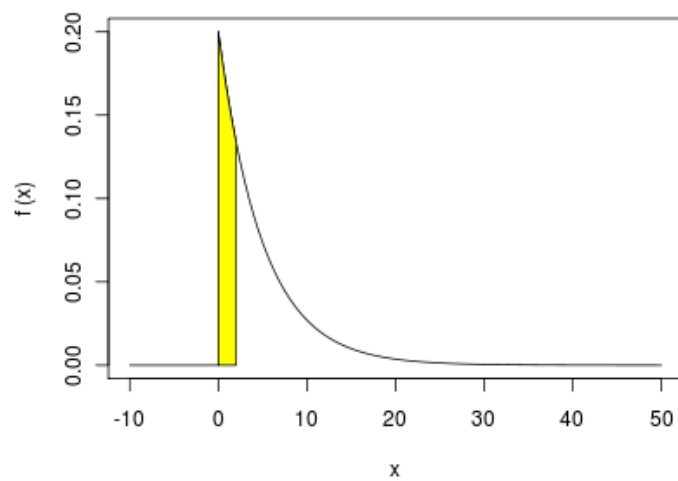
$$\begin{aligned} \mathbb{E}(X) &= \int_0^\infty x \alpha e^{-\alpha x} dx = \frac{1}{\alpha}, \\ \sigma^2 = \text{Var}(X) &= \int_0^\infty x^2 \alpha e^{-\alpha x} dx - \left(\frac{1}{\alpha}\right)^2 = \frac{1}{\alpha^2}. \end{aligned}$$

**Exemplo 75** Seja  $X \sim \text{Exp}(0,2)$  o intervalo de tempo (em min) entre emissões consecutivas de uma fonte radioativa.



A probabilidade de haver uma emissão em  $< 2$  min é

$$\mathbb{P}(X < 2) = \int_0^2 0,2e^{-0,2x} = 0,33$$



A probabilidade do intervalo ser  $\geq 7$  dado que foi  $\geq 5$

$$\begin{aligned} \mathbb{P}(X \geq 7 \mid X \geq 5) &= \frac{\mathbb{P}([X \geq 7] \cap [X \geq 5])}{\mathbb{P}(X \geq 5)} = \frac{\mathbb{P}(X \geq 7)}{\mathbb{P}(X \geq 5)} \\ &= \frac{\int_7^\infty 0,2e^{-0,2x} dx}{\int_5^\infty 0,2e^{-0,2x} dx} = 0,67 \end{aligned}$$

Usando R:  
 $\mathbb{P}(X < 2)$  é calculado com  
`pexp(2,.2)`  
e  $\mathbb{P}(X \geq 7)$  é calculado como  $1 - \mathbb{P}(X < 7)$  que é dado por  
 $1 - \text{pexp}(7,.2)$   
e  $\mathbb{P}(X \geq 5)$  é calculado por  
 $1 - \text{pexp}(5,.2)$ .  
O gráfico acima é dado por  
`plot(seq(-10,50),dexp(seq(-10,50),0.2),type="l")`.

Notemos que, no exemplo acima, ocorreu de  $\mathbb{P}(X \geq 7 \mid X \geq 5) = \mathbb{P}(X \geq 2)$ . Dizemos que uma variável aleatória  $X$  é **sem memória** se

$$\mathbb{P}(X > s + t \mid X > t) = \mathbb{P}(X > s) \quad (28)$$

para quaisquer  $s, t \geq 0$ .

**Exercício 66** Prove que  $X \sim \text{Exp}(\alpha)$  é sem memória.

**Exercício 67** Determine a f.d.a. de  $X \sim \text{Exp}(\alpha)$  e esboce o seu gráfico.

**Exercício 68** Seja  $X$  uma v.a. contínua que interpretamos como a vida útil (tempo) de um item, sejam  $F$  e  $f$  a f.d.a. e f.d.p. de  $X$ , respectivamente. A taxa de falhas é a função definida por

$$\lambda(t) = \frac{f(t)}{1 - F(t)}.$$

De um argumento que mostra que  $\lambda$  pode ser interpretada como a probabilidade (condiciona a idade) de que um item com idade  $t$  apresente falha (i.e.,  $\mathbb{P}(t < X < t + \Delta t \mid X > t)$ ).

Determine  $\lambda(t)$  no caso  $X \sim \text{Exp}(\alpha)$ .

**Exemplo 76** Suponha que um sistema contenha componentes cujo tempo de falha é  $T \sim \text{Exp}(1/5)$  em anos. Se 5 desses componentes são instalados em sistemas diferentes, qual é a probabilidade de que pelo menos 2 ainda estejam funcionando no final de 8 anos?

$$\mathbb{P}(T > 8) = \frac{1}{5} \int_8^\infty e^{-t/5} dt = 0,2$$

assim, se  $X$  é a quantidade de componentes funcionando após 8 anos

$$\mathbb{P}(X \geq 2) = \sum_{x=2}^5 \text{Bi}_{5,0.2}(x) = 1 - \sum_{x=0}^1 \text{Bi}_{5,0.2}(x) = 0,26.$$

## §8 Distribuição Normal e Teorema Central do Limite

A v.a.  $X$  tem distribuição *normal* com parâmetros  $\mu$  e  $\sigma^2$ , abreviado por  $X \sim N(\mu; \sigma^2)$ , se sua função densidade de probabilidade é dada por

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

Propriedades da Normal:

- $\mathbb{E}(X) = \mu$  e  $\text{Var}(X) = \sigma^2$ ;
- $f(x) \rightarrow 0$  quando  $x \rightarrow \pm\infty$ ;
- $\mu$  é ponto de máximo de  $f(x)$  e  $\mu - \sigma$  e  $\mu + \sigma$  são pontos de inflexão de  $f(x)$ ;
- a curva é simétrica com relação a  $x = \mu$ ;
- se  $X \sim N(\mu; \sigma^2)$  então  $aX + b \sim N(a\mu + b; a^2\sigma^2)$ , para  $a, b \in \mathbb{R}$ .

Para verificar a última propriedade listada acima, definimos  $Y = aX + b$  e temos

$$F_Y(x) = \mathbb{P}(Y \leq x) = \mathbb{P}(aX + b \leq x) = \mathbb{P}\left(X \leq \frac{x-b}{a}\right) = F_X\left(\frac{x-b}{a}\right)$$

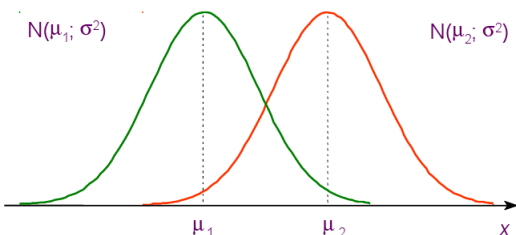
se  $a > 0$ , logo a densidade de  $Y$  é

$$F'_Y(x) = \frac{1}{a} F'_X\left(\frac{x-b}{a}\right) = \frac{1}{a\sigma\sqrt{2\pi}} e^{-\frac{((x-b)/a-\mu)^2}{2\sigma^2}} = \frac{1}{a\sigma\sqrt{2\pi}} e^{-\frac{(x-(b+a\mu))^2}{2a^2\sigma^2}}$$

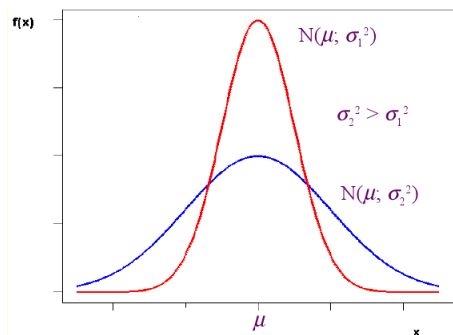
portanto  $Y \sim N(a\mu + b; a^2\sigma^2)$ , caso  $a < 0$  a mesma conclusão vale (verifique).

**Exercício 69** Prove essas propriedades.

O gráfico da densidade de duas normais com médias diferentes mas mesma variância é exibido a seguir



e se as médias são iguais mas variâncias distintas:



Usando R:

A densidade da normal,  $f(x)$ , com parâmetros  $\mu = m$  e  $\sigma^2 = s^2$  pode ser calculada com `dnorm(x,m,s)`.

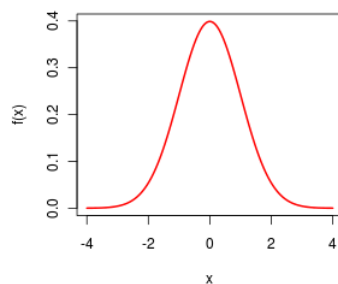
Por exemplo

```
x = seq(-4,4,length = 200)
```

```
y = dnorm(x,0,1)
```

```
plot(x,y,type = "l",col = "red")
```

desenha o gráfico da f.d.p. com média 0 e variância 1 no intervalo  $(-4,4)$ :



### Distribuição normal padrão $N(0;1)$

O problema com o qual nos deparamos agora é que

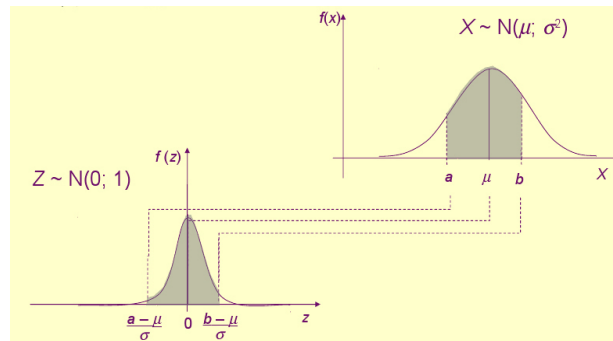
$$\mathbb{P}(a < X < b) = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

não tem solução analítica. Entretanto, decorre da última propriedade da Normal listada acima que se  $X \sim N(\mu; \sigma^2)$  então  $Z \sim N(0;1)$  para

$$Z = \frac{X - \mu}{\sigma}$$

portanto,

$$\mathbb{P}(a < X < b) = \mathbb{P}\left(\frac{a - \mu}{\sigma} < Z < \frac{b - \mu}{\sigma}\right)$$



e agora podemos usar uma tabela com a f.d.a. de  $N(0; 1)$  para determinar a f.d.a. de  $N(\mu; \sigma^2)$  para quaisquer parâmetros  $\mu$  e  $\sigma$ . Ademais, é costume usar  $\Phi(x)$  para denotar a f.d.a. de uma v.a. com distribuição normal.

Distribuição normal padrão  $N(0; 1)$  - Tabela f.d.a.

	0	1	2	3	4	5	6	7	8	9
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999

Para  $Z \sim N(0; 1)$ ,

- quanto é  $\mathbb{P}(Z \leq 0,32)$ ?

↓ segunda casa decimal

	0	1	2
0.0	0.5000	0.5040	0.5080
0.1	0.5398	0.5438	0.5478
0.2	0.5793	0.5832	0.5871
0.3	0.6179	0.6217	0.6255
0.4	0.6554	0.6591	0.6628
0.5	0.6915	0.6950	0.6985

← primeira casa decimal

$$\mathbb{P}(Z \leq 0,32) = 0,6255$$

- quanto é  $\mathbb{P}(0 < Z \leq 1,71)$ ?

1.6	0.9452	0.9463	0.9474
1.7	0.9554	0.9564	0.9573
1.8	0.9641	0.9649	0.9656

$$\mathbb{P}(0 < Z \leq 1,71) = \mathbb{P}(Z \leq 1,71) - \mathbb{P}(Z < 0) = 0,9564 - 0,5 = 0,4564$$

- quanto é  $\mathbb{P}(0,32 \leq Z \leq 1,71)$ ?

$$\mathbb{P}(0,32 \leq Z \leq 1,71) = \mathbb{P}(Z \leq 1,71) - \mathbb{P}(Z < 0,32) = 0,9564 - 0,6255 = 0,3309$$

- quanto é  $\mathbb{P}(Z \leq -1,71)$ ?

$$\mathbb{P}(Z \leq -1,71) = \mathbb{P}(Z \geq 1,71) = 1 - \mathbb{P}(Z < 1,71) = 1 - 0,9564 = 0,0436$$

- quanto é  $\mathbb{P}(-1,71 \leq Z \leq 1,71)$ ?

$$\mathbb{P}(-1,71 \leq Z \leq 1,71) = \mathbb{P}(Z \leq 1,71) - \mathbb{P}(Z < -1,71) = 0,9564 - 0,0436 = 0,9128$$

Ou seja, genericamente, se  $Z \sim N(0; 1)$  então para  $y \geq x \geq 0$  reais temos

- $\mathbb{P}(Z \leq x) = \Phi(x)$
- $\mathbb{P}(y \leq Z \leq x) = \Phi(x) - \Phi(y)$
- $\mathbb{P}(Z \leq -x) = \Phi(-x) = 1 - \Phi(x)$
- (intervalo central)  $\mathbb{P}(-x \leq Z \leq x) = \Phi(x) - \Phi(-x) = 2\Phi(x) - 1$

Usando R:

$\mathbb{P}(Z \leq x)$  é calculado com `pnorm(x, 0, 1)`.

$\mathbb{P}(Z \leq 0,32) = \text{pnorm}(0,32, 0, 1)$

$\mathbb{P}(0 < Z \leq 1,71) = \text{pnorm}(1,71, 0, 1) - \text{pnorm}(0, 0, 1)$

$\mathbb{P}(0,32 \leq Z \leq 1,71) = \text{pnorm}(1,71, 0, 1) - \text{pnorm}(0,32, 0, 1)$

$\mathbb{P}(Z \leq -1,71) = \text{pnorm}(-1,71, 0, 1)$

No caso  $X \sim N(m, s^2)$ ,  $\mathbb{P}(X \leq x)$  é calculado com `pnorm(x, m, s)`.

Como encontrar o valor  $z$  da distribuição  $N(0; 1)$  tal que  $\mathbb{P}(Z \leq z) = 0,975$ ?

	0	1	2	3	4	5	6	7	8	9
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817

$$z = 1,96$$

e  $z$  tal que  $\mathbb{P}(0 < Z \leq z) = 0,4664$ ?

$$0,5 + 0,4664 = 0,9664 \implies z = 1,83$$

e  $z$  tal que  $\mathbb{P}(Z \geq z) = 0,0228$ ?

$$1 - 0,0228 = 0,9772 \implies z = 2$$

**Exemplo 77** O tempo gasto no exame vestibular de uma universidade tem distribuição normal, com média 120 min e desvio padrão 15 min. Qual é a probabilidade com que um candidato termine o exame antes de 100 minutos?

Se  $X$  é o tempo gasto no exame vestibular, então  $X \sim N(120; 15^2)$  logo

$$\mathbb{P}(X < 100) = \mathbb{P}\left(Z \leq \frac{100 - 120}{15}\right) = \mathbb{P}(Z \leq -1,33) = 1 - \mathbb{P}(Z < 1,33) = 1 - 0,9082 = 0,0918$$

Qual deve ser o tempo de prova de modo a permitir que 95% dos vestibulandos terminem no prazo estipulado?

Devemos encontrar  $x$  tal que  $\mathbb{P}(X < x) = 0,95$ , ou seja, tal que

$$\mathbb{P}\left(Z \leq \frac{x - 120}{15}\right) = 0,95.$$

Pela tabela  $\mathbb{P}(Z \leq 1,64) = 0,95$  portanto

$$\frac{x - 120}{15} = 1,64$$

ou seja  $x = 144,6$  min.

Qual é o intervalo central de tempo, tal que 80% dos estudantes gastam para completar o exame?

$$\mathbb{P}(a \leq X \leq b) = 0,8 \Leftrightarrow \mathbb{P}\left(\frac{a - 120}{15} \leq Z \leq \frac{b - 120}{15}\right) = 0,8$$

Pela tabela  $\mathbb{P}(-1,28 \leq Z \leq 1,28) = 0,80$ , portanto  $a = 100,8$  e  $b = 139,2$  minutos.

Usando R:

No caso  $X \sim N(120, 15^2)$ ,  $\mathbb{P}(X \leq x)$  é calculado com `pnorm(x, 120, 15)`.

$\mathbb{P}(X < 100) = \text{pnorm}(100, 120, 15)$

**Exemplo 78** Um sistema considera que um sinal digital será transmitido quando a tensão exceder 0,9V. Na detecção do sinal o ruído tem distribuição  $N(0; 0,45)$ . Qual a probabilidade de detectar um sinal quando nada tiver sido enviado?

Se  $R \sim N(0; 0,45)$  é a tensão do ruído, então

$$\mathbb{P}(R > 0,9) = \mathbb{P}\left(\frac{R}{0,45} > \frac{0,9}{0,45}\right) = \mathbb{P}(Z > 2) = 1 - 0,97725 = 0,02275.$$

O intervalo central que inclui 99% de todas as leituras de ruído é dado por  $x$  tal que

$$\mathbb{P}(-x < R < x) = \mathbb{P}\left(\frac{-x}{0,45} < \frac{R}{0,45} < \frac{x}{0,45}\right) = \mathbb{P}\left(\frac{-x}{0,45} < Z < \frac{x}{0,45}\right) = 0,99.$$



De acordo com a tabela,  $x/0,45 = 2,58$ , ou seja,  $x = 1,16$ .

Suponha que quando um sinal é transmitido a média da v.a.  $R$  mude para 1,8V. Qual a probabilidade do sinal não ser detectado? Seja  $S$  a tensão quando um sinal é transmitido.

$$\mathbb{P}(S < 0,9) = \mathbb{P}\left(\frac{S - 1,8}{0,45} < \frac{0,9 - 1,8}{0,45}\right) = \mathbb{P}(Z < -2) = 0,02275.$$

Essa é a probabilidade com que um sinal é perdido.

Usando R:  
 $\mathbb{P}(R > 0,9) = 1 - \text{pnorm}(0.9, 0, .45)$   
 $\mathbb{P}(S < 0,9) = \text{pnorm}(0.9, 1.8, .45)$

**Exercício 70** Prove que  $X \sim N(0; 1)$  tem média 0 e variância 1. Deduza desse fato que a média e a variância de  $Y \sim N(\mu; \sigma^2)$ .

### Concentração em torno do média

Notemos que se  $k \in \mathbb{N}$ ,  $X \sim N(\mu; \sigma^2)$  e  $Z = (X - \mu)/\sigma \sim N(0; 1)$  então

$$\mathbb{P}(\mu - k\sigma \leq X \leq \mu + k\sigma) = \mathbb{P}(-k \leq Z \leq k) = \mathbb{P}(Z < k) - \mathbb{P}(Z < -k).$$

Para  $k = 1$

$$\mathbb{P}(\mu - \sigma \leq X \leq \mu + \sigma) = \mathbb{P}(-1 \leq Z \leq 1) = 0,6826895.$$

Para  $k = 2$

$$\mathbb{P}(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = \mathbb{P}(-2 \leq Z \leq 2) = 0,9544997.$$

Para  $k = 3$

$$\mathbb{P}(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = \mathbb{P}(-3 \leq Z \leq 3) = 0,9973002.$$

**Exercício 71** Compare os valores obtidos acima com os valores que se obtém a partir da desigualdade de Chebyshev (18).

### Teorema Central do Limite

Vimos que se  $X \sim N(\mu; \sigma^2)$  então  $aX + b \sim N(a\mu + b; a^2\sigma^2)$ , para quaisquer  $a, b \in \mathbb{R}$ . De fato, vale uma afirmação mais geral:

**Teorema 11** Se  $X_i \sim N(\mu_i; \sigma_i^2)$ , para  $i \in \{1, 2, \dots, n\}$ , são variáveis aleatórias independentes, então

$$c_1X_1 + c_2X_2 + \dots + c_nX_n \sim N(c_1\mu_1 + c_2\mu_2 + \dots + c_n\mu_n; c_1^2\sigma_1^2 + c_2^2\sigma_2^2 + \dots + c_n^2\sigma_n^2)$$

para quaisquer  $c_1, c_2, \dots, c_n \in \mathbb{R}$ .

**Corolário 7** Se  $\bar{X} = \frac{\sum_i X_i}{n}$  então

$$\bar{X} \sim N\left(\frac{\sum_i \mu_i}{n}; \frac{\sum_i \sigma_i^2}{n^2}\right).$$

Caso as variáveis tenham a mesma distribuição

$$\bar{X} \sim N\left(\mu; \frac{\sigma^2}{n}\right).$$

**Exemplo 79** Na fabricação de placas retangulares há pequenas perturbações de modo que o comprimento  $C$  de uma placa escolhida ao acaso tem distribuição  $N(2; 0, 1^2)$  e a largura  $L \sim N(5; 0, 2^2)$ ; em centímetros. Qual a probabilidade do perímetro exceder 15 cm?

Se  $Y$  é a v.a. pra o perímetro de uma placa escolhida ao acaso, então  $Y = 2C + 2L$  e pelo teorema acima  $Y \sim N(14; 0, 2)$  logo

$$\mathbb{P}(Y > 15) = \mathbb{P}\left(\frac{Y - 14}{\sqrt{0,2}} > \frac{15 - 14}{\sqrt{0,2}}\right) = \mathbb{P}(Z > 2, 236) = 0, 0129$$

**Exemplo 80** O engarrafamento de um refrigerante de 300ml tem variações de modo que o volume do líquido numa garrafa é uma variável aleatória com distribuição  $N(300, 25^2)$ . Numa inspeção, 10 garrafas são selecionadas e o volume de cada garrafa,  $V_1, V_2, \dots, V_{10}$  é medido, de modo que se a *média amostral*

$$\bar{V} = \frac{V_1 + V_2 + \dots + V_{10}}{10}$$

for menor que 290 (ml) então a engarrafadora é multada. Qual é a probabilidade de multa?

$\bar{V} \sim N(300; \frac{25^2}{10})$  de modo que

$$\mathbb{P}(\bar{V} < 290) = \mathbb{P}(Z < -1, 26) = 0, 1038$$

E se as v.a.  $X_1, \dots, X_{10}$  são independentes, com a mesma distribuição, mesma média, mesma variância mas a distribuição comum a elas não é normal?

Se  $(S_n)_n$  é uma sequência de v.a.s e  $Y$  v.a.

$$S_n \xrightarrow[n \rightarrow \infty]{\mathcal{D}} Y$$

significa que

$$\lim_{n \rightarrow \infty} \mathbb{P}(S_n \leq x) = \mathbb{P}(Y \leq x)$$

de modo que podemos usar  $Y$  como modelo probabilístico aproximado para  $S_n$  e quanto maior  $n$  melhor é a aproximação.

**Teorema 12 (Teorema Central do Limite (TCL))** Sejam  $X_1, X_2, \dots, X_n$  são variáveis aleatórias independentes e com a mesma distribuição, média  $\mu$  e variância  $\sigma^2$  finitas. Então

$$\frac{(X_1 + X_2 + \dots + X_n) - n\mu}{\sigma\sqrt{n}} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} Z$$

em que  $Z \sim N(0; 1)$ .

**Corolário 8** Se  $\bar{X} = \frac{\sum_i X_i}{n}$  então

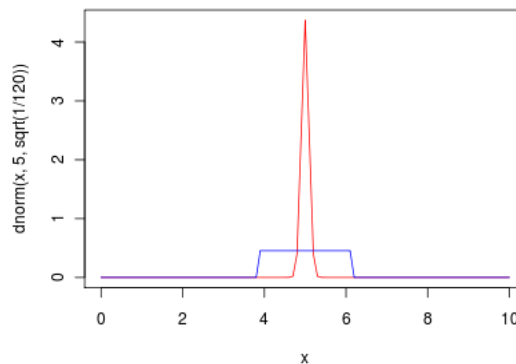
$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} Z \quad (29)$$

em que  $Z \sim N(0; 1)$ . Ou, ainda, se  $Y \sim N\left(\mu; \frac{\sigma^2}{n}\right)$ , então

$$\bar{X} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} Y. \quad (30)$$

Se  $X_1, X_2, \dots, X_n$  são variáveis aleatórias independentes e com a *mesma distribuição* da v.a.  $X$  então as chamamos de **amostra aleatória simples de  $X$  de tamanho  $n$**  e a v.a.  $\bar{X} = \sum_i X_i / n$  é a **média amostral**.

**Exemplo 81** A média amostral de uma amostra aleatória simples com 40 observações independentes de uma v.a.  $X \sim \text{Uniforme}([4, 6])$  tem distribuição aproximada por  $Y \sim N(5; 1/120)$



$$\begin{aligned} \mathbb{P}(X < 5) &= P(Y < 5) = 0,5 \\ \mathbb{P}(X < 4) &= 0 \quad \text{e} \quad \mathbb{P}(Y < 4) = 0,03 \times 10^{-30} \end{aligned}$$

**Exemplo 82** Lâmpadas produzidas numa fábrica têm vida útil em horas regida pela distribuição normal  $N(800; 40^2)$ . Qual a probabilidade de uma amostra aleatória simples de tamanho 16 ter vida útil média menor que 775 horas?

Vida útil média  $\bar{X} \approx Y \sim N(800; 40^2/16)$  (eq. (30)) portanto

$$\mathbb{P}(\bar{X} < 775) \approx \mathbb{P}(Y < 775) = \mathbb{P}(Z < -2,5) = 0,0062$$

**Exemplo 83** As chamadas telefônicas numa empresa têm duração em minutos que segue a distribuição exponencial com parâmetro  $1/3$ . Numa amostra aleatória com 50 chamadas qual é a probabilidade da média amostral não ultrapassar 4 min ?

