Introdução a Teoria das Probabilidades e Processos Aleatórios

Notas de Aula

Rafael Bassi Stern, com edições de Rafael Izbicki

Última revisão: 1 de março de 2018

Por favor, envie observações, comentários e correções para rafaelizbicki@gmail.com

Sumário

1 Preâmbulo Matemático

1.1 Análise Combinatória

Ser capaz de contar com precisão os possíveis resultados de um experimento é um ponto-chave na teoria das probabilidades. A Análise combinatória é uma ferramenta poderosa para atingir este objetivo. Esta seção contém uma breve introdução a este assunto.

1.1.1 Princípio de Contagem

Assuma que serão realizados r experimentos. Se o i-ésimo experimento possui n_i possíveis resultados, $i = 1, \dots, n$ então o número total de resultados dos r experimentos é

$$\Pi_{i=1}^{r} n_i := n_1 n_2 \dots n_r.$$

Exemplo 1. *Se lançamos uma moeda r vezes, o número total de resultados em cada experimento é dois: cara ou coroa. Segue do princípio de contagem que existem $\Pi_{i=1}^{r} 2 = 2^r$ resultados diferentes do experimento conjunto, i.e., existe um total de 2^r sequências de caras ou coroas.*

Exemplo 2. *Se lançamos uma moeda e depois um dado, existe um total de $2 \times 6 = 12$ resultados que podemos observar.*

1.1.2 Permutações

Considere um conjunto com n objetos. O número total de maneiras de que podemos ordenar estes objetos é dado por

$$n! := n \times (n-1) \times \dots \times 2 \times 1$$

Exemplo 3. *Existem 4!=24 maneiras de dispor quatro camisetas num armário.*

1.1.3 Combinações

Considere um conjunto com n objetos. O número total de grupos diferentes de tamanho $r \leq n$ que podem ser formados com estes n objetos é dado por

$$\binom{n}{r} := \frac{n!}{r!(n-r)!}$$

Exemplo 4. *Existem $\binom{10}{2} = 45$ formas de escolher duas questões de uma seção com 10 questões deste livro para estudar.*

1.1.4 Exercícios

Exercício 1. *Um restaurante oferece três entradas diferentes, cinco pratos principais e seis sobremesas. Quantas refeições distintas este restaurante oferece?*

Exercício 2. *Um aluno tem 2 livros de matemática, 3 livros de geografia e 4 livros de química. De quantas formas estes livros podem ser dispostos numa estante:*

- *Se o aluno não se importa se livros da mesma matéria não estejam um do lado do outro?*
- *Se o aluno quer que os livros sobre o mesmo assunto fiquem juntos?*

Exercício 3. *Quantas palavras diferentes podem ser formadas com duas letras A, três letras B e uma letra C?*

Exercício 4. *Um aluno precisa estudar para três provas esta semana. O professor de matemática deu 6 exercícios para ajudar ele a estudar para a prova, o professor de geografia deu 7 exercícios, e o professor de química deu 5. Considerando que este aluno não tem muito tempo, quantos conjuntos diferentes de exercícios ele pode pegar para resolver se ele quer escolher somente 2 exercícios de matemática, 4 exercícios de geografia e 1 exercício de química?*

Exercício 5. *Oito pessoas, chamadas A, B, $\ldots$, H, vã formar uma fila.*

- *De quantas maneiras estas pessoas podem ser colocadas na fila?*
- *De quantas maneiras estas pessoas podem ser colocadas se A e B precisam ficar um do lado do outro?*
- *De quantas maneiras estas pessoas podem ser colocadas se A e B precisam ficar um do lado do outro, e C e D também precisam ficar um d lado do outro?*

1.2 Teoria de Conjuntos (Básica)

Como vimos na seção anterior, a análise combinatória é usada para contar os possíveis resultados de um experimento. Na teoria das probabilidades, nos também temos de ser capazes de descrever com precisão os tais resultados. É aqui que a teoria de conjuntos se torna importante.

Um conjunto é uma coleção de objetos. Se um conjunto tem um número finito de objetos, $o_1, o_2, \ldots, o_n$, o denotamos por $\{o_1, o_2, \ldots, o_n\}$. Denotamos o conjunto dos números naturais por $\mathbb{N}$, o conjunto dos inteiros por $\mathbb{Z}$ e o conjunto dos reais por $\mathbb{R}$. Em probabilidade, conjuntos são fundamentais para descrever resultados de um experimento.

Exemplo 5 (Conjuntos)**.**

- *O conjunto de possíveis resultados de um dado de seis lados:* $\{1, 2, 3, 4, 5, 6\}$.
- *O conjunto dos resultados do lançamento de uma moeda:* $\{C, K\}$.
- *O conjunto de resultados do lançamento de duas moedas:* $\{(C, C), (C, K), (K, C), (K, K)\}$.
- *O conjunto de todos os números impares:* $\{2n + 1 : n \in \mathbb{N}\}$ *or* $\{1, 3, 5, 7, \ldots\}$.
- *O conjunto de números reais não negativos:* $\{x \in \mathbb{R} : x \geq 0\}$.
- *Círculo de raio 1:* $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$.

Definição 1 ($\in$ e $\notin$)**.** *Escrevemos $o \in S$ se o objeto o é um elemento do conjunto S e $o \notin S$, caso contrário.*

Exemplo 6 ($\in$ e $\notin$)**.**

- $C \in \{C, K\}$.
- $7 \notin \{1, 2, 3, 4, 5, 6\}$.
- $7 \in \{2n + 1 : n \in \mathbb{N}\}$.

Definição 2 (**conjunto vazio -** $\emptyset$)**.** *$\emptyset$ é o único conjunto sem elementos. Isto é, para todo objeto o, $o \notin \emptyset$.*

Definição 3 (**conjuntos disjuntos**).

- *Dois conjuntos A e B são* ***disjuntos*** *se, para todo $o \in A$ temos que $o \notin B$ e para todo $o \in B$, $o \notin A$.*
- *Uma sequência de conjuntos $(A_n)_{n \in \mathbb{N}}$ é disjunta se, para todo $i \neq j$, A_i é disjunto de A_j.*

Exemplo 7 (Conjuntos disjuntos).

- $\{1, 2\}$ *e* $\{3, 4\}$ *são disjuntos.*
- $\{1, 2\}$ *e* $\{2, 3\}$ *náo são disjuntos, pois* $2 \in \{1, 2\}$ *e* $2 \in \{2, 3\}$.

Definição 4 ($\subset$ e $=$). *Sejam A e B dois conjuntos. Dizemos que:*

- *$A \subset B$ se, para todo $o \in A$, $o \in B$.*
- *$A = B$ se $A \subset B$ e $B \subset A$.*

Exemplo 8 ($\subset$ e $=$).

- $\{1, 2\} \subset \{1, 2, 3, 4\}$.
- $\{n \in \mathbb{Z} : n \geq 1\} \subset \mathbb{N}$.
- $\{n \in \mathbb{Z} : n \geq 0\} = \mathbb{N}$.

Nos reservamos o símbolo Ω para o conjunto de todos os objetos que consideramos num certo modelo. Ω é frequentemente chamado de **espaço amostral** na teoria das probabilidades. Isto é, para todo conjunto A que consideramos naquele modelo, $A \subset \Omega$.

1.2.1 Operações com conjuntos

Definição 5 (complemento - ^{c}). *Seja A um conjunto. o é um elemento de A^c se e somente se $o \notin A$. Isto é, o complemento de A é definido formalmente como $A^c = \{o \in \Omega : o \notin A\}$.*

Exemplo 9 (^{c}).

- *Let* $\Omega = \{C, K\}$, $\{C\}^c = \{K\}$.
- *Let* $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\{1, 2\}^c = \{3, 4, 5, 6\}$.
- *Let* $\Omega = \mathbb{N}$, $\{n \in \mathbb{N} : n > 0\}^c = \{0\}$.

Definição 6 (união - $\cup$).

- *Sejam A e B dois conjuntos. $o \in \Omega$ é um elemento da união de A e B, $A \cup B$, se e somente se* ***ou*** *o é um elemento de A* ***ou*** *o é um elemento de B. Isto é $A \cup B = \{o \in \Omega : o \in A$* ***ou*** *$o \in B\}$.*
- *Seja $(A_n)_{n \in \mathbb{N}}$ uma sequência de conjuntos. $o \in \Omega$ é um elemento da união de $(A_n)_{n \in \mathbb{N}}$, $\cup_{n \in \mathbb{N}} A_n$, se e somente se existe um $n \in \mathbb{N}$ tal que $o \in A_n$. Isto é, $\cup_{n \in \mathbb{N}} A_n = \{o \in \Omega :$* ***existe*** *$n \in \mathbb{N}$ tal que $o \in A_n\}$*

Exemplo 10 ($\cup$).

- $\{C\} \cup \{K\} = \{C, K\}$.

- $\{1,2\} \cup \{2,3\} = \{1,2,3\}$.
- $\{1\} \cup \{3\} \cup \{5\} = \{1,3,5\}$.
- $\{n \in \mathbb{Z} : n > 0\} \cup \{n \in \mathbb{Z} : n < 0\} = \{n \in \mathbb{Z} : n \neq 0\}$.
- $\cup_{n \in \mathbb{N}}\{n\} = \mathbb{N}$.
- $\cup_{n \in \mathbb{N}}\{x \in \mathbb{R} : x \geq n\} = \{x \in \mathbb{R} : x \geq 0\}$.
- $\cup_{n \in \mathbb{N}}\{x \in \mathbb{R} : x \geq 1/(n+1)\} = \{x \in \mathbb{R} : x > 0\}$.

Definição 7 (interseçsão - $\cap$)**.**

- *Sejam A e B dois conjuntos. o é um elemento da interseçsão entre A e B, $A \cap B$, se e somente se $o \in \Omega$ é um elemento de A* ***e*** *o um elemento de B. Isto é, $A \cap B = \{o \in \Omega : o \in A$ **e** $o \in B\}$.*
- *Seja $(A_n)_{n \in \mathbb{N}}$ uma sequência de conjuntos. $o \in \Omega$ é um elemento da interseçsão de $(A_n)_{n \in \mathbb{N}}$, $\cap_{n \in \mathbb{N}} A_n$, se e somente se para todo $n \in \mathbb{N}$, $o \in A_n$. Isto é, $\cap_{n \in \mathbb{N}} A_n = \{o \in \Omega :$ **para todo** $n \in \mathbb{N}, o \in A_n\}$*

Exemplo 11 ($\cap$)**.**

- $\{C\} \cap \{K\} = \emptyset$.
- $\{1,2\} \cap \{2,3\} = \{2\}$.
- $(\{1,2\} \cap \{2,3\}) \cup \{5\} = \{2,5\}$.
- $\{n \in \mathbb{Z} : n \geq 0\} \cap \{n \in \mathbb{Z} : n \leq 0\} = \{0\}$.
- $\cap_{n \in \mathbb{N}}\{i \in \mathbb{N} : i \geq n\} = \emptyset$.
- $\cap_{n \in \mathbb{N}}\{x \in \mathbb{R} : x \leq n\} = \{x \in \mathbb{R} : x \leq 0\}$.

Teorema 1 (leis de DeMorgan's)**.** *Seja $(A_n)_{n \in \mathbb{N}}$ uma sequência de subconjuntos de Ω. Então, para todo $n \in \mathbb{N}$,*

- $(\cup_{i=1}^{n} A_i)^c = \cap_{i=1}^{n} A_i^c$
- $(\cap_{i=1}^{n} A_i)^c = \cup_{i=1}^{n} A_i^c$

Além disso,

- $(\cup_{i \in \mathbb{N}} A_i)^c = \cap_{i \in \mathbb{N}} A_i^c$
- $(\cap_{i \in \mathbb{N}} A_i)^c = \cup_{i \in \mathbb{N}} A_i^c$

Definição 8 (Partição)**.** *Seja $(A_n)_{n \in \mathbb{N}}$ uma sequência de conjuntos. Dizemos que $(A_n)_{n \in \mathbb{N}}$ particiona Ω se:*

- *para todo $i, j \in \mathbb{N}$ tal que $i \neq j$, A_i e A_j são disjuntos.*
- $\cup_{n \in \mathbb{N}} A_n = \Omega$.

1.2.2 Exercícios

Exercício 6. *Seja* $A = \{1, 3, 5\}$, $B = \{1, 2\}$ *e* $\Omega = \{1, 2, 3, 4, 5, 6\}$. *Encontre:*

a. $A \cup B$

b. $A \cap B^c$

c. $B \cup B^c$

d. $(A \cup B)^c$

e. $(A \cap B)^c$

Exercício 7. *Considere que um dado dia pode ser chuvoso - C - ou não chuvoso - NC. Estamos interessados no tempo para os próximos dois dias.*

a. *Como você escreveria o Ω formalmente?*

b. *Como você escreveria formalmente "O resultado é tal que os dois dias são chuvosos? Chame este conjunto de A.*

c. *Como você escreveria formalmente "O resultado é tal que pelo menos um dia é chuvoso? Chame este conjunto de B.*

d. *É verdadeiro que $A \subset B$?*

e. *Encontre B^c. Como você descreveria este conjunto em português?*

f. *É verdadeiro que A e B^c são disjuntos?*

Exercício 8. *As afirmações a seguir são verdadeiras ou falsas?*

a. $\emptyset \in \emptyset$.

b. $\emptyset \in \{\emptyset\}$.

c. $\{\emptyset\} \in \emptyset$.

d. $\emptyset \subset \emptyset$.

e. $\emptyset \subset \{\emptyset\}$.

f. $\{\emptyset\} \subset \emptyset$.

Exercício 9. *Demonstre as seguintes afirmações:*

a. *S e T são disjuntos se e somente se $S \cap T = \emptyset$.*

b. $S \cup T = T \cup S$

c. $S \cap T = T \cap S$

d. $S = (S^c)^c$

e. $S \cup S^c = \Omega$

f. $S \cap \Omega = S$

g. $S \cup \Omega = \Omega$

h. $(S \cap T)^c = S^c \cup T^c$

i. $(S \cup T)^c = S^c \cap T^c$

j. $(\{n\})_{n \in \mathbb{N}}$ *particiona* $\mathbb{N}$.

k. $(\{1, 2\}, \{3, 4\}, \emptyset, \emptyset, \emptyset, \ldots)$ *particiona* $\{1, 2, 3, 4\}$.

Exercício 10. *Demonstre a lei de DeMorgan.*

Exercício 11. *Numa classe de* 25 *alunos, considere as seguintes afirmações:*

- 14 *vão graduar em Ciências da Computação.*
- 12 *vão graduar em Engenharia.*
- 5 *vão graduar tanto em Ciências da Computação quanto em Engenharia.*

Quantos alunos não vão graduar nem em Ciências da Computação nem em Engenharia?

2 Teoria das Probabilidades

2.1 Os Axiomas da Probabilidade

Denotamos por Ω o **espaço amostral**, isto é, o conjunto de todos os possíveis resultados de um experimento que temos interesse. Por exemplo, pode-se estar interessado no resultado do lançamento de uma moeda ($\Omega = \{H, T\}$), o número de dias chuvosos no próximo ano ($\Omega = \{n \in \mathbb{N} : n \leq 366\}$) ou no tempo até certo transistor falhar ($\Omega = \{x \in \mathbb{R} : x > 0\}$).

Para todo $A \subset \Omega$, podemos não ter certeza se A é verdadeiro ou não. Por exemplo, $\{K\} \subset \{C, K\}$ e normalmente não podemos determinar se o resultado do lançamento de uma moeda será cara ou coroa antes de lançar a moeda. A probabilidade é a função definida nos subconjuntos de Ω que é normalmente interpretada de uma das formas abaixo:

- (Frequência) $P(A)$ denota a frequência relativa com que os resultados em A serão observados se o mesmo resultado for realizado, independentemente, muitas vezes.
- (Grau de confiança) $P(A)$ representa o grau de confiança que o avaliador tem em que o resultado será um elemento de A.

Eventos São subconjuntos de Ω aos quais nos atribuímos probabilidades. Denotamos por $\mathcal{F}$ o conjunto de todos os eventos. Para que P seja uma probabilidade, este deve satisfazer os Axiomas da Probabilidade.

Definição 9 (Axiomas da Probabilidade)**.** $\mathbb{P} : \mathcal{F} \to \mathbb{R}$ *é uma probabilidade se:*

1. *(Não negatividade) Para todo $A \in \mathcal{F}$, $\mathbb{P}(A) \geq 0$.*
2. *(Aditividade enumerável) Se $(A_n)_{n \in \mathbb{N}}$ é uma sequência de conjuntos disjuntos em $\mathcal{F}$, $\mathbb{P}(\cup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} \mathbb{P}(A_i)$.*
3. *(Normalização) $\mathbb{P}(\Omega) = 1$.*

A seguir, provaremos algumas consequências destes axiomas que serão usadas constantemente:

Lema 1. $\mathbb{P}(\emptyset) = 0$.

Demonstração. Seja $(A_n)_{n \in \mathbb{N}}$ uma sequência de subconjuntos de Ω tal que $A_0 = \Omega$ e $A_i = \emptyset$ para todo $i > 0$. Por construção, esta é uma sequência de conjuntos disjuntos e, portanto, pela aditividade enumerável:

$$\mathbb{P}(\cup_{n \in \mathbb{N}} A_i) = \sum_{n \in \mathbb{N}} \mathbb{P}(A_i) = \mathbb{P}(\Omega) + \sum_{n > 0} \mathbb{P}(\emptyset)$$

Observe que $\cup_{n \in \mathbb{N}} A_i = \Omega$ e, aplicando este fato na equação acima:

$$\mathbb{P}(\Omega) = \mathbb{P}(\Omega) + \sum_{n > 0} \mathbb{P}(\emptyset)$$

Concluímos que $\mathbb{P}(\emptyset) = 0$. $\square$

Lema 2. *Se $A_1, A_2, \ldots, A_n$ são disjuntos, então $\mathbb{P}(A_1 \cup A_2 \ldots \cup A_n) = \sum_{i=1}^{n} \mathbb{P}(A_i)$.*

Demonstração. Considere a sequência $(B_i)_{i\in\mathbb{N}}$ tal que $B_i = A_i$ para $i \leq n$ e $B_i = \emptyset$ para $i > n$. Observe que $(B_i)_{i\in\mathbb{N}}$ é uma sequência de conjuntos disjuntos. Consequentemente, pela aditividade enumerável:

$$\mathbb{P}(A_1 \cup A_2 \ldots \cup A_n) = \mathbb{P}(\cup_{i\in\mathbb{N}} B_i) = \sum_{i\in\mathbb{N}} \mathbb{P}(B_i) = \sum_{i=1}^{n} \mathbb{P}(A_i) + \sum_{i>n} \mathbb{P}(\emptyset)$$

Como $\mathbb{P}(\emptyset) = 0$, concluímos que $\mathbb{P}(A_1 \cup A_2 \ldots \cup A_n) = \sum_{i=1}^{n} \mathbb{P}(A_i)$. □

Lema 3. *Para todo evento A, $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$*

Demonstração. Observe que A e A^c são disjuntos e $A \cup A^c = \Omega$. Portanto:

$$1 = \mathbb{P}(\Omega) = \mathbb{P}(A \cup A^C) = \mathbb{P}(A) + \mathbb{P}(A^c)$$

Unindo os extremos, $1 = \mathbb{P}(A) + \mathbb{P}(A^c)$, that is, $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$. □

Lema 4. *Para todos os eventos A e B, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.*

Demonstração. Observe que $A^c \cap B$ e $A \cap B$ são disjuntos e $(A^c \cap B) \cup (A \cap B) = B$. Portanto,

$$\mathbb{P}(B) = \mathbb{P}(A^c \cap B) + \mathbb{P}(A \cap B) \tag{1}$$

$$\mathbb{P}(A^c \cap B) = \mathbb{P}(B) - \mathbb{P}(A \cap B) \tag{2}$$

Também observe que A e $A^c \cap B$ são disjuntos e $A \cup (A^c \cap B) = A \cup B$. Portanto,

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(A^C \cap B) \tag{3}$$

Usando Equação 2 na Equação 3:

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

□

Lema 5. *Se $A \subset B$, então $\mathbb{P}(B) \geq \mathbb{P}(A)$.*

Demonstração. Obseve que $A \cap B$ e $A^c \cap B$ são disjuntos e $(A \cap B) \cup (A^c \cap B) = B$. Portanto,

$$\mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(A^c \cap B)$$

Como $A \subset B$, concluímos que $A \cap B = A$. Portanto, $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(A^c \cap B)$. Pela não negatividade da probabilidade, $\mathbb{P}(A^c \cap B) \geq 0$ e, consequentemente, $\mathbb{P}(B) \geq \mathbb{P}(A)$. □

Exemplo 12. *Considere o lançamento de uma moeda. Seja $\Omega = \{C, K\}$ e $\mathcal{F} = \{\emptyset, \{C\}, \{K\}, \{C, K\}\}$. Segue dos axiomas da probabilidade que $\mathbb{P}(\emptyset) = 0$ e $\mathbb{P}(\{C, K\}) = 1$. Dizemos que a moeda é honesta se $\mathbb{P}(C) = 0.5$ e, consequentemente, $\mathbb{P}(K) = 1 - \mathbb{P}(C) = 0.5$.*

2.1.1 Exercícios

Exercício 12. *Se a probabilidade de eu ter uma mão pareada em um jogo de pocker é $\frac{1}{18}$, qual é a probabilidade de eu não ter uma mão pareada?*

Exercício 13. *Sejam A e B dois eventos disjuntos. É possível que $\mathbb{P}(A) = 0.4$ e $\mathbb{P}(B) = 0.7$? Porque?*

Exercício 14. *Considere o argumento incorreto a seguir:*

> *Considere uma moeda honesta e um dado de seis faces honesto. Seja C o conjunto de todos os possíveis resultados do lançamento da moeda, o que implica que $\mathbb{P}(C) = 1$. seja D o conjunto de todos os possíveis resultados do lançamento do dado, o que implica que $\mathbb{P}(D) = 1$. Agora $\mathbb{P}(C \cup D)$, que é a probabilidade de que ou C ou D ocorre é 1. Mesmo assim, C e D são também eventos disjuntos, portanto $\mathbb{P}(C \cup D) = \mathbb{P}(C) + \mathbb{P}(D) = 1 + 1$. Isto é, $1 + 1 = 1$.*

Qual é o espaço amostral do problema descrito? Onde estão os erros do argumento?

Exercício 15. *Sejam A e B dois eventos.*

a. *Se $\mathbb{P}(A) = 0.7$, $\mathbb{P}(B) = 0.4$ e $\mathbb{P}(A \cup B) = 0.8$, qual é a probabilidade de $\mathbb{P}(A \cap B)$?*

b. *Se $\mathbb{P}(A \cap B) = 0.25$, $\mathbb{P}(A \cup B) = 0.75$ e $\mathbb{P}(A) = 2\mathbb{P}(B)$, qual é o valor de $\mathbb{P}(A)$?*

c. *Prove que se $\mathbb{P}(A) = \mathbb{P}(B^c)$, então $\mathbb{P}(A^c) = \mathbb{P}(B)$.*

Exercício 16. *Mostre que $\mathbb{P}((A \cap B^c) \cup (A^c \cap B)) = \mathbb{P}(A) + \mathbb{P}(B) - 2\mathbb{P}(A \cap B)$.*

Exercício 17. *Prove que*

$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C)$$

Exercício 18.

Seja Ω o espaço amostral e A, B e C eventos. Prove que:

a. $\mathbb{P}(A \cup B) = 1 - \mathbb{P}(A^c \cap B^c)$.

b. $\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c)$.

c. $\max(\mathbb{P}(A), \mathbb{P}(B)) \leq \mathbb{P}(A \cup B) \leq \min(1, \mathbb{P}(A) + \mathbb{P}(B))$.

d. $\max(0, \mathbb{P}(A) + \mathbb{P}(B) - 1) \leq \mathbb{P}(A \cap B) \leq \min(\mathbb{P}(A), \mathbb{P}(B))$.

Exercício 19. *Sejam $A_1, A_2, \ldots \in \mathcal{F}$. Prove que*

$$\mathbb{P}(\cup_{n \geq 1} A_n) \leq \sum_{n \geq 1} \mathbb{P}(A_n)$$

Exercício 20. *Sejam A, B e C eventos de Ω. Usando operações de conjuntos, descreva os conjuntos a seguir:*

- *Pelo menos um de A, B ou C ocorre*
- *A, B e C ocorrem*
- *A e B ocorrem, mas C não*
- *Exatamente um de A, B ou C ocorre*

Exercício 21. *(Desafio). Seja $A_1, A_2, \ldots$ uma sequência de eventos com $A_i \subseteq A_{i+1}$ para todo $i \geq 1$. Mostre que*

$$\mathbb{P}(\lim_{n \longrightarrow \infty} A_n) = \lim_{n \longrightarrow \infty} \mathbb{P}(A_n),$$

onde

$$\lim_{n \longrightarrow \infty} A_n := \cup_{n \geq 1} A_n.$$

Dica: Defina $F_n := A_n (\cup_{i=1}^{n-1} A_i)^c$ e note que eles são disjuntos.

Exercício 22. *Assumindo que $\mathbb{P}(A_n) = 1$ para todo $n \in \mathbb{N}$, prove que $\mathbb{P}(\cap_{n \in \mathbb{N}} A_n) = 1$.*

2.2 *A Verdade sobre Probabilidade

Apesar de que quando apresentamos os axiomas da probabilidade nos assumimos que $\mathbb{P}$ é definido para *qualquer* subconjunto E do espaço amostral Ω, isso é, na verdade, forte demais para espaços não enumeráveis: assim, não é possível definir a distribuição uniforme em $(0, 1)$ (veja a seção 5) que satisfaz todos os axiomas. A probabilidade moderna (mais precisamente, a teoria da medida), nos mostra que só é possível definir tal $\mathbb{P}$ para alguns eventos. Nesta seção, apresentamos uma visão geral desta questão. Uma medida de probabilidade é, na verdade, definida em σ-álgebra do espaço amostral:

Definição 10 (σ-álgebra)**.**

Dizemos que o conjunto $\mathcal{F}$ de subconjuntos de Ω é uma σ-álgebra se

(i) $\Omega \in \mathcal{F}$

(ii) $A \in \mathcal{F}$ *implica que* $A^c \in \mathcal{F}$

(iii) $A_i \in \mathcal{F}$ *para todo* $i \in \mathbb{N}$ *implica que* $\cup_{i \in \mathbb{N}} A_i \in \mathcal{F}$

Exemplo 13 (σ-álgebras)**.** *Os conjuntos a seguir são σ-álgebras:*

- $\mathcal{F} = \{\emptyset, \Omega\}$,
- $\mathcal{F} = \{\emptyset, \{a\}, \{b, c\}, \Omega\}$ *quando* $\Omega = \{a, b, c\}$,
- $\mathcal{F} = \mathcal{P}(\Omega)$, *i.é., o conjunto das partes de Ω (i.é., $\mathcal{P}(\Omega) = \{E : E \subset \Omega\}$).*

Podemos agora escrever os afiomas gerais de probabilidade:

Definição 11. *Seja $\mathcal{F}$ uma σ-álgebra em Ω. Uma medida de probabilidade em $\mathcal{F}$ é uma função $\mathbb{P} : \mathcal{F} \to \mathbb{R}$ que satisfaz:*

1. *(Não negatividade) Para todo $A \in \mathcal{F}$, $P(A) \geq 0$.*
2. *(Aditividade enumerável) Se $(A_n)_{n \in \mathbb{N}}$ é uma sequência de conjuntos disjuntos em $\mathcal{F}$, $\mathbb{P}(\cup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} \mathbb{P}(A_i)$.*
3. *(Normalização) $\mathbb{P}(\Omega) = 1$.*

Alguns conjuntos $\mathcal{F}$ são mais úteis que os outros. Assim, é possível definir a medida contínua uniforme sobre uma σ-álgebra especial chamada de os *Borelianos* de $(0, 1)$, mas *não* é possível fazer isso sobre o conjunto das partes de $(0, 1)$. Esta é a razão pela qual σ-álgebras são necessárias na teoria das probabilidades moderna. Para mais informações sobre esta questão, veja por exemplo (Billingsley, 2008).

2.2.1 Exercícios

Exercício 23. *Seja* $\Omega = \{a, b, c, d\}$*.* $\{\emptyset, \{a\}, \{b, c\}, \Omega\}$ *é uma* σ*-álgebra de* Ω*?*

Exercício 24. *Seja* $\mathcal{F}$ *uma* σ*-álgebra. Prove que se* $A_i \in \mathcal{F}$ *para todo* $i \in \mathbb{N}$*, então* $\cap_{i \in \mathbb{N}} A_i \in \mathcal{F}$*.*

2.3 Variáveis aleatórias

Números sobre os quais não temos certeza ocupam uma posição especial na teoria das probabilidades. Variáveis aleatórias são funções de Ω para $\mathbb{R}$ e são usadas para representar a incerteza sobre números.

Definição 12. *Uma variável aleatóra* X *é uma função tal que* $X : \Omega \rightarrow \mathbb{R}$*.*

Exemplo 14. *Considere o lançamento de uma moeda.* $\Omega = \{C, K\}$*. O número de vezes que observamos cara é representado pela variável aleatória* X *tal que* $X(K) = 0$ *and* $X(C) = 1$*.*

Exemplo 15. *Considere que um dado dia pode ser chuvoso - C - ou não chuvoso - NC. Estamos interessados no tempo dos dois dias seguintes.* $\Omega = \{(NC, NC), (NC, C), (C, NC), (C, C)\}$*. O número de vezes em que os dois dias seguintes serão chuvosos é representado pela variável aleatória* X *definida na tabela 1.*

$w \in \Omega$	$X(w)$
(NC, NC)	0
$(NC, C), (C, NC)$	1
(C, C)	2

Tabela 1: Número de dias chuvosos (Exemplo 15)

Definição 13 (Função indicadora)**.** *A função indicadora é um tipo especial de variável aleatória. Seja um evento* $A \in \mathcal{F}$*. A função indicadora de* A *é denotada por* $I_A : \Omega \rightarrow \mathbb{R}$ *e definida como:*

$$I_A(w) = \begin{cases} 1 & \text{, se } w \in A \\ 0 & \text{, caso contrário} \end{cases}$$

Iremos discutir variáveis aleatórias em mais detalhes no Capítulo 3.

2.3.1 Exercícios

Exercício 25. *Sejam* A *e* B *dois eventos disjuntos. É possível escrever* $I_A + I_B$ *como uma única função indicadora?*

Exercício 26. *Sejam* A *e* B *dois eventos. É verdadeiro que* $I_{A \cup B} - I_A - I_B = 0$*? Porque?*

2.4 *Porque estes axiomas foram escolhidos?

Existem muitas possíveis justificativas para os axiomas da probabilidade. Uma explicação simples envolve tipos de apostas.

Suponha que temos um bilhete de loteria na qual você ganha \$1 se o evento A ocorre e \$0, caso contrário. Temos o seguinte jogo:

1. Você define um preço para este bilhete, $Pr(A)$.

2. Um oponente determina se você compra um bilhete dele ou vende um bilhete para ele.

Seu resultado monetário é $\alpha_A(I_A - Pr(A))$, onde α_A é -1 se o oponente faz você vender o bilhete e $\alpha_A = 1$ se faz você comprar. Você garantidamente perde se o oponente tem a opção de fazer você perder dinheiro, garantidamente. Em que situação você perde garantidamente?

Seja $Pr(A) < 0$. Neste caso, $I_A - Pr(A) > 0$ e, portanto, o oponente pode garantir que você perca definindo $\alpha_A = -1$.

Seja $Pr(\Omega) \neq 1$. Observe que $I_\Omega = 1$. Portanto $I_\Omega - Pr(\Omega) = 1 - Pr(\Omega)$.

- Se $Pr(\Omega) < 1$, $I_\Omega - Pr(\Omega) > 0$ e, consequentemente, o oponente pode garantir que você perca definindo $\alpha_A = -1$.
- Se $Pr(\Omega) > 1$, $I_\Omega - Pr(\Omega) < 0$ e, portanto, o oponente pode garantir que você perca definindo $\alpha_A = 1$.

Finalmente, sejam A e B eventos disjuntos. Observe que $I_{A\cup B} - I_A - I_B = 0$. Portanto, se o oponente define $\alpha_{A\cup B} = 1, \alpha_A = -1, \alpha_B = -1$, seu resltado final é:

$$\begin{aligned}
&\alpha_{A\cup B}(I_{A\cup B} - Pr(A\cup B)) + \alpha_A(I_A - Pr(A)) + \alpha_B(I_B - Pr(B)) = \\
&(I_{A\cup B} - I_A - I_B) + (Pr(A) + Pr(B) - Pr(A\cup B)) = \\
&Pr(A) + Pr(B) - Pr(A\cup B)
\end{aligned}$$

De forma similar, se $\alpha_{A\cup B} = -1, \alpha_A = 1, \alpha_B = 1$, então seu resultado final é $Pr(A\cup B) - Pr(A) - Pr(B)$. Portanto, se $Pr(A\cup B) \neq P(A) + P(B)$ existe uma opção tal que o oponente garante que você perca. Isto é, para que você evite uma perda certa Pr deve satisfazer:

1. $Pr(A) \geq 0$.
2. $Pr(\Omega) = 1$.
3. Se A e B são disjuntos, $Pr(A\cup B) = Pr(A) + Pr(B)$.

Estas regras se assemelham aos axiomas da probabilidade. Portanto, se acreditamos que a quantia que estamos dispostos a apostar pode aproximar um grau de confiança, o raciocínio acima pode justifcar os axiomas da probabilidade.

2.5 Resultados equiprováveis

Considere o caso específico em que Ω é u conjunto finito e para todos os $w_1, w_2 \in \Omega$, $\mathbb{P}(\{w_1\}) = \mathbb{P}(\{w_2\})$. Isto é, a probabilidade de qualquer resultado possível é a mesma. Equiprobabilidade é um **caso especial** e não se aplica a todos os modelos de probabilidade.

Lema 6. *Se todos os resultados são equiprováveis, então para todo $w \in \Omega$, $\mathbb{P}(\{w\}) = |\Omega|^{-1}$.*

Demonstração. Note que $\Omega = \cup_{w\in\Omega}\{w\}$. Também observe que, se $w_1 \neq w_2$, então $\{w_1\}$ é disjunto de $\{w_2\}$.

Portanto,

$$\begin{aligned}
\sum_{w\in\Omega} \mathbb{P}(\{w\}) &= P(\Omega) && \text{(Lema 2)}\\
\sum_{w\in\Omega} \mathbb{P}(\{w\}) &= 1 && \text{(Definição 9)}\\
|\Omega| \cdot \mathbb{P}(\{w\}) &= 1 && \text{(Resultados equiprováveis)}\\
\mathbb{P}(\{w\}) &= |\Omega|^{-1}
\end{aligned}$$

□

Lema 7. *Para todo evento A, $\mathbb{P}(A) = \frac{|A|}{|\Omega|}$.*

Demonstração. Observe que $A = \cup_{w\in A}\{w\}$. Como os eventos são disjuntos,

$$\begin{aligned}
\mathbb{P}(A) &= \sum_{w\in A} \mathbb{P}(\{w\}) && \text{(Lema 2)}\\
&= \sum_{w\in A} |\Omega|^{-1} && \text{(Lema 6)}\\
&= \frac{|A|}{|\Omega|}
\end{aligned}$$

□

Isto é, em problemas com resultados equiprováveis, a probabilidade de um conjunto é proporcional ao tamanho deste. Para determinar tamanhos de conjuntos, alguns princípios de contagem são de grande ajuda. Nos vimos um pouco de análise combinatória no Capítulo 1, e a seguir, uma breve revisão no contexto dos exemplos que usamos:

- Suponha que temos n objetos distintos. O número de maneiras de que podemos selecionar k entre eles, com reposição, é n^k.

  Exemplo 16. *Considere um dado de seis lados. Este pode assumir 6 valores diferentes. Portanto, se lançarmos o dado 3 vezes seguidamente, podemos observar 6^3 resultados diferentes.*

- Suponha que temos n objetos distintos. Se a ordem de seleção importa, o número de maneiras de que você consegue selecionar $k \leq n$ destes objetos é $\frac{n!}{(n-k)!}$.

  Exemplo 17. *Temos n participantes num certo torneio de xadrez. De quantas formas podemos formar a lista dos top 3? $\frac{n}{(n-3)!}$.*

- Suponha que temos n objetos distintos. Se a ordem de seleção não importa, o número de maneiras de selecioar $k \leq n$ destes objetos é $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.

  Exemplo 18. *O número de mãos de poker possíveis é $\binom{52}{5}$.*

2.5.1 Exercícios

Exercício 27. *Suponha que lançamos uma moeda honesta 3 vezes.*

a. Qual é o espaço amostral?

b. *Qual é a probabilidade de cada resultado?*

c. *Que conjunto corresponde a "obter pelo menos duas caras"? Qual a probabilidade deste conjunto?*

Exercício 28.

a. *Quantos números telefônicos distintos de 7 dígitos existem? Assuma que o primeiro dígito de um telefone não pode ser* 0 *ou* 1.

b. *Se eu sortear (com probabilidade igual) um número dentre todos os números válidos, qual é a probabilidade de eu sortear o número* $893 - 9240$*?*

c. *E um número da forma* $893 - 924z$*, onde* z *pode ser qualquer algarismo entre* 0 *e* 8*?*

d. *E um número tal que quaisquer* 2 *dígitos são diferentes?*

Exercício 29. *Suponha que há* 4 *pessoas numa sala. Qual é a probabilidade de que pelo menos dois deles v"ao celebrar aniversário no mesmo mês? Assuma que as probabilidades de se fazer aniversário em cada um dos meses são iguais e que as datas de aniversário de duas pessoas são independentes entre si. Você consegue imaginar uma situação em que assumir independência não seja razoável?*

Exercício 30. *Lançamos um dado honesto de seis lados duas vezes. Considere que todos os resultados são equiprováveis. Seja* X *uma variável aleatória que denota o resultado do primeiro dado. Seja* Y *uma variável aleatória que denota o resultado do segundo dado. Encontre* $P(|X - Y| \in \{0, 1, 2\})$.

Exercício 31. *(Desafio) Jogador A tem n+1 moedas, enquanto o jogador B tem n moedas. Os dois jogadores lançam todas as suas moedas simultaneamente e observam o número de caras que cada um obtem. Assumindo que todas as moedas são honestas, qual é a probabilidade de que A tenha mais caras que B?*

Exercício 32 (Desafio)**.** *Uma torre pode "capturar" a outra se as duas estão na mesma linha ou coluna do tabuleiro de xandez. Se colocamos 8 torres no tabuleiro tal que posições diferentes são equiprováveis, qual é a probabilidade de que nenhuma torre possa capturar outra torre?*

Exercício 33 (Desafio)**.** *Dois mágicos fazem o seguinte truque: Uma pessoa escolhe ao acaso* 5 *cartas de um baralho honesto e passa estas para o Mágico 1. Mágico 1 dispõe as 5 cartas da seguinte forma. A primeira carta está virada para baixo. As outras quatro cartas estão viradas para cima, colocadas na ordem que ele escolhe. Mágico 2 ola para as cartas e adivinha qual é a carta que está virada para baixo. O que os mágicos podem combinar para que o mágico 2 sempre responda corretamente?*

2.6 Probabilidade Condicional

Probabilidade condicional é uma extensão da probabilidade. Sejam A e B dois eventos, a probabilidade condicional de A dado B é denotada por $\mathbb{P}(A|B)$. Interpretamos esta probabilidade como a incerteza sobre A **assumindo que** B é verdadeiro. Neste sentido, probabilidade comum é um caso particular de probabilidade condicional, pois Ω é sempre verdadeiro e, portanto, $\mathbb{P}(A) = \mathbb{P}(A|\Omega)$. Além dos usuais axiomas da probabilidade (Definição 9), probabilidades condicionais também satisfazem um axioma extra:

Definição 14 (Axioma das Probabilidades Condicionais)**.**

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

Isto é, a probabilidade de A e B ocorrerem simultaneamente é a mesma que a probabilidade de A acontecer multiplicada pela probabilidade de B acntecer dado que A aconteceu.

