2021

Luciano Carlos de Lemos

DEMAT/UFOP

Universidade Federal de Ouro Preto
Instituto de Ciências Exatas e Biológicas
Mestrado Profissional em Matemática em Rede Nacional
PROFMAT

Dissertação

Números normais: uma aplicação da lei dos grandes números

Luciano Carlos de Lemos

Ouro Preto
2021

UFOP

UNIVERSIDADE FEDERAL DE OURO PRETO (UFOP)
INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS (ICEB)
DEPARTAMENTO DE MATEMÁTICA (DEMAT)
MESTRADO PROFISSIONAL EM MATEMÁTICA EM REDE NACIONAL (PROFMAT)

LUCIANO CARLOS DE LEMOS

Números Normais: uma aplicação da Lei dos Grandes Números

OURO PRETO - MG, BRASIL
2021

Luciano Carlos de Lemos

Números Normais: uma aplicação da Lei dos Grandes Números

Dissertação de mestrado apresentada como parte dos requisitos para obtenção do título de Mestre pelo programa de Mestrado Profissional em Matemática em Rede Nacional do Departamento de Matemática do Instituto de Ciências Exatas e Biológicas da Universidade Federal de Ouro Preto.

Orientador: Sávio Ribas.

Ouro Preto - MG, Brasil
2021

SISBIN - SISTEMA DE BIBLIOTECAS E INFORMAÇÃO

L556n Lemos, Luciano Carlos de .
Números normais [manuscrito]: uma aplicação da Lei dos Grandes Números. / Luciano Carlos de Lemos. - 2021.
87 f.: il.: color., tab..

Orientador: Prof. Dr. Sávio Ribas.
Dissertação (Mestrado Profissional). Universidade Federal de Ouro Preto. Departamento de Matemática. Programa de Pós-Graduação em Matemática.
Área de Concentração: Matemática com Oferta Nacional.

1. Matemática - Estudo e ensino. 2. Numeração. 3. Probabilidades. I. Ribas, Sávio. II. Universidade Federal de Ouro Preto. III. Título.

CDU 511:374

Bibliotecário(a) Responsável: Flavia Reis - CRB6-2431

MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE OURO PRETO
REITORIA
INSTITUTO DE CIENCIAS EXATAS E BIOLOGICAS
PROGRAMA DE POS-GRADUACAO EM MATEMATICA EM REDE NACIONAL

FOLHA DE APROVAÇÃO

Luciano Carlos de Lemos

Números normais: uma aplicação da Lei dos Grandes Números

Dissertação apresentada ao Programa de Mestrado Profissional em Matemática em Rede Nacional da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de Mestre em Matemática

Aprovada em 23 de novembro de 2021

Membros da banca

Prof. Dr. Sávio Ribas - Orientador Universidade Federal de Ouro Preto
Prof. Dr. Rodrigo Geraldo do Couto - Universidade Federal de Ouro Preto
Prof. Dra. Luana Amaral Gurgel - Faculdade COTEMIG
Prof. Dr. Danilo Vilela Avelar - Universidade Federal Fluminense

Prof. Dr. Sávio Ribas, orientador do trabalho, aprovou a versão final e autorizou seu depósito no Repositório Institucional da UFOP em 10/12/2021.

"Daria tudo que sei, pela metade do que ignoro."
(René Descartes)

Agradecimentos

A **Deus**, pela dádiva da vida e por me permitir realizar tantos sonhos nesta existência. Obrigado por Sua voz "invisível" que não me permitiu desistir, por Sua eterna compreensão e tolerância e principalmente por ter me dado uma família tão especial.

Ao **Prof. Sávio**, pela orientação, competência, profissionalismo e dedicação tão importantes. Tantas vezes que nos reunimos e, embora em algumas eu me apresentasse desestimulado, bastavam alguns minutos de conversa para retomar o ânimo. Obrigado por acreditar em mim e ter aceitado essa empreitada. Tenho certeza que não chegaria neste ponto sem o seu apoio.

Aos membros da banca examinadora, **Prof. Danilo Vilela Avelar, Profª Luana Amaral Gurgel e Prof. Rodrigo Geraldo do Couto**, que tão gentilmente aceitaram participar e colaborar com esta dissertação.

Aos **Professores do PROFMAT-UFOP**, pela dedicação, competência, apoio e todo conhecimento compartilhado.Agradeço também aos meus **colegas do PROFMAT**, por dividir comigo suas experiências, angústias e conhecimentos nas várias listas de exercícios, revisões e discussões sempre muito produtivas.

À minha família, avós, tios(as), primos(as), cunhados(as), sobrinhos(as) e minha sogra por apoiarem e compreenderem o meu distanciamento em vários momentos; em especial à **Denise**, que me incentivou e foi companheira de caminhada.

À minha **mãe Silvina** e ao meu **pai Carlos** deixo um agradecimento especial, por todas as lições de amor, companheirismo, amizade, caridade, dedicação, abnegação e compreensão que vocês me dão todos os dias. Sinto-me orgulhoso e privilegiado por ter pais tão especiais. E ao meu irmão **Leandro**, sempre pronto a me apoiar e incentivar em tudo nesta vida.

À minha amada esposa **Cláudia** e ao meu filho **Gabriel**, por todo amor, incentivo, apoio e compreensão. Obrigado por permanecerem ao meu lado, mesmo sem a atenção devida e depois de tantos momentos de lazer perdidos. Nada disso teria sentido se vocês não existissem na minha vida.

Por fim, a todos aqueles que contribuíram, direta ou indiretamente, para a realização desta dissertação, o meu sincero agradecimento.

Resumo

Neste trabalho, estudamos os números normais e pseudoaleatórios como uma aplicação da Lei dos Grandes Números. Para isso, primeiramente construímos toda a base teórica de probabilidade com o objetivo de fundamentar a Lei dos Grandes Números, um teorema da teoria da probabilidade que estabelece, sob certas hipóteses, que a média aritmética dos resultados observados através da realização da mesma experiência muitas vezes aproxima-se da esperança da variável aleatória. Em seguida, vemos que um problema fundamental na Matemática, na Estatística e na Computação é a geração de números aleatórios, que possuem diversas aplicações práticas, e ainda, que os melhores candidatos a números aleatórios são os números absolutamente normais, mas ainda não os conhecemos. Sendo assim, devemos nos contentar com números pseudoaleatórios, que são números "aparentemente" aleatórios, mas que são gerados de uma forma determinística. Na sequência apresentamos o problema de Monty-Hall, também conhecido popularmente como problema dos bodes, que foi proposto pela primeira vez em um programa de auditório nos EUA. Através de simulações e usando a Lei dos Grandes Números conseguimos conjecturar um resultado correto para esse problema. Atividades didáticas envolvendo tais resultados também foram abordadas e sua aplicação em sala de aula, com estudantes de Ensino Médio de uma escola de Manhumirim-MG e também de graduação em Matemática da UFOP, são descritas. Os dados foram coletados por meio de observações durante a aplicação da atividade e através das respostas dadas aos questionários. Os resultados mostram que os estudantes podem, em muitos casos, ter uma visão intuitiva do problema que os conduza ao erro na escolha da melhor estratégia de resolução. Daí percebemos que atividades como essas podem contribuir significativamente para o processo de aprendizagem dos estudantes.

Palavras-chave: Probabilidade; Lei dos Grandes Números; Números Normais; Números Aleatórios; Números Pseudoaleatórios; Problema de Monty-Hall.

Abstract

In this work, we study the normal and pseudorandom numbers as an application of the Law of Large Numbers. To do this, we first construct the entire theoretical basis of probability in order to substantiate the Law of Large Numbers, a theorem of probability theory that establishes, under certain hypotheses, that the arithmetic mean of the observed results after performing the same experiment many times approaches the expectation of the random variable. Next, we see that a fundamental problem in Mathematics, Statistics and Computing is the generation of random numbers, which has many practical applications, and further that the best candidates for random numbers are the absolutely normal numbers, but we do not know them yet. Therefore, we must be content with pseudorandom numbers, which are "apparently" random numbers, but which are generated in a deterministic way. Next, we present the Monty-Hall problem, also popularly known as the goat problem, which was first proposed in a talk show in the USA. Through simulations and using the Law of Large Numbers we were able to conjecture a correct result for this problem. Didactic activities involving such results have also been approached and their applications in the classroom, with high school students from a school in Manhumirim-MG and also with undergraduate students in Mathematics at UFOP, are described. Data were collected through observations during the application of the activity and through the answers given to the questionnaires. The results show that students can, in several cases, have an intuitive view of the problem that leads them to error in choosing the best solution strategy. Hence, we realized that activities like these can significantly contribute to the students' learning process.

Keywords: Probability; Law of Large Numbers; Normal Numbers; Random Numbers; Pseudorandom Numbers; Monty-Hall Problem.

Lista de Figuras

Lista de Tabelas

Sumário

Introdução

A Teoria da Probabilidade é um ramo bastante conhecido da Matemática e que possui diversas interpretações, mesmo para leigos em Matemática. Frequentemente, vemos na televisão notícias envolvendo probabilidades, como em assuntos relacionados a pesquisas eleitorais, previsão do tempo, ou mesmo em assuntos esportivos. Alguns indícios alegam que o surgimento da teoria das probabilidades teve início com os jogos de azar disseminados na Idade Média. Esse tipo de jogo é comumente praticado através de apostas, na ocasião também era utilizado no intuito de antecipar o futuro. Na verdade, a teoria da probabilidade não prevê o futuro e sim tenta obter informações sobre o futuro, com certa chance de ocorrer, baseado em eventos passados. Tal teoria se sustenta rigorosamente através dos Axiomas de Kolmogorov, que serão vistos no Capítulo 1. As bases da teoria das probabilidades e da análise combinatória foram estabelecidas por Pascal e Fermat no século XVII e aplicadas a situações relacionadas a apostas em jogos. Posteriormente, Bernoulli abordou a importância da Lei dos Grandes Números. Atualmente, a probabilidade é bastante usada também em Estatística [11].

O objetivo deste trabalho é apresentar algumas versões da Lei dos Grandes Números e, como aplicação, veremos os números normais e pseudoaleatórios. Para isso, temos que construir toda a base teórica anterior. Em seguida, vamos usar a geração de números pseudoaleatórios para simular o problema de Monty-Hall um número grande de vezes. Usando a Lei dos Grandes Números, vamos inferir uma resposta para esse problema. Vamos também apresentar essas aplicações em sala de aula para estudantes do ensino médio e superior.

No Capítulo 1, vamos definir formalmente um espaço de probabilidade. Para isso, em um "conjunto universo" que chamaremos de espaço amostral, vamos definir o conjunto de eventos através da σ-álgebra, e vamos precisar de uma função de probabilidade, que atribui um valor entre 0 e 1 a todo evento. Essa função de probabilidade satisfaz os axiomas de Kolmogorov e dela podemos extrair diversas propriedades. Vamos também definir probabilidade condicional e independência, além de apresentar alguns teoremas clássicos da probabilidade, como o Teorema da Multiplicação e o Teorema de Bayes.

No Capítulo 2, vamos apresentar o conceito de variáveis aleatórias, que são funções que determinam um característico numérico do resultado de um experimento. Vamos apresentar também a função de distribuição de variáveis aleatórias (tanto discretas quanto contínuas) e suas propriedades, além dos principais modelos de distribuição.

No Capítulo 3, definimos a esperança e a variância das variáveis aleatórias. Também introduzimos o conceito de momentos, além das desigualdades clássicas de Markov e Tchebychev.

No Capítulo 4, vamos definir dois tipos de convergência de sequências de variáveis aleatórias: convergência em probabilidade e convergência quase certa. A Lei Fraca diz respeito à convergência em probabilidade e é provada usando a desigualdade de Tchebychev com a hipótese das variáveis aleatórias serem não-correlacionadas (que é uma hipótese mais fraca que usar independência), enquanto a Lei Forte diz respeito à convergência quase certa. Daremos duas versões da Lei Forte, que são provadas usando o Lema de Borel-Cantelli e outros resultados mais profundos, supondo que as variáveis são independentes (em uma das versões necessitamos também que elas sejam identicamente distribuídas). A Lei Forte necessita de hipóteses mais restritivas que a Lei Fraca.

No Capítulo 5, finalmente vamos apresentar os números normais em uma base b e os números absolutamente normais, que são fortes candidatos a números aleatórios. Vamos mostrar que, embora tenham medida total (probabilidade 1), números absolutamente normais são de certa forma desconhecidos. Por isso, o problema de geração de números aleatórios é tratado somente com números pseudoaleatórios, isto é, determinísticos porém se comportam como aleatórios para a maioria das aplicações. Além disso, vamos apresentar o Problema de Monty-Hall e usar a Lei dos Grandes Números juntamente com um número gerado pseudo-aleatoriamente para obter uma possível resposta para o problema.

No Capítulo 6, exibimos dois problemas relacionados à Lei dos Grandes Números e à geração de números aleatórios, para aplicação em sala de aula. Aplicamos uma atividade em relação a esses problemas para estudantes de ensino médio e de ensino superior em Matemática. Nessa parte, tivemos a aprovação do projeto pela Comissão de Ética em Pesquisa da UFOP.

No Apêndice A, apresentamos outros tipos de integral (a saber: a Integral de Lebesgue e a de Stieltjes), suas vantagens em relação à Integral de Riemann e suas propriedades, que são usadas ao longo de todo o texto a partir do momento em que falamos sobre esperança de variáveis aleatórias.

No Apêndice B, calculamos a integral de uma função gaussiana.

Capítulo 1

Preliminares sobre Probabilidade

1.1 Definição de Probabilidade

Definição 1.1.1. *Seja Ω um conjunto não-vazio. Uma classe $\mathcal{A}$ de subconjuntos de Ω é chamada* σ-álgebra, *quando satisfaz as seguintes propriedades*

1. $\varnothing \in \mathcal{A}$,
2. *Se* $A \in \mathcal{A}$, *então* $A^c \in \mathcal{A}$,
3. *Se* $A_1, A_2, ... \in \mathcal{A}$, *então* $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$ *e* $\bigcap_{i=1}^{\infty} A_i \in \mathcal{A}$.

Exemplo 1.1.2. $\mathcal{A}_1 = \{\varnothing, \Omega\}$ *é uma σ-álgebra em Ω chamada de* σ-álgebra trivial. *O conjunto das partes* $\mathcal{A}_2 = \{A \subset \Omega\}$ *é a* σ-álgebra das partes de Ω.

Definição 1.1.3. Espaço amostral *é o conjunto Ω sobre o qual está definida a σ-álgebra.*

Definição 1.1.4. Evento *é um elemento (que é um conjunto) da σ-álgebra. Todo subconjunto $A \in \mathcal{A}$ será chamado evento. Por exemplo, Ω é o evento certo, $\varnothing$ o evento impossível. Um* evento aleatório *é aquele em que não há como prever o resultado.*

Exemplo 1.1.5. *Jogar um dado honesto e observar o número da face voltada para cima. Temos que o espaço amostral é $\Omega = \{1, 2, 3, 4, 5, 6\}$ e a σ-álgebra pode ser definida como a σ-álgebra das partes de Ω. O evento elementar "sair o número 2", significa o evento $A_1 = \{2\}$. "Sair número maior que 2", é um evento que consiste de 4 resultados $\{3, 4, 5, 6\}$.*

Definição 1.1.6. *A* σ-álgebra de Borel, *definida sobre $\Omega = [0, 1]$, é a menor σ-álgebra que contém todos os intervalos abertos, ou seja, um elemento de $\mathcal{B}$ é um conjunto formado por união enumerável, interseção enumerável e diferença de intervalos. Observemos que o complementar de conjuntos abertos é fechado e vice-versa, com isso não há perda de generalidade ao supor que* *os intervalos abertos são apenas intervalos. Um elemento da σ-álgebra de Borel é chamado de* boreliano.

Notação: $\mathcal{B}_{[0,1]} = \{A \subset [0,1] : A \; boreliano\}$ *é a σ-álgebra de Borel em* $[0,1]$.

A seguir, vamos definir probabilidade de acordo com os Axiomas de Kolmogorov.

Axioma 1.1.7. *Seja $\mathcal{P} : \mathcal{A} \to \mathbb{R}$.*

1. $\mathcal{P}(A) \geq 0$.

2. $\mathcal{P}(\Omega) = 1$.

3. *(Aditividade finita). Se $A_1, ..., A_n \in \mathcal{A}$ são disjuntos 2 a 2, então*

$$\mathcal{P}\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} \mathcal{P}(A_k).$$

3′. *(σ-aditividade). Se $A_1, A_2, ... \in \mathcal{A}$ são disjuntos 2 a 2, então*

$$\mathcal{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathcal{P}(A_n).$$

É possível provar, mas não o faremos, algumas relações entre esses axiomas. Por exemplo:

Proposição 1.1.8. *O Axioma 3′ implica o Axioma 3, isto é, se $\mathcal{P}$ é σ-aditiva, então é finitamente aditiva.*

Definição 1.1.9. *Uma função $\mathcal{P}$ definida numa σ-álgebra $\mathcal{A}$ e satisfazendo os Axiomas 1, 2 e 3′ chama-se uma* medida de probabilidade *em $\mathcal{A}$ ou simplesmente uma* probabilidade *em $\mathcal{A}$.*

Definição 1.1.10. *Considere a sequência de conjuntos $(A_n)_{n\geq 1}$. Dizemos que tal sequência é*

i) *crescente se $A_1 \subset A_2 \subset A_3 \subset \ldots$. Nesse caso, definimos $\lim A_n = \bigcup_{k\geq 1} A_k$ e dizemos que $(A_n)_n$ cresce para $\lim A_n$. Notação: $A_k \uparrow \lim A_n$.*

ii) *decrescente se $A_1 \supset A_2 \supset A_3 \supset \ldots$. Nesse caso, definimos $\lim A_n = \bigcap_{k\geq 1} A_k$ e dizemos que $(A_n)_n$ decresce para $\lim A_n$. Notação: $A_k \downarrow \lim A_n$.*

Definição 1.1.11. *a) Se a sequência de números reais $(a_n)_{n\geq 1}$ é crescente e $\lim a_n = a$, dizemos que a_n cresce para a e escrevemos $a_n \uparrow a$.*

b) Se a sequência de números reais $(a_n)_{n\geq 1}$ é decrescente e $\lim a_n = a$, dizemos que a_n decresce para a e escrevemos $a_n \downarrow a$.

Proposição 1.1.12. *Dados os Axiomas 1, 2, 3, o Axioma 3′ é equivalente à validade da continuidade no vazio, isto é, é equivalente à validade da seguinte afirmação: se a sequência* $(A_n)_{n\geq 1}$*, onde* $A_n \in \mathcal{A}\ \forall n$ *decrescer para o vazio então* $\mathcal{P}(A_n) \downarrow 0$ *quando* $n \to \infty$*. Com isso, podemos concluir que uma probabilidade finitamente aditiva é uma probabilidade se, e só se, é contínua no vazio.*

Dados os Axiomas de Kolmogorov e suas equivalências, podemos extrair diversas propriedades.

Proposição 1.1.13 (Propriedades de probabilidade)**.** *Seja* $\mathcal{P}$ *uma probabilidade em uma* σ*-álgebra* $\mathcal{A}$ *definida sobre* Ω*. Então, para todo* $A, A_1, A_2, A_3, \dots \in \mathcal{A}$*, valem as seguintes propriedades:*

P1. $\mathcal{P}(A^c) = 1 - \mathcal{P}(A)$*. Em particular:* $\mathcal{P}(\varnothing) = 1 - \mathcal{P}(\Omega) = 0$.

P2. $0 \leq \mathcal{P}(A) \leq 1$.

P3. Se $A_1 \subset A_2$ *então* $\mathcal{P}(A_1) \leq \mathcal{P}(A_2)$.

P4. $\mathcal{P}\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \mathcal{P}(A_i)$.

P5. $\mathcal{P}\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mathcal{P}(A_i)$.

P6. (Continuidade de probabilidade). Se $A_n \uparrow A$*, então* $\mathcal{P}(A_n) \uparrow \mathcal{P}(A)$*. Se* $A_n \downarrow A$*, então* $\mathcal{P}(A_n) \downarrow \mathcal{P}(A)$.

Demonstração. Vejamos:

$P1$. Segue dos Axiomas 2 e 3.

$P2$. Isso segue do Axioma 1 e de P1.

$P3$. Pela aditividade finita, $\mathcal{P}(A_2) = \mathcal{P}(A_1) + \mathcal{P}(A_2 - A_1) \geq \mathcal{P}(A_1)$, pelo Axioma 1.

$P4$. Pela aditividade finita, $\mathcal{P}(A_1 \cup A_2) = \mathcal{P}(A_1) + \mathcal{P}(A_2 \cap A_1^c) \leq \mathcal{P}(A_1) + \mathcal{P}(A_2)$, por P3, já que $A_2 \cap A_1^c \subset A_2$. Completa-se a prova por indução.

$P5$. Defina $B_1 = A_1, B_2 = A_2 - B_1, B_3 = A_3 - (B_1 \cup B_2), \dots, B_n = A_n - (B_1 \cup \cdots \cup B_{n-1})$. Temos que os B_n's são disjuntos 2 a 2, $\bigcup B_n = \bigcup A_n$ e $B_n \subset A_n$. Logo, pelo Axioma 3′ e por P3 temos que $\mathcal{P}(\bigcup A_n) = \mathcal{P}(\bigcup B_n) = \sum \mathcal{P}(B_n) \leq \sum \mathcal{P}(A_n)$.

$P6$. Supondo que $A_n \downarrow A$, ou seja, que $A_{n+1} \subset A_n \forall n$ e $\bigcap\limits_{n\geq 1} A_n = A$. Assim, $\mathcal{P}(A_n) \geq \mathcal{P}(A_{n+1})$, por $P3$, e $(A_n - A) \downarrow \varnothing \Rightarrow \mathcal{P}(A_n - A) \to 0$, pela continuidade no vazio. A aditividade finita implica $\mathcal{P}(A_n - A) = \mathcal{P}(A_n) - \mathcal{P}(A)$, pois $A \subset A_n$. Portanto, temos $\mathcal{P}(A_n) - \mathcal{P}(A) \to 0$ e $\{\mathcal{P}(A_n)\}_{n\geq 1}$ é decrescente, logo $\mathcal{P}(A_n) \downarrow \mathcal{P}(A)$.

Se $A_n \uparrow A$ (ou seja, $A_n \subset A_{n+1}, \forall n$ e $\bigcup\limits_{n\geq 1} A_n = A$), então $A_n^c \downarrow A^c$. Logo $\mathcal{P}(A_n^c) \downarrow \mathcal{P}(A^c)$, ou seja, $1 - \mathcal{P}(A_n) \downarrow 1 - \mathcal{P}(A)$, portanto $\mathcal{P}(A_n) \uparrow \mathcal{P}(A)$.

□

Proposição 1.1.14. *i) Se* $\mathcal{P}(A_n) = 0$ *para* $n \in \mathbb{N}$, *então* $\mathcal{P}\left(\bigcup\limits_{n=1}^{\infty} A_n\right) = 0$.

ii) Se $\mathcal{P}(A_n) = 1$ *para* $n \in \mathbb{N}$, *então* $\mathcal{P}\left(\bigcap\limits_{n=1}^{\infty} A_n\right) = 1$.

Demonstração. i) Pela propriedade P5, temos que

$$\mathcal{P}\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mathcal{P}(A_n) = \sum_{n=1}^{\infty} 0 = 0.$$

ii) Como $\mathcal{P}(A_n^c) = 0$ segue que $\mathcal{P}\left(\bigcup\limits_{n=1}^{\infty} A_n^c\right) = 0$ pelo item anterior. E, como

$$\bigcap_{n=1}^{\infty} A_n = \Omega - \bigcup_{n=1}^{\infty} A_n^c, \text{ temos que } \mathcal{P}\left(\bigcap_{n=1}^{\infty} A_n\right) = \mathcal{P}(\Omega) - \mathcal{P}\left(\bigcup_{n=1}^{\infty} A_n^c\right) = 1 - 0 = 1.$$

□

Definição 1.1.15. *Um* espaço de probabilidade *é um trio* $(\Omega, \mathcal{A}, \mathcal{P})$, *formado por*

(a) um conjunto Ω *não vazio;*

(b) uma σ*-álgebra* $\mathcal{A}$ *de subconjuntos de* Ω*;*

(c) uma probabilidade $\mathcal{P}$ *em* $\mathcal{A}$.

A partir de agora, a menos que explicitemos o contrário, todos os espaços de probabilidade serão dados pela tripla $(\Omega, \mathcal{A}, \mathcal{P})$, e os eventos serão eventos aleatórios pertencentes a $\mathcal{A}$.