$X \sim \text{Exp}(1/3)$  tem média 3 e variância 9. Então  $\bar{X} \approx Y \sim N(3; 9/50)$

$$\mathbb{P}(\bar{X} \leq 4) \approx \mathbb{P}(Z \leq 2,36) = 0,991.$$

**Exemplo 84 (Aproximação para a Binomial)** Sejam  $X_1, \dots, X_n$  v.a. independentes com distribuição Bernoulli( $p$ ). Então  $X = X_1 + \dots + X_n \sim \text{Binomial}(n, p)$  e temos as aproximações

$$X \approx Y \sim N(np; np(1-p)) \quad (31)$$

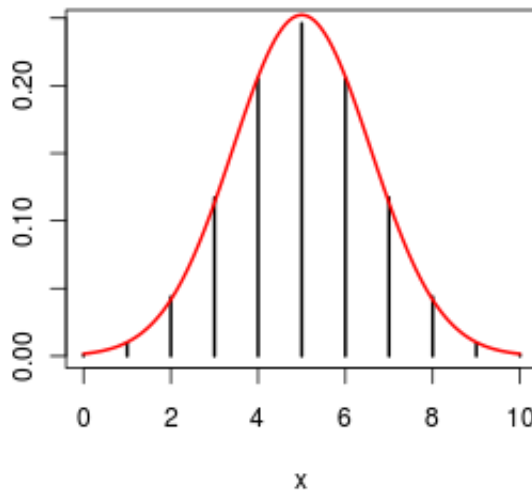
$$\frac{X - np}{\sqrt{np(1-p)}} \approx Z \sim N(0; 1) \quad (32)$$

$$\bar{X} \approx \bar{Y} \sim N\left(p; \frac{p(1-p)}{n}\right) \quad (33)$$

$$\frac{\bar{X} - p}{\sqrt{\frac{p(1-p)}{n}}} \approx Z \sim N(0; 1) \quad (34)$$

además

$$\mathbb{P}(k \leq X \leq l) \approx \mathbb{P}\left(\frac{k - np}{\sqrt{np(1-p)}} \leq Z \leq \frac{l - np}{\sqrt{np(1-p)}}\right) \quad (35)$$



Gráficos Bi(10; 0,5) por  $N(10, \sqrt{2,5})$

Por exemplo, se  $X \sim \text{Binomial}(225; 0,2)$  então por (35)

$$\mathbb{P}(39 \leq X \leq 48) \approx \mathbb{P}(-1 \leq Z \leq 0,5) = 0,5328072 \quad (36)$$

ainda

$$\mathbb{P}(39 \leq X \leq 48) = \sum_{j=39}^{48} \binom{225}{j} (0,2)^j (0,8)^{225-j} = 0,5852713.$$

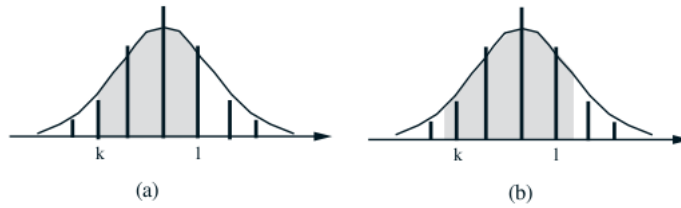
Entretanto

$$0.0417 = \mathbb{P}(X = 39) \approx \mathbb{P}(-1 \leq Z \leq -1) = 0$$

e aproximação seria melhor se fizéssemos

$$\mathbb{P}(X = 39) = \mathbb{P}(38,5 \leq X \leq 39,5) \approx \mathbb{P}(-1,083 \leq Z \leq -0,916) = 0,0403$$

que chamamos de correção de continuidade, o que melhora a aproximação



$$\mathbb{P}(k \leq X \leq l) \approx \mathbb{P}\left(\frac{k - 0,5 - np}{\sqrt{np(1-p)}} \leq Z \leq \frac{l + 0,5 - np}{\sqrt{np(1-p)}}\right) \quad (37)$$

agora, com a mesma técnica temos uma aproximação melhor que (36)

$$\mathbb{P}(39 \leq X \leq 48) \approx \mathbb{P}(-1,083 \leq Z \leq 0,583) = 0,5806491.$$

**Exemplo 85** Um sistema tem 100 componentes cada um com confiabilidade de 90% (independentemente um do outro). O sistema funciona corretamente se  $\geq 87$  componentes estão funcionando. Qual a confiabilidade do sistema aproximadamente? Se  $X$  é a v.a. que denota o no. de componentes em funcionamento então  $X \sim \text{Binomial}(100; 0,9)$  e por (37)

$$\mathbb{P}(X \geq 87) \approx \mathbb{P}(Z \geq -1,166) = 0,8781928$$

(o valor correto, de acordo com a distribuição binomial é 0,8761232)

**Exemplo 86** Um teste tem 200 perguntas com 4 alternativas cada, das quais apenas uma é correta. Qual a probabilidade aproximada que o estudante acerte por chute entre 25 e 30 questões para 80 das 200 questões.

Seja  $X \sim \text{Binomial}(80, 1/4)$  o número de respostas certas. Por (37)

$$\mathbb{P}(25 \leq X \leq 30) \approx \mathbb{P}(1,16 \leq Z \leq 2,71) = 0,1196602.$$

(o valor correto é 0,1192705)

**Exemplo 87 (Estimação via aproximação à Binomial)** (exemplo 64 revisitado)

Numa eleição, seja  $p$  a fração (desconhecida) da população que vota no candidato  $D$ . Para simplificar, assumimos que só há 2 respostas possíveis e um voto em  $D$  é ensaio de Bernoulli com parâmetro  $p$ .

Suponha que serão realizadas  $n$  entrevistas:  $V_i \sim \text{Bernoulli}(p)$  é a v.a. indicadora do  $i$ -ésimo voto ser para  $D$ , para  $1 \leq i \leq n$ . Então

$$S_n = \sum_{i=1}^n V_i \sim \text{Binomial}(n, p)$$

é o total de entrevistados a favor de  $D$ . Ademais, a razão  $\frac{S_n}{n}$  é uma estimativa (pontual) para  $p$ .

$$\begin{aligned} S_n &= V_1 + \dots + V_n \text{ é o no. de votos em } D, S_n \sim \text{Binomial}(n, p); \\ \bar{p} &= \frac{S_n}{n} \text{ é a proporção da amostra de votos em } D \approx N\left(p; \frac{p(1-p)}{n}\right) \\ p &= \text{é a proporção desconhecida da população votos em } D \end{aligned}$$

Queremos  $n$  para uma estimativa com erro de 4 (estimativa intervalar) pontos percentuais com 95% de certeza (100% impossível. Por que?), i.e., queremos  $n$  tal que  $|p - \bar{p}| \leq 0,04$  com probabilidade 0,95. (por que existe tal  $n$ ?)

Para  $\varepsilon = 0,04$

$$\begin{aligned} \mathbb{P}(|p - \bar{p}| \leq \varepsilon) &= \mathbb{P}(-\varepsilon \leq p - \bar{p} \leq \varepsilon) \\ &= \mathbb{P}\left(\frac{-\varepsilon}{\sqrt{p(1-p)/n}} \leq \frac{p - \bar{p}}{\sqrt{p(1-p)/n}} \leq \frac{\varepsilon}{\sqrt{p(1-p)/n}}\right) \\ &\approx \mathbb{P}\left(\frac{-\varepsilon}{\sqrt{p(1-p)/n}} \leq Z \leq \frac{\varepsilon}{\sqrt{p(1-p)/n}}\right) \\ &= \mathbb{P}\left(\frac{-\varepsilon\sqrt{n}}{\sqrt{p(1-p)}} \leq Z \leq \frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}\right) \\ &= 2\mathbb{P}\left(Z \leq \frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}\right) - 1 \end{aligned}$$

Para  $\gamma = 0,95$ , queremos  $z_\gamma$  (da tabela normal padrão) tal que  $2\mathbb{P}(Z \leq z_\gamma) - 1 = \gamma$  ou seja

$$\mathbb{P}(Z \leq z_\gamma) = \frac{1+\gamma}{2}$$

Por exemplo, para  $\gamma = 0,95$ , queremos  $z_\gamma$  tal que  $\mathbb{P}(Z \leq z_\gamma) = 1,95/2$ , donde tiramos  $z_\gamma = 1,96$ .

Descoberto tal  $z_\gamma$  precisamos escolher  $n$  de modo que

$$\frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}} = z_\gamma \text{ ou seja } n = \frac{1}{\varepsilon^2} p(1-p) z_\gamma^2$$

Notemos que  $p \in [0, 1]$ , portanto  $p(1 - p) \leq 1/4$  logo

$$\frac{\varepsilon\sqrt{n}}{\sqrt{4}} \leq \frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}$$

portanto é suficiente termos  $n$  tal que

$$\frac{\varepsilon\sqrt{n}}{\sqrt{4}} = z_\gamma$$

ou seja

$$n = \frac{z_\gamma^2}{4\varepsilon^2}. \quad (38)$$

De volta com os valores  $\varepsilon = 0,004$  e  $\gamma = 0,95$ , já sabemos que  $z_{0,95} = 1,96$  e por (38)

$$n = \frac{1,96^2}{4 \cdot 0,04^2} = 600,25$$

**Exemplo 88** No mesmo caso do exemplo anterior, para uma estimativa com erro de 3 pontos percentuais e 95% de grau de confiança

$$n = \frac{1,96^2}{4 \cdot 0,03^2} = 1067,111.$$

Analogamente,  $n$  para uma estimativa com erro de 3 pontos percentuais com 99% de grau de confiança então  $z_{0,99} = 2,57$  e

$$n = \frac{2,57^2}{4 \cdot 0,03^2} = 1834,69444.$$

**Exemplo 89 (Intervalo de confiança)** No exemplo acima provamos que

$$\mathbb{P}\left(\bar{p} - 1,96\sqrt{4/n} \leq p \leq \bar{p} + 1,96\sqrt{4/n}\right) = 0,95$$

e dizemos que  $(\bar{p} - 1,96\sqrt{4/n}, \bar{p} + 1,96\sqrt{4/n})$  é um *intervalo de confiança para  $p$  com grau de confiança 95%*. Notemos que  $p$  é um valor médio desconhecido e  $\bar{p}$  é uma variável aleatória, portanto o intervalo é aleatório.

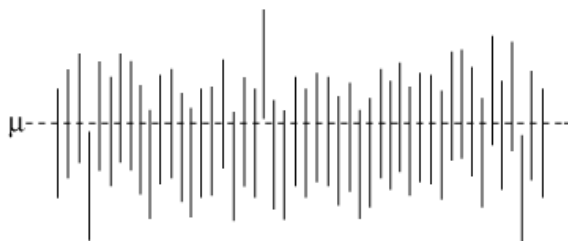
No caso geral, para cada amostra aleatória simples queremos uma estimativa intervalar  $(\bar{X} - \varepsilon, \bar{X} + \varepsilon)$  para a média  $\mu$  da população (desconhecida), variância conhecida, e com grau de confiança  $\gamma$ .

$$\begin{aligned} \gamma &= \mathbb{P}(\bar{X} - \varepsilon \leq \mu \leq \bar{X} + \varepsilon) \\ &= \mathbb{P}\left(-\frac{\varepsilon}{\sigma/\sqrt{n}} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{\varepsilon}{\sigma/\sqrt{n}}\right) \\ &\approx \mathbb{P}\left(-\frac{\varepsilon}{\sigma/\sqrt{n}} \leq Z \leq \frac{\varepsilon}{\sigma/\sqrt{n}}\right) \\ \Rightarrow n &= \left(\frac{z_\gamma}{\varepsilon}\right)^2 \sigma^2 \end{aligned}$$

e a estimativa intervalar para  $\mu$  com grau de confiança 0,95 é

$$\left( \bar{X} - z_\gamma \frac{\sigma}{\sqrt{n}}, \bar{X} + z_\gamma \frac{\sigma}{\sqrt{n}} \right)$$

Pode ser interpretado assim: num número grande de amostras do mesmo tamanho, se obtivermos um intervalo com grau de confiança 0,95 para cada uma delas



então 95% desses intervalos contém o parâmetro  $\mu$ .

**Exemplo 90** A renda per-capita domiciliar numa certa região tem desvio padrão 250 reais e média desconhecida. Se desejamos estimar a renda média da população com erro 50 reais e confiabilidade  $\gamma = 0,95$  quantos domicílios deveremos consultar?

Já sabemos que  $z_\gamma = 1,96$ , então

$$n = \left( \frac{z_\gamma}{\epsilon} \right)^2 \sigma^2 = \left( \frac{1,96}{50} \right)^2 250^2 = 96,04.$$

**Exemplo 91** Um provedor de internet monitora o a duração da conexão dos clientes a fim de dimensionar os seus servidores. A média e a distribuição desse tempo são desconhecidos mas o desvio padrão é  $\sqrt{50}$  minutos. Numa amostra de 500 conexões o valor médio foi 25 minutos; o que podemos dizer a respeito da média com grau de confiança 92%? Como o tamanho da amostra é razoavelmente grande, podemos usar o TCL e aproximar a distribuição por uma normal. Um intervalo de confiança para o tempo de conexão é

$$\left( \bar{X} - z_\gamma \frac{\sigma}{\sqrt{n}}, \bar{X} + z_\gamma \frac{\sigma}{\sqrt{n}} \right) = (24,45, 25,55).$$

Em virtude do uso do TCL, o intervalo acima é com grau de confiança aproximadamente 0,92.

Na prática não conhecemos  $\sigma^2$  e devemos substituí-lo por uma estimativa amostral, que pode ser

$$S^2 = \frac{1}{n-1} \sum (x_i - \bar{X})^2$$

**Exemplo 92** O tempo de reação de um remédio pode ser considerado como tendo distribuição normal. Num teste, 20 pacientes foram sorteados e os tempo anotados:



2,9 3,4 3,5 4,1 4,6 4,7 4,5 3,8 5,3 4,9  
 4,8 5,7 5,8 5,0 3,4 5,9 6,3 4,6 5,5 6,2

então, a variância amostral é  $S^2 = 0,992079$  e o intervalo a 95% é

$$\left( \bar{x} - z_{0,95} \sqrt{\frac{S^2}{n}}, \bar{x} + z_{0,95} \sqrt{\frac{S^2}{n}} \right) = (4,278843, 5,211157).$$

[aproximação de Stirling] Seja  $S_n = X_1 + \cdots + X_n$  com as v.a.  $X_i$  independentes e com distribuição Poisson(1). Então

$$\mathbb{P}(n-1 < S_n \leq n-1) = \mathbb{P}\left(\frac{-1}{\sqrt{n}} < \frac{S_n - n}{\sqrt{n}} \leq 0\right) \approx \int_{-1/\sqrt{n}}^0 \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \approx \frac{1}{\sqrt{2\pi n}}$$

portanto

$$\mathbb{P}(S_n = n) = \frac{e^{-n} n^n}{n!} \approx \frac{1}{\sqrt{2\pi n}}$$

donde

$$n! \approx n^{1/2+n} e^{-n} \sqrt{2\pi}.$$

## §9 Distribuição Conjunta

**Sobre notação:** daqui em diante

escrevemos  $P(X \leq k_1, X \leq k_2, \dots, X \leq k_j)$  para  $P([X \leq k_1] \cap [X \leq k_2] \cap \dots \cap [X \leq k_j])$ .

Se  $X, Y$  são v.a. sobre o mesmo espaço amostral, a **função de distribuição acumulada conjunta** é a função que associa as reais  $a, b$  o valor

$$F(a, b) \stackrel{\text{def}}{=} \mathbb{P}(X \leq a, Y \leq b).$$

A distribuição de  $X$  pode ser obtida a partir de  $F$  do seguinte modo:

$$F_X(a) = \mathbb{P}(X \leq a) = \mathbb{P}(X \leq a, Y < \infty).$$

Agora, se  $b_1 < b_2$  então  $[Y \leq b_1] \subset [Y \leq b_2]$ , portanto

$$[X \leq a] \cap [Y \leq b_1] \subset [X \leq a] \cap [Y \leq b_2]$$

e fazendo  $E_n = [X \leq a] \cap [Y \leq n]$  temos uma sequência crescente de eventos logo (veja exercício 25, página 17)

$$\lim_{n \rightarrow \infty} \mathbb{P}(E_n) = \mathbb{P}\left(\bigcup_{n \geq 0} E_n\right)$$

ou seja

$$\lim_{n \rightarrow \infty} F(a, n) = \mathbb{P}(X \leq a, Y < \infty).$$

Em resumo,

- $F_X(a) = \lim_{b \rightarrow \infty} F(a, b),$
- $F_Y(b) = \lim_{a \rightarrow \infty} F(a, b),$

são as **distribuições marginais** de  $X$  e  $Y$ , respectivamente.

**Exercício 72** Mostre que

$$\mathbb{P}(X > a, Y > b) = 1 - F_X(a) - F_Y(b) + F(a, b).$$

Se  $X$  e  $Y$  são v.a. discretas então

$$f(a, b) = \mathbb{P}(X = a, Y = b)$$

é a **função de massa de probabilidade conjunta**, e para tal

- $f(a, b) \geq 0;$
- $\sum_a \sum_b f(a, b) = 1;$

- $\mathbb{P}((X, Y) \in A) = \sum_{(a,b) \in A} f(a, b)$ , para qualquer conjunto enumerável  $A$  de pares de valores reais;

as **funções de probabilidade marginais** de  $X$  e de  $Y$  são, respectivamente

$$f_X(a) = \sum_b f(a, b) \quad \text{e} \quad f_Y(b) = \sum_a f(a, b).$$

Dizemos que  $X$  e  $Y$  são v.a. **conjuntamente contínuas** se existe uma **função de densidade de probabilidade conjunta**  $f$  tal que

$$F(a, b) = \int_{-\infty}^b \int_{-\infty}^a f(x, y) \, dx \, dy$$

e para tal  $f$

- $f(x, y) \geq 0$ ;
- $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \, dx \, dy = 1$ ;
- $\mathbb{P}((X, Y) \in A) = \iint_A f(x, y) \, dx \, dy$ , para qualquer região  $A$  do plano;

as **distribuições marginais** de  $X$  e de  $Y$  são, respectivamente

- $F_X(a) = \lim_{b \rightarrow \infty} F(a, b) = \int_{-\infty}^a \left( \int_{-\infty}^{\infty} f(x, y) \, dy \right) \, dx$ ,
- $F_Y(b) = \lim_{a \rightarrow \infty} F(a, b) = \int_{-\infty}^b \left( \int_{-\infty}^{\infty} f(x, y) \, dx \right) \, dy$ ,

e as densidades marginais

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y) \, dy \quad \text{e} \quad f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) \, dx.$$

**Exercício 73** *Mostre que*

$$f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y)$$

*quando as derivadas parciais existem.*

**Exemplo 93** *Dois refs de caneta são selecionados ao acaso dentre 3 azuis, 2 vermelhos e 3 verdes. Sejam  $X$  o número de refs azuis e  $Y$  o número de refs vermelhos selecionados. A distribuição conjunta é, para  $x, y \in \{0, 1, 2\}$  sujeito a condição  $x + y \leq 2$ ,*

$$f(x, y) = \frac{\binom{3}{x} \binom{2}{y} \binom{3}{2-x-y}}{\binom{8}{2}}$$

cujos valores está descrito na tabela abaixo

$x \backslash y$	0	1	2
0	3/28	3/14	1/28
1	9/28	3/14	0
2	3/28	0	0

A probabilidade  $\mathbb{P}((X, Y) \in \{(x, y) \mid x + y \leq 1\})$  é  $f(0, 0) + f(0, 1) + f(1, 0) = 9/14$ .

A distribuição marginal de  $X$  é dada pela soma da linha da tabela, por exemplo,

$$f_X(0) = 3/28 + 3/14 + 1/28.$$

Analogamente, a distribuição marginal de  $Y$  é dada pela soma da coluna, por exemplo

$$f_Y(1) = 3/14 + 3/14 + 0.$$

**Exemplo 94** Um banco resolveu apostar num serviço de drive-thru, além do atendimento convencional. Em um dia  $X$  é a proporção de tempo que o drive-thru está em uso e  $Y$  a proporção de tempo que o caixa convencional está em uso, assim  $(X, Y) \in \{(x, y) \mid 0 \leq x \leq 1, 0 \leq y \leq 1\}$ . Estudos indicam que a função de distribuição conjunta é

$$f(x, y) = \begin{cases} \frac{6}{5}(x + y^2) & \text{se } 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{caso contrário.} \end{cases}$$

A probabilidade de nenhuma das alternativas estar ocupada em mais de um quarto do tempo é  $\mathbb{P}(0 \leq X \leq \frac{1}{4}, 0 \leq Y \leq \frac{1}{4})$

$$= \int_0^{1/4} \int_0^{1/4} \frac{6}{5}(x + y^2) dx dy = \int_0^{1/4} \int_0^{1/4} \frac{6}{5}x dx dy + \int_0^{1/4} \int_0^{1/4} \frac{6}{5}y^2 dx dy = \frac{7}{640}.$$

Ademais

$$f_X(x) = \int_{-\infty}^{+\infty} \frac{6}{5}(x + y^2) dy = \int_0^1 \frac{6}{5}(x + y^2) dy = \frac{6}{5}x + \frac{2}{5}$$

é a função de densidade de probabilidade do tempo que o drive-thru está ocupado para  $x \in [0, 1]$ , e  $f_X(x) = 0$  para os outros valores de  $x$ . Também,

$$f_Y(y) = \int_{-\infty}^{+\infty} \frac{6}{5}(x + y^2) dx = \int_0^1 \frac{6}{5}(x + y^2) dx = \frac{6}{5}y^2 + \frac{3}{5}$$

é a função de densidade de probabilidade do tempo que o caixa convencional está ocupado para  $y \in [0, 1]$  e  $f_Y(y) = 0$  para os outros valores de  $y$ .

**Exemplo 95** Sejam  $X$  e  $Y$  as coordenadas de um ponto escolhido no círculo de raio 1 e centro na origem de um sistema de coordenadas com densidade conjunta

$$f(x, y) = \begin{cases} \frac{1}{\pi} & \text{se } x^2 + y^2 \leq 1 \\ 0 & \text{se } x^2 + y^2 > 1 \end{cases}$$

A densidade marginal de  $X$  é

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy = \int_{-\sqrt{1-x^2}}^{+\sqrt{1-x^2}} \frac{1}{\pi} dy = \frac{2}{\pi} \sqrt{1-x^2} \quad (\forall x, x^2 \leq 1),$$

e  $f_X(x) = 0$  nos outros casos. Analogamente,

$$f_Y(y) = \frac{2}{\pi} \sqrt{1-y^2} \quad (\forall y, y^2 \leq 1)$$

e  $f_Y(y) = 0$  nos outros casos.

A distância do ponto escolhido para a origem é a v.a.  $D = \sqrt{X^2 + Y^2}$  cuja distribuição é, para  $a \in [0, 1]$ , dada por  $F_D(a) = \mathbb{P}(\sqrt{X^2 + Y^2} \leq a) =$

$$= \iint_A f(x, y) dx dy = \iint_A \frac{1}{\pi} dx dy = \frac{1}{\pi} \iint_A dx dy = \frac{1}{\pi} \pi a^2 = a^2, \quad A = \{(x, y) : x^2 + y^2 \leq a^2\}.$$

A função de densidade de  $D$  é  $f_D(a) = \frac{2}{a}$  para  $a \in [0, 1]$ , portanto

$$\mathbb{E}(D) = \int_{-\infty}^{+\infty} a f_D(a) da = \int_0^1 2a^2 da = \frac{2}{3}.$$

## Variáveis aleatórias independentes

Dizemos que as v.a.  $X$  e  $Y$  são **independentes** se para quaisquer conjuntos  $A$  e  $B$  de números reais vale

$$\mathbb{P}([X \in A] \cap [Y \in B]) = \mathbb{P}(X \in A) \mathbb{P}(Y \in B). \quad (39)$$

Isso equivale a dizer que para quaisquer  $a, b \in \mathbb{R}$

$$\mathbb{P}([X \leq a] \cap [Y \leq b]) = \mathbb{P}(X \leq a) \mathbb{P}(Y \leq b). \quad (40)$$

como estabelece o seguinte exercício.

**Exercício 74 (Independência de v.a. conjuntamente distribuídas)** Prove que  $X$  e  $Y$  são v.a. independentes, com distribuição conjunta  $F(x, y)$  se e só se

$$F(x, y) = F_X(x) F_Y(y). \quad (41)$$

Decorre daí que  $X$  e  $Y$  independentes equivale a

$$f(x, y) = f_X(x) f_Y(y)$$

para quaisquer reais  $x$  e  $y$ .

**Exemplo 96** Se

$$f(x, y) = \begin{cases} 24xy, & \text{se } x \in [0, 1], y \in [0, 1], x + y \leq 1 \\ 0, & \text{caso contrário.} \end{cases}$$

é a distribuição conjunta de  $X$  e  $Y$  então  $f_X(3/4) = f_Y(3/4) = 9/16$  e  $f(3/4, 3/4) = 0$  portanto

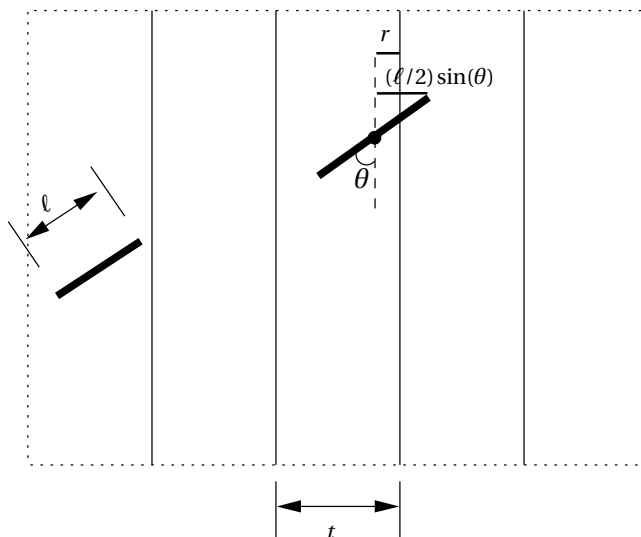
$$f_X(x) f_Y(y) \neq f(x, y)$$

ou seja, embora  $f(x, y) = g(x)h(y)$  as funções  $g$  e  $h$  não podem ser feitas distribuições marginais.

**Exercício 75** Prove que se  $f(x, y) = g(x)h(y)$  para todo  $x \in \mathbb{R}$  e todo  $y \in \mathbb{R}$  (no exemplo acima isso não vale pois as variáveis são relacionadas na região onde  $f$  é não-nula) então  $X$  e  $Y$  são independentes.

**Exemplo 97 (O problema da agulha de Buffon)** Suponha que temos um piso feito de tábuas paralelas de madeira, cada uma da mesma largura  $t$  cm, e nós deixamos cair uma agulha de comprimento  $\ell$  cm no chão,  $\ell < t$ . Qual é a probabilidade de que a agulha irá cruzar a divisória entre duas tábuas se encontram?

Seja  $\theta \in [0, \pi/2]$  o ângulo (agudo) com que a agulha cai em relação à vertical que passa pela centro da agulha. Seja  $r \in [0, t/2]$  a distância do centro da agulha até a divisória entre tábuas mais próxima. Assumimos que  $r$  e  $\theta$  são variáveis aleatórias independentes e uniformemente distribuídas.



Se  $r < \frac{\ell}{2} \sin(\theta)$  então a agulha cruzou uma das divisórias e a probabilidade desse evento é

$$\begin{aligned} \mathbb{P}\left(r < \frac{\ell}{2} \sin(\theta)\right) &= \iint_{\{(x,y): x < \frac{\ell}{2} \sin(y)\}} f_r(x) f_\theta(y) \, dx \, dy = \\ &= \int_0^{\pi/2} \int_0^{\frac{\ell}{2} \sin(y)} \frac{2}{\pi} \frac{2}{t} \, dx \, dy = \frac{4}{\pi t} \int_0^{\pi/2} \frac{\ell}{2} \sin(y) \, dy = \frac{4}{\pi t} \frac{\ell}{2} \int_0^{\pi/2} \sin(y) \, dy = \frac{2\ell}{\pi t}. \end{aligned}$$

Veja aqui uma simulação usada para determinar o valor aproximado de  $\pi$ .