Exemplo 19. *Considere o resultado de um dado honesto de seis faces.* $\Omega = \{1, 2, 3, 4, 5, 6\}$ *e tods os elementos de* Ω *são equiprováveis. Seja* $Impar$ *o evento correspondente ao resultado impar,* $Impar = \{1, 3, 5\}$*. Observe que* $\mathbb{P}(Impar) = \frac{1}{2}$ *e* $\mathbb{P}(\{3\}) = \frac{1}{6}$*. Uando a definição 14,*

$$\mathbb{P}(\{3\} \cap Impar) = \mathbb{P}(Impar)\mathbb{P}(\{3\}|Impar)$$

Como $\{3\} \subset Impar$*,* $\{3\} \cap Impar = \{3\}$ *e*

$$\mathbb{P}(\{3\}) = \mathbb{P}(Impar)\mathbb{P}(\{3\}|Impar)$$

Substituindo os valores por $\mathbb{P}(\{3\})$ *e* $\mathbb{P}(Impar)$*, obtemos* $\mathbb{P}(\{3\}|Impar) = \frac{1}{3}$*.*

Exemplo 20. *Considere o resultado de um dado honesto de seis faces.* $\Omega = \{1, 2, 3, 4, 5, 6\}$ *e tods os elementos de* Ω *são equiprováveis. Seja* $A = \{3, 4, 5\}$ *e* $B = \{1, 2, 3, 4\}$*. Usando a definição 14,*

$$\begin{aligned}\mathbb{P}(A \cap B) &= \mathbb{P}(A|B)\mathbb{P}(B) \\ \mathbb{P}(\{3, 4\}) &= \mathbb{P}(A|B)\mathbb{P}(\{1, 2, 3, 4\})\end{aligned}$$

Ptanto, $\mathbb{P}(A|B) = \frac{2}{4}$*. Poemos também calcular* $\mathbb{P}(B|A)$*.*

$$\begin{aligned}\mathbb{P}(A \cap B) &= \mathbb{P}(B|A)\mathbb{P}(A) \\ \mathbb{P}(\{3, 4\}) &= \mathbb{P}(A|B)\mathbb{P}(\{3, 4, 5\})\end{aligned}$$

Isto é, $\mathbb{P}(B|A) = \frac{2}{3}$ *e, portanto,* $\mathbb{P}(A|B) \neq \mathbb{P}(B|A)$*.*

Exemplo 21. *Seja* $A \subset B$*. Usando a definição 14,*

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

Como $A \subset B$, $\mathbb{P}(A \cap B) = \mathbb{P}(A)$. Portanto, $\mathbb{P}(B|A) = 1$. Similarmente,

$$\begin{aligned} \mathbb{P}(A \cap B) &= \mathbb{P}(B)\mathbb{P}(A|B) \\ \mathbb{P}(A) &= \mathbb{P}(B)\mathbb{P}(A|B) \end{aligned}$$

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

Definição 15. *Dois eventos A e B são independentes se $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.*

Lema 8. *A e B são independentes se e somente se $\mathbb{P}(A|B) = \mathbb{P}(A)$. Em outras palavras, A e B se e somente se a incerteza sobre A não muda se assumirmos que B é verdadeiro.*

Demonstração. Assuma $\mathbb{P}(A|B) = \mathbb{P}(A)$. Da definição 14, lembramos que $\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B)$. Usando a suposição inicial, $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ e A e B são independentes.

Assuma A e B independentes. Portanto, $\mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B)$. Juntando os dois extremos da desigualdade e dividindo por $\mathbb{P}(B)$, $\mathbb{P}(A|B) = \mathbb{P}(A)$. □

Exemplo 22. *Seja $\Omega = \{(C, C), (C, K), (K, C), (K, K)\}$ e todos os resultados, equiprováveis. A denota "O resultado do primeiro lançamento da moeda é cara", $A = \{(C, C), (C, K)\}$. Seja B o evento "O resultado do segundo lançamento é cara", $B = \{(C, C), (K, C)\}$. Observe que $A \cap B = \{(C, C)\}$ e $\mathbb{P}(A \cap B) = \frac{1}{4}$. Além disso, $\mathbb{P}(A) = \mathbb{P}(B) = \frac{1}{2}$. Portanto, como $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$, A e B são independentes.*

Um **erro comum** é acreditar que podemos concuir que dois conjuntos disjuntos são independentes ou vice-versa. Lembre que A e B são disjuntos se $A \cap B = \emptyset$, isto é $\mathbb{P}(A \cap B) = 0$. Em português, A e B são disjuntos se é impossível para A e B ocorrer simultaneamente. Por outro lado A e B são independentes se $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Portanto, por exemplo, se $\mathbb{P}(A) > 0$ e $\mathbb{P}(B) > 0$ dois eventos não podem ser independentemente disjunos e independentes.

Definição 16. *Os eventos $A_1, A_2, \ldots A_n$ são conjuntamente independentes se, para todo $I \subset \{1, 2, \ldots, n\}$,*

$$\mathbb{P}(\cap_{i \in I} A_i) = \prod_{i \in I} \mathbb{P}(A_i)$$

Teorema 2 (Regra da multiplicação)**.** *Seam $A_1, A_2, \ldots A_n$ evnos. Então*

$$\mathbb{P}(\cap_{i=1}^{n} A_i) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2) \ldots \mathbb{P}(A_n| \cap_{i=1}^{n-1} A_i)$$

Observação: é comum denotar $\mathbb{P}(\cap_{i=1}^{n} A_i)$ por $\mathbb{P}(A_1, \ldots, A_n)$.

2.6.1 Exercícios

Exercício 34. *Sejam A e B dois eventos tais que $\mathbb{P}(A) = \frac{1}{2}$, $\mathbb{P}(B) = \frac{1}{3}$ e $\mathbb{P}(A \cap B) = \frac{1}{4}$. Encontre:*

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

b. $\mathbb{P}(A|B^c)$

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

d. $\mathbb{P}(A^c|B^c)$

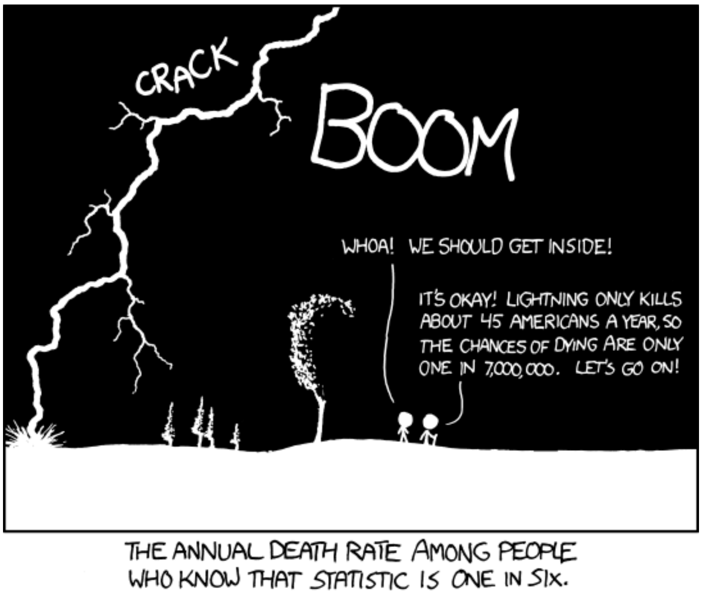


Figura 1

Exercício 35. *Explique como o quadrinho na Figura 1 de* http://xkcd.com/795/ *se relaciona com o conceito de probabilidade condicional.*

Exercício 36. *Considere que uma moeda seja lançada duas vezes. Eu anuncio que se observou cara pelo menos uma vez. Qual é a probabilidade de que foram 2 caras?*

Exercício 37. *A prescrição de um remédio contém as seguintes informações:*

- *Há 10% de chance de sentir dor de cabeça (evento H).*
- *Há 15% de chance de sentir nausea (evento N).*
- *Há 5% de chance de sentir dos dois efeitos colaterais.*

a. *Os eventos H e N são disjuntos? Porque?*

b. *Os eventos H e N são independentes? Porque?*

c. *Qal é a probabilidade de sentir pelo menos um dos dois efeitos colaterais?*

d. *Qual é a probabilidade de sentir exatamente um dos efeitos colaterais?*

e. *Qual é a probabilidade de não ter nem dor de cabeça, nem nausea?*

f. *Qual é a probabilidade de sentir dor de cabeça dado que você apresentou pelo menos um dos dois efeitos colaterais?*

Exercício 38. *Assma que uma carta é retirada de um baralho embaralhado. A denota que "a carta é um ás" e B denota que "o naipe da carta é ouros". Mostre que A e B são independentes.*

Exercício 39. *Duas cartas são retiradas de um baralho embaralhado:*

a. *Seja A o evento "o naipe da primeira carta é vermelho" e B o "o naipe da segunda carta é vermelho". Calcule $P(A \cap B)$. A e B são independentes?*

b. *Seja A o evento "o naipe da primeira carta é vermelho" e B o "a segunda carta é J, Q ou K". Clcule $P(A \cap B)$. A e B são independentes?*

Exercício 40. *Mostre que:*

1. *A e B são independentes se e somente se $\mathbb{P}(B|A) = \mathbb{P}(B)$.*
2. *A e B são independentes se e somente se $\mathbb{P}(A|B^c) = \mathbb{P}(A)$.*

Exercício 41.

a. *Se A é independente de B e B é independente de C, A e C são independentes? Porque?*

b. *Se A é ndependente de B, B é independente de C e A é independente de C, $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$?*

Exercício 42. *Prove o teorema 2.*

Exercício 43. *Uma moeda honesta é lançada 10 vezes. Qual a probabilidade de que não vão aparecer duas caras ou duas coroas consecutivamente?*

Exercício 44. *Seja A um evento. Mostre que $\mathbb{P}(\bullet|A)$ é uma medida de probabilidade.*

Exercício 45. *Seja X o resultado do lançamento de um dado de seis faces honesto. Encontre $\mathbb{P}((X-3)^2 = 1|X \neq 3)$.*

Exercício 46. *(Desafio) Jogador A tem n+1 moedas, enquanto o jogador B tem n moedas. Os dois jogadores lançam todas as suas moedas simultaneaente e observam o número de caras que cada um tem. Assumindo que todas as moedas são honestas, qual é a probabilidade de o número de caras nas moedas de A ser maior ou igual ao número de caras nas moedas de B? (Este problema é bastante difícil. Você não precisa saber como resolver este.)*

Exercício 47. *Prove que se $A_1, \ldots, A_n$ são independentes e a sua união é o espaço amostral, então $\mathbb{P}(A_i) = 1$ para pelo menos um i.*

2.7 A Lei da Probabilidade Total e o Teorema de Bayes

Exemplo 23. *Considere que há 2 moedas dentro de uma caixa. Você sabe que uma delas é honesta, e outra é viciada, com probabilidade $\frac{1}{3}$ de cara. Suponha que você retira ou a moeda honesta, ou a viciada, com a mesma probabilidade, lança esta moeda duas vzes e obtem duas caras.*

Denote por $C_{\frac{1}{2}}$ o evento em que você retira a moeda honesta, por $C_{\frac{1}{3}}$ o evento em que você retira a moeda viciada e por H_2 denote o evento em que você obtem duas caras. O problema diz que $\mathbb{P}(C_{\frac{1}{2}}) = \mathbb{P}(C_{\frac{1}{3}}) = \frac{1}{2}$. Também sabemos que $\mathbb{P}(H_2|C_{\frac{1}{2}}) = \frac{1}{4}$ e $\mathbb{P}(H_2|C_{\frac{1}{3}}) = \frac{1}{9}$. O que você acha sobre qual moeda que você escolheu depois que você realizou os lançamentos? Em outras palavras, você pode usar os valores de probabilidade anteriores e os axiomas da probabilidade para determinar $\mathbb{P}(C_{\frac{1}{2}}|H_2)$? E o valor de $\mathbb{P}(H_2)$?

A resposta é sim e, até o final da aula, nos seremos capazes de resolver este problema!

Nas aulas anteriores, demonstramos vários lemas simples que seguem dos axiomas da probabilidade. Nesta seção, nos usaremos estes lemas como tijolos para construir dois resultados gerais. Para apreciar integralmente a beleza dos resultados a seguir, tente por alguns minutos resolver da melhor forma que conseguir o Exemplo 23.

Teorema 3 (Lei da Probabilidade Total)**.** *Seja $(A_n)_{n\in\mathbb{N}}$ uma partição de Ω e B um evento.*

$$\mathbb{P}(B) = \sum_{n\in\mathbb{N}} \mathbb{P}(A_n)\mathbb{P}(B|A_n)$$

Demonstração. Oberve que

$$B = B \cap \Omega \tag{4}$$

Além disso, como $(A_n)_{n\in\mathbb{N}}$ é uma partição de Ω,

$$\Omega = \cup_{n\in\mathbb{N}} A_n \tag{5}$$

Usando a equação 5 na equação 4, obtemos:

$$\begin{aligned} B &= B \cap (\cup_{n\in\mathbb{N}} A_n) && (6)\\ &= \cup_{n\in\mathbb{N}}(B \cap A_n) && (7)\end{aligned}$$

e portanto,

$$\mathbb{P}(B) = \mathbb{P}(\cup_{n\in\mathbb{N}}(B \cap A_n)) \tag{8}$$

Como $(A_n)_{n\in\mathbb{N}}$ é uma partição, é também uma sequência disjunta. Portanto, para todo $i \neq j$, $A_i \cap A_j = \emptyset$ e

$$
\begin{aligned}
(B \cap A_i) \cap (B \cap A_j) &= B \cap (A_i \cap A_j) && (9)\\
&= B \cap \emptyset && (10)\\
&= \emptyset && (11)
\end{aligned}
$$

Iso é, $(B \cap A_n)_{n\in\mathbb{N}}$ também é uma sequência disjunta. Assim, dos axiomas da probabilidade (Definição 9),

$$\mathbb{P}(\cup_{n\in\mathbb{N}}(B \cap A_n)) = \sum_{n\in\mathbb{N}} \mathbb{P}(B \cap A_n) \tag{12}$$

Fnalmente, do axioma da probabilidade condicional (Definição 14), para todo $n \in \mathbb{N}$, $\mathbb{P}(B\cap A_n) = \mathbb{P}(A_n)\mathbb{P}(B|A_n)$. Uando as equações 8 e 12, concluimos que:

$$\mathbb{P}(B) = \sum_{n\in\mathbb{N}} \mathbb{P}(A_n)\mathbb{P}(B|A_n) \tag{13}$$

□

Lema 9. *Se $A_1, \ldots, A_n$ particiona Ω e B é um evento. $\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(A_n)\mathbb{P}(B|A_n)$.*

Demonstração. Se $A_1, \ldots, A_n$ paticiona Ω, a sequência $(C_i)_{i\in\mathbb{N}}$ tal que $C_i = A_i$ para todo $i \leq n$ e $C_i = \emptyset$ para $i > n$ também particiona Ω. A prova segue direto do Teorema 3 e $\mathbb{P}(\emptyset) = 0$. □

resultado acima permite relacionar probabilidade não condicional $\mathbb{P}(B)$ às probabilidades condicionais $\mathbb{P}(B|A_n)$.

Exemplo 24 (Continuação do exemplo 23)**.** *O problema descrito nos dá a probabilidade de obtermos cara uma vez que se sorteou a moeda, isto é, $\mathbb{P}(H_2|C_{\frac{1}{2}})$ e $\mathbb{P}(H_2|C_{\frac{1}{3}})$. Não obstante, ele não nos dá diretamente a probabilidade de obter H_2 sem saber qual moeda foi sorteada, $\mathbb{P}(H_2)$. Oberve que $C_{\frac{1}{2}}, C_{\frac{1}{3}}$ pariciona Ω e, portanto, segue do lema 9 que*

$$
\begin{aligned}
\mathbb{P}(H_2) &= \mathbb{P}(C_{\frac{1}{2}})\mathbb{P}(H_2|C_{\frac{1}{2}}) + \mathbb{P}(C_{\frac{1}{3}})\mathbb{P}(H_2|C_{\frac{1}{3}})\\
&= \frac{1}{2} \cdot \frac{1}{4} + \frac{1}{2} \cdot \frac{1}{9} = \frac{13}{72}
\end{aligned}
$$

Simples, não? Estamos começando a nos beneficiar de ter construído uma teoria rigorosa em vez de confiar só na intuição.

Teorema 4 (Teorema de Bayes)**.** *Seja $(A_i)_{i\in\mathbb{N}}$ uma partição de Ω e B um evento. Para todo $n \in \mathbb{N}$,*

$$\mathbb{P}(A_n|B) = \frac{\mathbb{P}(A_n)\mathbb{P}(B|A_n)}{\sum_{i\in\mathbb{N}} \mathbb{P}(A_i)\mathbb{P}(B|A_i)}$$

Demonstração. Lembramos do axioma da probabilidade condicional (definição 14) que

$$\mathbb{P}(A_n|B) = \frac{\mathbb{P}(A_n \cap B)}{\mathbb{P}(B)} \tag{14}$$

Usando o axioma da probabilidade condicional novamente, $\mathbb{P}(A_n \cap B) = \mathbb{P}(A_n)\mathbb{P}(B|A_n)$. Usando a equação 14,

$$\mathbb{P}(A_n|B) = \frac{\mathbb{P}(A_n)\mathbb{P}(B|A_n)}{\mathbb{P}(B)} \tag{15}$$

Finalmente, observe que $(A_i)_{i\in\mathbb{N}}$ é uma partição de Ω e, portanto, usando o Teorema 3, $\mathbb{P}(B) = \sum_{n\in\mathbb{N}} \mathbb{P}(A_n)\mathbb{P}(B|A_n)$. Colocando este valor na equação 15

$$\mathbb{P}(A_n|B) = \frac{\mathbb{P}(A_n)\mathbb{P}(B|A_n)}{\sum_{n\in\mathbb{N}} \mathbb{P}(A_n)\mathbb{P}(B|A_n)} \tag{16}$$

□

Lema 10. *Seja $A_1, \ldots, A_n$ partição de Ω e B um evento.*

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(A_i)\mathbb{P}(B|A_i)}{\sum_{i=1}^{n} \mathbb{P}(A_i)\mathbb{P}(B|A_i)}$$

Demonstração. Siga os passos do Teorema 4. No último passo, use Lema 9 em vez do Teorema 3. □

Exemplo 25 (Continuação do 23)**.** *A descrição do problema nos dá a probabilidade de obter duas caras dado o viés da moeda, $\mathbb{P}(H_2|C_{\frac{1}{2}}) = \frac{1}{4}$ e $\mathbb{P}(H_2|C_{\frac{1}{3}}) = \frac{1}{9}$. É menos intuitivo descobrir o que aprendemos sobre o viés da moeda ao observar duas caras, $\mathbb{P}(C_i|H_2)$. O Teorema de Bayes nos dá a resposta.*

Lembramos que $C_{\frac{1}{2}}, C_{\frac{1}{3}}$ éuma partição de Ω. Portanto, pelo Lema 10,

$$\begin{aligned}
\mathbb{P}(C_{\frac{1}{2}}|H_2) &= \frac{\mathbb{P}(C_{\frac{1}{2}})\mathbb{P}(H_2|C_{\frac{1}{2}})}{\mathbb{P}(C_{\frac{1}{2}})\mathbb{P}(H_2|C_{\frac{1}{2}}) + \mathbb{P}(C_{\frac{1}{3}})\mathbb{P}(H_2|C_{\frac{1}{3}})} \\
&= \frac{\frac{1}{2}\cdot\frac{1}{4}}{\frac{1}{2}\cdot\frac{1}{4} + \frac{1}{2}\cdot\frac{1}{9}} = \frac{9}{13}
\end{aligned}$$

2.7.1 Exercícios

Exercício 48. *Assuma que 5% dos homens e 0.25% das mulheres são daltônicos. Assuma que uma pessoa aleatória é selecionada e ele/ela é daltônico. Asmindo que há o mesmo número de homens e mulheres, qual a probabilidade de que esta pessoa seja um homem?*

Exercício 49. *Assuma que há 4 moedas honestas numa bolsa. Uma pessoa retira algumas moedas com igual probabilidade entre $\{1, 2, 3, 4\}$. A seguir,a pessoa lança todas as moedas que pegou. Qual é a probabilidade de que todas as moedas deem coroa, sem saber quantas moedas foram retiradas da bolsa?*

Exercício 50. *Histórias de casos médicos indicam que diferentes doenças podem resultar em sintomas identicos. Suponha que um certo conjunto de sintomas, H, ocorre somente quando uma das três doenças, I_1, I_2 ou I_3 ocorre. Assuma que a ocorrência simultânea de mais de uma destas doenças é impossível. Também assuma que:*

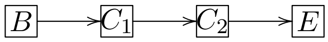
a. $\mathbb{P}(I_1) = 0.01$; $\mathbb{P}(I_2) = 0.05$; $\mathbb{P}(I_3) = 0.02$.

b. $\mathbb{P}(H|I_1) = 0.90$; $\mathbb{P}(H|I_2) = 0.95$; $\mathbb{P}(H|I_3) = 0.75$.

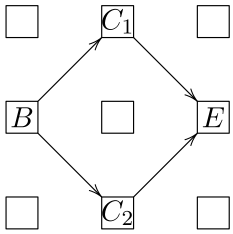
Assumindo que um doente mostra os sintomas H, qual é a probabilidade de que esta pessoa tenha a doença I_1?

Exercício 51. *Considere que cada componente $C_1, \ldots, C_n$ em um sistema falha com probabilidade p. Um sistema falha se não há um caminho operacional de B para E. Qual é a probabilidade de o sistema a seguir falhar?*

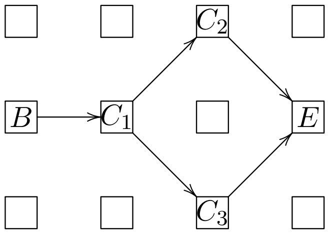
a. *Sistema em série:*



b. *Sistema em paralelo:*



c. *Sistema mixo:*



Exercício 52. *(Monty Hall) Temos 3 portas: A, B e C. Há um carro atrás de uma destas portas. Se você escolher a porta certa, você ganha o carro. A primeira vista, você acredita que é igualmente provável que o carro esteja atrás de qualquer uma das portas. Você escolhe a porta A. Em seguida, o apresentador do show abre a porta B, mostra que não há nada atrás desta e permite que você troque a porta que você escolheu para a C. É uma boa ideia trocar? Assuma que o apresentador sempre vai abrir a porta sem prêmio e oferecer a chance de trocar de porta, independente de você ter escolhido inicialmente a porta com prêmio ou sem prêmio.*

Exercício 53. *(Urna de Polya) Considere que uma urna tem 1 bola preta e 1 ola branca. Cada vez que você sorteia uma bola, você deve colocar ela de volta à urna e adicionar uma bola extra da mesma cor. Qual é a probabilidade de você tirar uma bola branca no seu 3ro sorteio?*

Exercício 54. *Você tem 12 bolas vermelhas, 12 amarelas e 2 urnas. Considere que as bolas são distribuídas entre as urnas, uma das urnas é selecionada ao acaso (as probabilidades de que você escolha cada uma das urnas são iguais) e uma bola é retirada da urna selecionada.*

a. *Assuma que você colocou 2 boas vermelhas e 4 bolas amarelas na primeira urna, e o restante, na segunda urna. Qual é a probabilidade de retirar uma bola amarela?*

b. *(Desafio) Qual distribuição das bolas entre as urnas maximiza a sua probabilidade de retirar uma bola amarela?*

Exercício 55. *Teoria das Probabilidades foi usada no caso criminal Povo v. Collins. Neste caso, a carteira de uma mulher foi roubada. Testemunhas afirmaram que um casal fugindo da cena do crime estava composto de um homem negro com barba e bigode e uma garota loira com rabo de cavalo. Testemunhas também disseram que o casal foi embora num carro amarelo. Malcolm e Janet Collins satisfaziam a descrição apresentada anteriormente.*

Um professor e matemática tabém testemunhou que, se um casal fosse escolhido ao acaso, teriamos as seguintes probabilidades:

Evento	Probabilidade
Homem com bigode	1/4
Garota loira	1/3
arotacom rabo de cavalo	1/10
Homem negro com barba	1/10
Casal interracial de carro	1/1000
Caro parcialmente amarelo	1/10

Portanto, a probabilidade de todas estas características seja observadas num casal escolhido ao acaso é 1 em $12,000,000$. A acusação afirmou que isso constitui uma prova além de qualqer dúvida razoável de que os acusados são culpados. Você acha que este argumento é convincente? Porque?

Exercício 56. *Prove que se* $(A_n)_{n\in\mathbb{N}}$ *e* $(B_n)_{n\in\mathbb{N}}$ *são duas sequências de eventos tais que* $\lim_{n\longrightarrow\infty}\mathbb{P}(A_n)=1$ *e* $\lim_{n\longrightarrow\infty}\mathbb{P}(B_n)=p$*, então* $\lim_{n\longrightarrow\infty}\mathbb{P}(A_n\cap B_n)=p$

3 Variáveis Aleatórias

3.1 Distribuição de variáveis aleatórias

Lembramos que uma variável aleatória é um número desconhecido, isto é, uma função de Ω para $\mathbb{R}$ (Definição 12). Uma variável aleatória discreta é uma variável aleatória que assume somente uma quantidade enumeravel de valores.

Exemplo 26. *Considere que uma pessoa lança uma moeda honesta uma vez. Os resultados possíveis são cara ou coroa e* $\Omega = \{C, K\}$. *Seja* X *o número observado de caras. Isto é,* $X(C) = 1$ *and* $X(H) = 0$.

Exemplo 27. *Suponha que uma pessoa lança uma moeda honesta até obter a primeira cara ou até fazer* 3 *lançamentos,* $\Omega = \{C, KC, KKC, KKK\}$. *Denote por* X *o número de lançamentos realizados.* X *é uma variável aleatória tal que* $X(C) = 1$, $X(KC) = 2$, $X(KKC) = 3$ *e* $X(KKK) = 3$. X *é discreta é assume um número finito de valores,* 1, 2 *ou* 3.

Exemplo 28. *Uma moeda honesta é lançada até se obter cara,* $\Omega = \{C, KC, KKC, KKKC, \ldots\}$. *Denote por* X *o número de lançamentos realizados.* X *é uma variável aleatória discreta e pode assumir um número infinito e enumerável de valores.* X *pode assumir qualquer valor em* $\mathbb{N} - \{0\}$. *Por exemplo,* $X(C) = 1, X(KC) = 2, X(KKC) = 3, \ldots$

Definição 17. *Seja* X *uma variável aleatória. Para* $x \in \mathbb{R}$, *definimos* $p_X(x) = \mathbb{P}(X = x) = \mathbb{P}(\{w \in \Omega : X(w) = x\})$. *Chamamos a função* $p_X : \mathbb{R} \to [0, 1]$ *a função de massa de probabilidade (fmp) ou* X.

Observação: O conjunto $\{w \in \Omega : X(w) = x\}$ é frequentemente denotado por $X^{-1}(\{x\})$.

Exemplo 29. *Considere o exemplo 26.*

$$p_X(0) = \mathbb{P}(X = 0) = \mathbb{P}(\{w \in \Omega : X(w) = 0\}) = \mathbb{P}(\{C\}) = 0.5$$
$$p_X(1) = \mathbb{P}(X = 1) = \mathbb{P}(\{w \in \Omega : X(w) = 1\}) = \mathbb{P}(\{K\}) = 0.5$$

Não é legal escrever $p_X(1)$ *em vez de* $\mathbb{P}(\{w \in \Omega : X(w) = 1\})$*? Este é o poder de uma boa notação...*

Exemplo 30. *Considere o exemplo 27. Observe que os elementos de* Ω *não são equiprováveis. Consequentemente, mesmo que* $\{C\}$ *e* $\{KC\}$ *sejam conjuntos unitários, suas probabilidades diferem.*

Com o propósito de obter as probabilidades dos subconjuntos unitários de Ω, *defina o evento* C_i *como "o i-ésimo lançamento resulta em cara". Observe que todos os* C_i *são conjuntamente independentes e, como a moeda é honesta,* $\mathbb{P}(C_i) = 0.5$. *Consequentemente,*

$$\begin{aligned}
\mathbb{P}(\{C\}) &= \mathbb{P}(C_1) = 0.5 \\
\mathbb{P}(\{KC\}) &= \mathbb{P}(C_1^c \cap C_2) = \mathbb{P}(C_1^c)\mathbb{P}(C^2) = 0.25 \\
\mathbb{P}(\{KKC\}) &= \mathbb{P}(C_1^c \cap C_2^c \cap C_3) = \mathbb{P}(C_1^c)\mathbb{P}(C_2^c)\mathbb{P}(C_3) = 0.125 \\
\mathbb{P}(\{KKK\}) &= \mathbb{P}(C_1^c \cap C_2^c \cap C_3^c) = \mathbb{P}(C_1^c)\mathbb{P}(C_2^c)\mathbb{P}(C_3^c) = 0.125
\end{aligned}$$

Assim, podemos calcular a pmf de X,

$$p_X(1) = \mathbb{P}(X = 1) = \mathbb{P}(\{w \in \Omega : X(w) = 1\}) = \mathbb{P}(\{H\}) = 0.5$$
$$p_X(2) = \mathbb{P}(X = 2) = \mathbb{P}(\{w \in \Omega : X(w) = 2\}) = \mathbb{P}(\{TH\}) = 0.25$$
$$p_X(3) = \mathbb{P}(X = 3) = \mathbb{P}(\{w \in \Omega : X(w) = 3\}) = \mathbb{P}(\{TTH, TTT\}) = 0.25$$

Exemplo 31. *Considere o Examplo 28. Observe que novamente os elementos de Ω não são equiprováveis. Defina o evento H_i como "o i-ésimo lançamento da moeda resulta em cara". Observe que todos H_i são conjuntamente independentes e, uma vez que a moeda é não viesada, $\mathbb{P}(H_i) = 0.5$. Assim,*

$$p_X(1) = \mathbb{P}(X = 1) = \mathbb{P}(\{w \in \Omega : X(w) = 1\}) = \mathbb{P}(H_1) = 0.5$$
$$p_X(2) = \mathbb{P}(X = 2) = \mathbb{P}(\{w \in \Omega : X(w) = 2\}) = \mathbb{P}(H_1^c \cap H_2) = 0.25$$
$$\ldots$$
$$p_X(n) = \mathbb{P}(X = n) = \mathbb{P}(\{w \in \Omega : X(w) = n\}) = \mathbb{P}(H_1^c \cap \ldots H_{n-1}^c \cap H_n) = \frac{1}{2^n}$$
$$\ldots$$

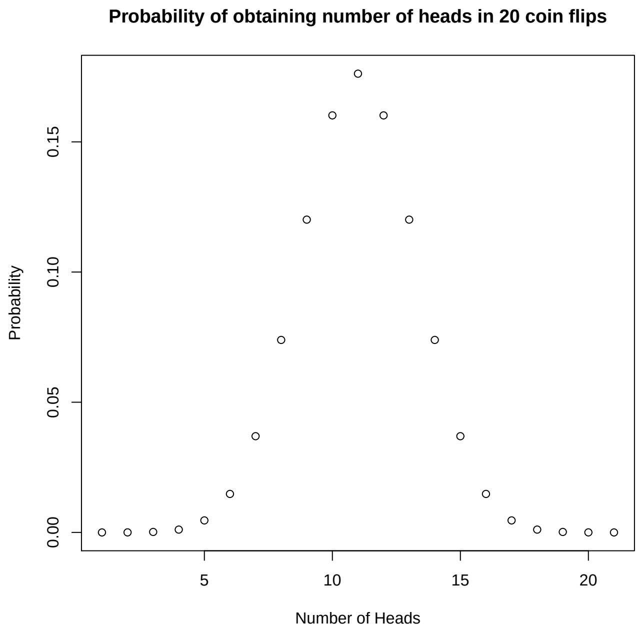


Figura 2

Lema 11. *Seja X uma variável aleatória discreta e p_X a fmp de X. Seja χ os valores possíveis de X.*

- *Para todo $x \in \chi$, $0 \leq p_X(x) \leq 1$.*
- *$\sum_{x \in \chi} p_X(x) = 1$.*

Demonstração.

- $p_X(x) = P(X = x) = P(\{w \in \Omega : X(w) = x\})$. Lembre que $0 \leq P(\{w \in \Omega : X(w) = x\}) \leq 1$.

- $\sum_{x\in\chi} p_X(x) = \sum_{x\in\chi} P(\{w \in \Omega : X(w) = x\})$. Observe que, para $x_1 \neq x_2$, os eventos $\{w \in \Omega : X(w) = x_1\}$ e $\{w \in \Omega : X(w) = x_2\}$ são disjuntos. Portanto, pela aditividade enumerável,

$$\sum_{x\in\chi} p_X(x) = P(\cup_{x\in\chi}\{w \in \Omega : X(w) = x\}) = P(\Omega) = 1$$

□

Existem muitas formas de representar uma fmp. Por exemplo, uma formula, uma tabela de valores ou simplesmente uma lista dos valores. Uma ferramenta útil de visualização é plotar os valores de probabilidade em função de χ, conforme mostrado na Figura 2.

Quaisquer probabilidades associadas a X podem ser calculadas usando a fmp. Podemos também usar *funções de distribuição acumulada*, i.e., $F_X(x) := \mathbb{P}(X \leq x)$, contudo deixamos isso para as variáveis aleatórias contínuas (Chapter 5).

Definição 18. *Seja $X_1, \ldots, X_n$ variáveis aleatórias discretas. Dizemos que elas são independentes se, para todo $x_1, \ldots, x_n \in \mathbb{R}$,*

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) := \mathbb{P}(\cap_{i=1}^n X_i = x_i) = \prod_{i=1}^n \mathbb{P}(X_i = x_i)$$

Portanto, para todo $x_1, \ldots, x_n$, $\{\omega \in \Omega : X_i(\omega) = x_i\}$ são conjuntamente independentes.

3.1.1 Exercícios

Exercício 57. *Seja X uma variável aleatória tal que $X \in \mathbb{N}$ e $p_X(i) = c \cdot 2^{-i}$. Encontre c.*

Exercício 58. *Considere que você lança 3 vezes uma moeda com probabilidade p de cara. Seja X o número de caras observadas. Encontre a fmp de X.*

Exercício 59. *Você lança um dado de quatro lados duas vezes. X_1 e X_2 deotam, respectivamente, os resultados do primeiro e do segundo lançamento. Seja $Y = X_1 + X_2$. Encontre a fmp de Y e de $2X_1$. Observe que a fmp da soma dos dois resultados é diferente da fmp da multiplicação do resultado de um dado por dois.*

Exercício 60. *Considere que você escolhe uma moeda, com probabilidades iguais, entre uma moeda honesta e uma moeda com probabilidade p de dar cara. A seguir, você lança a moeda escolhida 3 vezes. Seja X o número de caras que você observa. Encontre a fmp de X.*

Exercício 61. *Uma caixa tem 4 bolas azuis e 4 bolas vermelhas.*

a. Voê sorteia 3 bolas com reposição e X denota o número de bolas azuis sorteadas. Encontre a fmp de X.

b. Você sorteia 3 bolas sem reposição e Y deota o número de bolas azuis sortadas. Encontre a fmp de Y.

c. Descreva a diferença entre as fmp's de X e Y.

Exercício 62. *Sejam X e Y duas variáveis disretas independentes. e^X e e^Y são independentes?*

Exercício 63.

a. Sejam X e Y das variáveis aleatórias independentes. Mostre que, para quaisquer funções f e g, $f(X)$ é independente de $g(X)$.

b. Mostre um exemplo de variáveis aleatórias discretas X, Y e funções f e g tas que $f(X)$ é independente de $g(X)$ mas X não é independente de Y.

3.2 Valor Esperado (Condicional)

Definição 19. *Seja X uma variável aleatória discreta e A evento. O valor esperado de X dado A e denotado por $\mathbb{E}[X|A]$ e*

$$\mathbb{E}[X|A] = \sum_{w \in \Omega} X(w)\mathbb{P}(\{w\}|A)$$

Definição 20. *Seja X uma variável aleatória discreta. O valor esperado de X é denotado por $\mathbb{E}[X]$ e é igual a $\mathbb{E}[X|\Omega]$. Isto é,*

$$\mathbb{E}[X] = \sum_{w \in \Omega} X(w)\mathbb{P}(\{w\})$$

Lema 12 (Lei do estatístico inconsciente)**.** *Seja X uma variável aleatória discreta com fmp p_X e que assume valores em χ:*

$$\mathbb{E}[f(X)] = \sum_{x \in \chi} f(x) \cdot p_X(x)$$

Demonstração.

$$\begin{aligned}
\mathbb{E}[f(X)] &= \sum_{w \in \Omega} f(X(w))\mathbb{P}(\{w\}) \\
&= \sum_{x \in \chi} \sum_{w: X(w) = x} f(X(w))\mathbb{P}(\{w\}) \\
&= \sum_{x \in \chi} f(x) \sum_{w: X(w) = x} \mathbb{P}(\{w\}) \\
&= \sum_{x \in \chi} f(x) \cdot p_X(x)
\end{aligned}$$

□

Em particular, obseve que segue do Lema 12 que $\mathbb{E}[X] = \sum_{x \in \chi} x \cdot p_X(x)$. Isso é, $\mathbb{E}[X]$ é a média ponderada dos valores de χ usando p_X como pesos.

Exemplo 32 (Continuação do Exemplo 26)**.**

$$\mathbb{E}[X] = 0 \cdot p_X(0) + 1 \cdot p_X(1) = 1 \cdot 0.5 = 0.5$$

Exemplo 33 (Continuação do Exemplo 27)**.**

$$\begin{aligned}
\mathbb{E}[X] &= 1 \cdot p_X(1) + 2 \cdot p_X(2) + 3 \cdot p_X(3) \\
&= 1 \cdot 0.5 + 2 \cdot 0.25 + 3 \cdot 0.25 = 1.75
\end{aligned}$$

Lema 13. *Seja X umavariável aleatória discreta tal que $X \in \mathbb{N}$,*

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} \mathbb{P}(X \geq i)$$

Demonstração.

$$\mathbb{E}[X] = \sum_{j=0}^{\infty} j \cdot p_X(j) = \sum_{j=1}^{\infty} \sum_{i=1}^{j} p_X(j) = \sum_{i=1}^{\infty} \sum_{j=i}^{\infty} p_X(j) = \sum_{i=1}^{\infty} P(X \geq i)$$

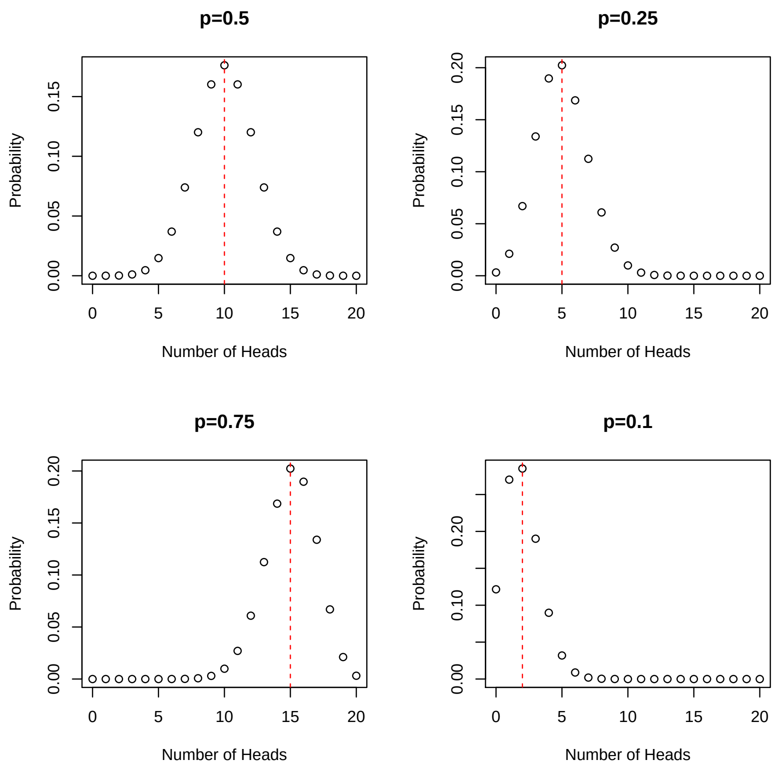


Figura 3: fmp's para o número de caras em 20 ançamentos de moeda com probabilidade p de caras. A linha vermelha pontilhada é o valor esperado.