1.2 Probabilidade Condicional e Independência

Nessa seção, vamos definir probabilidade condicional e com isso definir independência. Vamos apresentar também alguns resultados básicos da probabilidade.

Definição 1.2.1. *Se $B \in \mathcal{A}$ e $\mathcal{P}(B) > 0$, a probabilidade condicional de $A \in \mathcal{A}$ dado B, é definida por*

$$\mathcal{P}(A \mid B) = \frac{\mathcal{P}(A \cap B)}{\mathcal{P}(B)}.$$

Teorema 1.2.2 (Teorema da Multiplicação ou Teorema da Probabilidade Composta)**.** *São válidos:*

(i) $\mathcal{P}(A \cap B) = \mathcal{P}(A)\mathcal{P}(B \mid A) = \mathcal{P}(B)\mathcal{P}(A \mid B), \forall A, B \in \mathcal{A}$;

(ii) $\mathcal{P}(A_1 \cap A_2 \cap \cdots \cap A_n) = \mathcal{P}(A_1)\mathcal{P}(A_2 \mid A_1)\mathcal{P}(A_3 \mid A_1 \cap A_2)\ldots\mathcal{P}(A_n \mid A_1 \cap \cdots \cap A_{n-1})$, $\forall A_n \in \mathcal{A}$ *com* $n \in \mathbb{N}$.

Demonstração. O item (i) segue da definição e o item (ii) segue por indução. □

Exemplo 1.2.3. *Um lote contém 10 peças, sendo 7 boas e 3 defeituosas. Retiramos três peças, ao acaso e sem reposição, para inspeção. Qual a probabilidade de se obter três peças defeituosas?*

Seja D_i o evento "retirar uma peça defeituosa na i-ésima extração". Então (com D = "retirar 3 peças defeituosas") temos:

$$\mathcal{P}(D) = \mathcal{P}(D_1 \cap D_2 \cap D_3) = \mathcal{P}(D_1)\mathcal{P}(D_2 \mid D_1)\mathcal{P}(D_3 \mid D_1 \cap D_2) = \frac{3}{10} \cdot \frac{2}{9} \cdot \frac{1}{8} = \frac{1}{120}$$

Teorema 1.2.4 (Teorema da Probabilidade Total)**.** *Se a sequência (finita ou enumerável) de eventos aleatórios $A_1, A_2, \ldots$ formar uma partição de Ω, isto é, se os A_n's forem disjuntos 2 a 2 e $\bigcup A_n = \Omega$, então*

$$\mathcal{P}(B) = \sum_n \mathcal{P}(A_n)\mathcal{P}(B \mid A_n), \forall B \in \mathcal{A}.$$

Teorema 1.2.5 (Teorema de Bayes)**.** *Nas mesmas hipóteses do teorema anterior, vale*

$$\mathcal{P}(A_i \mid B) = \frac{\mathcal{P}(A_i)\mathcal{P}(B \mid A_i)}{\sum_j \mathcal{P}(A_j)\mathcal{P}(B \mid A_j)}.$$

Demonstração. Da definição de probabilidade condicional temos

$$\mathcal{P}(A_i \mid B) = \frac{\mathcal{P}(A_i \cap B)}{\mathcal{P}(B)}.$$

O numerador dessa expressão pode ser reescrito pela regra do produto, condicionado à A_n, isto é,

$$\mathcal{P}(A_i \cap B) = \mathcal{P}(B \cap A_i) = \mathcal{P}(B \mid A_i)\mathcal{P}(A_i).$$

Para completar a demonstração, basta notar que

$$\mathcal{P}(B) = \sum_j \mathcal{P}(B \cap A_j) = \sum_j \mathcal{P}(B \mid A_j)\mathcal{P}(A_j).$$

□

Exemplo 1.2.6. *Suponha que um jogador participa de um torneio de tênis onde sua probabilidade de vitória é 0,3 contra metade dos jogadores (chame-os de grupo 1), 0,4 contra um quarto do jogadores (chame-os de grupo 2) e 0,5 contra o um quarto dos jogadores restantes (chame-os de grupo 3). O jogador disputa uma partida contra um oponente selecionado aleatoriamente. Supondo que o jogador venceu a partida disputada, qual a probabilidade dele ter jogado contra um adversário do grupo 1?*

Considere que A_j é o evento "ter um adversário do grupo j". Temos $\mathcal{P}(A_1) = 0,5$; $\mathcal{P}(A_2) = 0,25$; $\mathcal{P}(A_3) = 0,25$. Além disso, se B é o evento "vencer uma partida" então $\mathcal{P}(B \mid A_1) = 0,3$; $\mathcal{P}(B \mid A_2) = 0,4$; $\mathcal{P}(B \mid A_3) = 0,5$. Usando o Teorema de Bayes, temos que

$$\mathcal{P}(A_1 \mid B) = \frac{\mathcal{P}(A_1)\mathcal{P}(B \mid A_1)}{\mathcal{P}(A_1)\mathcal{P}(B \mid A_1) + \mathcal{P}(A_2)\mathcal{P}(B \mid A_2) + \mathcal{P}(A_3)\mathcal{P}(B \mid A_3)}$$

$$= \frac{0,5 \cdot 0,3}{0,5 \cdot 0,3 + 0,25 \cdot 0,4 + 0,25 \cdot 0,5} = \frac{0,15}{0,375} = 0,4.$$

Ou seja, a probabilidade do jogador ter disputado uma partida contra o adversário do grupo 1, dado que ele venceu a partida é de 40%.

Definição 1.2.7. *Os eventos aleatórios A e B são independentes se*

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

Segue da definição de probabilidade condicional e de independência que se A e B são independentes com $\mathcal{P}(B) > 0$ então $\mathcal{P}(A \mid B) = \mathcal{P}(A)$, isto é, a probabilidade de A ocorrer não depende da ocorrência de B.

Eventos de probabilidade zero ou um são independentes de qualquer outro. De fato, se $\mathcal{P}(A) = 0$, então $\mathcal{P}(A \cap B) = 0$ pois $A \cap B \subset A$ e $\mathcal{P}(A) \cdot \mathcal{P}(B) = 0$, logo A e B são independentes $\forall B \in \mathcal{A}$. Se $\mathcal{P}(B) = 1$, então $\mathcal{P}(A \cap B) = \mathcal{P}(A - (A \cap B^c)) = \mathcal{P}(A) - \mathcal{P}(A \cap B^c)$ e, como $A \cap B^c \subset B^c$ implica $\mathcal{P}(A \cap B^c) \leq \mathcal{P}(B^c) = 0$, temos $\mathcal{P}(A \cap B^c) = 0$ e $\mathcal{P}(A \cap B) = \mathcal{P}(A) = \mathcal{P}(A) \cdot (B)$. Logo A e B são independentes $\forall B \in \mathcal{A}$.

Proposição 1.2.8. *O evento A é independente de si mesmo se, e só se, $\mathcal{P}(A) = 0$ ou 1.*

Demonstração. $\mathcal{P}(A) = \mathcal{P}(A \cap A) = \mathcal{P}(A) \cdot \mathcal{P}(A) \Leftrightarrow \mathcal{P}(A) = 0$ ou 1. □

Proposição 1.2.9. *Se A e B são independentes, então A e B^c também são independentes, assim como A^c e B e ainda A^c e B^c.*

Demonstração. Sejam A e B independentes. Então $\mathcal{P}(A \cap B^c) = \mathcal{P}(A - (A \cap B)) = \mathcal{P}(A) - \mathcal{P}(A \cap B) = \mathcal{P}(A) - \mathcal{P}(A)\mathcal{P}(B) = \mathcal{P}(A)(1 - \mathcal{P}(B)) = \mathcal{P}(A) \cdot \mathcal{P}(B^c)$. Análogo para os demais casos. □

Exemplo 1.2.10. *Seja* $\Omega = \{1, 2, 3, 4, 5, 6\}$ *o espaço amostral ao jogar um dado uma vez e seja* $\mathcal{A}$ *sua* σ*-álgebra das partes. Os eventos*

$$A = \{\textit{sair um número primo}\} \ \ \textit{e} \ \ B = \{\textit{sair um número par menor que } 5\}$$

são independentes, pois $\mathcal{P}(A) = \frac{3}{6} = \frac{1}{2}$ *(uma vez que 2, 3, 5 são primos),* $\mathcal{P}(B) = \frac{2}{6} = \frac{1}{3}$ *(uma vez que 2, 4 são pares menores que 5) e* $\mathcal{P}(A \cap B) = \frac{1}{6} = \mathcal{P}(A) \cdot \mathcal{P}(B)$ *(uma vez que 2 é o único primo par menor que 5). Por outro lado, o evento* $C = \{\textit{sair um número par}\}$ *não é independente de A, pois* $\mathcal{P}(C) = \frac{1}{2}$*,* $\mathcal{P}(A \cap C) = \frac{1}{6} \neq \frac{1}{2} \cdot \frac{1}{2} = \mathcal{P}(A) \cdot \mathcal{P}(C)$*.*

A seguir, vamos generalizar o conceito de independência para mais eventos.

Definição 1.2.11. *(a) Sejam* I *um conjunto de índices e* A_i *eventos aleatórios, com* $i \in I$*. Dizemos que os* A_i*'s são* independentes 2 a 2 *se* $\mathcal{P}(A_i \cap A_j) = \mathcal{P}(A_i)\mathcal{P}(A_j)$ $\forall i, j \in I$*,* $i \neq j$*.*

(b) Seja $n \geq 2$ *um inteiro. Dizemos que os eventos* $A_1, \ldots, A_n$ *são* coletivamente independentes *(ou simplesmente* independentes*) se*

$$\mathcal{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_m}) = \mathcal{P}(A_{i_1})\mathcal{P}(A_{i_2}) \ldots \mathcal{P}(A_{i_m})$$

$\forall 1 \leq i_1 < i_2 < \cdots < i_m \leq n, \forall m = 2, 3, \ldots, n.$

(c) Dizemos que os eventos da família infinita enumerável $A_1, A_2, \ldots$ *são* independentes *se* $\forall n \geq 2$*,* $A_1, A_2, \ldots, A_n$ *são independentes, isto é, se são independentes quando tomamos apenas os* n *primeiros termos da sequência, para todo* n*.*

(d) Seja I *um conjunto (enumerável ou não) de índices com pelo menos dois elementos. Dizemos que os eventos* A_i *dessa família, com* $i \in I$*, são* independentes *se toda subfamília finita deles é de eventos independentes, isto é, se* $A_{i_1}, A_{i_2}, A_{i_3}, \ldots, A_{i_m}$ *são independentes para todo subconjunto* $\{i_1, \ldots, i_m\} \subset I$ *com pelo menos dois elementos.*

É possível mostrar que independência 2 a 2 não implica independência coletiva (ver [6, Exemplo 7]).

Proposição 1.2.12. *Se os eventos* A_i*,* $i \in I$*, são independentes, então os eventos* B_i*,* $i \in I$*, também são independentes para cada* $B_i \in \{A_i, A_i^c\}$*.*

Não vamos demonstrar essa proposição, mas a prova pode ser encontrada em [6, Proposição 1.7]. A ideia é passar para o caso finito e usar a mesma ideia da Proposição 1.2.9.

Capítulo 2

Distribuição de variáveis aleatórias

2.1 Variáveis aleatórias

De maneira informal, uma variável aleatória determina um característico numérico do resultado de um experimento, isto é, é a função que "transfere" a ideia de que Ω é um conjunto qualquer e passa a considerá-lo como um conjunto de números reais.

Observação: Por $[X \leq x]$ vamos entender o conjunto $\{\omega \in \Omega; X(\omega) \leq x\}$.

Definição 2.1.1. *Uma* variável aleatória X *é uma função* $X : \Omega \to \mathbb{R}$ *tal que* $[X \leq x]$ *é evento aleatório para todo* $x \in \mathbb{R}$*, isto é,* X *é variável aleatória se* $[X \leq x] \in \mathcal{A}, \forall x \in \mathbb{R}$.

A seguir veremos alguns exemplos:

Exemplo 2.1.2. *Ao lançar uma moeda n vezes e observar a sequência de caras* (c) *e coroas* (k) *obtidas, os resultados possíveis são sequências de tamanho n de caras e coroas e podemos definir*

$$\Omega = \{\omega = (\omega_1, \ldots, \omega_n)\,;\ \omega_i \in \{c, k\},\ i = 1, \ldots, n\}$$
$$\textit{e } X(\omega) = \textit{número de } c's \textit{ em } \omega = \#\{i : \omega_i = c, 1 \leq i \leq n\}.$$

Com isso, o número de caras obtidas nos n lançamentos é um característico numérico da sequência de caras e coroas.

Exemplo 2.1.3. *Escolher um ponto ao acaso em* $[0, 1]$*, ou seja,* $\omega \in [0, 1]$*. Seja* X *o valor do ponto escolhido. Então*

$$\Omega = [0, 1], X(\omega) = \omega.$$

Exemplo 2.1.4. *Escolher um ponto ao acaso no quadrado unitário, ou seja, escolher um par ordenado* $(x, y) \in [0, 1] \times [0, 1]$*. Então* $\Omega = [0, 1] \times [0, 1]$ *e sendo a variável aleatória* X *o valor do produto das duas coordenadas, temos* $X(\omega) = xy$*, para todo* $\omega = (x, y) \in \Omega$.

Definição 2.1.5. *Seja X uma variável aleatória. Definimos a* função de distribuição *como*

$$F_X(x) = \mathcal{P}\,(X \leq x),\, x \in \mathbb{R}.$$

A função de distribuição descreve como probabilidades são associadas aos valores ou aos intervalos de valores de uma variável aleatória. Ela representa a probabilidade de uma variável aleatória ser menor ou igual a um valor real x. Desta forma, podemos definir variáveis aleatórias através da sua função de distribuição.

Exemplo 2.1.6. *Uma população de 1000 pessoas foi analisada num estudo para determinar a efetividade de uma vacina contra a COVID-19. No estudo, as pessoas recebiam uma dose da vacina e, após um mês, passavam por um novo teste. Caso ainda não tivessem completamente imunizadas, recebiam outra dose da vacina. Ao fim de 5 doses todas as pessoas foram consideradas imunizadas. Os resultados completos estão na tabela a seguir.*

Tabela 2.1: Imunização por doses

Doses	Freq.
1	245
2	288
3	256
4	145
5	66

Suponha que uma pessoa dessa população é sorteada ao acaso. Qual será a probabilidade dela ter ficado imunizada ao receber até 2 doses? O que precisamos obter é a função de distribuição no ponto 2, ou seja, calculamos a probabilidade acumulada de ocorrência de valores menores ou iguais a 2. Assim,

$$F\,(2) = \mathcal{P}\,(X \leq 2) = \mathcal{P}\,(X = 1) + \mathcal{P}\,(X = 2) = 0,245 + 0,288 = 0,533.$$

Logo, a probabilidade da pessoa sorteada ter ficado imunizada com até duas doses é de $53,3\%$.

Na sequência, vamos estudar propriedades da função de distribuição.

Proposição 2.1.7 (Propriedades). *A função de distribuição F de uma variável aleatória X tem três propriedades básicas:*

F1. *F é não decrescente.*

F2. *F é contínua à direita.*

F3. $0 \leq F\,(x) \leq 1$, $\lim_{x\to-\infty} F\,(x) = 0$ *e* $\lim_{x\to\infty} F\,(x) = 1$.

Demonstração.

$F1$. Se $x \leq y \Rightarrow [X \leq x] \subset [X \leq y] \Rightarrow F(x) = \mathcal{P}(X \leq x) \leq \mathcal{P}(X \leq y) = F(y)$.

$F2$. Se $x_n \downarrow x$, então $F(x_n) \downarrow F(x)$, isto é, $[X \leq x_n] \downarrow [X \leq x] \Rightarrow F(x_n) = \mathcal{P}(X \leq x_n) \downarrow \mathcal{P}(X \leq x) = F(x)$.

$F3$. Se $x \to -\infty$, então $[X \leq x] \downarrow \varnothing$, e assim $F(x) = \mathcal{P}(X \leq x) \downarrow 0$. Se $x \to +\infty$, então $[X \leq x] \uparrow \Omega$, e assim $F(x) = \mathcal{P}(X \leq x) \uparrow 1$. □

Uma função de distribuição é monótona não-decrescente e portanto tem um número finito ou enumerável de pontos de descontinuidade. Além disto, todas as descontinuidades são do tipo salto. É possível provar que o tamanho do salto em certo ponto x é igual a $\mathcal{P}(X = x)$. De fato,

$$\mathcal{P}(X = x) = \lim_{n\to\infty} \mathcal{P}\left(x - \frac{1}{n} < X \leq x\right) = F(x) - F\left(x^-\right),$$

onde x^- denota o limite à esquerda. Daí segue que F é contínua em x se, e só se $\mathcal{P}(X = x) = 0$.

É possível provar usando conceitos da Teoria da Medida que toda função satisfazendo $F1$, $F2$, $F3$ é função de distribuição de alguma variável aleatória. Para saber mais, veja [6, Seção 2.1].

Cada variável aleatória gera uma função de distribuição, mas uma função de distribuição pode corresponder a diversas variáveis aleatórias. Por exemplo, se $X : \Omega \to \mathbb{R}$ é uma variável aleatória contínua cuja densidade é dada por uma *função par* então $F_X = F_{-X}$, mas em geral $X \neq -X$.

2.2 Tipos de variáveis aleatórias

Definição 2.2.1 (Variável aleatória discreta)**.** *Seja X uma variável aleatória. Se o número de valores possíveis de X for enumerável (finito ou infinito), dizemos que X é uma variável aleatória* discreta*. A função $p(x_i) = \mathcal{P}(X = x_i)$, $i = 1, 2, \ldots$, é chamada* função de probabilidade da variável aleatória X *e satisfaz as seguintes condições:*

1. *$p(x_i) \geq 0$ para todo i;*
2. $\sum_{i=1}^{\infty} p(x_i) = 1$.

Exemplo 2.2.2. *Considere que uma moeda é lançada duas vezes, sendo que os possíveis resultados para cada lançamento são cara (c) e coroa (k), assim $\Omega = \{cc, ck, kc, kk\}$. Seja X a variável* *aleatória definida no espaço amostral que é igual ao número de coroas nos dois lançamentos. A variável X é discreta e sua distribuição de probabilidade será dada por*

$$\begin{aligned} p(0) &= \mathcal{P}(X=0) = \mathcal{P}(\{cc\}) = \frac{1}{4}, \\ p(1) &= \mathcal{P}(X=1) = \mathcal{P}(\{ck\}) + P(\{kc\}) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}, \\ p(2) &= \mathcal{P}(X=2) = \mathcal{P}(\{kk\}) = \frac{1}{4}, \end{aligned}$$

que satisfaz as condições da Definição 2.2.1.

Definição 2.2.3 (Variável aleatória contínua). *Dizemos que X é uma variável aleatória contínua se existe uma função $f : \mathbb{R} \to [0, \infty)$ denominada* função de densidade de X *tal que*

$$F_X(x) = \int_{-\infty}^{x} f(t)\,dt, \quad \forall x \in \mathbb{R}.$$

Dessa forma, para quaisquer $a, b \in \mathbb{R}$, com $a < b$ temos

$$\mathcal{P}(a < X \leq b) = F_X(b) - F_X(a) = \int_{a}^{b} f(t)\,dt.$$

Com isso, concluímos que quando X é uma variável aleatória contínua, a probabilidade de ocorrer um valor específico é zero, pois $\mathcal{P}(X=a) = \int_a^a f(x)\,dx = 0$. Pelo Teorema Fundamental do Cálculo, temos $F'_X(x) = f(x)$ para todo x. Além disso, $\int_{-\infty}^{\infty} f(x)dx = 1$, pela definição e pela propriedade $F3$. Reciprocamente, uma função $g(x) \geq 0$ é densidade de alguma variável aleatória se, e só se, $\int_{-\infty}^{\infty} g(x)\,dx = 1$, já que neste caso, a função F definida por

$$F(x) = \int_{-\infty}^{x} g(t)\,dt$$

é função de distribuição, pois satisfaz $F1$, $F2$ e $F3$.

Exemplo 2.2.4. *Vamos escolher ao acaso um número no intervalo $[0, 1]$. A probabilidade de ser o número $0,27$ é zero, pois como visto acima, todo ponto isolado em uma variável aleatória contínua tem probabilidade zero. Pela aditividade, a probabilidade de qualquer conjunto enumerável também é 0. Em particular, a probabilidade de escolher os racionais (que é enumerável) é 0.*

Exemplo 2.2.5. *Seja $A = \{x : -2 < x < 3\}$ e seja X uma variável aleatória tal que sua função densidade de probabilidade seja $f(x)$ definida abaixo, com k sendo uma constante. Qual* *deve ser o valor da constante k?*

$$f(x) = \begin{cases} k, & x \in A; \\ 0, & x \notin A. \end{cases}$$

Como f é uma função densidade de probabilidade ela deve satisfazer a condição que

$$\int_{-\infty}^{\infty} f(x)\,dx = 1 \Rightarrow \int_{-2}^{3} k dx = k \cdot (3 - (-2)) = 5k = 1 \Rightarrow k = \frac{1}{5}.$$

Exemplo 2.2.6. *Seja* $F_X(x) = \begin{cases} 0, & x < 0, \\ x, & 0 \leq x \leq 1, \\ 1, & x > 1. \end{cases}$

Como F_X é contínua, X tem densidade que é dada por $f(x) = F_X'(x) = \begin{cases} 1, & x \in (0,1) \\ 0, & x \notin [0,1]. \end{cases}$

Proposição 2.2.7. *Se X é variável aleatória em $(\Omega, \mathcal{A}, \mathcal{P})$, então o evento*

$$[X \in B] = \{\omega \in \Omega : \ X(\omega) \in \mathcal{B}\}$$

é evento aleatório para todo boreliano $\mathcal{B}$, isto é,

$$[X \in B] \in \mathcal{A}, \ \forall B \in \mathcal{B} = \sigma\text{-álgebra de Borel.}$$

Demonstração. A prova se dá utilizando a Teoria da Medida, mas podemos intuitivamente justificar a proposição: recordemos que a σ-álgebra $\mathcal{B}$, dos borelianos, é a menor σ-álgebra contendo os intervalos. Vamos, então, verificar a conclusão da proposição para um intervalo B:

(i) Se $B = (-\infty, b]$, então $[X \in B] \in \mathcal{A}$ pela Definição 2.1.1.

(ii) Se $B = (a, \infty)$, então $B = (-\infty, a]^c$ e $[X \in B] = [X \leq a]^c \in \mathcal{A}$, por (i). Neste caso, $\mathcal{P}(X \in B) = \mathcal{P}(X > a) = 1 - \mathcal{P}(X \leq a) = 1 - F_X(a)$.

(iii) Se $B = (a, b]$, então $[X \in B] = [a < X \leq b] = [X \leq b] - [X \leq a] \in \mathcal{A}$, por (i). $\mathcal{P}(X \in B) = \mathcal{P}(X \leq b) - \mathcal{P}(X \leq a) = F_X(b) - F_X(a)$.

(iv) Se $B = (a, b)$, então $B = \bigcup_{n=1}^{\infty} (a, b - \frac{1}{n}]$ e $[X \in B] = \bigcup_{n=1}^{\infty} (a < X \leq b - \frac{1}{n}] \in \mathcal{A}$, por (iii). Neste caso, $\mathcal{P}(X \in B) = \lim_{n \to \infty} \mathcal{P}(a < X \leq b - \frac{1}{n}] = \lim_{n \to \infty} [F_X\left(b - \frac{1}{n}\right) - F_X(a)] = F_X(b^-) - F_X(a)$.

Analogamente, podemos verificar que para todo intervalo B, $[X \in B] \in \mathcal{A}$ e $\mathcal{P}(X \in B)$ é determinada pela função de distribuição F_X. O mesmo resultado vale se $B = \bigcup_{i=1}^{n} B_i$, onde os B_i são intervalos disjuntos, uma vez que $[X \in B] = \bigcup_{i=1}^{n} [X \in B_i]$ e $\mathcal{P}(X \in B) = \sum_{i=1}^{n} \mathcal{P}(X \in B_i)$.

Com isso, a proposição vale para uniões finitas de intervalos, portanto, usando o Teorema de Extensão de Caratheodory [2, Capítulo 9], o resultado vale na σ-álgebra $\mathcal{B}$. □

Teoricamente, para verificar se certa propriedade é válida para todo boreliano, é suficiente verificar se é válida para toda união finita de intervalos e, se vale para B_n, para todo n, e $B_n \downarrow B$ ou $B_n \uparrow B$, então vale para B, ou seja, se continua válida também para limites monótonos. Enfatizamos outra implicação desta demonstração: as probabilidades $\mathcal{P}(X \in B)$ são determinadas pela função de distribuição F_X.