Usando R:  
podemos estimar  $\pi$  com

```
n <- 100000 #numero de pontos
l <- 1      #tamanho da agulha
t <- 2      #largura do piso
m <- 0      #contador de hits
for (i in 1:n) {
  r <- runif(1)
  theta <- runif(1, min=0, max=pi/2)
  if (r < 1/2 * sin(theta)) {
    m <- m +1
  }
}
estimativa <- 2*l*n/(t*m)

estimativa
```

e vizualizar uma simulação com

```
#Time-stamp: <Tue Oct 11 14:07:08 2011 jair@random>

L=0.35 #meia agulha
D=20   #tamanho do quadro
N=1000 #numero de lancamentos

#N pontos (R,T)
R=runif(N,min=L,max=D-L) #posicao vertical centro
C=cbind(runif(N,0,D),R)   #coordenadas (x,y) do centro no plano
T=runif(N,min=0,max=pi)   #angulo anti-horario

#as N coordenadas (x,y) dos extremos das agulhas
A=C+L*cbind(cos(T),sin(T))
B=C-L*cbind(cos(T),sin(T))

#numero de cruzamentos: uma agulha cruza dois pisos
#se a parte inteira das coordenadas y dos extremos
#diferem de 1
numbhits=sum(abs(trunc(A[,2])-trunc(B[,2]))>0)

#desenhos:

plot(C,type="n",axes=F,xlim=c(0,D),ylim=c(0,D))

#linhas do piso
for (i in 1:(D-1))
  abline(h=i,lty=2,col="sienna")

#desenho das agulhas
for (t in 1:N)
  lines(c(A[t,1],B[t,1]),c(A[t,2],B[t,2]),col="steelblue")

#titulo do grafico
title(main=paste(numbhits,"cruzamentos",sep=" ",",", estimativa " ,(4*L*N/numbhits)))
```

**Observação 9** *Esses conceitos desenvolvidos até aqui podem ser estendidos, de modo natural, ao caso de 3 ou mais v.a..*



**Exemplo 98** Sejam  $X, Y, Z$  v.a. independentes e uniformemente distribuídas sobre  $[0, 1]$ . Então

$$\begin{aligned}\mathbb{P}(X \geq YZ) &= \iiint_{\{(x,y,z): x \geq yz\}} f_{X,Y,Z}(x,y,z) \, dx \, dy \, dz \\ &= \int_0^1 \int_0^1 \int_{yz}^1 dx \, dy \, dz = \int_0^1 \int_0^1 (1-yz) \, dy \, dz = \int_0^1 \left(1 - \frac{z}{2}\right) dz = \frac{3}{4}.\end{aligned}$$

**Exercício 76** Mostre que se  $X$  e  $Y$  são v.a. independentes então

$$F_{X+Y}(a) = \int_{-\infty}^{+\infty} F_X(a-y) f_Y(y) \, dy \quad e \quad f_{X+Y}(a) = \int_{-\infty}^{+\infty} f_X(a-y) f_Y(y) \, dy.$$

Se  $X$  e  $Y$  têm distribuição conjunta  $f(x, y)$  e  $h$  é uma função a duas variáveis reais, então o valor médio de  $h(X, Y)$  é

$$\mathbb{E}(h(X, Y)) = \begin{cases} \sum_x \sum_y h(x, y) f(x, y) \\ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h(x, y) f(x, y) \, dx \, dy \end{cases}$$

nos casos discreto e contínuo, respectivamente.

**Exercício 77** O que é  $\mathbb{E}(h(X, Y))$  no caso  $h(X, Y) = X$ ?

Notemos que se  $X$  e  $Y$  são independentes então  $\mathbb{E}(g(X)h(Y)) = \mathbb{E}(g(X))\mathbb{E}(h(Y))$ , pra quaisquer funções  $g, h$ . De fato

$$\begin{aligned}\mathbb{E}(g(X)h(Y)) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x)h(y)f(x,y) \, dx \, dy \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x)h(y)f_X(x)f_Y(y) \, dx \, dy \\ &= \int_{-\infty}^{+\infty} g(x)f_X(x) \, dx \int_{-\infty}^{+\infty} h(y)f_Y(y) \, dy \\ &= \mathbb{E}(g(X))\mathbb{E}(h(Y)).\end{aligned}$$

Dedução análoga vale para distribuição conjunta discreta. Em particular, se  $X$  e  $Y$  são independentes então

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y). \quad (42)$$

**Exercício 78** Ache exemplos de v.a. que satisfazem (42) mas que não são independentes.

## Covariância

A covariância entre as variáveis  $X, Y$  é definida por

$$\begin{aligned}\text{Cov}(X, Y) &\stackrel{\text{def}}{=} \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) \\ &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).\end{aligned}$$

a segunda igualdade segue por linearidade da esperança.

Quando  $X$  e  $Y$  são independentes, por (42), temos  $\text{Cov}(X, Y) = 0$ . A recíproca não é verdadeira, se  $\text{Cov}(X, Y) = 0$  não necessariamente as v.a. são independentes, pode significar que as v.a. são de alguma forma relacionadas.

**Exemplo 99** Se  $X$  assume cada um dos valores  $-1, 0$  e  $1$  com probabilidade  $1/3$  e  $Y$  é a variável indicadora do evento  $X = 0$ , então  $\text{Cov}(X, Y) = 0$  (verifique), entretanto não são variáveis independentes (verifique).

**Exemplo 100** Sejam  $I_A$  e  $I_B$  v.a. indicadoras de ocorrência dos eventos  $A$  e  $B$ , respectivamente

$$\text{Cov}(I_A, I_B) = \mathbb{P}(I_A \cap I_B) - \mathbb{P}(I_A)\mathbb{P}(I_B) = \mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(B)(\mathbb{P}(A | B) - \mathbb{P}(A))$$

portanto,  $\text{Cov}(I_A, I_B) > 0$  se e só se  $\mathbb{P}(A | B) > \mathbb{P}(A)$ , ou seja, o evento  $B$  faz ser mais provável  $A$  (e vice-versa).

A covariância satisfaz as seguintes propriedades, cuja verificação fica como exercício.

1.  $\text{Cov}(X, X) = \text{Var}(X)$ ;
2.  $\text{Cov}(X, Y) = \text{Cov}(Y, X)$ ;
3.  $\text{Cov}(cX, Y) = c\text{Cov}(X, Y)$  para qualquer constante  $c$ ;
4.  $\text{Cov}\left(\sum_{i=1}^n X_i, \sum_{j=1}^m Y_j\right) = \sum_{i=1}^n \sum_{j=1}^m \text{Cov}(X_i, Y_j)$ .

**Exercício 79** Prove as propriedades.

**Exercício 80** Prove que  $\text{Cov}(X, Y) = \pm \text{Var}(X)\text{Var}(Y)$  se e só se  $Y = aX + b$  para  $a \neq 0$  e  $b$  reais.

Das propriedades acima, conseguimos deduzir que

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \text{Cov}\left(\sum_{i=1}^n X_i, \sum_{j=1}^n X_j\right) = \sum_{i=1}^n \text{Var}(X_i) + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \text{Cov}(X_i, X_j).$$

Por conseguinte, se a soma é de v.a. independentes então

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i).$$

**Exercício 81** Use a identidade acima para obter a variância de  $X \sim \text{Binomial}(n, p)$ .

## §10 Distribuição e esperança condicionais

Se  $X$  e  $Y$  são v.a. discretas e  $x \in \mathbb{R}$  é tal que  $f_X(x) > 0$  então a **distribuição condicional** de  $Y$  dado  $X = x$  é a função em  $y \in \mathbb{R}$  dada por

$$F(y | x) \stackrel{\text{def}}{=} \mathbb{P}(Y \leq y | X = x) \quad (43)$$

a **função (de massa/de densidade) de probabilidade** condicional de  $Y$  dado que  $X = x$  é

$$f(y | x) \stackrel{\text{def}}{=} \frac{f(x, y)}{f_X(x)}.$$

Notemos que no caso discreto

$$\mathbb{P}(Y = y | X = x) = \frac{\mathbb{P}(Y = y, X = x)}{\mathbb{P}(X = x)} = \frac{f(x, y)}{f_X(x)}$$

para qualquer valor real  $y$ . A última fração é uma função de  $y$ , com  $x$  fixo e satisfaz as condições de função de massa de probabilidade.

Notemos também que, caso as v.a. sejam independentes vale  $f(y | x) = f_Y(y)$  e vice-versa, i.e., a igualdade implica independência.

Considerações análogas levam à definição de  $f(x | y)$ . Naturalmente,

$$\mathbb{E}(Y | X = x) = \begin{cases} \sum_y y f(y | x) & \text{no caso discreto} \\ \int_{-\infty}^{+\infty} y f(y | x) dy & \text{no caso contínuo.} \end{cases} \quad (44)$$

Notemos que quando as v.a. são independentes  $\mathbb{E}(Y | X = x) = \mathbb{E}(Y)$  (verifique).

Se  $\mu(x) \stackrel{\text{def}}{=} \mathbb{E}(Y | X = x)$  então  $\mu(X)$  é a **esperança condicional** de  $Y$  dado  $X$ , denotada

$$\mathbb{E}(Y | X)$$

que embora sugira uma média é, de fato, uma variável aleatória, a qual tem a seguinte propriedade

**Exercício 82**  $\mathbb{E}(\mathbb{E}(Y | X)) = \mathbb{E}(Y)$ .

Logo

$$\mathbb{E}(Y) = \begin{cases} \sum_x \mathbb{E}(Y | X = x) \mathbb{P}(X = x) & \text{no caso discreto} \\ \int_{-\infty}^{+\infty} \mathbb{E}(Y | X = x) f_X(x) dx & \text{no caso contínuo.} \end{cases} \quad (45)$$

**Exemplo 101** De volta ao exemplo 94, a distribuição de  $Y$  condicionada a  $X = 8$  é

$$f(y | 0, 8) = \frac{f(0, 8, y)}{f_X(0, 8)} = \frac{1,2(0,8 + y^2)}{1,2 \cdot 0,8 + 0,4}, \quad \forall y \in (0, 1)$$

A probabilidade do caixa convencional estar ocupado metade do tempo, dado que  $X = 8$ , é

$$\mathbb{P}(Y \leq 0,5 \mid X = 8) = \int_{-\infty}^{0,5} f(y \mid 0,8) dy = 0,39.$$

Usando a distribuição marginal temos que  $\mathbb{P}(Y \leq 0,5) = 0,35$ . Ademais  $\mathbb{E}(Y) = 6$  e

$$\mathbb{E}(Y \mid X = 8) = \int_{-\infty}^{+\infty} y f(y \mid 0,8) dy = 0,574.$$

**Exemplo 102** Sejam  $X, Y \sim \text{Binomial}(n, p)$  independentes. Então

$$\begin{aligned} \mathbb{P}(X + Y = m) &= \sum_{i=0}^n \mathbb{P}(X = i, Y = m - i) = \sum_{i=0}^n \mathbb{P}(X = i) \mathbb{P}(Y = m - i) \\ &= \sum_{i=0}^n \binom{n}{i} \binom{n}{m-i} p^m (1-p)^{n-m+i} = \binom{2n}{m} p^m (1-p)^{2n-m} \end{aligned}$$

pois  $\binom{a+b}{k} = \sum_{i=0}^n \binom{a}{i} \binom{b}{k-i}$ . Desse fato temos

$$\begin{aligned} \mathbb{P}(X = k \mid X + Y = m) &= \frac{\mathbb{P}(X = k, X + Y = m)}{\mathbb{P}(X + Y = m)} \\ &= \frac{\mathbb{P}(X = k, Y = m - k)}{\mathbb{P}(X + Y = m)} \\ &= \frac{\mathbb{P}(X = k) \mathbb{P}(Y = m - k)}{\mathbb{P}(X + Y = m)} \\ &= \frac{\binom{n}{k} p^k (1-p)^{n-k} \binom{n}{m-k} p^{m-k} (1-p)^{n-m+k}}{\binom{2n}{m} p^m (1-p)^{2n-m}} \\ &= \frac{\binom{n}{k} \binom{n}{m-k}}{\binom{2n}{m}} \end{aligned}$$

ou seja,  $X$  condicionada a  $X + Y = m$  tem distribuição hipergeométrica com parâmetros  $2n, n, m$ , logo

$$\mathbb{E}(X \mid X + Y = m) = \frac{m}{2}.$$

**Exemplo 103 (Esperança de v.a. geométrica)** Seja  $X$  o número de lançamentos de uma moeda até sair cara, o que ocorre com probabilidade  $p$ , seja  $Y$  a v.a. indicadora de “cara no primeiro lançamento”. Assim,

$$\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X \mid Y)) = \mathbb{E}(X \mid Y = 0)(1 - p) + \mathbb{E}(X \mid Y = 1)p$$

mas  $\mathbb{E}(X \mid Y = 1) = 1$  e  $\mathbb{E}(X \mid Y = 0) = \mathbb{E}(X + 1)$ , portanto

$$\mathbb{E}(X) = \mathbb{E}(X)(1 - p) + (1 - p) + p$$

que resolvendo para  $\mathbb{E}(X)$  resulta em  $\mathbb{E}(X) = 1/p$ , como já sabíamos.

Do mesmo modo, temos

$$\mathbb{E}(X^2) = \mathbb{E}(X^2 \mid Y = 0)(1 - p) + \mathbb{E}(X^2 \mid Y = 1)p = \mathbb{E}(X + 1)^2(1 - p) + p$$

que resolvendo para  $\mathbb{E}(X^2)$  resulta em

$$\mathbb{E}(X^2) = \frac{2 - p^2}{p}$$

e com isso

$$\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{2 - p^2}{p} - \frac{1}{p^2} = \frac{1 - p^2}{p}.$$

**Exemplo 104 (Soma de v.a. com número aleatório de termos)** Num determinado dia do cotidiano de uma loja  $N$  pessoas entram na loja, com  $\mathbb{E}(N) = 50$ , e o gasto dessa pessoas são dados pelas v.a. independentes  $X_1, X_2, \dots, X_N$  com  $\mathbb{E}(X_i) = \mu = 80$  reais, para todo  $i$ , e independentes de  $N$ .

A quantia gasta num determinado dia é  $X = \sum_{i=1}^N X_i$  cuja média é

$$\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X \mid N)) = \mathbb{E}\left(\mathbb{E}\left(\sum_{i=1}^N X_i \mid N\right)\right)$$

Agora,

$$\mathbb{E}\left(\sum_{i=1}^n X_i \mid N = n\right) = \mathbb{E}\left(\sum_{i=1}^n X_i\right) = n\mu$$

por causa da independência, por fim

$$\mathbb{E}\left(\mathbb{E}\left(\sum_{i=1}^N X_i \mid N\right)\right) = \mathbb{E}(N\mu) = \mathbb{E}(N)\mu = 4.000$$

reais.

**Exercício 83** Considere o espaço amostral das sequências definidas pelas permutações de  $\{1, 2, 3\}$

$$\Omega = \{(1, 2, 3), (1, 3, 2), (2, 1, 3), (2, 3, 1), (3, 1, 2), (3, 2, 1), (1, 1, 1), (2, 2, 2), (3, 3, 3)\}.$$

Seja  $X_i(\omega)$  a  $i$ -ésima coordenada de  $\omega$ , para  $i = 1, 2, 3$ . Defina  $N$  a variável aleatória igual a  $X_2$ . Prove que

- para todo  $i, j \in \{1, 2, 3\}$ , vale  $\mathbb{P}(X_i = j) = 1/3$ ;
- as variáveis  $X_i$  não são independentes;
- $\sum_{i=1}^N \mathbb{E}X_i = 4$  e conclua que

$$\mathbb{E}\left(\sum_{i=1}^N X_i\right) \neq \sum_{i=1}^N \mathbb{E}X_i.$$

**Exemplo 105** Sejam  $U_1, U_2, \dots$  v.a. com distribuição Uniforme  $((0, 1))$  independentes e  $m(x)$  o valor esperado para  $N(x)$  que é o menor  $n$  para o qual  $U_1 + \dots + U_n > x$ . Certamente,

$$\mathbb{E}(N(x) \mid U_1 = y) = \begin{cases} 1, & \text{se } y > x \\ 1 + m(x - y), & \text{se } y \leq x. \end{cases}$$

Portanto,

$$m(x) = \int_0^1 \mathbb{E}(N(x) \mid U_1 = y) dy = 1 + \int_0^x m(x - y) dy = 1 + \int_0^x m(u) du.$$

Derivando os extremos dessa cadeia de igualdades obtemos  $m'(x) = m(x)$ , ou seja,  $m(x) = ae^x$  para alguma constante real  $a$  e como  $m(0) = 1$  temos

$$m(x) = e^x$$

é o número esperado de termos para que a soma de v.a. uniformes em  $(0, 1)$  ultrapasse o valor  $x$ .

**Exercício 84** Mostre que para constantes  $a, b \in \mathbb{R}$

$$\mathbb{E}(aX + bY \mid Z) = a\mathbb{E}(X \mid Z) + b\mathbb{E}(Y \mid Z).$$

**Exercício 85** Mostre que  $\mu(X) = \mathbb{E}(Y \mid X)$  satisfaz

$$\mathbb{E}(\mu(X)g(X)) = \mathbb{E}(Yg(X))$$

para qualquer função  $g$  pra qual as esperanças acima existam.

**Exercício 86** Uma galinha bota  $N$  ovos, em que  $N \sim \text{Poisson}(\lambda)$ . Cada ovo vinga com probabilidade  $p$  e independente dos outros ovos. Calcule  $\mathbb{E}(N \mid K)$ ,  $\mathbb{E}(K)$  e  $\mathbb{E}(K \mid N)$ , em que  $K$  é o número de pintinhos.

## §11 Cadeias de Markov

Uma coleção  $\{X_i : i \in T\}$  de variáveis aleatórias que assumem valor num conjunto de estados  $S$  é um *processo estocástico*. O índice  $i$  é interpretado como tempo e o valor de  $X_i$  é o estado do processo no instante  $i$ . O processo é dito de *tempo-discreto* ou *tempo-contínuo* dependendo de  $T$  ser enumerável ou não, respectivamente. Cadeias de Markov são um caso particular de processo estocástico, no nosso caso de tempo discreto e variáveis discretas.

No que segue fixamos  $T = \mathbb{N}$  e a menos que seja dito o contrário  $S \subset \mathbb{N}$ . Nessas condições um processo estocástico é uma cadeia de Markov se vale a condição de Markov de que o estado futuro depende somente do estado atual e é independente dos estados passados

$$\mathbb{P}(X_{n+1} = j \mid X_n = i, X_{n-1} = s_{n-1}, \dots, X_0 = s_0) = \mathbb{P}(X_{n+1} = j \mid X_n = i) \quad (46)$$

e ela é dita *homogênea* se essa probabilidade não depende do tempo  $n$ , que é o caso que estudaremos a seguir. Como todas as cadeias desse capítulo são homogêneas nós omitiremos esse adjetivo daqui em diante.

Uma sequência  $(\rho_i)_{i \in S}$  com  $\rho_i \geq 0$ , de modo que  $\sum_{i \in S} \rho_i = 1$  é chamada *vetor de probabilidades* ou *distribuição de  $X$*  no caso em que  $\rho_i = \mathbb{P}(X = i)$ . Chamamos  $X_0$  de *estado inicial* e sua função de massa de *distribuição inicial*.

Uma cadeia de Markov é caracterizada por

1. conjuntos de estados  $S$ ,
2. distribuição inicial  $(\rho_i)_{i \in S}$ ,
3. probabilidades de transição do estado  $i$  para o estado  $j$ , para todo  $(i, j) \in S^2$ ,

$$p(i, j) \geq 0 \quad \text{com} \quad \sum_{j \in S} p(i, j) = 1$$

de modo que um o process estocástico  $\{X_t\}_{t \in \mathbb{N}}$  com valores em  $S$  é uma **cadeia de Markov homogênea** com distribuição inicial  $\rho$  e probabilidade transição  $p$  se a distribuição conjunta de  $X_0, \dots, X_n$ , para todo  $n \in \mathbb{N}$ , é

$$\mathbb{P}(X_0 = s_0, X_1 = s_1, \dots, X_n = s_n) = \rho_{s_0} \prod_{i=0}^{n-1} p(s_i, s_{i+1}) \quad (47)$$

para toda escolha de  $s_0, s_1, \dots, s_n \in S$ . Dessa distribuição conjunta temos

$$\mathbb{P}\left(X_{n+1} = s_{n+1} \mid \bigcap_{\ell=0}^n [X_\ell = s_\ell]\right) = \frac{\mathbb{P}\left(\bigcap_{\ell=0}^{n+1} [X_\ell = s_\ell]\right)}{\mathbb{P}\left(\bigcap_{\ell=0}^n [X_\ell = s_\ell]\right)} = \frac{\rho_{s_0} \prod_{i=0}^{n+1} p(s_i, s_{i+1})}{\rho_{s_0} \prod_{i=0}^n p(s_i, s_{i+1})} = p(s_n, s_{n+1})$$

e

$$\mathbb{P}\left(X_{n+1} = s_{n+1} \mid X_n = s_n\right) = \frac{\mathbb{P}(X_{n+1} = s_{n+1}, X_n = s_n)}{\mathbb{P}(X_n = s_n)} = \frac{\sum_{s_0, \dots, s_{n-1}} \rho_{s_0} \prod_{i=0}^n p(s_i, s_{i+1})}{\sum_{s_0, \dots, s_{n-1}} \rho_{s_0} \prod_{i=0}^{n-1} p(s_i, s_{i+1})} = p(s_n, s_{n+1})$$

que é a condição de Markov (46). Ademais, fixado  $n$ , temos a partir dos itens 2 e 3 acima que

$$\begin{aligned} \sum_{s_0 \in S} \sum_{s_1 \in S} \cdots \sum_{s_{n-1} \in S} \sum_{s_n \in S} \rho_{s_0} \prod_{i=0}^{n-1} p(s_i, s_{i+1}) &= \sum_{s_0 \in S} \sum_{s_1 \in S} \cdots \sum_{s_{n-1} \in S} \rho_{s_0} \prod_{i=0}^{n-2} p(s_i, s_{i+1}) \sum_{s_n \in S} p(s_{n-1}, s_n) = \\ &= \sum_{s_0 \in S} \sum_{s_1 \in S} \cdots \sum_{s_{n-1} \in S} \rho_{s_0} \prod_{i=0}^{n-2} p(s_i, s_{i+1}) = \cdots = \sum_{s_0 \in S} \sum_{s_1 \in S} \rho_{s_0} p(s_0, s_1) = \sum_{s_0 \in S} \rho_{s_0} = 1 \end{aligned}$$

portanto 47 é uma atribuição de probabilidades válida no espaço das realizações das  $n$  variáveis. A mesma dedução vale para qualquer valor de  $n$  portanto a definição da distribuição é consistente (o Teorema de Kolmogorov garante a existência das variáveis).

**Exercício 87** Prove que o evento

$$X_{i_1} = s_{i_1}, X_{i_2} = s_{i_2}, \dots, X_{i_r} = s_{i_r}$$

ocorre com mesma probabilidade em todo espaço das realizações das variáveis com pelo menos  $r$  variáveis.

**Exercício 88 (Propriedade de Markov)** Prove que se  $\{X_t\}_{t \in \mathbb{N}}$  é uma cadeia de Markov com respeito a  $(\rho_i)_{i \in S}$  e  $\{p(i, j)\}_{i, j \in S}$  então condicionado a  $X_m = i$  o processo  $\{X_{t+m}\}_{t \in \mathbb{N}}$  é uma cadeia de Markov com respeito a  $(\delta_i(j))_{j \in S}$  e  $\{p(i, j)\}_{i, j \in S}$  e independente de  $X_0, \dots, X_m$ . Acima,  $\delta_i(j) = 1$  se e só se  $i = j$ .

Convencionamos que quando é dada uma definição de uma cadeia de Markov, as probabilidades de transição que não são dadas explicitamente são iguais a zero. Ademais, usualmente não damos a distribuição inicial  $\rho_s$  porque ou esse parâmetro não é relevante no interesse do momento ou é trivial (i.e.,  $\mathbb{P}(X_0 = s) = 1$  para algum  $s$ ).

A seguir daremos alguns exemplos de cadeias de Markov.

**Exemplo 106** Numa pilha com  $n$  cartas de baralho distintas numeradas de 1 a  $n$ , definimos uma cadeia de Markov tomando um estado para cada permutação  $\pi$ , em que  $\pi(i)$  é a posição da  $i$ -ésima carta;  $X_0$  é a permutação identidade com probabilidade 1.

Uma transição entre estados é obtida retirando a carta topo e colocando-a numa posição arbitrária escolhida uniformemente entre as cartas restantes na pilha, logo

$$p(\pi, \sigma) = \mathbb{P}(X_{t+1} = \sigma \mid X_t = \pi) = \frac{1}{n}$$

para todo  $t \in \mathbb{N}$  e com a permutação  $\sigma$  obtida a partir da permutação  $\pi$  pelo processo descrito acima (nos casos em que  $\sigma$  não pode ser obtida pelo processo descrito vale  $p(\pi, \sigma) = 0$ , como convencionamos). Um problema interessante é estimar o valor  $n$ , caso exista, para o qual a distribuição de  $X_n$  seja aproximadamente uniforme.

**Exemplo 107** Sejam  $Y_i \sim \text{Bernoulli}(p)$  variáveis aleatórias independentes,  $X_0 = 0$  e para  $i > 0$ , tomemos  $X_i = Y_i + Y_{i-1}$ . Então

$$\mathbb{P}(X_{i+1} = 2 \mid X_i = 1, X_{i-1} = 2) = 0,$$

enquanto que

$$\mathbb{P}(X_{i+1} = 2 \mid X_i = 1, X_{i-1} = 0) = p,$$

logo  $\{X_i\}$  não é uma cadeia de Markov por não satisfazer (46).

**Exemplo 108** Definimos uma cadeia de Markov se tomamos o conjunto de estados como sendo os inteiros positivos,  $S = \mathbb{N} \setminus \{0\}$ , o estado inicial  $X_0 = 1$  com probabilidade 1 e as transições probabilidades

$$p(i, 1) = \frac{1}{i+1}$$

$$p(i, i+1) = \frac{i}{i+1}$$

para todo inteiro  $i \geq 1$ .



**Exemplo 109 (passeio aleatório)** Tomemos o conjunto de estados como os inteiros,  $S = \mathbb{Z}$ , o estado inicial  $X_0 = 0$  com probabilidade 1 e as transições de estados têm as probabilidades

$$\begin{aligned} p(i, i-1) &= 1-p \\ p(i, i+1) &= p \end{aligned}$$

para todo inteiro  $i$  e algum  $p \in (0, 1)$ . Essa cadeia é um passeio aleatório pelos inteiros.

**Exemplo 110** Seja  $G$  um grafo finito. Definimos uma cadeia de Markov tomando  $S$  como o conjunto de vértices do grafo e as transições são definidas de modo que se  $X_t = v$  então  $X_{t+1}$  é qualquer vértice adjacente a  $v$  com probabilidade uniforme. Esse tipo de cadeia de Markov é conhecida como passeio aleatório num grafo.

**Exemplo 111 (Cadeia de Markov não-homogênea)** Suponha  $Y_1, Y_3, Y_5, \dots, Y_{2i+1}, \dots$  v.a. independentes e identicamente distribuídas com

$$\mathbb{P}(Y_{2k+1} = 1) = \mathbb{P}(Y_{2k+1} = -1) = \frac{1}{2}$$

para todo  $k \geq 0$ . Agora, tomemos  $Y_{2k} = Y_{2k+1}Y_{2k-1}$  para todo  $k > 0$ .