□

Lema 14. *Linearidade do Valor Esperado*

$$\mathbb{E}\left[\sum_{i=1}^{n} c_i X_i \middle| A\right] = \sum_{i=1}^{n} c_i \mathbb{E}[X_i \big| A]$$

Demonstração.

$$\begin{aligned}
\mathbb{E}\left[\sum_{i=1}^{n} c_i X_i \middle| A\right] &= \sum_{w \in \Omega} \sum_{i=1}^{n} c_i X_i(w) \mathbb{P}(\{w\} \big| A) \\
&= \sum_{i=1}^{n} c_i \sum_{w \in \Omega} X_i(w) \mathbb{P}(\{w\} \big| A) \\
&= \sum_{i=1}^{n} c_i \mathbb{E}[X_i \big| A]
\end{aligned}$$

□

Exemplo 34 (20 lançamentos de moeda)**.** *Considere que você lança* 20 *moedas com probabilidade* p *de cara e conta o número de caras. A Figura 3 mostra as fmp's para diferentes valores de* p*. Suas médias correspondem a linhas verticais vrmelhas.*

Lema 15 (Lei da esperança total)**.** *Seja $A_1, \dots, A_n$ uma partição de Ω e X uma v.a. discreta.*

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X|A_i] \cdot \mathbb{P}(A_i)$$

Demonstração.

$$\begin{aligned}
\sum_{i=1}^{n} \mathbb{E}[X|A_i] \cdot \mathbb{P}(A_i) &= \sum_{i=1}^{n} \sum_{w \in \Omega} X(w) \mathbb{P}(\{w\}|A_i) \mathbb{P}(A_i) \\
&= \sum_{i=1}^{n} \sum_{w \in \Omega} X(w) \mathbb{P}(\{w\} \cap A_i) && \text{(Definição 14)} \\
&= \sum_{w \in \Omega} X(w) \sum_{i=1}^{n} \mathbb{P}(\{w\} \cap A_i) \\
&= \sum_{w \in \Omega} X(w) \mathbb{P}\left(\bigcup_{i=1}^{n} (\{w\} \cap A_i) \right) && \text{(Lema 2 and } A_i \text{ disjoint)} \\
&= \sum_{w \in \Omega} X(w) \mathbb{P}(\{w\}) = \mathbb{E}[X] && \text{(Como } (A_i)_{i=1}^n \text{ particiona } \Omega, \cup_{i=1}^n A_i = \Omega)
\end{aligned}$$

□

3.2.1 Exercícios

Exercício 64. *Lembre que I_A denota a função indicadora (definição 13) do evento A. Se $\mathbb{P}(A) = 0.2$, $\mathbb{P}(B) = 0.9$ e $\mathbb{P}(A \cap B) = 0.1$,*

a. Calcule a fmp de $10I_A + 5I_B$.

b. Calcule $\mathbb{E}[10I_A + 5I_B]$.

Exercício 65. *Considere que você lança 200 vezes uma moeda com probabilidade p de ocorrer cara. Seja X o número total de caras observadas. Determine $\mathbb{E}[X]$.*

Dica: Denote por H_i o evento em que o i-ésimo lançamento resulta em cara. Observe que $X = \sum_{i=1}^{200} I_{H_i}$.

Exercício 66. *Seja X o número de caras observadas em dois lançamentos de uma moeda honesta. Determine $\mathbb{E}[X^2]$ e $\mathbb{E}[X]^2$.*

Exercício 67. *Seja $X_1, \dots, X_n$ uma variável aleatória discreta tal que, para todo $i \in \{1, \dots, n\}$, $\mathbb{E}[X_i] = \mu \in \mathbb{R}$.*

a. Seja $p_i \geq 0$ tal que $\sum_{i=1}^n p_i = 1$. Encontre, $\mathbb{E}[\sum_{i=1}^n p_i X_i]$.

b. Seja $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$, comunmente chamado "média amostral". Encontre $\mathbb{E}[\bar{X}]$.

Exercício 68. *Considere que você escolhe entre um dado de quatro lados e um dado de seis lados com probabilidade igual. A seguir, você lança o dado escolhido 1000 vezes. Seja X a soma dos resultados dos 1000 lançamentos de dado. Calcule $\mathbb{E}[X]$.*

Exercício 69. *Considere uma urna com bolas numeradas de 1 a n. Se amostramos uma bola por vez com resposição, qual é a fmp da variável aleatória X : número de bolas retiradas até que a mesma bola é retirada duas vezes pela primeira vez? Qual é a a sua esperança? Dica: use o Lema 13.*

Exercício 70. *Suponha que você tem seis chaves. Você não tem certeza de qual delas abre a porta que você precisa abrir, por isso você começa a testar cada uma das chaves. Qual é o número médio de chaves que você precisa testar até abrir a porta?*

Exercício 71. *(Desafio) Na iteração* 1*, uma colônia de bactérias tem uma única bactéria. A cada iteração nova, uma bactéria na colônia pode ou morrer (com probabilidade* $1-p$*) ou se dividir em das bactérias (com probabilidade* p*). Seja* X_i *o úmero de bactérias na colônia na iteração* i*.*

a. *Encontre* $\mathbb{E}[X_1]$.

b. *Encontre* $\mathbb{E}[X_n|X_1 = 0]$ *e* $\mathbb{E}[X_n|X_1 = 2]$.

c. *Encontre* $\lim_{n\to\infty} \mathbb{E}[X_n]$.

Exercício 72. *Prove que:*

- *Se, para todo* $\omega \in \Omega$, $X(\omega) = c$ *para algum* $c \in \Re$, *então* $\mathbb{E}[X] = c$.
- *Se, para todo* $\omega \in \Omega$, $0 \leq X(\omega) = Y(\omega)$, *enão* $0 \leq \mathbb{E}[X] \leq \mathbb{E}[Y]$

3.3 Variância

Seja X uma variável aleatória. Na subseção anterior, nos definimos o valor esperado de X, $\mathbb{E}[X]$. Vimos que, intuitivamente, $\mathbb{E}[X]$ é o valor central em torno do qual se dispersam os possíves valores de X.

Nesta subseção apresentamos a variância de X, $Var[X]$. A variância de X é uma medida da concentração dos possíveis valores de X em torno de $\mathbb{E}[X]$.

Definição 21 (Variância). *A variância de uma v.a. discreta X 'definida como $\mathbb{E}[(X - \mathbb{E}[X])^2]$ e deotada por $Var[X]$.*

Exemplo 35. *Seja H evento em que o resultado de um lançamento de uma moeda é cara. Considere que $\mathbb{P}(H) = p$. Vamos calcular $Var[I_H]$. Observe que $p_{I_H}(0) = 1 - p$ e $p_{I_H}(1) = p$. Portanto,*

$$\mathbb{E}[I_H] = 0 \cdot (1 - p) + 1 \cdot p = p$$

Portanto, $Var[I_H] = E[(I_H - p)^2]$. If $I_H = 0$, então $(I_H - p)^2 = p^2$ e if $I_H = 1$, logo $(I_H - p)^2 = (1 - p)^2$. Concluimos que

$$\begin{aligned}
Var[I_H] &= \mathbb{E}[(I_H - p)^2] \\
&= (1 - p) \cdot p^2 + p \cdot (1 - p)^2 \\
&= p(1 - p)(p + 1 - p) = p(1 - p)
\end{aligned}$$

Observe que $p(1 - p)$ é uma parábola com raízes 0 e 1 e que ela assume seu valor máximo em 0.5. Em outras palavras, a variância de I_H é minimizada se $p = 0$ ou $p = 1$, pois nestes casos nos sabemos ao certo se iremos observar cara ou não. A variância é maximizada para uma moeda honesta e, neste sentido, os lançamentos de uma moeda honesta oscilam mais em torno do valor esperado.

Exemplo 36. *Seja X o número de caras observadas em dois lançamentos de uma moeda com probabilidade p de cara. Seja H_i o evento em que o i-ésimo lançamento resulta em cara. Vamos calcular $Var[X]$. A fmp de X é:*

$$\begin{aligned}
p_X(0) &= \mathbb{P}(\{TT\}) = \mathbb{P}(H_1^c \cap H_2^c) = (1 - p)^2 \\
p_X(1) &= \mathbb{P}(\{TH, HT\}) = \mathbb{P}(H_1^c \cap H_2) + \mathbb{P}(H_1 \cap H_2^c) = 2p(1 - p) \\
p_X(2) &= \mathbb{P}(\{HH\}) = \mathbb{P}(H_1 \cap H_2) = p^2
\end{aligned}$$

Ptanto,

$$\mathbb{E}[X] = 0 \cdot (1 - p)^2 + 1 \cdot 2p(1 - p) + 2 \cdot p^2 = 2p$$

Assim, $Var[X] = \mathbb{E}[(X - 2p)^2]$ e

$$\begin{aligned}
\mathbb{E}[(X - 2p)^2] &= (0 - 2p)^2 \cdot (1 - p)^2 + (1 - 2p)^2 \cdot 2p(1 - p) + (2 - 2p)^2 \cdot p^2 \\
&= 2p(1 - p) \cdot (2p(1 - p) + (1 - 2p)^2 + 2(1 - p)p) = 2p(1 - p)
\end{aligned}$$

Isto é, a variância de X é uma parábola com mínimos em $p = 0$ e $p = 1$ e máximo em $p = 0.5$.

Lema 16. $Var[aX + b] = a^2 Var[X]$

Demonstração.

$$\begin{aligned}
Var[aX+b] &= \mathbb{E}[(aX+b-\mathbb{E}[aX+b])^2] && \text{(Definição 21)}\\
&= \mathbb{E}[(aX+b-a\mathbb{E}[X]+b)^2] && \text{(Lema 14)}\\
&= \mathbb{E}[a^2(X-\mathbb{E}[X])^2]\\
&= a^2\mathbb{E}[(X-\mathbb{E}[X])^2] = a^2Var[X]
\end{aligned}$$

□

O lema seguinte requentemente pode ser útil para cálculo de variâncias:

Lema 17.

$$Var[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

Demonstração.

$$\begin{aligned}
Var[X] &= \mathbb{E}[(X-E[X])^2] && \text{(Definição 21)}\\
&= \mathbb{E}[X^2 - 2\mathbb{E}[X]X + \mathbb{E}[X]^2]\\
&= \mathbb{E}[X^2] - 2\mathbb{E}[X]\mathbb{E}[X] + \mathbb{E}[X]^2 && (\mathbb{E}[X]\in\mathbb{R}, \text{ Lema 14})\\
&= \mathbb{E}[X^2] - \mathbb{E}[X]^2
\end{aligned}$$

□

Exemplo 37. *No Exemplo 36, temos*

$$\mathbb{E}[X^2] = 0^2\cdot(1-p)^2 + 1^2\cdot 2p(1-p) + 2^2\cdot p^2 = 2p^2+2p.$$

Segue do Lema 17 que

$$Var[X] = 2p^2+2p-(\mathbb{E}[X])^2 = 2p^2+2p-(2p)^2 = 2p(1-p),$$

que bate com o valor encontrado no Exemplo 36.

Exemplo 38. *Cnsidere dois lançamentos de uma moeda com probabilidade p de cara. Seja H_i se i-ésimo lançamento é cara. Vamos mostrar que I_{H_1} e I_{H_2} são independentes,*

$$\begin{aligned}
\mathbb{P}(I_{H_1}=0\cap I_{H_2}=0) &= (1-p)^2 = \mathbb{P}(I_{H_1}=0)\mathbb{P}(I_{H_2}=0)\\
\mathbb{P}(I_{H_1}=0\cap I_{H_2}=1) &= p(1-p) = \mathbb{P}(I_{H_1}=0)\mathbb{P}(I_{H_2}=1)\\
\mathbb{P}(I_{H_1}=1\cap I_{H_2}=0) &= (1-p)p = \mathbb{P}(I_{H_1}=1)\mathbb{P}(I_{H_2}=0)\\
\mathbb{P}(I_{H_1}=1\cap I_{H_2}=1) &= p^2 = \mathbb{P}(I_{H_1}=1)\mathbb{P}(I_{H_2}=1)
\end{aligned}$$

Como I_{H_1} e I_{H_2} assume valores em $\{0,1\}$ e para todo $i_1, i_2 \in \{0,1\}$, $P(I_{H_1}=i_1\cap I_{H_2}=i_2)$ e igual a $P(I_{H_1}=i_1)P(I_{H_2}=i_2)$, cncluímos que I_{H_1} e I_{H_2} são independentes.

Lema 18. *Se X e Y são independentes, $E[XY]=E[X]E[Y]$.*

Demonstração.

$$\begin{aligned}
E[XY] &= \sum_{x\in Im(X), y\in Im(Y)} xyP(X=x\cap Y=y) \\
&= \sum_{x\in Im(X)}\sum_{y\in Im(Y)} xyP(X=x)P(Y=y) \\
&= \sum_{x\in Im(X)} xP(X=x)\sum_{y\in Im(Y)} yP(Y=y) \\
&= \sum_{x\in Im(X)} xp_X(x)\sum_{y\in Im(Y)} yp_Y(y) \\
&= E[X]E[Y]
\end{aligned}$$

□

Observação: $E[XY]=E[X]E[Y]$ não implica que X e Y são independentes. Vamos ver um contra-exemplo na lição de casa futura.

Lema 19. *Se X e Y são independentes, $Var[X+Y]=Var[X]+Var[Y]$.*

Demonstração.

$$\begin{aligned}
Var[X+Y] &= E[(X+Y-E[X+Y])^2] \\
&= E[(X-E[X]+Y-E[Y])^2] \\
&= E[(X-E[X])^2]+E[(Y-E[Y])^2]+E[2(X-E[X])(Y-E[Y])] \\
&= Var[X]+Var[Y]+E[2(X-E[X])(Y-E[Y])]
\end{aligned}$$

Portanto, falta mostrar que $E[2(X-E[X])(Y-E[Y])]=0$.

$$\begin{aligned}
E[2(X-E[X])(Y-E[Y])] &= 2E[XY-E[X]Y-XE[Y]+E[X]E[Y]] \\
&= 2(E[XY]-E[X]E[Y]-E[X]E[Y]+E[X]E[Y]]) \\
&= 2(E[XY]-E[X]E[Y])
\end{aligned}$$

Como X e Y são independentes, $E[XY]=E[X]E[Y]$, o que completa a demonstração. □

Exemplo 39 (Continuação dos Exemplos 36 e 38)**.** *Mostramos que $X=I_{H_1}+I_{H_2}$ e I_{H_1} and I_{H_2} são independentes. Portanto, $Var[X]=Var[I_{H_1}]+Var[I_{H_2}]$. Do Exemplo 35, $Var[I_{H_1}]=Var[I_{H_2}]=p(1-p)$. Portanto, $Var[X]=2p(1-p)$, o que confirma novamente os cálculos do Exemplo 36.*

Lema 20. *Se X é uma v.a. discreta, $Var[X]=0$ se e somente se, X é constante (i.é, existe $c\in\mathbb{R}$ tal que $\mathbb{P}(X=c)=1$).*

Demonstração. Assuma que existe $c\in\mathbb{R}$ tal que $\mathbb{P}(X=c)=1$. Então $\mathbb{E}[X]=c$, e $\mathbb{E}[X^2]=c^2$, que implica que $Var[X]=c^2-c^2=0$.

Agora, seja $\mathcal{X}$ o conjunto de todos os valores que X assume. Se $Var[X]=0$, temos que

$$\sum_{x\in\mathcal{X}}(x-\mathbb{E}[X])^2p_X(x)=0.$$

Agora, como $p_X(x) > 0$ para todo $x \in \mathcal{X}$, isso implica que $x - \mathbb{E}[X])^2 = 0$ para todo $x \in \mathcal{X}$, i.e., $x = \mathbb{E}[X]$ para todo $x \in \mathcal{X}$, o que conclui a demonstração. $\square$

3.3.1 Exercícios

Exercício 73. *Mostre que* $Var[X] \geq 0$.

Exercício 74. X *assume valores em* $\{-1, 0, 1\}$*. Seja* $p_X(-1) = p_X(1) = \frac{1-p}{2}$ *e* $p_X(0) = p$*. Encontre* $Var[X]$*. Para que valores de* p *a variância é maximizada? Para que valores de* p *ela é minimizada?*

Exercício 75. *Considere que* X *assume valores em* $\{-a, a\}$ *e* $p_X(-a) = p_X(a) = 0.5$*. Encontre* $\mathbb{E}[X]$ *e* $Var[X]$*. Qua é o comportamento de* $Var[X]$ *como função de* a*?*

Exercício 76. *Considere* 2 *lançamentos de uma moeda com probabilidade* p *de caras. Seja* X *o úmero de caras observadas. Encontre* $Var[X]$*.*

Exercício 77. *Considere que eu retiro duas bolas sem reposição e com probabilidade igual de uma caixa com* 3 *blas rosas e* 3 *bolas laranjas. Seja* X *o número de bolas laranjas que eu retiro. Encontre* $\mathbb{E}[X]$ *e* $Var[X]$*.*

Exercício 78. *Se* $\mathbb{E}[X^2] = 4$, $\mathbb{E}[X] = 1$, $\mathbb{E}[Y^2] = 9$, $\mathbb{E}[Y] = 2$ *e* X *e* Y *ão independentes, encontre* $Var[X+2Y+3]$*.*

Exercício 79. *Seja* X *uma v.a. discreta e* $d \in \mathbb{R}$*,*

a. *Mostre que* $\mathbb{E}[(X-d)^2] = Var[X] + (\mathbb{E}[X] - d)^2$*.*

b. *Prove que* $\mathbb{E}[(X-d)^2]$ *é minimizado quando* $d = \mathbb{E}[X]$*.*

Exercício 80. *Sejam* $X_1, \ldots, X_n$ *v.a.s independentes tais que, para todo* $i \in \{1, \ldots, n\}$, $\mathbb{E}[X_i] = \mu$ *e* $Var[X_i] = \sigma^2$*. Seja* $p_i \geq 0$ *tal que* $\sum_{i=1}^n p_i = 1$*.*

a. *Mostre que* $p_1 X_1, \ldots, p_n X_n$ *são conjuntamente independentes.*

b. *Encontre* $Var[\sum_{i=1}^n p_i X_i]$*.*

c. *Lembramos que* $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$*. Encontre* $Var[\bar{X}]$*.*

d. *Mostre que, para qualquer* $p_i, \ldots, p_n$, $Var[\sum_{i=1}^n p_i X_i] \geq Var[\bar{X}]$*.*

e. *Combine os resultados dos Exercícios 67 e 80. Dê razões em português de porque, entre todas as possíveis variáveis aleatórias da forma* $\sum_{i=1}^n p_i X_i$, $\bar{X}$ *provê os valores que são "ma próximos" de* μ*.*

3.4 Covariância e o Espaço Vetorial das Variáveis Aleatórias

Definição 22 (Covariância)**.** *Sejam X e Y duas variáveis aleatórias discretas.*

$$Cov[X,Y] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

Lema 21. $Cov[X,Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$.

Demonstração.

$$\begin{aligned} Cov[X,Y] &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\ &= \mathbb{E}[XY - X\mathbb{E}[Y] - Y\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[Y]] \\ &= \mathbb{E}[XY] - \mathbb{E}[Y]\mathbb{E}[X] - \mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y] && \text{(Lema 14)} \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \end{aligned}$$

□

Lema 22 (Propriedades da covariância)**.** *Se X e Y são v.a.'s discretas,*

a. $Cov[X,X] = Var[X] \geq 0$. *Portanto $Cov[X,X] = 0$ se X é uma variável aleatória constante.*

b. $Cov[X,Y] = Cov[Y,X]$.

c. $Cov[aX + bY, Z] = aCov[X,Z] + bCov[Y,Z]$.

Demonstração. a.

$$\begin{aligned} Cov[X,X] &= \mathbb{E}[X \cdot X] - \mathbb{E}[X]\mathbb{E}[X] && \text{(Lema 21)} \\ &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = Var[X] && \text{(Lema 17)} \end{aligned}$$

A conclusão segue do Lema 20.

b.

$$\begin{aligned} Cov[X,Y] &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] && \text{(Lema 21)} \\ &= \mathbb{E}[YX] - \mathbb{E}[Y]\mathbb{E}[X] = Cov[Y,X] && \text{(Lema 21)} \end{aligned}$$

c.

$$\begin{aligned} Cov[aX + bY, Z] &= \mathbb{E}[(aX + bY)Z] - \mathbb{E}[aX + bY]\mathbb{E}[Z] && \text{(Lema 21)} \\ &= \mathbb{E}[aXZ + bYZ] - \mathbb{E}[aX + bY]\mathbb{E}[Z] \\ &= a\mathbb{E}[XZ] + b\mathbb{E}[YZ] - (a\mathbb{E}[X] + b\mathbb{E}[Y])\mathbb{E}[Z] && \text{(Lema 14)} \\ &= a(\mathbb{E}[XZ] - \mathbb{E}[X]\mathbb{E}[Z]) + b(\mathbb{E}[YZ] - \mathbb{E}[Y]\mathbb{E}[Z]) \\ &= aCov[X,Z] + bCov[Y,Z] && \text{(Lema 21)} \end{aligned}$$

□

Definição 23. *Seja Ω um conjunto enumerável, P a função probabilidade em Ω.*

1. *Defina $\mathcal{V} = \{X : \Omega \to \mathbb{R}$ tal que $\mathbb{E}[X] = 0\}$.*

2. *Para todo $X \in \mathcal{V}$ e $Y \in \mathcal{V}$, nos definimos $(X + Y) : \Omega \to \mathbb{R}$ ta que $(X + Y)(w) = X(w) + Y(w)$.*

3. *Para todo $X \in \mathcal{V}$ e $a \in \mathbb{R}$, nos definimos $(aX) : \Omega \to \mathbb{R}$ tal que $(a \cdot X)(w) = aX(w)$.*

Lema 23. *$(\mathcal{V}, +, \cdot)$ é um espaço vetorial sobre $\mathbb{R}$.*

Demonstração. Para provar isso, é suficiente demonstrar os itens abaixo:

1. Como $\mathbb{E}[0] = 0$, $0 \in \mathcal{V}$.

2. Se $X \in \mathcal{V}$ and $Y \in \mathcal{V}$, $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y] = 0$. Potanto, $X + Y \in \mathcal{V}$.

3. Se $a \in \mathbb{R}$ e $X \in \mathcal{V}$, $\mathbb{E}[aX] = a\mathbb{E}[X] = 0$. Prtanto, $aX \in \mathcal{V}$.

□

Lema 24. *Cov é m produto interno em $\mathcal{V}$. Portanto, $\sqrt{Var[X]}$ é uma norma em $\mathcal{V}$.*

Demonstração. Segue diretamente do Lema 22. □

Lema 25 (Cauchy-Schwarz para variáveis aleatórias)**.**

$$|Cov[X, Y]| \leq \sqrt{Var[X]}\sqrt{Var[Y]}.$$

A igualdade ocorre se e somente se existem $a, b \in \mathbb{R}$ com $a \neq 0$ tais que $Y = aX + b$.

Demonstração. Seja $V = X - \mathbb{E}[X]$ e $W = Y - \mathbb{E}[Y]$. Como $V, W \in \mathcal{V}$ e, como pelo Lema 24, covariância é um produto interno, segue que

$$\begin{aligned}
|Cov[X, Y]| = |Cov[V, W]| &\leq \sqrt{Cov[V, V]}\sqrt{Cov[W, W]} && \text{(desigualdade de Cauchy-Schwarz)} \\
&= \sqrt{Var[V]}\sqrt{Var[W]} && \text{(Lema 22)} \\
&= \sqrt{Var[X]}\sqrt{Var[Y]}
\end{aligned}$$

Agora, da desigualdade de Cauchy-Schwarz sabemos que a igualdade ocorre se e somente se existe $b \in \mathbb{R}$, $b \neq 0$, tal que $W = bV$. Em outras palavras, a igualdade ocorre se e somente se existem $b \in \mathbb{R}$, $b \neq 0$, tais que $Y - \mathbb{E}[Y] = b(X - \mathbb{E}[X])$, o que conclui a demonstração. □

Lema 26 (Teorema de Pitágoras para variáveis aleatórias)**.**

$$Var[X + Y] = Var[X] + Var[Y] + 2Cov[X, Y]$$

Portanto, se $Cov[X, Y] = 0$, $Var[X + Y] = Var[X] + Var[Y]$.

Demonstração.

$$\begin{aligned}
Var[X + Y] &= Cov[X + Y, X + Y] && \text{(Lema 22)} \\
&= Cov[X, X] + Cov[Y, Y] + 2Cov[X, Y] && \text{(Lema 24)} \\
&= Var[X] + Var[Y] + 2Cov[X, Y] && \text{(Lema 22)}
\end{aligned}$$

$\square$

Definição 24 (Correlação)**.**

$$Corr[X,Y] = \frac{Cov[X,Y]}{\sqrt{Var[X]}\sqrt{Var[Y]}}$$

Seja $< \cdot, \cdot >$ um produto interno e $\|\cdot\|$ a norma gerada pelo produto interno. Lembramos das aulas de álgebra linear que $\frac{<v_1,v_2>}{\|v_1\|\|v_2\|}$ é o cosseno do ângulo entre v_1 e v_2. Prtanto, usando o Lema 24, podemos interpretar a $Corr[X,Y]$ como o cosseno do ângulo entre variáveis aleatórias X e Y. Em outras palavras, $Corr[X,Y]$ é a medida da associação linear entre X e Y.

Lema 27.

$$|Corr[X,Y]| \leq 1$$

Demonstração. Segue diretamente da aplicação do Lema 25 à Definição 24. $\square$

Lema 28. *Seja X a variável aleatória discreta.*

a. Seja $b \in \mathbb{R}$, $Cov[b,X] = 0$.

b. Seja $a \neq 0$,

$$Corr[aX+b,X] = \begin{cases} 1 & , \textit{ if } a > 0 \\ -1 & , \textit{ if } a < 0 \end{cases}$$

Demonstração.

a. $Cov[b,X] = E[bX] - E[b]E[X] = bE[X] - bE[X] = 0$

b.

$$\begin{aligned} Corr[aX+b,X] &= \frac{Cov[aX+b,X]}{\sqrt{Var[aX+b]}\sqrt{Var[X]}} \\ &= \frac{aCov[X,X] + Cov[b,X]}{\sqrt{Var[aX+b]}\sqrt{Var[X]}} && \text{(Lema 22)} \\ &= \frac{aVar[X]}{\sqrt{aVar[X]}\sqrt{Var[X]}} && \text{(Lemas 22 e 16)} \\ &= \frac{a}{|a|} \end{aligned}$$

$\square$

Sob hipótese de independência, a covariância e a correlação devem ser iguais a zero:

Lema 29. *Se X e Y são independentes, então $Cov(X,Y) = Corr(X,Y) = 0$.*

Demonstração. Segue do fato de que se X e Y são independentes, então $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ $\square$

O oposto, porém, não é verdadeiro; veja exemplos nos exercícios.

Segue do Lema 25 qu a correlação tem valor absoluto máximo se e somente se X e Y são linearmente dependentes. É por isso que dizemos que acovariância é uma medida de associação linear.

Covariâncias permitem-nos calcular variâncias de somas de variáveis aleatórias que não são independentes de forma fácil:

Lema 30. *Sejam $X_1, \ldots, X_n$ vaiáveis aleatórias. Temos que*

$$Var[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} Var(X_i) + \sum\sum_{i \neq j} Cov[X_i, X_j]$$

Lema 30 é a generalização do Lema 26, e nos deixamos a sua demonstração para o leitor.

Exemplo 40 (Ross)**.** *Um grupo de n pessoas joga os seus chapéus no meio de uma sala. Os chapéus se misturam, e então cada pessoa pega um chapéu ao acaso. Seja X o número esperado de pessos que vão pegar o seu próprio chapéu. Nos temos que X pode ser escrito como a soma $X = \sum_{i=1}^{n} X_i$, onde X_i é um se a i-ésima pessoa pega o próprio chapéu e zero caso contrário. Agora, $\mathbb{E}[X_i] = 1 \cdot p_{X_i}(1) = 1/n$. Segue que $\mathbb{E}[X] = n \cdot 1/n = 1$. Podemos também calcular a variância de X. Primeiro, note que $X_i^2 = X_i$, e portanto $Var(X_i) = \mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2 = 1/n - (1/n)^2$. Além disso, para $i \neq j$, X_iX_j é uma variável aleatória que pode assumir valores 0 e 1. Mais,*

$$\mathbb{P}(X_iX_j = 1) = \mathbb{P}(X_i = 1, X_j = 1) = \mathbb{P}(X_j = 1)\mathbb{P}(X_i = 1|X_j = 1) = \frac{1}{n} \cdot \frac{1}{n-1}.$$

Segue que $\mathbb{E}[X_iX_j] = 1 \cdot \frac{1}{n} \cdot \frac{1}{n-1}$, e portanto

$$Cov[X_i, X_j] = \mathbb{E}[X_iX_j] - \mathbb{E}[X_i]\mathbb{E}[X_j] = \frac{1}{n} \cdot \frac{1}{n-1} - \frac{1}{n} \cdot \frac{1}{n} = \frac{1}{n^2(n-1)}.$$

Finalmente, observamos que $Cov[X_i, X_j]$ é a mesma para todo $i \neq j$. Como há $2\binom{n}{2}$ destes pares, segue do Lema 30 que

$$Var[X] = Var\left[\sum_{i=1}^{n} X_i\right] = n\left(\frac{1}{n} - \frac{1}{n^2}\right) + 2\binom{n}{2}\frac{1}{n^2(n-1)} = 1$$

3.4.1 Exercícios

Exercício 81. *Sejam $\mathbb{P}(A) = p_A$, $\mathbb{P}(B) = p_B$ and $\mathbb{P}(A \cap B) = p_{A \cap B}$.*

a. Encontre $Cov[I_A, I_B]$ e $Corr[I_A, I_B]$.

b. Encontre um valor numérico para $Corr[I_A, I_B]$ quando $A = B$.

c. Encontre um valor numérico para $Corr[I_A, I_B]$ quando A é independente de B.

d. Encontre um valor numérico para $Corr[I_A, I_B]$ quando A e B paticionam Ω.

e. Dê uma interpretação dos itens acima em português.

Solução:

a.

$$\begin{aligned} Cov[I_AI_B] &= E[I_AI_B] - E[I_A]E[I_B] && \text{(Lema 21)} \\ &= E[I_{A \cap B}] - E[I_A]E[I_B] \\ &= p_{A \cap B} - p_Ap_B \end{aligned}$$

$$\begin{aligned} Corr[I_A I_B] &= \frac{Cov[I_A, I_B]}{\sqrt{Var[I_A]}\sqrt{Var[I_B]}} & \text{(Lema 24)} \\ &= \frac{p_{A\cap B} - p_A p_B}{\sqrt{p_A(1-p_A)}\sqrt{p_B(1-p_B)}} \end{aligned}$$

b. Quando $A = B$, $p_A = p_B = p_{A\cap B}$,

$$\begin{aligned} Corr[I_A I_B] &= \frac{p_{A\cap B} - p_A p_B}{\sqrt{p_A(1-p_A)}\sqrt{p_B(1-p_B)}} \\ &= \frac{p_A - p_A^2}{\sqrt{p_A(1-p_A)}\sqrt{p_A(1-p_A)}} = 1 \end{aligned}$$

c. Quando A é independente de B, $p_{A\cap B} = p_A p_B$,

$$\begin{aligned} Corr[I_A I_B] &= \frac{p_{A\cap B} - p_A p_B}{\sqrt{p_A(1-p_A)}\sqrt{p_B(1-p_B)}} \\ &= \frac{p_A p_B - p_A p_B}{\sqrt{p_A(1-p_A)}\sqrt{p_B(1-p_B)}} = 0 \end{aligned}$$

d. Quando A e B paticionam Ω, $p_{A\cap B} = 0$ e $p_A + p_B = 1$,

$$\begin{aligned} Corr[I_A I_B] &= \frac{p_{A\cap B} - p_A p_B}{\sqrt{p_A(1-p_A)}\sqrt{p_B(1-p_B)}} \\ &= \frac{-p_A(1-p_A)}{\sqrt{p_A(1-p_A)}\sqrt{(1-p_A)p_A}} = -1 \end{aligned}$$

e. I_A e I_B tem máxima associação linear quando $A = B$. Neste casp, I_A e I_B sempre assumem o mesmo valor. Quando A e B são independentes, I_A e I_B não estão associados linearmente. Finalmente, quando A e B particionam Ω, I_A e I_B nunca assumem o mesmo valor. Neste caso, I_A e I_B tem a associação linear negativa máxima.

Exercício 82. *Seja $\Omega = \{-1, 0, 1\}$ e todos os possíveis resultados igualmente prováveis. Seja $X(w) = w$ e $Y(w) = w^2$.*

a. Encontre $Corr[X, Y]$. X e Y têm associação linear?

b. X e Y são independentes?

Solução:

a. Observe da Definição 24 que, se $Cov[X, X^2] = 0$, então $Corr[X, X_2] = 0$.

$$\begin{aligned} Cov[X, Y] &= E[XY] - E[X]E[Y] & \text{(Lema 21)} \\ &= E[X^3] - E[X]E[X^2] \end{aligned}$$

$$
\begin{aligned}
E[X] &= \frac{-1+0+1}{3} = 0 \\
E[X^2] &= \frac{(-1)^2+0^2+1^2}{3} = \frac{2}{3} \\
E[X^3] &= \frac{-1^3+0^3+1^3}{3} = 0
\end{aligned}
$$

Portanto, $Cov[X,Y] = 0$ e $Corr[X,Y] = 0$. X e Y não têm associação linear.

b.

$$
\begin{aligned}
P(X=0 \cap Y=0) &= P(\{w \in \Omega : w = 0\} \cap \{w \in \Omega : w^2 = 0\}) \\
&= P(\{0\} \cap \{0\}) = P(\{0\}) = \frac{1}{5}
\end{aligned}
$$

$$
\begin{aligned}
P(X=0)P(Y=0) &= P(\{w \in \Omega : w = 0\})P(\{w \in \Omega : w^2 = 0\}) \\
&= P(\{0\})P(\{0\}) = \frac{1}{25}
\end{aligned}
$$

Portanto, $P(X=0 \cap Y=0) \neq P(X=0)P(Y=0)$ e X e Y não são independentes.

Exercício 83. *Sejam $X_1, \ldots, X_n, Y_1, \ldots, Y_m$ vaiáveis aleatórias e sejam $a_1, \ldots, a_n, b_1, \ldots, b_m$ números reais. Prove que*

$$
Cov(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{m} b_j Y_j) = \sum_{j=1}^{m} a_i b_j Cov(X_i, Y_j).
$$

Exercício 84. *Assuma que $X_1, \ldots, X_n$ são variáveis aleatórias* independentes e identicamente distribuídas *(a.k.a., iid) com variância $\sigma^2 < \infty$. Mostre que $Cov[X_i - \bar{X}, \bar{X}] = 0$.*

Exercício 85. *Assuma que $X_1, \ldots, X_n$ são variáveis identicamente distribuídas com covariância $Cov[X_i, X_j] = \rho$ para $i \neq j$, variância σ^2 e média μ. Determine a variância de $\bar{X}$.*

Exercício 86. *Sejam $X_1, X_2, \ldots$ independentes com média comum μ e vaiância comum σ^2. Para $n = 1, 2, \ldots$, seja $Y_n = X_n + X_{n+1} + X_{n+2}$. Para $j = 0, 1, \ldots$, encontre $Cov(Y_n, Y_{n+j})$.*

8u

Exercício 87. *(Desafio) Demonstre a desigualdade triangular para variáveis aleatórias:*

$$
\sqrt{Var[X+Y]} \leq \sqrt{Var[X]} + \sqrt{Var[Y]}
$$

4 Processos de Bernoulli (um exemplo longo. . .)

4.1 Distribuição de Bernoulli

Dizemos que a distribuição de uma variável aleatória X é Bernoulli se X pode assumir valores em $\{0, 1\}$. Em alguns contextos, 0 é interpretado como falha e 1 como sucesso. Dizemos que X tem distribuição de Bernoulli com parâmetro p se $\mathbb{P}(X = 1) = p$ e escrevemos $X \sim \text{Ber}(p)$.

Exemplo 41. *Seja H o evento em que observamos cara no lançamento de uma moeda honesta. I_H asume valores em $\{0, 1\}$ e a sua distribuição é Bernoulli. Mais, como a moeda é honesta, $I_H \sim Ber(p)$.*

Lema 31. *Se $X \sim Ber(p)$, $\mathbb{E}[X] = p$ e $Var[X] = p(1 - p)$.*

Demonstração.

$$\mathbb{E}[X] = 0 \cdot (1 - p) + 1 \cdot p = p$$

e

$$\begin{aligned} Var[X] &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\ &= (0 \cdot (1 - p) + 1 \cdot p) - p^2 \\ &= p(1 - p) \end{aligned}$$

□

Definição 25. *Dizemos que uma sequência de variáveis aleatórias $X_1, X_2, \ldots$ é um Processo de Bernoulli com parâmetro p se as variáveis aleatórias são conjuntamente independentes e tais que, para cada $i \in \mathbb{N}$, $X_i \sim Bernoulli(p)$.*

Exemplo 42 (Programação)**.** *Escreva um código que gera números de acordo com a distribuição Bernoulli(p).*

```
import random
def rbernoulli(p):
  return random.random() < p
```

4.2 Distribuição Binomial

Definição 26. *Seja $X_1, X_2, \ldots$ um Processo de Bernoulli com parâmetro p. Dizemos que $\sum_{i=1}^{n} X_i \sim Binomial(n, p)$. Isto é, $\sum_{i=1}^{n} X_i$ tem distribuição Binomial com parâmetros n e p.*

Podemos interpretar a distribuição Binomial da seguinte forma. Considere que realizamos n experimentos **independentes** e cada experimento pode ser um sucesso com probabilidade p ou falhar com probabilidade $1 - p$. Seja X o número de sucessos que obtemos depois de realizar os n experimentos. X em distribuição Binomial(n, p), onde n é o número de experimentos e p é a probabilidade de ucesso em cada experimento.

Lema 32. *Se X tem distribuição Binomial(n, p), então, para $0 \leq i \leq n$, $\mathbb{P}(X = i) = \binom{n}{i} p^i (1 - p)^{n-i}$.*

Demonstração. Para obter $X = i$, devemos observar i sucessos e $n-i$ falhas. Observe que, como os experimentos são independentes, a probabilidade de qalquer resultado com i sucessos e $n-i$ fracassos é $p^i(1-p)^i$.