Definição 2.2.8. *A probabilidade $\mathcal{P}_X$, definida na σ-álgebra de Borel por $\mathcal{P}_X(B) = \mathcal{P}(X \in B)$, é chamada* distribuição de X.

Nos casos discreto e contínuo, podemos descrever a distribuição por meio da função de probabilidade ou densidade, como veremos na próxima seção.

2.3 Principais modelos de distribuição

Nesta seção, vamos introduzir os principais modelos de distribuição, tanto no caso discreto quanto no caso contínuo. Existem vários outros modelos que não serão tratados aqui mas que também são tão importantes quanto os apresentados a seguir, como o de Poisson, o modelo exponencial, o modelo gama, etc.

Definição 2.3.1 (Modelo Uniforme Discreto)**.** *Seja X uma variável aleatória discreta, cujos possíveis valores são representados somente por $x_1, x_2, \ldots$. Dizemos que X segue o modelo* Uniforme Discreto *se atribui a mesma probabilidade a cada um desses valores, isto é,*

$$\mathcal{P}(X = x_i) = \mathcal{P}(X = x_j) \text{ ou } p(x_i) = p(x_j) \text{ para todo } i, j.$$

Aqui, vale a pena ressaltar que $\{x_1, x_2, \ldots\}$ deve ser finito, não pode ser infinito enumerável. De fato, se fosse, teríamos $1 = \mathcal{P}(\Omega) = \sum p(x_1)$. Mas essa última soma é 0 se $p(x_1) = 0$ e é ∞ se $p(x_1) > 0$, o que é um absurdo. Dessa forma, seja n o número de elementos em tal conjunto finto. Temos $\mathcal{P}_X(B) = \dfrac{|B|}{n}$ para todo $B \in \mathcal{A}$.

Usaremos a notação $X \sim U\{x_1, \ldots, x_n\}$ para indicar que X possui distribuição Uniforme Discreto entre os n valores.

Definição 2.3.2 (Modelo Uniforme Contínuo)**.** *Uma variável aleatória X tem distribuição* Uniforme Contínua *no intervalo $B = [a, b]$, $a < b$, se sua função densidade de probabilidade é dada por:*

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & a \leq x \leq b; \\ 0, & \textit{caso contrário.} \end{cases}$$

Assim, a probabilidade de B é dada por

$$\mathcal{P}_X(B) = \int_B f(x)\, dx, \quad \forall B \in \mathcal{B}.$$

Usaremos a notação $X \sim U[a, b]$ para indicar que X possui distribuição Uniforme Contínua no intervalo dado.

Definição 2.3.3 (Modelo Normal)**.** *Dizemos que uma variável aleatória contínua X tem distribuição* Normal *com parâmetros μ e σ^2 se sua função densidade é dada por*

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \ para - \infty < x < \infty.$$

Posteriormente, vamos mostrar que μ é a média (ou esperança) de X e σ^2 é a variância de X. Para saber o valor da probabilidade de X estar em um intervalo $[a, b]$, devemos resolver a integral da função densidade em tal intervalo, ou seja,

$$\mathcal{P}(a \leq X \leq b) = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx.$$

Notemos que essa integral é da forma $\int_\alpha^\beta e^{-y^2} dy$, que não possui primitiva em termos de funções elementares. Mas ainda assim podemos mostrar que, no caso da integral definida em $\mathbb{R}$, temos $\mathcal{P}(-\infty < X < \infty) = 1$ (ver Apêndice B). Para outros intervalos podemos usar métodos numéricos e obter aproximações.

Usaremos a notação $X \sim N(\mu, \sigma^2)$ para indicar que X tem distribuição Normal com parâmetros μ e σ^2.

Definição 2.3.4 (Modelo binomial)**.** *Dizemos que X tem* distribuição binomial *com parâmetros n e p, onde n é um inteiro positivo e $0 < p < 1$, se*

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

Notação: $X \sim b(n, p)$.

Esta é a distribuição que atribuímos, por exemplo, ao número de caras obtidas em

n lançamentos de uma moeda tendo probabilidade p de dar cara.

Definição 2.3.5. *Chamamos de* ensaios de Bernoulli *com parâmetro p os ensaios binomiais cuja probabilidade de sucesso é a mesma p para todo ensaio.*

Definição 2.3.6 (Modelo geométrico)**.** *Dizemos que X tem* distribuição geométrica *com parâmetro p se sua função de probabilidade é dada por*

$$p(k) = (1-p)^{k-1} p, \quad k = 1, 2, \ldots.$$

Notação: $X \sim geom(p)$.

A distribuição geométrica indica a probabilidade de obter o primeiro sucesso em uma sequência de ensaios de Bernoulli com probabilidade p de sucesso exatamente no k-ésimo ensaio.

2.4 Vetores aleatórios

Definição 2.4.1. *(a) Denomina-se* vetor aleatório *(ou variável aleatória n-dimensional), um vetor $\tilde{X} = (X_1, X_2, \ldots, X_n)$ cujas componentes são variáveis aleatórias definidas no mesmo espaço de probabilidade $(\Omega, \mathcal{A}, \mathcal{P})$.*

(b) A função de distribuição $F = F_{\tilde{X}} = F_{X_1,\ldots,X_n}$ *de um vetor aleatório $\tilde{X} = (X_1, \ldots, X_n)$ é dada por:*

$$F(\tilde{x}) = F(x_1, \ldots, x_n) = \mathcal{P}(X_1 \leq x_1, \ldots, X_n \leq x_n), \forall \tilde{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

F é também chamada função de distribuição conjunta *das variáveis aleatórias $X_1, \ldots, X_n$.*

Quanto a função de distribuição de um vetor aleatório, podemos traçar propriedades análogas àquelas da Proposição 2.1.7.

Proposição 2.4.2. *Seja F a função de distribuição de um vetor aleatório $(X_1, \ldots, X_n)$. São válidos:*

F1. $F(x_1, \ldots, x_n)$ é não-decrescente em cada uma das variáveis.

F2. $F(x_1, \ldots, x_n)$ é contínua a direita em cada uma das variáveis.

F3. Para todo i, $\lim_{x_i \to -\infty} F(x_1, \ldots, x_i, \ldots, x_n) = 0$ e $\lim_{x_i \to +\infty} F(x_1, \ldots, x_i, \ldots, x_n) = 1$.

Dado um intervalo $(a, b]$ e $g : \mathbb{R}^k \to \mathbb{R}$ uma função qualquer, o operador diferença é definido por

$$\Delta_{a,b} g(x_1, \ldots, x_n) = g(x_1, \ldots, x_{n-1}, b) - g(x_1, \ldots, x_{n-1}, a).$$

F4. $\Delta_{a_1,b_1}, \ldots, \Delta_{a_n,b_n} F(x_1, \ldots, x_n) \geq 0, \forall (a_k, b_k], a_k < b_k, k = 1, \ldots, n.$

Essa quarta propriedade é de fundamental importância, pois sem ela podemos encontrar uma função que satisfaz $F1$, $F2$ e $F3$ porém apresenta probabilidade negativa (ver [6, Exemplo 14]), não sendo portanto, uma função de distribuição de nenhum vetor aleatório. Essa propriedade vale para cada componente do vetor. Por outro lado, pode-se verificar (ver [4, parágrafo 2.5]) que uma função que satisfaz as propriedades $F1$, $F2$, $F3$ e $F4$ é uma função de distribuição de um vetor aleatório.

Definição 2.4.3. *A probabilidade definida em $\mathcal{B}^n$ por $\mathcal{P}\left(\tilde{X} \in B\right)$ é chamada distribuição de $\tilde{X}$ ou distribuição conjunta de $X_1, \cdots, X_n$.*

Notação. $\mathcal{P}_{\tilde{X}}(B) = \mathcal{P}\left(\tilde{X} \in B\right)$, *$\mathcal{P}_{\tilde{X}}$ é a distribuição de $\tilde{X}$.*

Proposição 2.4.4. *(a) Se o vetor aleatório $\tilde{X}$ é discreto, então*

$$\mathcal{P}_{\tilde{X}}(B) = \sum_{i: x_i \in B} \mathcal{P}\left(\tilde{X} = \tilde{x_i}\right), \forall B \in \mathcal{B}^n.$$

(b) Se $\tilde{X}$ é contínuo com densidade $f(x_1, \ldots, x_n)$, então

$$\mathcal{P}_{\tilde{X}}(B) = \mathcal{P}\left(\tilde{X} \in B\right) = \int \cdots \int_B f(x_1, \ldots, x_n)\, dx_1 \ldots dx_n.$$

2.5 Independência de variáveis aleatórias

Vamos lembrar da Definição 1.2.11, que é equivalente à definição a seguir:

Definição 2.5.1. *As variáveis aleatórias $X_1, \ldots, X_n$ são (coletivamente) independentes se*

$$\mathcal{P}(X_1 \in B_1, X_2 \in B_2, \ldots, X_n \in B_n) = \prod_{i=1}^{n} \mathcal{P}(X_i \in B_i), \forall B_i \in \mathcal{B}, i = 1, \ldots, n.$$

A seguir, vamos obter critérios de independência em geral e um específico para o caso contínuo.

Proposição 2.5.2 (Critério de independência). *(a) Se $X_1, \ldots, X_n$ são independentes, então*

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i), \forall (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

(b) Por outro lado, se existem funções $F_1, \ldots, F_n$ tais que

$$\lim_{x_i \to +\infty} F_i(x_i) = 1, \forall i$$

e

$$F_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \prod_{i=1}^{n} F_i(x_i), \forall (x_1,\ldots,x_n) \in \mathbb{R}^n,$$

então $X_1,\ldots,X_n$ *são independentes e* $F_i = F_{X_i}, \forall i = 1,\ldots,n.$

Demonstração. (a) Supondo $X_1,\ldots,X_n$ independentes. Então

$$\begin{aligned} F_{X_1,\ldots,X_n}(x_1,\ldots,x_n) &= \mathcal{P}(X_1 \le x_1,\ldots,X_n \le x_n) \\ &= \mathcal{P}(X_1 \in (-\infty,x_1],\ldots,X_n \in (-\infty,x_n]) \\ &= \prod_{i=1}^{n} \mathcal{P}(X_i \in (-\infty,x_i]) = \prod_{i=1}^{n} \mathcal{P}(X_i \le x_i) = \prod_{i=1}^{n} F_{X_i}(x_i), \forall (x_1,\ldots,x_n). \end{aligned}$$

(b) A prova pode ser encontrada em [6, Proposição 2.4.b].

□

Proposição 2.5.3 (Critério para independência no caso contínuo). *(a) Se* $X_1,\ldots,X_n$ *são independentes e possuem densidades* $f_{X_1},\ldots,f_{X_n}$*, então a função*

$$f(x_1,\ldots,x_n) = \prod_{i=1}^{n} f_{X_i}(x_i), (x_1,\ldots,x_n) \in \mathbb{R}^n,$$

é densidade conjunta das variáveis aleatórias $X_1,\ldots,X_n$*, isto é,* $f = f_{X_1,\cdots,X_n}$.

(b) Por outro lado, se $X_1,\ldots,X_n$ *têm densidade conjunta* f *satisfazendo*

$$f(x_1,\ldots,x_n) = \prod_{i=1}^{n} f_i(x_i), \forall (x_i,\ldots,x_n) \in \mathbb{R}^n,$$

onde $f_i(x) \ge 0$ *e* $\int_{-\infty}^{\infty} f_i(x)\,dx = 1$*,* $\forall i$*, então* $X_1,\ldots,X_n$ *são independentes e* f_i *é a densidade de* X_i*, para* $i = 1,\ldots,n$*.*

Demonstração. (a) Se $X_1,\ldots,X_n$ são independentes, então

$$\begin{aligned} F_{X_1,\ldots,X_n}(x_1,\ldots,x_n) &= \prod_{i=1}^{n} F_{X_i}(x_i) = \prod_{i=1}^{n} \int_{-\infty}^{x_i} f_{X_i}(t_i)\,dt_i \\ &= \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f_{X_1}(t_1)\ldots f_{X_n}(t_n)\,dt_1\ldots dt_n. \end{aligned}$$

Logo $\prod_{i=1}^{n} f_{X_i}$ é densidade conjunta de $X_1,\ldots,X_n$.

(b)

$$\begin{aligned} F_{X_1,\ldots,X_n}(x_1,\ldots,x_n) &= \int_{-\infty}^{x_n}\cdots\int_{-\infty}^{x_1} f(t_1,\ldots,t_n)\,dt_1\ldots dt_n \\ &= \int_{-\infty}^{x_n}\cdots\int_{-\infty}^{x_1} f_1(t_1)\ldots f_n(t_n)\,dt_1\ldots dt_n \\ &= \prod_{i=1}^{n}\int_{-\infty}^{x_i} f_i(t_i)\,dt_i. \end{aligned}$$

Definindo $F_i(x_i) = \int_{-\infty}^{x_i} f_i(t_i)\,dt_i$, temos, por hipótese, $\lim_{x_i\to+\infty} F_i(x_i) = 1$.

A Proposição 2.5.2(b) implica que $X_1,\ldots,X_n$ são independentes e $F_i = F_{x_i}$, logo f_i é densidade de X_i.

$\square$

Capítulo 3

Esperança e Variância de variáveis aleatórias

3.1 Esperança e suas propriedades

Definição 3.1.1. *Seja X uma variável aleatória discreta, cujos valores possíveis são $x_1, x_2, x_3, \ldots$ e com função de probabilidade dada por $p(x_i)$. A* esperança *de X é definida por*

$$E(X) = \sum_i x_i p(x_i) = \sum_i x_i \mathcal{P}(X = x_i).$$

Este valor está bem definido quando a soma não depende da ordem dos termos, em particular quando $\sum\limits_i |x_i| p(x_i) < \infty$, isto é, quando a série converge absolutamente. A esperança de X também é chamada de *média* de X, ou *valor esperado* de X, ou *expectância* de X. De fato, $E(X)$ é uma *média ponderada*, onde os pesos são as probabilidades $p(x_i)$, isto é, $E(X)$ é uma média dos valores possíveis de X, ponderada conforme a distribuição de X.

Intuitivamente, uma explicação desta definição se encontra na interpretação de probabilidade como limite de frequências relativas, ou seja, interpretando novamente X como um característico numérico do resultado de um experimento, e admitindo que vamos repetir o experimento n vezes, independentemente, e observar os valores desse característico numérico. Nesses n experimentos, se n é grande, as observações tomarão o valor x_i com frequência relativa de aproximadamente $p(x_i)$, $\forall i$, isto é, x_i aparecerá mais ou menos $np(x_i)$ vezes nas n observações. Portanto, o valor médio observado nesses n ensaios do experimento, isto é, a média aritmética dos n valores observados, será aproximadamente igual a

$$\frac{\sum_i [x_i \cdot np(x_i)]}{n} = \sum_i x_i p(x_i).$$

Este valor será o limite quando $n \to \infty$, isto é, o valor médio obtido em n ensaios do experimento convergirá para $E(X)$ quando $n \to \infty$. Logo, podemos dizer que "esperamos" obter a longo prazo um valor médio $E(X)$.

Vamos apresentar através de exemplos, os cálculos da *esperança* de alguns dos modelos teóricos discretos que foram definidos na Seção 2.3 do capítulo anterior.

Exemplo 3.1.2. *Considere a variável aleatória X com modelo Uniforme Discreto (ver Definição 2.3.1) entre os valores 1 e n. Aplicando a definição de esperança matemática e a expressão para a soma dos termos de uma progressão aritmética temos*

$$E(X) = \sum_{i=1}^{n} x_i \mathcal{P}(X = x_i) = \sum_{i=1}^{n} x_i \cdot \frac{1}{n} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{n+1}{2}.$$

Em particular, a esperança ao jogar um dado honesto é $\dfrac{(1+2+3+4+5+6)}{6} = 3,5.$

Exemplo 3.1.3. *Seja X a variável aleatória com distribuição Binomial de parâmetros n e p, conforme Definição 2.3.4. Temos*

$$\begin{aligned}
E(X) &= \sum_{k=0}^{n} k \mathcal{P}(X = k) \\
&= \sum_{k=1}^{n} k \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} \frac{n!}{(n-k)!(k-1)!} p^k (1-p)^{n-k} \\
&= np \sum_{k=1}^{n} \frac{(n-1)!}{(n-k)!(k-1)!} p^{k-1} (1-p)^{n-k}.
\end{aligned}$$

Substituindo nesta última expressão, $k-1$ por j, obtemos

$$E(X) = np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j (1-p)^{n-1-j} = np,$$

uma vez que a somatória é igual a 1, pois corresponde a somar todas as probabilidades de uma variável Binomial com parâmetros $n-1$ e p.

Exemplo 3.1.4. *Lembrando que chamamos de ensaios de Bernoulli, os ensaios Binomiais cuja probabilidade de sucesso é a mesma* p *para todo o ensaio. Então para uma variável aleatória* X *com distribuição Bernoulli de parâmetro* p*, temos*

$$E(X) = 0 \times (1-p) + 1 \times p = p.$$

Em particular, no caso de uma moeda honesta, se X *denota a variável aleatória satisfazendo* $X(c) = 0$ *e* $X(k) = 1$*, temos* $E(X) = \dfrac{1}{2}$.

Na sequência vamos definir esperança para o caso contínuo.

Definição 3.1.5. *Seja* X *uma variável aleatória qualquer e* F *sua função de distribuição. A* esperança *de* X *é definida por*

$$E(X) = \int_{-\infty}^{\infty} x dF(x),$$

quando a integral imprópria de Riemann-Stieltjes está bem definida. Se X *tem densidade* f*, então a integral acima pode ser reescrita, pelo item 7 do Apêndice A, como*

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

Se a densidade f for integrável a Riemann então esta última integral também será de Riemann [10, Teorema 6.17]. Em outras palavras, podemos continuar a trabalhar com a integral de Riemann no caso contínuo.

No caso discreto, já vimos que $E(X) = \sum_i x_i p(x_i)$. Pelo item 6 do Apêndice A, esta definição concorda com a Definição 3.1.5.

Vamos agora, apresentar através de exemplos, os cálculos da esperança de alguns dos modelos contínuos que foram definidos anteriormente.

Exemplo 3.1.6. *Sendo* X *uma variável aleatória com distribuição uniforme no intervalo* $[a, b]$*, ou seja,* $X \sim U[a, b]$*, então* $f(x) = \dfrac{1}{b-a}$ *para* $a \leq x \leq b$ *e* $f(x) = 0$ *em caso contrário. A esperança de* X *fica dada por*

$$E(X) = \int_a^b x \frac{1}{b-a} dx = \frac{1}{b-a} \cdot \frac{x^2}{2}\Big|_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}.$$

Exemplo 3.1.7. *Seja* X *uma variável aleatória com distribuição normal com parâmetros* μ *e* σ^2*. A esperança de* X *é dada por*

$$E(X) = \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx.$$

Fazendo a mudança de variáveis $y = \dfrac{x-\mu}{\sigma\sqrt{2}}$, $dy = \dfrac{1}{\sigma\sqrt{2}}dx$, *temos*

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} (y\sigma\sqrt{2}+\mu)\frac{1}{\sigma\sqrt{2\pi}}e^{-y^2}\sigma\sqrt{2}dy \\ &= \int_{-\infty}^{\infty} (y\sigma\sqrt{2}+\mu)\frac{1}{\sqrt{\pi}}e^{-y^2}dy \\ &= \sigma\sqrt{\frac{2}{\pi}}\int_{-\infty}^{\infty} ye^{-y^2}dy + \frac{\mu}{\sqrt{\pi}}\int_{-\infty}^{\infty} e^{-y^2}dy. \end{aligned}$$

A primeira integral da linha anterior vale 0 *pois é uma integral de função ímpar sobre um intervalo simétrico. A segunda integral vale* $\sqrt{\pi}$ *pelo Apêndice B. Logo,* $E(X) = \mu$.

Definição 3.1.8. *Se* $E(X)$ *é finita, dizemos que* X *é* integrável*.*

Observações.

(A) Nesse texto, além das esperanças finitas, vamos admitir a existência de esperanças infinitas, desde que seu valor seja "bem comportado", isto é, tome o valor $+\infty$ ou $-\infty$.

(B) A esperança está bem definida se $\int_0^\infty xdF(x)$ ou $\int_{-\infty}^0 xdF(x)$ for finita. De fato, sendo

$$\int_{-\infty}^{\infty} xdF_X(x) = \underbrace{\int_{-\infty}^{0} xdF_X(x)}_{I} + \underbrace{\int_{0}^{\infty} xdF_X(x)}_{II},$$

temos $I \leq 0,\ II \geq 0$ e

(i) se I e II são finitas, então $E(X) = I + II$, ou seja, X é integrável,

(ii) se I é finita e $II = +\infty$, então $E(X) = +\infty$,

(iii) se $I = -\infty$ e II é finita, então $E(X) = -\infty$,

(iv) se $I = -\infty$ e $II = +\infty$, então $E(X)$ não está definida.

Dessa forma, X é integrável se, e só se, $\int_{-\infty}^{\infty} |x|dF_X(x) < \infty$. Mais adiante veremos que essa última integral vale $E|X|$.

Proposição 3.1.9. $E(X) = \int_0^\infty (1-F(x))dx - \int_{-\infty}^0 F(x)dx$.

Demonstração. Vamos provar que $\int_0^\infty xdF(x) = \int_0^\infty (1-F(x))dx$. De forma completamente análoga, é possível mostrar que $\int_{-\infty}^0 xdF(x) = -\int_{-\infty}^0 F(x)dx$. Fazendo a integração por partes com $u = x$ e $dv = dF(x)$, temos $du = dx$ e $v = F(x)$, logo

$$\forall b > 0,\ \int_0^b xdF(x) = [xF(x)]_0^b - \int_0^b F(x)dx = bF(b) - \int_0^b F(x)dx = \int_0^b [F(b) - F(x)]dx.$$

A ideia agora é fazer $b \to \infty$ e passar o limite para dentro da integral, obtendo $F(b) \to 1$. Feito isso, obtemos o resultado desejado. Como $F(b) \leq 1$ e $1 - F(x) \geq 0$, temos

$$\int_0^b [F(b) - F(x)]dx \leq \int_0^\infty [1 - F(x)]dx, \ \forall b > 0,$$

de modo que

$$\lim_{b\to\infty} \int_0^b xdF(x) \leq \int_0^\infty [1 - F(x)]dx.$$

Por outro lado, seja $\lambda > 0$. Se $b > \lambda$, então

$$\begin{aligned}\int_0^\infty [F(b) - F(x)]dx &\geq \int_0^\lambda [F(b) - F(x)]dx \\ &= \int_0^\lambda [F(b) - 1]dx + \int_0^\lambda [1 - F(x)]dx \\ &= \lambda[F(b) - 1] + \int_0^\lambda [1 - F(x)]dx,\end{aligned}$$

e portanto

$$\lim_{b\to\infty} \int_0^b [F(b) - F(x)]dx \geq \int_0^\lambda [1 - F(x)]dx + \lim_{b\to\infty} \lambda[F(b) - 1] = \int_0^\lambda [1 - F(x)]dx.$$

Já que isto vale para todo $\lambda > 0$, temos

$$\int_0^\infty xdF(x) \geq \lim_{\lambda\to\infty} \int_0^\lambda [1 - F(x)]dx = \int_0^\infty [1 - F(x)]dx.$$

□

Corolário 3.1.10. *Se X tomar somente valores não-negativos, ou seja, $X(\omega) \geq 0, \ \forall \omega \in \Omega$, então $F_X(x) = 0$ para $x < 0$ e*

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

Demonstração. A prova é imediata, pois $1 - F_X(x) = \mathcal{P}(X > x)$. □

Corolário 3.1.11. *Se a variável aleatória X assumir somente valores inteiros não-negativos, então*

$$E(X) = \sum_{n=0}^\infty \mathcal{P}(X > n) = \sum_{n=1}^\infty \mathcal{P}(X \geq n).$$

Demonstração. Pela proposição temos

$$E(X) = 1 - F(0) + 1 - F(1) + 1 - F(2) + \cdots = \sum_{n=0}^{\infty}[1 - F(n)] = \sum_{n=0}^{\infty}\mathcal{P}[X > n].$$

Como X assume apenas valores inteiros, temos

$$\mathcal{P}(X > n) = \mathcal{P}(X \geq n + 1),$$

logo

$$\sum_{n=0}^{\infty}\mathcal{P}(X > n) = \sum_{n=0}^{\infty}\mathcal{P}(X \geq n + 1) = \sum_{n=1}^{\infty}\mathcal{P}(X \geq n).$$

□

Corolário 3.1.12. *Seja $A \in \mathcal{A}$ e seja*

$$I_A(\Omega) = \begin{cases} 1 & \textit{se } \omega \in A, \\ 0 & \textit{se } \omega \notin A, \end{cases}$$

a função indicadora do evento A. Então $E(I_A) = \mathcal{P}(A)$.