A sequência de v.a.  $\{Y_i\}_{i \geq 1}$  não define uma cadeia de Markov pois

$$\mathbb{P}(Y_{2k+1} = 1 \mid Y_{2k} = -1) = \frac{1}{2}$$

enquanto que

$$\mathbb{P}(Y_{2k+1} = 1 \mid Y_{2k} = -1, Y_{2k-1} = -1) = 0.$$

No entanto,  $Z_n = (Y_n, Y_{n+1})$  é uma cadeia de Markov não homogênea pois, por exemplo

$$\mathbb{P}(Z_{n+1} = (1, 1) \mid Z_n = (1, 1)) = \begin{cases} 1/2 & \text{se } n \text{ par} \\ 1 & \text{se } n \text{ ímpar.} \end{cases}$$

**Exemplo 112 (Cadeia de Markov não-homogênea)** Consideremos uma caixa com  $v$  bolas vermelhas e  $a$  bolas azuis e o processo de retirar aleatoriamente uma bola da caixa sem reposição. Seja  $X_i$  o número de bolas vermelhas na caixa na  $i$ -ésima rodada. A probabilidade de transição não depende só do número de bolas vermelhas (estado), mas também depende do momento  $i$ .

**Exemplo 113 (ruína do jogador)** Em um jogo, o jogador ganha 1 real com probabilidade  $p$ , ou perde 1 real com probabilidade  $1-p$ , de modo independente a cada rodada de até que sua fortuna seja 0 ou  $n$ , nesse momento ele pára de jogar.

Denotamos a quantia do jogador no instante  $t$  pela variável  $X_t$  e temos que

$$\begin{aligned} p(0, 0) &= p(n, n) = 1 \\ p(i, i+1) &= p \text{ e } p(i, i-1) = 1-p, \end{aligned}$$

para  $1 \leq i \leq n$ , são as transições de estado de uma cadeia de Markov.

Um problema interessante para o jogador é conhecer a probabilidade de chegar ao estado  $n$  antes do estado 0, esse é um problema clássico da probabilidade conhecido como o problema da ruína do jogador.

Seja  $P_i$  a probabilidade de o jogador chegar antes ao estado  $n$  dado que  $X_0 = i$ . Então

$$P_i = pP_{i+1} + (1-p)P_{i-1}$$

para todo  $i \in \{1, \dots, n\}$ , que equivale a

$$pP_i + (1-p)P_i = pP_{i+1} + (1-p)P_{i-1}$$

para todo  $i \in \{1, \dots, n\}$ . Para facilitar usaremos  $q = 1 - p$ . Segue da igualdade acima que

$$\begin{aligned} P_{i+1} - P_i &= \frac{q}{p}(P_i - P_{i-1}) \\ &= \left(\frac{q}{p}\right)^2 (P_{i-1} - P_{i-2}) \\ &= \dots \\ &= \left(\frac{q}{p}\right)^i (P_1 - P_0). \end{aligned}$$

donde concluímos, usando que  $P_0 = 0$ , que

$$P_{i+1} = P_i + \left(\frac{q}{p}\right)^i P_1 \quad (48)$$

para todo  $i \in \{1, \dots, n\}$ . Iterando

$$\begin{aligned} P_{i+1} &= P_i + \left(\frac{q}{p}\right)^i P_1 \\ &= P_{i-1} + \left(\left(\frac{q}{p}\right)^{i-1} + \left(\frac{q}{p}\right)^i\right) P_1 \\ &= P_{i-2} + \left(\left(\frac{q}{p}\right)^{i-2} + \left(\frac{q}{p}\right)^{i-1} + \left(\frac{q}{p}\right)^i\right) P_1 \\ &= \dots \\ &= \left(1 + \frac{q}{p} + \dots + \left(\frac{q}{p}\right)^{i-2} + \left(\frac{q}{p}\right)^{i-1} + \left(\frac{q}{p}\right)^i\right) P_1 \end{aligned}$$

e de  $P_n = 1$  temos

$$1 = \left(1 + \frac{q}{p} + \dots + \left(\frac{q}{p}\right)^{n-2} + \left(\frac{q}{p}\right)^{n-1}\right) P_1.$$

Se  $p \neq q$  então a soma acima é de uma progressão geométrica cuja valor é  $1 - (q/p)/1 - (q/p)^n$  logo

$$P_1 = \begin{cases} \frac{1}{n} & \text{se } p = q \\ \frac{1 - (q/p)}{1 - (q/p)^n} & \text{se } p \neq q. \end{cases}$$

e substituindo em (48)

$$P_i = \begin{cases} \frac{i}{n} & \text{se } p = q \\ \frac{1 - (q/p)^i}{1 - (q/p)^n} & \text{se } p \neq q. \end{cases}$$

Notemos que a probabilidade do jogador ficar “infinitamente rico”, i.e.,  $P_i(n)$  quando  $n \rightarrow \infty$ , é 0 se  $p \leq 1/2$  pois  $\frac{i}{n}$ ,  $\frac{1 - (q/p)^i}{1 - (q/p)^n}$  tendem a 0. Por outro lado, se  $p > 1/2$  então  $(q/p)^n \rightarrow 0$  quando  $n \rightarrow \infty$  logo  $P_i \rightarrow 1 - (q/p)^i > 0$ .

Usando R:

```
simulaRuina=function(n=10,p=1/2,estado=5)
{
  cat(estados)
  m<-0
  while ( estado < n & estado > 0) {
    proxEstado <- estado + sample(x=c(-1,1), size=1, prob=c(1-p,p))
    cat(" -> ",proxEstado)
    m <- m+1
    estado <- proxEstado
  }
  cat("\n",m," passos","\n")
}

simulaRuina()

simulaRuina(10,1/3,3)
```

simulaRuina() simula transições aleatórias até atingir o estado 10 ou 0 a partir do 5;  
 simulaRuina(10,1/3,3) simula até atingir o estado 10 ou 0 a partir do 3 com probabilidade de ganho 1/3;  
 também devolve o número de passos efetuados.

**Exemplo 114** Consideremos a cadeia de Markov sobre  $S = \{0, 1, \dots, n\}$  com estado inicial  $X_0 = 0$  e as transições de estados têm as probabilidades

$$p(0, 1) = 1 = p(n, n)$$

$$p(j, j + 1) = 1/2$$

$$p(j, j - 1) = 1/2$$

para todo inteiro  $j \geq 1$ . Vamos estimar o número esperado de passos até a cadeia chegar ao estado  $n$ . Seja  $Y_j$  o número de passos para atingir  $n$  a partir de  $j$ . Se o estado atual é  $j > 0$  e o próximo  $j - 1$  então  $Y_j = 1 + Y_{j-1}$ , senão  $Y_j = 1 + Y_{j+1}$ , portanto, por (45)

$$\mathbb{E}(Y_j) = \mathbb{E}(Y_{j-1} + 1) \frac{1}{2} + \mathbb{E}(Y_{j+1} + 1) \frac{1}{2} = \frac{1}{2} (\mathbb{E}Y_{j-1} + \mathbb{E}Y_{j+1}) + 1$$

e temos

$$\begin{cases} 2\mathbb{E}(Y_j) = \mathbb{E}(Y_{j-1}) + \mathbb{E}(Y_{j+1}) + 2, & \text{se } 0 < j < n \\ \mathbb{E}(Y_0) = \mathbb{E}(Y_1) + 1, \\ \mathbb{E}(Y_n) = 0. \end{cases}$$

cuja solução é  $\mathbb{E}(Y_j) = n^2 - j^2$  (verifique). Finalmente,  $\mathbb{E}(Y_0) = n^2$ . Pela desigualdade de Markov, equação (24), a probabilidade da cadeia não chegar no estado  $n$  em  $2n^2$  passos é

$$\mathbb{P}(Y_0 \geq 2n^2) \leq \frac{\mathbb{E}(Y_0)}{2n^2} = \frac{1}{2} \quad (49)$$

e, analogamente, tende a zero a probabilidade da cadeia não chegar no estado  $n$  em  $n^2\omega(n)$  passos qualquer que seja  $\omega(n)$  que tenda ao infinito com  $n$ , por exemplo,  $\omega(n) = \log \log n$ .

Usando R:

```
simulaMCN = function (n , estado =0)
{
  cat ( estado )
  m <-0

  while ( estado < n ) {

    if ( estado > 0) {
      proxEstado <- estado + sample (x=c ( -1 ,1) , size =1 , prob =c (1 / 2 ,1 / 2)
    )}
    else { proxEstado <- 1 }
    cat ( " -> " , proxEstado )

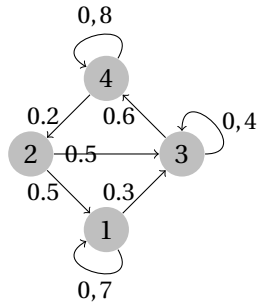
    m <- m +1
    estado <- proxEstado
  }
  cat ( "\ n" ,m , " passos " , " \n")
}
```

simulaMCN(10) simula até atingir o estado 10 a partir do 0;  
 simulaMCN(10,3) simula até atingir o estado 10 a partir do 3;  
 também devolve o número de passos efetuados.

**Exemplo 115** Suponha que a ocorrência de chuva num dia é determinada pela condição do clima nos dois dias anteriores do seguinte modo: amanhã chove com probabilidade 0,7 se chove nos dois dias anteriores; amanhã chove com probabilidade 0,5 se chove hoje mas não choveu ontem; amanhã chove com probabilidade 0,4 se não chove hoje mas choveu ontem; amanhã chove com probabilidade 0,2 se não chove nos dois dias anteriores.

Identificamos cada uma das quatro situações acima com seguintes estados **1** se choveu hoje e ontem; **2** se choveu hoje mas não choveu ontem; **3** se não choveu hoje mas choveu ontem; e **4** se não choveu nem hoje e nem ontem.

O estado no dia  $X_{n+1}$  depende da condição nos dias anteriores; assim a transição de  $X_n = 1$  para  $X_{n+1} = 3$ , por exemplo, ocorre quando não chove amanhã ( $n+1$ ) mas choveu hoje, dado que choveu hoje ( $n$ ) e ontem ( $n-1$ ), nessa configuração a probabilidade de não-chuva amanhã se choveu hoje e ontem é  $1 - 0,7 = 0,3$ . O seguinte diagrama ilustra as transições de estados com suas probabilidades.



**Exercício 89** Um dado é lançado repetidamente. Quais das seguintes sequências de variáveis formam uma cadeia de Markov?

1.  $X_n$  é o maior resultado até a  $n$ -ésima rodada;
2.  $Y_n$  é a quantidade de 6 em  $n$  rodadas;
3. no instante  $r$ ,  $Z_r$  é o tempo desde o 6 mais recente;
4. no instante  $r$ ,  $W_r$  é o tempo até o próximo 6.

**Exercício 90** Se  $\{X_n\}_n$  é uma cadeia de Markov, quais das seguintes sequências é uma cadeia de Markov?

1.  $\{X_{m+r}\}_{r \geq 0}$ , para  $m$  fixo;
2.  $\{X_{2m}\}_{m \geq 0}$ ;
3.  $\{(X_n, X_{n+1})\}_{n \geq 0}$ .

**Exercício 91 (Equação de Wald e ruína do jogador)** Sejam  $Z_1, Z_2, \dots$  v.a. independentes, identicamente distribuídas, e com esperança  $\mathbb{E}(Z)$  finita. Chamamos a v.a.  $N$  de tempo de parada para a sequência  $\{Z_i\}_{i \geq 1}$  se o evento  $[N = n]$  é independente de  $Z_t$  para  $t > n$ , para todo  $n$ .

Prove que se  $N$  tem esperança  $\mathbb{E}(N)$  finita, então

$$\mathbb{E} \left( \sum_{i=1}^N Z_i \right) = \mathbb{E}(N) \mathbb{E}(Z). \quad (50)$$

No exemplo 113, seja  $i$  o estado inicial  $D_i$  o tempo de duração do jogo, ou seja, o tempo esperado até a cadeia atingir o estado 0 ou o estado  $n$ . Assuma que as esperanças sejam finitas e use a equação (50) para determinar  $D_i$ .

## Representação matricial, transiência e recorrência

A matriz  $P = (p_{i,j})$  definida pelas probabilidades de transição  $p_{i,j} = p(i, j)$  é dita **matriz de transições** da cadeia de Markov. Ainda, para  $k \in \mathbb{N}$ , se

$$p_{i,j}^{(k)} \stackrel{\text{def}}{=} \mathbb{P}(X_{t+k} = j \mid X_t = i)$$

é a transição do estado  $i$  para o estado  $j$  em  $k$  passos então

$$p_{i,j}^{(k)} = \sum_{s \in S} p_{i,s} p_{s,j}^{(k-1)}$$

para todo inteiro  $k > 1$  e com  $p_{i,j}^{(1)} = p_{i,j}$  e *convencionamos*  $p_{i,j}^{(0)} = 1$  se  $i = j$ , caso contrário  $p_{i,j}^{(0)} = 0$ . Logo a  $k$ -ésima potência da matriz  $P$  é a matriz de transição em  $k$  passos

$$P^k = (p_{i,j}^{(k)}). \quad (51)$$

Notemos que  $\sum_j p_{i,j}^{(k)} = 1$ , para qualquer  $k \in \mathbb{N}$ . Uma matriz cujas entradas são não-negativas e as linhas somam 1 é uma *matriz estocástica*.

**Exercício 92** Verifique que se  $P$  é estocástica então  $P^n$  também é estocástica.

Notemos que  $\mathbb{P}(A, B \mid C) = \mathbb{P}(A \mid B, C)\mathbb{P}(B \mid C)$  e usando esse fato podemos deduzir

$$\begin{aligned} p_{i,j}^{(k+t)} &= \mathbb{P}(X_{k+t} = j \mid X_0 = i) \\ &= \sum_{s \in S} \mathbb{P}(X_{k+t} = j, X_k = s \mid X_0 = i) \\ &= \sum_{s \in S} \mathbb{P}(X_{k+t} = j \mid X_k = s, X_0 = i) \mathbb{P}(X_k = s \mid X_0 = i) \\ &= \sum_{s \in S} p_{s,j}^{(t)} p_{i,s}^{(k)} \end{aligned}$$

portanto, vale a seguinte identidade que é bastante útil.

**Lema 10 (Identidade de Chapman–Kolmogorov)** Se  $P = (p_{i,j})$  é uma matriz de transição,

$$p_{i,j}^{(k+t)} = \sum_{s \in S} p_{i,s}^{(k)} p_{s,j}^{(t)} \quad (52)$$

para quaisquer  $i, j \in S$ , ou ainda,  $P^{k+t} = P^k P^t$ .

A evolução da cadeia é determinada pela matriz  $P$  e boa parte dos estudos das cadeias de Markov são reduzidas ao estudo das propriedades algébricas desses elementos.

**Exemplo 116** No caso do exemplo 115 dado que choveu na segunda-feira e na terça-feira, qual é a probabilidade de chover na quinta?

A matriz toda está descrita abaixo

$$P = \begin{pmatrix} 0,7 & 0 & 0,3 & 0 \\ 0,5 & 0 & 0,5 & 0 \\ 0 & 0,4 & 0 & 0,6 \\ 0 & 0,2 & 0 & 0,8 \end{pmatrix}$$

A matriz de transição em 2 passos é a matriz

$$P^2 = \begin{pmatrix} 0,49 & 0,12 & 0,21 & 0,18 \\ 0,35 & 0,20 & 0,15 & 0,30 \\ 0,20 & 0,12 & 0,20 & 0,48 \\ 0,10 & 0,16 & 0,10 & 0,64 \end{pmatrix}$$

e chover na quinta equivale ao processo estar no estado 0 ou no estado 1, logo a probabilidade é  $p_{0,0}^{(2)} + p_{0,1}^{(2)} = 0,61$ .

Usando R:

```
#Algoritmo para simulacao de cadeia de Markov nos primeiros N passos,
#seja M[,] a matriz de transicoes
#
#1. Escolha um estado inicial X0 = j de acordo com a distribuicao inicial.
#2. i <- 1
#3. Repita N vezes
#4.     Gerar Y_j com distribuicao M[j,]
#5.     Xi <- Yj,
#6.     j <- Xi
#7.     i <- i+1
```

```
simulaMC=function(N=100)
{
  # matriz transicao
  trans <- matrix(c(
    0.7,0,0.3,0,
    0.5,0,0.5,0,
    0,0.4,0,0.6,
    0,0.2,0,0.8
  ), ncol=4, byrow=TRUE)

  # estado inicial
  estado = sample(x=1:ncol(trans), size=1, prob=c(1/4,1/4,1/4,1/4))
  cat(estado)

  # N passos da cadeia
  for (i in 1:N) {
    proxEstado <- sample(x=1:ncol(trans), size=1, prob=trans[estado,])
    cat(" -> ",proxEstado)
    estado <- proxEstado
  }
  cat("\n")
}
```

simulaMC() simula transições aleatórias até atingir 100 passos; simulaMC(10) simula até atingir 10 passos.

**Exercício 93** Se  $\{X_n\}_n$  é uma cadeia de Markov com matriz de transições

$$P = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 1/2 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

e  $f$  é dada por  $f(0) = 0$  e  $f(1) = f(2) = 1$ , então  $Y_n = f(X_n)$  é uma cadeia de Markov?

**Exercício 94** Mostre que a sequência de v.a.  $\{Y_i\}_{i \geq 1}$  do exemplo 111 é de variáveis independentes 2-a-2, conclua que  $p_{i,j}^{(n)} = 1/2$  e que vale a identidade de Chapman–Kolmogorov (52) mesmo não sendo uma cadeia de Markov.

### Classificação de estados

Os estados de uma cadeia de Markov são classificados em dois tipos fundamentais. Um estado é *recorrente* (ou, persistente) se, uma vez tendo estado nele, é certo que a cadeia eventualmente volta a ele, caso contrário o estado é dito *transiente* (ou, transitório). Quando o estado é recorrente, a cadeia visita-o infinitas vezes pois uma vez que a cadeia esteja em tal estado é como um reinício, ela voltará a ele com probabilidade 1, e no caso transiente o estado é visitado um número finito de vezes. Notemos que isso leva a conclusão de que se  $|S|$  é finito então pelo menos um estado é recorrente.

Para  $n \in \mathbb{N}$ , seja  $f_{i,j}^{(n)}$  a probabilidade do evento “primeira passagem pelo estado  $j$  a partir do estado  $i$  em  $n$  passos”

$$X_0 = i, X_1 \neq j, X_2 \neq j, \dots, X_{n-1} \neq j, X_n = j$$

com  $f_{i,j}^{(0)} = 0$ , assim  $f_{i,i}^{(n)}$  é a probabilidade do primeiro retorno a  $i$  em  $n$  passos e definimos

$$f_{i,i} \stackrel{\text{def}}{=} \sum_{n \geq 0} f_{i,i}^{(n)}$$

que é a probabilidade de eventualmente retornar ao estado  $i$ . Com essa definição, um estado  $i$  é **recorrente** se e somente se  $f_{i,i} = 1$ , caso contrário  $f_{i,i} < 1$  e chamamos  $i$  de **transiente**.

Seja  $I_n$  a v.a. indicadora do evento  $[X_n = i]$ . Então a soma para todo  $n \geq 0$  das variáveis  $I_n$  é o número de vezes que a cadeia visita o estado  $i$ ,

$$V_i \stackrel{\text{def}}{=} \sum_{n \geq 0} I_n$$

que afirmamos acima ser finito se e só se  $n$  é transiente. De fato, pelo exercício 49,

$$\mathbb{E}(V_i | X_0 = i) = \sum_{k \geq 1} \mathbb{P}(V_i \geq k | X_0 = i)$$

e, saindo de  $i$ , visitar  $j$  pelo menos  $k$  vezes equivale a visitar  $j$  e em seguida visitar  $j$  pelo menos  $k - 1$  vezes, o que ocorre com probabilidade  $f_{i,j}(f_{j,j})^{k-1}$  logo, fazendo  $j = i$  temos

$$\mathbb{E}(V_i | X_0 = i) = \sum_{k \geq 1} (f_{i,i})^k = \begin{cases} \frac{f_{i,i}}{1-f_{i,i}} & \text{se } f_{i,i} < 1 \\ \infty & \text{se } f_{i,i} = 1 \end{cases}$$

Essa classificação de estados pode ser caracterizada pelas probabilidades de transições da maneira que deduziremos a seguir.

**Lema 11** O estado  $i$  é

- transiente se e só se  $\sum_{n \geq 0} p_{i,i}^{(n)} < \infty$ .



- *recorrente se e só se*  $\sum_{n \geq 0} p_{i,i}^{(n)} = \infty$ .

DEMONSTRAÇÃO: Pela linearidade da esperança,

$$\mathbb{E}(V_i | X_0 = i) = \sum_{n \geq 0} \mathbb{E}(I_n | X_0 = i) = \sum_{n \geq 0} \mathbb{P}(X_n = i | X_0 = i) = \sum_{n \geq 0} p_{i,i}^{(n)}$$

e as caracterizações seguem das considerações acima a respeito da esperança condicionada.  $\square$

No caso  $f_{i,i} = 1$ , temos uma função de massa probabilidade  $\{f_{i,i}^{(n)}\}_{n \geq 0}$  para o tempo de retorno ao estado  $i$ ,

$$T_i \stackrel{\text{def}}{=} \min\{n \geq 1 : X_n = i\} \quad (53)$$

condicionado a  $X_0 = i$ , cuja média é chamada de *tempo médio de recorrência* (ou, tempo médio de retorno) para  $i$

$$\mu_i \stackrel{\text{def}}{=} \sum_{n \geq 0} n f_{i,i}^{(n)}.$$

No caso de  $i$  transiente convencionamos  $\mu_i = \infty$ .

Dado que o retorno ao estado inicial  $i$  é certo, é natural perguntar se o tempo médio de retorno é finito. No caso  $\mu_i = \infty$  o estado  $i$  é dito **recorrente nulo**, no caso de média finita,  $\mu_i < \infty$ , o estado  $i$  é dito **recorrente positivo**.

**Exemplo 117** No exemplo 108,  $S = \mathbb{N} \setminus \{0\}$ ,  $X_0 = 1$ ,  $p(i, 1) = \frac{1}{i+1}$  e  $p(i, i+1) = \frac{i}{i+1}$  para todo inteiro  $i \geq 1$ . A probabilidade de não retornar ao estado 1 nos  $n-1$  primeiros passos é

$$\prod_{j=1}^{n-1} \frac{j}{j+1} = \frac{1}{n}$$

portanto (exerc. 25) a probabilidade de nunca voltar ao estado 1 é 0 e a probabilidade de voltar ao 1 no  $n$ -ésimo passo é

$$f_{1,1}^{(n)} = \frac{1}{n} \frac{1}{n+1}$$

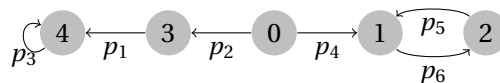
e, a partir disso, o tempo médio de recorrência do estado 1 é

$$\mu_1 = \sum_{n \geq 0} n f_{1,1}^{(n)} = \sum_{n \geq 0} \frac{1}{n+1} = \infty.$$

Em resumo, o estado 1 é **recorrente nulo**.

Os estados 0 e  $n$  da cadeia do exemplo 113, quando são atingidos a cadeia não muda mais de estado; nesses casos chamamos tais estados de **absorvente**, que é um estado recorrente positivo.

**Exemplo 118** Na cadeia representada pelo esquema abaixo o estado 4 é absorvente, os estados 0 e 3 são transiente e os estados 1 e 2 são recorrentes



**Exercício 95** Prove que numa cadeia de Markov sobre um conjunto finito de estados, pelo menos um estado é recorrente e todo estado recorrente é positivo

**Exercício 96** Prove que

$$p_{j,k}^{(n)} = \sum_{t=1}^n f_{j,k}^{(t)} p_{k,k}^{(n-t)}. \quad (54)$$

Para tal, deduza da propriedade de Markov que

$$\begin{aligned} & \mathbb{P}(X_0 = j, X_1 \neq k, X_2 \neq k, \dots, X_{t-1} \neq k, X_t = k, X_n = k) = \\ & \mathbb{P}(X_0 = j, X_1 \neq k, X_2 \neq k, \dots, X_{n-1} \neq k, X_t = k \mid X_0 = j) \mathbb{P}(X_n = k \mid X_t = k). \end{aligned}$$

**Exercício 97** Seja  $s$  um estado absorvente numa cadeia de Markov tal que para todo estado  $i$  da cadeia  $p_{i,s}^{(n)} > 0$ , para algum  $n = n(i)$ . Mostre que todos os estados, a não ser  $s$ , são transientes.

**Exercício 98** Considere a variável aleatória  $T_i$  definida acima com a convenção de que se  $i$  não é visitado então  $T_i = \infty$ . Verifique que

$$\mu_i = \mathbb{E}(T_i \mid X_0 = i).$$

Verifique também que o estado  $i$  é transiente se e só se  $\mathbb{P}(T_i = \infty \mid X_0 = i) > 0$  e, nesse caso,  $\mathbb{E}(T_i \mid X_0 = j) = \infty$ .

**Exercício 99** Considere a variável aleatória  $V_j$  definida acima e defina

$$\eta_{i,j} \stackrel{\text{def}}{=} \mathbb{P}(V_j = \infty \mid X_0 = j).$$

Prove que

$$\eta_{i,i} = \begin{cases} 1, & \text{se } i \text{ recorrente,} \\ 0, & \text{se } i \text{ transiente.} \end{cases}$$

e que

$$\eta_{i,j} = \begin{cases} \mathbb{P}(T_j < \infty \mid X_0 = j), & \text{se } i \text{ recorrente,} \\ 0, & \text{se } i \text{ transiente.} \end{cases}$$

**Exercício 100** Seja  $A$  um subconjunto de estados e defina

$$\begin{aligned} T_A & \stackrel{\text{def}}{=} \min\{n \geq 1 : X_n \in A\} \\ \eta_j & \stackrel{\text{def}}{=} \mathbb{P}(T_A < \infty \mid X_0 = j). \end{aligned}$$

Mostre que  $\eta_j = 1$  se  $j \in A$  e

$$\eta_j = \sum_{s \in S} p_{j,s} \eta_s$$

se  $j \notin A$ . Agora defina

$$\xi_j \stackrel{\text{def}}{=} \mathbb{E}(T_A \mid X_0 = j)$$

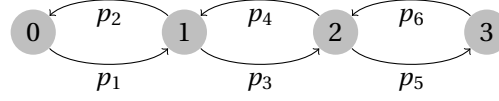
e mostre que  $\xi_j = 0$  se  $j \in A$  e que

$$\xi_j = 1 + \sum_{s \in S} p_{j,s} \xi_s$$

se  $j \notin A$ .

## Periodicidade

Numa cadeia de Markov que pode ser representada como



se  $X_0 = 0$  então  $p_{0,0}^{(n)} > 0$  só pode ocorrer na cadeia em instantes  $n$  da forma  $n = 2k$ ,  $n = 4k$ ,  $n = 6k$ . Para o estado 1, as probabilidades de retorno positivas só ocorrem para  $n = 2k$ ,  $n = 4k$ . Para o estado 2 ocorre o mesmo fenômeno que no estado 1 e o caso do estado 3 é semelhante ao caso do estado 0.