Portanto $\mathbb{P}(X = i) = c_{n,i}p^i(1-p)^{n-i}$, onde $c_{n,i}$ é o número de resultados com i sucessos e $n-i$ fracassos. Note que $c_{n,i}$ corresponde ao número de anagramas de $\underbrace{SSS\ldots}_{\text{i times}}\underbrace{FF\ldots}_{\text{n-i times}}$. Há $n!$ permutações entre as letras mas as $i!$ permutações entre os S's são a mesma coisa, tal como as $(n-i)!$ permutações entre os F's. Portanto $c_{n,i} = \frac{n!}{i!(n-i)!} = \binom{n}{i}$ e $\mathbb{P}(X = i) = \binom{n}{i}p^i(1-p)^{n-i}$. □

O resultado a seguir frequentemente é útil quando realizamos cálculos com binomiais.

Lema 33 (Teorema Binomial)**.**

$$\sum_{i=0}^{n}\binom{n}{i}a^ib^{n-i} = (a+b)^n$$

Exemplo 43. *Seja X variável aleatória com distribuição Binomial(n,p). Vamos usar o Teorema Binomial para fazer um teste de sanidade e provar que $\mathbb{P}(X \in \{0,1,\ldots,n\}) = 1$.*

$$\begin{aligned}
\mathbb{P}(X \in \{0,1,\ldots,n\}) &= \sum_{i=0}^{n}\mathbb{P}(X = i) \\
&= \sum_{i=0}^{n}\binom{n}{i}p^i(1-p)^{n-i} \\
&= (p+1-p)^n = 1
\end{aligned}$$

Lema 34. *Se X tem distribuição Binomial(n,p), então $\mathbb{E}[X] = np$ e $Var[X] = np(1-p)$.*

Demonstração.

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{i=0}^{n} i\cdot P(X = i) \\
&= \sum_{i=1}^{n} i\binom{n}{i}p^i(1-p)^{n-i} && \text{(Cancelando i=0)} \\
&= \sum_{i=1}^{n} i\frac{n!}{(n-i)!i!}p^i(1-p)^{n-i} \\
&= np\sum_{i=1}^{n}\frac{(n-1)!}{(n-i)!(i-1)!}p^{i-1}(1-p)^{n-i} && \text{(Pois } i! = i\cdot(i-1)!\text{)} \\
&= np\sum_{i=1}^{n}\binom{n-1}{i-1}p^{i-1}(1-p)^{n-i} \\
&= np\sum_{j=0}^{n-1}\binom{n-1}{j}p^j(1-p)^{n-1-j} && \text{(Calling j = i+1)} \\
&= np(p+1-p)^{n-1} = np && \text{(Teorema Binomial, Lema 33)}
\end{aligned}$$

Além disso,

$$
\begin{aligned}
\mathbb{E}[X(X-1)] &= \sum_{i=0}^{n} i(i-1)\cdot\mathbb{P}(X=i) \\
&= \sum_{i=2}^{n} i(i-1)\frac{n!}{(n-i)!i!}p^i(1-p)^{n-i} && \text{(cancelando i=0 e i=1)} \\
&= n(n-1)p^2\sum_{i=0}^{n}\frac{(n-2)!}{(n-i)!(i-2)!}p^{i-2}(1-p)^{n-i} && \text{(Pois } i! = i(i-1)\cdot(i-2)!\text{)} \\
&= n(n-1)p^2\sum_{j=0}^{n-2}\binom{n-2}{j}p^j(1-p)^{n-2-j} && \text{(Chamando j = i+2)} \\
&= n(n-1)p^2(p+1-p)^{n-2} = n(n-1)p^2 && \text{(Teorema Binomial, Lema 33)}
\end{aligned}
$$

Nos provamos que

$$
\mathbb{E}[X] = np
$$
$$
\mathbb{E}[X(X-1)] = n(n-1)p^2
$$

Lembramos que, do Lema 17

$$
Var[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2
$$

e, portanto,

$$
\begin{aligned}
Var[X] &= \mathbb{E}[X^2] - \mathbb{E}[X] + \mathbb{E}[X] - \mathbb{E}[X]^2 \\
&= \mathbb{E}[X(X-1)] + \mathbb{E}[X] - \mathbb{E}[X]^2 && \text{(Aditividade da Esperança, Lema 19)} \\
&= n(n-1)p^2 + np - n^2p^2 \\
&= n^2p^2 - np^2 + np - n^2p^2 \\
&= np(1-p)
\end{aligned}
$$

Como demonstração alternativa, seja S_i o evento em que o experimento i foi um sucesso. Sabemos que $X = \sum_{i=1}^{n} I_{S_i}$ e que I_{S_i} são variáveis aleatórias independentes com distribuição Bernoulli(p). Portanto,

$$
\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}[\sum_{i=1}^{n} I_{S_i}] \\
&= \sum_{i=1}^{n}\mathbb{E}[I_{S_i}] && \text{(Aditividade da Esperança, Lema 19)} \\
&= np && \text{(Propriedades de Bernoulli(p), Lema 31)}
\end{aligned}
$$

e

$$
\begin{aligned}
Var[X] &= Var[\sum_{i=1}^{n} I_{S_i}] \\
&= \sum_{i=1}^{n} Var[I_{S_i}] && \text{(Aditividade da Variância, Lema 19)} \\
&= np(1-p) && \text{(Propriedades de Bernoulli(p), Lema 31)}
\end{aligned}
$$

□

Exemplo 44 (Programação)**.** *screva um código que gera um número de acordo com distribuição Binomial(n,p). Considere que rbernoulli(p) é a função do Exemplo 42.*

```
def rbinom(n,p):
  sum = 0
  foreach ii in range(n):
    sum += rbernoulli(p)
  return(sum)
```

4.2.1 Exercícios

Exercício 88. *Eu lanço uma moeda com probabilidade* p *de cara* 5 *vezes. Qual é a probabilidade de que exatamente* 3 *resultados são caras?*

Solução: Seja X o número de caras. $X \sim \text{Binomial}(5, p)$. Portanto, pelo Lema 32,

$$P(X = 3) = \binom{5}{3} p^3 (1-p)^2 = 10p^3(1-p)^2$$

Exercício 89. *Eu lanço uma moeda honesta* 7 *vezes. Qual é a probabilidade de que* 1 *o mais resultados são caras?*

Solução: Seja X o número de caras. $X \sim \text{Binomial}(7, p)$. Observe que

$$
\begin{aligned}
P(X \geq 1) &= 1 - P(X = 0) \\
&= 1 - \binom{7}{0}(1-p)^7 = 1 - (1-p)^7
\end{aligned}
$$

Exercício 90. *Os itens produzidos por uma máquina apresentam defeito independentemente, com probabilidade* p*. Em uma batelada de* 1000 *itens, quantos esperamos que terão defeitos? Qual é a variância do número de itens com defeito?*

Solução: Seja X o número de itens com defeito em uma batelada de 1000. $X \sim \text{Binomial}(1000, p)$. Portanto, usando Lema 34, $E[X] = 1000p$.

Exercício 91. *Eu lanço a moeda honesta* 4 *vezes e uma moeda com probabilidade* p *de cara* 8 *vezes. Seja* X *o número total de caras. Calcule* $\mathbb{E}[X]$ *and* $Var[X]$*.*

Solução: Sejam Y e Z respectivamente o número total de caras da moeda honesta e da moeda viciada. Por definição $X = Y + Z$. Além disso, $Y \sim$ Binomial$(4, 0.5)$ e $Z \sim$ Binomial$(8, p)$ e X e Y são independentes. Portanto,

$$E[X + Y] = E[X] + E[Y] = 4 \cdot 0.5 + 8 \cdot p = 2 + 8p$$

$$Var[X + Y] = Var[X] + Var[Y] = 4 \cdot 0.5 \cdot 0.5 + 8p(1 - p) = 1 + 8p(1 - p)$$

Exercício 92. *Considere que uma caixa tem $o \geq 1$ bolas laranjas e $p \geq 1$ bolas rosas. Suponha que eu retiro 2 bolas sem resposição da caixa. Seja X o número total de bolas rosas. A distribuição de X é uma Binomial? Se sim, encontre os parâmetros da distribuição.*

Solução: A f X is

$$\begin{aligned}
p_X(0) &= \frac{o}{o+p} \cdot \frac{o-1}{o+p-1} \\
p_X(1) &= \frac{o}{o+p} \cdot \frac{p}{o+p-1} + \frac{p}{o+p} \cdot \frac{o}{o+p-1} = \frac{2op}{(o+p)(o+p-1)} \\
p_X(2) &= \frac{p}{o+p} \cdot \frac{p-1}{o+p-1}
\end{aligned}$$

A seguir, vamos determinar $E[X]$ e $Var[X]$.

$$\begin{aligned}
E[X] &= 1 \cdot \frac{2op}{(o+p)(o+p-1)} + 2 \cdot \frac{p}{o+p} \cdot \frac{p-1}{o+p-1} \\
&= \frac{2p(o+p-1)}{(o+p)(o+p-1)} = 2 \cdot \frac{p}{o+p}
\end{aligned}$$

$$\begin{aligned}
E[X^2] &= 1^2 \cdot \frac{2op}{(o+p)(o+p-1)} + 2^2 \cdot \frac{p}{o+p} \cdot \frac{p-1}{o+p-1} \\
&= \frac{2p(o+2(p-1))}{(o+p)(o+p-1)}
\end{aligned}$$

$$\begin{aligned}
Var[X] &= E[X^2] - E[X]^2 \\
&= \frac{2p(o+2(p-1))}{(o+p)(o+p-1)} - \frac{4p^2}{(o+p)^2} \\
&= \frac{2p}{o+p}\left(\frac{o+2p-2}{o+p-1} - \frac{2p}{o+p}\right) \\
&= \frac{2p}{o+p}\left(\frac{(o+2p-2)(o+p) - 2p(o+p-1)}{(o+p)(o+p-1)}\right) \\
&= \frac{2p}{o+p}\left(\frac{o(o+p-2)}{(o+p)(o+p-1)}\right)
\end{aligned}$$

Como $Im(X) = \{0, 1, 2\}$, X só pode ser uma binomial com $n = 2$. Assuma que $X \sim$ Binomial$(2, a)$. Usando Lema 34, $E[X] = 2a$ e, portanto, $a = \frac{p}{o+p}$. Assim, $Var[X] = 2 \cdot \frac{p}{o+p} \cdot \frac{o}{o+p}$. Isto é,

$$\frac{2p}{o+p}\left(\frac{o(o+p-2)}{(o+p)(o+p-1)}\right) = 2\cdot\frac{p}{o+p}\cdot\frac{o}{o+p}$$
$$\frac{o+p-2}{o+p-1} = 1$$

Uma contradição. Portanto, X não tem distribuição Binomial.

Exercício 93. *Assuma que você lança uma moeda n vezes e observa um número total de caras igual a X. Digamos que a probabilidade de cara é p. Se você observa exatamente i caras, $i \in \{0, 1, \ldots, n\}$, qual é o valor de p quemaximiza a probabilidade $\mathbb{P}(X = i)$? Este é um exemplo da técnica de* inferência estatística*: dado o valor observado de uma variável aleatória, você tenta inferir qual é a distribuição de probabilidade da qual este valor veio.*

Exercício 94.

a. *Uma moeda honesta é lançada* 1001 *vezes. Encontre a probabilidade de que observamos estritamente mais caras do que coroas.*

b. *Uma moeda honesta é lançada* 1000 *vezes. Encontre a probabilidade de que observamos estritamente mais caras do que coroas.*

Solução:

a. Seja X_{1001} o número total de resultados que observamos em 1001 lançamentos da moeda. Como a moeda é honesta, $P(X_{1001} \leq 500) = P(X_{1001} \geq 501)$. Prtanto, $P(X_{1001} \geq 501) = \frac{1}{2}$.

b. Seja X_{1000} o número total de resultados que observamos em 1000 lançamentos da moeda. Como a moeda é honesta, $P(X_{1000} \leq 499) = P(X_{1000} \geq 501)$. Portanto,

$$\begin{aligned}
P(X_{1000} \leq 499) + P(X_{1000} = 500) + P(X_{1000} \geq 501) &= 1\\
2P(X_{1000} \geq 501) &= 1 - P(X_{1000} = 500)\\
P(X_{1000} \geq 501) &= \frac{1 - P(X_{1000} = 500)}{2}
\end{aligned}$$

Usando a fmp Binomial, $P(X_{1000} \geq 501) = \frac{1-\binom{1000}{500}(0.5)^{500}}{2}$.

Exercício 95 (Desafio)**.** *Seja $X \sim$ Binomial$(1000, 0.5)$ e $Y \sim$ Binomial$(1001, 0.5)$ duas variáveis aleatórias independentes. Determine a probabilidade de que $Y > X$.*

Solução: Verifique a solução do Exercício 31.

4.3 Distribuição Geométrica

Seja $X_1, X_2, \ldots$ um processo de Bernoulli com parâmetro p. Defina Y como o menor índice i tal que $X_i = 1$. Isto é, para todo $w \in \Omega$, $Y(w) = \min_{i \in \mathbb{N}}\{i : X_i(w) = 1\}$. Y pode ser interpretado como o número de experimentos em um Processo de Bernoulli até observarmos um 1. Dizemos que Y tem distribuição Geométrica com parâmetro p e escrevemos $Y \sim \text{Geom}(p)$.

Exemplo 45. *Considere que uma moeda honesta é lançada repetidamente. Seja Y o número de lançamentos até observarmos a primeira cara. $Y \sim$ Geom(0.5).*

Se Y tem distribuição Geometrica, Y pode assumir qualquer valor maior ou igual a 1. Isto é, ao contrário da distribuição Binomial, a Geométrica pode assumir um número infinito de valores.

Lema 35. *Seja $Y \sim$ Geom(p). $p_Y(i) = p(1-p)^{i-1}$.*

Demonstração. Observe que só existe uma forma de atingirmos $Y = i$. Os primeiros $i-1$ experimentos devem ser falhas e o experimento seguinte deve ser um sucesso. Como os experimentos são independentes, a probabilidade deste evento é $(1-p)^{i-1}p$. □

Lema 36 (Série Geométrica)**.** *Se $0 < q < 1$,*

$$\sum_{i=j}^{\infty} q^i = \frac{q^j}{1-q}$$

Exemplo 46. *Podemos fazer um teste de sanidade e mostrar que a fmp da Geométrica soma 1. Observe que:*

$$\begin{aligned}
\sum_{i=1}^{\infty} p(1-p)^{i-1} &= p\sum_{j=0}^{\infty}(1-p)^j \\
&= \frac{p}{1-(1-p)} = 1 && \textit{Série Geométrica (Lema 36)}
\end{aligned}$$

Observe que $\mathbb{P}(X \geq i)$ corresponde à probabilidade de que observamos pelo menos $i-1$ falhas. O lema a seguir demonstra uma forma de calcular $\mathbb{P}(X \geq i)$ usando a série geométrica.

Lema 37. *Seja $X \sim$ Geom(p). Temos que $\mathbb{P}(X \geq i) = (1-p)^{i-1}$.*

Demonstração.

$$\begin{aligned}
\mathbb{P}(X \geq i) &= \sum_{j=i}^{\infty} \mathbb{P}(X = i) \\
&= \sum_{j=i}^{\infty} (1-p)^{j-1}p \\
&= p\sum_{j=i}^{\infty} (1-p)^{j-1} \\
&= \frac{p(1-p)^{i-1}}{p} = (1-p)^{i-1} && \text{Série Geométrica (Lema 36)}
\end{aligned}$$

□

Lema 38. *Se X tem distribuição Geom(p), então $\mathbb{E}[X] = \frac{1}{p}$ e $Var[X] = \frac{1-p}{p^2}$.*

Demonstração.

$$
\begin{aligned}
\mathbb{E}[X] &= \sum_{i=1}^{\infty} i\mathbb{P}(X = i) \\
&= \sum_{i=1}^{\infty} ip(1-p)^{i-1} \\
&= \sum_{i=1}^{\infty} -p\frac{\partial(1-p)^i}{\partial p} \\
&= -p\frac{\partial}{\partial p}\sum_{i=1}^{\infty}(1-p)^i \\
&= -p\frac{\partial}{\partial p}\frac{1-p}{p} && \text{Série Geométrica (Lema 36)} \\
&= \frac{1}{p}
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}[X(X+1)] &= \sum_{i=1}^{\infty} i(i+1)\mathbb{P}(X = i) \\
&= \sum_{i=1}^{\infty} i(i+1)p(1-p)^{i-1} \\
&= \sum_{i=1}^{\infty} p\frac{\partial^2(1-p)^{i+1}}{\partial p^2} \\
&= p\frac{\partial^2}{\partial p^2}\sum_{i=1}^{\infty}(1-p)^{i+1} \\
&= p\frac{\partial^2}{\partial p^2}\frac{(1-p)^2}{p} && \text{Série Geométrica (Lema 36)} \\
&= \frac{2}{p^2}
\end{aligned}
$$

Finalmente,

$$
\begin{aligned}
Var[X] &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 && \text{(Lema 17)} \\
&= \mathbb{E}[X^2] + \mathbb{E}[X] - \mathbb{E}[X] - \mathbb{E}[X]^2 \\
&= \mathbb{E}[X^2 + X] - \mathbb{E}[X] - \mathbb{E}[X]^2 && \text{Aditividade da Esperança (Lema 14)} \\
&= \frac{2}{p^2} - \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2}
\end{aligned}
$$

□

Exemplo 47 (Programação)**.** *Escreva um código que gera um número de acordo com a distribuição Geometric(p). Considere que rbernoulli(p) é a função do Exemplo 42.*

```
def rgeom(p):
  ii = 1
  while(not(rbernoulli(p))):
    ii += 1
  return(ii)
```

4.3.1 Exercícios

Exercício 96. *Numa certa população, 10% das pessoas têm típo sanguíneo O, 40% têm típo sanguíneo A, 45% têm típo sanguíneo B, e 5% têm típo sanguíneo AB. Seja Y o número de doadores que entram num banco de sangue num certo dia até o primeiro doador potencial para um paciente com tipo sanguíneo B (i.é. até um doador ter tipo sanguíneo O ou B).*

a. *Encontre $P(Y = 1)$.*

b. *Encontre $P(Y \geq 4)$.*

c. *Encontre $E[Y]$.*

d. *Encontre $Var[Y]$.*

Solução: Um doador é compatível com tipo B se ele é do tipo O (10%) ou do tipo B (45%). Portanto, a probabilidade de que um dado doador seja compatível com o tipo B é $10\%+45\% = 55\%$. Assim, $Y \sim \text{Geom}(0.55)$.

a. $P(Y = 1) = 0.55$

b. $P(Y \geq 4) = (1 - 0.55)^3 \approx 0.091$

c. $E[Y] = \frac{1}{0.55} \approx 1.82$

d. $Var[Y] = \frac{1-0.55}{0.55^2} \approx 1.49$

Exercício 97. *Seja $X \sim$ Geom(p). Determine $\mathbb{E}[X(X + 1)(X + 2)]$.*

Solução:

$$\begin{aligned}
E[X(X+1)(X+2)] &= \sum_{i=1}^{\infty} i(i+1)(i+2)(1-p)^{i-1}p \\
&= -p\sum_{i=1}^{\infty} \frac{\partial^3 (1-p)^{i+2}}{\partial p^3} \\
&= -p\frac{\partial^3}{\partial p^3}\sum_{i=1}^{\infty}(1-p)^{i+2} \\
&= -p\frac{\partial^3}{\partial p^3}\frac{(1-p)^3}{p}\sum_{i=1}^{\infty}p(1-p)^{i-1} \\
&= -p\frac{\partial^3}{\partial p^3}\frac{(1-p)^3}{p} = -p\frac{-6}{p^4} = \frac{6}{p^3}
\end{aligned}$$

Exercício 98. *Seja $X \sim$ Geom(p). Determine $\mathbb{P}(X \geq t + s|X > s)$.*

Solução: Usamos o Lema 37 várias vezes.

$$\begin{aligned} P(X \geq t+s|X > s) &= \frac{P(X \geq t+s \cap X > s)}{P(X > s)} \\ &= \frac{P(X \geq t+s)}{P(X > s)} \\ &= \frac{(1-p)^{t+s}}{(1-p)^{s+1}} = (1-p)^{t-1} = P(X \geq t) \end{aligned}$$

Exercício 99. *Suponha que uma pessoa está esperando num ponto de ônibus. Este pessoa acredita que o evento de o ônibus passar em cada intervalo de cinco minutos é independente de todos os outros intervalo de cinco minutos, e que a probabilidade de o ônibus passar em um dado intervalo de cinco minutos é p, onde p é conhecido. Esta crença é diferente da crença em que os ônibus operam em horários fixos. Se esperou-se já 20 minutos, a chance de o ônibus passar nos próximos cinco minutos é maior, menor, ou igual? Porque?*

Solution: Como o evento de o ônibus passar em cada intervalo é independente dos outros intervalos, o primeiro intervalo em que o ônibus passa tem distribuição Geométrica com parâmetro p. No último exercício, mostramos que a distribuição Geométrica não tem memória, isto é, independentemente de quanto se esperou, a probabilidade de o ônibus passar no próximo intervalo é sempre igual a p.

Exercício 100. *Determine a esperança de uma variável aleatória geométrica usando o Lema 13.*

Exercício 101. *Considere o seguinte jogo: uma moeda honesta é lançada até aparecer primeira cara. Se n lançamentos são feitos no total, você recebe o prêmio de 2^n dólares. Quanto dinheiro você espera ganhar neste jogo?*

4.4 Distribuição Binomial Negativa

Considere uma sequência de experimentos de Bernoulli independentes, todos cm a mesma probabilidade de sucesso, p. Lembramos que a distribuição Geométrica envolve realizar experimentos até obtermos o primeiro sucesso. A Binomial Negativa é uma generalização da distribuição Geométrica e corresponde ao número de experimentos até que os primeiros r sucessos sejam observados. Neste sentido, se nos considerarmos $X_1, \ldots, X_r$ variáveis independentes tais que $X_i \sim \text{Geom}(p)$, então $\sum_{i=1}^{r} X_i$ tem distribuição Binomial Negativa(r, p).

Lema 39. *Se Y tem distribuição Binomial Negativa(r,p), então $p_Y(i) = \binom{i-1}{r-1}(1-p)^{i-r}p^r$.*

Demonstração. Observe que $\mathbb{P}(Y = i)$ corresponde ao evento em que o r-ésimo sucesso ocorre exatamente no i-ésimo experimento. Por definição, este evento é verdadeiro se e somente se foram observados r sucessos e $i-r$ falhas nos primeiros i experimentos e o último experimento foi um sucesso. A probabilidade de cada resultado com trials and the last trial is a sucess. A probabilidade de cada resultado com r sucessos e $i-r$ falhas em i experimentos é $p^r(1-p)^{i-r}$. Além disso, existem $\binom{i-1}{r-1}$ permutações possíveis das $i-r$ falhas e dos $r-1$ primeiros sucessos. Como todas estas permutações são eqiprováveis, $\mathbb{P}(Y = i) = \binom{i-1}{r-1}p^r(1-p)^{i-r}$. □

Exemplo 48. *Observe que se $Y \sim$ Binomial Negativa$(1,p)$, $\mathbb{P}(Y = i) = \binom{i-1}{1-1}p^1(1-p)^{i-1} = p(1-p)^{i-1}$. Esta expressão é exatamente a fmp de uma distribuição Geométrica e, portanto, $Y \sim$ Geom(p). Isto é, a distribuição Geométrica é um caso particular da Binomial Negativa quando $r = 1$.*

Lema 40. *Se Y tem distribuição Binomial Negativa(r,p), então $\mathbb{E}[Y] = \frac{r}{p}$ e $Var[Y] = \frac{r(1-p)}{p^2}$.*

Demonstração. Lembre que se Y tem distribuição Binomial Negativa(r, p), então existem $X_1, \ldots, X_r$ variáveis independentes tais que $X_i \sim$ Geom(p) e $Y = \sum_{i=1}^{r} X_i$.

$$\mathbb{E}[Y] = \mathbb{E}[\sum_{i=1}^{r} X_i] = \sum_{i=1}^{r} \mathbb{E}[X_i] = \sum_{i=1}^{r} \frac{1}{p} = \frac{r}{p}$$

Similarmente, como os X_i's são independentes,

$$Var[Y] = Var[\sum_{i=1}^{r} X_i] = \sum_{i=1}^{r} Var[X_i] = \sum_{i=1}^{r} \frac{1-p}{p^2} = \frac{r(1-p)}{p^2}$$

□

Exemplo 49 (Programação)**.** *Escreva um código que gera um número de acordo com a distribuição Binomial Negativa (r, p). Considere as funções rbernoulli(p) e rgeom(p) nos Exemplos 42 e 47.*

```
def rnbinom(r,p):
  sum = 0
  for ii in xrange(r):
    sum += rgeom(p)

def rnbinom(r,p):
  number_of_ones = value = 0
  while(number_of_ones < r):
    value += 1
    number_of_ones += rbernoulli(p)
  return value
```

4.4.1 Exercícios

Exercício 102. *Um poço exploratório tem probabilidade de 10% de atingir petróleo. Considere que poços diferentes são independentes:*

a. *Qual é a probabilidade de que a* $3a$ *vez que um poço atinge o petróleo ocorre na* $8a$ *tentativa?*

b. *Qual é o número esperado de tentativas até que o petróleo é atingido pela* $5th$ *vez?*

c. *Qual é a variância do número de tentativas até que a empresa atinge opetróleo pela* $8a$ *vez?*

Solução:

a. Seja X o número de tentativas até atingir o petróleo pela terceira vez. Como as tentativas são independentes, $X \sim$ Binomial Negativa$(3, 0.1)$. Portanto, usando o Lema 39, $P(X = 8) = \binom{7}{2}(0.9)^5(0.1)^3$.

b. Seja Y o número de tentativas até atingir o petróleo pela 5a vez. Como as tentativas são independentes, $X \sim$ Binomial Negativa$(5, 0.1)$. Portanto, usando o Lema 40, $E[Y] = \frac{5}{0.1} = 50$.

c. Seja Z o número de tentativas até atingir o petróleo pela 8a vez. Como as tentativas são independentes, $Z \sim$ Binomial Negativa$(8, 0.1)$. Portanto, usando o Lema 40, $Var[Z] = \frac{50.9}{(0.1)^2} = 450$.

Exercício 103. *Considere que lanço uma moeda 5 vezes. Seja X o número total de caras. Depois destes lançamentos, eu lanço a mesma moeda até obter 2 caras. Seja Y o número total de lançamentos no segundo caso. Quais são as distribuiçÕes de X e Y? $\mathbb{P}(X = 2) = \mathbb{P}(Y = 5)$? Interprete este resultado.*

Solução: X corresponde ao número de sucessos em 5 lançamentos independentes. Portanto $X \sim \text{Binomial}(5, p)$. Y corresponde ao número de lançamentos até que observamos 2 caras. Portanto, $Y \sim \text{Binomial Negativa}(2, p)$. Usando Lemas 32 e 39,

$$P(X = 2) = \binom{5}{2}p^2(1-p)^3 \neq \binom{4}{1}p^2(1-p)^3 = P(Y = 5)$$

Obter 2 sucessos em 5 lançamentos é mais provável que esperar até o 5o lançamento para obter 2 sucessos.

Exercício 104. *Considere os mesmos X e Y do Exercício 103. Também considere que eu escolhi uma moeda, com probabilidades iguais, entre uma moeda honesta e uma moeda com probabilidade 0.1 de cara, e não contei qual das duas escolhi. Seja F o evento em que a moeda é honesta. Calcule $\mathbb{P}(F|X = 2)$ e $\mathbb{P}(F|Y = 2)$.*

Solução:

$$\begin{aligned} P(F|X = 2) &= \frac{P(F)P(X = 2|F)}{P(F)P(X = 2|F) + P(F^c)P(X = 2|F^c)} && \text{Teorema de Bayes (Teorema 4)} \\ &= \frac{0.5 \cdot \binom{5}{2}(0.5)^2(1-0.5)^3}{0.5 \cdot \binom{5}{2}p^2(1-p)^3 + 0.5 \cdot \binom{5}{2}p^2(1-p)^3} \\ &= \frac{(0.5)^5}{(0.5)^5 + p^2(1-p)^3} \end{aligned}$$

$$\begin{aligned} P(F|Y = 5) &= \frac{P(F)P(Y = 5|F)}{P(F)P(Y = 5|F) + P(F^c)P(Y = 5|F^c)} && \text{Teorema de Bayes (Teorema 4)} \\ &= \frac{0.5 \cdot \binom{4}{1}(0.5)^2(1-0.5)^3}{0.5 \cdot \binom{4}{1}p^2(1-p)^3 + 0.5 \cdot \binom{4}{1}p^2(1-p)^3} \\ &= \frac{(0.5)^5}{(0.5)^5 + p^2(1-p)^3} \end{aligned}$$

That is, $P(F|X = 2) = P(F|Y = 5)$.

Exercício 105. *Argumente que se $X \sim$ Binomial Negativa(r, p) e $Y \sim$ Binomial(n, p), então $\mathbb{P}(X > n) = \mathbb{P}(Y < r)$.*

Exercício 106. *Você tem duas jarras com balas, cada uma com n balas. Toda vez que você quer uma bala, você escolhe ao acaso se vai pegar ela da jarra A ou da jarra B. Qual é a probabilidade de que, quando você esvaziar a primeira jarra, sobrarem exatamente i balas na outra?*

4.5 Distribuição Hipergeométrica

Uma população tem N indivíduos e k deles tem uma certa característica de interesse. Considere uma amostra de tamanho n retirada sem reposição desta população, de tal forma que todos os grupos de tamanho n são equiprováveis. Seja X o número de membros da população com a característica de interesse que foram amostrados. Dizemos que X tem distribuição Hipergeométrica com parâmetros N, n e k e escrevemos $X \sim \text{Hipergeométrica}(N, n, k)$.

Lema 41. *Se $X \sim$ Hipergeométrica(N, n, k), então* $\mathbb{P}(X = i) = \frac{\binom{k}{i}\binom{N-k}{n-i}}{\binom{N}{n}}$.

Demonstração. Lembramos que existem $\binom{N}{n}$ maneiras de escolher um grupo de n da populaçao de tamanho N. Como todos os grupos são equiprováveis, segue do Lema 7 que é suficiente determinar o número de grupos de tamanho n com i indivíduos com a característica de interesse. Observe que tal grupo deve conter i indivíduos com a característica de interesse e $n - i$ indivíduos sem a característica. Isto é, queremos determinar de quantas formas é possível selecionar um grupo de i indivíduos entre k com a característica e um grupo de $n - i$ indivíduos entre os $N - k$ que não têm a característica de interesse. Este número é $\binom{k}{i}\binom{N-k}{n-i}$. Concluímos que $P(X = i) = \frac{\binom{k}{i}\binom{N-k}{n-i}}{\binom{N}{n}}$. □

Lema 42. *Considere que existem N indivíduos e k deles têm uma dada propriedade. Uma maostra de tamanho n é retirada da população sem reposição e todos os grupos de tamanho n são igualmente prováveis. Seja M_i o evento em que o i-ésimo indivíduo sorteado tem a propriedade de interesse. Para todo $1 \leq n \leq N$, $P(M_n) = \frac{k}{N}$.*

Demonstração. Escolhemos o primeiro indivíduo com probabilidade igual entre os N e, portanto, $P(M_1) = \frac{k}{N}$. Seja X_n o número de indivíduos com a propriedade de interesse que sorteamos apos n sorteios. Como a amostra é selecionada sem reposição, observe que $P(M_{n+1}|X_n = j) = \frac{k-j}{N}$. Também observe que $X_n \sim$ Hipergeométrica(N, n, k). Portanto,

$$\begin{aligned}
\mathbb{P}(M_{n+1}) &= \sum_{j=0}^{n} \mathbb{P}(M_{n+1}|X_n = j)\mathbb{P}(X_n = j) && \text{(Teorema 3)} \\
&= \sum_{j=0}^{n} \frac{k-j}{N} \cdot \frac{\binom{k}{j}\binom{N-k}{n-j}}{\binom{N}{n}} && \text{(Lema 41)} \\
&= \sum_{j=0}^{n} \frac{k-j}{N} \cdot \frac{\frac{k!}{j!(k-j)!} \cdot \frac{(N-k)!}{(n-j)!(N-k-n+j)!}}{\frac{N!}{n!(N-n)!}} \\
&= \sum_{j=0}^{n} \frac{\frac{k!}{j!(k-j-1)!} \cdot \frac{(N-k)!}{(n-j)!(N-k-n+j)!}}{\frac{(N-1)!}{n!(N-n)!}} \\
&= \frac{k}{N} \sum_{j=0}^{n} \frac{\frac{(k-1)!}{j!(k-j-1)!} \cdot \frac{(N-k)!}{(n-j)!(N-k-n+j)!}}{\frac{(N-1)!}{(n-1)!(N-n)!}} \\
&= \frac{k}{N} \cdot \sum_{j=0}^{n} \frac{\binom{k-1}{j}\binom{N-k}{n-j}}{\binom{N-1}{n}} = \frac{k}{N}
\end{aligned}$$

A última igualdade vale pois $\frac{\binom{k-1}{j}\binom{N-k}{n-j}}{\binom{N-1}{n}}$ é a fmp de uma Hipergeométrica$(N - 1, n, k - 1)$ e a soma de uma fmp para todos os valores possíveis é 1 (Lema 11). □

Lema 43. *Se $X \sim$ Hipergeométrica(N, n, k), então* $\mathbb{E}[X] = \frac{nk}{N}$ *e* $Var[X] = \frac{nk}{N}\frac{(N-k)}{N}\frac{N-n}{N-1}$.

Demonstração. Seja M_i o evento em que o i-ésimo elemento da amostra tem a propriedade de interesse. Também observe que $X = \sum_{i=1}^{n} I_{M_i}$ e $I_{M_i} \sim$ Bernoulli$(P(M_i))$. Dessa forma,

$$
\begin{aligned}
E[X] &= E[\sum_{i=1}^{n} I_{M_i}] \\
&= \sum_{i=1}^{n} E[I_{M_i}] && \text{Aditividade do Valor Esperado (Lema 14)} \\
&= \sum_{i=1}^{n} P(M_i) && \text{Propriedades da Bernoulli (Lema 31)} \\
&= \sum_{i=1}^{n} \frac{k}{N} = \frac{nk}{N} && \text{(Lema 42)}
\end{aligned}
$$

Não obstante, como a amostragem é sem reposição, os I_{H_i} não são independentes. Portanto, não podemos usar o mesmo truque para calcular a variância. Em vez disso, determinamos $\mathbb{E}[X(X-1)]$,

$$
\begin{aligned}
\mathbb{E}[X(X-1)] &= \sum_{i=0}^{n} i(i-1)\mathbb{P}(X=i) = \sum_{i=1}^{n} \frac{i(i-1)\binom{k}{i}\binom{N-k}{n-i}}{\binom{N}{n}} \\
&= \sum_{i=1}^{n} \frac{i(i-1)\frac{k!}{i!(k-i)!}\binom{N-k}{n-i}}{\frac{N!}{n!(N-n)!}} \\
&= \sum_{i=1}^{n} \frac{(i-1)\frac{k(k-1)!}{(i-1)!(r-i)!}\binom{N-k}{n-i}}{\frac{N(N-1)!}{n(n-1)!(N-n)!}} \\
&= \frac{kn}{N} \sum_{i=1}^{n} \frac{(i-1)\binom{k-1}{i-1}\binom{N-k}{n-i}}{\binom{N-1}{n-1}} \qquad (17)
\end{aligned}
$$

Seja $Y \sim$ Hipergeométrica$(N-1, n-1, k-1)$. Observe que $\mathbb{P}(Y = i-1) = \frac{\binom{k-1}{i-1}\binom{N-k}{n-i}}{\binom{N-1}{n-1}}$. Portanto a equação 17 simplifica para

$$
\begin{aligned}
\mathbb{E}[X(X-1)] &= \frac{kn}{N} \sum_{i=1}^{n} (i-1)\mathbb{P}(Y = i-1) \\
&= \frac{kn}{N} \sum_{j=0}^{n-1} jP(Y=j) \\
&= \frac{kn}{N} \mathbb{E}[Y] = \frac{kn}{N} \frac{(k-1)(n-1)}{N-1}
\end{aligned}
$$

Finalmente, lembre que

$$
\begin{aligned}
Var[X] &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 && \text{(Lema 17)} \\
&= \mathbb{E}[X^2] - \mathbb{E}[X] + \mathbb{E}[X] - \mathbb{E}[X]^2 \\
&= \mathbb{E}[X(X-1)] + \mathbb{E}[X] - \mathbb{E}[X]^2 && \text{Aditividade do Valor Esperado (Lema 14)} \\
&= \frac{kn}{N} \cdot \frac{(k-1)(n-1)}{N-1} + \frac{kn}{N} - \left(\frac{kn}{N}\right)^2 \\
&= \frac{kn}{N} \cdot \frac{N-k}{N} \cdot \frac{N-n}{N-1}
\end{aligned}
$$

□

Exemplo 50 (Programação)**.** *Escreva um código que gera um número de acordo com a Hipergeométrica*(N, n, k).

```
def rhyper(N,n,k):
k = float(k)
number_of_ones = 0
for(ii in range(n)):
if(k > 0 and rbernoulli(k/N)):
k -= 1
number_of_ones += 1
return number_of_ones
```

4.5.1 Exercícios

Exercício 107. *Para vender um lote de brinquedos no Brasil, o lote deve passar um teste de segurança dos brinquedos. Este teste consiste em verificar a segurança de uma amostra sem reposição dos brinquedos do lote. Se algum dos brinquedos amostrados não é seguro, nenhum dos produtos do lote pode ser vendido. Assuma que um lote tem* 100 *produtos e* 10 *deles são deietuosos e não seguros. Qual é o menor tamanho amostral tal que a probabilidade de que o lote não passa no teste é maior que* 95%*?*

Solução: Seja Y o número de brinquedos defituosos na amostra de tamanho n. $Y \sim$ Hipergeométrica$(100, n, 10)$. O lote não passa no teste se $Y \geq 1$. Portanto, queremos que $\mathbb{P}(Y = 0) \leq 0.05$.

$$
\begin{aligned}
P(Y=0) &= \frac{\binom{10}{0}\binom{90}{n}}{\binom{100}{n}} = \frac{\frac{90!}{n!(90-n)!}}{\frac{100!}{n!(100-n)!}} \\
&= \frac{90!(100-n)!}{100!(90-n)!}
\end{aligned}
$$

Depois de alguns cálculos na força bruto, determinei que $n = 25$ é o menor tamanho amostral tal que $P(Y = 0) \leq 0.05$.

Exercício 108. *Uma caixa tem* 5 *bolas laranjas e* 8 *bolas rosas.*

a. *Suponha que* 3 *bolas são retirads sem reposição. Seja X o número total de bolas rosas. Encontre a distribuição de X, $\mathbb{E}[X]$ e $Var[X]$.*

b. *Suponha que* 3 *bolas são retiradas com reposição. Seja Y o número total de bolas rosas. Encontre a distribuição de Y, $\mathbb{E}[Y]$ e $Var[Y]$.*

c. *Compare $Var[X]$ com $Var[Y]$.*

Solução:

a. $X \sim$ Hipergeométrica$(13, 3, 8)$, $E[X] = 3 \cdot \frac{8}{13} = \frac{24}{13}$ and $Var[X] = \frac{24}{13} \cdot \frac{5}{13} \cdot \frac{10}{12} = \frac{100}{169}$.

b. $Y \sim$ Binomial$(3, \frac{8}{13})$, $E[Y] = 3 \cdot \frac{8}{13} = \frac{24}{13}$ and $Var[Y] = 3 \cdot \frac{8}{13} \cdot \frac{5}{13} = \frac{120}{169}$.

c. $Var[X] \leq Var[Y]$. Amostragem sem reposição reduz a variância.