Demonstração. Pelo corolário anterior e, como I_A só assume os valores 0 e 1, segue que

$$E(I_A) = \sum_{n=0}^{\infty}\mathcal{P}(I_A > n) = \mathcal{P}(I_A > 0) = \mathcal{P}(I_A = 1) = \mathcal{P}(A).$$

□

Exemplo 3.1.13. *Seja X o número de lançamentos necessários até obter a primeira cara ao se lançar uma moeda independentemente várias vezes e p a probabilidade de obter cara em um dado lançamento. Então X toma apenas os valores $1, 2, 3, \ldots$ e $\mathcal{P}(X \geq n)$ é a probabilidade de não obter cara nos lançamentos $1, 2, \ldots, n-1$, isto é, a probabilidade de sair coroa nos lançamentos 1 até $n-1$, que vale $(1-p)^{n-1}$. Logo*

$$E(X) = \sum_{n=1}^{\infty}\mathcal{P}(X \geq n) = \sum_{n=1}^{\infty}(1-p)^{n-1} = \sum_{n=0}^{\infty}(1-p)^{n} = \frac{1}{1-(1-p)} = \frac{1}{p}.$$

Portanto, vemos que a esperança do tempo de espera até o primeiro sucesso em uma sequência de ensaios de Bernoulli (ou seja, a esperança do modelo geométrico) é $\frac{1}{p}$, onde p é a probabilidade de sucesso em cada ensaio. Para uma moeda honesta, essa esperança vale 2.

Como $|X| \geq 0$, o Corolário 3.1.10 implica que

$$\begin{aligned} E|X| &= \int_0^\infty \mathcal{P}(|X| > x)dx = \int_0^\infty [\mathcal{P}(X > x) + \mathcal{P}(X < -x)]dx \\ &= \int_0^\infty [1 - F_X(x) + F_X((-x)^-)]dx. \end{aligned}$$

Vale lembrar que $F_X((-x)^-)$ é o limite da função $F_X(y)$ quando $y \uparrow (-x)$. Logo, $F_X((-x)^-) = F_X(-x)$ quando $-x$ é o ponto de continuidade de F_X. Por isso, $F_X((-x)^-)$ e $F_X(-x)$ são funções monótonas (decrescentes em x) iguais, exceto em um número finito ou enumerável de pontos. Com isso concluímos que

$$\int_0^\infty F_X((-x)^-)dx = \int_0^\infty F_X(-x)dx = \int_{-\infty}^0 F_X(x)dx.$$

Resumindo, temos

$$E|X| = \int_0^\infty [1 - F_X(x)]dx + \int_{-\infty}^0 F_X(x)dx.$$

Então, pela prova da Proposição 3.1.9,

$$\begin{aligned} E|X| &= \int_0^\infty x dF_X(x) - \int_{-\infty}^0 x dF_X(x) \\ &= \int_0^\infty |x| dF_X(x) - \int_{-\infty}^0 |x| dF_X(x) = \int_{-\infty}^\infty |x| dF_X(x). \end{aligned} \tag{3.1}$$

Assim provamos que $E|X| = \int_{-\infty}^\infty |x| dF_X(x)$, e que X é integrável se, e somente se, $E|X| < \infty$.

Proposição 3.1.14 (Propriedades da Esperança)**.** *Sejam X, Y variáveis aleatórias e a, b números reais. São válidos:*

E1. *Se $X = c$, ou seja, $X(\omega) = c \ \forall \omega \in \Omega$, então $E(X) = c$.*

E2. *Se $X \leq Y$ então $E(X) \leq E(Y)$, se as esperanças estão bem definidas.*

E3. Linearidade.

- *E3.i.* *Se $E(X)$ está bem definida, então $E(aX + b) = aE(X) + b$ para todo $a, b \in \mathbb{R}$. Aqui estamos convencionando que $0 \cdot \infty = 0$.*
- *E3.ii.* *$E(aX + bY) = aE(X) + bE(Y)$, quando o termo à direita da igualdade tem sentido, isto é, não temos um caso $+\infty - \infty$.*

E4. Desigualdade de Jensen. *Seja φ uma função convexa definida na reta, isto é, uma função que satisfaz $t\varphi(a)+(1-t)\varphi(b) \geq \varphi(at+(1-t)b)$. Se a variável aleatória X é integrável, então*

$$E\varphi(X) \geq \varphi(E(X)).$$

Escreve-se: $E\varphi(X) = E(\varphi(X))$.

Demonstração. $E1$. Como X é uma variável aleatória discreta, então

$$E(X) = c\mathcal{P}(X = c) = c \cdot 1 = c.$$

$E2$. Se $X \leq Y$, então $Y \leq z$ implica $X \leq z$, logo $[Y \leq z] \subset [X \leq z]$. Portanto, $F_Y(z) \leq F_X(z)$ e $1 - F_Y(z) \geq 1 - F_X(z)$. Pela Proprosição 3.1.9,

$$E(Y) = \int_0^{\infty}(1-F_Y(z))dz - \int_{-\infty}^{0} F_Y(z)dz \geq \int_0^{\infty}(1-F_X(z))dz - \int_{-\infty}^{0} F_X(z)dz = E(X).$$

$E3$. i. Se $a = 0$, então $E(aX+b) = E(b) = b = aE(X)+b$. Se $a > 0$, então

$$F_{aX+b}(x) = \mathcal{P}(aX+b \leq x) = \mathcal{P}\left(X \leq \frac{x-b}{a}\right) = F_X\left(\frac{x-b}{a}\right),$$

logo tomando $y = \dfrac{x-b}{a}$ obtemos

$$\begin{aligned}
E(aX+b) &= \int_0^{\infty}(1 - F_{aX+b}(x))dx - \int_{-\infty}^{0} F_{aX+b}(x)dx \\
&= \int_0^{\infty}\left(1 - F_X\left(\frac{x-b}{a}\right)\right)dx - \int_{-\infty}^{0} F_X\left(\frac{x-b}{a}\right)dx \\
&= a\int_{-b/a}^{\infty}(1 - F_X(y))dy - a\int_{-\infty}^{-b/a} F_X(y)dy \\
&= a\int_0^{\infty}(1 - F_X(y))dy - a\int_{-\infty}^{0} F_X(y)dy + a\int_{-b/a}^{0}(1 - F_X(y))dy \\
&\quad + a\int_{-b/a}^{0} F_X(y)dy \\
&= aE(X) + a\int_{-b/a}^{0} dy = aE(X) + b.
\end{aligned}$$

Analogamente, temos o caso $a < 0$, e $E3.\,i.$ está provado. Para $E3.\,ii.$, resta provar $E(X+Y) = E(X) + E(Y)$ se o termo à direita está bem definido. Veremos mais tarde quando considerarmos esperanças de funções de vetores aleatórios.

$E4$. Dado x_0 e o ponto $\varphi(x_0)$ do gráfico de φ, pela convexidade, a reta L tangente ao gráfico no ponto $(x_0, \varphi(x_0))$ fica abaixo da curva φ, como na figura a seguir.

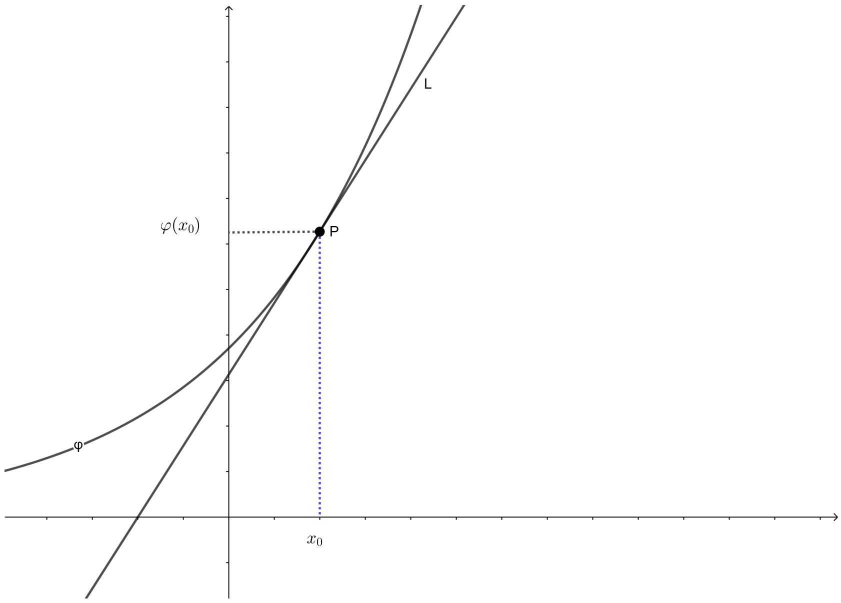

Figura 3.1: Gráfico de uma função convexa

Seja λ de forma que $y - \varphi(x_0) = \lambda(x - x_0)$ é a equação desta reta L. Então

$$\varphi(x) \geq L(x) = \varphi(x_0) + \lambda(x - x_0), \ \forall x.$$

Portanto, pelas propriedades anteriores, segue que

$$E(\varphi(X)) \geq E(L(X)) = \varphi(x_0) + \lambda(E(X) - x_0).$$

Tomando-se $E(X) = x_0$, vem $E(\varphi(X)) \geq \varphi(E(X))$. □

Proposição 3.1.15 (Critério para integrabilidade). *Seja X uma variável aleatória qualquer. Então*

$$\sum_{n=1}^{\infty} \mathcal{P}(|X| \geq n) \leq E|X| \leq 1 + \sum_{n=1}^{\infty} \mathcal{P}(|X| \geq n),$$

e portanto X é integrável se, e somente se, $\sum\limits_{n=1}^{\infty} \mathcal{P}(|X| \geq n) < \infty$.

Demonstração. Seja $\lfloor x \rfloor$ o maior inteiro menor ou igual a x, a função piso de x. Se $x \geq 0$, então a variável aleatória $\lfloor |X| \rfloor$ assume o valor k quando $k \leq |X| < k+1$ e

$$0 \leq \lfloor |X| \rfloor \leq |X| \leq \lfloor |X| \rfloor + 1,$$

logo, por (E2) e (E3),

$$0 \leq E\lfloor |X| \rfloor \leq E|X| \leq 1 + E\lfloor |X| \rfloor .$$

Mas pelo Corolário 3.1.11,

$$E\lfloor |X| \rfloor = \sum_{n=1}^{\infty} \mathcal{P}(\lfloor |X| \rfloor \geq n) = \sum_{n=1}^{\infty} \mathcal{P}(|X| \geq n),$$

logo

$$\sum_{n=1}^{\infty} \mathcal{P}(|X| \geq n) \leq E|X| \leq 1 + \sum_{n=1}^{\infty} \mathcal{P}(|X| \geq n).$$

□

3.2 Esperanças de funções de variáveis aleatórias

Sejam X uma variável aleatória, $\varphi(x)$ uma função real mensurável (ver Definição A.1.1) e $Y = \varphi(X)$. Então Y é uma variável aleatória cuja esperança é dada por

$$E(Y) = \int y dF_{\varphi(X)}(y) = \int_0^{\infty} [1 - F_{\varphi(X)}(y)]dy - \int_{-\infty}^{0} F_{\varphi(X)}(y)dy,$$

pela Proposição 3.1.9.

Teorema 3.2.1. *Seja X uma variável aleatória, $\varphi(x)$ uma função polinomial (que é mensurável). Então*

$$E(\varphi(x)) = \int y dF_{\varphi(x)}(y) = \int \varphi(x) dF_X(x),$$

onde a existência de uma das integrais implica a existência da outra e a igualdade das duas.

Demonstração. Prova-se o caso geral com a Teoria da Medida. Mas já provamos na Equação (3.1) que para $\varphi(x) = |x|$ tal resultado é válido, e podemos provar o teorema para polinômios usando somente a integral de Riemann, baseando-nos na Proposição 3.1.9. Mostremos que

$$E(X^k) = \int x^k dF_X(x), \; para \; k = 1, 2, \ldots$$

Vamos abreviar $F = F_X$ e supor inicialmente k par. Então

$$\begin{aligned}
E(X^k) &= \int_0^\infty \mathcal{P}(X^k > t)dt \ \ (\text{pois } X^k \geq 0) \\
&= \int_0^\infty \mathcal{P}(X > \sqrt[k]{t})dt + \int_0^\infty \mathcal{P}(X < -\sqrt[k]{t})dt \\
&= \int_0^\infty [1 - F(\sqrt[k]{t})]dt + \int_0^\infty F((-\sqrt[k]{t})^-)dt \ \ (\text{fazendo } t = s^k) \\
&= \int_0^\infty [1 - F(s)]ks^{k-1}ds + \int_0^\infty F((-s)^-)ks^{k-1}ds.
\end{aligned}$$

Observe que $F((-s)^-)$ e $F(-s)$ são iguais exceto em um número enumerável de pontos, assim podemos desconsiderar o segundo sinal negativo. Fazendo $u = -s$ na segunda integral, temos

$$E(X^k) = \int_0^\infty [1 - F(s)]ks^{k-1}ds - \int_{-\infty}^0 F(u)ku^{k-1}du. \tag{3.2}$$

Por outro lado, de forma similar à prova da Proposição 3.1.9, é possível mostrar que

$$\begin{aligned}
\int_{-\infty}^\infty x^k dF(x) &= \int_0^\infty [1 - F(x)]dx^k - \int_{-\infty}^0 F(x)dx^k \\
&= k\left\{\int_0^\infty [1 - F(x)]x^{k-1}dx - \int_{-\infty}^0 F(x)x^{k-1}dx\right\}.
\end{aligned}$$

Logo $E(X^k) = \int x^k dF_X(x)$ para k par. Para k ímpar, a prova é análoga. Para φ polinômio em geral, o resultado segue pela linearidade da esperança e da integral de Stieltjes. □

Vamos diferenciar os casos discreto e contínuo do teorema anterior.

Caso discreto. Se X tiver função de probabilidade $p(x_i)$, então

$$E(\varphi(X)) = \sum_i \varphi(x)p(x_i).$$

Caso contínuo. Se X tiver densidade $f(x)$, então

$$E(\varphi(X)) = \int \varphi(x)f(x)dx.$$

De forma mais geral, é possível afirmar o seguinte:

Teorema 3.2.2. *Seja $\tilde{X} = (X_1, \ldots, X_n)$ um vetor aleatório e $\varphi : \mathbb{R}^n \to \mathbb{R}$, $\mathcal{B}$-mensurável.*

Então

$$E\varphi(\tilde{X}) = \int y dF_{\varphi(\tilde{X})}(y) = \int_{\mathbb{R}^n} \varphi dF_{\tilde{X}} = \int \cdots \int \varphi(x_1, \ldots, x_n) dF_{\tilde{X}}(x_1, \ldots, x_n),$$

onde a última integral é uma integral n-dimensional de Stieltjes, assim como a penúltima.

Corolário 3.2.3. $E(X^k) = k\left\{\int_0^\infty [1 - F(x)]x^{k-1}dx - \int_{-\infty}^0 F(x)x^{k-1}dx\right\}$*, para* $k = 1, 2, \ldots$. *Essa é a fórmula (3.2).*

Exemplo 3.2.4. *Suponha que* $X \sim U[0,1]$ *e* $Y = min\left(X, \frac{1}{2}\right)$. *Então* $Y = \varphi(X)$*, onde*

$$\varphi(x) = min\left(x, \frac{1}{2}\right) = \begin{cases} x, & se\ x < \frac{1}{2} \\ \frac{1}{2}, & se\ x \geq \frac{1}{2}. \end{cases}$$

Usando o Teorema 3.1.9 e o item e) do Apêndice A.2, segue que a esperança de Y vale

$$\begin{aligned} E(Y) &= \int \varphi(x) dF_X(x) = \int \varphi(x) f_X(x) dx = \int_0^1 \varphi(x) dx \\ &= \int_0^{1/2} x dx + \int_{1/2}^1 \frac{1}{2} dx = \frac{1}{8} + \frac{1}{4} = \frac{3}{8}. \end{aligned}$$

O teorema anterior nos auxilia na definição de momentos a seguir:

Definição 3.2.5. *Seja X uma variável aleatória. O valor $E(X-b)^k$, se existe, é chamado k-ésimo* momento *de X em torno de b, para $b \in \mathbb{R}$, $k \in \mathbb{Z}_+^*$.*

O k-ésimo momento em torno de zero, $E(X^k)$, é chamado de k-ésimo momento de X ou momento de ordem k de X.

Se X é integrável, então o k-ésimo momento em torno da média, $E(X - E(X))^k$, se chama k-ésimo momento central de X.

Notemos que o primeiro momento é a *esperança* e o primeiro momento central é nulo, pois $E(X - E(X)) = 0$.

Definição 3.2.6. *O segundo momento central é chamado* variância *de X.*

Notação. $Var(X)$

Da definição anterior, segue que

$$\begin{aligned} Var(X) &= E(X - E(X))^2 = E(X^2 - 2XE(X) + (E(X))^2) \\ &= E(X^2) - 2E(X)E(X) + (E(X))^2 \\ &= E(X^2) - (E(X))^2. \end{aligned}$$

Além disso, se $X = c$, então $Var(X) = 0$, pois sendo $E(X) = c$, temos $Var(X) = E(X - c)^2 = E(0) = 0$.

Se $a, b \in \mathbb{R}$ então temos $Var(X + b) = Var(X)$ e $Var(aX + b) = a^2 Var(X)$, pois sendo $E(aX + b) = aE(X) + b$, temos que $Var(aX + b) = E(aX + b - aE(X) - b)^2 = E(a^2(X - E(X))^2) = a^2 Var(X)$.

Definição 3.2.7. *A raiz quadrada da variância é chamada de* desvio-padrão *de X, ou seja,*

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

A ideia do desvio-padrão é manter a mesma unidade da variável aleatória. Por exemplo, se X denota uma variável aleatória medida em metros (m), então $Var(X)$ terá a unidade em m^2, e assim σ_X terá a unidade em m.

Definição 3.2.8. *Seja $t > 0$. Definimos $E(|X|^t)$ como o t-*ésimo momento absoluto *de X.*

Da definição anterior, segue que momentos absolutos possuem a seguinte propriedade:

Proposição 3.2.9. *Sejam X uma variável aleatória e $t > 0$. A função*

$$f(t) = E^{1/t}|X|^t$$

é não decrescente em t. Notação: $E^{1/t}|X|^t = [E(|X|^t)]^{1/t}$.

Demonstração. Sejam $0 < s < t$. Temos que a função $\varphi(y) = |y|^{t/s}$ é convexa, pois $t/s > 1$. Se Y é integrável, a desigualdade de Jensen (Proposição 4 (E4)) implica que

$$E(|Y|^{t/s}) \geq |E(Y)|^{t/s}.$$

Tomando $Y = |X|^s$: segue que se $|X|^s$ é integrável, então

$$E(|X|^t) \geq E^{t/s}|X|^s,$$

isto é, $E^{1/t}|X|^t \geq E^{1/s}|X|^s$.

Se $|X|^s$ não é integrável, então $|X|^t$ também não o é, pois $s < t$ implica $|X|^s \leq 1 + |X|^t$ e, portanto

$$E(|X|^s) = +\infty \Rightarrow +\infty \leq 1 + E(|X|^t) \Rightarrow E(|X|^t) = +\infty.$$

De qualquer forma, segue que

$$E^{1/s}|X|^s \leq E^{1/t}|X|^t \text{ para } 0 < s < t.$$

□

Corolário 3.2.10. *Se $E(|X|^t)$ é finita para algum $0 < t < \infty$, então $E(|X|^s)$ é finito para todo s tal que $0 < s < t$.*

Corolário 3.2.11. *Se X e Y são variáveis aleatórias em $(\Omega, \mathcal{A}, \mathcal{P})$ tais que $E(|X|^t) < \infty$ e $E(|Y|^t) < \infty$, então $E(|X+Y|^t) < \infty$.*

Demonstração. Pela desigualdade triangular, $|X+Y| \leq |X| + |Y| \leq 2\max(|X|, |Y|)$. Portanto,

$$|X+Y|^t \leq 2^t \max(|X|^t, |Y|^t) \leq 2^t(|X|^t + |Y|^t),$$

logo

$$E(|X+Y|^t) \leq 2^t(E(|X|^t) + E(|Y|^t)).$$

□

Exemplo 3.2.12. *Se $E(X^2) < \infty$ então X é integrável. Além disso, se o k-ésimo momento é finito, então todos os momentos de ordem menor que k também são finitos.*

Proposição 3.2.13 (Desigualdade de Markov)**.** *Seja X uma variável aleatória não-negativa. Para todo $\lambda > 0$, vale que*

$$\mathcal{P}(X \geq \lambda) \leq \frac{1}{\lambda} E(X).$$

Demonstração. Pelo Corolário 3.1.10, temos que

$$E(X) = \int_0^\infty \mathcal{P}(X > x)dx = \int_0^\lambda \mathcal{P}(X > x)dx \; + \int_\lambda^\infty \mathcal{P}(X > x)dx$$
$$\geq \int_0^\lambda \mathcal{P}(X \geq \lambda)dx \; + \; 0 = \lambda \mathcal{P}(X \geq \lambda),$$

uma vez que $\mathcal{P}(X > x) \geq \mathcal{P}(X \geq \lambda)$ para x no intervalo $[0, \lambda]$. O resultado segue dividindo ambos os lados por λ. □

Proposição 3.2.14 (Desigualdade de Tchebychev)**.** *Se X é integrável, então para todo $\lambda > 0$ temos*

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

Demonstração. Pela Desigualdade de Markov, como $(X - E(X))^2$ é uma variável aleatória não-negativa, temos

$$\mathcal{P}(|X - E(X)| \geq \lambda) = \mathcal{P}((X - E(X))^2 \geq \lambda^2) \leq \frac{1}{\lambda^2} E(X - E(X))^2 = \frac{Var(X)}{\lambda^2}.$$

□

Proposição 3.2.15 (Desigualdade de Markov generalizada)**.** *Seja X uma variável aleatória qualquer. Então para todo $t > 0$,*

$$\mathcal{P}(|X| \geq \lambda) \leq \frac{E|X|^t}{\lambda^t},\ \forall \lambda > 0.$$

Demonstração. Sendo $|X|$ uma variável aleatória não-negativa, usamos o mesmo argumento da demonstração da proposição anterior. □

É importante salientar que se X é integrável então sua média μ minimiza o segundo momento $E(X-c)^2,\ c \in \mathbb{R}$, isto é,

$$Var(X) = E(X-\mu)^2 = \min_{c \in \mathbb{R}} E(X-c)^2.$$

Definição 3.2.16. *Uma* mediana *de uma variável aleatória X é um valor m que satisfaz*

$$\mathcal{P}(X \geq m) \geq \frac{1}{2} \text{ e } \mathcal{P}(X \leq m) \geq \frac{1}{2}.$$

Da definição, segue que se m é mediana de X, então m minimiza $E|X-c|,\ c \in \mathbb{R}$, isto é,

$$E|X-m| = \min_{c \in \mathbb{R}} E|X-c|.$$

Agora já temos todas as ferramentas necessárias para a demonstração da linearidade da esperança (propriedade (E3)). Necessitamos provar que $E(X+Y) = E(X) + E(Y)$, supondo que o termo à direita seja bem-definido. Por isso, sejam $\varphi(x,y) = x+y,\ \varphi_1(x,y) = x,\ \varphi_2(x,y) = y$. Pelo Teorema 3.2.1,

$$E(X+Y) = E\varphi(X,Y) = \int\int (x+y)dF_{X,Y}(x,y).$$

Pela linearidade da integral múltipla de Stieltjes, obtemos

$$E(X+Y) = \int\int x dF_{X,Y}(x,y) + \int\int y dF_{X,Y}(x,y) = E\varphi_1(X,Y) + E\varphi_2(X,Y) = E(X)+E(Y).$$

Proposição 3.2.17. *Se $X_1, \ldots, X_n$ são variáveis aleatórias independentes e integráveis, então $\prod_{i=1}^{n} X_i$ é integrável e $E(X_1 X_2 \ldots X_n) = \prod_{i=1}^{n} E(X_i)$.*

Demonstração. Vamos provar o caso $n = 2$ e o restante segue por indução, uma vez que $E(X_1 \ldots X_{n-1} X_n)$ será igual a $E(X_1 \ldots X_{n-1})E(X_n)$. Seja $\varphi(x,y) = xy$. A independência de X e Y implica

$$\begin{aligned} E(XY) &= E\varphi(X,Y) = \int\int \varphi(x,y)dF_X(x)dF_Y(y) \\ &= \int \left[\int xdF_X(x)\right] ydF_Y(y) = \int (E(X))ydF_Y(y) \\ &= E(X)\cdot E(Y). \end{aligned}$$

□

É importante ressaltar que $E(XY) = E(X)\cdot E(Y)$ não implica X e Y independentes, como vemos no seguinte exemplo.