O período de um estado  $i$  numa cadeia de Markov é

$$\tau(i) = \text{mdc} \left\{ n : p_{i,i}^{(n)} > 0 \right\}$$

Caso  $\tau(i) > 1$  dizemos que  $i$  é **periódico**, nesse caso  $p_{i,i}^{(n)} = 0$  a menos que  $n$  seja múltiplo de  $\tau_i$ , e caso  $\tau(i) = 1$  dizemos que  $i$  é **aperiódico**. No exemplo acima todos os estados são periódicos de período  $t = 2$ .

**Observação 12** Suponha que, como no exemplo acima, numa cadeia de Markov todos os estados tenham o mesmo período  $t > 1$  e que para cada par de estados  $(i, j)$  existe um instante  $n$  par ao qual  $p_{i,j}^{(n)} > 0$  (de fato, essas 2 hipóteses são, em certo sentido, redundantes, veja o lema 13).

Para todo estado  $k$  da cadeia devem existir instantes  $m$  e  $n$  tais que

$$p_{0,k}^{(n)}, p_{k,0}^{(m)} > 0 \quad (55)$$

e como  $p_{0,0}^{(n+m)} \geq p_{0,k}^{(n)} p_{k,0}^{(m)} > 0$  devemos ter que  $t$  divide  $n + m$ . Fixando  $m$  concluímos que todo  $n$  para o qual  $p_{0,k}^{(n)} > 0$  é da forma  $a + vt$  para  $0 \leq a < t$  inteiro. Assim, podemos particionar  $S$  em  $S_0, S_1, \dots, S_{t-1}$  (no exemplo acima  $S_0 = \{0, 2\}$  e  $S_1 = \{1, 3\}$ ) para os valores de  $a$  acima de modo que se  $k \in S_a$  então  $p_{0,k}^{(n)} = 0$  a menos que  $n = a + vt$ . Agora consideramos os estados na ordem  $S_0, \dots, S_{t-1}, S_0$  ciclicamente e um passo da cadeia sai de um estado para um estado na classe  $a$  direita, e a cada  $t$  passos a cadeia está de volta à mesma classe.

## Classificação das cadeias

A condição  $p_{i,j}^{(n)} > 0$  para algum  $n$  equivale a  $f_{i,j} > 0$  e significa que a partir do estado  $i$  a cadeia eventualmente atinge o estado  $j$ ; nesse caso dizemos que  $j$  é **acessível** a partir de  $i$  e escrevemos  $i \rightarrow j$ . Quando  $j$  não é acessível a partir de  $i$  a probabilidade da cadeia chegar em  $j$  saindo de  $i$  é

$$\mathbb{P} \left( \bigcup_{n \geq 0} [X_n = j] \mid X_0 = i \right) \leq \sum_{n \geq 0} \mathbb{P}(X_n = j \mid X_0 = i) = 0.$$

Escrevemos  $i \leftrightarrow j$  se  $i \rightarrow j$  e  $j \rightarrow i$  e dizemos que os estados  $i$  e  $j$  se **comunicam**. Deixamos para o leitor a verificação de que  $\leftrightarrow$  define uma relação de equivalência sobre  $S$ , i.e.,

1.  $i \leftrightarrow i$ ,
2. se  $i \leftrightarrow j$ , então  $j \leftrightarrow i$ ,
3. se  $i \leftrightarrow k$  e  $k \leftrightarrow j$  então  $i \leftrightarrow j$ ,

portanto, ela particiona  $S$  em classes de equivalência, que chamaremos *classe de comunicação*. A importância dessa partição é pelo fato enunciado no exercício a seguir.

**Lema 13** *Recorrência e transiência são propriedades de classe de equivalência de  $\leftrightarrow$ , isto é, dois estados que se comunicam tem a mesma classificação. Ademais, estados que se comunicam têm o mesmo período.*

DEMONSTRAÇÃO: Sejam  $i$  e  $j$  estados que se comunicam e sejam  $n$  e  $m$  inteiros positivos tais que  $p_{i,j}^{(n)} > 0$  e  $p_{j,i}^{(m)} > 0$ . Para todo  $t$

$$p_{i,i}^{(t+n+m)} \geq p_{i,j}^{(n)} p_{j,j}^{(t)} p_{j,i}^{(m)} \quad (56)$$

portanto,  $p_{i,i}^{(t+n+m)} > 0$  sempre que  $p_{j,j}^{(t)} > 0$ .

Se fizermos  $t = 0$  na equação (56), resulta que  $p_{i,i}^{(n+m)} > 0$  portanto  $\tau(i)$  divide  $n + m$ . O lado esquerdo da equação (56) é nulo a menos nos períodos múltiplos de  $\tau(i)$ , portanto  $p_{j,j}^{(t)} > 0$  só quando  $t$  é múltiplo de  $\tau(i)$ , portanto  $\tau(j)$  divide  $\tau(i)$ . Trocando os papéis de  $i$  e  $j$  concluiremos que  $\tau(i)$  divide  $\tau(j)$ . Portanto, os períodos são iguais.

Se somamos os dois lados (56) sobre todo natural  $t$  temos que se  $\sum_t p_{j,j}^{(t)}$  não converge, então também não converge o lado esquerdo, portanto se  $j$  é recorrente então também será o estado  $i$ . Trocando os papéis de  $i$  e  $j$  concluiremos que se  $i$  é recorrente então  $j$  é recorrente. Ademais, se  $j$  é transiente então também será  $i$  e vice-versa.  $\square$

Um conjunto  $C \subset S$  é **irredutível** se  $i \leftrightarrow j$  para todos  $i, j \in C$ , portanto, cada classe de comunicação é um subconjunto irredutível maximal. Quando um conjunto não é irredutível, dizemos *redutível*. Quando há uma única classe de comunicação dizemos que a cadeia é uma **cadeia de Markov irredutível**.

**Exemplo 119** *Uma cadeia de Markov com estados  $\{1, 2\}$  e matriz de transições*

$$P = \begin{pmatrix} 1 & 0 \\ 1/2 & 1/2 \end{pmatrix}$$

*não corresponde a uma cadeia irredutível pois para todo inteiro  $k > 0$*

$$P^k = \begin{pmatrix} 1 & 0 \\ \frac{2^k - 1}{2^k} & \frac{1}{2^k} \end{pmatrix}$$

*portanto  $1 \nleftrightarrow 2$ .*

**Exemplo 120** Uma cadeia com estados  $\{1, 2, 3\}$  e matriz de transições

$$P = \begin{pmatrix} 1/3 & 2/3 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

é redutível pois  $P^k$  é da forma

$$\begin{pmatrix} \frac{x}{3^k} & \frac{3^k - x}{3^k} & 0 \\ \frac{y}{3^{k-1}} & \frac{3^{k-1} - y}{3^k} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

com  $0 < x < 3^k$  e com  $0 < y < 3^{k-1}$  e para todo inteiro  $k > 0$ .

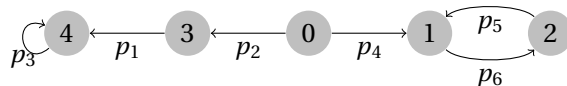
No exemplo 115 é fácil ver a partir do diagrama que a cadeia é irredutível. A cadeia do exemplo 113, claramente, não é irredutível por causa dos estados absorventes 0 e  $n$ . Toda cadeia com pelo menos dois estados e com pelo menos um deles absorvente é uma cadeia redutível.

Num conjunto irredutível de estados todos os estados são do mesmo tipo, logo podemos classificar um conjunto de estados, ou a própria cadeia de Markov, irredutível como

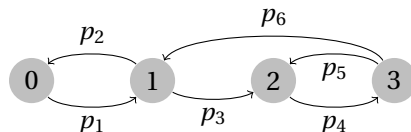
- **aperiódica** se todos os seus estados o forem;
- **periódica** se todos os seus estados o forem;
- **transiente** se todos os seus estados o forem;
- **recorrente** se todos os seus estados o forem.

**Observação 14** Notemos que a Observação 12 se aplica a uma classe de comunicação, ou a uma cadeia periódica, pois todos os estados dela têm o mesmo período.

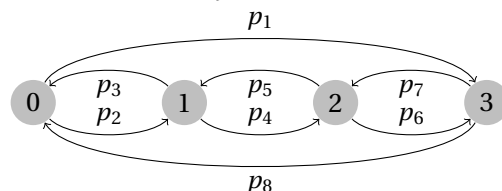
**Exemplo 121**



é uma cadeia redutível, a classe  $\{1, 2\}$  é periódica de período 2, o estado 4 é absorvente.



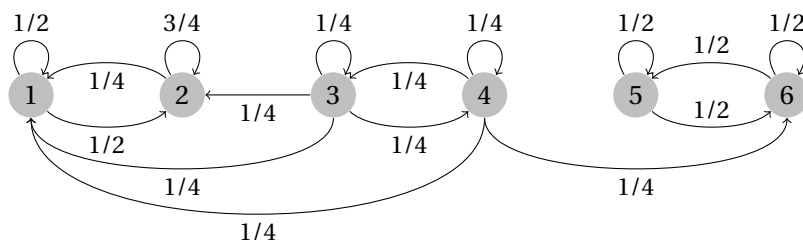
é uma cadeia aperiódica e irredutível.



é uma cadeia irredutível, recorrente e periódica.

**Exemplo 122** No caso da cadeia com matriz de transições

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ 1/4 & 3/4 & 0 & 0 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 & 0 & 0 \\ 1/4 & 0 & 1/4 & 1/4 & 0 & 1/4 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \end{pmatrix}$$



A cadeia é redutível e o conjunto dos estados têm três classes de equivalência de comunicação  $S = \{\{1, 2\}, \{3, 4\}, \{5, 6\}\}$ . Os conjuntos  $\{1, 2\}$  e  $\{5, 6\}$  são irreduzíveis e recorrentes;  $\{3, 4\}$  é transiente pois uma vez que a cadeia deixa o conjunto ela não volta mais a ele. Todos os estados têm período 1, pois  $p_{i,i} > 0$ , portanto a cadeia é aperiódica. Ainda

$$f_{1,1}^{(1)} = \frac{1}{2} \quad e \quad f_{1,1}^{(n)} = \frac{1}{2} \left(\frac{3}{4}\right)^{n-2} \frac{1}{4} \quad (n \geq 2)$$

portanto

$$\mu_1 = \frac{1}{2} + \frac{1}{8} \sum_{n \geq 2} n \left(\frac{3}{4}\right)^{n-2} = \frac{1}{2} + \frac{1}{6} \sum_{n \geq 2} n \left(\frac{3}{4}\right)^{n-1} = \frac{1}{2} - \frac{1}{6} + \frac{1}{6} \sum_{n \geq 2} n \left(\frac{3}{4}\right)^{n-1} = \frac{1}{3} + \frac{1}{6} 12 = \frac{7}{3}.$$

**Exemplo 123 (passeio aleatório, exemplo 109)** Vejamos o caso do passeio aleatório do exemplo 109, onde

$$p_{i,i+1} = p = 1 - p_{i,i-1}.$$

Claramente, a cadeia é irreduzível logo todos os estados são ou recorrentes ou transientes. Fixemos no estado 0 e vamos determinar  $\sum_{n \geq 0} p_{0,0}^{(n)}$ .

Como é impossível estar num estado par com um número ímpar de movimentos

$$p_{0,0}^{(2n-1)} = 0, \quad \text{para } n \geq 1.$$

Após  $2n$  passos a cadeia estará de volta em 0 se e só se metade dos passos foi num sentido (e.g., do estado atual para o maior, o que ocorre com probabilidade  $p$ ) e metade no sentido oposto, portanto

$$p_{0,0}^{(2n)} = \binom{2n}{n} p^n (1-p)^n = \frac{(2n)!}{n!} n! p^n (1-p)^n \sim \frac{(4p(1-p))^n}{\sqrt{\pi n}}$$

onde  $a_n \sim b_n$  significa que  $\lim_{n \rightarrow \infty} (a_n/b_n) = 1$  e  $n! \sim n^{n+1/2} e^{-n} \sqrt{2\pi}$ . Agora, se  $a_n \sim b_n$  então  $\sum_n a_n$  converge se, e só se,  $\sum_n b_n$  converge, portanto precisamos estudar a convergência de

$$\sum_{n \geq 1} \frac{(4p(1-p))^n}{\sqrt{\pi n}}.$$

No numerador  $4p(1-p) \leq 1$  e vale a igualdade se e somente se  $p = 1/2$ . Portanto a série é infinita se e só se  $p = 1/2$  e como a cadeia é irredutível e o 0 recorrente, a cadeia é recorrente. Se  $p \neq 1/2$  então a cadeia é transiente.

**Exercício 101** Considere a cadeia de Markov com estados  $\mathbb{N}$  e transições  $p_{0,j} = a_j$  para todo  $j \in \mathbb{N}$ , e  $p_{i,i} = p$ , e  $p_{i,i-1} = 1-p$  para todo  $i \neq 0$ . Classifique os estados e determine os tempos médios de recorrência.

Essa classificação, num certo sentido, diz-nos que basta estudarmos as cadeias irredutíveis. Se  $X_0$  está em alguma classe de equivalência de comunicação  $C_\ell$  que é recorrente, então a evolução da cadeia se restringe a essa classe pois, assumindo que  $i \in C_\ell$  e  $j \notin C_\ell$  de modo que  $i \rightarrow j$ , por conseguinte  $j \not\rightarrow i$ , temos de  $[X_1 = j] \subset \bigcup_{n \geq 1} [X_n \neq i]$

$$\mathbb{P}\left(\bigcup_{n \geq 1} [X_n \neq i] \mid X_0 = i\right) \geq p_{i,j} > 0$$

contrariando o fato de  $i$  ser recorrente. Se, por outro lado,  $X_0$  no conjunto  $T$  dos estados transientes implica em ou a cadeia ficar em  $T$  (consulte seção XV.8 do Feller para a probabilidade desse evento) ou se mover para algum  $C_\ell$  e não sair mais dessa classe.

**Exemplo 124** Considere uma cadeia de Markov com matriz de transição

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/4 & 3/4 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

cujas classes de comunicação são  $\{0, 1\}$  que é recorrente,  $\{3\}$  que é absorvente e  $\{2\}$  que é transiente. Se o início é no estado 2, então a probabilidade  $p$  com que a cadeia entra na classe recorrente  $\{0, 1\}$  é, condicionando em  $X_1$ , dada por

$$p = \frac{1}{4}1 + \frac{1}{4}1 + \frac{1}{4}p + \frac{1}{4}0 = \frac{1}{2} + \frac{1}{4}p$$

logo  $p = \frac{2}{3}$ ; e com probabilidade  $\frac{1}{3}$  a cadeia é absorvida em 3.

No caso da matriz de transição

$$Q = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ 1/3 & 2/3 & 0 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 1/3 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 0 & 1/3 & 1/6 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

as classes de comunicação são  $\{0,1\}$ ,  $\{2,3\}$  e  $\{4,5\}$ , respectivamente, recorrente aperiódica, transitiva e recorrente periódica. Se  $p_i$  é a probabilidade de entrar na primeira classe descrita acima a partir do estado  $i$ ,  $i = 2,3$ , então, condicionando em  $X_1$ ,

$$\begin{aligned} p_2 &= \frac{1}{2} \cdot 1 + 0 \cdot 1 + 0p_2 + \frac{1}{3}p_3 + \frac{1}{6} \cdot 0 + \frac{1}{6} \cdot 0 \\ p_3 &= \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 1 + \frac{1}{6}p_2 + 0p_3 + \frac{1}{3} \cdot 0 + \frac{1}{6} \cdot 0 \end{aligned}$$

cuja solução é  $p_2 = \frac{8}{17}$  e  $p_3 = \frac{7}{17}$ .

Uma classe de comunicação é dita *fechada* se  $j \in C$  sempre que  $i \rightarrow j$  e  $i \in C$ . Acima verificamos que toda classe recorrente é fechada.

**Exercício 102** Mostre que toda classe fechada finita é recorrente.

## Distribuição invariante e Convergência ao equilíbrio

Recordemos a matriz de transição e transição em dois passos do exemplo 116

$$P = \begin{pmatrix} 0,7 & 0 & 0,3 & 0 \\ 0,5 & 0 & 0,5 & 0 \\ 0 & 0,4 & 0 & 0,6 \\ 0 & 0,2 & 0 & 0,8 \end{pmatrix} \text{ e } P^2 = \begin{pmatrix} 0,49 & 0,21 & 0,21 & 0,18 \\ 0,35 & 0,20 & 0,15 & 0,30 \\ 0,20 & 0,12 & 0,20 & 0,48 \\ 0,10 & 0,16 & 0,10 & 0,64 \end{pmatrix}$$

Se formos um pouco mais adiante obtemos

$$\begin{aligned} P^6 &= \begin{pmatrix} 0.285709 & 0.144552 & 0.162561 & 0.407178 \\ 0.270935 & 0.147320 & 0.156915 & 0.424830 \\ 0.240920 & 0.150912 & 0.147320 & 0.460848 \\ 0.226210 & 0.153616 & 0.141610 & 0.478564 \end{pmatrix} \\ P^{10} &= \begin{pmatrix} 0.2553440 & 0.1491648 & 0.1519040 & 0.4435872 \\ 0.2531733 & 0.1495093 & 0.1511255 & 0.4461919 \\ 0.2486079 & 0.1502124 & 0.1495093 & 0.4516704 \\ 0.2464373 & 0.1505568 & 0.1487306 & 0.4542752 \end{pmatrix} \\ P^{20} &= \begin{pmatrix} 0.2500461 & 0.1499928 & 0.1500164 & 0.4499447 \\ 0.2500274 & 0.1499957 & 0.1500098 & 0.4499671 \\ 0.2499880 & 0.1500019 & 0.1499957 & 0.4500144 \\ 0.2499693 & 0.1500048 & 0.1499890 & 0.4500369 \end{pmatrix} \\ P^{30} &= \begin{pmatrix} 0.2500004 & 0.1499999 & 0.1500001 & 0.4499995 \\ 0.2500002 & 0.1500000 & 0.1500001 & 0.4499997 \\ 0.2499999 & 0.1500000 & 0.1500000 & 0.4500001 \\ 0.2499997 & 0.1500000 & 0.1499999 & 0.4500003 \end{pmatrix} \end{aligned}$$



$$P^{34} = \begin{pmatrix} 0.2500001 & 0.15 & 0.15 & 0.4499999 \\ 0.2500000 & 0.15 & 0.15 & 0.4500000 \\ 0.2500000 & 0.15 & 0.15 & 0.4500000 \\ 0.2500000 & 0.15 & 0.15 & 0.4500000 \end{pmatrix}$$

$$P^{35} = P^{36} = \dots P^{40} = \dots = \begin{pmatrix} 0.25 & 0.15 & 0.15 & 0.45 \\ 0.25 & 0.15 & 0.15 & 0.45 \\ 0.25 & 0.15 & 0.15 & 0.45 \\ 0.25 & 0.15 & 0.15 & 0.45 \end{pmatrix}$$

Ainda, se fizermos  $\pi^{(0)} = \rho$ , a distribuição inicial, então  $\pi^{(1)} = \pi^{(0)}P$  é a distribuição de  $X_1$ . Por exemplo, para  $\pi^{(0)} = (1/4 \ 1/4 \ 1/4 \ 1/4)$ , a distribuição uniforme em  $S$ , então temos  $\pi^{(1)} = (0,285 \ 0,15 \ 0,165 \ 0,4)$  de modo que  $\pi_i^{(1)} = \mathbb{P}(X_1 = i)$ . Analogamente,  $\pi^{(2)}$  é a distribuição de  $X_2$  e, em geral, para  $n \in \mathbb{N}$

$$\pi^{(n+1)} = \pi^{(n)}P$$

é a distribuição de  $X_{n+1}$ . Notemos que

$$\pi^{(n+1)} = \pi^{(n)}P = (\pi^{(n-1)}P)P = \dots = \pi^{(0)}P^n$$

portanto, para  $n \geq 36$  o vetor  $\pi^{(n)}$  não muda e é igual a

$$\pi = (0.25 \ 0.15 \ 0.15 \ 0.45).$$

Além disso  $\pi^{(0)}P^n$  na coluna  $j$  vale

$$\begin{aligned} (\pi^{(0)}P^n)_j &= \pi_1^{(0)}p_{1,j}^{(n)} + \pi_2^{(0)}p_{2,j}^{(n)} + \pi_3^{(0)}p_{3,j}^{(n)} + \pi_4^{(0)}p_{4,j}^{(n)} \\ &= (\pi_1^{(0)} + \pi_2^{(0)} + \pi_3^{(0)} + \pi_4^{(0)})\pi_j \\ &= \pi_j \end{aligned}$$

para  $n \geq 35$  pois  $p_{i,j}^{(n)} = \pi_j$  para todo  $i$ . Na última igualdade usamos que  $\pi_1^{(0)} + \pi_2^{(0)} + \pi_3^{(0)} + \pi_4^{(0)} = 1$ , portanto a conclusão é que  $\pi$  não depende da distribuição inicial  $\pi^{(0)}$ , isto é, independentemente do estado inicial, se  $n$  for suficientemente grande então a probabilidade da cadeia de Markov estar em qualquer um dos estados é dada pelo vetor

$$\pi = (0.25 \ 0.15 \ 0.15 \ 0.45).$$

Para  $\pi$  e  $P$  acima valem

$$\pi P = \pi \quad \text{e} \quad \pi_j = \lim_{n \rightarrow \infty} p_{i,j}^{(n)}. \quad (57)$$

Um vetor de probabilidades  $\pi = (\pi_j)_{j \in S}$  que satisfaz  $\pi = \pi P$  é chamado de **distribuição invariante** ou **distribuição estacionária** da cadeia de Markov com matriz de transições  $P$ . O limite é o que chamamos de convergência ao equilíbrio.

**Exemplo 125** Uma cadeia redutível com matriz de transição

$$P = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

tem distribuição estacionária  $\pi = (p \quad 1-p)$  para qualquer  $p \in (0,1)$  e, por exemplo, para  $p = 1/2$  temos que  $p_{1,1}^{(n)} \not\rightarrow 1/2$ . O mesmo acontece com a matriz de transição

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

e  $\pi = (1/2 \quad 1/2)$  é estacionário mas não há convergência. A cadeia é periódica. Considere uma cadeia de Markov com quatro estado e matriz de transições

$$\begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

nesse caso,  $(1, 0, 1, 0)$  e  $(0, 1, 0, 1)$  são vetores estacionários.

Quando  $S$  é finito não é difícil ver que a convergência e a invariância estão intimamente relacionadas pois se  $p_{i,j}^{(n)} \rightarrow \pi_j$  então temos uma distribuição invariante pois

$$\pi_j = \lim_{n \rightarrow \infty} p_{i,j}^{(n)} = \lim_{n \rightarrow \infty} \sum_{k \in S} p_{i,k}^{(n)} p_{k,j} = \sum_{k \in S} \lim_{n \rightarrow \infty} p_{i,k}^{(n)} p_{k,j} = \sum_{k \in S} \pi_k p_{k,j}$$

e

$$\sum_{i \in S} \pi_i = \sum_{i \in S} \lim_{n \rightarrow \infty} p_{i,j}^{(n)} = \lim_{n \rightarrow \infty} \sum_{i \in S} p_{i,j}^{(n)} = 1.$$

Quando  $S$  não é finito isso não vale necessariamente, por exemplo, no caso assimétrico ( $p \neq 1/2$ ) do passeio aleatório em  $\mathbb{Z}$ , exemplo 123, temos  $p_{i,j}^{(n)} \rightarrow 0$  pois a cadeia é transiente, entretanto,  $\pi$  não é uma distribuição.

**Teorema 13** Uma cadeia de Markov irredutível tem uma distribuição estacionária  $\pi = (\pi_j)_j$  se, e somente se, todos os estados são recorrentes positivos. O vetor estacionário é único e satisfaz

$$\pi_j = \frac{1}{\mu_j}. \quad (58)$$

Ainda, se a cadeia for aperiódica então

$$\pi_j = \lim_{n \rightarrow \infty} p_{i,j}^{(n)}. \quad (59)$$

Usando R:

```
invariante = function() {
# Para determinar pi que satisfaz pi %*% trans = pi resolvemos o
# sistema (t(trans) - I) %* pi = 0 onde t() e' a matriz transposta e
# pi agora e' um vetor coluna. No lado direito trocamos a ultima linha
# por uma linha de 1's, e no lado esquerdo trocamos o vetor nulo por
# (0,0,...,0,1), isso representa a restricao da soma das entradas de pi
# ser 1. Entao, resolvemos para pi.

#####
  trans <- matrix(c(
    0 , 0.1 , 0.4 , 0.5,
    0 , 0.1 , 0.4 , 0.5,
    0.1 , 0.4 , 0.5 , 0,
    0 , 0.1 , 0.4 , 0.5
  ), ncol=4, byrow=TRUE)

  n <- nrow(trans)
  a <- t(trans)
  for(i in 1:n) { a[i, i] <- a[i, i] - 1 }
  a[n, ] <- 1
  b <- rep(0, n)
  b[n] <- 1
  solve(a, b)
}
```

No caso da cadeia ser periódica de período  $d$  então

$$\lim_{n \rightarrow \infty} p_{i,i}^{(nd)} = \frac{d}{\mu_i} \quad (60)$$

e ser for recorrente

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=0}^{n-1} p_{i,j}^{(m)} = \frac{1}{\mu_j} = \pi_j \quad (61)$$

com  $\pi$  (único) vetor de estacionário.

**Observação 15** Se uma sequência converge  $\lim_{n \rightarrow \infty} a_n = a$ , então a sequência das médias parciais também converge

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} a_i = a$$

portanto de (59) temos

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=0}^{n-1} p_{i,j}^{(t)} = \pi_j$$

e se  $I_t$  é a v.a. indicadora do evento  $[X_t = i]$  então

$$\sum_{t=0}^{n-1} p_{i,j}^{(t)} = \mathbb{E}(V_i^{[0..n-1]} \mid X_0 = i)$$

em que  $V_i^{[0..n-1]}$  é o número de visitas a  $i$  no intervalo de tempo  $0, \dots, n-1$  (veja o lema 11), ou seja,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}(V_i^{[0..n-1]} \mid X_0 = i) = \pi_j.$$

Em palavras,  $\pi_j$  é a fração média de visitas ao estado  $j$  por unidade de tempo.

**Teorema 14 (Teorema ergódico)** Para qualquer cadeia de Markov irredutível

$$\mathbb{P} \left( \lim_{n \rightarrow \infty} \frac{V_i^{[0..n-1]}}{n} = \frac{1}{\mu_i} \right) = 1.$$

**Exercício 103** Uma matriz quadrada com entradas não-negativas é duplamente estocástica se é estocástica e a soma das entradas de cada coluna é 1. Mostre que se  $M$  é quadrada de ordem  $n$  e duplamente estocástica então o vetor uniforme  $(1/n, 1/n, \dots, 1/n)$  é invariante para  $M$ .