Exercício 109. *Qual é a distribuição do número de caras que obtenho nos primeiros* 10 *lançamentos de uma moeda com viés p dado que nos primeiros* 20 *lançamentos eu obtive* 11 *caras?*

Seja X_{1-10} o número de caras nos primeiros 10 lançamentos, X_{11-20} o número de caras nos próximos 10 lançamentos e X_{1-20} o número de caras nos primeiros 20 lançamentos. X_{1-10} e X_{11-20} são independentes, $X_{1-10} \sim$ Binomial$(10, p)$, $X_{11-20} \sim$ Binomial$(10, p)$ e $X_{1-20} \sim$ Binomial$(20, p)$.

$$\begin{aligned}
P(X_{1-10} = i | X_{1-20} = 11) &= \frac{P(X_{1-10} = i \cap X_{1-20} = 11)}{P(X_{1-20} = 11)} \\
&= \frac{P(X_{1-10} = i \cap X_{11-20} = 11 - i)}{P(X_{1-20} = 11)} \\
&= \frac{P(X_{1-10} = i)P(X_{11-20} = 11 - i)}{P(X_{1-20} = 11)} \\
&= \frac{\binom{10}{i} p^i (1-p)^{10-i} \binom{10}{11-i} p^{11-i} (1-p)^{i-1}}{\binom{20}{11} p^{11} (1-p)^{20-11}} \\
&= \frac{\binom{10}{i}\binom{10}{11-i}}{\binom{20}{11}}
\end{aligned}$$

A distribuição do número de caras que observo nos primeiros 10 lançamentos da moeda dado que nos primeiros 20 lançamentos eu observei 11 caras é Hipergeométrica$(20, 10, 11)$.

Exercício 110 (Challenge)**.** *Escreva um código que retorna, com igual probabilidade, uma das permutações de* $(0, \ldots, n-1)$.

Solução: O codigo a seguir é $O(n)$. Porque ele funciona?

```
import random
def rpermutation(n):
permutation = range(n)
for(ii in range(n)):
jj = random.randint(ii,n-1)
permutation[ii],permutation[jj] = permutation[jj],permutation[ii]
return permutation
```

4.6 Distribuição de Poisson

A distribuição de Poisson é normalmente usada para modelar a ocorrência de eventos raros e pode ser usada como uma aproximação da distribuição Binomial(n, p) quando n é grande e p é pequeno. Formalmente, se X tem distribuição de Poisson com parâmetro λ, escrevemos $X \sim$ Poisson(λ). A Poisson é definida de tal forma que $X \in \mathbb{N}$ tem fmp:

$$\mathbb{P}(X = i) = \frac{e^{-\lambda}\lambda^i}{i!}$$

Lema 44 (Expansão de Taylor de e^x)**.**

$$e^x = \sum_{i=0}^{\infty} \frac{x^i}{i!}$$

Lema 45. *Se $X \sim Poisson(\lambda)$, então $\mathbb{E}[X] = \lambda$ e $Var[X] = \lambda$.*

Demonstração.

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{i=0}^{\infty} i\mathbb{P}(X = i) \\
&= \sum_{i=1}^{\infty} i\frac{e^{-\lambda}\lambda^i}{i!} \\
&= \sum_{i=1}^{\infty} \frac{ie^{-\lambda}\lambda\lambda^{i-1}}{i(i-1)!} \\
&= \lambda\sum_{i=1}^{\infty} \frac{e^{-\lambda}\lambda^{i-1}}{(i-1)!} \\
&= \lambda\sum_{j=0}^{\infty} \frac{e^{-\lambda}\lambda^{j}}{j!} \\
&= \lambda\sum_{j=0}^{\infty} \mathbb{P}(X = j) = 1
\end{aligned}$$

A seguir, determinamos $\mathbb{E}[X(X-1)]$ para achar $Var[X]]$

$$\begin{aligned}
\mathbb{E}[X(X-1)] &= \sum_{i=0}^{\infty} i(i-1)\mathbb{P}(X = i) \\
&= \sum_{i=2}^{\infty} i(i-1)\frac{e^{-\lambda}\lambda^i}{i!} \\
&= \lambda^2\sum_{i=2}^{\infty} \frac{e^{-\lambda}\lambda^{i-2}}{(i-2)!} \\
&= \lambda^2\sum_{j=0}^{\infty} \mathbb{P}(X = j) = \lambda^2
\end{aligned}$$

□

Finalmente,

$$\begin{aligned}
Var[X] &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\
&= \mathbb{E}[X(X-1)] + \mathbb{E}[X] - \mathbb{E}[X]^2 \\
&= \lambda^2 + \lambda - \lambda^2 = \lambda
\end{aligned}$$

4.6.1 Exercícios

Exercício 111. *A demanda média por um certo produto é* 4 *por dia. O proprietário da loja quer ter 99% de ceretza de que não vai faltar este produto durante o dia. Quantas unidades do produto ele deve ter no estoque?*

Solução: Seja X a demanda pelo produto num certo dia. Da descrição do problema, podemos modelar $X \sim$ Poisson(4). Nos queremos encontrar os valores de i tais que $P(X \leq i) \geq 0.99$. Observe que $P(X \leq i) = \sum_{j=0}^{i} \frac{e^{-4}4^j}{j!}$. Calculando esta soma para diferentes valores de i, obtemos,

i	$P(X \leq i)$
0	0.02
1	0.09
2	0.24
3	0.43
4	0.63
5	0.79
6	0.90
7	0.95
8	0.98
9	0.99

Portanto, o proprietário da loja deve ter pelo menos 9 produtos em estoque.

Exercício 112. *Numa linha de transmissão, há uma falha na isolação a cada 2.5 milhas, em média. Qual é a probabilidade de ter* 2 *faltas em menos de* 7.5 *milhas?*

Solução: Seja X o número de falhas na isolação em 7.5 milhas. Da descrição do problema, pode-se modelar $X \sim$ Poisson(3). Desejamos encontrar $P(X \geq 2)$.

$$\begin{aligned} P(X \geq 2) &= 1 - P(X < 2) \\ &= 1 - P(X = 0) - P(X = 1) \\ &= 1 - \frac{e^{-3}3^0}{0!} - \frac{e^{-3}3^1}{1!} \\ &\approx 1 - 0.05 - 0.15 = 0.8 \end{aligned}$$

Exercício 113. *O número médio de terremotos de magnitude oito ou mais na escala de Richter é* 1 *por ano. Seja X o número de terremotos de magnitude oito ou mais no ano seguinte. Encontre $\mathbb{P}(X = 0)$ e $Var[X]$.*

Solução: Da descrição do problema, podemos modelar $X \sim$ Poisson(1). Portanto, $P(X = 0) = e^{-1}$ e $Var[X] = 1$.

4.7 Revisão das Distribuições Discretas Especiais

1. **Variável Aleatória Binomial** - $X \sim \text{Binomial}(n, p)$.

   X conta o número de sucessos em n experimentos independentes (Bernoulli), cada um com probabilidade de "sucesso" p.

2. **Variável Aleatória Geométrica** - $X \sim \text{Geom}(p)$.

   Considere uma série de experimentos independentes (Bernoulli), cada um com probabilidade de "sucesso" p. X é o número de experimentos até o primeiro "sucesso".

3. **Variável Aleatória Binomial Negativa** - $X \sim \text{Binomial Negativa}(r, p)$.

   Considere uma série de experimentos independentes (Bernoulli), cada um com probabilidade de "sucesso" p. X é o número de experimentos até obtermos r "sucessos". Por definição, Binomial Negativa$(1, p)$ é o mesmo que $Geom(p)$.

4. **Variável Aleatória Hipergeométrica** - $X \sim \text{Hipergeométrica}(N, n, k)$.

   Uma amostra de tamanho n é escolhida (sem respoição) de um grupo de tamanho N que tem k "sucessos" (e $N - k$ "falhas"). X é o número de "sucessos" na amostra.

5. **Variável Aleatória Poisson** - $X \sim \text{Poisson}(\lambda)$.

   A família de distribuições de Poisson frequentemente dá um bom modelo para o número de eventos (em particular, eventos raros) que ocorrem num período de tempo fixo (ou outra unidade fixa). Por exemplo, o número de clientes chegando em uma hora, número de reivindicações de seguro em um mês, número de terremotos num ano, número de erros de digitação numa página. λ é o número médio de eventos.

Variável Aleatória (Y)	fmp: $p_Y(y) = P(Y = y)$	$E[Y]$	$Var[Y]$
Binomial(n, p)	$\binom{n}{y} p^y (1-p)^{n-y}$ $y \in \{0, 1, 2, \ldots, n\}$	np	$np(1-p)$
Hipergeométrica(N, n, k)	$\frac{\binom{k}{y}\binom{N-k}{n-y}}{\binom{N}{n}}$ $\max(0, n-N+k) \leq y \leq \min(n, k)$	$n \cdot \frac{k}{N}$	$n \cdot \frac{k}{N} \cdot (1 - \frac{k}{N}) \cdot \frac{N-n}{N-1}$
Geométrica(p)	$p(1-p)^{y-1}$ $y \in \{1, 2, 3, \ldots\}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Binomial Negativa(r, p)	$\binom{y-1}{r-1} p^r (1-p)^{y-r}$ $y \in \{r, r+1, r+2, \ldots\}$	$r \cdot \frac{1}{p}$	$r \cdot \frac{1-p}{p^2}$
Poisson(λ)	$\frac{e^{-\lambda}\lambda^y}{y!}$ $y \in \mathbb{N}$	λ	λ

Tabela 2

4.7.1 Exercícios

Exercício 114. *Você consegue dar um exemplo de uma variável aleatória X tal que $\mathbb{E}[X(X+1)] = 10$ e $\mathbb{E}[X] = 5$?*

Exercício 115. *Sejam A e B eventos tais que $\mathbb{P}(A) = 0.4$, $\mathbb{P}(B) = 0.8$ e $\mathbb{P}(A \cap B) = 0.3$. Encontre o valor de $\mathbb{E}[I_A + I_B]$ e de $Var[I_A + I_B]$.*

Exercício 116. *Um sistema é composto de* 4 *componentes e funciona quando pelo menos* 3 *deles estão funcionando. A cada minuto, cada um dos componentes está funcionando de forma independente com probabilidade* p*. Seja* X *a variável aleatória que representa o número de minutos até que o sistema funcione pela primeira vez. Encontre a distribuição de* X*. Qual é a probabilidade de que o sistema irá funcionar em* 5 *minutos ou menos?*

Exercício 117.

a. *Seja* $X \sim$ *Bernoulli*(p)*. Calcule* $\mathbb{E}[e^{tX}]$.

b. *Sejam* $X_1, \ldots, X_n$ *independentes com distribuição Bernoulli*(p)*. Calcule* $\mathbb{E}[e^{t\sum_{i=1}^n X_i}]$.

c. *Seja* $Y \sim$ *Binomial*(n,p)*. Use os itens anteriores para calcular* $\mathbb{E}[e^{tY}]$.

Exercício 118. *Mensagens chegam em um servidor com uma taxa média de* 6 *mensagens por hora. Qual é a probabilidade de que, em uma dada hora,* 4 *ou mais mensagens cheguem no servidor?*

Exercício 119. *Eu conto a você que peguei aleatoriamente uma moeda entre uma moeda honesta e outra que tem probabilidade de cara igual a* 0.25*, sendo que cada uma delas tinha igual probabilidade de ser pega. Qual a probabilidade da moeda pega ser a honesta dado que eu joguei ela* 5 *vezes até obter* 2 *caras?*

Exercício 120. *Existem* N *bolas brancas dentro de uma caixa. Uma pessoa pinta, de maneira independente e com probabilidade* p*, cada bola da caixa usando tinta vermelha. Seja* X *a variável aleatória que denota o número de bolas vermelhas na caixa. Qual é a distribuição de* x*? Em seguida, você pega uma amostra sem reposição de tamanho* n *de bolas da caixa. Chame o número de bolas vermelhas que você obtiver de* Y*. Qual a distribuição de* Y *dado que* $X = k$*? Qual a distribuição de* Y *se você não sabe o valor de* X*?*

Exercício 121. *Uma população tem tamanho* 20*.* 10 *indivíduos foram infectados com uma certa doesnça. Um certo teste detecta os indivíduos infectados com probabilidade* p *e sempre diz que um indivíduo saudável não está infectado. Se* 2 *indivíduos são amostrados sem reposição, qual a probabilidade de que o teste não detecta nenhum indivíduo infectado?*

Exercício 122. *Seja* $X \sim$ *Poisson*(λ).

a. *Encontre* $f(t) = \mathbb{E}[t^X]$.

b. *Calcule* $\frac{\partial f}{\partial t}(0)$.

c. *Calcule* $\frac{\partial^2 f}{\partial t^2}(0)$.

d. *Compare os itens anteriores com* $\mathbb{E}[X]$ *e* $\frac{1}{2} \cdot \mathbb{E}[X^2]$.

Exercício 123. *Uma pessoa* A *lança uma moeda honesta* 5 *vezes. A pessoa* B *lança uma moeda honesta* 6 *vezes. Sejam* X *e* Y *respectivamente o número de caras obtidas pelas pessoas* A *e* B*. Determine* $\mathbb{P}(X > Y)$.

Exercício 124. *Uma urna contem 3 bolas pretas e 4 bolas brancas. Você pega 4 bolas ao acaso. Se exatamente duas destas bolas são pretas, você para. Caso contrário, você devolve estas bolas à urna e repete todo o procedimento. Isso continua até você retirar exatamente duas bolas pretas. Qual é a probabilidade de que você fará exatamente* n *retiradas?*

5 Variáveis Aleatórias Contínuas

5.1 Introdução

Variáveis aleatórias contínuas podem assumir um número não enumerável de valores. Elas são usadas, frequentemente, para modelar quantidades como tempo, Continuous random variables can assume an uncountable number of values. They are commonly used to model quantities such as time, retorno de ações, altura, peso, etc ... Enquanto a disribuição de uma variável aleatória discreta é definida pela fmp, a distribuição de uma variável aleatória contínua é definida pela sua função de densidade de probabilidade.

Definição 27. *Seja X uma variável aleatória contínua. Denotamos a função de densidade de probabilidade de X por $f_X : \mathbb{R} \to \mathbb{R}$. Ela satisfaz as seguintes propriedades:*

1. $f_X(x) \geq 0$.
2. $\int_{-\infty}^{\infty} f_X(x)dx = 1$.
3. $\int_a^b f_X(x)dx = \mathbb{P}(a \leq X \leq b)$.

Lema 46. *Seja X uma variável aleatória contínua com densidade $f_X(x)$. Para todo $x \in \mathbb{R}$, $\mathbb{P}(X = x) = 0$.*

Demonstração. Observe que $\mathbb{P}(X = x) = \mathbb{P}(x \leq X \leq x)$. Usando a Definição 27, $\mathbb{P}(x \leq X \leq x) = \int_x^x f_X(x)dx = 0$. □

Exemplo 51. *Suponha que X tem a pdf:*

$$f_X(x) = \begin{cases} 0 & , \text{ if } x < 0 \\ 3x^2 & , \text{ if } 0 \leq x \leq 1 \\ 0 & , \text{ if } x > 0 \end{cases}$$

f_X satisfaz as propriedades (1) e (2):

$$\begin{aligned} \mathbb{P}(X \in \mathbb{R}) &= \int_{-\infty}^{\infty} f_X(x)dx \\ &= \int_0^1 3x^2 \\ &= x^3\Big|_0^1 = 1 \end{aligned}$$

Podemos usar f_X para determinar $\mathbb{P}(X < 0.5)$,

$$\begin{aligned} \mathbb{P}(X < 0.5) &= \int_{-\infty}^{0.5} 3x^2dx \\ &= \int_0^{0.5} 3x^2dx \\ &= x^3\Big|_0^{0.5} = 0.125 \end{aligned}$$

Exemplo 52. *Suponha que $X \geq 1$ e $f_X(x) = \frac{c}{x^3}$ para algum $c \in \mathbb{R}$. Qual é o único valor de c tal que f_X é uma função de densidade legítima?*

$$
\begin{aligned}
\int_{-\infty}^{\infty} f_X(x)dx &= \int_{-\infty}^{\infty} \frac{c}{x^3}dx \\
&= -\frac{c}{3}x^{-3}|_1^{\infty} = \frac{c}{3}
\end{aligned}
$$

Para que f_X seja legítima, $\int_{-\infty}^{\infty} f_X(x)dx = 1$. Portanto $c = 3$.

Exemplo 53. *Suponha que $X \geq 0$ é uma variável aleatória contínua com densidade $f_X(x) = ce^{-\lambda x}$, para algum $c > 0$.*

$$
\begin{aligned}
\int_{-\infty}^{\infty} f_X(x)dx &= \int_{0}^{\infty} ce^{-\lambda x}dx \\
&= -\frac{c}{\lambda}e^{-\lambda x}|_0^{\infty} \\
&= \frac{c}{\lambda} = 1
\end{aligned}
$$

Portanto, $c = \lambda$.

Definição 28. *Seja X uma variável aleatória contínua com densidade $f_X(x)$. O valor esperado de $g(X)$ é*

$$
\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x)f_X(x)dx
$$

Comentário: Todas as regras para $\mathbb{E}[\cdot]$ que vimos no caso discreto podem ser aplicadas para variáveis aleatórias contínuas. Por exemplo, a aditividade dos valores esperados (Lema 14) ainda se aplica.

Exemplo 54. *Considere o Exemplo 51.*

$$
\begin{aligned}
\mathbb{E}[X] &= \int_{-\infty}^{\infty} xf_X(x)dx \\
&= \int_{0}^{1} x3x^2dx \\
&= \int_{0}^{1} 3x^3dx \\
&= \frac{3}{4}x^4|_0^1 = \frac{3}{4}
\end{aligned}
$$

Além disso,

$$
\begin{aligned}
\mathbb{E}[X^2] &= \int_{-\infty}^{\infty} x^2 f_X(x)dx \\
&= \int_0^1 x^2 3x^2 dx \\
&= \int_0^1 3x^4 dx \\
&= \frac{3}{5}x^5\big|_0^1 = \frac{3}{5}
\end{aligned}
$$

Portanto, $Var[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{3}{5} - \frac{9}{16} = 0.0375$.

Definição 29. *A função de distribuição acumulada (fda) de uma variável aleatória* X *é a função* $F_X : \mathbb{R} \to \mathbb{R}$,

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

Lema 47. *A fda* F *de uma variável aleatória* X *deve satisfazer as seguintes propriedades, mesmo se* X *não é contínua:*

1. F *é não decrescente.*

2. $\lim_{x \to -\infty} F(x) = 0$.

3. $\lim_{x \to \infty} F(x) = 1$.

4. F *é contínua à direita.*

Demonstração.

1. Note que, para todo $t_1 \leq t_2$, $\{\omega \in \Omega : X(\omega) \leq t_1\} \subseteq \{\omega \in \Omega : X(\omega) \leq t_2\}$. Segue do Lema **??** que $F(t_1) := \mathbb{P}(X \leq t_1) \leq \mathbb{P}(X \leq t_2) := F(t_2)$.

2. Seja $(y_n)_{n\in\mathbb{N}}$ uma sequência tal que $y_n \longrightarrow -\infty$. Defina $B_n := \{\omega \in \Omega : X(\omega) \leq y_n\}$. Por construção, $B_{n+1} \subseteq B_n$, and $\cap_{n\in\mathbb{N}} B_n = \emptyset$. Do exercício 21, segue que

$$\lim_{x\to-\infty} F(x) = \lim_{n\to\infty} F(y_n) = \lim_{n\to\infty} \mathbb{P}(B_n) = \mathbb{P}(\cap_{n\in\mathbb{N}} B_n) = \mathbb{P}(\emptyset) = 0$$

3. Deixamos esta demonstração como exercício.

4. Para provar que F é contínua à direita, é necessário mostrar qu, para todo t_0, $F(t_0) = \lim_{t\downarrow t_0} F(t)$. Seja $(t_n)_{n\in\mathbb{N}}$ uma sequência tal que $t_n \downarrow t_0$ (i.e., $t_n \geq t_{n+1}$ para todo n). Defina $B_n := \{\omega \in \Omega : X(\omega) \leq t_n\}$. Por construção, $B_{n+1} \subseteq B_n$, and $\cap_{n\in\mathbb{N}} B_n = \{\omega \in \Omega : X(\omega) \leq t_0\}$. Do Exercício 21, segue que

$$\lim_{t\downarrow t_0} F(x) = \lim_{n\to\infty} F(t_n) = \lim_{n\to\infty} \mathbb{P}(B_n) = \mathbb{P}(\cap_{n\in\mathbb{N}} B_n) = \mathbb{P}(X \leq t_0) = F(t_0).$$

□

Para variáveis aleatórias contínuas, F é sempre contínua. Contudo, para variáveis aleatórias discretas, F tem discontinuidades à esquerda nos valores da imagem de X.

Exemplo 55. *Considere Exemplo 51. Para* $0 \leq x \leq 1$,

$$
\begin{aligned}
F(x) &= \int_{-\infty}^{x} f_X(x)dx \\
&= \int_{0}^{x} 3x^2 dx \\
&= x^3\big|_0^x = x^3
\end{aligned}
$$

Para $x \leq 0$, $F(x) = 0$ e para $x \geq 1$, $F(x) = 1$.

Lema 48. *Seja X uma variável aleatória contínua com a função de distribuição acumulada F_X. Para $b \geq a$, $F_X(b) - F_X(a) = \mathbb{P}(a \leq X \leq b)$.*

Demonstração. Usando a definição de fda

$$
\begin{aligned}
F(b) - F(a) &= \int_{-\infty}^{b} f_X(x)dx - \int_{-\infty}^{a} f_X(x)dx \\
&= \int_{a}^{b} f_X(x)dx = \mathbb{P}(a \leq X \leq b)
\end{aligned}
$$

□

Lema 49. *Seja $F_X(x)$ a fda de X e $f_X(x)$ a fdp de X.*

$$
\frac{\partial F_X(x)}{\partial x} = f_X(x)
$$

Demonstração.

$$
\begin{aligned}
\frac{\partial F_X(x)}{\partial x} &= \frac{\partial}{\partial x} \int_{-\infty}^{x} f_X(y)dy \\
&= f_X(x) && \text{Teorema Fundamental do Cálculo}
\end{aligned}
$$

□

Lema 50. *Seja X uma variável aleatória contínua com densidade f_X. Então, para todo conjunto $A \in \mathcal{F}$,*

$$
\mathbb{P}(X \in A) = \int_A f_X(x)dx.
$$

Definição 30. *Seja X uma variável aleatória contínua. Denotamos a **mediana** de X por $Med[X]$ e definimos ela como $m \in \mathbb{R}$ tal que $\mathbb{P}(X \leq m) = F(m) = 0.5$. Isto é,*

$$
\int_{-\infty}^{m} f_X(x)dx = 0.5
$$

Exemplo 56. *Considere Exemplo 51.*

$$\begin{aligned}
\mathbb{P}(X \leq m) &= \int_0^m f(x)dx \\
&= \int_0^m 3x^2 dx \\
&= x^3\Big|_0^m = m^3
\end{aligned}$$

Portanto, $Med[X]$ *corresponde a* m *tal que* $m^3 = 0.5$*.* $Med[X] = (0.5)^{\frac{1}{3}}$*.*

5.1.1 Exercícios

Exercício 125. *Seja* $0 \leq X \leq 1$ *uma variável aleatória contínua tal que* $f_X(x) = c$ *para* $0 \leq x \leq 1$*. Encontre* c*,* $\mathbb{E}[X]$*,* $Var[X]$*,* $Med[X]$ *e* $\mathbb{E}[e^{tX}]$ *e* $\mathbb{P}(0.25 \leq X \leq 0.75)$*.*

Exercício 126. *Seja* $X \geq 0$ *uma variável aleatória contínua tal que* $f_X(x) = ce^{-x}$ *para* $x \geq 0$*. Encontre* c*,* $\mathbb{E}[X]$*,* $Var[X]$ *e* $\mathbb{P}(X \geq 1)$*.*

Exercício 127. *Seja* $0 \leq X \leq n$*, para algum* $n \in \mathbb{N}, n > 0$ *Se* $f_X(x) = c$*, qual é o valor de* c*? Calcule* $\mathbb{P}(\leq 1)$*.*

Exercício 128. *Seja* $X \geq 1$*. Pode existir* $c > 0$ *tal que a densidade de* X *é* $f_X(x) = cx^{-1}$*, para* $x \geq 1$*?*

Exercício 129 (Desafio)**.** *Seja* X *uma variável aleatória contínua. Encontre* $d \in \mathbb{R}$ *que minimiza* $\mathbb{E}[|X - d|]$*. (Isto não é necessário para o curso).*

Portanto, para todo $d \in \mathbb{R}$, $E[|X - d|] \geq E[|X - m|]$. Conclua que m minimiza $E[|X - d|]$.

Tal como no caso das variáveis aleatórias discretas, várias distribuições contínuas são frequentemente usadas. Vamos explorar agora algumas delas.

5.2 Distribuição Uniforme

Definição 31. *Dizemos que X tem distribuição uniforme no intervalo (a,b) e denotamos por $X \sim U(a,b)$ se a densidade de X é tal que:*

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

Lema 51. *Se $X \sim U(a,b)$, então para todo conjunto (c,d) tal que $(c,d) \subset (a,b)$, $\mathbb{P}(c \leq X \leq d) = \frac{d-c}{b-a}$. Isto é, a probabilidade de que X esta dentro de um intervalo que é subconjunto de (a,b) é proporcional à largura do intervalo. Assim, quaisque dois intervalos de mesma largura têm a mesma probabilidade.*

Demonstração.

$$\begin{aligned} P(c \leq X \leq d) &= \int_c^d \frac{1}{b-a} dx \\ &= \frac{x}{b-a}\Big|_c^d = \frac{d-c}{b-a} \end{aligned}$$

□

Lema 52. *Se $X \sim U(a,b)$, $\mathbb{E}[X] = \frac{a+b}{2}$ and $Var[X] = \frac{(b-a)^2}{12}$.*

Demonstração.

$$\begin{aligned} \mathbb{E}[X] &= \int_{-\infty}^{+\infty} x \cdot f_X(x) dx \\ &= \int_a^b \frac{x}{b-a} dx \\ &= \frac{x^2}{2(b-a)}\Big|_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{b+a}{2} \end{aligned}$$

$$\begin{aligned} \mathbb{E}[X^2] &= \int_{-\infty}^{+\infty} x^2 \cdot f_X(x) dx \\ &= \int_a^b x^2 \frac{1}{b-a} dx \\ &= \frac{x^3}{3(b-a)}\Big|_a^b = \frac{b^3-a^3}{3(b-a)} = \frac{a^2+ab+b^2}{3} \end{aligned}$$

$$\begin{aligned} Var[X] &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\ &= \frac{a^2+ab+b^2}{3} - \frac{(a+b)^2}{4} \\ &= \frac{4a^2+4ab+4b^2-3a^2-3b^2-6ab}{12} = \frac{(a-b)^2}{12} \end{aligned}$$

$\square$

5.2.1 Exercícios

Exercício 130. *Seja* $X \sim U(0,1)$*. Find* $\mathbb{P}(X \in (-0.5, 0.25) \cup X \in (0.5, 0.75))$*.*

Exercício 131. *Seja* $X \sim U(0,a)$ *e* $Y \sim U(0,b)$*. Se* $f_X(a/2) > f_Y(b/2)$*,* $a < b$*? Porque?*

Exercício 132. *Seja* X *uma variável contínua com fda* $F_X(x)$*. Seja* $U \sim Uniforme(0,1)$*. Seja* $Y = F_X^{-1}(U)$*. Encontre a fda de* Y*.*

Exercício 133 (Programação)**.** *Na linguagem de programação Python, você pode gerar uma Uniforme(0,1) usando a função random.random(). Considere que a função* `Inv_F(x)` *lhe dá a inversa da fda* F*. Use o Exercício 132 para escrever um código curto que gera uma variável aleatória que tem a fda* F*.*

Exercício 134 (Programação)**.** *Use o exercício 132 para escrever uma função que simula uma variável aleatória com distribuição* $U(a,b)$ *usando a função random.random().*

5.3 Distribuição Exponencial

Definição 32. *Dizemos que a variável aleatória* X *tem distribuição exponencial com parâmetro* λ *e denotamos por* $X \sim Exp(\lambda)$ *se a densidade de* X *é:*

$$f_X(x) = \begin{cases} \lambda^{-1} e^{-\frac{x}{\lambda}} & , \text{ if } x \geq 0 \\ 0 & , \text{ caso contrário} \end{cases}$$

Lema 53. *Se* $X \sim Exp(\lambda)$*, então* $F_X(x) = 1 - e^{-\frac{x}{\lambda}}$*. Portanto,* $\mathbb{P}(X > x) = e^{-\frac{x}{\lambda}}$*.*

Demonstração.

$$\begin{aligned} F_X(x) &= \int_0^x \lambda^{-1} e^{-\frac{y}{\lambda}} dy \\ &= -e^{-\frac{y}{\lambda}} \Big|_0^x = 1 - e^{-\frac{x}{\lambda}} \end{aligned}$$

$\square$

Lema 54. *Se* $X \sim Exp(\lambda)$*, então* $\mathbb{E}[X] = \lambda$ *e* $Var[X] = \lambda^2$*.*

Demonstração.

$$\begin{aligned} \mathbb{E}[X] &= \int_{-\infty}^{\infty} x \cdot f_X(x) dx \\ &= \int_0^{\infty} x \lambda^{-1} e^{-\frac{x}{\lambda}} dx \\ &= -xe^{-\frac{x}{\lambda}} \Big|_0^{\infty} - \int_0^{\infty} -e^{-\frac{x}{\lambda}} dx && \text{Integração por partes} \\ &= 0 - \lambda e^{-\frac{x}{\lambda}} \Big|_0^{\infty} = \lambda \end{aligned}$$

Uma segunda forma de derivar a variância da distribuição exponencial é usando o fato que, se X é uma distribuição contínua não negativa $\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X > x)dx$ (esta é a versão contínua do Lema 13 – tente demonstrar isso!). Do Lema 53,

$$\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X > x)dx = \int_0^\infty e^{-\frac{x}{\lambda}}dx = -\lambda e^{-\frac{x}{\lambda}}|_0^\infty = \lambda$$

Muito mais fácil, certo?

Para derivar a variância, calculamos

$$\begin{aligned}
\mathbb{E}[X^2] &= \int_{-\infty}^{\infty} x^2 \cdot f_X(x)dx \\
&= \int_0^\infty x^2\lambda^{-1}e^{-\frac{x}{\lambda}}dx \\
&= -x^2e^{-\frac{x}{\lambda}}|_0^\infty - \int_0^\infty -2xe^{-\frac{x}{\lambda}}dx && \text{Integração por partes} \\
&= 0 - 2\lambda xe^{-\frac{x}{\lambda}}|_0^\infty - \int_0^\infty -2\lambda e^{-\frac{x}{\lambda}}dx && \text{Integração por partes} \\
&= -2\lambda^2e^{-\frac{x}{\lambda}}|_0^\infty = 2\lambda^2
\end{aligned}$$

$$Var[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = 2\lambda^2 - \lambda^2 = \lambda^2$$

□

Lema 55. *Se $X \sim Exp(\lambda)$, então X não tem memória. Isto é, $\mathbb{P}(X > t + s|X > t) = \mathbb{P}(X > s)$.*

Demonstração.

$$\begin{aligned}
\mathbb{P}(X > t + s|X > t) &= \frac{\mathbb{P}(X > t + s \cap X > t)}{\mathbb{P}(X > t)} && \text{Probabilidade Condicional (Definição 14)} \\
&= \frac{\mathbb{P}(X > t + s)}{\mathbb{P}(X > t)} \\
&= \frac{e^{-\frac{t+s}{\lambda}}}{e^{-\frac{t}{\lambda}}} && \text{Fda da Exponencial (Lema 53)} \\
&= e^{-\frac{s}{\lambda}} = \mathbb{P}(X > s) && \text{Fda da Exponencial (Lema 53)}
\end{aligned}$$

□

Note que isso é muito similar ao que acontece com a distribuição Geométrica (Exercício 98).

5.3.1 Exercícios

Exercício 135. *Seja $X \sim Exp(1)$. Encontre $\mathbb{E}[X^3]$.*

Exercício 136. *Uma lâmpada dura, em média 300 horas. Qual é a probabilidade de que ela vai durar mais que 700 dado que ela já durou 300 horas? Assuma que seu tempo de duração tem distribuição exponencial.*

Exercício 137. *A moda de uma distribuição contínua é o valor que maximiza f_X. Qual é a moda da distribuição exponencial?*

Exercício 138 (Programação)**.** *Use o Lema 53 e o Exercício 133 para simular uma Exponencial(λ).*

Exercício 139. *Sejam $X_1 \sim$ Exponencial(λ_1) e $X_2 \sim$ Exponencial(λ_2) variáveis aleatórias independentes. Seja $Z = \min(X_1, X_2)$. Encontre a fda de Z. Você reconhece esta fda? Generalize este resultado para $X_1, \ldots, X_n$ tal que $X_i \sim$ Exponencial(λ_i) e $Z = \min(X_1, \ldots, X_n)$.*

Dica*: para qualquer $a \in \Re$, $\min\{x, y\} > a$ se e somente se tanto x quanto y são maiores que a.*

$F_Z(z)$ é a fda de uma distribuição Exponencial com parâmetro $\left(\sum_{i=1}^n \lambda_i^{-1}\right)^{-1}$. Assim, $Z \sim \text{Exp}(\left(\sum_{i=1}^n \lambda_i^{-1}\right)^{-1})$.

Exercício 140. *Temos um sistema com n componentes em série. Lembramos, este sistema falha quando qualquer um dos componentes falha. Assuma que cada componente dura em média 1 ano antes de falhar. Quanto tempo você espera que o sistema dure até falhar? Use o Exercício 139.*

5.4 Distribuição Gama

Definição 33. $\Gamma : \mathbb{R}^+ \to \mathbb{R}^+$ *é chamada de função Gama e é tal que:*

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$$

Lema 56. *A Γ função Gama satisfaz as seguintes propriedades:*

1. *Para $a \geq 1$, $\Gamma(a) = (a-1)\Gamma(a-1)$*
2. *Se $n \in \mathbb{N}$, $\Gamma(n) = (n-1)!$*

Demonstração.

1.
$$\begin{aligned}
\Gamma(a) &= \int_0^\infty t^{a-1} e^{-t} dt \\
&= -t^{a-1} e^{-t}\Big|_0^\infty - \int_0^\infty -(a-1) t^{a-2} e^{-t} dt && \text{Integração por partes} \\
&= 0 + (a-1) \int_0^\infty t^{a-2} e^{-t} dt = (a-1)\Gamma(a-1) && \text{Função Gama (Definição 33)}
\end{aligned}$$

2. Usando o item anterior e iterando em n, observe que

$$\Gamma(n) = (n-1) \cdot (n-2) \ldots 2 \cdot 1 \cdot \Gamma(1) = (n-1)! \cdot \Gamma(1)$$

Assim, basta mostrar que $\Gamma(1) = 1$.

$$\Gamma(1) = \int_0^\infty e^{-t} dt = -e^{-t}\Big|_0^\infty = 1$$

□

Definição 34. *Dizemos que uma variável aleatória X tem distribuição Gama com parâmetros(k, λ) e denotamos por $X \sim$ Gama(k, λ) se a densidade de X é*

$$f_X(x) = \begin{cases} \frac{1}{\Gamma(k)\lambda^k} x^{k-1} e^{-\frac{x}{\lambda}} & , \textit{if } x > 0 \\ 0 & , \textit{caso contrário} \end{cases}$$

A figura 4 mostra algumas densidades possíveis para a distribuiçã Gama.

Lema 57. *TA densidade da distribuição Gama integra para* 1.

Demonstração.

$$\begin{aligned}
\int_0^\infty \frac{1}{\Gamma(k)\lambda^k} x^{k-1} e^{-\frac{x}{\lambda}} dx &= \int_0^\infty \frac{1}{\Gamma(k)\lambda^k} (\lambda t)^{k-1} e^{-t} \lambda dt && \text{Fazemos } x = \lambda t \\
&= \int_0^\infty \frac{1}{\Gamma(k)} t^{k-1} e^{-t} dy = 1 && \text{Função Gama (Definição 33)}
\end{aligned}$$

□

Observe que, no caso especial em que $k = 1$, a distribuição Gama corresponde à distribuição Exponencial.

Lema 58. *Para todo* $a > 0$, $\mathbb{E}[X^a] = \frac{\Gamma(k+a)\lambda^a}{\Gamma(k)}$. *If* $X \sim Gama(k, \lambda)$, *então* $\mathbb{E}[X] = k\lambda$ *e* $Var[X] = k\lambda^2$.

Demonstração.

$$\begin{aligned}
\mathbb{E}[X^a] &= \int_0^\infty x^a \frac{1}{\Gamma(k)\lambda^k} x^{k-1} e^{-\frac{x}{\lambda}} dx \\
&= \int_0^\infty \frac{1}{\Gamma(k)\lambda^k} x^{k+a-1} e^{-\frac{x}{\lambda}} dx \\
&= \frac{\Gamma(k+a)\lambda^a}{\Gamma(k)} \int_0^\infty \frac{1}{\Gamma(k+a)\lambda^{k+a}} x^{k+a-1} e^{-\frac{x}{\lambda}} dx && \text{Densidade de uma Gama}(k+1, \lambda) \\
&= \frac{\Gamma(k+a)\lambda^a}{\Gamma(k)} && \text{Propriedades da função Gama (Lema 56)}
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[X] &= \frac{\Gamma(k+1)\lambda^1}{\Gamma(k)} = \frac{k\Gamma(k)\lambda^1}{\Gamma(k)} = k\lambda && \text{Use } a = 1 \text{ no item anterior} \\
\mathbb{E}[X^2] &= \frac{\Gamma(k+2)\lambda^2}{\Gamma(k)} = \frac{k(k+1)\Gamma(k)\lambda^2}{\Gamma(k)} = k(k+1)\lambda^2 && \text{Use } a = 2 \text{ no item anterior} \\
Var[X] &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = k(k+1)\lambda^2 - k^2\lambda^2 = k\lambda^2
\end{aligned}$$

□

5.4.1 Exercícios

Exercício 141. *Seja* X *o tempo que um aluno leva para resolver este exercício. Assuma que* X *tem distribuição Gama. Se* $\mathbb{E}[X] = 6$ *minutos e* $Var[X] = 12$ *minutos*^{2}, *quais são os parâmetros da distribuição Gamma?*

Exercício 142. *Seja* $X \sim Gama(k, \lambda)$. *Encontre* $\mathbb{E}[X^3]$.