Exemplo 3.2.18. *Sejam X e Y variáveis aleatórias tomando os valores $-1, 0, 1$, com distribuição uniforme nos pontos $(-1,-1), (-1,1), (1,-1), (1,1)$ e $(0,0)$, isto é, a probabilidade de cada um desses pontos é $\frac{1}{5}$. A função de probabilidade é a da seguinte tabela.*

$Y \backslash X$	-1	0	1
-1	$1/5$	0	$1/5$
0	0	$1/5$	0
1	$1/5$	0	$1/5$

Tabela 3.1: Distribuição conjunta

Então,

$$E(X) = (-1)\cdot\frac{2}{5} + 0\cdot\frac{1}{5} + 1\cdot\frac{2}{5} = 0 = E(Y),$$

e

$$E(XY) = \sum_{i,j} ijp(i,j) = \frac{1}{5}(1-1-1+1+0) = 0.$$

Portanto, $E(XY) = E(X)\cdot E(Y)$, mas X e Y não são independentes. De fato, temos, por exemplo,

$$\mathcal{P}(X=0,Y=0) = p(0,0) = \frac{1}{5} \neq \frac{1}{25} = \frac{1}{5}\cdot\frac{1}{5} = \mathcal{P}(X=0)\cdot P(Y=0).$$

Definição 3.2.19. *Sejam X e Y variáveis aleatórias integráveis. A* covariância *entre X e Y é definida por*

$$Cov(X,Y) = E[(X-E(X))(Y-E(Y))],$$

se esta esperança existe.

Pela linearidade da esperança, temos

$$Cov(X,Y) = E(XY - YE(X) - XE(Y) + E(X)\cdot E(Y)) = E(XY) - E(X)\cdot E(Y),$$

o que implica que a covariância entre duas variáveis integráveis existe se, e somente se, existe a esperança $E(XY)$.

Definição 3.2.20. *Se $Cov(X,Y) = 0$, dizemos que X e Y são* não-correlacionadas.

Se X e Y são independentes e integráveis, então são não-correlacionadas, pois neste caso $E(XY) = E(X)\cdot E(Y)$ pela Proposição 3.2.17. Pelo Exemplo 3.2.18, covariância zero não necessariamente implica independência.

Proposição 3.2.21. *Sejam $X_1, \dots, X_n$ variáveis aleatórias integráveis tais que $Cov(X_i, X_j) = 0$ para $i \neq j$. Então*

$$Var(X_1 + \cdots + X_n) = \sum_{i=1}^{n} VarX_i.$$

Demonstração.

$$\begin{aligned}
Var(X_1 + \cdots + X_n) &= E(X_1 + \cdots + X_n - E(X_1 + \cdots + X_n))^2 \\
&= E((X_1 - E(X_1)) + \cdots + (X_n - E(X_n)))^2 \\
&= E\left[\sum_{i=1}^{n}(X_i - E(X_i))^2 + 2\sum_{i<j}(X_i - E(X_i))(X_j - E(X_j))\right] \\
&= \sum_{i=j}^{n} VarX_i + 2\sum_{i<j} Cov(X_i, X_j) \\
&= \sum_{i=1}^{n} VarX_i.
\end{aligned}$$

□

Corolário 3.2.22. *Se $X_1, \dots, X_n$ são independentes e integráveis, então*

$$Var(X_1 + \cdots + X_n) = \sum_{i=1}^{n} VarX_i.$$

Capítulo 4

A Lei dos Grandes Números

Consideremos certo experimento básico, com a variável aleatória X representando o valor de um característico numérico do resultado. Pensemos na realização deste experimento n vezes, onde n é um número grande, de tal maneira que as realizações sejam independentes. Suponhamos que depois de cada realização do experimento registre-se o valor de X; este será o valor *observado*. Informalmente, os n observados formam "uma amostra aleatória da variável aleatória X". A Lei dos Grandes Números afirma que a média aritmética dos n valores observados é aproximadamente igual a $E(X)$ quando n é grande; de fato, ela afirma que esta média aritmética das observações *converge*, em certo sentido, para média $E(X)$, quando $n \to \infty$.

Um *experimento composto* consiste em realizar o experimento básico, de forma sucessiva e independente, n vezes. Assim, um resultado possível do experimento composto é uma sequência de n resultados possíveis do experimento básico. Portanto, se Ω_0 é o espaço amostral do experimento básico, então o espaço amostral para o experimento composto é o conjunto de sequências de comprimento n de elementos de Ω_0, isto é,

$$\Omega_n = \Omega_0^n = \{(\omega_1, \dots, \omega_n);\ \omega_i \in \Omega_0,\ i = 1, \dots, n\}.$$

Na verdade, estamos interessados em *passar o limite* no espaço Ω_n quando $n \to \infty$ para aplicar a convergência afirmada pela Lei dos Grandes Números. Por isso, o espaço amostral do experimento composto que vamos considerar consiste nas sequências infinitas de elementos de Ω_0, isto é,

$$\Omega = \{(\omega_1, \omega_2, \dots);\ \omega_i \in \Omega_0,\ i = 1, 2, \dots\} = \Omega_0 \times \Omega_0 \times \cdots = \Omega_0^\infty.$$

Aqui, Ω_n é o resultado das n primeiras realizações do experimento básico. É importante registrar o resultado do n-ésimo experimento básico. Para isso, sendo $\omega = (\omega_1, \omega_2, \dots)$, os valores observados serão $X(\omega_1)$, $X(\omega_2)$, $\dots$. Assim, é conveniente representar por

X_n o valor observado do n-ésimo ensaio, isto é, X_n é função do resultado ω do experimento global, com

$$X_n(\omega) = X(\omega_n),$$

e no decorrer do experimento serão registrados os valores das variáveis aleatórias $X_1, X_2, \ldots$. Como estamos tratando de repetições do mesmo experimento, segue que X_n tem a mesma distribuição de X. Além disso, as variáveis aleatórias X_n são *independentes*, e portanto a sequência $X_1, X_2, \ldots$ é formada por variáveis aleatórias independentes e identicamente distribuídas. Se X_1 é integrável, então todas elas o são, e $E(X_n) = E(X_1)$ para todo n. Neste caso, temos a seguinte versão da Lei dos Grandes Números: Se $X_1, X_2, \ldots$ são independentes, identicamente distribuídas e integráveis, então

$$\frac{X_1 + \cdots + X_n}{n} \to E(X_1).$$

A seguir, vamos definir o tipo de convergência a ser considerado. Sejam $Y, Y_1, Y_2, \ldots$ variáveis aleatórias definidas em um mesmo espaço de probabilidade $(\Omega, \mathcal{A}, \mathcal{P})$.

Definição 4.0.1. *Y_n converge para Y* em probabilidade *se para todo $\epsilon > 0$,*

$$\mathcal{P}(|Y_n - Y| \geq \epsilon) \to 0 \quad \textit{quando} \quad n \to \infty.$$

Notação. $Y_n \xrightarrow{P} Y$.

Definição 4.0.2. *Y_n converge para Y* quase certamente *se*

$$\mathcal{P}(Y_n \to Y \ \ \textit{quando}\ n \to \infty) = 1,$$

isto é, se o evento $A_0 = \{\omega; Y_n(\omega) \to Y(\omega)\}$ tem probabilidade 1.
Notação. $Y_n \xrightarrow{qc} Y$.

Costuma-se dizer que $Y_n(\omega)$ converge para $Y(\omega)$ para "quase todo" ω. Interpretando $\omega \in \Omega$ como um resultado possível de um experimento, a sequência $Y_n(\omega)$ converge para $Y(\omega)$ para quase todo ω quando $Y_n \xrightarrow{qc} Y$.

Por outro lado, convergência em probabilidade apenas afirma que para valores grandes de n as variáveis Y_n e Y são aproximadamente iguais com probabilidade bem alta. Convergência em probabilidade é mais fraca que convergência quase certa, em vista da proposição e do exemplo a seguir.

Proposição 4.0.3. *Se $Y_n \xrightarrow{qc} Y$, então $Y_n \xrightarrow{P} Y$.*

Demonstração. Suponha que $Y_n \xrightarrow{qc} Y$ e seja $\epsilon > 0$. Vamos provar que

$$\mathcal{P}(|Y_n - Y| \geq \epsilon) \to 0.$$

Seja $A_0 = \{\omega; Y_n(\omega) \to Y(\omega)\}$, onde $\mathcal{P}(A_0) = 1$. Para todo $\omega \in A_0$, $|Y_n(\omega) - Y(\omega)| < \epsilon$ para todo n suficientemente grande. Seja A_n o evento "para todo $k \geq n,\ |Y_k - Y| < \epsilon$", isto é,

$$A_n = \bigcap_{k=n}^{\infty} (|Y_k - Y| < \epsilon).$$

Se $\omega \in A_0$, então $\omega \in A_n$ para algum n. Além disso, temos $A_n \subset A_{n+1}$, logo

$$A_0 \subset \bigcup_{n\geq 1} A_n = \lim_{n\to\infty} A_n.$$

Portanto, $1 = \mathcal{P}(A_0) \leq \mathcal{P}(\bigcup_{n\geq 1} A_n)$ e, por continuidade de probabilidade, $\mathcal{P}(A_n) \uparrow 1$. Mas $A_n \subset [|Y_n - Y| < \epsilon]$, logo $\mathcal{P}(|Y_n - Y| < \epsilon) \to 1$ e $\mathcal{P}(|Y_n - Y| \geq \epsilon) = 1 - \mathcal{P}(|Y_n - Y| < \epsilon) \to 0$. □

A recíproca da proposição anterior não é necessariamente verdadeira.

Exemplo 4.0.4. *Seja $X \sim U[0,1]$. Consideremos os seguintes intervalos:*

$$I_1 = [0,1],\ I_2 = [0,1/2],\ I_3 = [1/2,1],\ I_4 = [0,1/4],\ \ldots,\ I_7 = [3/4,1],\ I_8 = [0,1/8],\ \ldots$$

isto é, para $m = 0,1,2,\ldots$ e $i = 0,1,\ldots,2^m - 1$, temos

$$I_{2^m+i} = \left[\frac{i}{2^m}, \frac{i+1}{2^m}\right].$$

Então os 2^m intervalos de comprimento $\frac{1}{2^m}$ cobrem o intervalo $[0,1]$, ao mesmo tempo que seu comprimento fica cada vez menor.

Seja Y_n o indicador do evento $[X \in I_n]$, ou seja,

$$Y_n = \begin{cases} 1 \text{ se } X \in I_n \\ 0 \text{ se } X \notin I_n. \end{cases}$$

Afirmamos que a sequência $Y_1, Y_2, \ldots$ converge em probabilidade para a variável aleatória constante $Y = 0$. De fato, seja $0 < \epsilon \leq 1$. Temos

$$\mathcal{P}(|Y_n - 0| \geq \epsilon) = \mathcal{P}(Y_n = 1) = \mathcal{P}(X \in I_n),$$

e esta probabilidade, que é igual ao comprimento de I_n, converge para zero quando $n \to \infty$. Se $\epsilon > 1$, é impossível que $|Y_n - 0| \geq \epsilon$, logo $\mathcal{P}(|Y_n - 0| \geq \epsilon) = \mathcal{P}(\emptyset) = 0$.

Afirmamos agora que Y_n não converge quase certamente para zero. De fato, não converge em ponto algum, pois qualquer que seja o valor de X, este valor pertence a um ou dois dos 2^n

intervalos de comprimento $\frac{1}{2^n}$*, para todo* n*. Logo, se* $\omega \in \Omega$*,* $Y_n(\omega)$ *assume o valor* 1 *para um número infinito de* $n's$*, assim como assume o valor* 0 *também para um número infinito de* $n's$ *e isso implica que para cada* $\omega \in \Omega$*,* $Y_n(\omega)$ *não converge.*

Agora estamos aptos a formular a Lei dos Grandes Números de uma maneira mais formal, separando em duas classes de lei: fraca e forte, de acordo com o tipo de convergência.

Definição 4.0.5. *Sejam* $X_1, X_2, \ldots$ *variáveis aleatórias integráveis em* $(\Omega, \mathcal{A}, \mathcal{P})$*, e sejam* $S_1, S_2, \ldots$ *as somas parciais, definidas por* $S_n = X_1 + \cdots + X_n$*. Observamos que* $S_1, S_2, \ldots$ *também são variáveis aleatórias em* $(\Omega, \mathcal{A}, \mathcal{P})$*.*

Dizemos que $X_1, X_2, \ldots$ *satisfazem a* Lei Fraca dos Grandes Números *se*

$$\frac{S_n - E(S_n)}{n} \xrightarrow{P} 0,$$

ou equivalentemente, se para todo $\epsilon > 0$*, vale*

$$\mathcal{P}\left(\left|\frac{X_1 + \cdots + X_n - (E(X_1) + \cdots + E(X_n))}{n}\right| \geq \epsilon\right) \to 0.$$

Dizemos que $X_1, X_2, \ldots$ *satisfazem a* Lei Forte dos Grandes Números *se*

$$\frac{S_n - E(S_n)}{n} \xrightarrow{qc} 0,$$

ou equivalentemente, se

$$\frac{(X_1 - E(X_n)) + (X_2 - E(X_2)) + \cdots + (X_n - E(X_n))}{n} \xrightarrow{qc} 0.$$

Pela Proposição 4.0.3, se a sequência $X_1, X_2, \ldots$ satisfaz a Lei Forte, então satisfaz a Lei Fraca, mas a recíproca não é necessariamente verdade pelo Exemplo 4.0.4. Com isso, a Lei Forte é realmente mais "forte".

Intuitivamente, este conceito mais geral da Lei dos Grandes Números pode ser expresso assim: uma sequência de variáveis aleatórias satisfaz a Lei dos Grandes Números se, quando n é grande, a média aritmética dos primeiros n observados é aproximadamente igual à média aritmética das suas esperanças, ou seja, $\dfrac{S_n}{n}$ é aproximadamente igual a $\dfrac{E(S_n)}{n} = \dfrac{E(X_1) + \cdots + E(X_n)}{n}$.

4.1 A Lei Fraca dos Grandes Números

Teorema 4.1.1 (Lei Fraca de Tchebychev). *Sejam $X_1, X_2, \ldots$ variáveis aleatórias não correlacionadas com variâncias finitas e uniformemente limitadas (isto é, existe c finito tal que para todo n, $Var(X_n) \leq c$). Então $X_1, X_2, \ldots$ satisfazem a Lei Fraca dos Grandes Números:*

$$\frac{S_n - E(S_n)}{n} \xrightarrow{P} 0.$$

Demonstração. Seja $\epsilon > 0$. Queremos mostrar que

$$\mathcal{P}\left(\frac{|S_n - E(S_n)|}{n} \geq \epsilon\right) \to 0 \text{ quando } n \to \infty.$$

Pela Proposição 3.2.21, temos $Var(S_n) = Var(X_1 + \cdots + X_n) = \sum_{i=1}^{n} Var(X_i) \leq nc$. Pela desigualdade de Tchebychev temos

$$\mathcal{P}(|S_n - E(S_n)| \geq \epsilon n) \leq \frac{Var(S_n)}{\epsilon^2 n^2} \leq \frac{c}{\epsilon^2 n} \to 0.$$

□

Corolário 4.1.2 (Lei Fraca de Bernoulli). *Consideremos uma sequência de ensaios binomiais independentes, tendo a mesma probabilidade p de "sucesso" em cada ensaio. Se S_n é o número de sucessos nos primeiros n ensaios, então*

$$\frac{S_n}{n} \xrightarrow{P} p.$$

Demonstração. Seja

$$X_n = \begin{cases} 1 \text{ se o } n\text{-ésimo ensaio é sucesso,} \\ 0 \text{ se o } n\text{-ésimo ensaio é fracasso,} \end{cases}$$

onde $\mathcal{P}(X_n = 1) = p$ e $\mathcal{P}(X_n = 0) = 1 - p$. Então $X_1, X_2, \ldots$ são independentes, identicamente distribuídas e integráveis, com média $\mu = p$. Como $Var(X_n) = p(1-p)$ é uniformemente limitada, a Lei Fraca de Tchebychev implica que

$$\frac{S_n - np}{n} \xrightarrow{P} 0.$$

ou, equivalentemente,

$$\frac{S_n}{n} \xrightarrow{P} p.$$

□

4.2 O Lema de Borel-Cantelli

Vamos apresentar agora o Lema de Borel-Cantelli, que será uma peça importante na prova da Lei Forte. Mas antes, precisamos definir os limites superior e inferior de uma sequência de eventos.

Definição 4.2.1. *Seja $A_1, A_2, \ldots$ uma sequência de eventos. O* limite superior *da sequência é definido por*

$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k;$$

o limite inferior *é definido por*

$$\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k.$$

Por conveniência, vamos abreviar a notação como $\limsup A_n$ *e* $\liminf A_n$.

Intuitivamente, temos que o evento $\limsup A_n$ é o evento "ocorrência de um número infinito dos A_n", ou seja, ocorre A_n para infinitos n's. O evento $\liminf A_n$ também tem uma interpretação intuitiva, é o evento "ocorrência de A_n para todo n suficientemente grande".

Como a ocorrência de A_n para todo n suficientemente grande implica que A_n ocorre infinitas vezes, temos $\liminf A_n \subset \limsup A_n$. É comum definir o limite de A_n, $\lim A_n$, como $\liminf A_n$ se $\liminf A_n = \limsup A_n$.

Por vezes, vamos denotar $\limsup A_n = [A_n \text{ infinitas vezes}]$.

Proposição 4.2.2 (Lema de Borel-Cantelli)**.** *Considere a sequência de eventos $A_1, A_2, \ldots$ definida em $(\Omega, \mathcal{A}, \mathcal{P})$.*

(a) Se $\sum_{n=1}^{\infty} \mathcal{P}(A_n) < \infty$, então $\mathcal{P}(A_n \text{ infinitas vezes}) = 0$.

(b) Se $\sum_{n=1}^{\infty} \mathcal{P}(A_n) = \infty$, e os A_n são independentes, então $\mathcal{P}(A_n \text{ infinitas vezes}) = 1$.

O item (b) pode não ser verdadeiro se a hipótese de independência for removida. Por exemplo, seja $A_n = A$ para todo n, onde $0 < \mathcal{P}(A) < 1$. Então $\sum \mathcal{P}(A_n) = \infty$, mas o evento $[A_n \text{ infinitas vezes}] = A$ e $\mathcal{P}(A_n \text{ infinitas vezes}) = \mathcal{P}(A) < 1$.

Demonstração. (a) Se $\sum \mathcal{P}(A_n) < \infty$, então $\sum_{k=n}^{\infty} \mathcal{P}(A_k) \to 0$ quando $n \to \infty$. Mas

$[A_n \text{ infinitas vezes}] \subset \bigcup\limits_{k=n}^{\infty} A_k$ para todo n, logo

$$\mathcal{P}(A_n \text{ infinitas vezes}) \leq \mathcal{P}\left(\bigcup_{k=n}^{\infty} A_k\right) \leq \sum_{k=n}^{\infty} \mathcal{P}(A_k) \to 0.$$

Portanto, $\mathcal{P}(A_n \text{ infinitas vezes}) = 0$.

(b) Basta provar que $\mathcal{P}\left(\bigcup\limits_{k=n}^{\infty} A_k\right) = 1 \ \forall n$ pois $[A_n \text{ infinitas vezes}] = \bigcap\limits_{n=1}^{\infty} \bigcup\limits_{n=k}^{\infty} A_k$ e a interseção de um número enumerável de eventos de probabilidade um, é também de probabilidade um. Seja $B_n = \bigcup\limits_{k=n}^{\infty} A_k$. Então $B_n \supset \bigcup\limits_{k=n}^{n+m} A_k$ para todo m, e

$$B_n^c \subset \left(\bigcup_{k=n}^{n+m} A_k\right)^c = \bigcap_{k=n}^{n+m} A_k^c.$$

Aplicando a probabilidade na expressão anterior, obtemos para todo m,

$$1 - \mathcal{P}(B_n) = \mathcal{P}(B_n^c) \leq \mathcal{P}\left(\bigcap_{k=n}^{n+m} A_k^c\right) = \prod_{k=n}^{n+m} \mathcal{P}(A_k^c) = \prod_{k=n}^{n+m} (1 - \mathcal{P}(A_k)).$$

Figura 4.1: Gráfico das funções $f(x) = 1 - x$ e $g(x) = e^{-x}$

Pelo gráfico, temos $1 - p \leq e^{-p}$ para $0 \leq p \leq 1$, logo

$$1 - \mathcal{P}(B_n) \leq \prod_{k=n}^{n+m} e^{-\mathcal{P}(A_k)} = e^{\left\{-\sum\limits_{k=n}^{n+m} \mathcal{P}(A_k)\right\}} \to 0$$

quando $m \to \infty$, pois $\sum\limits_{k=n}^{n+m} \mathcal{P}(A_k) \to +\infty$ quando $m \to \infty$. Logo $\mathcal{P}(B_n) = 1 \ \forall n$, como queríamos demonstrar. $\square$

Corolário 4.2.3. *Consideremos uma sequência de ensaios binomiais independentes com probabilidade p_n de sucesso no n-ésimo ensaio. Seja*

$$X_n = \begin{cases} 1 \text{ se o } n\text{-ésimo ensaio é sucesso} \\ 0 \text{ se é fracasso.} \end{cases}$$

Então, vale o seguinte:

1. *Se* $\sum_{n=1}^{\infty} p_n = +\infty$, *então* $\mathcal{P}\left(\sum_{n=1}^{\infty} X_n = \infty\right) = 1$.
2. *Se* $\sum_{n=1}^{\infty} p_n < +\infty$, *então* $\mathcal{P}\left(\sum_{n=1}^{\infty} X_n < \infty\right) = 1$.