**Observação 16** No caso finito a existência de um vetor estacionário é garantido para qualquer matriz estocástica. Esse fato decorre do Teorema do Ponto Fixo de Brower: Para toda função contínua  $f: T \rightarrow T$ , onde  $T \subset \mathbb{R}^n$  é compacto e convexo, existe  $x \in T$  tal que  $f(x) = x$ . Agora, consideremos para todo  $\mathbf{v} = (v_1, v_2, \dots, v_n) \in \mathbb{R}^n$  a norma  $L^1$

$$\|\mathbf{v}\|_1 = \sum_{i=1}^n |v_i| \quad (62)$$

e o conjunto (compacto e convexo)

$$T = \{\mathbf{v} \in \mathbb{R}^n : \mathbf{v} \geq 0 \text{ e } \|\mathbf{v}\|_1 = 1\}. \quad (63)$$

Seja  $P$  uma matriz estocástica e considere a função linear de  $T$  em  $T$  dada por  $\mathbf{x} \mapsto \mathbf{x}P$ . Por ser linear a função é contínua, portanto, pelo Teorema do Ponto Fixo de Brower deduzimos a existência de um vetor estacionário.

## Cadeias redutíveis

Os resultados da seção anterior aplicam-se a qualquer classe de comunicação recorrente. Se  $C_\ell$  é classe de comunicação recorrente e aperiódica de um processo, então a submatriz formada pelos  $i, j \in C_\ell$  é estocástica logo  $p_{i,j}^{(n)} \rightarrow 1/\mu_j$  quando  $n \rightarrow \infty$ .

**Exemplo 126** Por exemplo,

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/4 & 3/4 & 0 & 0 \\ 0 & 0 & 1/3 & 2/3 \\ 0 & 0 & 2/3 & 1/3 \end{pmatrix}$$

tem as classes recorrentes  $\{0, 1\}$  com a submatriz estocástica irredutível

$$P_1 = \begin{pmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{pmatrix}$$

e  $\{2, 3\}$  com a submatriz estocástica irredutível

$$P_2 = \begin{pmatrix} 1/3 & 2/3 \\ 2/3 & 1/3 \end{pmatrix}.$$

A matriz  $P_1$  admite o vetor estacionário  $\pi^{(1)} = (1/3 \ 2/3)$  e a matriz  $P_2$ , o vetor estacionário  $\pi^{(2)} = (1/2 \ 1/2)$  e, além disso

$$\lim_{n \rightarrow \infty} P^n = \begin{pmatrix} \pi_0^{(1)} & \pi_1^{(1)} & 0 & 0 \\ \pi_0^{(1)} & \pi_1^{(1)} & 0 & 0 \\ 0 & 0 & \pi_0^{(2)} & \pi_1^{(2)} \\ 0 & 0 & \pi_0^{(2)} & \pi_1^{(2)} \end{pmatrix} = \begin{pmatrix} 1/3 & 2/3 & 0 & 0 \\ 1/3 & 2/3 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \end{pmatrix}.$$

Ainda, se  $i, j \in C_\ell$ , com  $C_\ell$  classe de comunicação recorrente e periódica, então

$$\frac{1}{n} \sum_{m=0}^{n-1} p_{i,j}^{(m)} \rightarrow \frac{1}{\mu_j}$$

quando  $n \rightarrow \infty$ .

Por outro lado, se  $j$  é um estado transiente então  $p_{i,j}^{(n)} \rightarrow 0$  quando  $n \rightarrow \infty$  para qualquer estado inicial  $i$ .

**Exemplo 127** De volta ao exemplo 124, a classe recorrente  $\{0, 1\}$  com respeito a matriz  $P$  tem distribuição estacionária  $\pi = (1/3 \ 2/3)$ . Ademais essa classe é atingida a partir do estado 2 com probabilidade  $p = 2/3$ , portanto  $p_{2,0}^{(n)} \rightarrow \frac{2}{3} \cdot \frac{1}{3} = \frac{2}{9}$  e  $p_{2,1}^{(n)} \rightarrow \frac{2}{3} \cdot \frac{2}{3} = \frac{4}{9}$  logo

$$\lim_{n \rightarrow \infty} P^n = \begin{pmatrix} 1/3 & 2/3 & 0 & 0 \\ 1/4 & 3/4 & 0 & 0 \\ 2/9 & 4/9 & 0 & 1/3 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

No que diz respeito a matriz  $Q$ , temos  $\{0, 1\}$  recorrente aperiódico com  $\pi = (2/4 \ 3/5)$ ,  $\{2, 3\}$  transiente com probabilidade de sair de 2 (respec., 3) e ir ser absorvido por  $\{0, 1\}$  sendo  $p_2 = 8/17$  (respec.,  $7/17$ ), e  $\{4, 5\}$  recorrente periódico

$$\lim_{n \rightarrow \infty} Q^n = \begin{pmatrix} 2/5 & 3/5 & 0 & 0 & 0 & 0 \\ 2/5 & 3/5 & 0 & 0 & 0 & 0 \\ \frac{8}{17} \cdot \frac{2}{5} & \frac{8}{17} \cdot \frac{3}{5} & 0 & 0 & & \\ \frac{7}{17} \cdot \frac{2}{5} & \frac{7}{17} \cdot \frac{3}{5} & 0 & 0 & & \\ 0 & 0 & 0 & 0 & & \\ 0 & 0 & 0 & 0 & & \end{pmatrix}$$

em que os espaços vazios indicam que o limite não existe. Entretanto,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=0}^{n-1} Q^m = \begin{pmatrix} 2/5 & 3/5 & 0 & 0 & 0 & 0 \\ 2/5 & 3/5 & 0 & 0 & 0 & 0 \\ \frac{8}{17} \cdot \frac{2}{5} & \frac{8}{17} \cdot \frac{3}{5} & 0 & 0 & \frac{9}{17} \cdot \frac{1}{2} & \frac{9}{17} \cdot \frac{1}{2} \\ \frac{7}{17} \cdot \frac{2}{5} & \frac{7}{17} \cdot \frac{3}{5} & 0 & 0 & \frac{10}{17} \cdot \frac{1}{2} & \frac{10}{17} \cdot \frac{1}{2} \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \end{pmatrix}.$$

## Exemplo: 2-SAT

O 2-SAT (*2-Satisfiability Problem*) é o seguinte problema; dado uma fórmula booleana 2-CNF (cada cláusula é uma conjunção de 2 literais  $C_i = \ell_{i_1} \vee \ell_{i_2}$ ), digamos

$$\Phi = C_1 \wedge C_2 \wedge \cdots \wedge C_k$$

sobre um conjunto de variáveis  $V = \{x_1, x_2, \dots, x_n\}$ , determinar se existe uma valoração  $v: V \rightarrow \{0, 1\}$  que satisfaça  $\Phi$ , isto é, faça  $\Phi = 1$ .

É sabido<sup>1</sup> que 2-SAT pode ser resolvido em tempo polinomial. A seguir, daremos um algoritmo aleatorizado bem simples, de tempo polinomial e que responde corretamente com alta probabilidade.

**Dado** : cláusulas  $\mathcal{C} = \{\{\ell_{1_1}, \ell_{1_2}\}, \dots, \{\ell_{k_1}, \ell_{k_2}\}\}$  e as variáveis  $V = \{x_1, \dots, x_n\}$  de uma fórmula 2-CNF  $\Phi$  e  $t \in \mathbb{N}$ .

**Devolve**: 1 se existe valoração de  $V$  que satisfaz  $\Phi$ , ou 0 se não encontrou uma.

```

1 repita
2   para  $i$  de 1 até  $n$  faça  $x_i \leftarrow_R \{0, 1\}$ ;
3   repita
4     escolha, independente e uniformemente, um literal  $\ell$  de uma cláusula
       não satisfeita;
5     inverta o valor da variável do literal  $\ell$ ;
6     se  $\Phi = 1$  então responda 1
7   até completar  $2n^2$  rodadas;
8 até completar  $t$  rodadas;
9 responda 0.
```

**Algoritmo 1:** 2-SAT.

Quando existe uma valoração válida o algoritmo não a encontra com probabilidade menor que  $(1/2)^t$ .

Se não existe valoração que satisfaça todas as cláusulas o algoritmo termina após  $2tn^2$  rodadas, caso contrário, seja  $v$  uma valoração que satisfaz  $\Phi$  e  $v_i$  a valoração construída pelo algoritmo após a  $i$ -ésima rodada do laço interno, linha 3.

Denotemos por  $X_i$  o número de valores em comum que tomam as valorações  $v$  e  $v_i$ , isto é, a quantidade de variáveis da fórmula que têm o mesmo valor binário nas duas valorações. O algoritmo termina com  $X_i = n$  ou quando encontra alguma valoração diferente que  $v$  que satisfaça  $\Phi$ .

Notemos que valem  $\mathbb{P}(X_{i+1} = 1 \mid X_i = 0) = 1$  e

$$\mathbb{P}(X_{i+1} = j + 1 \mid X_i = j) \geq 1/2$$

$$\mathbb{P}(X_{i+1} = j - 1 \mid X_i = j) \leq 1/2$$

<sup>1</sup>Aspvall, Plass, Tarjan, *A linear-time algorithm for testing the truth of certain quantified Boolean formulas*, Inform. Process. Lett., 1979, no. 3, pg. 121–123

$(\mathbb{P}(X_{i+1} = j + 1 \mid X_i = j))$  vale 1 quando os dois literais envolvem variáveis que discordam de  $v$ ). As variáveis  $\{X_i\}_i$  não definem uma cadeia de Markov, entretanto, a cadeia do exemplo 114 é um cenário de pior caso, isto é,  $\mathbb{E}X_j \leq \mathbb{E}Y_j$ , onde  $Y_j$  é a variável aleatória do exemplo, por isso, usando (49) temos que

$$\mathbb{P}(X_0 \geq 2n^2) \leq \mathbb{P}(Y_0 \geq 2n^2) \leq \frac{1}{2}$$

portanto em cada rodada do laço interno a probabilidade de não encontrar uma valoração que satisfaz  $\Phi$  é no máximo  $1/2$ . A probabilidade de não ser encontrada uma valoração nas  $t$  rodadas do laço externo, linha 1, é no máximo  $(1/2)^t$ . Se a fórmula é satisfatível então o laço interno encontrará a valoração certa com alta probabilidade se for repetido  $n^2 \log n$  vezes.

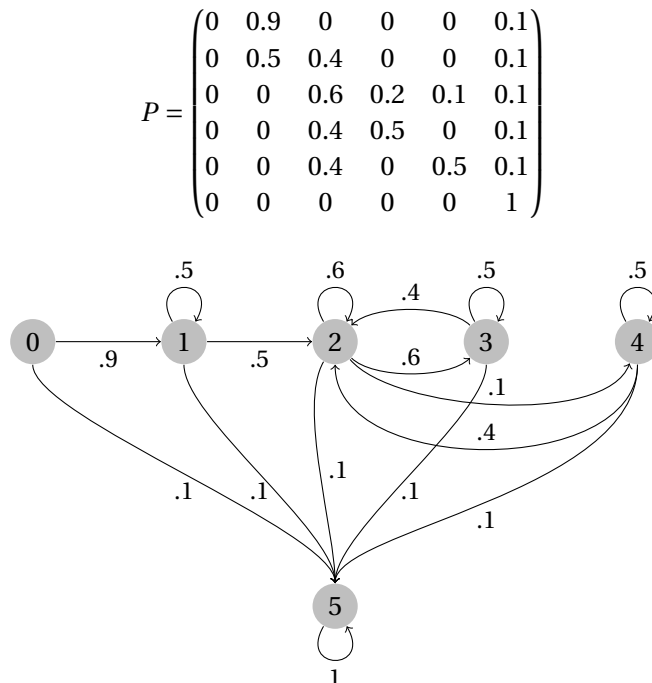
**Exercício 104** Por que  $\{X_n\}_n$  acima não define uma cadeia de Markov?

**Exercício 105** Prove a afirmação  $\mathbb{E}X_j \leq \mathbb{E}Y_j$  feita no exemplo acima.

### Exemplo: Modelo de fecundidade

As mudanças nos padrões sociais afetam a taxa de crescimento populacional. Modelos para analisar os efeitos das mudanças na fecundidade média de uma população levam em conta a idade, a situação conjugal, o número de filhos e várias outras informações a respeito do público feminino da população. Essas características compõem cada estado de um processo cujo interesse é determinar o tempo médio do processo nas categorias de maior fertilidade. As probabilidades de transição são inferidas de dados demográficos.

Num modelo simplificado consideramos os estados 0  $\equiv$  pré-puberil, 1  $\equiv$  solteira, 2  $\equiv$  casada, 3  $\equiv$  Divorciada, 4  $\equiv$  Viúva, 5  $\equiv$  morte ou emigração.



Denotemos por  $e_i$  o tempo médio da cadeia no estado 2 dado que  $X_0 = i$ . Então

$$\begin{cases} e_0 = 0.9e_1 + 0.1e_5 \\ e_1 = 0.5e_1 + 0.4e_2 + 0.1e_5 \\ e_2 = 1 + 0.6e_2 + 0.2e_3 + 0.1e_4 + 0.1e_5 \\ e_3 = 0.4e_2 + 0.5e_3 + 0.1e_5 \\ e_4 = 0.4e_4 + 0.5e_4 + 0.1e_5 \\ e_5 = 0 \end{cases}$$

cuja (única) solução é  $e_0 = 4.5$ ,  $e_1 = 5$ ,  $e_2 = 6.25$ ,  $e_3 = e_4 = 5$ .

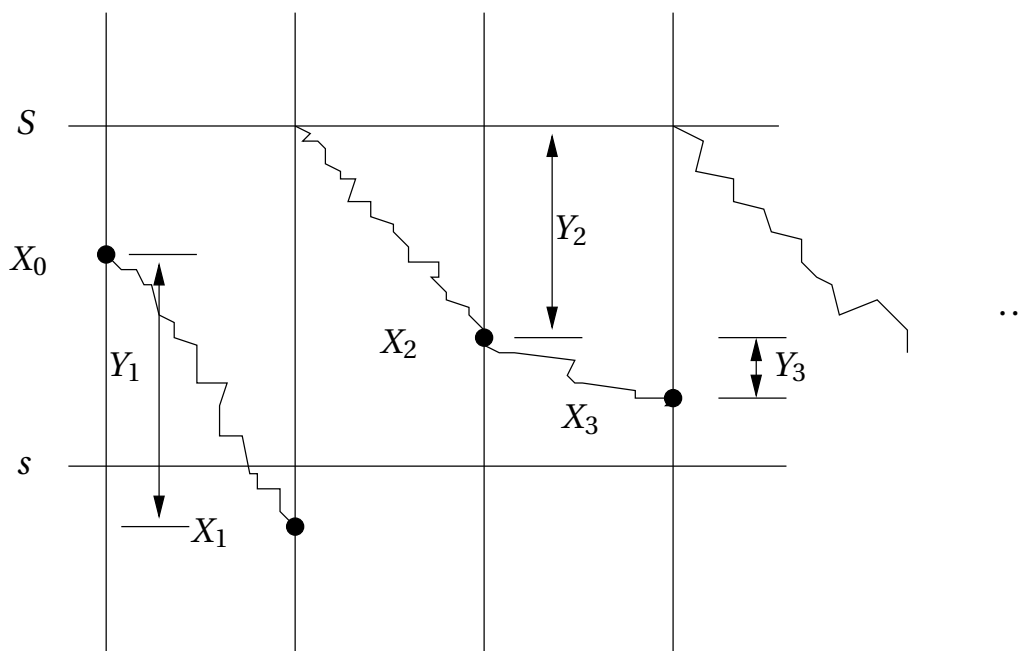
Notemos que, como esperado, a distribuição estacionária é  $(0 \ 0 \ 0 \ 0 \ 1)$ .

### Exemplo: Controle de estoque

Uma mercadoria é armazenada e ao final dos períodos  $n = 0, 1, \dots$  o estoque é reabastecido. A demanda pela mercadoria no período  $n$  é uma variável aleatória  $Y_n$  com  $Y_1, Y_2, \dots$  independentes e identicamente distribuídas.

A política de reabastecimento envolve escolher dois parâmetros inteiros e positivos  $s$  e  $S$  de modo que se ao final de um período o estoque é no máximo  $s$  então ele é abastecido até atingir  $S$ , caso contrário não é abastecido.

A variável aleatória  $X_n$  é a quantidade de mercadoria ao fim do período  $n$  antes do reabastecimento e caso seja negativa a demanda é atendida após o reabastecimento.





Os níveis de estoque em dois períodos consecutivos são relacionados por

$$X_{n+1} = \begin{cases} X_n - Y_{n+1} & s < X_n \leq S \\ S - Y_{n+1} & X_n \leq s \end{cases}$$

e as probabilidades das transições

$$p_{i,j} = \mathbb{P}(X_{n+1} = j \mid X_n = i) = \begin{cases} \mathbb{P}(Y_{n+1} = i - j) & s < X_n \leq S \\ \mathbb{P}(Y_{n+1} = S - j) & X_n \leq s \end{cases}$$

Nesse caso, gostaríamos de saber qual é, a longo prazo, a fração de períodos que a demanda não é atendida

$$\lim_{n \rightarrow \infty} \sum_{j < 0} \mathbb{P}(X_n = j)$$

e qual é, a longo prazo, o nível médio de estoque

$$\lim_{n \rightarrow \infty} \sum_{j > 0} j \mathbb{P}(X_n = j)$$

e sob certas condições  $\mathbb{P}(X_n = j)$  converge.

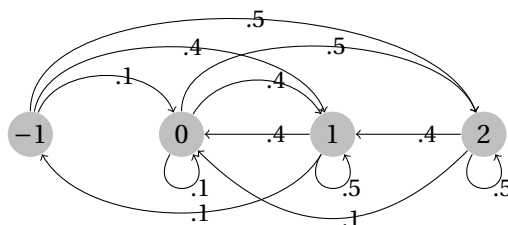
Para um exemplo numérico tomemos

$$\mathbb{P}(Y_n = 0) = 0.5, \quad \mathbb{P}(Y_n = 1) = 0.4, \quad \mathbb{P}(Y_n = 2) = 0.1$$

e  $s = 0$  e  $S = 2$ . Assim,  $X_n \in \{-1, 0, 1, 2\}$ .

Quando  $X_n = 1$  então não é necessário reabastecimento do estoque, logo  $X_{n+1} = 0$  caso  $Y_{n+1} = 1$ , o que ocorre com probabilidade 0.4, logo  $p_{1,0} = 0.4$ . Agora, se  $X_n = 0$  então há reabastecimento para  $S = 2$  e  $X_{n+1} = 0$  caso  $Y_{n+1} = 2$ , logo  $p_{0,0} = 0.1$ . A matriz fica da seguinte forma, lembrando que os estados são  $\{-1, 0, 1, 2\}$

$$P = \begin{pmatrix} 0 & 0.1 & 0.4 & 0.5 \\ 0 & 0.1 & 0.4 & 0.5 \\ 0.1 & 0.4 & 0.5 & 0 \\ 0 & 0.1 & 0.4 & 0.5 \end{pmatrix}$$



que, claramente, é irreduzível, recorrente e aperiódica, cujo vetor estacionário é

$$(4/90 \quad 21/90 \quad 40/90 \quad 25/90).$$

## Reversibilidade

Numa cadeia de Markov, dado o estado presente, o futuro e o passado são independentes. Seja  $\{X_i\}_{i \geq 0}$  uma cadeia de Markov irredutível com respeito a matriz estocástica  $P$  e a distribuição inicial  $\pi$ , e  $\pi$  uma distribuição invariante. Tomemos a matriz  $Q = q_{i,j}$  dada por

$$q_{i,j} = \frac{\pi_j}{\pi_i} p_{j,i}.$$

Essa matriz é estocástica pois, pela invariância de  $\pi$

$$\sum_{j \in S} q_{i,j} = \frac{1}{\pi_i} \sum_{j \in S} \pi_j p_{j,i} = 1.$$

Ainda,

$$\sum_{j \in S} \pi_i q_{i,j} = \sum_{j \in S} \pi_j p_{j,i} = \pi_i$$

ou seja,  $\pi$  é invariante com relação a  $Q$ .

Se tomarmos  $Y_n = X_{N-n}$  para  $0 \leq n \leq N$  então

$$\mathbb{P}(Y_0 = i_0, Y_1 = i_1, \dots, Y_N = i_N) = \mathbb{P}(X_0 = i_N, X_1 = i_{N-1}, \dots, X_N = i_0) =$$

$$\pi_{i_N} p_{i_N, i_{N-1}} \cdots p_{i_1, i_0} = \pi_{i_0} q_{i_0, i_1} \cdots q_{i_{N-1}, i_N}$$

logo  $\{Y_n\}_{0 \leq n \leq N}$  é uma cadeia de Markov com relação a  $Q$  e  $\pi$ . Ademais,

$$q_{i_N, i_{N-1}} \cdots q_{i_1, i_0} = \frac{1}{\pi_{i_0}} \pi_{i_N} p_{i_0, i_1} \cdots p_{i_{N-1}, i_N} > 0$$

portanto, a cadeia é irredutível, chamada *tempo-reverso* da cadeia  $\{X_i\}_{0 \leq i \leq N}$ .

**Exercício 106** Mostre que se  $P$  é uma matriz estocástica e  $\lambda$  um vetor não-negativo tal que para todos  $i, j$

$$\lambda_i p_{i,j} = \lambda_j p_{j,i} \tag{64}$$

então  $\lambda P = \lambda$ .

**Exercício 107** Prove que  $\{X_n\}_{n \geq 0}$  irredutível, com matriz de transições  $P$  e distribuição inicial  $\lambda$  é reversível se e só se vale (64).

**Exemplo 128** Consideremos um grafo conexo  $G$  com peso positivo  $w_{i,j}$  na aresta em cada aresta  $\{i, j\} \in E(G)$ . Uma partícula move-se sobre o grafo com a seguinte estratégia: se a partícula encontra-se no vértice  $i$  no instante  $t$  então no instante  $t+1$  a partícula estará no estado  $j$  com probabilidade

$$p_{i,j} = \frac{w_{i,j}}{\sum_{j \in V(G)} w_{i,j}}$$

de modo que  $w_{i,j} = 0$  caso  $\{i, j\}$  não seja uma aresta. A equação (64) é equivalente a

$$\pi_i = \frac{\sum_{j \in V(G)} w_{i,j}}{\sum_{i \in V(G)} \sum_{j \in V(G)} w_{i,j}}.$$

No caso mais simples,  $w_{i,j} \in \{0, 1\}$  e  $w_{i,j} = 1$  se e só se  $\{i, j\}$  é aresta,

$$\pi_i = \frac{d(i)}{2|E(G)|}.$$

Veremos que esse sempre é o caso, portanto passeios aleatórios em grafos são reversíveis.

## Demonstração do Teorema 13 no caso finito

Vamos começar considerando o caso em que  $P$  é uma matriz estocástica positiva. Digamos que  $P$  é  $n \times n$  e definimos  $\epsilon = \min P > 0$ . Vamos provar que  $\lim_{k \rightarrow \infty} P^k = Q$  com  $Q$  uma matriz estocástica com as colunas constantes.

**Exercício 108** Seja  $P$  uma matriz estocástica com todas entradas positivas. Tomemos  $\epsilon = \min P$  e  $\mathbf{c}$  um vetor coluna qualquer com  $\min \mathbf{c} = a_0$  e  $\max \mathbf{c} = b_0$ . Sejam  $a_1 = \min P\mathbf{c}$  e  $b_1 = \max P\mathbf{c}$ . Prove que  $a_1 \geq a_0$ ,  $b_1 \leq b_0$  e  $b_1 - a_1 \leq (1 - 2\epsilon)(b_0 - a_0)$  (Sugestão: tome  $\mathbf{c}'$  o vetor obtido a partir de  $\mathbf{c}$  trocando todas as coordenadas por  $a_0$  menos a coordenada  $b_0$ ; tome  $\mathbf{c}''$  o vetor obtido a partir de  $\mathbf{c}$  trocando todas as coordenadas por  $b_0$  menos a coordenada  $a_0$ . Use o fato de  $P\mathbf{c}' \leq P\mathbf{c} \leq P\mathbf{c}''$ .)

Denotemos por  $\mathbf{e}_j$  o vetor *coluna* com todas as entradas nulas a menos da posição  $j$  que é 1. A seguir, vamos usar o exercício anterior para definir um vetor estocástico  $\pi$  que são as linhas de  $Q$ , o qual mostraremos depois ser estacionário.

Fixamos  $j \in \{1, 2, \dots, n\}$ , uma coluna de  $P$ , e definimos

$$\begin{aligned} a_k &= \min P^k \mathbf{e}_j \\ b_k &= \max P^k \mathbf{e}_j \end{aligned}$$

o mínimo e o máximo, respectivamente, da  $j$ -ésima coluna de  $P^k$ , tomamos  $a_0 = 0$  e  $b_0 = 1$ . Para  $k \in \mathbb{N}$  temos

$$a_{k+1} = \min P^{k+1} \mathbf{e}_j = \min P(P^k \mathbf{e}_j) \text{ e } a_k = \min(P^k \mathbf{e}_j).$$

Analogamente, temos

$$b_{k+1} = \max P^{k+1} \mathbf{e}_j = \max P(P^k \mathbf{e}_j) \text{ e } b_k = \max(P^k \mathbf{e}_j).$$

Do exercício 108 temos a sequência de máximos e mínimos

$$b_0 \geq b_1 \geq \dots \geq b_k \geq \dots \geq a_k \geq a_{k-1} \geq \dots \geq a_0$$

de modo que

$$b_k - a_k \leq (1 - 2\epsilon)(b_{k-1} - a_{k-1}) \leq (1 - 2\epsilon)^k$$

pois  $\epsilon > 0$ . Da equação anterior  $\lim_{k \rightarrow \infty} b_k - a_k = 0$  e definimos

$$\pi_j = \lim_{k \rightarrow \infty} b_k = \lim_{k \rightarrow \infty} a_k.$$

Com isso provamos que todas as entradas da  $j$ -ésima coluna de  $P^k$  convergem para um valor constante  $\pi_j$ . Feito isso para cada coluna  $j$ , o vetor que procuramos é

$$\pi = (\pi_1, \pi_2, \dots, \pi_n) \quad (65)$$

e com isso

$$\lim_{k \rightarrow \infty} p_{i,j}^{(k)} = \pi_j \quad (\forall i) \quad (66)$$

e tomamos  $Q = \mathbf{1}^T \pi$ , a matriz com cada linha igual a  $\pi$  que queríamos provar. Como  $P > 0$  temos  $a_1 > 0$ , portanto  $\pi_j > 0$ . Também,  $\pi_j < 1$ .

**Exercício 109** Prove que temos  $\sum_j \pi_j = 1$  no vetor dado acima.

Estenderemos esse resultado para matrizes não negativas usando o fato de que para a matriz de uma cadeia de Markov irreduzível e aperiódica temos  $P^t$  positiva para algum  $t \in \mathbb{N}$ .

**Proposição 17** Se uma cadeia Markov é irreduzível e aperiódica então existe  $K_0 \in \mathbb{N}$  tal que para todo  $i, j \in S$ , se  $k \geq K_0$  então  $p_{i,j}^{(k)} > 0$ .