Exercício 143. *Seja* $X \sim Gama(k, \lambda)$. *Seja* $c > 0$ *e* $Y = cX$.

a. Quanto valem $\mathbb{E}[Y]$ *e* $Var[Y]$*?*

b. Se você soubesse que Y *tem distribuição Gama, quais seriam os parâmetros desta distribuição?*

c. Use o fato que $\frac{\partial P(Y \leq y)}{\partial y} = f_Y(y)$ *para mostrar que* $f_Y(y) = \frac{1}{c} f_X(\frac{y}{c})$.

d. Qual é a distribuição de Y*?*

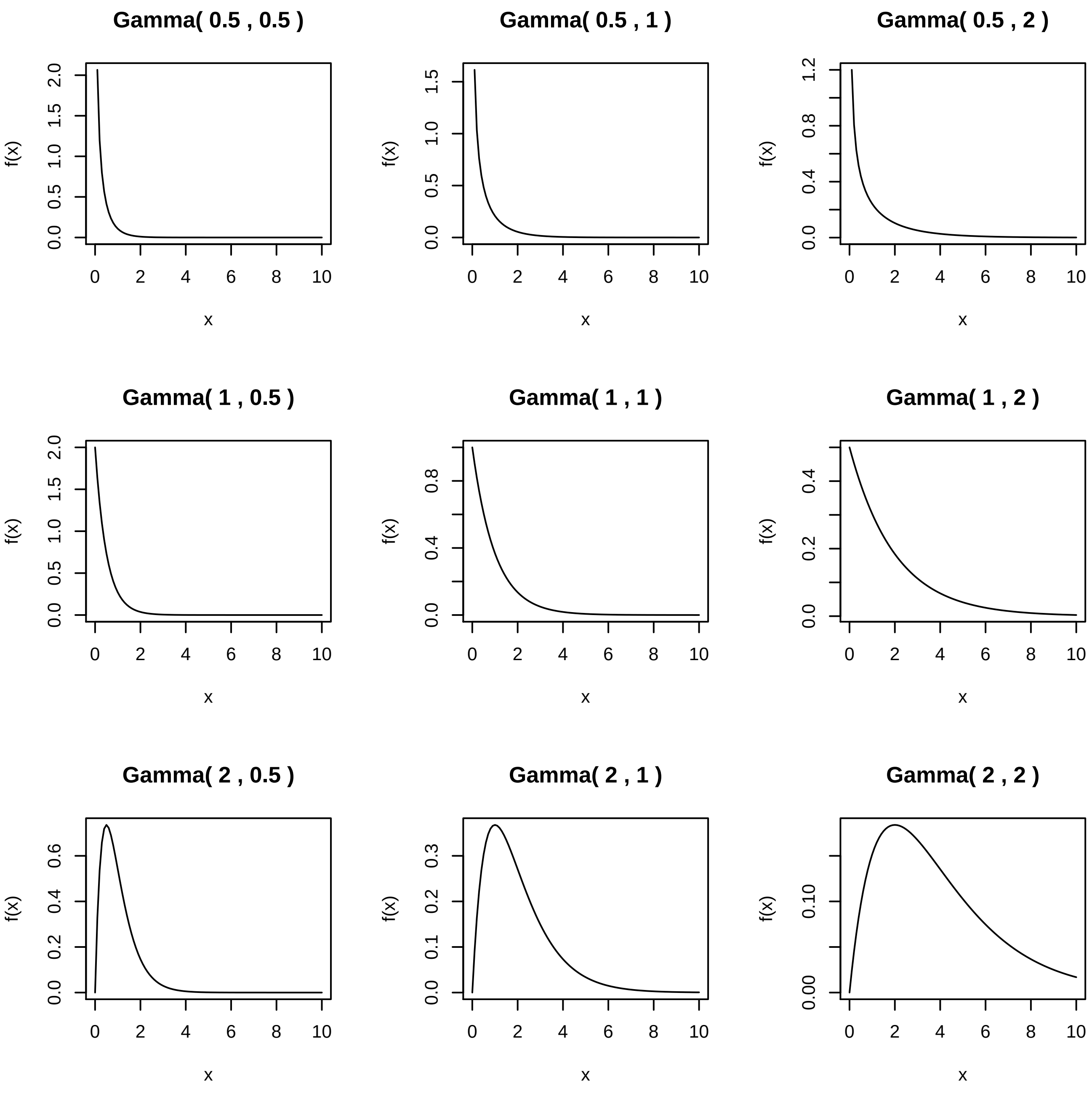


Figura 4: densidades de algumas distribuições Gama

5.5 Distribuição Beta

Definição 35. *Dizemos que X tem distribuição Beta com parâmetros (α, β) e denotamos por $X \sim Beta(\alpha, \beta)$ se a densidade de X é*

$$f_X(x) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot x^{\alpha-1}(1-x)^{\beta-1} & \text{, if } 0 < x < 1 \\ 0 & \text{, caso contrário} \end{cases}$$

Lema 59. *A densidade na Definição 35 é uma fdp válida. Em particular,*

$$\int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} dx = 1$$

Demonstração.

$$\begin{aligned}
\Gamma(\alpha)\Gamma(\beta) &= \int_0^\infty x^{\alpha-1}e^{-x}dx \int_0^\infty y^{\beta-1}e^{-y}dy && \text{(Definição 33)} \\
&= \int_0^\infty \int_0^\infty x^{\alpha-1}y^{\beta-1}e^{-(x+y)}dxdy \\
&= \int_0^\infty \int_0^1 (tu)^{\alpha-1}(t(1-u))^{\beta-1}e^{-t}t \cdot dudt && t = x+y,\ u = x(x+y)^{-1} \\
&= \int_0^\infty t^{\alpha+\beta-1}e^{-t}dt \int_0^1 u^{\alpha-1}(1-u)^{\beta-1}du \\
&= \Gamma(a+b) \int_0^1 u^{\alpha-1}(1-u)^{\beta-1}du && \text{(Definição 33)}
\end{aligned}$$

Assim, $\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}du = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(a+b)}$ and $\int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}dx = 1$. □

Como a distribuição Beta assume valores em $(0, 1)$, ela é frequentemente usada para modelar frequências relativas e razões. É uma distribuição flexível e pode assumir muitas formas diferentes. A Figura 5 mostra algumas densidades que a distribuição Beta pode assumir.

Lema 60. *Se $X \sim Beta(\alpha, \beta)$, então para todo $c, d \geq 0$ $\mathbb{E}[X^c(1-X)^d] = \frac{\Gamma(\alpha+c)\Gamma(\beta+d)}{\Gamma(\alpha+\beta+c+d)}$. Portanto, $\mathbb{E}[X] = \frac{\alpha}{\alpha+\beta}$ e $Var[X] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.*

Demonstração.

$$\begin{aligned}
\mathbb{E}[X^c(1-X)^d] &= \int_{-\infty}^\infty x^c(1-x)^d f_X(x)dx \\
&= \int_0^1 x^c(1-x)^d \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}dx \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\Gamma(\alpha+c)\Gamma(\beta+d)}{\Gamma(\alpha+\beta+c+d)} \int_0^1 \frac{\Gamma(\alpha+\beta+c+d)}{\Gamma(\alpha+c)\Gamma(\beta+d)} x^\alpha(1-x)^{\beta-1}dx \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\Gamma(\alpha+c)\Gamma(\beta+d)}{\Gamma(\alpha+\beta+c+d)} \cdot 1 && \text{(Lema 59)}
\end{aligned}$$

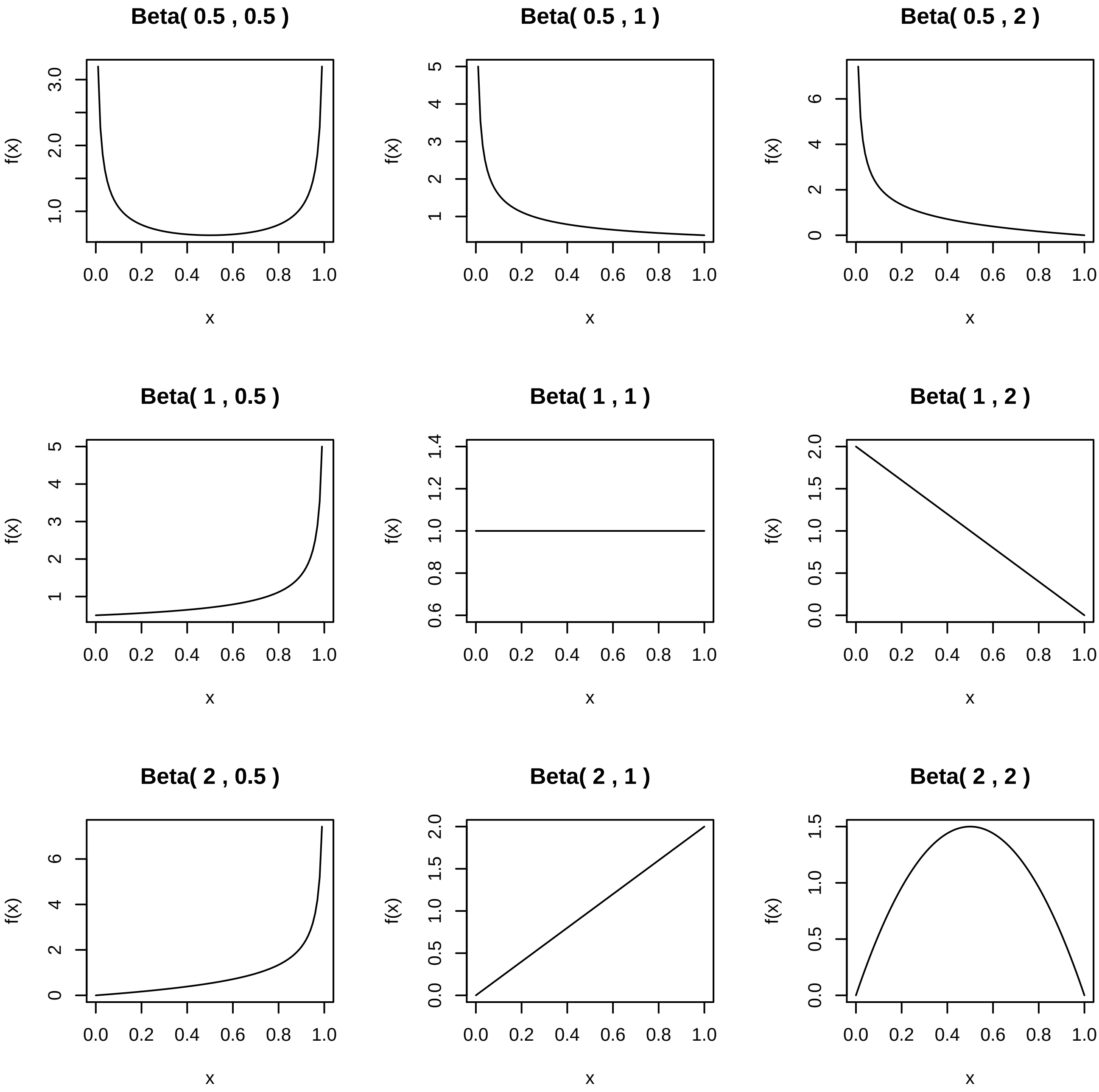


Figura 5: Densidades de algumas distribuições Beta

$$
\begin{aligned}
\mathbb{E}[X] &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\Gamma(\alpha+1)\Gamma(\beta+0)}{\Gamma(\alpha+\beta+1+0)} \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)} \cdot \frac{\alpha\Gamma(\alpha)}{(\alpha+\beta)\Gamma(\alpha+\beta)} = \frac{\alpha}{\alpha+\beta} && \text{(Lema 56)} \\
\mathbb{E}[X^2] &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\Gamma(\alpha+2)\Gamma(\beta+0)}{\Gamma(\alpha+\beta+2+0)} \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)} \cdot \frac{\alpha(\alpha+1)\Gamma(\alpha)}{(\alpha+\beta+1)(\alpha+\beta)\Gamma(\alpha+\beta)} = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} && \text{(Lema 56)} \\
Var[X] &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\
&= \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} - \frac{\alpha^2}{(\alpha+\beta)^2} \\
&= \frac{(\alpha+\beta)(\alpha+1)\alpha - \alpha^2(\alpha+\beta+1)}{(\alpha+\beta+1)(\alpha+\beta)^2} = \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}
\end{aligned}
$$

□

5.5.1 Exercícios

Exercício 144. *Seja $X \sim Beta(\alpha, \beta)$. Ache $\mathbb{E}[X(1-X)]$.*

Exercício 145. *Seja $X \sim Beta(1, \beta)$. Ache $F_X(x)$.*

Exercício 146. *Seja $X \sim Beta(\alpha, 1)$. Ache $F_X(x)$.*

Exercício 147. *Seja $X \sim Beta(\alpha, \beta)$*

a. *Seja $Y = 1 - X$. Quais são os valores de $\mathbb{E}[Y]$ e $Var[Y]$?*

b. *Se você soubesse que Y tem distribuição Beta, quais seriam os parâmetros desta distribuição?*

c. *Use o fato de que $\frac{\partial d\mathbb{P}(Y \leq y)}{\partial y} = f_Y(y)$ para mostrar que $f_Y(y) = f_X(1-y)$.*

d. *Qual é a distribuição de Y?*

5.6 Distribuição Normal

Definição 36. *Dizemos que X tem distribuição Normal com parâmetros (μ, σ^2) e denotamos por $X \sim N(\mu, \sigma^2)$ se a densidade de X é*

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Se $\mu = 0$, e $\sigma = 1$, dizemos que X tem distribuição normal padrão.

Definição 37. *Seja $Z \sim N(0, 1)$.*

$$\phi(z) = \mathbb{P}(Z \leq z) = F_Z(z)$$

ϕ não tem solução analítica. As duas formas de obter valores aproximados de ϕ são:

1. *Muitas linguagens de programação têm livrarias estatísticas com estes valores. Por exemplo, R, python, C, ...*

2. Consultar uma tabela da normal padrão, disponível em muitos livros de Estatística.

Lema 61. *Se $X \sim N(\mu, \sigma^2)$, then $\frac{(X-\mu)}{\sigma} \sim N(0,1)$.*

Demonstração. Seja $Z = \frac{(X-\mu)}{\sigma}$.

$$\begin{aligned} F_Z(z) = \mathbb{P}(Z \leq z) = \mathbb{P}\left(\frac{(X-\mu)}{\sigma} \leq z\right) \\ = \mathbb{P}(X \leq \sigma z + \mu) = F_X(\sigma z + \mu) \end{aligned}$$

Assim,

$$\begin{aligned} f_Z(z) = \frac{\partial F_Z(z)}{\partial z} &= \frac{\partial F_X(\sigma z + \mu)}{z} && \text{(Lema 49)} \\ &= \frac{\partial F_X(\sigma z + \mu)}{\partial(\sigma z + \mu)} \cdot \frac{\partial(\sigma z + \mu)}{\partial z} && \text{Regra da Cadeia} \\ &= f_X(\sigma z + \mu) \cdot \sigma \\ &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\sigma z + \mu - \mu)^2}{\sigma^2}} \cdot \sigma = \frac{1}{\sqrt{2\pi}} e^{-z^2} \end{aligned}$$

□

Segue do Lema 61 que qualquer parametrização da distribuição normal pode ser obtida com mudança apropriada de escala da normal padrão. A Figura 6 apresenta a densidade da distribuição normal padrão.

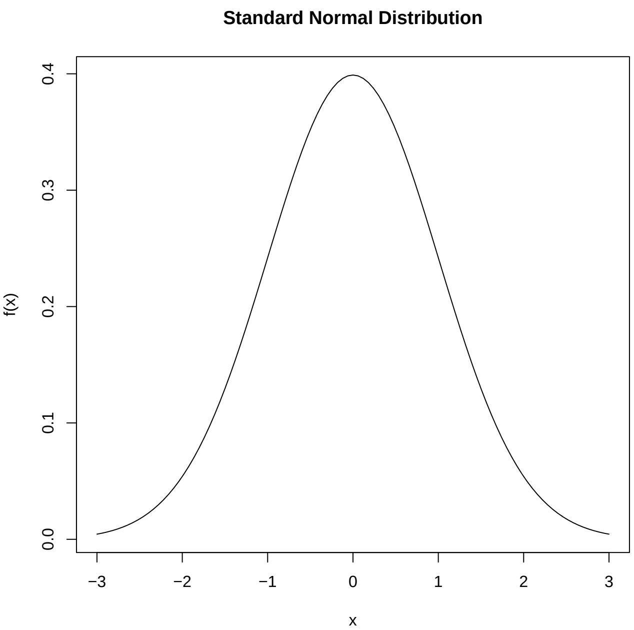


Figura 6

Lema 62. *Se $X \sim N(\mu, \sigma^2)$, então $\mathbb{E}[X] = \mu$ e $Var[X] = \sigma^2$.*

Demonstração. Considere que $X \sim \mathrm{N}(0,1)$.

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} x e^{-x^2} dx$$

Como xe^{-x^2} é uma função impar, $\mathbb{E}[X] = 0$.

Lembramos que $Var[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$. Como $\mathbb{E}[X] = 0$, $Var[X] = \mathbb{E}[X^2]$.

$$\begin{aligned}
\mathbb{E}[X^2] &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} x^2 e^{-x^2} dx \\
&= -\frac{1}{2\sqrt{2\pi}} x e^{-x^2} \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2} dx && \text{integração por partes} \\
&= 0 + 1 && \text{fdp da normal padrão (Definições, 27 36)}
\end{aligned}$$

A seguir, considere que $X \sim \mathrm{N}(\mu, \sigma^2)$,

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}\left[\sigma \cdot \frac{X-\mu}{\sigma} + \mu\right] \\
&= \sigma \cdot \mathbb{E}\left[\frac{X-\mu}{\sigma}\right] + \mu \\
&= \sigma \cdot 0 + \mu = \mu && \text{(Lema 61)}
\end{aligned}$$

$$\begin{aligned}
Var[X] &= Var\left[\sigma \cdot \frac{X-\mu}{\sigma} + \mu\right] \\
&= \sigma^2 \cdot Var\left[\frac{X-\mu}{\sigma}\right] \\
&= \sigma^2 \cdot 1 = \sigma^2 && \text{(Lema 61)}
\end{aligned}$$

□

Lema 63 (Aproximação da Binomial e Teorema do Limite Central)**.** *Considere que $X \sim Binomial(n, p)$ e que n é "grande" ($n = 30$ frequentemente é suficiente para um bom resultado).*

$$\mathbb{P}\left(\frac{X - np}{\sqrt{np(1-p)}} \leq z\right) \approx \phi(z)$$

Esta é uma instância específica de uma regra mais genérica. Se $X_1, X_2, \ldots$ é uma sequência de variáveis aleatórias independentes com a mesma distribuição, $\mathbb{E}[X_i] = \mu$ e $Var[X_i] = \sigma^2$, então, para n grande,

$$\mathbb{P}\left(\frac{\sum_{i=1}^{n}(X_i - \mu)}{\sqrt{n}\sigma} \leq z\right) \approx \phi(z)$$

Demonstração. Vamos deixar esta afirmação mais precisa e também prová-la no Capítulo 8. □

5.6.1 Exercícios

Exercício 148. *Seja $X \sim N(\mu, \sigma^2)$. Determine $\mathbb{E}[X^3]$.*

Solution: Seja $Z = \frac{X-\mu}{\sigma}$.

$$\begin{aligned} E[X^3] &= E[(Z \cdot \sigma + \mu)^3] \\ &= E[\sigma^3 Z^3 + 3\mu\sigma^2 Z^2 + 3\mu^2\sigma Z + \mu^3] \\ &= \sigma^3 E[Z^3] + 3\mu\sigma^2 E[Z^2] + 3\mu^2\sigma E[Z] + \mu^3 \end{aligned}$$

Do Lema 61, $Z \sim \text{N}(0,1)$. Portanto, usando o Lema 62, $E[Z] = 0$ e $E[Z^2] = 1$. Além disso,

$$\begin{aligned} E[Z^3] &= \int_{-\infty}^{\infty} z^3 \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \\ &= 0 && \text{Pois } z^3 \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \text{ é uma função impar} \end{aligned}$$

Portanto, $E[X^3] = \mu^3 + 3\mu\sigma^2$

Exercício 149. *Seja $Y \sim N(\mu, \sigma^2)$. Sejam $a, b \in \mathbb{R}$. Mostre que*

$$\mathbb{P}(a < Y < b) = \phi\left((b-\mu)/\sigma\right) - \phi\left((a-\mu)/\sigma\right),$$

onde ϕ é a distribuição acumulada da distribuição normal padrão (i.é., média zero e variância 1).

Isso mostra que podemos calcular qualquer probabilidade relacionada a uma variável aleatória normal calculando a distribuição acumulada da distribuição normal padrão. Isto era muito importante quando os computadores não estavam disponíveis, pois permite calcular todas as probabilidades de interesse tendo acesso a uma única tabela. Não são muitas as distribuições que têm esta propriedade.

Exercício 150. *Seja X o número de caras em 10000 lançamentos de uma moeda honesta. Dê uma aproximação para $\mathbb{P}(X \leq 5050)$.*

5.7 Revisão das Distribuições Contínuas Especiais

1. **A Variável Aleatória Uniforme** - $X \sim \mathrm{U}(a,b)$.

   Todos os subconjuntos de (a,b) com o comesmo comprimento são equiprováveis. A distribuição é normalmente usada quando todos os pontos em (a,b) são "igualmente prováveis".

2. **A Variável Aleatória Exponencial** - $X \sim \mathrm{Exp}(\lambda)$.

   Este distribuição é usada, frequentemente, para modelar o tempo até um certo evento ocorrer. É a única distribuição contínua com propriedade de perda de memória.

3. **A Variável Aleatória Gama** - $X \sim \mathrm{Gama}(k,\lambda)$.

   A Gama(1, λ) é uma Exponencial(λ) e, assim, a distribuição Gama é a generaliação da Exponencial. Se k é um número natural, $X = \sum_{i=1}^{k} Y_i$, onde Y_i são variáveis aleatórias independentes e $Y_i \sim$ Exponencial(λ).

4. **A Variável Aleatória Beta** - $X \sim \mathrm{Beta}(\alpha,\beta)$.

   Como a distribuição Beta assume valores em $(0,1)$, ela é frequentemente usada para modelar frequências e razões.

5. **A Variável Aleatória Normal (Gaussiana)** - $X \sim \mathrm{N}(\mu,\sigma^2)$.

   Usando o Teorema do Limite Central (Lema 63), a distribuição Normal é frequentemente usada para aproximar a distribuição da média de uma sequência de variáveis aleatórias independentes e identicamente distribuídas.

Variável Aleatória (Y)	fdp: $f_Y(y)$	$\mathbb{E}[Y]$	$Var[Y]$
Uniforme(a,b)	$\frac{1}{b-a}$ $y \in (a,b)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponencial(λ)	$\frac{1}{\lambda}e^{-\frac{y}{\lambda}}$ $y \in \mathbb{R}^+$	λ	λ^2
Gama($k\lambda$)	$\frac{1}{\Gamma(k)\lambda^k}y^{k-1}e^{-\frac{y}{\lambda}}$ $y \in \mathbb{R}^+$	$k\lambda$	$k\lambda^2$
Beta(α,β)	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$ $y \in (0,1)$	$\frac{\alpha}{\alpha+\beta}$	$r \cdot \frac{1-p}{p^2}$
Normal(μ,σ^2)	$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(z-\mu)^2}{2\sigma^2}}$ $y \in \mathbb{R}$	μ	σ^2

Tabela 3: Revisão das Distribuições Contínuas

6 Funções Geradoras de Momentos e Desigualdades de Probabilidade

6.1 Funções Geradoras de Momentos

Nesta seção, introduzimos funções geradoras de momentos, que são uma ferramenta muito útil para estudar as propriedades de uma variável aleatória.

Definição 38. *Seja X uma variável aleatória. A função geradora de momentos de X é $M_X : \mathbb{R} \to \mathbb{R}$ tal que:*

$$M_X(t) = \mathbb{E}[e^{tX}]$$

Exemplo 57. *No Exercício 117 determinamos a função geradora de momentos de $X \sim$ Bernoulli(p).*

$$\begin{aligned} M_X(t) = \mathbb{E}[e^{tX}] &= e^{t\cdot 0}\mathbb{P}(X = 0) + e^{t\cdot 1}\mathbb{P}(X = 1) \\ &= (1-p) + e^t p \end{aligned}$$

Além disso, se $Y \sim$ Binomial(n, p), sabemos que $Y = \sum_{i=1}^n X_i$, onde X_i são variáveis aleatórias independentes com distribuição Bernoulli(p). Portanto,

$$\begin{aligned} M_Y(t) = \mathbb{E}[e^{tY}] &= \mathbb{E}[e^{t\sum_{i=1}^n X_i}] \\ &= \mathbb{E}[\prod_{i=1}^n e^{tX_i}] \\ &= \prod_{i=1}^n \mathbb{E}[e^{tX_i}] && \textit{Independência (Lema 18)} \\ &= \prod_{i=1}^n ((1-p) + e^t p) && \textit{Função Geradora de Momentos da Bernoulli} \\ &= ((1-p) + e^t p)^n \end{aligned}$$

Exemplo 58. *Se $X \sim$ Geom(p), $M_X(t) = \frac{pe^t}{1-e^t(1-p)}$ para $0 < e^t(1-p) < 1$.*

$$\begin{aligned} M_X(t) = \mathbb{E}[e^{tX}] &= \sum_{i=1}^{\infty} e^{ti}\mathbb{P}(X = i) \\ &= \sum_{i=1}^{\infty} e^{ti} p(1-p)^{i-1} && \textit{fmp da Geométrica (Lema 35)} \\ &= pe^t \sum_{i=1}^{\infty} e^{t(i-1)}(1-p)^{i-1} \\ &= pe^t \sum_{j=0}^{\infty} (e^t(1-p))^j && (j = i-1) \\ &= \frac{pe^t}{1-e^t(1-p)} && \textit{Série Geométrica (Lema 45)} \end{aligned}$$

Lema 64. *Se $M_X(t)$ é a função geradora de momentos de X, então $\frac{\partial^n M_X}{\partial t^n}(0) = \mathbb{E}[X^n]$ para todo $n = 1, 2, \ldots$.*

Demonstração. Usando expansão de Taylor de e^x (Lema 44), observe que, quando $\mathbb{E}[e^{tX}]$ está definida:

$$\begin{aligned}
M_X(t) &= \mathbb{E}\left[\sum_{i=0}^{\infty}\frac{(tX)^i}{i!}\right]\\
&= \sum_{i=0}^{\infty}\frac{t^i\mathbb{E}[X^i]}{i!}
\end{aligned}$$

Assim,

$$\frac{\partial^n M_X}{\partial t^n} = \sum_{i=n}^{\infty}\frac{t^{i-n}\mathbb{E}[X^i]}{(i-n)!}$$

E, portanto,

$$\frac{\partial^n M_X}{\partial t^n}(0) = \sum_{i=n}^{\infty}\frac{0^{i-n}\cdot\mathbb{E}[X^i]}{(i-n)!}$$

Assim, todos os termos que correspondem a $i > n$, se tornam 0 e $\frac{\partial^n M_X}{\partial t^n}(0) = \mathbb{E}[X^n]$.
Uma prova alternativa consiste em calcular $\frac{\partial^n}{\partial t^n}M_X(t) = \mathbb{E}\left[\frac{\partial^n}{\partial t^n}e^{tX}\right]$. Deixamos isso como exercício. □

Exemplo 59 (Continuação do Exemplo 57)**.** *Seja $X \sim$ Bernoulli(p). Nos sabemos que $M_X(t) = (1-p)+pe^t$. portanto,*

$$\begin{aligned}
\frac{\partial M_X(t)}{\partial t} &= pe^t\\
\frac{\partial^2 M_X(t)}{\partial t^2} &= pe^t
\end{aligned}$$

Concluímos que $\mathbb{E}[X] = \frac{\partial M_X(t)}{\partial t}(0) = p$ e $\mathbb{E}[X^2] = \frac{\partial^2 M_X(t)}{\partial t^2}(0) = p$. Assim, usando o Lema 17,

$$Var[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = p - p^2 = p(1-p)$$

Estes resultados estão de acordo com os nossos cálculos anteriores para a Bernoulli (Lema 31).

Lema 65. *Se $X \sim N(\mu,\sigma^2)$, então $M_X(t) = \mathbb{E}[e^{tX}] = e^{t\mu+\frac{(t\sigma)^2}{2}}$.*

Demonstração. Seja $Z \sim N(0,1)$.

$$\begin{aligned}
M_Z(t) = \mathbb{E}[e^{tZ}] &= \int_{-\infty}^{\infty} e^{tz}f_Z(z)dz\\
&= \int_{-\infty}^{\infty} e^{tz}\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}dz\\
&= e^{\frac{t^2}{2}}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2+2tz+t^2}{2}}dz\\
&= e^{\frac{t^2}{2}}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{(z+t)^2}{2}}dz\\
&= e^{\frac{t^2}{2}}\cdot 1 = e^{\frac{t^2}{2}} && \text{(Definição 36)}
\end{aligned}$$

Se $X \sim N(\mu, \sigma^2)$, então

$$\begin{aligned}
\mathbb{E}[e^{tX}] &= \mathbb{E}\left[e^{t(\sigma\frac{X-\mu}{\sigma}+\mu)}\right] \\
&= e^{t\mu}\mathbb{E}\left[e^{t\sigma\frac{X-\mu}{\sigma}}\right] \\
&= e^{t\mu}M_{\frac{X-\mu}{\sigma}}(t\sigma) \\
&= e^{t\mu}M_Z(t\sigma) && \text{(Lema 61)} \\
&= e^{t\mu}e^{\frac{(t\sigma)^2}{2}} = e^{t\mu+\frac{(t\sigma)^2}{2}}
\end{aligned}$$

□

Uma propriedade importante das funções geradoras de momentos é que se $M_X(t) = M_Y(t)$ para todo t, então a distribuição de X deve ser a mesma que a distribuição de Y. Isto é, as funções geradoras de momentos determinam a distribuição de forma única. Apesar de que não vamos demonstrar este fato, vamos usá-lo em alguns exercícios.

Exemplo 60. *Se $X \sim N(\mu_1, \sigma_1^2)$ e $Y \sim N(\mu_2, \sigma_2^2)$ são duas variáveis aleatórias independentes, Lema 65 implica que*

$$M_{X+Y}(t) = \mathbb{E}[e^{t(X+Y)}] = \mathbb{E}[e^{tX}]\mathbb{E}[e^{tY}] = M_X(t) \cdot M_Y(t) = e^{t\mu_1+\frac{(t\sigma_1)^2}{2}} \cdot e^{t\mu_2+\frac{(t\sigma_2)^2}{2}} = e^{t(\mu_1+\mu_2)+\frac{(t(\sigma_1+\sigma_2))^2}{2}},$$

que é a função geradora de momentos de uma variável aleatória com média $\mu_1 + \mu_2$ e variância $\sigma_1^2 + \sigma_2^2$. Como a função geradora de momentos caracteriza a distribuição de uma variável aleatória de forma única, segue que $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

6.1.1 Exercícios

Exercício 151. *Seja $X \sim$ Binomial Negativa(r, p). Encontre $M_X(t)$. Dica: Pense sobre a derivação da Binomial Negativa no Exemplo 57.*

Exercício 152. *Seja $X \sim$ Poisson(λ). Encontre $M_X(t)$.*

Exercício 153. *Seja $X \sim$ Exponencial(λ) e $t < \lambda^{-1}$. Encontre $M_X(t) = \mathbb{E}[e^{tX}]$.*

Exercício 154. *Seja $X \sim$ Gama(k, λ) e $M_X(t) = \mathbb{E}[e^{tX}]$. Encontre $M_X(t)$.*

Exercício 155. *Sejam $X_1, \ldots, X_n$ variáveis aleatórias independentes. Mostre que $M_Y(t) = (M_{X_1}(t))^n$, onde $Y = \sum_{i=1}^n X_i$.*

Exercício 156. *Se $X_i \sim$ Poisson(λ_i) for $i = 1, \ldots, n$ são independentes, qual é a distribuição de $\sum_{i=1}^n X_i$?*

Exercício 157. *Assuma que X tem a sua função geradora de momentos dada por$M_X(t) = e^{(3t+8t^2)}$. Determine a função geradora de momentos e o valor esperado de $Y = 2 * X + 3$.*

6.2 Desigualdades de Probabilidade

Seja X uma variável aleatória. Um evento extremo é da forma $\{X \geq i\}$, isto é, observar valores tão ou mais extremos que i. Estes eventos podem ser de interesse em muitas situações:

Exemplo 61 (Eventos caudais)**.**

a. *Um servidor aguenta no máximo n conexões por minuto. O número médio de usuários que acessam o servidor por minuto é λ. O evento em que o servidor é sobrecarregado num dado momento é um evento extremo.*

b. *Tomamos n indivíduos de uma população de N em que k têm uma doença. Seja X o número de indivíduos doentes na amostra. O evento em que a amostra não tem precisão ϵ, isto é $\{|\frac{X}{n} - \frac{k}{N}| > \epsilon\}$ é um evento extremo.*

c. *Um sistema tem n componentes. O sistema falha se pelo menos k componentes falhares. Seja X o número de componentes que falham. O evento em que o sistema falha $\{X \geq k\}$ é um evento extremo.*

Para algumas distribuições, pode ser dificil calcular exatamente o valor de $\mathbb{P}(X \geq i)$. Nestes casos, é comum limitar superiormente a probabilidade do evento extremo. A seguir, veremos uma introdução aos métodos que são normalmente usados para obter estes limites superiores.

Lema 66 (Desigualdade de Markov)**.** *Seja X uma variável aleatória. Se $X \geq 0$, então*

$$\mathbb{P}(X \geq M) \leq \frac{\mathbb{E}[X]}{M}$$

para todo $M > 0$

Demonstração.

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{x \in Im(X)} x p_X(x) \\
&\geq \sum_{x \in Im(X) \cap x \geq M} x p_X(x) && \text{Since } X \geq 0 \\
&\geq \sum_{x \in Im(X) \cap x \geq M} M p_X(x) && \text{Pois todos os x's incluídos são tais que } x \geq M \\
&\geq t \sum_{x \in Im(X) \cap x \geq M} p_X(x) = t\mathbb{P}(X \geq M)
\end{aligned}$$

Portanto, $\mathbb{P}(X \geq M) \leq \frac{\mathbb{E}[X]}{M}$. □

Exemplo 62. *Se X é o tempo que leva até um ônibus chagar num certo ponto, e nos sabemos que $\mathbb{E}[X] \leq 10$, então a desigualdade de Markov implica que*

$$\mathbb{P}(X > 20) \leq 10/20 = 1/2.$$

Lema 67 (desigualdade de Chebyshev)**.** *Seja X uma variável aleatória,*

$$\mathbb{P}(|X - \mathbb{E}[X]| > M) \leq \frac{Var[X]}{M^2}$$

Demonstração.

$$\begin{aligned}
\mathbb{P}(|X-\mathbb{E}[X]|>M) &= \mathbb{P}((X-\mathbb{E}[X])^2>M^2) \\
&\leq \frac{\mathbb{E}[(X-\mathbb{E}[X])^2]}{M^2} && \text{Desigualdade de Markov (Lema 66)} \\
&= \frac{Var[X]}{M^2}
\end{aligned}$$

□

Exemplo 63. *Sejam $X_1, \ldots, X_n$ variáveis aleatórias independentes e identicamente distribuídas com média $\mu < \infty$ e variância $\sigma^2 < \infty$. A desigualdade de Chebyshev implica que*

$$\mathbb{P}(|\bar{X}-\mu|>\epsilon) \leq \frac{\sigma^2}{n\epsilon^2}.$$

É equivalente a dizer que

$$\mathbb{P}(|\bar{X}-\mu|\leq\epsilon) \geq 1-\frac{\sigma^2}{n\epsilon^2}.$$

Isto é, se n é grande, $\bar{X}$ é próximo de μ com probabilidade elevada.

Lema 68 (Desigualdade de Chernoff)**.** *Seja X uma variável aleatória e $t>0$,*

$$\mathbb{P}(X\geq\epsilon) \leq \frac{M_X(t)}{e^{t\epsilon}}$$

Demonstração.

$$\begin{aligned}
\mathbb{P}(X\geq\epsilon) &= \mathbb{P}(e^{tX}\geq e^{t\epsilon}) && e^{tx}\text{é uma função crescente} \\
&\leq \frac{\mathbb{E}[e^{tX}]}{e^{t\epsilon}} && \text{Desigualdade de Markov (Lema 66)} \\
&= \frac{M_X(t)}{e^{t\epsilon}}
\end{aligned}$$

□

A desigualdade de Chernoff é, em geral, mais justa que a desigualdade de Markov, mas ela usa a função geradora de momentos de X.

Exemplo 64. *Se $X\sim N(0,1)$, desigualdade de Chernoff diz que, para todo $\epsilon, t>0$,*

$$\mathbb{P}(Z\geq\epsilon) \leq e^{-t\epsilon}e^{-t^2/2}.$$

Como $e^{-t\epsilon}e^{-t^2/2}$ é minimizado quando $t=\epsilon$, a desigualdade de Chernoff implica que

$$\mathbb{P}(Z\geq\epsilon) \leq e^{-\epsilon^2/2}.$$

Assim, $\mathbb{P}(Z\geq 2)\leq e^{-2^2/2}\approx 0.14$.

Lema 69 (Desigualdade de Jensen)**.** *Se g é uma função convexa, então $\mathbb{E}[g(X)]\geq g(\mathbb{E}[X])$. De forma similar, se g é côncava, então $\mathbb{E}[g(X)]\leq g(\mathbb{E}[X])$.*

Demonstração. Seja $L(x) = a + bx$ a linha tangente a $g(x)$ no ponto $\mathbb{E}[X]$. Se g é convexa, por definição g está abaixo da linha $L(x)$, i.é., $g(y) \geq L(y)$ para todo y. Segue que $g(X) \geq L(X)$, e portanto, do Exercício **??**, $\mathbb{E}[g(X)] \geq \mathbb{E}[L(X)]$. O lema segue do fato de que $\mathbb{E}[L(X)] = a + b\mathbb{E}[X] = L(\mathbb{E}[X]) = g(\mathbb{E}[X])$. □

Exercício 158. *Seja X uma variável aleatória não negativa com média 25. A desigualdade de Jensen implica que $\mathbb{E}[\log X] \leq \log \mathbb{E}[X] \approx 3.2$, e $\mathbb{E}[e^{-X}] \geq e^{-\mathbb{E}[X]} \approx 1.4 \times 10^{-11}$.*

6.2.1 Exercícios

Exercício 159. *Um sistema tem 1000 componentes. Cada componente falha independentemente com probabilidade 0.1. Dê um limite superior para a probabilidade de que pelo menos 120 componentes vão falhar.*

Exercício 160. *Usuários acessam um servidor a uma taxa média de 50 por minuto. Dê um limite superior para a probabilidade de que, em um certo minuto, o servidor será acessado por 70 usuários ou mais.*

Exercício 161. *Cada aluno acerta uma questão, independentemente, com probabilidade 0.8. Seja X o número de alunos de 20 que acertaram a questão. Dê um limite superior para $\mathbb{P}(X \leq 11)$.*

Exercício 162. *Usando a desigualdade de Chernoff, limite $\mathbb{P}(X \geq \epsilon)$ onde $X \sim Poission(\lambda)$.*

Exercício 163. *Use a desigualdade de Jensen para provar que a variância de uma variável aleatória é sempre não negativa.*

Exercício 164. *Seja X uma variável aleatória com média μ e variância σ^2. Mostre que $\mathbb{P}(|X - \mu| \geq k\sigma) \leq 1/k^2$*

Exercício 165. *Sejam $X_1 \sim N(0,1)$ e $X_2 \sim Exp(1)$ variáveis aleatórias independentes. Usando a desigaldade de Chebyshev, limite superiormente $\mathbb{P}(|\sum_i X_i - 1| > 3)$ e $\mathbb{P}(|\bar{X} - 1/2| > 2)$*

7 Distribuições Conjuntas

Nos últimos Capítulos, vimos distribuições para um único número desconhecido. A seguir, vamos ver distribuições de dois ou mais números desconhecidos.