Em outras palavras,

$$\sum p_n < \infty \Rightarrow \text{um número finito de sucessos, quase certamente}$$
$$\sum p_n = \infty \Rightarrow \text{um número infinito de sucessos, quase certamente.}$$

Demonstração. Seja $A_n = [X_n = 1]$. Então $\mathcal{P}(A_n) = p_n$ e os A_n são independentes. Pelo Lema de Borel-Cantelli,

$$\sum p_n < \infty \Rightarrow \mathcal{P}(A_n \text{ infinitas vezes}) = 0 \text{ (item (a))}$$
$$\sum p_n = \infty \Rightarrow \mathcal{P}(A_n \text{ infinitas vezes}) = 1 \text{ (item (b)).}$$

Mas $[A_n$ infinitas vezes$]^c$ = "apenas um número finito de sucessos entre todos os ensaios". Portanto,

$$[A_n \text{ infinitas vezes}] = \left[\sum X_n = \infty\right],$$
$$[A_n \text{ infinitas vezes}]^c = \left[\sum X_n < \infty\right].$$

□

Exemplo 4.2.4. *Ao colocar um macaco diante de um computador, é razoável supor que haja uma probabilidade positiva, embora muito pequena, dele digitar as obras completas de Shakespeare sem erro. Vamos chamar o primeiro ensaio de* sucesso *se ele realizar essa façanha, e de* fracasso *quando ele comete o primeiro erro. No final do primeiro ensaio, que provavelmente chegará logo, vamos alimentá-lo ou levá-lo para dar uma volta (para garantir a independência dos ensaios) e começar o segundo ensaio. O processo segue assim sucessivamente. Como* $p_n = p > 0$ *para todo* n*, temos* $\sum p_n = \infty$. *Pelo corolário anterior, há probabilidade* 1 *dele escrever as obras completas de Shakespeare um número infinito de vezes.*

Exemplo 4.2.5. *O exemplo anterior pode ser adaptado para loterias. Se jogarmos na loteria com uma frequência determinada (não necessariamente os mesmos números), com probabilidade* 1 *vamos ganhar infinitas vezes. O grande problema é o tempo que isso vai levar. Provavelmente, seguindo esse processo, pode ser que a primeira vez que ganharmos já teríamos morrido ou falido, pois levaríamos milhares ou milhões de anos ou teríamos gastado todo o dinheiro disponível no mundo.*

4.3 A Lei Forte dos Grandes Números

Proposição 4.3.1 (Desigualdade de Kolmogorov). *Sejam* $X_1, \dots, X_n$ *variáveis aleatórias independentes tais que* $E(X_k) = 0$ *e* $Var(X_k) < \infty$, $k = 1, \dots, n$, *e sejam* $S_k = X_1 + \cdots + X_k$ *e* $\lambda > 0$. *É válida a seguinte desigualdade:*

$$\mathcal{P}\left(\max_{1\leq k\leq n} |S_k| \geq \lambda\right) \leq \frac{1}{\lambda^2}Var(S_n) = \frac{1}{\lambda^2}\sum_{k=1}^{n} Var(X_k).$$

Demonstração. Seja $A = [\max S_k^2 \geq \lambda^2]$, e sejam

$$\begin{aligned}
&A_1 = [S_1^2 \geq \lambda^2],\\
&A_2 = [S_1^2 < \lambda^2, S_2^2 \geq \lambda^2],\\
&\quad \dots\\
&A_k = [S_1^2 < \lambda^2, \dots, S_{k-1}^2 < \lambda^2, S_k^2 \geq \lambda^2], \text{ para } 2 \leq k \leq n,
\end{aligned}$$

de forma que $A = \bigcup\limits_{k=1}^{n} A_k$ e os $A_{k's}$ são dois a dois disjuntos. Intuitivamente, A_k é o evento que indica que a primeira vez que $S_j^2 \geq \lambda^2$ ocorre para $j = k$. Dessa forma, temos $I_A = \sum\limits_{k=1}^{n} I_{A_k}$, logo

$$S_n^2 \geq S_n^2 I_A = \sum_{k=1}^{n} S_n^2 I_{A_k} \Rightarrow E(S_n^2) \geq \sum_{k=1}^{n} E(S_n^2 I_{A_k}). \tag{4.1}$$

Além disso,

$$S_n^2 = (S_n - S_k + S_k)^2 = (S_n - S_k)^2 + S_k^2 + 2(S_n - S_k)S_k \geq S_k^2 + 2(S_n - S_k)S_k.$$

Multiplicando por I_{A_k} e tomando a esperança na desigualdade acima, temos que

$$E(S_n^2 I_{A_k}) \geq E(S_k^2 I_{A_k}) + 2E((S_n - S_k)S_k I_{A_k}).$$

Notemos que $S_kI_{A_k}$ só depende de $X_1, \ldots, X_k$, enquanto $S_n - S_k = X_{k+1} + \cdots + X_n$ depende de $X_{k+1}, \ldots, X_n$, logo $S_kI_{A_k}$ e $S_n - S_k$ são independentes. Pela Proposição 3.2.17, segue que $E((S_n - S_k)S_kI_{A_k}) = E(S_n - S_k)E(S_kI_{A_k}) = 0$ pois $E(S_n) = E(S_k)$. Como $S_k^2 \geq \lambda^2$ em A_k e usando o Corolário 3.1.12 segue que

$$E(S_n^2 I_{A_k}) \geq E(S_k^2 I_{A_k}) \geq E(\lambda^2 I_{A_k}) = \lambda^2 E(I_{A_k}) = \lambda^2 \mathcal{P}(A_k).$$

Pela desigualdade em (4.1), obtemos pela aditividade da probabilidade que

$$E(S_n^2) \geq \sum_{k=1}^{n} \lambda^2 \mathcal{P}(A_k) = \lambda^2 \mathcal{P}(A),$$

de onde segue a proposição. $\square$

Teorema 4.3.2 (Primeira Lei Forte de Kolmogorov). *Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes e integráveis, e suponha que*

$$\sum_{n=1}^{\infty} \frac{Var(X_n)}{n^2} < +\infty.$$

Então os X_n satisfazem a Lei Forte dos Grandes Números, isto é,

$$\frac{X_1 + \cdots + X_n}{n} - \frac{(E(X_1) + \cdots + E(X_n))}{n} \xrightarrow{qc} 0.$$

Demonstração. Se $E(X_n) \neq 0$ então tome $Y_n = X_n - E(X_n)$. Então $E(Y_n) = 0$, $Var(Y_n) = Var(X_n)$ e as Y_n satisfazem as hipóteses do teorema. Se o teorema vale para o caso $E(Y_n) = 0$, então

$$\frac{Y_1 + \cdots + Y_n}{n} \xrightarrow{qc} 0,$$

isto é,

$$\frac{(X_1 - E(X_1)) + \cdots + (X_n - E(X_n))}{n} \xrightarrow{qc} 0.$$

Vamos supor, portanto $E(X_n) = 0$ para todo n. Queremos mostrar que $\frac{S_n}{n} \xrightarrow{qc} 0$, onde $S_n = X_1 + \cdots + X_n$. Basta mostrar que

$$M_n = \max_{2^n < k \leq 2^{n+1}} \frac{|S_k|}{k} \xrightarrow{qc} 0, \quad \text{quando} \quad n \to \infty.$$

Usando a desigualdade de Kolmogorov e o Lema de Borel-Cantelli respectivamente, vamos provar que

(i) $\sum_{n=1}^{\infty} \mathcal{P}\left(M_n \geq \frac{1}{m}\right) < \infty, \forall m = 1, 2, \ldots,$ e

(ii) $M_n \xrightarrow{qc} 0$.

Prova de (i). Seja m fixo. Como $\dfrac{|S_k|}{k} < \dfrac{|S_k|}{2^n}$ para $k > 2^n$ temos

$$\begin{aligned}\mathcal{P}\left(\max_{2^n<k\le 2^{n+1}} \frac{|S_k|}{k} \ge \frac{1}{m}\right) &\le \mathcal{P}\left(\max_{2^n<k\le 2^{n+1}} |S_k| \ge \frac{2^n}{m}\right)\\ &\le \mathcal{P}\left(\max_{1\le k\le 2^{n+1}} |S_k| \ge \frac{2^n}{m}\right) \le \frac{m^2}{4^n}\sum_{k=1}^{2^{n+1}} Var(X_k),\end{aligned}$$

pela desigualdade de Kolmogorov. Seja

$$A_n = \left[\max_{2^n<k\le 2^{n+1}} \frac{|S_k|}{k} \ge \frac{1}{m}\right] = \left[M_n \ge \frac{1}{m}\right].$$

Temos

$$\begin{aligned}\sum_{n=1}^{\infty} \mathcal{P}(A_n) &\le m^2 \sum_{n=1}^{\infty}\left(\frac{1}{4^n}\sum_{k=1}^{2^{n+1}} Var(X_k)\right)\\ &= m^2 \sum_{k=1}^{\infty} \sum_{n\ge\lceil \log_2(k)\rceil-1}\left(\frac{1}{4^n} Var(X_k)\right)\\ &= m^2 \sum_{k=1}^{\infty}\left(Var(X_k)\sum_{n\ge\lceil \log_2(k)\rceil-1}\frac{1}{4^n}\right).\end{aligned}$$

Como

$$\sum_{n=j}^{\infty}\frac{1}{4^n} = \frac{1}{4^j}\left(1+\frac{1}{4}+\left(\frac{1}{4}\right)^2+\dots\right) = \frac{1}{4^j}\cdot\frac{4}{3}$$

temos

$$\sum_{n\ge\lceil \log_2(k)\rceil-1}\frac{1}{4^n} = \frac{4}{3}\cdot\frac{1}{4^{n\ge\lceil \log_2(k)\rceil-1}} \le \frac{4}{3}\cdot\frac{1}{4^{\log_2(k)-1}} = \frac{16}{3k^2}.$$

Portanto, por hipótese temos

$$\sum_{n=1}^{\infty}\mathcal{P}(A_n) \le \frac{16m^2}{3}\sum_{k=1}^{\infty}\frac{Var(X_k)}{k^2} < \infty.$$

Prova de (ii). Com a mesma notação de (i), temos $\mathcal{P}(A_n \text{ infinitas vezes}) = 0$ por Borel-Cantelli. Em outras palavras, para todo m fixo, a probabilidade de que M_n assuma um valor $\ge 1/m$ infinitas vezes é 0. Isto significa que para todo m, a probabilidade de que M_n assuma um valor $\ge 1/m$ para somente um número finito de n's é 1. Considere

$B_n =$ "M_n assume um valor $\geq \frac{1}{m}$ para somente um número finito de n's".

Temos que $\mathcal{P}(B_m) = 1\ \forall m$ e $\mathcal{P}\left(\bigcap\limits_{m=1}^{\infty} B_m\right) = 1$.

Agora basta notar que o evento $\bigcap\limits_{m=1}^{\infty} B_m$ é equivalente ao evento $[M_n \to 0]$. De fato, temos $M_n \geq 0$ e portanto:

$$\begin{aligned}\omega \in \bigcap_{m=1}^{\infty} B_m &\Leftrightarrow \forall m,\ M_n(\omega) \geq 1/m \ \text{ para somente um número finito de } \ n's \\ &\Leftrightarrow \forall m,\ 0 \leq M_n(\omega) < 1/m \ \text{ para todo } \ n \ \text{ suficientemente grande} \\ &\Leftrightarrow M_n(\omega) \to 0 \ \text{ quando } \ n \to \infty.\end{aligned}$$

□

Lema 4.3.3. *Seja X uma variável aleatória integrável com função de distribuição F. Então,*

$$\sum_{n=1}^{\infty}\left\{\frac{1}{n^2}\int_{-n}^{n} x^2 dF(x)\right\} < \infty.$$

Lembramos aqui que $\int\limits_{-n}^{n} = \int_{(-n,n]}$.

Demonstração. Vamos observar primeiramente, usando o teste da integral, que para $j = 1, 2, \ldots$

$$\sum_{n=j}^{\infty}\frac{1}{n^2} \leq \frac{1}{j^2} + \int_{j}^{\infty}\frac{1}{x^2}dx = \frac{1}{j^2} + \left[-\frac{1}{x}\right]_j^{\infty} = \frac{1}{j^2} + \frac{1}{j} \leq \frac{2}{j}.$$

Como

$$\int_{-n}^{n} x^2 dF(x) = \sum_{-n+1}^{n}\int_{j-1}^{j} x^2 dF(x),$$

temos

$$\begin{aligned}\sum_{n=1}^{\infty}\left\{\frac{1}{n^2}\int_{-n}^{n} x^2 dF(x)\right\} &= \sum_{n=1}^{\infty}\sum_{j=-n+1}^{n}\left\{\frac{1}{n^2}\int_{j-1}^{j} x^2 dF(x)\right\} \\ &= \sum_{j=1}^{\infty}\sum_{n=j}^{\infty}\left\{\frac{1}{n^2}\int_{j-1}^{j} x^2 dF(x)\right\} + \sum_{j=-\infty}^{0}\sum_{n=|j|+1}^{\infty}\left\{\frac{1}{n^2}\int_{j-1}^{j} x^2 dF(x)\right\} \\ &\leq 2\sum_{j=1}^{\infty}\int_{j-1}^{j}\frac{x^2}{j}dF(x) + 2\sum_{j=-\infty}^{0}\int_{j-1}^{j}\frac{x^2}{|j|+1}dF(x).\end{aligned}$$

Como $\frac{x^2}{j} \leq x$ em $(j-1, j]$, para $j \geq 1$, e $\frac{x^2}{|j|+1} \leq |x|$ em $(j-1, j]$, para $j \leq 0$, temos

$$\begin{aligned}\sum_{n=1}^{\infty} \left\{ \frac{1}{n^2} \int_{-n}^{n} x^2 dF(x) \right\} &\leq 2 \sum_{j=1}^{\infty} \int_{j=1}^{j} x dF(x) + 2 \sum_{j=-\infty}^{0} \int_{j-1}^{j} |x| dF(x) \\ &= 2 \sum_{j=-\infty}^{\infty} \int_{j-1}^{j} |x| dF(x) \\ &= 2 \int_{-\infty}^{\infty} |x| dF(x) = 2E(|X|) < \infty.\end{aligned}$$

□

Lema 4.3.4. *Seja* $(a_n)_{n\in\mathbb{N}}$ *uma sequência de números reais. Se* $\lim a_n = 0$ *então*

$$\lim \frac{a_1 + a_2 + \cdots + a_n}{n} = 0.$$

Demonstração. Seja $\epsilon > 0$ e tome $n_0 \in \mathbb{N}$ de forma que se $n > n_0$ então $|a_n| < \epsilon$. Pela desigualdade triangular, se $n > n_0$ temos

$$\begin{aligned}\left| \frac{a_1 + \cdots + a_n}{n} \right| &\leq \left| \frac{a_1 + \cdots + a_{n_0}}{n} \right| + \left| \frac{a_{n_0+1} + \cdots + a_n}{n} \right| \\ &\leq \left| \frac{a_1 + \cdots + a_{n_0}}{n} \right| + \frac{\epsilon + \cdots + \epsilon}{n} \\ &\leq \left| \frac{a_1 + \cdots + a_{n_0}}{n} \right| + \frac{n - n_0}{n}\epsilon\end{aligned}$$

Tomando $n \to \infty$, segue que $\left| \frac{a_1 + \cdots + a_n}{n} \right| \to 0.$ □

Teorema 4.3.5 (A Lei Forte de Kolmogorov)**.** *Sejam* $X_1, X_2, \ldots$ *variáveis aleatórias independentes, identicamente distribuídas e integráveis, com* $E(X_n) = \mu$*. Então*

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow{qc} \mu.$$

Demonstração. Como no teorema anterior, basta supor $\mu = 0$.

Definimos os *truncamentos* de X_n como

$$Y_n = X_n I_{[-n < X_n \leq n]} = \begin{cases} X_n & \text{se } -n < X_n \leq n \\ 0 & \text{caso contrário,} \end{cases}$$

isto é, $Y_n = X_n I_{[-n<X_n\leq n]}$.

Definimos $Z_n = X_n - Y_n$, de modo que $X_n = Y_n + Z_n$ e

$$\frac{X_1+\cdots+X_n}{n} = \frac{Y_1+\cdots+Y_n}{n} + \frac{Z_1+\cdots+Z_n}{n}.$$

Vamos dividir a demonstração em 3 partes:

(a) $\dfrac{Z_1+\cdots+Z_n}{n} \xrightarrow{qc} 0$ (usaremos Borel-Cantelli),

(b) $\dfrac{Y_1+\cdots+Y_n}{n} - \dfrac{E(Y_1)+\cdots+E(Y_n)}{n} \xrightarrow{qc} 0$ (Lema 4.3.3 e Teorema 4.3.2),

(c) $\dfrac{E(Y_1)+\cdots+E(Y_n)}{n} \to 0$ (pelo Teorema da Convergência Dominada).

Notemos que (a), (b) e (c) implicam o teorema, pois se

$$A = \left[\frac{Z_1+\cdots+Z_n}{n} \to 0\right] \text{ e } B = \left[\frac{Y_1+\cdots+Y_n}{n} - \frac{E(Y_1)+\cdots+E(Y_n)}{n} \to 0\right],$$

então somando os três termos na interseção dos eventos quase certos de (a) e (b), segue que

$$\mathcal{P}(A) = 1, \mathcal{P}(B) = 1, \mathcal{P}(A\cap B) = 1 \text{ e } \frac{X_1+\cdots+X_n}{n} \to 0 \text{ em } A\cap B.$$

Prova de (a). Notemos que $Z_n \neq 0 \Leftrightarrow Y_n \neq X_n \Leftrightarrow X_n \notin (-n,n]$. Logo,

$$\mathcal{P}(Z_n \neq 0) = \mathcal{P}(X_n \notin (-n,n]) \leq \mathcal{P}(|X_n| \geq n).$$

Seja, $A_n = [Z_n \neq 0]$. Como as $X_{n's}$ são identicamente distribuídas, temos pelo Corolário 3.1.11 que

$$\sum_{n=1}^{\infty} \mathcal{P}(A_n) \leq \sum_{n=1}^{\infty} (\mathcal{P}|X_n| \geq n) = \sum_{n=1}^{\infty} \mathcal{P}(|X_1| \geq n) = E(|X_1|) < \infty.$$

Pelo Lema de Borel-Cantelli segue que $\mathcal{P}(A_n$ infinitas vezes$) = \mathcal{P}(Z_n \neq 0$ infinitas vezes$) = 0$. Tomando o complementar, obtemos

$$\mathcal{P}(Z_n \neq 0 \text{ para apenas um número finito de } n) = 1,$$

isto é,

$$\mathcal{P}(Z_n = 0 \text{ para todo } n \text{ suficientemente grande}) = 1.$$

Pelo Lema de Borel-Cantelli, se $Z_n = 0$ para n suficientemente grande, então $Z_n \to 0$ e, pelo lema anterior, $\dfrac{Z_1+\cdots+Z_n}{n} \to 0$, logo $\mathcal{P}\left(\dfrac{Z_1+\cdots+Z_n}{n} \to 0\right) = 1$.

Prova de (b). Seja F a função de distribuição de todos os $X_{n's}$, ou seja $F = F_{X_n}$. Vamos mostrar que valem as hipóteses da primeira Lei Forte de Kolmogorov para as variáveis aleatórias Y_n. Temos

$$Var(Y_n) \leq E(Y_n^2) = E(X_n^2 I_{[-n<X_n\leq n]}) = \int x^2 I_{[-n,n]}(x) dF(x) = \int_{-n}^{n} x^2 dF(x),$$

logo

$$\sum_{n=1}^{\infty} \frac{Var(Y_n)}{n^2} \leq \sum_{n=1}^{\infty} \frac{1}{n^2} \int_{-n}^{n} x^2 dF(x) < \infty.$$

Daí segue que as $Y_{n's}$ satisfazem a primeira Lei Forte, o que prova (b).

Prova de (c). Pelo Lema 4.3.4 basta mostrar que $E(Y_n) \to 0$. Notemos que $|X_1 I_{[-n<X_1\leq n]}| \leq |X_1|$, que é integrável por hipótese. Além disso, $[-n < X_1 \leq n] = \{\omega; -n < X_1(\omega) \leq n\} \uparrow \Omega$, de modo que $I_{[-n<X_1\leq n]} \to 1\ \forall\omega$, ou seja, $X_1 I_{[-n<X_1\leq n]} \to X_1$. Como as $X_{n's}$ são identicamente distribuídas, segue pelo Teorema da Convergência Dominada que

$$E(Y_n) = E(X_n I_{[-n<X_n\leq n]}) = E(X_1 I_{[-n<X_1\leq n]}) \to E(X_1) = 0.$$

□

Corolário 4.3.6 (Lei Forte de Borel, 1909)**.** *Sendo $X_1, X_2, \ldots$ variáveis aleatórias independentes e identicamente distribuídas tais que $\mathcal{P}(X_n = 1) = p,\ \mathcal{P}(X_n = 0) = 1 - p$. Portanto $\frac{S_n}{n} \xrightarrow{qc} p$, onde $S_n = X_1 + \cdots + X_n$.*

Capítulo 5

Números Normais e pseudo-aleatoriedade

5.1 Números normais

Seja $x \in [0,1]$ escrito na base $b \geq 2$ como $x = 0, x_1x_2x_3\ldots$, isto é, $x = \sum\limits_{n=1}^{\infty} \frac{x_n}{b^n}$. Notemos que tal expansão pode não ser única: de fato, alguns racionais possuem duas expansões. Por exemplo, na base 10, temos $0,1999\cdots = 0,2 = 0,2000\ldots$. Quando este for o caso, tomemos a expansão que termina com $0's$.

Seja $A = (a_{k-1}\ldots a_0)_b = a_{k-1}b^{k-1} + \cdots + a_1 b + a_0$ um número inteiro com no máximo k dígitos na expansão em base b. Denotamos A como um bloco de tamanho k. Seja $N(A; n; x)$ o número de vezes em que o bloco A aparece entre os primeiros n dígitos do número x.

Exemplo 5.1.1. *Fixado* $b = 10,\ N(12; 20; 0,12345678910111213141 5\ldots) = 2$.

Fixado $b = 2,\ N(1; 10; 0,101001000100001\ldots) = 4$.

Definição 5.1.2. *Dizemos que x é* normal *na base b se para todo $k \geq 1$ e todo bloco A de tamanho k, a frequência relativa de aparições de A na expansão de x é a mesma. Mais formalmente, x é normal na base b se*

$$\lim_{n\to\infty} \frac{N(A; n; x)}{n} = \frac{1}{b^k}.$$

Exemplo 5.1.3. *Números racionais não são normais em nenhuma base, pois: (i) a expansão "termina" e a partir de um determinado ponto todos os algarismos são iguais a zero; (ii) a expansão sendo uma dízima periódica, ela se repete, o que implica em convergência para* 1 *do bloco que se repete. Para outros blocos, a convergência pode ser para outros valores. Por exemplo, no número racional* $0,123123123123123\ldots$*, a frequência dos blocos* 123, 231 *e* 312 *tende a* 1*, enquanto a frequência do bloco* 1234 *é* 0*; além disso, a frequência do bloco* 23 *tende a* $\frac{1}{3}$*, enquanto a frequência do bloco* 24 *é* 0.

Foi provado em [5] que o número $0,12345678910111213141516\ldots$ formado pela concatenação dos números naturais após a vírgula é normal na base 10. Analogamente, pode-se provar que números desse mesmo formato em outras bases também são normais nessas bases.

O próximo resultado, devido a Borel, é dos principais teoremas desse trabalho.

Teorema 5.1.4 (Borel [3]). *Fixada uma base b, "quase todos" os números reais são normais na base b, onde usamos a expressão "quase todos" no sentido de Lebesgue.*

Demonstração. Sejam $\Omega = [0,1]$, $\mathcal{A}$ a σ-álgebra dos borelianos de $[0,1]$ e $\mathcal{P}$ a probabilidade uniforme em $[0,1]$, que coincide com a medida de Lebesgue (ver Apêndice A). Por simplicidade, vamos provar somente alguns casos desse teorema, mas deixando claro que o método funciona em qualquer caso, bastando poucas modificações no argumento. A primeira simplificação que faremos é considerar $b = 2$ (para outras bases, o resultado é completamente análogo). Vamos considerar também blocos de tamanhos 1 e 2.

Para blocos de tamanho 1 na base 2, as únicas possibilidades são 0 e 1. Seja $x = 0,x_1x_2x_3\cdots = \sum\limits_{n=1}^{\infty} \dfrac{x_n}{2^n} \in [0,1]$, onde $x_i \in \{0,1\}$. Vamos definir a seguinte sequência de variáveis aleatórias:

$$X_1(x) = \begin{cases} 0 \text{ se } x \in [0,1/2), \\ 1 \text{ se } x \in [1/2,1], \end{cases}$$

$$X_2(x) = \begin{cases} 0 \text{ se } x \in [0,1/4) \cup [1/2,3/4), \\ 1 \text{ se } x \in [1/4,1/2) \cup [3/4,1], \end{cases}$$

$$\cdots$$

Notemos que $X_n(x) = x_n$. Além disso, $\mathcal{P}(X_n = 0) = 1/2 = \mathcal{P}(X_n = 1)$, logo as variáveis aleatórias X_n são identicamente distribuídas. Nesse caso, essas variáveis também são independentes. De fato, por exemplo, $\mathcal{P}(X_1 = 1 \text{ e } X_2 = 0) = \mathcal{P}([1/2,3/4)) = 1/4 = \mathcal{P}(X_1 = 1)\mathcal{P}(X_2 = 0)$ (para os demais caso é completamente análogo). Além disso, $E(X_n) = 1/2$ para todo n, logo a Lei Forte (Corolário 4.3.6) nos garante que $\frac{X_1+\cdots+X_n}{n} \xrightarrow{qc} \frac{1}{2}$, isto é, $\frac{x_1+\cdots+x_n}{n} \to \frac{1}{2}$ para quase todo x.