DEMONSTRAÇÃO: Da cadeia ser irreduzível temos que para cada  $i, j \in S$  existe  $k(i, j)$  tal que  $p_{i,j}^{(k(i,j))} > 0$ . Da cadeia ser aperiódica temos que para cada  $i \in S$  existe  $k_0(i)$  tal que  $p_{i,i}^{(k)} > 0$  para todo  $k \geq k_0(i)$ . Com essas constantes, se  $t \geq 0$  então

$$p_{i,i}^{(k_0(i)+t)} > 0 \text{ e } p_{i,j}^{(k(i,j))} > 0$$

portanto, por (52), para todo  $t \in \mathbb{N}$

$$p_{i,j}^{(k(i,j)+k_0(i)+t)} = \sum_{\ell \in S} p_{i,\ell}^{(k_0(i)+t)} p_{\ell,j}^{(k(i,j))} \geq p_{i,i}^{(k_0(i)+t)} p_{i,j}^{(k(i,j))} > 0.$$

Tomamos

$$K_0 = \max\{k(i, j) + k_0(i) \mid i, j \in S\}$$

o que prova a proposição.  $\square$

O vetor  $\pi$  é estacionário

$$\pi_j = \lim_{n \rightarrow \infty} p_{i,j}^{(n)} = \lim_{n \rightarrow \infty} \sum_{k \in S} p_{i,k}^{(n)} p_{k,j} = \sum_{k \in S} \lim_{n \rightarrow \infty} p_{i,k}^{(n)} p_{k,j} = \sum_{k \in S} \pi_k p_{k,j}.$$

Agora, provaremos que se o estado inicial é algum elemento de  $S = \{1, 2, \dots, n\}$  com probabilidade dada por  $\rho = \rho^{(0)}$  então  $\rho P^t \rightarrow \pi$  quando  $t \rightarrow \infty$ , ou seja, qualquer distribuição inicial a longo prazo se aproxima da distribuição estacionária.

Consideremos os vetores de probabilidades

$$\varrho^{(0)} = \begin{pmatrix} q_1^{(0)} & \dots & q_n^{(0)} \end{pmatrix}$$

e

$$\varrho^{(t)} = \begin{pmatrix} q_1^{(t)} & \dots & q_n^{(t)} \end{pmatrix}$$

e a partir de  $\varrho^{(t)}$  obtemos a distribuição  $X_{t+1}$  por

$$\varrho^{(t+1)} = \varrho^{(t)} P = \varrho P^t.$$

Agora, seja  $K_0 = K_0(P)$  a constante dada pela proposição 17. Assim,  $P^{K_0}$  é uma matriz com todas as entradas positivas e para  $\epsilon = \min P^{K_0}$  e para  $t \in \mathbb{N}$

$$b_{tK_0} - a_{tK_0} \leq (1 - \epsilon)^t$$

portanto, a subsequência converge,  $\lim_{t \rightarrow \infty} b_{tK_0} - a_{tK_0} = 0$ . Ademais, as diferenças  $b_t - a_t$  nunca aumentam com  $t$ , ou seja, a seqüência  $(b_t - a_t)_t$  é não-crescente, portanto converge para 0, portanto,  $\lim_{t \rightarrow \infty} P^t = Q$

Seja  $v$  um vetor *qualquer* com  $v = vP$ . Para mostrar a unicidade de  $\pi$  é suficiente mostrar que  $v$  é um múltiplo escalar de  $\pi$ .

Se  $v = vP$  então  $v = vP^t$ , portanto

$$v = \lim_{t \rightarrow \infty} vP^t = v \lim_{t \rightarrow \infty} P^t = vQ$$

logo  $v_j = \sum_i v_i \pi_j = \pi_j \sum_i v_i$  e podemos concluir então que  $v = (\sum_i v_i) \pi$ .

Agora, vamos provar que  $\lim_{t \rightarrow \infty} \varrho P^t = \pi$  para qualquer vetor de probabilidades  $\varrho = (q_1, q_2, \dots, q_n)$ .

Como acima,

$$\lim_{t \rightarrow \infty} \varrho P^t = \varrho Q$$

donde  $(\varrho Q)_j = (\sum_i q_i) \pi_j$  é a  $j$ -ésima coordenada de  $\varrho Q$ . Como  $\varrho$  é um vetor de probabilidades a soma das coordenadas é 1, logo

$$(\varrho Q)_j = \pi_j$$

ou seja,  $\varrho Q = \pi$  como queríamos provar.

Resta provar que  $\pi_j = 1/\mu_j$ . Como  $S$  é finito  $\mu_j < \infty$ .

Primeiro, definimos  $\mu_{i,j} = \mathbb{E}(T_j \mid X_0 = i)$  para  $i \neq j$ , e se consideramos o primeiro passo da cadeia temos que vale

$$\mu_{i,j} = p_{i,j} + \sum_{k \neq j} p_{i,k} (\mu_{k,j} + 1) = 1 + \sum_{k \neq j} p_{i,k} \mu_{k,j}.$$

De modo análogo  $\mu_{i,i} = \mu_i = 1 + \sum_k p_{i,k} \mu_{k,j}$ . Esse conjunto de equações pode ser escrito de forma matricial como

$$M = PM + U - D$$

em que  $U$  é a matriz com todas as entradas iguais a 1 e  $D = (d_{i,j})$  é a matriz diagonal  $d_{i,i} = \mu_i$ . Reescrevendo temos

$$D - U = (P - Id)M$$

em que  $Id$  é a matriz identidade. Se multiplicarmos por  $\pi$  a direita nos dois lados da igualdade temos, no lado direito, que  $\pi(P - Id) = 0$  pois  $\pi$  é estacionário, portanto  $\pi D = \pi U$ , mas  $\pi U = (1 \ 1 \ \cdots \ 1)$ , logo  $(\pi D)_j = 1$  para cada  $j$ , isto é,

$$\pi_j \mu_j = 1.$$

## §12 Passeio aleatório em grafos

Seja  $G = (V, E)$  um grafo finito e conexo. O leitor não familiarizado com as nomenclaturas elementares da Teoria dos Grafos pode consultar estas notas.

Um **passeio aleatório** em  $G$  é uma sequência  $v_0, v_1, v_2, \dots$  de vértices de  $V$  de modo que  $v_{i+1}$  é escolhido uniformemente em  $N(v_i) = \{u \in V : \{u, v_i\} \in E\}$ , a vizinhança de  $v_i$ , para todo  $i \in \mathbb{N}$ . Dizendo de outro modo, é uma cadeia de Markov  $\{X_t\}_{t \geq 0}$  homogênea com  $V$  como conjunto de estados e

$$p_{v,u} = \mathbb{P}(X_{t+1} = u \mid X_t = v) = \frac{1}{d(v)}$$

para todo  $u \in N(v)$  e todo  $t \in \mathbb{N}$ , onde  $d(v) = |N(v)|$  é o grau do vértice  $v$ .

Se  $A = A(G)$  é a matriz de adjacências de  $G$ , isto é  $A = (a_{v,u})$  com  $a_{v,u} \in \{0, 1\}$  e  $a_{v,u} = 1$  sse e só se  $\{v, u\}$  é aresta, e  $D = (d_{u,v})$  é a matriz diagonal  $d_{v,v} = 1/d(v)$  então a matriz de transições do passeio aleatório em  $G$  é

$$P = AD.$$

**Exemplo 129** Seja  $G$  o grafo sobre o conjunto de vértice  $\{1, 2, \dots, n\}$  com todas as  $\binom{n}{2}$  arestas, chamado de grafo completo, e consideremos um passeio aleatório em  $G$ , assim

$$p_{v,u} = \frac{1}{n-1}$$

e a matriz de transição é  $P = \frac{1}{n-1} A$ .

Nesse caso, o número esperado de passos para atingir  $v$  a partir de  $u$  é  $n-1$ . De fato, a probabilidade de atingir  $v$  em 1 passo é  $f_{u,v}^{(1)} = 1/(n-1)$ , a probabilidade de atingir  $v$  em 2 passos é  $f_{u,v}^{(2)} = (n-2)/(n-1)^2$ , em 3 passos  $f_{u,v}^{(3)} = (n-2)^2/(n-1)^3$ , e assim por diante.

O número esperado de passos é

$$\sum_k k f_{u,v}^{(k)} = \sum_{k \geq 1} k \frac{1}{n-1} \left( \frac{n-2}{n-1} \right)^{k-1} = \frac{1}{n-1} \frac{1}{\left(1 - \frac{n-2}{n-1}\right)^2} = n-1.$$

O número esperado de passos para visitar todos os vértices do grafo pode ser estimado da seguinte forma. Seja  $t_i$  o instante em que pela primeira vez temos exatamente  $i$  vértices visitados, portanto,  $t_{i+1} - t_i$  é uma variável aleatória geométrica que conta o número de passos enquanto espera-se para conhecer um novo vértice, evento que ocorre com probabilidade  $(n-i)/(n-1)$ , logo  $\mathbb{E}(t_{i+1} - t_i) = (n-1)/(n-i)$  e  $t_n$  é o número de passos até visitar todos os vértices

$$\mathbb{E}(t_n) = \sum_{i=1}^{n-1} \mathbb{E}(t_{i+1} - t_i) = \sum_{i=1}^{n-1} \frac{n-1}{n-i} = (n-1) \sum_{i=1}^{n-1} \frac{1}{i} = (n-1) H_{n-1}$$

onde  $H_n$  denota o  $n$ -ésimo número harmônico,  $H_n = \sum_{i=1}^n 1/i = \ln n + \gamma + \Theta(n^{-1})$ , onde  $\gamma \approx 0,577$  é a constante de Euler–Mascheroni.

Um passeio aleatório num grafo conexo é uma cadeia de Markov irredutível e em certos casos aperiódica. O próximo resultado caracteriza tais passeios aleatórios aperiódicos.

**Teorema 15** *Se  $G$  é conexo com pelo menos dois vértices então um passeio aleatório em  $G$  define uma cadeia de Markov irredutível. Ainda, tal cadeia de Markov é periódica se e só se  $G$  é bipartido.*

DEMONSTRAÇÃO: Se  $G$  é conexo então para dois vértices  $u$  e  $v$  quaisquer que estão a distância  $k$  vale  $p_{u,v}^{(k)} > 0$ . Como  $G$  é conexo a distância entre quaisquer dois vértices é algum  $k \in \mathbb{N}$ , portanto a cadeia de Markov definida pelo passeio aleatório é irredutível.

Seja  $G$  um grafo conexo e bipartido com bipartição  $\{A, B\}$ . Então, todos os passeios de  $u$  para  $v$  têm a mesma paridade no número de arestas, para o caso  $u, v \in A$  ou  $u, v \in B$  e ímpar caso  $u \in A$  e  $v \in B$  ou  $v \in A$  e  $u \in B$ . Logo

- se  $u, v \in A$  ou  $u, v \in B$  então  $p_{u,v}^{(k)} > 0$  exceto quando  $k$  é ímpar,
- se  $u \in A$  e  $v \in B$ , ou  $v \in A$  e  $u \in B$ , então  $p_{u,v}^{(k)} > 0$  exceto quando  $k$  é par.

Portanto a cadeia tem período 2.

Agora, suponhamos  $G$  não-bipartido, então contém um circuito de comprimento ímpar  $C$ . Tomemos  $P$  um passeio de  $u$  a  $v$ , vértices que distam  $k$  em  $G$ , com o menor número de arestas. Um passeio de  $u$  para  $v$  com  $k + 2r$  arestas existe para todo  $r \in \mathbb{N}$ , basta repetir  $r$  vezes alguma aresta do passeio  $P$ . Como  $G$  é conexo existe um passeio de algum vértice de  $P$  até algum vértice de  $C$ , portanto, podemos usar as arestas de  $C$  para obter passeios de  $u$  para  $v$  que têm a paridade oposta a de  $k$ . Logo,  $p_{u,v}^{(t)} > 0$  para todo  $t$  suficientemente grande, ou seja, a cadeia é aperiódica.  $\square$

**Corolário 18** *Se  $G = (V, E)$  é conexo, não-bipartido e com pelo menos dois vértices então a cadeia de Markov dada por um passeio aleatório em  $G$  admite um vetor estacionário. Ademais, a distribuição estacionária é única e dada por*

$$\pi = (\pi_v)_{v \in V}, \text{ com } \pi_v = \frac{d(v)}{2|E|}. \quad (67)$$

DEMONSTRAÇÃO: A cadeia é irredutível e aperiódica, portanto admite um único vetor estacionário pelo teorema 13. Agora, basta verificar que (67) é estacionário.

A soma dos graus dos vértices de um grafo é  $2|E|$ , logo  $\sum_{v \in V} \pi_v = 1$ . Resta verificar que  $\pi = \pi P$ , onde  $P$  é a matriz de transição, mas

$$(\pi P)_v = \sum_{u \in V} \pi_u p_{u,v} = \sum_{u \in N(v)} \frac{d(u)}{2|E|} \frac{1}{d(u)} = \sum_{u \in N(v)} \frac{1}{2|E|} = \frac{d(v)}{2|E|} = \pi_v$$

portanto, o vetor dado em (67) é o vetor estacionário.  $\square$

**Observação 19** Se  $G = (V, E)$  é bipartido então podemos contornar essa propriedade indesejada (no sentido do corolário acima) acrescentando laços aos vértices do grafo com probabilidade de transição  $1/2$  (ou seja,  $p_{v,v} = 1/2$ ) e dividir por 2 a probabilidade das outras arestas, ou seja, se  $P$  é a matriz de transição da cadeia no grafo original, a nova matriz de transição é

$$Q = \frac{P + Id}{2}$$

onde  $Id$  é a matriz identidade  $|V| \times |V|$ . Essa transformação apenas “reduz a velocidade” do passeio.

**Exemplo 130** De volta ao grafo completo  $G$  do exemplo 129, vamos usar a estratégia de acrescentar laço e o número esperado de passos até um passeio aleatório passar por todos os vértices de  $G$ . Seja  $P$  a matriz de transição da cadeia de Markov do passeio em  $G$  e consideremos a modificação acima, acrescentando um laço em cada vértice ficamos com a matriz de transição do novo passeio

$$Q = \frac{P + Id}{2}.$$

Seja  $T$  o número de passos para o passeio definido por  $Q$  visitar todos os vértices e  $q_i = (n - i)/n$  a probabilidade de visitar um vértice novo se outros  $i$  vértices já foram visitados, então o número esperado de passos até o passeio visitar o  $i + 1$ -ésimo vértice é  $1/q_i$ . Assim,

$$\mathbb{E}(T) = \sum_{i=1}^n \frac{n}{i} = nH_n.$$

### $s - t$ conexidade em grafos

O problema da  $s - t$  conexidade em grafos é: dado um grafo  $G = (V, E)$  e dois vértices  $s, t$  em  $G$  decidir se há um passeio entre esses dois vértices. Esse problema pode ser resolvido em tempo linear no tamanho do grafo,  $|V| + |E|$ , e usando espaço  $\Omega(|V|)$ .

A pergunta que interessa nesse caso é se o problema pode ser resolvido usando espaço logarítmico. Essa pergunta foi respondida afirmativamente em [Reingold, *Undirected connectivity in log-space*, J. ACM 58, 2008], até então o único progresso significativo foi o algoritmo aleatorizado que apresentaremos a seguir devido a [Aleliunas, Karp, Lipton, Lovasz, Rackoff, *Random walks, universal traversal sequences, and the complexity of maze problems*, SFCS '79].

**Dado** : grafo  $G$  com  $n \geq 2$  vértices e dois vértices  $s$  e  $t$ .

**Devolve**: *sim*, se há  $s - t$  passeio em  $G$ , ou *não* caso não haja ou há mas não é encontrado, o que ocorre com probabilidade  $4/n$ .

$v \leftarrow s$ ;

**repita**

**se**  $v = t$  **então responda** *sim*;

  escolha  $u \in N(v)$  uniformemente;

$v \leftarrow u$ ;

**até complete**  $n^4$  rodadas;

**responda** *não*.

**Algoritmo 2:**  $s - t$ -conexidade.



Notemos que o espaço extra gasto pelo algoritmo é para manter o vértice atual e o contador de rodadas para o laço, portanto, o espaço utilizado é  $O(\log n)$  para um grafo com  $n$  vértices.

Se o algoritmo devolve *sim* então o grafo contém um passeio entre os vértices  $s$  e  $t$ . Agora, se o algoritmo responde *não*, então ou não há caminho, ou o algoritmo não foi capaz de encontrá-lo em tempo. Portanto, se não existe caminho a resposta está correta, caso contrário a resposta pode estar errada e vamos limitar a probabilidade de erro.

Seja  $T$  o tempo para um passeio aleatório no grafo  $G$  visitar todos os vértices de  $G$ . O **tempo de cobertura** de  $G$  (*cover time*) é

$$\max_{v \in V(G)} \mathbb{E}(T \mid X_0 = v).$$

**Lema 20** *O tempo de cobertura de um grafo  $G = (V, E)$  conexo é no máximo  $4|V||E|$ .*

DEMONSTRAÇÃO: Tomemos um subgrafo acíclico maximal de  $G = (V, E)$  (ou, uma árvore geradora) e consideremos um passeio  $W = v_0, v_1, \dots, v_{2|V|-2} = v_0$  nas  $|V| - 1$  arestas dessa árvore de modo que cada aresta seja percorrida exatamente duas vezes, uma vez em cada direção. O tempo de cobertura de  $G$  é limitado superiormente pelo tempo esperado para um passeio percorrer a sequência  $W$  que é

$$\sum_{i=0}^{2|V|-3} \mathbb{E}(T_{v_{i+1}} \mid X_0 = v_i) < (2|V| - 2)(2|E|) < 4|V||E|$$

pois  $\mathbb{E}(T_{v_{i+1}} \mid X_0 = v_i) < 2|E|$  para toda aresta  $\{v_i, v_{i+1}\}$  em  $G$ . De fato, calculando  $\mu_v$  de dois modos distintos

$$\frac{2|E|}{d(v)} = \frac{1}{\pi_v} = \mu_v = \sum_{u \in N(v)} (1 + \mathbb{E}(T_v \mid X_0 = u)) \frac{1}{d(u)}$$

portanto,  $2|E| = \sum_{u \in N(v)} (1 + \mathbb{E}(T_v \mid X_0 = u))$ .  $\square$

Com isso conseguimos estimar a probabilidade de erro da seguinte maneira. Seja  $G$  um grafo e  $s$  e  $t$  dois vértices de  $G$  que são ligados por um passeio. O algoritmo erra se em  $n^4$  rodadas não consegue achar um  $s - t$  caminho em  $G$ . Denotemos por  $C(G)$  o tempo de cobertura de  $G$ , usando a desigualdade de Markov

$$\mathbb{P}(\text{erro}) = \sum_{k > n^4} f_{s,t}^{(k)} \leq \frac{\mathbb{E}(T_t \mid X_0 = s)}{n^4}$$

logo

$$\mathbb{P}(\text{erro}) \leq \frac{C(G)}{n^4} \leq \frac{4n|E|}{n^4} < \frac{4}{n}.$$

## Passeio aleatório em grafos regulares

Um grafo é regular se todos os vértices têm o mesmo grau. Nesse caso temos um bom tanto de informação que podemos retirar da matriz de transição que é duplamente estocástica (veja o exercício 103). A começar, que essa matriz é simétrica, portanto, todos os seus autovalores são reais, ademais, o Teorema Espectral garante a existência de uma família ortonormal de autovetores, tantos quanto é a dimensão da matriz. Um passeio aleatório num grafo regular converge

para a distribuição uniforme (exercício 103) e nesta seção mostraremos que a velocidade dessa convergência é ditada pelo segundo maior autovalor da matriz de transição.

Um grafo  $G = (V, E)$  é  $d$ -regular,  $d \in \mathbb{N}$ , se todos os vértices de  $V$  têm grau  $d$ . No que segue, vamos assumir que todo grafo é sobre  $V = \{1, 2, \dots, n\}$ , é conexo e  $d$ -regular.

Nesse caso, a matriz de transição de um passeio em  $G$  é

$$P = \frac{1}{d} A$$

para  $A = A(G)$  a matriz de adjacências. A matriz  $P$  é simétrica, portanto, seus autovalores são reais. Sejam

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \quad (68)$$

os autovalores de  $P$ . Do Teorema de Perron–Frobenius temos que se  $G$  é conexo então

- $\lambda_1 > \lambda_2$ ;
- $\lambda_1 = -\lambda_n$  se e somente se  $G$  é bipartido.

Não é difícil constatar que o vetor  $\pi$  com todas as coordenadas iguais a  $1/|V|$  é um autovetor de  $P$  associado ao autovalor 1, ou seja,  $P\pi = 1\pi$  e como  $P$  é simétrica

$$\pi = \pi P$$

que, em outras palavras, quer dizer que a distribuição uniforme sobre  $V$  é a distribuição estacionária do passeio aleatório em  $G$ , como já havíamos dito.

Ainda, se  $k$  é tal que  $|\pi_k| = \max_v |\pi_v|$  então

$$|\lambda_1 \pi_k| = \left| \sum_{v=1}^n \pi_v p_{v,k} \right| \leq \sum_{v=1}^n |\pi_v| |p_{v,k}| \leq |\pi_k| \sum_{v=1}^n |p_{v,k}| = |\pi_k|$$

ou seja,  $\lambda_1 \leq 1$ , portanto  $\lambda_1 = 1$  já que 1 é autovalor.

Seja  $G$  um grafo conexo, não-bipartido,  $d$ -regular e com  $n$  vértices. Fixemos

$$\lambda = \max\{|\lambda_2|, |\lambda_n|\} \quad (69)$$

e  $\varrho = \varrho^{(0)}$  uma distribuição inicial sobre  $V$ , seja  $\varrho^{(t)}$  a distribuição de probabilidade sobre os vértices de  $G$  no instante  $t \in \mathbb{N}$ , isto é, a distribuição de  $X_t$ . Vamos mostrar que  $\varrho^{(t)}$  converge para  $\pi$  com velocidade controlada por  $\lambda$ .

Seja  $\xi_i$ ,  $1 \leq i \leq n$ , uma base ortonormal de autovetores de  $P$ , de modo que  $\xi_i$  é autovetor associado ao autovalor  $\lambda_i$ . Notemos que  $\xi_1 = \sqrt{n}\pi$ .

Se  $\varrho = a_1 \xi_1 + \dots + a_n \xi_n$  então  $\varrho^{(t)} = P^t \varrho = a_1 \mathbf{v}_1 P^t + \dots + a_n \mathbf{v}_n P^t = a_1 \lambda_1^t \xi_1 + \dots + a_n \lambda_n^t \xi_n$  e

$$\|\varrho^{(t)} - a_1 \xi_1\|_2^2 = (a_1 \lambda_1^t - a_1)^2 + \sum_{i>1} (a_i \lambda_i^t)^2 \leq \lambda^{2t} \sum_{i>1} a_i^2 \leq \lambda^{2t} \|\varrho^{(0)}\|_2^2$$

donde segue que

$$\|\varrho^{(t)} - a_1 \xi_1\|_2 \leq \lambda^t \|\varrho^{(0)}\|_2 \leq \lambda^t \|\varrho^{(0)}\|_1 = \lambda^t$$

e como  $\lambda < 1$  temos  $\varrho^{(t)}$  converge para  $a_1 \xi_1$  quando  $t \rightarrow \infty$ , logo  $a_1 \xi_1$  é igual a  $\pi$ , pelo teorema 13.

Com isso, provamos

**Lema 21** *Sejam  $G$  um grafo com  $n$  vértices, conexo,  $d$ -regular e não-bipartido,  $P = A(G)/d$  a matriz de transição de um passeio aleatório em  $G$  e*

$$\lambda = \max \{|\alpha| : \alpha \text{ é autovalor de } P \text{ e } \alpha < 1\}.$$

*Então para todo vetor de probabilidades  $\varrho$  e todo  $t \in \mathbb{N}$*

$$\|\varrho^{(t)} - \pi\|_2 \leq \lambda^t$$

*onde  $\pi = (1/n, 1/n, \dots, 1/n)$  é o vetor estacionário do passeio aleatório.*

**Exercício 110** *Seja  $\rho = (p_1, \dots, p_n) \in \mathbb{R}^n$  um vetor, então  $\|\rho\|_\infty$  é uma norma dada por*

$$\|\rho\|_\infty = \max \{|p_i| : 1 \leq i \leq n\}.$$

*Prove que  $\|\rho\|_\infty \leq \|\rho\|_2$  para todo  $\rho \in \mathbb{R}^n$ .*

**Teorema 16** *Seja  $\{X_t\}_{t \in \mathbb{N}}$  um passeio aleatório num grafo  $G$  com  $n$  vértices, conexo,  $d$ -regular e não-bipartido e com segundo maior autovalor  $\lambda$ . Então, para todo  $\delta > 0$  existe  $k = k(\lambda, \delta)$  tal que*

$$\left| \mathbb{P}(X_k = v) - \frac{1}{n} \right| < \delta$$

*para todo vértice  $v$ .*

DEMONSTRAÇÃO: Seja  $G$  como no enunciado. Vamos mostrar que

$$\|\varrho A^k - \pi\|_\infty = \left| \mathbb{P}(X_k = v) - \frac{1}{n} \right| < \delta.$$

Seja  $\varrho = \varrho^{(0)}$  uma distribuição inicial qualquer, então a distribuição de  $X_k$  é dada pelo vetor  $\varrho P^k$ ,

$$\mathbb{P}(X_k = v) = (\varrho P^k)_v$$

em que  $P$  é a matriz de transição do passeio aleatório.

Pelo lema 21 e exercício acima

$$\|\varrho A^k - \pi\|_\infty \leq \|\varrho A^k - \pi\|_2 \leq \lambda^{k+1}$$

portanto, se  $k > \log_\lambda(\delta/\lambda)$  então  $\|\varrho A^k - \pi\|_\infty < \delta$ .  $\square$

**Exercício 111** *Verifique se vale o resultado anterior para  $G$  bipartido, conexo e  $d$ -regular com o passeio aleatório dado por  $Q = (P + \text{Id})/2$ , onde  $P = A(G)/d$ .*

**Exercício 112** Sejam  $G$  um grafo com  $n$  vértices, conexo,  $d$ -regular e  $\mu_1 \geq \mu_2 \geq \dots \geq \mu_n$  os autovalores da matriz de adjacências de  $G$ . Consideremos as matrizes estocásticas  $P = A(G)/d$  e  $Q = (P + \text{Id})/2$  cujos autovalores são, respectivamente,  $\lambda_1 \geq \lambda_2 \geq \dots, \lambda_n$  e  $\nu_1 \geq \nu_2 \geq \dots \geq \nu_n$ . Prove que

$$\lambda_i = \frac{\mu_i}{d} = 2\nu_i - 1$$

para todo inteiro  $1 \leq i \leq n$ .

**Exercício 113** Sejam  $G$  um grafo com  $n$  vértices,  $m$  arestas e  $\mu_1 \geq \mu_2 \geq \dots \geq \mu_n$  os autovalores da matriz de adjacências de  $G$ . Prove que

$$\sum_{i=1}^n \mu_i^2 = 2m. \quad (70)$$

(Sugestão:  $\sum_i \mu_i^2$  é a soma da diagonal principal da matriz  $A^2$  e  $A$  é diagonalizável.)