7.1 Distribuição de Vetores Aleatórios

Definição 39. *Sejam X e Y duas variáveis aleatórias discretas. A fmp conjunta de X e Y é $p_{X,Y}(x,y) := \mathbb{P}(X = x, Y = y)$. De forma similar, se X e Y são duas variáveis aleatórias com distribuições contínuas, a fdp conjunta de X e Y é $f_{X,Y}(x,y)$. Seja A um subconjunto de $\mathbb{R}^2$,*

$$\mathbb{P}((X,Y) \in A) = \mathbb{P}(\{\omega \in \Omega : (X(\omega), Y(\omega)) \in A\}) = \begin{cases} \sum_{(x,y)\in A} p_{X,Y}(x,y) & \text{, se } X \text{ e } Y \text{ são discretas.} \\ \int_{(x,y)\in A} f_{X,Y}(x,y)d(x,y) & \text{, se } X \text{ e } Y \text{ são contínuas.} \end{cases}$$

Exemplo 65. *Sejam X e Y variáveis aleatórias contínuas, com a seguinte fdp*

$$f_{X,Y}(x,y) = \begin{cases} cxy & \text{, if } x > 0,\ y > 0,\ x + y < 1 \\ 0 & \text{, caso contrário} \end{cases}$$

Qual é o valor de c? Observe que $\mathbb{P}((X,Y) \in \mathbb{R}^2) = 1$. Além disso, observe que

$$\begin{aligned}
\mathbb{P}((X,Y) \in \mathbb{R}^2) &= \int_{(x,y)\in\mathbb{R}^2} f_{X,Y}(x,y)d(x,y) \\
&= \int_0^1 \int_0^{1-x} cxydydx \\
&= c\int_0^1 x \int_0^{1-x} ydy \\
&= c\int_0^1 x\frac{(1-x)^2}{2}dx \\
&= \frac{c}{2}\int_0^1 (x - 2x^2 + x^3)dx \\
&= \frac{c}{2}\left(\frac{1}{2} - \frac{2}{3} + \frac{1}{4}\right) \\
&= \frac{c}{2} \cdot \frac{1}{12} = \frac{c}{24}
\end{aligned}$$

Conclua que $c = 24$.

Exemplo 66. *Considere que X e Y are são variáveis aleatórias discretas com a fmp conjunta dada na Tabela 4.*

X/Y	0	1
0	0.2	0.4
1	0.3	0.1

Tabela 4: Tabela de $p_{X,Y}(x,y)$

$$\sum_{x\in\{0,1\}}\sum_{y\in\{0,1\}} p_{X,Y}(x,y) = 0.2+0.4+0.3+0.1 = 1$$

Também podemos calcular $\mathbb{P}(X=1)$.

$$\begin{aligned}\mathbb{P}(X=1) &= \sum_{\{(x,y)\in\{0,1\}^2:x=1\}} p_{X,Y}(x,y) \\ &= \sum_{y\in\{0,1\}} p_{X,Y}(1,y) = 0.3+0.1 = 0.4\end{aligned}$$

Observe que $X \sim$ *Bernoulli*(0.4). *De forma similar,* $Y \sim$ *Bernoulli*(0.5).

Exemplo 67. *Considere Exemplo 65. Também podemos calcular* $\mathbb{P}(Y < X)$.

$$\begin{aligned}\mathbb{P}(X<Y) &= \int_{\{(x,y)\in\mathbb{R}^2:x<y\}} f_{X,Y}(x,y)d(x,y) \\ &= \int_0^1\int_0^x \mathbb{I}_{\{x+y<1\}} 24xydydx \\ &= \int_0^1\int_0^{\min(1-x,x)} 24xydydx \\ &= \int_0^{0.5}\int_0^x 24xydydx + \int_{0.5}^1\int_0^{1-x} 24xydydx \\ &= \int_0^{0.5} 24x\frac{x^2}{2}dx + \int_{0.5}^1 24x\frac{(1-x)^2}{2}dx \\ &= 3x^4|_0^{0.5} + \int_{0.5}^1 12(x-2x^2+x^3)dx \\ &= \frac{3}{16} + 12(\frac{x^2}{2} - \frac{2x^3}{3} + \frac{x^4}{4})|_{0.5}^1 \\ &= \frac{3}{16} + 1 - 3(\frac{1}{2} - \frac{1}{3} + \frac{1}{16}) = \frac{1}{2}\end{aligned}$$

Lema 70. *Se* X *e* Y *são variáveis aleatórias discretas com distribuição conjunta* $p_{X,Y}(x,y)$*, então a função de massa marginal de* X *é* $p_X(x) = \sum_{y\in Im[Y]} p_{X,Y}(x,y)$. *De forma similar, se* X *e* Y *são variáveis aleatórias continuas com densidade conjunta* $f_{X,Y}(x,y)$*, então a densidade marginal de* X *é* $f_X(x) = \int_{Im[Y]} f_{X,Y}(x,y)dy$.

Demonstração. Se X e Y são discretos,

$$\begin{aligned}p_X(x) &= \mathbb{P}(X=x) \\ &= \mathbb{P}(X=x, Y\in Im[Y]) \\ &= \sum_{\{(a,b):a=x,b\in Im[Y]\}} p_{X,Y}(a,b) && \text{(Definição 39)} \\ &= \sum_{b\in Im[Y]} p_{X,Y}(x,b)\end{aligned}$$

A demonstração do caso contínuo é deixada como exercício. □

Exemplo 68. *Considere que*

$$f_{X,Y}(x,y) = \begin{cases} c & , \text{ se } 0 \leq x \leq 1 \text{ e } 0 \leq y \leq 1 \\ 0 & , \text{ caso contrário} \end{cases}$$

Observe que

$$\begin{aligned} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dydx &= \int_0^1\int_0^1 cdydx \\ &= \int_0^1 cy\Big|_0^1 dx \\ &= \int_0^1 cdx = cx\Big|_0^1 = c \end{aligned}$$

Portanto, $c = 1$*. Observe que, if* $0 < x < 1$

$$\begin{aligned} F_X(x) &= \mathbb{P}(X \leq x) \\ &= \mathbb{P}(X \leq x \cap Y \in \mathbb{R}) \\ &= \int_{-\infty}^{x}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dydx \\ &= \int_0^x\int_0^1 1dydx \end{aligned}$$

De forma similar, se $x < 0$*, então* $F_X(x) = 0$ *e se* $x > 1$*, então* $F_X(x) = 1$*. Portanto* $X \sim$ *Uniform*$(0,1)$*.*

Note que a mesma conclusão segue do Lema 70:

$$f_X(x) = \int_{Im[Y]} f_{X,Y}(x,y)dy = \int_0^1 1dy = 1,$$

que é a densidade de uma variável aleatória uniforme em $(0,1)$*.*

Definição 40. *Sejam* X *e* Y *variávis aleatórias. A distribuição acumulada conjunta de* (X,Y) *é*

$$F_{X,Y}(x,y) = \mathbb{P}(X \leq x, Y \leq y)$$

Lembramos da Definição **??** que X e Y são independentes se e somente se $p_{X,Y}(x,y) = p_X(x)p_Y(y)$. A independência é definida de forma similar para variáveis aleatórias contínuas:

Definição 41. *Duas variáveis aleatórias contínuas* X *e* Y *são independentes se e somente se* $f_{X,Y}(x,y) = f_X(x)f_Y(y)$*.*

Lema 71. *Sejam* X *e* Y *duas variáveis aleatórias.* X *e* Y *são independentes se e somente se para todos* $A, B \subset \Re$*,* $\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$*.*

Lema 72. *Duas variáveis contínuas,* X *e* Y*, são independentes se e somente se a distribuição conjunta pode ser escrita como* $f_{X,Y}(x,y) = g(x)h(y)$ *para alguma função* g *e* h*.*

Demonstração. Se X e Y são independentes por definição $f_{X,Y}(x,y) = f_X(x)f_Y(y)$, que prova a condição suficiente. Além disso, se $f_{X,Y}(x,y) = g(x)h(y)$, então $f_X(x) = \int f_{X,Y}(x,y)dy = \int g(x)h(y)dy = g(x)C$, onde C não depende de x. Como $g(x)C$ deve integrar para um, segue que $g(x)$ é proporcional à densidade de X. De forma similar, $h(y)$ deve ser proporcional a densidade de Y. Segue que $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ e portanto X e Y são independentes. □

Em geral, um vetor aleatório é uma função $\mathbf{X} : \Omega \longrightarrow \mathbb{R}^d$ que é tipicamente representada como $\mathbf{X} = (X_1, \dots, X_d)$. Isto é, cada coordenada é uma variável aleatória. Um vetor aleatório é caracterizado quanto pela sua função de distribuição acumulada, tanto pela sua função de densidade (multivariada):

Definição 42. *A distribuição acumulada conjunta de $\mathbf{X} = (X_1, \dots, X_d)$ é*

$$F_{\mathbf{X}}(\mathbf{x}) := \mathbb{P}(X_1 \le x_1, \dots, X_d \le x_d),$$

onde $\mathbf{x} = (x_1, \dots, x_d)$. Se as componentes de $\mathbf{X}$ são contínuas, a fdp de $\mathbf{X}$, $f_{\mathbf{X}}(\mathbf{x})$, é uma função não negativa tal que

$$\mathbb{P}(\mathbf{X} \in B) = \int_B f_{\mathbf{X}}(x_1, \dots, x_d) dx_1 \dots dx_d$$

para qualquer conjunto $B \subset \mathbb{R}^d$. Além disso, se as componentes de $\mathbf{X}$ são discretas, a fmp de $\mathbf{X}$, $p_{\mathbf{X}}(\mathbf{x})$, é uma função não negativa tal que

$$\mathbb{P}(\mathbf{X} \in B) = \sum_{(x_1, \dots, x_d) \in B} p_{\mathbf{X}}(x_1, \dots, x_d)$$

para qualquer conjunto $B \subset \mathbb{R}^d$.

Lema 73. *Se $\mathbf{X} = (X_1, \dots, X_d)$ é um vetor aleatório, então*

1. $$\lim_{\substack{x_j \longrightarrow \infty \\ \forall j \neq i}} F_{\mathbf{X}}(\mathbf{x}) = F_{X_i}(x_i).$$

2. *If $X_1, \dots, X_n$ são variáveis aleatórias contínuas,*

$$\frac{\partial^d}{\partial x_1 \dots \partial x_d} F_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{X}}(\mathbf{x}).$$

Demonstração. Para a demonstração de 1., veja Exercício 166. 2. é a versão multivariada do Lema 49, e a sua demonstração segue a mesma linha. □

Exemplo 69. *Considere novamente a densidade conjunta no Exemplo 65:*

$$f_{XY}(x,y) = \begin{cases} 24xy & \text{, se } x > 0,\ y > 0,\ x + y < 1 \\ 0 & \text{, caso contrário} \end{cases}$$

Primeiro, note que X e Y não são independentes pois $f_{XY}(x,y) = 24xy\mathbb{I}_{0<x+y<1}$ não pode ser decomposta tal como no Lema 72. Agora, seja $\mathbf{X} = (X,Y)$. Nos vamos determinar a função de distribuição acumulada conjunta de $\mathbf{X}$ para $1 > u > 0$ e $v > 1$,

$$\begin{aligned} F_{\mathbf{X}}(u,v) = \mathbb{P}(X \le u, Y \le v) &= \int_0^u \int_0^v f_{XY}(x,y) dy dx = \int_0^u \int_0^{1-x} 24xy dy dx \\ = 24 \int_0^u x \frac{(1-x)^2}{2} dx &= 12 \left[\frac{x^2}{2} - \frac{2x^3}{3} + \frac{x^4}{4} \bigg|_{x=0}^{u} \right] = 12 \left[\frac{u^2}{2} - \frac{2u^3}{3} + \frac{u^4}{4} \right], \end{aligned}$$

e portanto, para $0 < u < 1$*,*

$$F_X(u) = \lim_{v \longrightarrow \infty} F_{\mathbf{X}}(u, v) = 12 \left[\frac{u^2}{2} - \frac{2u^3}{3} + \frac{u^4}{4} \right]$$

que podemos derivar com respeito a u *para obter a distribuição marginal de* Y*:*

$$f_X(u) = \frac{dF_X(u)}{du} = 12 \left(u - 2u^2 + u^3 \right)$$

for $0 < u < 1$*. Esta é a mesma função que obtemos calculando a marginal de* $f_{(X,Y)}$ *para* Y*:*

$$f_X(u) = \int_0^1 f_{(X,Y)}(u, y) dy = \int_0^{1-u} 24uydy = 12u(1-u)^2$$

Deixamos os outros casos para o leitor.

Temos a seguinte versão da lei do estatístico inconsciente (lembre-se do Lema 12) para vetores aleatórios:

Lema 74. *Seja* $\mathbf{X} = (X_1, \ldots, X_d)$ *um vetor aleatório. Definimos a função* $g : \Re^d \longrightarrow \Re$*. Então* $g(\mathbf{X})$ *é uma variável aleatória com esperança*

$$\mathbb{E}[g(\mathbf{X})] = \sum_{\mathbf{x} \in \chi} g(\mathbf{x}) p_{\mathbf{X}}(\mathbf{x})$$

se $\mathbf{X}$ *tem distribuição discreta (onde* $\chi \subset \mathbb{R}^d$ *denotam os valores que* $\mathbf{x}$ *assume), e*

$$\mathbb{E}[g(\mathbf{X})] = \int \ldots \int g(\mathbf{x}) f_{\mathbf{X}}(x_1, \ldots, x_d) dx_1 \ldots dx_n$$

Demonstração. Vamos provar o lema para o caso discreto. Da definição de esperança temos que

$$\begin{aligned}
\mathbb{E}[g(\mathbf{X})] &= \sum_{w \in \Omega} g(\mathbf{X}(w)) \mathbb{P}(\{w\}) = \\
&= \sum_{\mathbf{x} \in \chi} \sum_{w : \mathbf{X}(w) = \mathbf{x}} g(\mathbf{X}(w)) \mathbb{P}(\{w\}) \\
&= \sum_{\mathbf{x} \in \chi} g(\mathbf{x}) \sum_{w : \mathbf{X}(w) = \mathbf{x}} \mathbb{P}(\{w\}) \\
&= \sum_{\mathbf{x} \in \chi} g(\mathbf{x}) \cdot p_{\mathbf{X}}(\mathbf{x})
\end{aligned}$$

□

Exemplo 70. *No Exemplo 65, temos que*

$$\begin{aligned}
\mathbb{E}[XY] &= \int_{0<x+y<1} (xy) 24xydydx = 24 \int_0^1 \int_0^{1-x} (xy) xydydx = \int_0^1 x^2 \frac{(1-x)^3}{3} dx \\
&= \frac{\Gamma(3)\Gamma(4)}{3\Gamma(3+4)} \int_0^1 \frac{\Gamma(3+4)}{\Gamma(3)\Gamma(4)} x^{3-1} (1-x)^{4-1} dx = 1/(3*5*4*3).
\end{aligned}$$

7.1.1 Exerçicios

Exercício 166. *Mostre que* $\lim_{y \to \infty} F_{X,Y}(x, y) = F_X(x)$*. Dica: use Exercício 21.*

Solution:

$$
\begin{aligned}
F_X(x) &= P(X \leq x) \\
&= P(X \leq x \cap Y \in \mathbb{R}) \\
&= \lim_{y \to \infty} P(X \leq x, Y \leq y) \\
&= \lim_{y \to \infty} F_{X,Y}(x, y)
\end{aligned}
$$

Exercício 167. *Considere Exemplo 66. Mostre que X e Y não são independentes.*

Solução: Para mostrar que X e Y não são independentes, basta mostrar que $p_{X,Y}(x, y) \neq p_X(x)p_Y(y)$. Observe que

$$
\begin{aligned}
p_{X,Y}(1, 1) &= 0.1 \\
p_X(1) &= p_{X,Y}(1, 0) + p_{X,Y}(1, 1) = 0.3 + 0.1 = 0.4 \\
p_Y(1) &= p_{X,Y}(0, 1) + p_{X,Y}(1, 1) = 0.4 + 0.1 = 0.5
\end{aligned}
$$

Como $0.1 \neq 0.4 \cdot 0.5$, concluímos que X e Y não são independentes.

Exercício 168. *Considere Exemplo 68. Mostre que X e Y são independentes.*

Solução: Usando a Definição 41, basta mosrar que, para todos $x, y \in \mathbb{R}$, $f_{X,Y}(x, y) = f_X(x)f_Y(y)$. Lembramos do Exemplo 68 que $f_X(x) = I_{(0,1)}(x)$. De forma similar, $f_Y(y) = I_{(0,1)}(y)$. Assim,

$$
f_{X,Y}(x, y) = I_{(0,1)^2}(x, y) = I_{(0,1)}(x)I_{(0,1)}(y) = f_X(x)f_Y(y)
$$

e X e Y são independentes.

Exercício 169. *Encontre um exemplo de X e Y contínuos tais que X e Y Não são independentes*

Solução: Considere Exemplo 65. Se $0 < x_1 < 1$,

$$
\begin{aligned}
F_X(x_1) &= P(X \leq x_1) \\
&= P(X \leq x_1 \cap Y \in \mathbb{R}) \\
&= \int_{-\infty}^{x_1} \int_{-\infty}^{\infty} f_{X,Y}(x, y)dydx \\
&= \int_0^{x_1} \int_0^{1-x} 24xydydx \\
&= \int_0^{x_1} 24x\frac{(1-x)^2}{2}dx
\end{aligned}
$$

Portanto,

$$
\begin{aligned}
f_X(x_1) &= \frac{dF_X(x_1)}{dx_1} \\
&= \frac{d\int_0^{x_1} 24x\frac{(1-x)^2}{2}dx}{dx_1} \\
&= 24x_1\frac{(1-x_1)^2}{2}
\end{aligned}
$$

De forma similar, pela simetria de $f_{X,Y}(x,y)$, observe que,

$$f_Y(y_1) = 24y_1\frac{(1-y_1)^2}{2}$$

Finalmente, observe que

$$\begin{aligned}
f_{X,Y}(0.75, 0.75) &= 0 && (0.75 + 0.75 > 1) \\
f_X(0.75)f_Y(0.75) &= 24 \cdot 0.75\frac{(1-0.75)^2}{2} \cdot 24 \cdot 0.75\frac{(1-0.75)^2}{2}
\end{aligned}$$

Portanto, $f_{X,Y}(0.75, 0.75) \neq f_X(0.75)f_Y(0.75)$. Conclua, da Definição 41 que X e Y não são independentes.

Exercício 170. *Sejam $X \sim Exp(\alpha)$ e $Y \sim Exp(\beta)$ variáveis aleatórias independentes, $c, d > 0$. Encontre $\mathbb{P}(cX > dY)$.*

Solução: Seja $A = \{(x,y) \in \mathbb{R}^2 : cx > dy\}$.

$$\begin{aligned}
P(cX > dY) &= \int_A f_{X,Y}(x,y)d(x,y) \\
&= \int_A f_X(x)f_Y(y)d(x,y) && \text{(Definição 41)} \\
&= \int_0^\infty \int_0^{\frac{cx}{d}} \frac{1}{\alpha}e^{-\frac{x}{\alpha}}\frac{1}{\beta}e^{-\frac{y}{\beta}}dydx \textbf{***ATTN*** redo calc} \\
&= \int_0^\infty -\alpha e^{-\frac{x}{\alpha}}e^{-\frac{y}{\beta}}\Big|_0^{\frac{cx}{d}} dx \\
&= \int_0^\infty \alpha e^{-\frac{x}{\alpha}}(1 - e^{-\frac{cx}{d\beta}})dx \\
&= \int_0^\infty \alpha e^{-\frac{x}{\alpha}}dx - \int_0^\infty \alpha e^{-\frac{x}{\alpha}}e^{-\frac{cx}{d\beta}}dx \\
&= 1 - \int_0^\infty \alpha e^{-\frac{x}{\alpha}}e^{-\frac{cx}{d\beta}}dx \\
&= 1 - \alpha\int_0^\infty e^{-x\left(\frac{1}{\alpha}+\frac{c}{d\beta}\right)}dx \\
&= 1 - \alpha e^{-x\left(\frac{1}{\alpha}+\frac{c}{d\beta}\right)}\left(\frac{1}{\alpha}+\frac{c}{d\beta}\right)^{-1}\Big|_0^\infty \\
&= 1 - \alpha\left(\frac{1}{\alpha}+\frac{c}{d\beta}\right)^{-1}
\end{aligned}$$

Exercício 171. *Sejam $X \sim Uniforme(a,b)$ e $Y \sim Uniforme(c,d)$ variáveis aleatórias independentes e $a < c < b < d$. Encontre $\mathbb{P}(X < Y)$.*

Solução: Let $A = \{(x, y) \in \mathbb{R}^2 : x < y\}$.

$$\begin{aligned}
P(X < Y) &= \int_A f_{X,Y}(x, y) d(x, y) \\
&= \int_A f_X(x) f_Y(y) d(x, y) && \text{(Definição 41)} \\
&= \int_c^d \int_a^{\min(b,y)} \frac{1}{b-a} \frac{1}{d-c} dx dy \\
&= \int_c^b \int_a^y \frac{1}{(b-a)(d-c)} dx dy + \int_b^d \int_a^b \frac{1}{(b-a)(d-c)} dx dy \\
&= \int_c^b \frac{y-a}{(b-a)(d-c)} dy + \int_c^b \frac{1}{d-c} dy \\
&= \frac{(b^2-c^2) - 2a(b-c)}{2(b-a)(d-c)} + \frac{b-c}{d-c}
\end{aligned}$$

Exercício 172. *Sejam $X, Y, Z \sim U(0, 1)$ variáveis aleatórias independentes. Determine $\mathbb{P}(X \geq YZ)$.*

Exercício 173. *X e Y têm distribuição uniforme sobre o círculo unitário, isto é,*

$$f_{X,Y}(x, y) = \begin{cases} c & , \textit{se } x^2 + y^2 \leq 1 \\ 0 & , \textit{caso contrário} \end{cases}$$

a. Encontre c.

b. Encontre $P(X < Y)$.

c. Encontre $F_X(x)$ e $f_X(x)$.

d. Encontre $E[X]$ e $Var[X]$.

Exercício 174. *Considere a densidade conjunta de (X, Y)*

$$f(x, y) = \begin{cases} \frac{1}{2}xy & \textit{se } 0 < y < x < 2 \\ 0 & \textit{caso contrário} \end{cases}$$

- *Encontre as densidades marginais de X e Y, e calcule $\mathbb{E}[X]$, $\mathbb{E}[Y]$, $\mathbb{E}[XY]$, $Var[XY]$ e $Cov[X, Y]$.*
- *X e Y são independentes?*

Exercício 175. *Demonstre o Lema 71 para o caso discreto.*

Exercício 176. *Prove que se X e Y são independentes, então $f(X)$ é independente de $g(Y)$ para quaisquer funções que assumem valores nos reais f e g.*

7.2 *Alguns Modelos Multivariados

Tal como no caso univariado, há muitos modelos de probabilidade multivariados bem conhecidos. Aqui, vamos ilustrar dois deles.

Definição 43. *A esperança de um vetor aleatório $\mathbf{X} = (X_1, \ldots, X_d)$ é definida por $\mathbb{E}[\mathbf{X}] = (\mathbb{E}[X_1], \ldots, \mathbb{E}[X_d])$.*

Definição 44. *A matriz de variâncias e covariâncias de um vetor aleatório* $\mathbf{X} = (X_1, \ldots, X_d)$*,* $\mathbb{V}[\mathbf{X}]$*, é a matriz* $d \times d$ *cujo componente* (i, j) *é dado por* $Cov(X_i, X_j)$*.*

Lema 75. *Seja* $\mathbf{X} = (X_1, \ldots, X_d)$ *e* $\mathbf{a}$ *um vetor* d *dimensional. Então* $\mathbb{E}[\mathbf{a} \cdot \mathbf{X}] = \mathbf{a} \cdot \mathbb{E}[\mathbf{X}]$ *e* $VAR[\mathbf{a} \cdot \mathbf{X}] = \mathbf{a} \times VAR[\mathbf{X}] \times \mathbf{a}^t$*, onde "·"representa o produto escalar entre dois vetores, e* $\times$ *é o produto matricial.*

7.2.1 Distribuição Normal Multivariada

Definição 45. *Seja* $\mathbf{\Sigma}$ *uma matriz* $d \times d$ *não negativa definida e* $\boldsymbol{\mu} \in \mathbb{R}^d$*. Dizemos que* $\mathbf{X} \sim$ *Normal Multivariada*$(\boldsymbol{\mu}, \mathbf{\Sigma})$ *se, para todo* $\mathbf{x} \in \mathbb{R}^d$*,*

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^d |\mathbf{\Sigma}|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})\mathbf{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})'\right).$$

Lema 76. *Se* $\mathbf{X} \sim$ *Normal Multivariada*$(\boldsymbol{\mu}, \mathbf{\Sigma})$*, então* $\mathbb{E}[\mathbf{X}] = \boldsymbol{\mu}$ *e* $Var[\mathbf{X}] = \mathbf{\Sigma}$*.*

7.2.2 Distribuição Multinomial

A distribuição multinomial é a generalização da distribuição binomial. Suponha que lançamos um dado com d faces n vezes. Se os lançamentos são independentes e cada face tem probabilidade p_i, $i = 1, \ldots, d$, então $(X_1, \ldots, X_d) \sim \text{Multinomial}(n, (p_1, \ldots, p_d))$, onde X_i é o número de vezes que a face i é observada.

Definição 46. *Seja* $\boldsymbol{p} = (p_1, \ldots, p_d)$ *tal que* $p_i > 0$ *para todo* i *e* $\sum_{i=1}^{d} p_i = 1$ *e* $n \in \mathbb{N}$*. Dizemos que* $\mathbf{X} = (X_1, \ldots, X_d) \sim$ *Multinomial*$(n, \boldsymbol{p})$ *se*

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \frac{n!}{x_1! \ldots x_n!} \prod_{i=1}^{d} p_i^{x_i},$$

para todo $x_i \geq 0$ *com* $\sum_{i=1}^{n} x_i = n$*.*

Lema 77. *Se* $\mathbf{X} = (X_1, \ldots, X_d) \sim$ *Multinomial*$(n, \boldsymbol{p})$*, então* $\mathbb{E}[\mathbf{X}] = n\boldsymbol{p}$ *e a componente* (i, j) *de* $Var[\mathbf{X}]$ *é* $-np_ip_j$ *para* $i \neq j$ *e* $np_i(1 - p_i)$ *caso contrário.*

Exemplo 71. *Há três candidatos a uma eleição, A, B e C. Se 10% da população vota em A, 40% vota em B e 50% vota em C, então a probabilidade de que em uma amostra de tamanho* $n = 10$*, retirada desta população com reposição, obteremos 2 eleitores de A, 3 de B e 5 de C é dada por*

$$\frac{10!}{2!3!5!}(0.1)^2(0.4)^3(0.5)^5.$$

7.2.3 Exercícios

Exercício 177. *Prove o Lema 75*

Exercício 178. *Se* $\mathbf{X} \sim$ *Normal Multivariada*$(\boldsymbol{\mu}, \mathbf{\Sigma})$ *e* $\mathbf{a}$ *é um vetor* d *dimensional, qual é a esperança e a variância de* $\mathbf{a} \cdot \mathbf{X}$*?*

Exercício 179. *Suponhamos que* $\mathbf{X} = (X_1, X_2)$ *tem distribuição normal multivariada com média* μ *e matriz de variâncias e covaroâncias* Σ*, onde* $(\Sigma)_{i,j} = 0$ *se* $i \neq j$*. Mostre que* X_1 *é independente de* X_2*, e encontre a respectiva distribuição. Este exercício mostra que, para distribuições normais multivariadas, a covariância zero implica em independência. Nos vimos que isso não é verdade, de modo geral, para as outras distribuições.*

Exercício 180. *Discuta porque a distribuição multinomial realmente modela o número de vezes cada face do dado é observada.*

7.3 Distribuições Condicionais

Definição 47. *Se X e Y são variáveis aleatórias discretas, definimos a* massa de probabilidade condicional *de X dado que $Y = y$ por*

$$p_{X|Y}(x|y) = \mathbb{P}(X = x|Y = y) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)},$$

para todos os valores de y tais que $\mathbb{P}(Y = y) > 0$.

Note que, para um dado y, $g(x) = p_{X|Y}(x|y)$ é uma fmp n sentido em que a estudamos antes.

Exemplo 72. *No Exemplo 66,*

$$p_{X|Y}(0|0) = \frac{0.2}{0.5} = 0.4 = 1 - p_{X|Y}(1|0)$$

e

$$p_{X|Y}(0|1) = \frac{0.4}{0.5} = 0.8 = 1 - p_{X|Y}(1|1).$$

Remark: Podemos extender de forma trivial a Definição 47 para vetores aleatórios.

Definição 48. *Se X e Y são variáveis aleatórias contínuas com densidade conjunta $f(x, y)$, definimos a* densidade de probabilidade condicional *de X dado $Y = y$ por*

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)},$$

para todos os valores de y tais que $f_Y(y) > 0$.

Note que, para um dado y, $g(x) = f_{X|Y}(x|y)$ é uma fdp n sentido em que a estudamos antes.

Exemplo 73. *Seja densidade conjunta de X e Y definida por*

$$f(x, y) = \begin{cases} \frac{15}{2}x(2 - x - y) & \text{se } 0 < x, y < 1 \\ 0 & \text{caso contrário} \end{cases}$$

Para calcular a densidade condicional de $X|Y = y$ quando $0 < y < 1$, primeiro observamos que

$$f_Y(y) = \int_0^1 f(x, y)dx = \int_0^1 \frac{15}{2}x(2 - x - y)dx = \frac{15}{2}(\frac{2}{3} - \frac{y}{2}).$$

Segue da Definição 48 que, para $0 < x, y < 1$,

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)} = \frac{x(2 - x - y)}{\frac{2}{3} - \frac{y}{2}} = \frac{6x(2 - x - y)}{4 - 3y}.$$

Distribuições condicionais podem ser usadas para checar independência:

Lema 78. *Sejam X e Y duas variáveis aleatórias discretas. X é independente de Y se e somente se $\mathbb{P}(X = x|Y = y) = \mathbb{P}(X = x)$ para todos x e y. Analogamente, sejam X e Y duas variáveis aleatórias contínuas. X é independente de Y se e somente se $f(x|y) = f(x)$ para todos x e y.*

O Teorema de Bayes também pode ser usado para variáveis discretas e contínuas:

Lema 79. *Sejam X e Y variáveis aleatórias contínuas, e N e M variáveis aleatórias discretas. Então valem as seguintes versões do Teorema de Bayes:*

1.
$$\mathbb{P}(N=n|M=m)=\frac{\mathbb{P}(M=m|N=n)\mathbb{P}(N=n)}{\sum_n \mathbb{P}(M=m|N=n)\mathbb{P}(N=n)}$$

2.
$$f(x|y)=\frac{f(y|x)f(x)}{\int_{\mathbb{R}} f(y|x)f(x)dx}$$

Também é possível falar sobre distribuições condicionais de vetores aleatórios que são parcialmente contínuos e parcialmente discretos. Assim, se X é contínua e N é discreta, então

$$f(x|N=n)=\frac{\mathbb{P}(N=n|X=x)f(x)}{\int_{\mathbb{R}} \mathbb{P}(N=n|X=x)f(x)dx}$$

e

$$\mathbb{P}(N=n|X=x)=\frac{f(x|N=n)\mathbb{P}(N=n)}{\sum_n f(x|N=n)\mathbb{P}(N=n)}.$$

Observe que $\mathbb{P}(N=n)=\int_{\mathbb{R}} \mathbb{P}(N=n|X=x)f(x)dx$ e $f(y)=\int_{\mathbb{R}} f(y|x)f(x)dx$ são versões contínuas da Lei das Probabilidades Totais (Lema 3).

Como as probabilidades condicionais também são probabilidades, também vale que, para toda variável aleatória Z, $\mathbb{P}(N=n|Z=z)=\int_{\mathbb{R}} \mathbb{P}(N=n|Z=z,X=x)f(x|Z=z)dx$ e $f(y|Z=z)=\int_{\mathbb{R}} f(y|Z=z,x)f(x|Z=z)dx$.

Exemplo 74. *Assuma que escolhemos um número $U \sim$ Unif$(0,1)$. EnTão lançamos uma moeda com probabilidade de cara U. Seja X o indicador de cara. Temos que*

$$\mathbb{P}(X=1)=\int_0^1 \mathbb{P}(X=1|u)f(u)du=\int_0^1 u\cdot 1du=1/2.$$

Segue do Teorema de Bayes que

$$f(u|X=1)=\frac{\mathbb{P}(X=1|u)f(u)}{1/2}=2u.$$

Isto é, $U|X=1 \sim$ Beta$(2,1)$

Observação: Note que na verdade não precisamos calcular o denominador no Teorema de Bayes para encontrar a distribuição de $U|X=1$ no último exemplo. Isso ocorre pois o Teorema afirma que $f(u|x)=C\mathbb{P}(X=x|u)f(u)$, onde C é constante em u (mais precisamente, é a única constante que faz com que $f(u|x)$ seja uma densidade genuina que integra para um). segue que $f(u|x)=c\cdot u$, e a única constante que faz com que esta densidade integre para um é 2.

Também é possível definir independência condicional de variaveis aleatórias:

Definição 49. *Dizemos que duas variáveis aleatórias discretas X e Y são independentes dado uma variável aleatória Z se $\mathbb{P}(X=x,Y=y|Z=z)=\mathbb{P}(X=x|Z=z)\mathbb{P}(Y=y|Z=z)$*

A independência condicional é definida de forma análoga para vetores aleatórios contínuos.

Exemplo 75. *Considere Exemplo 74 novamente, mas agora lançando a moeda duas vezes. Sejam X_1 e X_2 as indicadoras de que o resultado é caras. É razoável assumir que X_1 é independente de X_2 dado que a probabilidade de caras é $U=u$. Segue que*

$$\mathbb{P}(X_2=1|X_1=1)=\int_0^1 \mathbb{P}(X_2=1|X_1=1,u)f(u|X_1=1)du.$$

Segue do Exercício 74 que

$$\mathbb{P}(X_2 = 1|X_1 = 1) = \int_0^1 u2udu = \frac{2}{3} \neq \frac{1}{2} = \int_0^1 \mathbb{P}(X_2 = 1|u)f(u)du = \mathbb{P}(X_2 = 1),$$

i.e., X_1 e X_2 não são independentes.

7.3.1 Esperança Condicional de Variáveis Aleatórias

Aplicando a Definição 19 a variáveis aleatórias, temos que a esperança condicional de uma variável aleatória discreta X que assume valores em χ dado que outra variável aleatória discreta Y assume o valor y é dada por

$$\begin{aligned}\mathbb{E}[X|Y = y] := \sum_{w\in\Omega} X(w)P(\{w\}|Y = y) &= \sum_{x\in\chi}\sum_{w:X(w)=x} X(w)P(\{w\}|Y = y)\\ &= \sum_{x\in\chi} x \sum_{w:X(w)=x} P(\{w\}|Y = y) = \sum_{x\in\chi} xP(X = x|Y = y)\\ &\sum_{x\in\chi} xp_{X|Y}(x|y)\end{aligned}$$

De forma análoga, quando X e Y são contínuas,

$$\mathbb{E}[X|Y = y] = \int_{x\in\chi} xf_{X|Y}(x|y)dx$$

Exemplo 76. *Assuma que a densidade conjunta de X e Y é dada por $f_{X,Y}(x, y) = e^{-x/y}e^{-y}/y$ para $0 < x, y < \infty$. A densidade condicional de X dado que $Y = y$ é dada por*

$$f_{X|Y}(x|y) = \frac{e^{-x/y}e^{-y}/y}{\int_0^\infty e^{-x/y}e^{-y}/ydx} = \frac{e^{-x/y}e^{-y}/y}{e^{-y}/y\int_0^\infty e^{-x/y}dx} = 1/ye^{-x/y}.$$

Segue que $\mathbb{E}[X|Y = y] = \int_0^\infty x1/ye^{-x/y}dx = y$. Note que $X|Y = y \sim Exp(1/y)$.

Usamos a notação $\mathbb{E}[X|Y]$ para denotar a função em Y dada por $g(y) = \mathbb{E}[X|Y = y]$. Como Y é uma variável aleatória, $g(Y) = \mathbb{E}[X|Y]$ também o é. Um resultado muito útil para calcular esperanças é o seguinte

Teorema 5. *Sejam X e Y duas variáveis aleatórias. Então $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|Y]]$.*

Demonstração. Provaremos o teorema no caso de X discreto. Lembramos do Lema 15 que

$$\mathbb{E}[X] = \sum_i \mathbb{E}[X|A_i]\cdot\mathbb{P}(A_i)$$

onde $(A_i)_i$ é a partição de Ω. Agora, seja $A_i = \{\omega \in \Omega : Y(\omega) = i\}$ para todo $i \in \chi$, a imagem de Y. Como $(A_i)_i$ é uma partição de Ω,

$$\mathbb{E}[X] = \sum_i \mathbb{E}[X|A_i]\cdot\mathbb{P}(A_i) = \sum_{i\in\chi} \mathbb{E}[X|Y = i]p_Y(i),$$

o que conclui a prova (porque?). □

Exemplo 77. *N pessoas decidem doar dinheiro a uma instituição. Seja X_i a quantia doada pela i-ésima pessoa. Assumindo que X_i é binomial com parâmetros n_B e p_B, que N é geométrica com parâmetro p_G, e que todas as* *variáveis aleatórias são independentes, podemos calcular a soma esperada de dinheiro que a instituição vai receber, $T := \sum_{i=1}^{N} X_i$ usando o Teorema 5. Primeiro, determinamos a esperança condicional*

$$\mathbb{E}[T|N = n] = \mathbb{E}[\sum_{i=1}^{N} X_i|N = n] = \mathbb{E}[\sum_{i=1}^{n} X_i|N = n] = \sum_{i=1}^{n} \mathbb{E}[X_i|N = n] = nn_B p_B.$$

Assim $\mathbb{E}[T|N] = Nn_B p_B$. Segue que

$$\mathbb{E}[T] = \mathbb{E}[\mathbb{E}[T|N]] = n_B p_B \mathbb{E}[N] = \frac{n_B p_B}{p_G}$$

7.3.2 Variância Condicional de Variáveis Aleatórias

Também podemos definir a variância condicional de variáveis aleatórias. Temos que

$$\mathbb{V}[X|Y = y] = \mathbb{E}[(X - \mathbb{E}[X|Y = y])^2|Y = y] = \mathbb{E}[X^2|Y = y] - \mathbb{E}^2[X|Y = y]$$

A very useful result for computing variances is the following

Teorema 6. *Let X and Y be two random variables. Then $\mathbb{V}[X] = \mathbb{V}[\mathbb{E}[X|Y]] + \mathbb{E}[\mathbb{V}[X|Y]]$.*

Demonstração. First, notice that

$$\mathbb{V}[\mathbb{E}[X|Y]] = \mathbb{E}[(\mathbb{E}[X|Y])^2] - \mathbb{E}^2[\mathbb{E}[X|Y]] = \mathbb{E}[(\mathbb{E}[X|Y])^2] - \mathbb{E}^2[X]. \tag{18}$$

We also have that $\mathbb{V}[X|Y] = \mathbb{E}[X^2|Y] - \mathbb{E}^2[X|Y]$, and hence,

$$\mathbb{E}[\mathbb{V}[X|Y]] = \mathbb{E}[\mathbb{E}[X^2|Y]] - \mathbb{E}[\mathbb{E}^2[X|Y]] = \mathbb{E}[X^2] - \mathbb{E}[(\mathbb{E}[X|Y])^2] \tag{19}$$

By adding Equations 18 and 19 and noticing that $Var[X] = \mathbb{E}[X^2] - \mathbb{E}^2[X]$ we conclude the proof. $\square$

Exemplo 78. *In Example 77, it is possible to compute the variance of T using Theorem 6. First, we have already shown that $\mathbb{E}[T|N] = Nn_B p_B$. It follows that $\mathbb{V}[\mathbb{E}[T|N]] = \mathbb{V}[Nn_B p_B] = (n_B p_B)^2 \mathbb{V}[N] = (n_B p_B)^2 \frac{1-p_G}{p_G^2}$. Moreover, $\mathbb{V}[T|N = n] = \mathbb{V}[\sum_{i=1}^{n} X_i|N = n] = \sum_{i=1}^{n} \mathbb{V}[X_i] = nn_B p_B(1 - p_B)$. Hence, $\mathbb{E}[\mathbb{V}[T|N]] = \mathbb{E}[Nn_B p_B(1 - p_B)] = n_B p_B(1 - p_B)\frac{1}{p_G}$. From Theorem 6, we conclude that*

$$\mathbb{V}[T] = n_B p_B(1 - p_B)\frac{1}{p_G} + (n_B p_B)^2 \frac{1 - p_G}{p_G^2}$$

7.3.3 Exercises

Exercício 181. *Prove Lemma 78.*

Exercício 182. *Define conditional independence for continuous random vectors.*

Exercício 183. *Prove that if X and Y are independent Poisson random variables with rates λ_1 and λ_2 respectively, then $X|Z = n$ is binomial with parameters n and $\frac{\lambda_1}{\lambda_1+\lambda_2}$, where $Z = X + Y$.*

Exercício 184. *Consider the joint density of (X, Y)*

$$f(x, y) = \begin{cases} xe^{-x(y+1)} & \textit{if } x > 0 \textit{ and } y > 0 \\ 0 & \textit{otherwise} \end{cases}$$

- *Find the marginal densities of X given $Y = y$ and $\mathbb{E}[X|Y = y]$. What is the expectation of X?*
- *Compute $\mathbb{P}(0 < X < 1|Y = 2)$, $\mathbb{P}(X < Y|X < 2Y)$, and $\mathbb{P}(1 < X + Y < 2)$.*

Exemplo 79. *Say $\mathbf{X} = (X_1, X_2)$ has a multivariate normal distribution with mean μ and variance covariance vector Σ. Compute the conditional distribution of $X_1|X_2 = x_2$.*

Exercício 185. *Prove that if $\mathbf{X} \sim$ Multinomial$(n, (p_1, \ldots, p_d))$, then $X_1 \sim$ Binomial(n, p_1).*

Exercício 186. *Prove that if X is a discrete random variable, then, for every $a \in \mathbb{R}$, $\mathbb{E}[X] = \mathbb{E}[X|X < a]\mathbb{P}(X < a) + \mathbb{E}[X|X > a]\mathbb{P}(X > a)$.*

Exercício 187. *(Ross, pg 379) Ten hunters are waiting for ducks to fly by. When a flock of ducks flies overhead, the hunters fire at the same time, but each chooses his target at random, independently of the others. If each hunter independently hits his target with probability .6, compute the expected number of ducks that are hit. Assume that the number of ducks in a flock is a Poisson random variable with mean 6.*

Exercício 188. *Consider Example 74 again. Say that instead of flipping the coin only once, we flip it n times. Without computing the (unconditional) distribution of X, the total number of heads you observe, compute $\mathbb{E}[X]$, $\mathbb{V}[X]$ and $f(u|x)$. Assume the flips are independent conditionally on U.*

Exercício 189. *If $N|Y \sim$ Poisson(Y) and $Y \sim$ Gamma(k, λ), what is the distribution of $Y|N = n$?*

7.4 Univariate Transformations

Lema 80. *Let X be a discrete random variable and f be a function. If $Y = f(X)$, then*

$$p_Y(y) = \sum_{\{x: f(x)=y\}} p_X(x)$$

Demonstração.

$$\begin{aligned}
p_Y(y) &= \mathbb{P}(f(X) = y) \\
&= \sum_{x \in Im[X]} \mathbb{P}(f(X) = y|X = x)\mathbb{P}(X = x) && \text{Law of Total Probability (Lemma 3)} \\
&= \sum_{x \in Im[X]: f(x)=y} \mathbb{P}(X = x) && \mathbb{P}(f(X) = y|X = x) = 1, \text{ if f(x)=y, and 0 otherwise}
\end{aligned}$$

□

Exemplo 80. *Let X be a discrete random variable with the following pmf: Let $g(x) = x^2$. Using Lemma 80,*

x	-2	-1	0	1	2	3
$p_X(x)$	0.1	0.2	0.3	0.1	0.25	0.05

$p_{g(X)}(0) = p_X(0) = 0.3$, $p_{g(X)}(1) = p_X(-1) + p_X(1) = 0.3$, $p_{g(X)}(4) = p_X(-2) + p_X(2) = 0.35$ *and* $p_{g(X)}(9) = p_X(3) = 0.05$. *Hence,*

x	0	1	4	9
$p_{X^2}(x)$	0.3	0.3	0.35	0.05

Lema 81. *Let g be a strictly increasing and differentiable function and X be a countinuous random variable. If $Y = g(x)$, then*

$$f_Y(y) = f_X(g^{-1}(y)) \cdot \frac{\partial g^{-1}(y)}{\partial y}$$

Demonstração.