Para blocos de tamanhos maiores, só precisamos de uma pequena modificação no argumento. Por exemplo, se queremos saber a frequência relativa do bloco 10, consideramos a seguinte sequência:

$$y_n = \begin{cases} 1 \text{ se } x_n = 1 \text{ e } x_{n+1} = 0, \\ 0 \text{ caso contrário.} \end{cases}$$

Por exemplo, se $x = 0,11100010110001\ldots$ então $y_1 = y_2 = 0,\ y_3 = 1,\ y_4 = y_5 = y_6 = 0,\ y_7 = 1, \ldots$. Para identificar a frequência relativa, basta calcular o valor de $\frac{y_1+\cdots+y_n}{n}$ quando n é grande. Para isso, seja $Y_n = X_n(1-X_{n+1})$ para $n \geq 1$. Podemos verificar que essa sequência de variáveis aleatórias é identicamente distribuída de forma análoga ao caso anterior, mas elas não são independentes. De fato, temos $\mathcal{P}(Y_n = 1) = \mathcal{P}(X_n = 1 \text{ e } X_{n+1} = 0) = 1/4$, logo $\mathcal{P}(Y_n = 0) = \frac{3}{4}$. Por outro lado, $\mathcal{P}(Y_1 = 1 \text{ e } Y_2 = 1) = \mathcal{P}(x_1 = 1 \text{ e } x_2 = 0 \text{ e } x_2 = 1 \text{ e } x_3 = 0) = 0 \neq \frac{1}{16} = \mathcal{P}(Y_1 = 1)\mathcal{P}(Y_2 = 1)$. Informalmente, podemos pensar que x_n influencia em Y_{n-1} e em Y_n. Mas isso pode ser contornado tomando duas subsequências de Y_n: as de índice par (Y_{2n}) e as de índice ímpar (Y_{2n+1}). Similar ao caso anterior, é possível mostrar que separadamente, cada uma dessas subsequências são independentes. Pela Lei Forte, obtemos

$$\frac{Y_1 + Y_3 + \cdots + Y_{2n-1}}{n} \xrightarrow{qc} \frac{1}{4}$$

$$\frac{Y_2 + Y_4 + \cdots + Y_{2n}}{n} \xrightarrow{qc} \frac{1}{4}.$$

O restante do argumento segue da seguinte forma: se calcularmos a frequência até um índice par, teremos

$$\frac{Y_1 + Y_2 + \cdots + Y_{2n}}{2n} = \frac{1}{2}\left(\frac{Y_1 + Y_3 + \cdots + Y_{2n-1}}{n} + \frac{Y_2 + Y_4 + \cdots + Y_{2n}}{n}\right) \xrightarrow{qc} \frac{1}{4}.$$

Se calcularmos a frequência até um índice ímpar, usamos o índice par anterior para mostrar que os limites são iguais. De fato, usando a desigualdade triangular, obtemos

$$\begin{aligned}\left|\frac{Y_1 + \cdots + Y_{2n+1}}{2n+1} - \frac{Y_1 + \cdots + Y_{2n}}{2n}\right| &= \frac{|2n(Y_1 + \cdots + Y_{2n+1}) - (2n+1)(Y_1 + \cdots + Y_{2n})|}{2n(2n+1)} \\ &= \frac{|2nY_{2n+1} - (Y_1 + \cdots + Y_{2n})|}{2n(2n+1)} \leq \frac{1}{2n+1} \to 0.\end{aligned}$$

□

Definição 5.1.5. *Se x é normal na base b para todo $b \geq 2$, então x é dito* absolutamente normal.

Corolário 5.1.6. *Quase todos os números do intervalo $[0,1]$ são absolutamente normais.*

Demonstração. Seja $A_b = [x \text{ é normal na base } b]$. Pelo Teorema 5.1.4, temos $\mathcal{P}(A_b) = 1$ para todo $b \geq 2$. Pela Proposição 1.1.14, segue que $\mathcal{P}(\bigcap_{b\geq 2} A_b) = 1$. □

A noção contra intuitiva do último resultado vem do fato de que ninguém até hoje consegue encontrar um exemplo de número absolutamente normal, apesar de quase todos os números reais serem dessa forma.

5.2 Números aleatórios e pseudoaleatórios

Um problema fundamental na Matemática, na Estatística e na Computação é a geração de números aleatórios, que possuem diversas aplicações práticas, dentre as quais podemos destacar o Método de Monte-Carlo [8, 9], que é uma classe de métodos estatísticos que se baseia em amostragens aleatórias muito grandes para obter resultados numéricos. A princípio, podem ser utilizados para resolver quaisquer problemas com interpretação probabilística . Números absolutamente normais são candidatos ideais a serem números aleatórios, mas ainda não os conhecemos concretamente. Sendo assim, devemos nos contentar com números pseudoaleatórios, que são números "aparentemente" aleatórios, mas que são gerados de uma forma determinística. Intuitivamente, todos os números que conseguimos pensar são pseudoaleatórios.

5.3 Blocos de dígitos iguais

Nessa seção, usamos o website 4Devs Ferramentas Online [1] para gerar o seguinte número binário com 200 dígitos:

$$\begin{gathered}
00100011111001100011011000101101101101101111000100 \\
11101001000110110001101010001111001010110010001000 \\
01001000111001100010010101110100101111101000010001 \\
10111100100110001100010011000001001011110000000101
\end{gathered}$$

Notemos que os 10 últimos dígitos são 0000000101, ou seja, no número gerado há um bloco de tamanho 7 formado apenas por zeros. Parece contra-intuitivo que exista um bloco tão grande formado pelo mesmo dígito, mas veremos a seguir que não é. De fato, vamos dividir esse número em 40 blocos de 5 dígitos cada. Vamos chamar de sucesso a ocorrência do bloco 00000 e fracasso a ocorrência de quaisquer dos 31 blocos possíveis de tamanho 5 com pelo menos um dígito 1. Temos portanto um modelo de Bernoulli com parâmetro $\frac{1}{32}$, e a média de tempo até o primeiro sucesso é dado pelo modelo geométrico: no caso, tal média é de 32 blocos pelo Exemplo 3.1.13. Como há 40 blocos, existe uma alta probabilidade do bloco 00000 aparecer pelo menos uma vez no número gerado.

Fazendo um paralelo com os números normais na base 2, o bloco 00000 também deve aparecer com uma frequência $\frac{1}{32}$ para quase todo número gerado aleatoriamente.

Por isso podemos pensar que os números (absolutamente) normais são os melhores exemplos de números aleatórios.

No Capítulo 6, veremos uma atividade relacionada a esse problema contra-intuitivo de encontrar blocos formados pelos mesmos dígitos.

5.4 O problema de Monty-Hall

Também conhecido popularmente como problema dos bodes, é um famoso jogo que foi proposto pela primeira vez em um programa de auditório nos EUA. O jogo consistia no seguinte: havia três portas numeradas de 1 a 3. Atrás de uma delas havia um prêmio e atrás das outras duas havia bodes. Somente o apresentador do programa sabia onde estava o prêmio. Inicialmente, o jogador escolhe uma das portas, mantendo-a fechada. Em seguida, o apresentador abre uma das outras duas portas revelando um dos bodes. Posteriormente, o apresentador pergunta ao jogador se ele escolhe manter a mesma porta ou trocar de porta, sendo que agora só há duas portas disponíveis para o prêmio. A pergunta natural é: qual é a melhor estratégia? Trocar ou não trocar? A probabilidade de ganhar o prêmio se altera de alguma forma?

Intuitivamente, podemos pensar que após a abertura da porta contendo um dos bodes, teremos duas portas e chances iguais para cada uma delas, isto é, uma probabilidade igual a $\frac{1}{2}$ a partir desse momento. Com isso, não importaria se fizéssemos ou não a troca de porta. Mas essa resposta está errada, como veremos a seguir de duas formas distintas: Na primeira, vamos simular o problema diversas vezes e tentar adivinhar a resposta com base na lei dos grandes números. Na segunda, vamos de fato resolver analiticamente o problema.

Para fazer a simulação do problema, usamos o website 4Devs Ferramentas Online [1], que consegue gerar números pseudo-aleatórios de diversas formas. Para isso, geramos 100 números em $\{1, 2, 3\}$ para designar onde está o prêmio, e posteriormente geramos 100 números no mesmo conjunto para designar qual foi a porta escolhida inicialmente. Os números gerados estão compilados nas tabelas a seguir:

2	3	1	2	2	1	2	3	2	1
2	1	1	3	1	2	1	3	1	1
1	3	2	1	1	3	3	3	3	2
2	1	1	2	2	3	3	3	2	1
1	1	2	2	1	2	3	1	1	1
1	1	1	3	2	1	3	1	3	2
3	2	2	1	1	1	3	3	3	2
3	1	3	2	2	3	3	2	1	3
2	1	1	2	1	2	3	3	3	3
1	2	1	3	2	1	3	2	1	3

Tabela 5.1: Porta onde está o prêmio

2	1	3	3	2	2	3	1	2	1
2	1	1	2	1	3	2	3	3	1
3	1	1	2	2	3	2	2	2	2
3	1	2	2	3	1	1	2	2	3
3	3	3	1	2	1	1	1	3	3
2	2	3	2	1	2	2	2	1	1
3	3	1	2	1	1	2	2	3	3
1	2	3	3	2	1	1	2	2	3
1	3	2	1	1	2	2	3	3	2
3	3	2	3	2	2	3	3	1	3

Tabela 5.2: Porta que escolhemos inicialmente

A próxima tabela apresenta a diferença (em módulo) entre os respectivos valores das duas tabelas anteriores. Notemos que se tal diferença for 0 na tabela a seguir e trocarmos de porta, então perdemos o prêmio. Além disso, se tal diferença for distinta de 0 e trocarmos de porta, então ganhamos o prêmio.

0	2	2	1	0	1	1	2	0	0
0	0	0	1	0	1	1	0	2	0
2	2	1	1	1	0	1	1	1	0
1	0	1	0	1	2	2	1	0	2
2	2	1	1	1	1	2	0	2	2
1	1	2	1	1	1	1	1	2	1
0	1	1	1	0	0	1	1	0	1
2	1	0	1	0	2	2	0	1	0
1	2	1	1	0	0	1	0	0	1
2	1	1	0	0	1	0	1	0	0

Tabela 5.3: Módulo da diferença entre as respectivas entradas das Tabelas 5.1 e 5.2

Se der 0 aqui, quer dizer que escolhemos a porta certa inicialmente, e ao trocar de porta vamos perder. Se der diferente de 0, quer dizer que escolhemos a porta errada inicialmente, mas ao trocar vamos ganhar. Observemos que há 33 zeros e 67 uns e dois na tabela anterior. Isso significa que ao trocar de porta vamos conseguir o prêmio em 67 das 100 simulações. Notemos também que se tivéssemos feito uma quantidade menor de simulações, então os resultados poderiam se alterar um pouco, mas ainda assim os números indicariam uma tendência a trocar de porta. Por exemplo, considerando os 10 primeiros números gerados (1ª linha), teríamos ganhado 6 vezes e perderíamos 4 vezes ao trocar de porta. Considerando os 30 primeiros números gerados (3 primeiras linhas), teríamos ganhado 18 e perdido 12 vezes ao trocar de porta.

Tamanho da simulação	10	30	100
Ao trocar de porta, ganhamos o prêmio em	6	18	67
Não trocando de porta, ganhamos o prêmio em	4	12	33

Tabela 5.4: Compilação dos resultados das simulações

Com os resultado apresentados e pela garantia dada pela Lei dos Grandes Números, parece razoável concluir que trocar de porta é mais vantajoso pois oferece uma possibilidade maior de ganhar o prêmio; ao que tudo indica, trocar de porta apresenta uma probabilidade de ganhar o prêmio em $\frac{2}{3}$ das vezes a medida que o número de simulações fica grande o suficiente.

Agora vamos resolver analiticamente o problema. A melhor estratégia é trocar de porta, pois nesse caso nossa chance de ganhar o prêmio seria de $\frac{2}{3}$. De fato, se não trocarmos, a única chance de ganhar seria ter escolhido a porta certa no primeiro passo, que corresponde a uma probabilidade igual a $\frac{1}{3}$. Com isso, ao trocar de porta, nossa chance passa a ser o complementar: $\frac{2}{3}$. Outra forma de analisar o problema é a seguinte: suponhamos que dizemos sempre "sim" à troca de portas. Vamos supor sem perda de generalidade que o prêmio esteja na porta 1. Temos três possibilidades: Se escolhermos a porta 1 inicialmente, então o apresentador irá abrir alguma das portas 2 e 3. Ao trocar de porta, perdemos o prêmio. Se escolhermos a porta 2 inicialmente, então o apresentador irá abrir a porta 3 obrigatoriamente. Ao trocar de porta, vencemos. De forma análoga, se escolhermos a porta 3 inicialmente e trocarmos, então vencemos. Logo respondendo "sim" à troca, vencemos em 2 das 3 possibilidades, que são equiprováveis. Ao dizer "não" à troca, temos o complementar de chances para vencer: $\frac{1}{3}$. Alternativamente, podemos analisar o problema com 50 portas: em 49 delas há bodes e na restante há um prêmio. Inicialmente, escolhemos uma porta e dentre as 49 portas restantes o apresentador abre 48, revelando bodes. Nesse momento, com apenas duas portas disponíveis, ele nos dá a escolha de trocar ou manter a nossa porta inicial. Agora sim fica claro que a melhor estratégia é trocar de porta. No Capítulo 6, veremos uma atividade relacionada ao problema de Monty-Hall.

Capítulo 6

Aplicações em sala de aula

Neste capítulo, apresentaremos as atividades práticas realizadas com estudantes do Ensino Médio e com alunos da graduação em Matemática. Devido ao período de pandemia do novo Coronavírus no ano de 2021, todas as orientações foram feitas de modo não presencial, utilizando de ferramentas digitais de comunicação. O objetivo principal foi apresentar a Lei dos Grandes Números para os estudantes, aplicar duas atividades relacionadas aos números aleatórios e ao Problema de Monty-Hall e mostrar aos estudantes que esses são problemas com respostas altamente contra intuitivas. Essa pesquisa só foi realizada após anuência da escola envolvida, do PROFMAT-UFOP, dos coordenadores do Bacharelado e da Licenciatura em Matemática na UFOP, dos estudantes e dos pais ou responsáveis pelos menores de idade (após obtenção do termo de consentimento livre e esclarecido por parte dos participantes e dos pais ou responsáveis pelos estudantes menores de idade) e só ocorreu após a aprovação do projeto junto ao Comitê de Ética em Pesquisa da Universidade Federal de Ouro Preto.

A mesma atividade foi aplicada nos dois grupos de estudantes, com direcionamentos de acordo com a escolaridade de cada um deles, e com o objetivo de observar suas percepções. A atividade foi aplicada em um único encontro com cada grupo através de videoconferência utilizando o *Google Meet* como aplicativo de comunicação, o *Power Point* para apresentação, o website *4Devs: Ferramentas Online* [1] para simulações e geração de sequências de números aleatórios e o aplicativo *Whiteboard*, que funciona como um quadro branco, para complementação das explicações e apoio.

6.1 Justificativa e Relevância

A Lei dos Grandes Números é um importante resultado em Teoria das Probabilidades, que possui diversas aplicações nas demais áreas do conhecimento. Nesse projeto, apresentamos duas dessas aplicações: a primeira na identificação de números pseudoaleatórios e a segunda para abordar o problema de Monty-Hall, com o objetivo de instigar os estudantes a buscar soluções alternativas para problemas práticos; em particular, usando a Lei dos Grandes Números.

Gerar e identificar números aleatórios é um problema fundamental da Matemática, Estatística e Computação, que possui inúmeras aplicações práticas, dentre as quais citamos o Método de Monte Carlo [8, 9]. Por outro lado, o problema de Monty-Hall é um problema que desperta a curiosidade dos alunos, visto que é um jogo de fácil entendimento.

6.2 Roteiro da atividade

A pesquisa foi aplicada a 28 alunos do Ensino Médio de uma escola em Manhumirim-MG, onde o mestrando Luciano Carlos de Lemos leciona a disciplina Matemática, e a 02 alunos do curso de Matemática (Bacharelado) da UFOP. Para cada um dos públicos-alvo separadamente, foi suficiente um encontro com duração de cerca de 120 minutos para a apresentação dos conceitos e dos problemas, para realização da atividade, preenchimento dos questionários e para a apresentação da resolução.

6.2.1 Objetivos

- Introduzir conceitos de probabilidade a estudantes do Ensino Médio e Superior;
- Introduzir a Lei dos Grandes Números a esses estudantes;
- Apresentar os números normais e pseudoaleatórios, juntamente com sua importância;
- Verificar se um número dado pode ser considerado aleatório;
- Apresentar o problema de Monty-Hall e instigar as respostas dos estudantes;
- Exibir as soluções corretas aos estudantes.

6.2.2 Metodologia

A atividade foi conduzida pelo mestrando, que mediou as discussões de modo a levar os alunos a concluírem que todos os números que conseguimos pensar são pseudoaleatórios, isto é, números "aparentemente" aleatórios, mas que são gerados de uma forma determinística; o que é uma noção contra intuitiva desse problema, já que é possível provar (ver Corolário 5.1.6), que "quase todos" os números reais são absolutamente normais, e que esses são os melhores candidatos a números aleatórios. No caso do Problema de Monty-Hall, espera-se que os alunos concluam após as simulações que a melhor estratégia é trocar de porta. Em seguida será apresentada uma solução analítica para o problema. O registro das atividades foi feito através da plataforma *Google Meet*, tanto com a gravação quanto com o arquivo com as mensagens de texto. As impressões dos alunos, algumas de suas falas e respostas a perguntas feitas pelo mestrando também foram registradas através do Chat do *Google Meet* ou através de manifestações durante as atividades. Também foram aplicados questionários com o objetivo de obtermos informações sobre as impressões dos participantes acerca das atividades.

6.2.3 Descrição das Atividades

Atividade 1: Números normais e pseudoaleatórios

Os alunos foram apresentados aos números normais e aos pseudoaleatórios, além de suas propriedades e utilidades. Explicamos o problema de gerar e identificar números aleatórios. Após isso, tentamos criar um número binário aleatório grande da seguinte forma: No ensino médio, convidamos cada um dos 28 estudantes a escrever uma sequência binária de 8 dígitos, e em seguida concatenamos as sequências de todos os estudantes obtendo uma sequência com aproximadamente 200 dígitos. No ensino superior, convidamos a única estudante presente no momento para que escrevesse uma sequência binária de 100 dígitos. A seguir, indicamos aos participantes que respondessem à 1ª pergunta relativa a esse problema. Após isso, apresentamos um cálculo de forma a mostrar que a sequência obtida possui propriedades que a tornam provavelmente pseudoaleatória. Com isso, os participantes foram convidados a responder às demais perguntas relativas a esse problema.

Perguntas: Números normais e pseudoaleatórios

1. Você acha que o número obtido a partir da concatenação parece aleatório? Ou você o classificaria como pseudoaleatório? Por quê?

2. Após os cálculos apresentados, você mantém a resposta da pergunta 1?

3. O que você achou dessa aplicação da matemática?

4. Você consegue pensar em outra aplicação da geração de números aleatórios? Qual?

Atividade 2: O Problema de Monty-Hall

Os alunos foram apresentados às regras do problema de Monty-Hall, também conhecido popularmente como problema dos bodes, um famoso jogo que foi proposto pela primeira vez em um programa de auditório nos EUA. O jogo consistia no seguinte: havia três portas numeradas de 1 a 3. Atrás de uma delas havia um prêmio e atrás das outras duas havia bodes. Somente o apresentador do programa sabia onde estava o prêmio. Inicialmente, o jogador escolhe uma das portas, mantendo-a fechada. Em seguida, o apresentador abre uma das outras duas portas, revelando um dos bodes. Posteriormente, o apresentador pergunta ao jogador se ele escolhe manter a mesma porta ou trocar de porta, sendo que agora só há duas portas disponíveis para o prêmio. Neste momento, os estudantes são convidados a responder à 1ª pergunta relativa a esse problema. Em seguida, apresentamos uma visão probabilística para o problema, simulando computacionalmente o problema 100 vezes, daí os estudantes foram guiados à 2ª e 3ª perguntas. Após discussão e a apresentação de uma solução analítica, os estudantes responderam à 4ª pergunta.

Perguntas: O Problema de Monty-Hall

1. Qual seria a sua estratégia? Mudar de porta ou manter? Por quê?
2. Após ver o resultado gerado pela simulação, qual seria a sua estratégia?
3. Ao trocar de porta, qual é a probabilidade de ganhar o prêmio? Por quê?
4. Você consegue pensar em outra aplicação da Lei dos Grandes Números para "prever" a solução de algum problema? Qual?

Durante toda a atividade os estudantes foram incentivados a fazerem perguntas, a interagirem e a indicarem suas dúvidas.

6.3 Análise dos Questionários

Realizamos agora a análise dos questionários aplicados aos estudantes durante as atividades. Tais questionários pretendiam nos permitir uma coleta das impressões e opiniões dos estudantes em relação ao tema proposto.

A mesma atividade e os mesmos questionários foram aplicados aos estudantes do Ensino Médio e da graduação em Matemática, com o objetivo de verificar as semelhanças e diferenças na percepção que cada grupo teria a partir da aplicação das atividades.

Na Atividade 1, o questionário teve como objetivo verificar se os estudantes teriam algum conhecimento sobre o conceito números aleatórios e pseudoaleatórios, bem como da importância e aplicabilidade da geração de números aleatórios.

Na Atividade 2, o questionário pretendia na primeira pergunta registrar o solução intuitiva dada pelos estudantes ao Problema de Monty-Hall, e nas outras perguntas, após a simulação, registrar suas conclusões e outras aplicações para a Lei dos Grandes Números.

6.3.1 Análise das respostas dos estudantes do Ensino Médio

Atividade 1: Números normais e pseudoaleatórios

A 1ª pergunta da Atividade 1 tinha por objetivo investigar se os estudantes apresentavam algum conhecimento empírico sobre números aleatórios e pseudoaleatórios.

1. Você acha que o número obtido a partir da concatenação parece aleatório? Ou você o classificaria com pseudoaleatório? Por quê?

   Catorze estudantes responderam que parece aleatório, apresentando como justificativa o fato de que "não existe uma ordem lógica na escolha dos números" ou ainda, "porque foi formado por vários números intuitivos de cada aluno(a)". Outros catorze estudantes responderam que parece pseudoaleatório, sendo que uma aluna justificou sua resposta dizendo: "Parece pseudoaleatório, porque apesar da escolha ser individual, a maioria dos alunos podem optar por padrões de repetição como a alternância dos números ou a repetição dos números em conjuntos pares".

A 2ª pergunta pretendia, após a apresentação da definição de números aleatórios e pseudoaleatórios e de alguns cálculos com o número formado anteriormente, verificar se os estudantes mudariam a opinião expressa na pergunta 1.

2. Após os cálculos apresentados, você mantém a resposta da pergunta 1?

   Dos 14 que responderam, na pergunta anterior, que o número parece aleatório, 6 mudariam a resposta e 8 manteriam. Dos que responderam se tratar de um número pseudoaleatório, 10 manteriam a resposta e 4 mudariam.

As perguntas seguintes foram com objetivo de verificar qual a percepção dos estudantes dessa aplicação e se conseguiam pensar em outras aplicações.

3. O que você achou dessa aplicação da Matemática?

   Algumas das respostas dos estudantes foram: "interessante"; "complicada de entender"; "fora do comum"; "incrível ver com a matemática pode ser a base para a resolução de problemas de diversas áreas"; "que apesar de certa aleatoriedade ainda conseguimos achar sequências de padrões"; etc.

4. Você consegue pensar outra aplicação da geração de números aleatórios? Qual?

   Treze alunos não conseguiram pensar em outra aplicação; mas quinze alunos disseram que sim, principalmente para "criação de senhas e códigos", "programação de computador", "jogos do tipo bingo", etc.

Atividade 2: O Problema de Monty-Hall

A 1ª pergunta dessa atividade objetivava registrar a intuição dos estudantes em relação à solução do problema.

1. Qual seria sua estratégia? Mudar de porta ou manter? Por quê?

   Nove participantes mudariam, por achar que assim teriam mais chances. Cinco acham que tanto faz (mudar ou manter), pois a chances são iguais. Catorze manteriam a 1ª escolha, pois "acreditam na intuição", e alguns relataram que "o apresentador poderia estar tentando induzí-los ao erro".

A 2ª e a 3ª perguntas tinham como objetivo ver a percepção dos estudantes em relação ao problema após simularmos o mesmo 100 vezes. Cabe ressaltar que a simulação gerou o seguinte resultado: ao trocar de porta, o jogador vence 67 vezes; não trocando, vence 33 vezes.

2. Após a simulação, qual seria sua estratégia?

   Três alunos responderam que permaneceriam com a sua escolha inicial, pois ainda acham que as chances continuariam as mesmas. Mas 25 trocariam de porta, pois viram na simulação que das 100 vezes em que trocassem de porta, ganhariam em 67 delas e perderiam em 33 vezes.

3. Ao trocar de porta, qual é a probabilidade de ganhar o prêmio? Por quê?

   Algumas das respostas foram: 67 vezes em 100; $\frac{2}{3}$; $66,6\%$; pois foi o que viram na simulação.

Após uma apresentação analítica da resolução do problema, os estudantes foram convidados a responder à 4ª pergunta, com o objetivo de verificarmos se eles conseguiam pensar outra aplicação da Lei dos Grandes Números na "previsão" da solução de problemas.