## Passeios aleatórios em grafos expansores

Da discussão anterior podemos concluir que quanto menor for o segundo autovalor de um grafo, que denotamos por  $\lambda$ , mais rápido um passeio aleatório converge para a distribuição uniforme, entretanto  $\lambda$  não pode ser arbitrariamente pequeno. De (70) temos

$$(n-1)\lambda^2 + \lambda_1^2 \geq 2m$$

portanto, no caso  $d$ -regular  $\lambda = \Omega(\sqrt{d})$ .

No que segue,  $G^n$  é um grafo com  $n$  vértices, conexo e  $d$ -regular para  $d$  fixo. Dizemos que  $G^n$  é um grafo  $\epsilon$ -expansor se  $\lambda \leq d - \epsilon$ .

Seja  $W$  um subconjunto de vértices de  $G^n$ . Denotamos por  $N(W)$  o subconjunto dos vértices de  $V \setminus W$  que são adjacentes a algum vértice de  $W$ . Dizemos que  $G$  é  $c$ -vértice expansor se para todo  $W$  com  $|W| \leq n/2$  vale que  $|N(W)| \geq c|W|$ , daí vem o adjetivo *expansor*. Essas definições são equivalentes no seguinte sentido: se  $G^n$  é  $\epsilon$ -expansor então é  $\epsilon/(2d)$ -vértice-expansor. Por outro lado, se  $G^n$  é  $c$ -vértice-expansor então é  $c^2/(4+2c^2)$ -expansor. Logo, grafos expansores são caracterizados combinatorialmente por possuir alta conexidade, que é equivalente a dizer que a *distância espectral* (diferença  $d - \lambda$  entre os dois maiores autovalores da matriz de adjacências do grafo) é grande.

Grafos expansores *esparcos* são objetos aparentemente contraditórios, mas a existência desses grafos segue de métodos probabilísticos usuais, como mostrou Pinsker em 1973 que quase todo grafo  $d$ -regular,  $d \geq 3$ , é expansor. Embora abundantes, a construção explícita desses grafos não é simples; de fato, em geral a construção e a prova da expansão usam ferramentas profundas e complexas da matemática. Sortear um grafo e testar se é expansor é inviável, o sorteio usa muitos bits aleatórios (uma das principais aplicações desses grafos que veremos aqui é o uso para economizar bits aleatórios em algoritmos aleatorizados) e o problema de decisão está em coNP.

No que segue vamos assumir a possibilidade de construir explicitamente grafos expansores tais que dados um vértice  $x$  e  $i \in \{1, 2, \dots, d\}$  o vértice  $v_i(x)$  pode ser determinado em espaço logarítmico. Para as construções explícitas o leitor pode consultar [este documento](#). Uma ótima referência sobre grafos expansores e aplicações é

O nosso interesse agora é no uso desses grafos para desaleatorização em BPP que segue a seguinte estratégia. Suponha que temos um algoritmo que decide  $L \in \text{BPP}$  com probabilidade de erro menor que  $\epsilon$  usando  $n$  bits aleatórios para entradas de tamanho fixo. Então, podemos construir um algoritmo que decide a mesma linguagem com probabilidade de erro menor que  $\epsilon^k$  usando  $kn$  bits tomando  $k$  rodadas independentes do algoritmo. Agora, suponha que temos um grafo expensor 3-regular com  $2^n$  vértices. Um gerador pseudoaleatório consiste em determinar um vértice desse grafo uniformemente, de modo que temos os  $n$  bits aleatórios necessários para uma rodada do algoritmo. Para determinar um desses vértices usamos o lema 21, que diz que é suficiente começamos um passeio aleatório de comprimento  $\approx n$ , cada passo do passeio precisa de 2 bits genuinamente aleatórios. Como o algoritmo é executado  $k$  vezes, o número de bits aleatórios usados é  $O(n + k)$ .

Antes de expormos formalmente a idéia acima chamamos a atenção para um resultado conhecido estreitamente relacionado com os resultados discutidos nessa seção.

### ***Expander mixing lemma***

É sabido que se  $X, Y \subseteq V(G^n)$  então a distribuição de arestas entre  $X$  e  $Y$  em  $G^n$  satisfaz

$$|E(X, Y) - \frac{d}{n}|X||Y|| \leq \lambda \sqrt{|X||Y|} \quad (71)$$

DEMONSTRAÇÃO:[Esboço de prova de (71)] Seja  $\xi_i$ ,  $1 \leq i \leq n$ , uma base ortonormal de autovetores de  $A$ . Sejam  $\mathbf{x}$  e  $\mathbf{y}$  os vetores (linha) característicos de  $X$  e  $Y$ , respectivamente. Então  $|E(X, Y)| = \mathbf{x}A\mathbf{y}^T$ , onde  $T$  denota o transposto. Escrevendo na base de autovetores  $|E(X, Y)| = \sum_i \lambda_i \alpha_i \beta_i$ , onde  $\alpha_i$  e  $\beta_i$  são os produtos escalares  $\langle \mathbf{x}, \xi_i \rangle$  e  $\langle \mathbf{y}, \xi_i \rangle$ , respectivamente. O resultado segue de  $\sum_i \lambda_i \alpha_i \beta_i = d|X||Y|/n + \sum_{i \neq 1} \lambda_i \alpha_i \beta_i$ .  $\square$

Portanto,  $\lambda$  pequeno garante uma distribuição de arestas como num grafo aleatório com densidade de arestas  $d/n$ . Grafos com  $\lambda$  pequeno são ditos pseudoaleatórios. Linial e Bilu provaram uma recíproca desse resultado, se 71 vale para algum  $\lambda > 0$  então segundo maior autovalor de  $A(G)$  é  $O(\lambda \log(d/\lambda))$ .

De (71) temos

$$\left| \frac{|E(X, Y)|}{nd} - \frac{|X||Y|}{n^2} \right| \leq \frac{\lambda}{d} \quad (72)$$

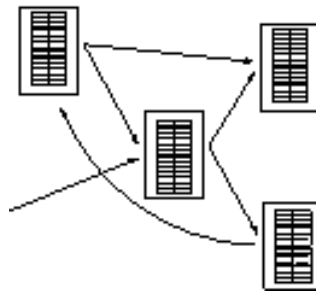
ou seja, a probabilidade de sortear um par de vértices e cair em  $X \times Y$  é próxima a probabilidade de sortear uma aresta e cair em  $E(X, Y)$ . Escrevendo de outro modo, consideremos a vizinhança de cada vértice de  $G$  indexada por  $\{1, 2, \dots, d\}$  e denotemos por  $v_i(x)$  o  $i$ -ésimo vizinho do vértice  $x$  de acordo com essa indexação. Assim, a equação (71) pode ser interpretada como

$$\left| \mathbb{P}_{(x,i) \in \mathbb{R} V \times [d]}([x \in X] \cap [v_i(x) \in Y]) - \mathbb{P}_{(x,y) \in \mathbb{R} X \times Y}([x \in X] \cap [y \in Y]) \right| \leq \frac{\lambda}{d}.$$

## Passeio aleatório na WEB: o Google PageRank

*“To test the utility of PageRank for search, we built a web search engine called Google”*

O grafo *web* é um grafo dirigido definido pelas páginas *web* e as ligações (*links* ou *hyperlinks*) entre as páginas. Conhecer a estrutura desse grafo é importante para o desenvolvimento de algoritmos eficientes para a *web*, por exemplo.



Um marco no projeto e desenvolvimento desses algoritmos é o algoritmo *PageRank* [Page, Brin, Motwani, Winograd, The PageRank Citation Ranking: Bringing Order to the Web] desenvolvido como parte da ferramenta de busca na *web* batizada *Google* pelos fundadores da empresa com mesmo nome.

O *PageRank* é um algoritmo para classificação (*ranking*) de páginas na *web*. A idéia que o motivou é modelar a importância relativa de uma página de modo que uma busca resulte em resultados com relevância. A própria empresa explica a idéia da seguinte maneira:

“O coração do nosso software é o PageRank(TM), um sistema para dar notas para páginas na web, desenvolvido pelos nossos fundadores Larry Page e Sergey Brin na Universidade de Stanford. E enquanto nós temos dúzias de engenheiros trabalhando para melhorar todos os aspectos do Google no dia a dia, PageRank continua a ser a base para todas nossas ferramentas de busca na web.

### Explicações sobre o PageRank

A classificação das páginas (PageRank) confia na natureza excepcionalmente democrática da Web, usando sua vasta estrutura de links como um indicador do valor de uma página individual. Essencialmente, o Google interpreta um link da página A para a página B como um voto da página A para a página B. Mas o Google olha além do volume de votos, ou links, que uma página recebe; analisa também a página que dá o voto. Os votos dados por páginas "importantes" pesam mais e ajudam a tornar outras páginas "importantes."

([http://www.google.com.br/why\\_use.html](http://www.google.com.br/why_use.html))

Assim, uma página tem uma classificação alta se é referenciada por páginas com classificação alta.

O que nos interessa no momento é que o modelo adotado no *PageRank* pode ser interpretado como um passeio aleatório no grafo *web*: um internauta absorto, começa a navegar na *web* a partir de uma página qualquer, e segue a navegação por um dos *links* da página atual escolhido uniformemente; depois de muito tempo nessa tarefa as páginas começam a repetir e o internauta entediado pára o processo e recomeça-o a partir de alguma outra página.

O modelo simplificado do *PageRank* é descrito da seguinte maneira. Seja  $G = (V, E)$  o grafo da *web*, ou seja,  $V$  é o conjuntos formado pelas páginas *web*, as quais serão consideradas sem perda de generalidade  $\{1, 2, \dots, n\} = V$ , e  $(a, b) \in E$  se na página  $a$  há um *link* para a página  $b$ . Denotamos por  $N^+(a)$  o conjunto dos vértices  $b$  tais que  $(a, b)$  é uma aresta de  $G$ , e por  $N^-(a)$  denotamos o conjunto dos vértices  $b$  tais que  $(b, a)$  é uma aresta de  $G$ . No grafo *web* da figura 1  $|N^+(3)| = 2$  e  $|N^-(3)| = 3$ .

A classificação é dada por um vetor  $\mathbf{r}$  onde  $r_a$  é a classificação da página  $a$  e satisfaz

$$r_a = \sum_{b \in N^-(a)} \frac{r_b}{|N^+(b)|} \quad (73)$$

ou seja, se  $b$  aponta para  $a$  então  $b$  contribui com  $1/|N^+(b)|$  de sua relevância para a relevância de  $a$ . Seja  $P$  a matriz

$$p_{a,b} = \begin{cases} \frac{1}{|N^+(a)|} & \text{se } (a, b) \text{ é aresta} \\ 0 & \text{caso contrário,} \end{cases}$$

então de (73)

$$\mathbf{r} = \mathbf{r}P$$

ou seja,  $\mathbf{r}$  é um autovetor à esquerda de  $P$ . No grafo *web* da figura 1 está descrita a matriz de transição  $P$  para aquele grafo.

Se  $P$  for uma matriz estocástica então o vetor  $\mathbf{r}$  é um vetor estacionário e dessa forma,  $\mathbf{r}$  pode ser calculado escolhendo uma distribuição inicial  $\mathbf{r}^{(0)}$  e fazendo  $\mathbf{r}^{(k+1)} = \mathbf{r}^{(k)}P$ . Como vimos, sob certas hipóteses  $\mathbf{r}^{(k)}$  converge para  $\mathbf{r}$ . Esse método, conhecido como método das potências, é usado há muito tempo para calcular autovetores associado ao maior autovalor. Entretanto, na atual situação não sabemos se o vetor converge pois não temos garantia que a matriz  $P$  seja irreduzível (garante  $\lambda_1 > \lambda_2$ , portanto convergência) ou estocástica (garante o autovalor  $\lambda_1 = 1$ , portanto o método converge para o vetor estacionário).

**Exemplo 131** Por exemplo, para

$$P = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

e  $\mathbf{r}^{(0)} = (0, 1)$  temos  $\mathbf{r}^{(2)} = \mathbf{r} = (0, 0)$ , No caso de um circuito dirigido, um vetor inicial pode não convergir.

No exemplo da figura 1 abaixo a matriz  $P$  não é estocástica. Os vértices sem arestas que saem, ou seja os vértices  $v$  tais que  $|N^+(v)| = 0$  são chamados de pendentes (*dangling*) e por causa deles a matriz não é estocástica. No exemplo da figura 1 o vértice 6 é pendente. Na *web* de fato há vários vértices pendentes, por exemplo, os documentos em pdf disponíveis em páginas *web* são vértices pendentes do grafo.

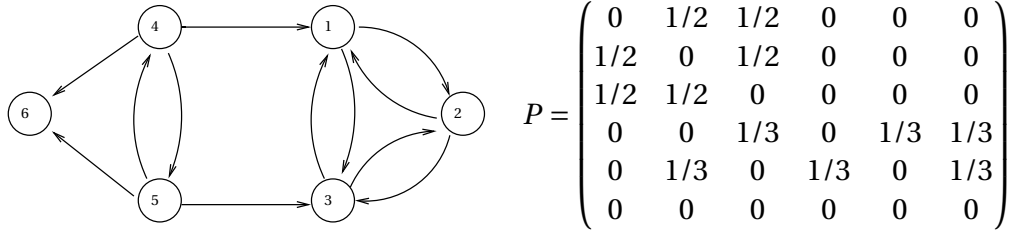


Figura 1: Exemplo de grafo *web*.

Seja  $n$  o número de vértices em  $G$ . Definimos uma matriz auxiliar  $A$  pondo para cada vértice pendente  $a$ , o que corresponde a uma página sem *links*, a linha  $a$  com entradas  $1/n$  e tomamos

$$Q = P + A$$

isso significa que um passeio aleatório que chega numa página sem saída continua em qualquer outra página com igual probabilidade.

A matriz  $Q$  é estocástica. A matriz  $Q$  referente a matriz  $P$  da figura 1 é dada a seguir na figura 2 ao lado do grafo *web* que corresponde à modificação no grafo que reflete a modificação na matriz  $P$ . No grafo da figura 2 as arestas tracejadas são as arestas incluídas e correspondem à possibilidade de navegação do internauta que chega a uma página sem saída e recomeça a navegação de qualquer lugar uniformemente.

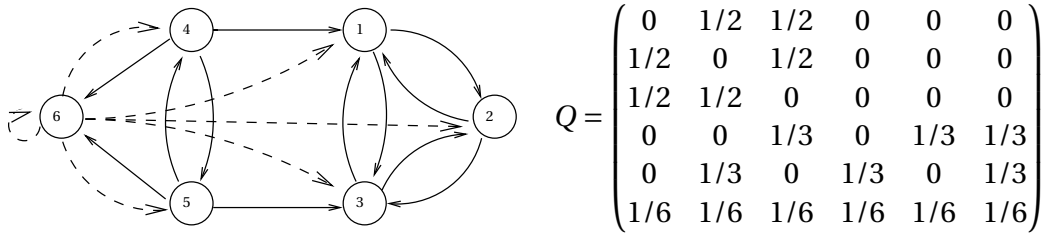


Figura 2: Modelo de grafo *web* modificado, sem vértices pendentes. As arestas tracejadas são as arestas incluídas artificialmente.

No exemplo dado na figura 2, o maior autovalor da matriz  $Q$  é 1 com multiplicidade 1 e autovetor associado  $(1, 1, 1, 0, 0, 0)$ . Notemos que os vértices 4, 5, 6 tem classificação 0. Isso decorre do fato de não haver aresta que sai de  $\{1, 2, 3\}$  e chega em qualquer outro vértice diferente desses. Esse conjunto é chamado de sorvedouro (*rank sink*).

Para lidar com esses sorvedouros basta garantir que a matriz seja irredutível pois se  $p_{a,b}^{(k)} > 0$  para algum  $k$ , para todo  $a, b \in V$  então não há sorvedouros no grafo. Para garantir uma matriz irredutível<sup>2</sup> tomamos  $p \in (0, 1)$  e consideramos um passeio aleatório que segue as transições de  $Q$  com probabilidade  $p$  ou que com probabilidade  $1 - p$  vai pra qualquer outra página *web* (como o comportamento do internauta absorto que ficou entediado), ou seja,

$$R = pQ + (1 - p) \frac{1}{n} \mathbf{1} \quad (74)$$

<sup>2</sup>Em linguagem de Teoria dos Grafos, o grafo dirigido tem que ser fortemente conexo, caso contrário o vetor estacionário pode ter todas as coordenadas nulas fora de uma componente fortemente conexa do grafo.



onde  $\mathbf{1}$  é a matriz com todas as entradas iguais a 1. A matriz obtida de  $Q$  do exemplo da figura 2

$$R = \begin{pmatrix} \frac{1-p}{6} & \frac{p}{2} + \frac{1-p}{6} & \frac{p}{2} + \frac{1-p}{6} & \frac{1-p}{6} & \frac{1-p}{6} & \frac{1-p}{6} \\ \frac{p}{2} + \frac{1-p}{6} & \frac{1-p}{6} & \frac{p}{2} + \frac{1-p}{6} & \frac{1-p}{6} & \frac{1-p}{6} & \frac{1-p}{6} \\ \frac{p}{2} + \frac{1-p}{6} & \frac{p}{2} + \frac{1-p}{6} & \frac{1-p}{6} & \frac{1-p}{6} & \frac{1-p}{6} & \frac{1-p}{6} \\ \frac{1-p}{6} & \frac{1-p}{6} & \frac{p}{3} + \frac{1-p}{6} & \frac{1-p}{6} & \frac{p}{3} + \frac{1-p}{6} & \frac{p}{3} + \frac{1-p}{6} \\ \frac{1-p}{6} & \frac{p}{3} + \frac{1-p}{6} & \frac{1-p}{6} & \frac{p}{3} + \frac{1-p}{6} & \frac{1-p}{6} & \frac{p}{3} + \frac{1-p}{6} \\ \frac{p}{6} + \frac{1-p}{6} & \frac{p}{6} + \frac{1-p}{6} & \frac{p}{6} + \frac{1-p}{6} & \frac{p}{6} + \frac{1-p}{6} & \frac{p}{6} + \frac{1-p}{6} & \frac{p}{6} + \frac{1-p}{6} \end{pmatrix}$$

É sabido que o método iterativo para computar o vetor estacionário descrito, o *método das potências*, converge com velocidade  $|\lambda_2/\lambda_1|$  onde  $\lambda_1 > \lambda_2$  são os dois maiores autovalores da matriz  $R$ . Ainda, é sabido que  $\lambda_1 = 1$  e que  $|\lambda_2| \leq p$ .

O parâmetro  $p$  é conhecido como fator de amortecimento (*damping factor*). No trabalho que originou o algoritmo os autores do PageRank estabeleceram o valor  $p = 0,85$  após testes. Atualmente, a Google não divulga o valor desse fator. No mesmo trabalho, os autores reportam de 50 a 100 iterações do método das potências até a condição de parada do método das potências. O critério tradicional de parada é da forma  $\|\mathbf{p}^{(t+1)} - \mathbf{p}^{(t)}\|_2 < \epsilon$  para uma tolerância  $\epsilon > 0$  pequena; notemos que não é necessário conhecer as grandezas do vetor estacionário, só é preciso determinar a ordem das coordenadas, o que pode ser usado para diminuir o número de iterações. Para  $p = 0,85$  foi reportado que com 29 iterações  $\|\pi^{k+1} - \pi^k\|_2 < 10^{-2}$ , e no caso de 50 a 100 iterações a tolerância é de  $10^{-3}$  a  $10^{-7}$ .

Para concluirmos esta seção vão mostrar uma alternativa para escrever essa matriz, uma vez que atordoantes 1 trilhão de páginas indexadas foi reportado pela Google em julho de 2008. A matriz  $P$  é esparsa pela natureza da *web*: páginas com poucos *links*, 52 em média, e muitos vértices pendentes, a matriz  $Q$  é mais densa e a matriz  $R$  é positiva. A matriz  $A$  pode ser escrita como  $\mathbf{u}^T \mathbf{a}$  onde  $\mathbf{u}$  é o vetor com todas as entradas iguais a  $1/n$ . De fato, pode ser qualquer vetor de probabilidades, o vetor uniforme foi a escolha original, mas atualmente sabe-se que favorece *link spamming* e esse parâmetro também não é divulgado pela Google. O vetor  $\mathbf{a}^T$  é o transposto do vetor característico dos vértices pendentes. Com essas definições

$$\mathbf{p}^{(t)} = \mathbf{p}^{(t-1)} R = p \mathbf{p}^{(t-1)} Q + (1-p) \frac{1}{n} \mathbf{p}^{(t-1)} \mathbf{1} = p \mathbf{p}^{(t-1)} Q + (1-p) \frac{1}{n} \mathbf{1} = p \mathbf{p}^{(t-1)} Q + p \mathbf{p}^{(t-1)} \mathbf{u}^T \mathbf{a} + (1-p) \mathbf{u}^T \mathbf{e}$$

onde  $\mathbf{e} = (1, 1, \dots, 1)$ , portanto só precisamos armazenar a matriz  $P$  e os vetores  $\mathbf{a}$ ,  $\mathbf{e}$  e  $\mathbf{u}$ .

**Exercício 114** Mostre que a matriz  $R$  é irredutível.

**Exercício 115** O grafo dirigido com vértices  $\{0, 1, \dots, n-1\}$  e arestas  $(i, i+1 \bmod n)$  é um circuito dirigido com  $n$  vértices. Analise o comportamento de  $\mathbf{r}^{(k)}$  para  $k \in \mathbb{N}$  com  $\mathbf{r}^{(0)} = (1, 0, \dots, 0)$ .

## Exercícios complementares

**Exercício 116** Seja  $P$  uma matriz de transições de uma cadeia de Markov. Defina a matriz  $P' = (P + \text{Id})/2$ . Prove que a cadeia com matriz de transições  $P'$  é aperiódica.

**Exercício 117** Considere a generalização natural do algoritmo para o 2-SAT, seção , para 3-SAT. Prove que tal algoritmo é exponencial,  $O(2^n)$  em que  $n$  é o número de variáveis.

**Exercício 118** Por dia, uma dentre  $n$  pessoas é requisitada, a pessoa  $i$  requisitada com probabilidade  $p_i$ . As pessoas são organizadas numa lista ordenada e a cada requisição a pessoa escolhida é colocada no topo da lista, o resto da lista fica inalterada.

1. Numa modelagem com cadeia de Markov, quais são os estados?
2. Qual é a distribuição invariante? (Dica: Não precisa fazer conta.)

**Exercício 119** Uma pessoa tem  $n$  guarda-chuvas que usa no trajeto de casa ao trabalho e vice-versa. Se ela está em casa no começo de dia e está chovendo então escolhe um guarda-chuva para ir ao escritório, caso haja disponível. Se ela está no escritório no fim do dia e está chovendo então escolhe um guarda-chuva para ir para casa, caso haja disponível. Quando não está chovendo ela não escolhe usar guarda-chuva. Assumindo que no início do dia e no fim do dia chove de modo independente e com probabilidade  $p$ , qual é a fração das viagens que ela faz sob chuva e sem guarda-chuva?

**Exercício 120** Um grupo de  $n$  processadores é arranjado em lista ordenada e quando uma tarefa chega é atribuída ao primeiro da lista; se ele está ocupado tenta-se o segundo da lista, e assim por diante. Uma vez que a tarefa é processada ou que não é encontrado um processador disponível pra ela, ela deixa o sistema. Nesse momento, podemos reordenar os processadores antes que a próxima tarefa chegue e a estratégia é: se um processador atendeu a tarefa ele avança uma posição na lista, trocando-o com o processador imediatamente a frente; se todos falharam ou o primeiro atendeu a ordem permanece a mesma. Assumamos que no processador  $i$  atende a tarefa com probabilidade  $p_i$ .

1. Defina uma cadeia de Markov para analisar o problema.
2. Mostre que a cadeia é reversível.
3. Descubra as probabilidades limite.

**Exercício 121** Num tabuleiro de xadrez o cavalo começa num dos quatro cantos e se move com igual probabilidade para qualquer uma das posições legais, de acordo com as regras do seu movimento. Qual é o número esperado de movimentos até voltar para a posição de saída?

**Exercício 122** Como podemos gerar valores para vinte variáveis aleatórias independentes e com distribuição uniforme em  $(0, 1)$  condicionadas a soma delas serem menor que dois?

**Exercício 123** Uma coloração de um grafo é uma atribuição de cores aos vértices do grafo. Um grafo é  $k$ -colorível se há uma coloração com  $k$  cores sem que vértices adjacentes recebam a mesma cor. Seja  $G$  um grafo 3-colorível.

Mostre que  $G$  tem uma coloração com 2 cores sem que haja triângulo com os três vértices da mesma cor.

Considere o seguinte algoritmo para achar a coloração afirmada no parágrafo acima: o algoritmo começa com uma coloração de duas cores arbitrária. Enquanto houver triângulo monocromático, escolha um deles e sorteie um de seus vértices, mude a cor do vértice sorteado. Determine um limitante superior para o número esperado de recolorações até o algoritmo terminar.

**Exercício 124** Seja  $X_n$  a soma dos resultados de  $n$  lançamentos independentes de um dado. Mostre que para  $k \geq 2$

$$\lim_{n \rightarrow \infty} \mathbb{P}(k \text{ divide } X_n) = \frac{1}{k}.$$

**Exercício 125** Numa cidade com  $n+1$  habitantes, uma pessoa conta um boato para uma segunda pessoa, que por sua vez conta o boato para uma terceira pessoa e assim por diante. Em cada passo o ouvinte do boato é escolhido aleatoriamente de maneira uniforme dentre as outras pessoas na cidade. Como esse processo pode ser modelado como uma cadeia de Markov? Qual a probabilidade do boato ser contado  $r$  vezes sem repetir nenhuma pessoa? Escreva um código em R para determinar o número de rodadas até que todos tenham ouvido o boato com probabilidade 0.999, para  $n = 128$ .

**Exercício 126** Um gato e um rato passeiam aleatoriamente num grafo conexo e não-bipartido de modo independente. Eles começam ao mesmo tempo em vértices diferentes e ambos dão um passo a cada instante, também de modo independente. O gato come o rato se em algum instante eles estão no mesmo vértice. Mostre que o número esperado de passos até eles se encontrarem é  $O(m^2 n)$ , em que  $n$  é o número de vértice e  $m$  o número de arestas de  $G$ . (Dica: considere a cadeia de Markov com estados  $(u, v)$  sendo as posições de cada animal no grafo, num instante.)

**Exercício 127** O grafo *lollipop* é um grafo completo com  $n$  vértices e um caminho com  $n$  vértices e um único vértice  $u$  em comum. Chamamos de  $v$  o único vértice de grau um, o vértice final do caminho.

Mostre que se um passeio aleatório começa em  $v$  então o número esperado de passos até visitar todos os vértices é  $O(n^2)$ .

Mostre que se um passeio aleatório começa em  $u$  então o número esperado de passos até visitar todos os vértices é  $O(n^3)$ .