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \leq y) \\ &= \mathbb{P}(g(X) \leq y) \\ &= \mathbb{P}(X \leq g^{-1}(y)) && g \text{ is strictly increasing} \end{aligned}$$

Hence, using Lemma 49,

$$\begin{aligned} f_Y(y) &= \frac{\partial F_Y(y)}{\partial y} \\ &= \frac{\partial \mathbb{P}(X \leq g^{-1}(y))}{\partial y} \\ &= \frac{\partial \mathbb{P}(X \leq g^{-1}(y))}{\partial g^{-1}(y)} \cdot \frac{\partial g^{-1}(y)}{\partial y} && \text{Chain rule} \\ &= f_X(g^{-1}(y)) \cdot \frac{\partial g^{-1}(y)}{\partial y} \end{aligned}$$

□

Exemplo 81. *Let $g(x) = ax + b$, $a > 0$, and $Y = g(X)$, then $f_Y(y) = \frac{1}{a} f_X(\frac{y-b}{a})$. This follows through Lemma 81 and observing that $g^{-1}(y) = \frac{y-b}{a}$.*

Exemplo 82. *Let $X \sim Exp(\lambda)$ and $g(x) = x^2$. Let $Y = g(X)$. Observe that $X \geq 0$ and $g(x) : \mathbb{R}^+ \to \mathbb{R}$ is strictly increasing. Also observe that $g^{-1}(y) = \sqrt{y}$. Hence,*

$$\begin{aligned} f_Y(y) &= f_X(\sqrt{y}) \cdot \frac{\partial \sqrt{y}}{\partial y} && \textit{(Lemma 81)} \\ &= \frac{1}{\lambda} e^{-\frac{\sqrt{y}}{\lambda}} \cdot \frac{1}{2\sqrt{y}} && \textit{(Table 3)} \\ &= \frac{1}{2\lambda\sqrt{y}} e^{-\frac{\sqrt{y}}{\lambda}} \end{aligned}$$

Exemplo 83. *Let $Z \sim N(0, 1)$ and $g(z) = z^2$. What is the distribution of $g(Z)$? Observe that g is not strictly increasing on $\mathbb{R}$. Hence, we cannot use Lemma 81. We can follow similar steps, though:*

$$\begin{aligned} \mathbb{P}(Z^2 \leq y) &= \mathbb{P}(-\sqrt{y} \leq Z \leq \sqrt{y}) \\ &= 1 - \mathbb{P}(Z \leq -\sqrt{y} \cup Z \geq \sqrt{y}) \\ &= 1 - (\mathbb{P}(Z \leq -\sqrt{y}) + \mathbb{P}(Z \geq \sqrt{y})) && \textit{Additivity (Lemma 2)} \\ &= 1 - 2\mathbb{P}(Z \leq -\sqrt{y}) && \textit{Simmetry of the standard normal around } 0 \\ &= 1 - 2F_Z(-\sqrt{y}) \end{aligned}$$

Hence,

$$\begin{aligned}
f_{Z^2}(y) &= \frac{\partial \mathbb{P}(Z^2 \leq y)}{\partial y} && \textit{(Lemma 49)} \\
&= \frac{\partial(1 - 2F_Z(-\sqrt{y}))}{\partial(-\sqrt{y})} \cdot \frac{\partial(-\sqrt{y})}{\partial y} && \textit{Chain rule} \\
&= -2f_Z(-\sqrt{y}) \cdot \frac{-1}{2\sqrt{y}} \\
&= \frac{1}{\sqrt{\pi}\sqrt{2}} \cdot \frac{1}{\sqrt{y}} e^{-\frac{(\sqrt{y})^2}{2}} && \textit{(Table 3)} \\
&= \frac{1}{\Gamma(0.5)2^{0.5}} y^{0.5-1} e^{-\frac{y}{2}} && \Gamma(0.5) = \sqrt{\pi}
\end{aligned}$$

Hence, $Z^2 \sim Gamma(0.5, 2)$. This is also called a Chi-square distribution with one degree of freedom.

Lema 82. *Let $g(x) = ax + b$, $a \neq 0$, and $Y = g(X)$, then $f_Y(y) = \frac{1}{|a|} f_X(\frac{y-b}{a})$.*

Demonstração. If $a > 0$, the proof follows from Example 81. If $a < 0$,

$$\begin{aligned}
F_Y(y) &= \mathbb{P}(Y \leq y) \\
&= \mathbb{P}(aX + b \leq y) \\
&= \mathbb{P}\left(X > \frac{y-b}{a}\right) \\
&= 1 - F_X\left(\frac{y-b}{a}\right)
\end{aligned}$$

Using Lemma 49,

$$\begin{aligned}
f_Y(y) &= \frac{\partial F_Y(y)}{\partial y} \\
&= \frac{\partial(1 - F_X(\frac{y-b}{a}))}{\partial y} \\
&= -\frac{1}{a} f_X\left(\frac{y-b}{a}\right) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)
\end{aligned}$$

□

7.4.1 Exercises

Exercício 190. *Use Lemma 82 to prove the following results:*

a. If $X \sim Uniform(a, b)$, then $cX + d \sim Uniform(\min(ca + d, cb + d), \max(ca + d, cb + d))$.

b. If $X \sim Gamma(k, \lambda)$, then If $c > 0$, $cX \sim Gamma(k, c\lambda)$.

c. If $X \sim Beta(\alpha, \beta)$, then $1 - X Beta(\beta, \alpha)$

d. If $X \sim N(\mu, \sigma^2)$, then $aX + b \sim N(a\mu + b, \sigma^2 a^2)$.

Exercício 191. *Let X be a continuous random variable with distribution function F, and let $Y := F(X)$. Show that Y has uniform distribution over $(0, 1)$.*

7.5 Bivariate Transformations

Lema 83 (Convolution of discrete random variables)**.** *Let X and Y be independent random variables and $Z = X+Y$.*

$$p_Z(z) = \sum_{x \in Im[X]} p_X(x) p_Y(z-x)$$

Demonstração.

$$\begin{aligned}
p_Z(z) &= \mathbb{P}(X+Y=z) \\
&= \sum_{x \in Im[X]} \mathbb{P}(X+Y=z|X=x)\mathbb{P}(X=x) && \text{Law of Total Probability (Lemma 3)} \\
&= \sum_{x \in Im[X]} \mathbb{P}(Y=z-x)\mathbb{P}(X=x) && \text{Independence of } X \text{ and } Y \\
&= \sum_{x \in Im[X]} p_Y(z-x)p_X(x)
\end{aligned}$$

□

Lema 84 (Convolution of continuous random variables)**.** *Let X and Y be independent random variables and $Z = X+Y$.*

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx$$

Demonstração.

$$\begin{aligned}
F_Z(z) &= \mathbb{P}(Z \leq z) = \mathbb{P}(X+Y \leq z) \\
&= \int_{\{(x,y):x+y\leq z\}} f_{X,Y}(x,y) d(x,y) \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f_{X,Y}(x,y) dxdy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f_X(x) f_Y(y) dxdy && \text{Independence of } X \text{ and } Y
\end{aligned}$$

$$\begin{aligned}
f_Z(z) &= \frac{\partial F_Z(z)}{\partial z} && \text{(Lemma 49)} \\
&= \frac{\partial \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f_X(x) f_Y(y) dxdy}{\partial z} \\
&= \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx && \text{Fundamental Theorem of Calculus}
\end{aligned}$$

□

Exemplo 84. *If $X \sim Binomial(n,p)$ and $Y \sim Binomial(m,p)$ are independent, then $X+Y \sim Binomial(n+m,p)$.*

This is because

$$\begin{aligned}
p_{X+Y}(z) &= \sum_{x=0}^{n} p_X(x)p_Y(z-x) && \textit{(Lemma 83)}\\
&= \sum_{x=0}^{\min(n,z)} \binom{n}{x} p^x(1-p)^{n-x}\binom{m}{z-x}p^{z-x}(1-p)^{n-z+x} && \textit{(Table 2)}\\
&= p^z(1-p)^{n+m-z}\sum_{x=0}^{\min(n,z)}\binom{n}{x}\binom{m}{z-x} = \binom{n+m}{z}p^z(1-p)^{n+m-z}
\end{aligned}$$

Exemplo 85. *If $X \sim$ Negative Binomial(r_1,p) and $Y \sim$ Negative Binomial(r_2,p) are independent, then $X+Y \sim$ Negative Binomial(r_1+r_2,p). This is because*

$$\begin{aligned}
p_{X+Y}(z) &= \sum_{x=r_1}^{\infty} p_X(x)p_Y(z-x) && \textit{(Lemma 83)}\\
&= \sum_{x=r_1}^{r_2-1-z}\binom{x}{r_1-1}p^{r_1}(1-p)^{x-r_1}\binom{z-x}{r_2-1}p^{r_2}(1-p)^{z-x-r_2} && \textit{(Table 2)}\\
&= p^{r_1+r_2}(1-p)^{z-r_1-r_2}\sum_{x=r_1}^{r_2-1-z}\binom{x}{r_1-1}\binom{z-x}{r_2-1} = \binom{z}{r_1+r_2-1}p^z(1-p)^{n+m-z}
\end{aligned}$$

Exemplo 86. *If $X \sim$ Poisson(λ_1) and $Y \sim$ Poisson(λ_2) are independent, then $X+Y \sim$ Poisson$(\lambda_1+\lambda_2)$. This is because*

$$\begin{aligned}
p_{X+Y}(z) &= \sum_{x=0}^{\infty} p_X(x)p_Y(z-x) && \textit{(Lemma 83)}\\
&= \sum_{x=0}^{z}\frac{e^{-\lambda_1}\lambda_1^x}{x!}\cdot\frac{e^{-\lambda_2}\lambda_2^{z-x}}{(z-x)!} && \textit{(Table 2)}\\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{z!}\sum_{x=0}^{z}\frac{z!}{x!(z-x)!}\lambda_1^x\lambda_2^{z-x}\\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{z!}\sum_{x=0}^{z}\binom{z}{x}\lambda_1^x\lambda_2^{z-x}\\
&= \frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^z}{z!} && \textit{Binomial Theorem (Lemma 33)}
\end{aligned}$$

Exercício 192. *If $X \sim N(a,b^2)$ and $Y \sim N(c,d^2)$ are independent, then $Z = X+Y \sim N(a+c,b^2+d^2)$. Indeed, using Exercise 190, $X = a+M$, $Y = c+N$, where M and N are independent and $M \sim N(0,b^2)$ and $N \sim N(0,d^2)$.*

Hence $X+Y=a+c+(M+N)$. Let $O=M+N$.

$$\begin{aligned}
f_O(o) &= \int_{-\infty}^{\infty} f_M(m) f_N(o-m)dm && \textit{Convolution (Lemma 84)}\\
&= \int_{-\infty}^{\infty} \frac{1}{b\sqrt{2\pi}} e^{-\frac{m^2}{2b^2}} \cdot \frac{1}{d\sqrt{2\pi}} e^{-\frac{(o-m)^2}{2d^2}} dm\\
&= \frac{1}{bd2\pi}\int_{-\infty}^{\infty} e^{-\frac{m^2}{2b^2}} e^{-\frac{(o-m)^2}{2d^2}} dm\\
&= \frac{1}{bd2\pi}\int_{-\infty}^{\infty} e^{-\frac{(b^2+d^2)m^2-2b^2om}{2b^2d^2}} e^{-\frac{o^2}{2d^2}} dm\\
&= \frac{1}{bd2\pi}\int_{-\infty}^{\infty} e^{-\frac{(b^2+d^2)m^2-2b^2om+\frac{b^4o^2}{b^2+d^2}}{2b^2d^2}} e^{\frac{o^2b^2}{2d^2(b^2+d^2)}} e^{-\frac{o^2}{2d^2}} dm && \textit{Completing the square...}\\
&= \frac{1}{bd2\pi} e^{-\frac{o^2}{b^2+d^2}} \int_{-\infty}^{\infty} e^{-\frac{\left(m-\frac{b^2o}{b^2+d^2}\right)^2}{\frac{2b^2d^2}{b^2+d^2}}} dm\\
&= \frac{1}{bd2\pi} e^{-\frac{o^2}{b^2+d^2}} \frac{\sqrt{2\pi}bd}{\sqrt{b^2+d^2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\frac{bd}{\sqrt{b^2+d^2}}} e^{-\frac{\left(m-\frac{b^2o}{b^2+d^2}\right)^2}{\frac{2b^2d^2}{b^2+d^2}}} dm\\
&= \frac{1}{\sqrt{(b^2+d^2)}\sqrt{2\pi}} e^{-\frac{o^2}{b^2+d^2}} \cdot 1 && \textit{Normal density integrates to 1.}
\end{aligned}$$

A useful lemma for computing bivariate transformations is the following:

Lema 85. *Let X_1 and X_2 be two continuous random variables, and let $Y_1=g_1(X_1,X_2)$ and $Y_2=g_2(X_1,X_2)$, where g_1 and g_2 are continuous functions. If $y_1=g_1(x_1,x_2)$ and $y_2=g_2(x_1,x_2)$ can be uniquely solved for x_1 and x_2 in terms of y_1 and y_2, g_1 and g_2 have continuous partial derivatives, and*

$$J(x_1,x_2) = \begin{vmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} \end{vmatrix} \neq 0$$

for every x_1, x_2, then

$$f_{Y_1,Y_2}(y_1,y_2) = \frac{f_{X_1,X_2}(x_1,x_2)}{|J(x_1,x_2)|},$$

where $x_1 = G_1(y_1,y_2)$, $x_2=G_2(y_1,y_2)$, and G_1 and G_2 are the inverse transformations of g_1 and g_2 in the sense that $x_1=G_1(g_1(x_1,x_2),g_2(x_1,x_2))$ and $x_2=G_2(g_1(x_1,x_2),g_2(x_1,x_2))$.

Exemplo 87. *Let $X_1 \sim Gamma(\alpha,\lambda)$ and $X_2 \sim Gamma(\beta,\lambda)$. We will find the joint distribution of $Y_1=X_1+X_2$ and $Y_2=X_1/(X_1+X_2)$. Here, $g_1(x_1,x_2)=x_1+x_2$ and $g_2(x_1,x_2)=x_1/(x_1+x_2)$. We have*

$$J(x_1,x_2) = \begin{vmatrix} 1 & 1 \\ x_2/(x_1+x_2)^2 & -x_1/(x_1+x_2)^2 \end{vmatrix} = -\frac{1}{x_1+x_2}$$

As $x_1=y_1y_2$ and $x_2=y_1(1-y_2)$, it follows from Lemma 85 that

$$f_{Y_1,Y_2}(y_1,y_2) = f_{X_1,X_2}(y_1y_2, y_1(1-y_2))\frac{1}{y_1y_2+y_1(1-y_2)} = Ce^{y_1/\lambda}(y_1/\lambda)^{\alpha+\beta-1}y_1^{\alpha-1}(1-y_2)^{\beta-1},$$

where C is a constant. It follows that Y_1 is $Gamma(\alpha+\beta,\lambda)$ and Y_2 is $Beta(\alpha,\beta)$, and both random variables are independent.

7.5.1 Exercises

Exercício 193. *Let X and Y have joint distribution given by*

$$f(x,y)=\begin{cases}\frac{1}{x^2y^2} & \text{if } x\geq 1, y\geq 1\\ 0 & \text{otherwise}\end{cases}$$

- *Compute the joint density of $U=XY$ and $V=\frac{X}{Y}$*
- *What are the marginal distributions of U and V?*

Exercício 194. *Use the techniques developed in this section to show that if $X\sim Gamma(a,\lambda)$ and $Y\sim Gamma(b,\lambda)$ are independent, then $Z=X+Y\sim Gamma(a+b,\lambda)$. As a consequence, if X and Y are independent, $X\sim Exp(\lambda)$ and $Y\sim Exp(\lambda)$, then $X+Y\sim Gamma(2,\lambda)$. More generally, if $X_1,\dots,X_n$ are independent random variable such that $X_i\sim Exp(\lambda)$, then, $\sum_{i=1}^n X_i\sim Gamma(n,\lambda)$.*

Solution:

$$\begin{aligned}
f_Z(z) &= \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dx && \text{Convolution (Lemma 84)}\\
&= \int_0^z \frac{1}{\Gamma(a)\lambda^a}x^{a-1}e^{-\frac{x}{\lambda}}\cdot\frac{1}{\Gamma(b)\lambda^b}(z-x)^{b-1}e^{-\frac{z-x}{\lambda}}dx && x\geq 0, z-x\geq 0\\
&= \frac{e^{-\frac{z}{\lambda}}}{\Gamma(a)\lambda^a\Gamma(b)\lambda^b}\int_0^z x^{a-1}(z-x)^{b-1}dx\\
&= \frac{e^{-\frac{z}{\lambda}}z^{a+b-2}}{\Gamma(a)\Gamma(b)\lambda^{a+b}}\int_0^z \left(\frac{x}{z}\right)^{a-1}\left(1-\frac{x}{z}\right)^{b-1}dx\\
&= \frac{e^{-\frac{z}{\lambda}}z^{a+b-2}}{\Gamma(a)\Gamma(b)\lambda^{a+b}}\int_0^1 y^{a-1}(1-y)^{b-1}zdy && y=\frac{x}{z}\\
&= \frac{e^{-\frac{z}{\lambda}}z^{a+b-1}}{\Gamma(a+b)\lambda^{a+b}}\int_0^1 \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}y^{a-1}(1-y)^{b-1}dy\\
&= \frac{e^{-\frac{z}{\lambda}}z^{a+b-1}}{\Gamma(a+b)\lambda^{a+b}}\cdot 1 && \text{Beta density integrates to 1}
\end{aligned}$$

Exercício 195. *Let X and Y be i.i.d. exponential with rate 1. Compute the joint density of*

- $U=X$ *and* $V=\frac{X}{Y}$
- $U=X+Y$ *and* $V=\frac{X}{X+Y}$

7.5.2 Joint Moment Generating Functions

Definição 50. *Let $\mathbf{X}=(X_1,\dots,X_d)$ be a random vector. We define the joint moment generating function of $\mathbf{X}$, $M_{\mathbf{X}}:\mathbb{R}^d\longrightarrow\mathbb{R}$, to be $M_{\mathbf{X}}(t_1,\dots,t_d)=\mathbb{E}[e^{\sum_{i=1}^d t_iX_i}]$.*

As in the univariate case, $M_{\mathbf{X}}$ characterizes the distribution of $\mathbf{X}$.

Exemplo 88. *Let X and Y be independent normal random variables with the same mean μ and the same variance*

σ^2. *Let* $Z = X + Y$ *and* $W = X - Y$. *The joint moment generating function of* Z *and* W *is*

$$\begin{aligned}
M_{(Z,W)}(t_1,t_2) &= \mathbb{E}\left[e^{t_1 Z + t_2 W}\right] = \mathbb{E}\left[e^{t_1(X+Y)+t_2(X-Y)}\right] = \mathbb{E}\left[e^{(t_1+t_2)X+(t_1-t_2)Y}\right] \qquad \textit{independence of } X \textit{ and } Y \\
&= \mathbb{E}\left[e^{(t_1+t_2)X}\right]\mathbb{E}\left[e^{(t_1-t_2)Y}\right] = M_X(t_1+t_2)M_Y(t_1-t_2) \\
&= e^{2\mu(t_1+t_2)+\sigma^2(t_1-t_2)^2/2}e^{2\mu(t_1-t_2)+\sigma^2(t_1-t_2)^2/2} \\
&= e^{4\mu(t_1)+\sigma^2 t_1^2}e^{\sigma^2 t_2^2},
\end{aligned}$$

which is the joint moment generating function of two independent normal random variables, the first with mean 2μ *and variance* $2\sigma^2$, *and the second with mean 0 and variance* $2\sigma^2$. *It follows from the fact that the joint mgf characterizes the distribution of a random vector that* $Z \sim N(2\mu, 2\sigma^2)$ *and* $W \sim N(0, 2\sigma^2)$, *and both random variables are independent.*

7.5.3 Exercises

Exercício 196. *If* $\mathbf{X} \sim$ *Multivariate Normal*$(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, *then*

$$M_{\mathbf{X}}(t_1,\ldots,t_d) = e^{\mu'\mathbf{t}+\frac{1}{2}\mathbf{t}'\Sigma\mathbf{t}},$$

where $\mathbf{t} = (t_1,\ldots,t_d)$. *Using this fact, prove that if* Σ *is a diagonal matrix, then all components of* $\mathbf{X}$ *are independent. Also compute the distribution of* $\mathbf{a}\cdot\mathbf{X}$ *when* $\mathbf{a}$ *is a d-dimensional vector (recall that in Exercise 178 you've already computed its mean and variance.)*

8 Convergence

Several notions of convergence are available in probability. We now explore some of them. First, we recall the definition of limit of a numerical sequence.

Definição 51. *Let $(a_n)_{n\in\mathbb{N}}$ be a sequence of numbers, $a_n \in \mathbb{R}$ for every $n \in \mathbb{N}$, and let $a \in \mathbb{R}$ We say* $\lim_{n\longrightarrow\infty} a_n = a$ *if, for every $\xi > 0$, there exists n_0 such that $n > n_0$ implies that $|a_n - a| < \xi$.*

8.1 Convergence in Probability

Definição 52. *A sequence of random variables $(X_n)_{n\in\mathbb{N}}$ converge in probability to a random variable X if for every $\epsilon > 0$,*

$$\lim_{n\longrightarrow\infty} \mathbb{P}(|X_n - X| \geq \epsilon) = 0.$$

This is denoted by

$$X_n \xrightarrow{\mathbb{P}} X.$$

In words, convergence in probability means that for every small number ϵ, for large n the probability that X_n is further from X than ϵ is close to zero.

Notice that convergence in probability means that, for every ϵ, the *numeric* sequence $a_n^\epsilon = \mathbb{P}(|X_n - X| \geq \epsilon)$ converges to zero.

A classical example of convergence in probability is the *weak law of large numbers*:

Teorema 7 (Weak Law of Large Numbers)**.** *Let $X_1, X_2, \dots$ be independent and identically distributed random variables with finite mean μ and variance $\sigma^2 < \infty$. Let $\bar{X}_n$ denote the sample means of the first n random variables. Then, for every $\epsilon > 0$,*

$$\lim_{n\longrightarrow\infty} \mathbb{P}(|\bar{X}_n - \mu| > \epsilon) = 0.$$

That is,

$$\bar{X}_n \xrightarrow{\mathbb{P}} \mu,$$

i.e., $\bar{X}_n$ converges in probability to the constant random variable μ.

Demonstração. By Example 63, for every $\epsilon > 0$,

$$0 \leq \mathbb{P}(|\bar{X} - \mu| > \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}.$$

The Theorem follows from the squeeze theorem and the fact that

$$\frac{\sigma^2}{n\epsilon^2} \xrightarrow{n\longrightarrow\infty} 0.$$

□

Lema 86. *Assume that X_n is a sequence of random variables such that $\mathbb{E}[X_n] \longrightarrow a$ and $Var[X_n] \longrightarrow 0$. Then $\mathbb{E}[X_n] \xrightarrow{\mathbb{P}} a$.*

Demonstração. Fix $\epsilon > 0$. Let n_0 be such that, for every $n > n_0$, $|\mathbb{E}[X_n] - a| < \epsilon/2$. Then, by Chebychev's inequality,

$$\mathbb{P}(|X_n - a| > \epsilon) \leq \mathbb{P}(|X_n - \mathbb{E}[X_n]| > \epsilon/2) \leq 4\frac{Var[X_n]}{\epsilon^2} \longrightarrow 0,$$

which concludes the proof of the Lemma. $\square$

8.2 Almost Sure Convergence

Definição 53. *A sequence of random variables $(X_n)_{n\in\mathbb{N}}$ defined over a domain Ω converge almost surely to a random variable X if*

$$\mathbb{P}(\{\omega \in \Omega : \lim_{n\longrightarrow\infty} X_n(\omega) = X(\omega)\}) = 1$$

This is denoted by

$$X_n \xrightarrow{a.s.} X.$$

In words, convergence in probability means that the set of ω's where the *numeric sequence* $X_1(\omega), X_2(\omega), \ldots$ converges to the number $X(\omega)$ has probability one.

Exemplo 89. *Let $\Omega = (0,1)$, and consider $X_n(\omega) = \omega/n$. For every $\omega \in \Omega$, $\lim_n X_n(\omega) = 0$. It follows that $\mathbb{P}(\{\omega \in \Omega : \lim_{n\longrightarrow\infty} X_n(\omega) = 0\}) = \mathbb{P}(\Omega) = 1$ for every probability measure $\mathbb{P}$ in Ω. That is, X_n converges almost surely to zero for every $\mathbb{P}$.*

Exemplo 90. *Let $\Omega = (0,1)$, and consider $X_n(\omega) = \omega/n\mathbb{I}_{\{\omega<0.5\}} + (1-\omega/n)\mathbb{I}_{\{\omega\geq 0.5\}}$. For every $\omega \in \Omega$, $\lim_n X_n(\omega) = \mathbb{I}_{\{\omega\geq 0.5\}}$. It follows that $\mathbb{P}(\{\omega \in \Omega : \lim_{n\longrightarrow\infty} X_n(\omega) = \mathbb{I}_{\{\omega\geq 0.5\}}\}) = \mathbb{P}(\Omega) = 1$ for every probability measure $\mathbb{P}$ in Ω. That is, X_n converges almost surely to a Bernoulli random variable with parameter $\mathbb{P}(\mathbb{I}_{\{\omega\geq 0.5\}} = 1)$.*

A classical example of almost sure convergence is the *strong law of large numbers*:

Teorema 8 (Strong Law of Large Numbers)**.** *Let $X_1, X_2, \ldots$ be independent and identically distributed random variables with finite mean μ. Let $\bar{X}_n$ denote the sample means of the first n random variables. Then*

$$\bar{X}_n \xrightarrow{a.s.} \mu,$$

i.e., $\bar{X}_n$ converges almost surely to the constant random variable μ.

Demonstração. $\square$

8.3 Convergence In Distribution

Definição 54. *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables* not necessarily defined over the same probability space *and X be another random variable. We say $(X_n)_{n\in\mathbb{N}}$ converges in distribution to X if, for all $x \in \mathbb{R}$ where F_X is continuous,*

$$\lim_{n\longrightarrow\infty} F_{X_n}(x) = F_X(x).$$

This is denoted by

$$X_n \xrightarrow{\mathcal{D}} X.$$

In words, convergence in distribution means that, for large n, probabilities computed with F_X are close to those computed with F_{X_n}. Notice that, for every x where F_X is continuous, it is the *numeric sequence* $(F_{X_n}(x))_{n\in\mathcal{N}}$ that needs to converge.

Exemplo 91. *Let $X_1, X_2, \dots$ be i.i.d.$\sim U(a,b)$. Define $Y_n = \max\{X_1, \dots, X_n\}$. We have that*

$$F_{Y_n}(y) = \mathbb{P}(Y_n \le y) = \mathbb{P}(X_1 \le y, \dots, X_n \le y) = \Pi_{i=1}^n \mathbb{P}(X_i \le y) = \begin{cases} 0 & \text{if } y \le a \\ \left(\frac{y-a}{b-a}\right)^n & \text{if } a < y < b \\ 1 & \text{if } b \le y \end{cases}$$

Now, because $\left(\frac{y-a}{b-a}\right)^n \longrightarrow 0$ when $a < y < b$, it follows that $F_{Y_n}(y) \longrightarrow F(y)$ for every y where F is continuous, where

$$F(y) = \begin{cases} 0 & \text{if } y < b \\ 1 & \text{if } b \le y \end{cases}$$

Hence, Y_n converges in distribution to a degenerate distribution in b (that is, a random variable X such that $\mathbb{P}(X = b) = 1$).

In order to prove that a sequence of random variables converge in distribution to another random variable, it is enough to prove that their moment generating functions converge around a neighborhood of zero:

Lema 87. *Let $(Y_n)_{n\in\mathbb{N}}$ be a sequence of random variables and Y be another random variable. If*

$$M_{Y_n}(t) \xrightarrow{n\longrightarrow\infty} M_Y(t)$$

for all t at some open interval around 0, then

$$\lim_{n\longrightarrow\infty} F_{Y_n}(x) = F_Y(x)$$

for all $x \in \mathbb{R}$ where F_Y is continuous.

A classical example of convergence in distribution is the *central limit theorem*, a key result in statistics:

Teorema 9 (Central Limit Theorem (CLT))**.** *Let $X_1, X_2, \dots$ be independent and identically distributed each having finite mean μ and variance σ^2. Let*

$$Y_n = \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}},$$

and $Y \sim N(0,1)$. Then

$$Y_n \xrightarrow{\mathcal{D}} Y.$$

Demonstração. We will prove the theorem assuming $M_{X_n}(t)$ exists and is finite. Let $M = M_{X_n}$. We assume that $\mu = 0$ and $\sigma^2 = 1$, and leave the proof to the reader for the other cases. We have that $M_{X_n/\sqrt{n}}(t) = M(t/\sqrt{n})$. It follows from Exercise 155 that

$$M_{(\sum_{i=1}^n X_i)/\sqrt{n}}(t) = (M(t/\sqrt{n}))^n. \tag{20}$$

Now, let $L(t) := \log M(t)$. Notice that $L(0) = 0$, $L'(0) = M'(0)/M(0) = \mu = 0$ and $L''(0) = (M(0)M''(0) - (M'(0))^2)/(M(0))^2 = \mathbb{E}[X_n^2] = 1$. Now, notice that

$$\lim_{n\longrightarrow\infty} nL(t/\sqrt{n}) = \lim_{n\longrightarrow\infty} \frac{L'(t/\sqrt{n})n^{-3/2}t}{-2n^{-2}} = \lim_{n\longrightarrow\infty} \frac{L'(t/\sqrt{n})t}{2n^{-1/2}} = \qquad \text{L'Hospital Rule twice}$$
$$\lim_{n\longrightarrow\infty} \frac{L''(t/\sqrt{n})t^2}{2} = \frac{t^2}{2}.$$

It follows that

$$\lim_{n \longrightarrow \infty} M_{(\sum_{i=1}^n X_i)/\sqrt{n}}(t) = \qquad \text{Equation 20}$$
$$\lim_{n \longrightarrow \infty} (M(t/\sqrt{n}))^n = \lim_{n \longrightarrow \infty} \exp(nL(t/\sqrt{n})) = \exp\left(\frac{t^2}{2}\right).$$

The proof is concluded by noticing that $\exp\left(\frac{t^2}{2}\right)$ is the moment generating function of a normal distribution with mean zero and variance 1 (Lemma 65) and by using Lemma 87. □

Exemplo 92. *If $X_1, \ldots, X_n$ with $n = 500$ are i.i.d. with mean $\mu =$ and variance $\sigma^2 = 1$, the central limit theorem states that*

$$\mathbb{P}(\bar{X} < 0.05) = \mathbb{P}\left(\frac{n\bar{X} - n\mu}{\sigma\sqrt{n}} < \frac{n \cdot 0.05 - n\mu}{\sigma\sqrt{n}}\right) \approx \mathbb{P}\left(Z < \frac{n \cdot 0.05 - n\mu}{\sigma\sqrt{n}}\right) = \mathbb{P}(Z < 1.11) \approx 0.86,$$

where $Z \sim N(0, 1)$.

8.4 Relationships Between the Various Convergence Notions

8.4.1 Exercises

Exercício 197. *Let X_i, $i = 1, 2, \ldots$, be random variables such that $\mathbb{P}(X_i = i) = 1 - \mathbb{P}(X_i = 0) = \frac{1}{i}$. X_i converges in probability to some random variable? In case the answer is yes, prove the result. Does the mean $\mathbb{E}[X_i]$ converge? Where to?*

Exercício 198. *Let $X_i \sim Ber(p)$, $0 < p < 1$, $i = 1, 2, \ldots$, be independent random variables. Let $Y_n = X_1 \times \ldots \times X_n$. Compute $\mathbb{P}(Y_n > 0)$ and show that Y_n converges in probability to zero.*

Exercício 199. *Prove that if $Y_n \xrightarrow{a.s.} Y$ and $X_n \xrightarrow{a.s.} X$, where all random variables are defined in the same space, then $X_n Y_n \xrightarrow{a.s.} XY$. Hint: use Exercise* **??**.

Solution: Notice that

$$\mathbb{P}(\{\omega \in \Omega : X_n(\omega) \longrightarrow X(\omega)\}) = \mathbb{P}(\{\omega \in \Omega : Y_n(\omega) \longrightarrow Y(\omega)\}) = 1$$

.

It follows from Exercise **??** that

$$\mathbb{P}(\{\omega \in \Omega : X_n(\omega) \longrightarrow X(\omega)\} \cap \{\omega \in \Omega : Y_n(\omega) \longrightarrow Y(\omega)\}) = 1.$$

The result follows from the fact that

$$\{\omega \in \Omega : X_n(\omega) \longrightarrow X(\omega)\} \cap \{\omega \in \Omega : Y_n(\omega) \longrightarrow Y(\omega)\} \subset \{\omega \in \Omega : X_n(\omega)Y_n(\omega) \longrightarrow X(\omega)Y(\omega)\}$$

and probability is monotonic.

Exercício 200. *Using Lemma 87, prove that if $X_n \sim N(\mu, 1/n)$, $n = 1, 2, \ldots$, then X_n converges in distribution to X such that $\mathbb{P}(X = \mu) = 1$. Prove this using the definition of convergence in distribution as well.*

Exercício 201. *Prove that if $Y_n \xrightarrow{\mathbb{P}} Y$ and $X_n \xrightarrow{\mathbb{P}} X$, then $X_n + Y_n \xrightarrow{\mathbb{P}} X + Y$.*

Solution: For every $\epsilon > 0$,

$$0 \le \Pr\left(|(X_n, Y_n) - (X, Y)| \ge \varepsilon\right) \le \Pr\left(|X_n - X| + |Y_n - Y| \ge \varepsilon\right)$$
$$\le \Pr\left(|X_n - X| \ge \tfrac{\varepsilon}{2}\right) + \Pr\left(|Y_n - Y| \ge \tfrac{\varepsilon}{2}\right) \xrightarrow{n\longrightarrow\infty} 0$$

Exercício 202. *Using the Central Limit Theorem, show that if $X_1, \ldots, X_n$ are i.i.d. Gamma(k, λ) random variables, then, for n large,*

$$P\left(\sqrt{n}(\bar{X} - k\lambda) \le 1\right) \approx \phi\left(\frac{1}{\sqrt{k}\lambda}\right),$$

where ϕ is the cumulative distribution of the standard normal distribution.

Exercício 203. *Consider $\Omega = (0, 1)$ and $\mathbb{P}$ to be the uniform measure, i.e., $\mathbb{P}([a, b]) = b - a$ for every $0 \le a \le b \le 1$. Define the sequence of random variables $(X_n)_{n\ge1}$ by $X_n(\omega) = \mathbb{I}_{\{0\le\omega\le\frac{n+1}{2n}\}}$. Does X_n converge almost surely? If so, where to?*

Exercício 204. *Let $(X_n)_n$ be a sequence of random variables such that $X_n \sim Exp(\lambda)$. Prove that X_n converges in probability to zero.*

Exercício 205. *Let $X_n = X + Y_n$, where $\mathbb{E}[Y_n] = 1/n$ and $Var[Y_n] = 2/n$. Prove that X_n converges in probability to X.*

Exercício 206. *Let $X, X_1, X_2, \ldots$ be independent random variables with $X_n \sim Ber(1/2 + 1/n)$ and $X \sim Ber(1/2)$. Prove that $X_n \xrightarrow{\mathcal{D}} X$, but X_n does not converge in probability to X*

9 Bibliography

Referências

P. Billingsley. *Probability and measure*. John Wiley & Sons, 2008. 14