4. Você consegue pensar outra aplicação da Lei dos Grandes Números para "prever" a solução de algum problema? Qual?

   Dez alunos não conseguiram pensar em outra aplicação, mas 18 responderam que sim, principalmente, nas previsões do tempo e desastres naturais, jogos de loteria, pesquisas de intenção de voto, estatísticas em geral.

Analisando as respostas dos questionários e a participação dos estudantes durante as atividades, observamos que eles demonstraram uma satisfatória compreensão do conteúdo e que as atividades despertaram neles a curiosidade para problemas com soluções muito das vezes contra intuitivas.

6.3.2 Análise das respostas dos estudantes da Graduação em Matemática

Todos os alunos da graduação em Matemática da UFOP foram convidados para participar da pesquisa, porém somente dois alunos efetivamente aceitaram o convite, sendo que a Atividade 1 foi realizada apenas por uma estudante. Os objetivos dos questionários foram os mesmos citados para os alunos de Ensino Médio, porém com uma expectativa de que as respostas fossem mais elaboradas, visto a maior experiência desses alunos.

Atividade 1: Números normais e pseudoaleatórios

1. Você acha que o número obtido a partir da concatenação parece aleatório? Ou você o classificaria com pseudoaleatório? Por quê?

   A participante respondeu "aleatório. Pois os algarismos foram escolhidos aleatoriamente".

2. Após os cálculos apresentados, você mantém a resposta da pergunta 1?

   A participante disse que “não”.

3. O que você achou dessa aplicação da Matemática?

   A estudante disse: “Achei curiosa, pois é contra intuitivo. À princípio se espera não obter nenhuma relação do número descrito, uma vez que supôs ser aleatório”.

4. Você consegue pensar outra aplicação da geração de números aleatórios? Qual?

   Segundo a participante, “talvez em corretor de código de erros”[1].

Atividade 2: O Problema de Monty-Hall

Dessa atividade participaram dois estudantes.

1. Qual seria sua estratégia? Mudar de porta ou manter? Por quê?

   Um respondeu que manteria, pois aparentemente não muda a probabilidade. Já o outro respondeu que mudaria, pois a chance de escolher o bode na primeira escolha é $\frac{2}{3}$.

Aqui, simulamos o problema 100 vezes e o resultado foi: o participante ganha 66 vezes ao trocar de porta e ganha 34 vezes ao não trocar.

2. Após a simulação, qual seria sua estratégia?

   Os dois responderam que mudariam de porta.

3. Ao trocar de porta, qual é a probabilidade de ganhar o prêmio? Por quê?

   Um deles respondeu “$\frac{2}{3}$, pois você tem que ter errado a primeira escolha e acertar a segunda escolha”; já o outro respondeu “66%, de acordo com a simulação a chance é essa”.

[1] Códigos corretores de erros

4. Você consegue pensar outra aplicação da Lei dos Grandes Números para "prever" a solução de algum problema? Qual?

   Segundo os participantes, pode ser aplicada em "jogos de baralho, como tapão, pifê, truco, poker" e no "jogo de tampinhas, em que um objeto é colocado embaixo das tampas e as tampas são embaralhadas rapidamente e uma determinada pessoa tem que descobrir onde está o objeto".

Analisando as respostas dos questionários e a participação dos estudantes, apesar de uma amostra muito pequena, percebemos que o conteúdo abordado ainda traz uma série de compreensões intuitivas que muito nos levam ao erro, e que mesmo os alunos da graduação necessitam de um aprofundamento no estudo da Teoria das Probabilidades e de suas aplicações.

Capítulo 7

Conclusões

Nesse estudo, fizemos um compilado de resultados sobre a Teoria das Probabilidades, possibilitando assim demonstrar alguns teoremas mais complexos, como a Lei dos Grandes Números. Desse modo, conseguimos mostrar também algumas aplicações pouco conhecidas como os Números Normais e outras aplicações com resultados bastante contra intuitivos, como o Problema de Monty-Hall.

Nas atividades aplicadas em sala de aula, pudemos perceber o quanto importante é trabalhar atividades que estimulam os estudantes a procurar soluções não muito óbvias. Assim, contribuímos com a formação desses estudantes e com suas percepções sobre diversos conteúdos e aplicações.

Pessoalmente, pude retomar e aprofundar conteúdos vistos em várias disciplinas do PROFMAT, que foram fundamentais para o desenvolvimento desse estudo. Em minha prática profissional, como professor de Matemática no Ensino Médio, esse estudo contribuiu para um maior domínio de alguns conteúdos e principalmente por acrescentar formas diferentes de abordá-los, sempre buscando um maior interesse e melhor entendimento dos estudantes.

Assim, posso dizer que as experiências vivenciadas no decorrer desse estudo e durante todo curso PROFMAT me proporcionaram um desenvolvimento pessoal e profissional muito valioso.

Apêndice A

Generalizações da integral de Riemann

No Cálculo, aprendemos a integral de Riemann, que é definida como o limite das somas das áreas de diversos pequenos retângulos quando a maior das bases tende a zero, caso esse limite exista. Em outras palavras, se $f : [a, b] \to \mathbb{R}$ é uma função e tomamos uma partição $x_0 = a < x_1 < \cdots < x_{n-1} < x_n = b$, então definimos

$$\int_a^b f(x)dx = \lim_{\max\{x_{j+1}-x_j\}\to 0} \sum_{i=0}^{n-1} f(x_i^*)(x_{i+1} - x_i)$$

quando esse limite existir, onde $x_i^* \in [x_i, x_{i+1}]$ é um ponto qualquer. É possível provar [7, Teorema 6 do Capítulo IX] que se f é contínua então a integral de Riemann está bem definida no intervalo $[a, b]$. Mais geralmente, é possível provar (ver [7, Teorema 20 do Capítulo IX]) que a integral de Riemann está bem definida se f é limitada e o conjunto dos seus pontos de descontinuidade tem medida nula. Além disso, dada a hipótese f é limitada, vale a volta da afirmação anterior. Posteriormente, usando o conceito de limites, podemos estender a ideia de integrabilidade de funções para intervalos infinitos: $[a, \infty)$, $(-\infty, b]$ e $(-\infty, \infty)$, que podem ser vistos como os limites do intervalo $[a, b]$ quando $b \to \infty$ e a é deixado fixo, quando $a \to -\infty$ e b é deixado fixo, e quando $a \to -\infty$ e $b \to \infty$ simultaneamente.

Nesse apêndice, vamos apresentar duas generalizações das integrais de Riemann, a saber, a integral de Lebesgue e a integral de Riemann-Stieltjes. Por fim, poderemos misturar as duas definições e considerar a integral de Stieltjes-Lebesgue, que será usada nos Capítulos 3 e 4.

A.1 A integral de Lebesgue

A integral de Lebesgue é uma generalização da integral de Riemann e apresenta algumas vantagens com relação à esta. Existem funções que não são Riemann-integráveis, como $f : [0, 1] \to \mathbb{R}$ dada por

$$f(x) = \begin{cases} 0 & \text{se } x \in \mathbb{Q}, \\ 1 & \text{se } x \in \mathbb{R} - \mathbb{Q}, \end{cases}$$

mas que terão um valor bem definido ao considerar a integral de Lebesgue (no caso da função acima, a integral valerá 1, uma vez que "quase todos" os números em $[0, 1]$ são irracionais). A seguir, vamos construir a integral de Lebesgue sobre um espaço de probabilidade $(\Omega, \mathcal{A}, \mathcal{P})$.

Seja $E \in \mathcal{A}$ um evento. Uma função $\varphi : E \to \mathbb{R} \cup \{\pm\infty\}$ é dita *função simples* se φ assume apenas um conjunto finito de valores. Em outras palavras,

$$\varphi(x) = \sum_{i=1}^{n} a_i X_{E_i}(x),$$

onde $a_i \in \mathbb{R} \cup \{\pm\infty\}$, os E_i's são eventos que podemos supor 2 a 2 disjuntos satisfazendo $\bigcup_{i=1}^{n} E_i = E$, e X_{E_i} é a *função indicadora* do evento E_i, ou seja,

$$X_{E_i}(x) = \begin{cases} 1 & \text{se } x \in E_i \\ 0 & \text{se } x \notin E_i \end{cases}.$$

Convencionando que $\pm\infty \cdot 0 = 0$, definimos a integral de Lebesgue da função simples φ sobre E como

$$\int_E \varphi(x) d\mathcal{P} = \sum_{i=1}^{n} a_i \mathcal{P}(E_i).$$

Definição A.1.1. *Dizemos que uma função $f : \Omega \to \mathbb{R}$ é $\mathcal{A}$-mensurável (ou apenas* mensurável *quando não houver dúvidas) se, para todo $\alpha \in \mathbb{R}$, o conjunto $\{x \in \Omega; f(x) > \alpha\}$ pertence à σ-álgebra $\mathcal{A}$.*

Usando as propriedades de $\mathcal{A}$, é possível provar que o sinal $>$ pode ser substituído por $<$, $\geq$ ou $\leq$. Além disso, são exemplos de funções mensuráveis:

1. funções indicadoras de eventos,

2. funções simples,

3. funções contínuas sobre $\mathbb{R}$ são $\mathcal{B}$-mensurável,

4. funções monótonas são $\mathcal{B}$-mensurável sobre $\mathbb{R}$,

5. combinações lineares de funções mensuráveis, quadrado, produto e módulo de funções mensuráveis,

6. parte positiva $f^+(x) = \max\{f(x), 0\}$ e parte negativa $f^-(x) = \max\{-f(x), 0\}$ de uma função mensurável $f(x) = f^+(x) - f^-(x)$,

7. limites, supremos e ínfimos, limsups e liminfs de sequências de funções mensuráveis, etc.

Seja $f : \Omega \to [0, \infty)$ é uma função mensurável e não-negativa. Também é possível provar que existe uma sequência de funções simples $(\varphi_n)_{n\in\mathbb{N}}$ satisfazendo $0 \leq \varphi_n(x) \leq \varphi_{n+1}(x)$ para todo $n \in \mathbb{N}$ e todo $x \in \Omega$ de forma que $f(x) = \lim_n \varphi_n(x)$. Nesse caso, a integral de Lebesgue de f é definida por

$$\int f d\mathcal{P} = \lim \int \varphi_n(x) d\mathcal{P}.$$

Se $E \in \mathcal{A}$, definimos

$$\int_E f d\mathcal{P} = \lim \int \varphi_n(x) X_E(x) d\mathcal{P}.$$

É possível provar que a integral definida dessa forma é linear e satisfaz que se $0 \leq f \leq g$ então $\int_E f d\mathcal{P} \leq \int_E g d\mathcal{P}$.

Exemplo A.1.2. *Seja* $f(x) = \begin{cases} 0 & se\ x \in \mathbb{Q}, \\ 1 & se\ x \in \mathbb{R} - \mathbb{Q}, \end{cases}$ *que é uma função $\mathcal{B}$-mensurável sobre* $[0, 1]$. *Notemos que* $[0, 1] = [\mathbb{Q} \cup (\mathbb{R} - \mathbb{Q})] \cap [0, 1]$, *que* $\mathcal{P}(\mathbb{Q} \cap [0, 1]) = 0$ *e* $\mathcal{P}((\mathbb{R} - \mathbb{Q}) \cap [0, 1]) = 1$, *logo*

$$\int_{[0,1]} f(x) d\mathcal{P} = \int_{[0,1]\cap\mathbb{Q}} 0 d\mathcal{P} + \int_{[0,1]\cap(\mathbb{R}-\mathbb{Q})} 1 d\mathcal{P} = 1.$$

Definição A.1.3. *Sejam* $f, f_1, f_2, \cdots : \Omega \to \mathbb{R}$. *Dizemos que a sequência* $(f_n)_n$ converge pontualmente *para f se, para todo $x \in \Omega$, vale que $f_n(x) \to f(x)$ quando $n \to \infty$.*

Um importante resultado sobre convergência de integrais de Lebesgue é o seguinte:

Teorema A.1.4 (da Convergência Monótona)**.** *Se $(f_n)_{n\in\mathbb{N}}$ é uma sequência monótona crescente de funções mensuráveis não-negativas que converge pontualmente para f, então $\int f d\mathcal{P} = \lim \int f_n d\mathcal{P}$. Em outras palavras, podemos trocar a integral de posição com o limite se houver uma "monotonicidade" da sequência.*

Demonstração. Ver [6, Teorema 3.3]. □

Para funções $f : \Omega \to \mathbb{R}$ em geral (não necessariamente não-negativas), basta considerar a $f = f^+ - f^-$, onde as partes positiva f^+ e negativa f^- de f são função não-negativas e estender a definição por linearidade, isto é, $\int f d\mathcal{P} = \int f^+ d\mathcal{P} - \int f^- d\mathcal{P}$. Algumas propriedades que seguem diretamente são a linearidade da integral e a desigualdade triangular. Dizemos que f é *integrável* se as integrais $\int f^+ d\mathcal{P}$ e $\int f^- d\mathcal{P}$ são finitas.

Nesse sentido, podemos enunciar o segundo resultado importante dessa seção:

Teorema A.1.5 (da Convergência Dominada)**.** *Seja $(f_n)_{n\in\mathbb{N}}$ uma sequência de funções que converge quase certamente para uma função mensurável f. Se existe uma função integrável g tal que $|f_n(x)| < g(x)$ para todo n e para todo $x \in \Omega$, então f é integrável e $\int f d\mathcal{P} = \lim_n \int f_n d\mathcal{P}$. Em outras palavras, podemos trocar a integral de posição com o limite se houver uma "dominação" da sequência.*

Demonstração. Ver [6, Teorema 3.4]. □

O Teorema da Convergência Dominada é usado algumas vezes no Capítulo 4.

É importante ressaltar que quando ambas estão definidas, as integrais de Lebesgue e de Riemann de fato coincidem.

A.2 A integral de Stieltjes

Sejam $\varphi : [a, b] \to \mathbb{R}$ uma função contínua e $F : \mathbb{R} \to [0, 1]$ uma função de distribuição. Define-se a *integral de Riemann-Stieltjes* de φ em relação a F (ou ponderada por F), como o limite de "somas ponderadas de Riemann" da forma

$$\lim_{\max\{x_{j+1}-x_j\}\to 0} \sum_{i=0}^{n-1} \varphi(x_i^*)[F(x_{i+1}) - F(x_i)], \tag{A.1}$$

quando esse limite existe, onde $a = x_0 < x_1 < x_2 < \cdots < x_n = b$, x_i^* é um ponto arbitrário de $[x_i, x_{i+1}]$. Tal limite existe e é finito sob as condições descritas, e é representado por

$$\int_a^b \varphi(x)\, dF(x).$$

A função φ é chamada de integrando, F de integrador.

A hipótese de que F é uma função de distribuição não é necessária: basta que F seja uma função de variação limitada (ver [6, Seção 3.1]). Em particular, a definição da integral de Riemann-Stieltjes se estende para funções F monótonas, mas não usamos isso nesse texto.

Por outro lado, a definição da integral de Riemann-Stieltjes pode ser estendida a outras funções φ além das contínuas. Para uma função φ qualquer, define-se

$$\int_a^b \varphi(x)\, dF(x)$$

de forma análoga ao que fazemos com a integral de Riemann. O problema desta definição é que até funções bem simples deixam de possuir integrais, como vemos no seguinte exemplo.

Exemplo A.2.1. *Seja F a função de distribuição definida por*

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

e consideremos a integral $\int_{-1}^{1} F(x)dF(x)$, ou seja, F é ao mesmo tempo integrador e integrando. Essa integral não está bem definida pois o limite (A.1) não existe. De fato, se $x_i \neq 0$ para todo i, teremos $x_j < 0 < x_{j+1}$ para algum j, logo $F(x_{j+1}) - F(x_j) = 1$. Assim, o somatório assume como valor ou 0 ou 1, dependendo do valor escolhido para x_j^* ser negativo ou positivo, respectivamente.

Dada essa deficiência da definição, a integral de Riemann-Stieltjes mostra-se insuficiente para nossos propósitos, e temos que utilizar uma integral mais geral, a saber, a integral de *Lebesgue-Stieltjes*, que, de agora em diante, será chamada simplesmente *integral de Stieltjes*. A integral de Stieltjes será portanto a união das definições das integrais de Riemann-Stieltjes e de Lebesgue, isto é, será a integral de Riemann-Stieltjes com a interpretação mais geral dos conjuntos e funções mensuráveis dada pela integral de Lebesgue.

No que se segue, o integrando φ é uma função real mensurável, e o integrador F é uma função de distribuição. A notação $\int_a^b \varphi dF$ significa $\int_a^b \varphi(x)dF(x)$. Quando não aparecem limites de integração, a integral é sobre toda a reta:

$$\int \varphi dF = \int_{-\infty}^{\infty} \varphi dF = \lim_{a \to -\infty, b \to +\infty} \int_a^b \varphi dF.$$

Quando o integrando é contínuo em $[a, b]$, a integral de Stieltjes torna-se uma simples integral de Riemann-Stieltjes, e podemos utilizar as propriedades desta. Em particular, isso facilita nossa discussão sobre as propriedades da esperança, já que $\varphi(x) = x$

é contínua.

Proposição A.2.2. *São válidos:*

a) *Teorema Fundamental do Cálculo:* $\int_a^b dF(x) = F(b) - F(a)$.

b) *A integral de Stieltjes é* linear, *tanto no integrando quanto no integrador. Em outras palavras, para* $\varphi(x) = \alpha f(x) + \beta g(x)$ *temos*

$$\int_a^b \varphi dF = \alpha \int_a^b f dF + \beta \int_a^b g dF$$

e

$$\int \varphi dF = \alpha \int f dF + \beta \int g dF,$$

e para $H(x) = \alpha F(x) + \beta G(x)$ *temos*

$$\int_a^b \varphi dH = \alpha \int_a^b \varphi dF + \beta \int_a^b \varphi dG$$

e

$$\int \varphi dH = \alpha \int \varphi dF + \beta \int \varphi dG,$$

onde valem as equações acima desde que as integrais estejam bem definidas e as somas tenham sentido.

c) *A integral de Stieltjes é* aditiva, *isto é, se* $a < b < c$ *então*

$$\int_a^c \varphi dF = \int_a^b \varphi dF + \int_b^c \varphi dF.$$

Isto vale também quando os intervalos são infinitos. Por exemplo,

$$\int \varphi dF = \int_{-\infty}^a \varphi dF + \int_a^\infty \varphi dF.$$

Novamente, estas equações são válidas quando as integrais estão bem definidas e as somas têm sentido.

d) *Quando* F *é a função de distribuição de uma variável aleatória discreta* X, *a integral de Stieltjes reduz-se a uma série: se* $\mathcal{P}(X = x_i) = p(x_i) > 0$ *e* $\sum_i p(x_i) = 1$, *isto é, se* p *é a função de probabilidade de* X, *então* $p(x_i)$ *é o salto de* F *em* x_i *e*

$$\int \varphi dF = \int_{-\infty}^\infty \varphi(x) dF(x) = \sum_i \varphi(x_i) p(x_i).$$

Observação: *uma explicação intuitiva desta propriedade resulta da interpretação da diferencial $dF(x)$ como igual a $p(x_i)$ no ponto x_i e zero nos pontos x que não são pontos de salto de F (notemos que F cresce apenas em seus pontos de salto).*

Quando a região de integração é um intervalo finito, temos

$$\int_a^b \varphi dF = \sum_{i:a<x_i\leq b} \varphi(x_i)p(x_i).$$

e) *Quando F é a função de distribuição de uma variável aleatória contínua tendo densidade f, então f é a derivada de F (em quase toda parte). Temos $dF(x) = f(x)dx$ logo,*

$$\int_a^b \varphi dF = \int_a^b \varphi(x)f(x)dx \quad e \quad \int \varphi dF = \int_{-\infty}^{\infty} \varphi(x)f(x)dx.$$

Já que F é contínua à direita, a integral de Riemann-Stieltjes em $[a,b]$, se existe, ignora um eventual salto de F no ponto a, mas leva em consideração um eventual salto de F no ponto b. Por exemplo, seja

$$F_1(x) = \begin{cases} 0, \text{se } x < a \\ \frac{1}{2}, \text{se } a \leq x < b \\ 1, \text{se } b \leq x. \end{cases}$$

Vemos que a função de distribuição F_1, embora possua dois pontos de salto pertencentes ao intervalo $[a,b]$, salta somente uma vez no intervalo.

Se φ for uma função contínua em $[a,b]$, a integral de Riemann-Stieltjes de φ em $[a,b]$, ponderada por F_1, será

$$\int_a^b \varphi dF_1 = \frac{1}{2}\varphi(b).$$

Portanto, a integração leva em conta o salto em b e ignora o salto em a. Com isso, utilizaremos o símbolo $\int_a^b$ para representar a integral sobre o intervalo $(a,b]$, isto é,

$$\int_a^b \varphi dF = \int_{(a,b]} \varphi dF \tag{A.2}$$

onde o termo à direita é a integral de Stieltjes em $(a,b]$.

Desta maneira, quando a integral de Riemann-Stieltjes existir, será igual à integral de Stieltjes e o termo à esquerda indicará as duas integrais.

No caso da esperança, podemos refrasear os Teoremas A.1.4 e A.1.5 como a seguir:

Corolário A.2.3 (Teorema da Convergência Monótona para esperanças)**.** *Sejam* X, X_1, $X_2, \ldots$ *variáveis aleatórias em* $(\Omega, \mathcal{A}, \mathcal{P})$. *Se* $0 \leq X_n \uparrow X$, *isto é,* $X_n(\omega) \geq 0$ *e* $X_n(\omega) \uparrow X(\omega)$ *para todo* $\omega \in \Omega$, *então* $E(X_n) \uparrow E(X)$.

Corolário A.2.4 (Teorema da Convergência Dominada para esperanças)**.** *Sejam* Y, X, $X_1, X_2, \ldots$ *variáveis aleatórias em* $(\Omega, \mathcal{A}, \mathcal{P})$ *tais que* Y *é integrável,* $|X_n| \leq Y \ \ \forall n$, *e* $X_n \to X$ *(isto é,* $X_n(\omega) \to X(\omega) \ \ \forall \omega$*). Então* X *e* X_n *são integráveis e* $E(X_n) \to E(X)$.

Apêndice B

A integral de uma função gaussiana

Uma função gaussiana é uma função da forma $f(x) = ae^{-(x-b)^2/c^2}$, onde $a, b, c \in \mathbb{R}$ e $ac \neq 0$. É possível provar que a primitiva dessa função não pode ser explicitada em termos de uma combinação finita de funções elementares. Por outro lado, podemos usar a série de potências da função exponencial para calcular sua integral, e com isso obter aproximações para o real valor da integral definida. Ainda assim, é possível encontrar a integral de uma função gaussiana sobre $\mathbb{R}$.

De fato, seja

$$I = \int_{-\infty}^{\infty} e^{-y^2} dy.$$

Vamos calcular I^2 usando duas integrais, o que se transforma em uma integral dupla, assim:

$$I^2 = \int_{-\infty}^{\infty} e^{-y^2} dy \int_{-\infty}^{\infty} e^{-x^2} dx = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-y^2} e^{-x^2} dydx = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-y^2-x^2} dydx.$$

Usando as coordenadas polares e observando que $x^2 + y^2 = r^2$, temos que a região de integração é todo o plano xy, portanto $0 \leq r < \infty$ e o ângulo $\theta \in [0, 2\pi]$. Assim a integral

$$I^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-y^2-x^2} dydx = \int_{0}^{2\pi}\int_{0}^{\infty} e^{-r^2} r\, dr\, d\theta$$

pode ser calculada utilizando o método de substituição de variáveis. Fazendo $u = -r^2$, $du = -2rdr$, $-\dfrac{du}{2r} = dr$ e $-\infty < u \leq 0$, temos

$$\int_{0}^{\infty} e^{-r^2} r\, dr = -\frac{1}{2}\int_{-\infty}^{0} e^{u} du = -\frac{1}{2} e^{-r^2}|_{-\infty}^{0} = -\frac{1}{2}(0-1) = \frac{1}{2}.$$

Teremos então:

$$\int_0^{2\pi} \int_0^{\infty} e^{-r^2} r\, dr\, d\theta = \int_0^{2\pi} \frac{1}{2} d\theta = \frac{1}{2}(2\pi - 0) = \pi.$$

Portanto, concluímos que $I^2 = \pi$, o que implica $I = \sqrt{\pi}$ já que $I > 0$.

Referências Bibliográficas

[1] 4DEVS, FERRAMENTAS ONLINE; **Gerador de Números Aleatórios.** https://www.4devs.com.br/gerador_de_numeros_aleatorios. Acesso em: 24 de set. de 2021.

[2] BARTLE, R.; **Elements of integration and Lebesgue measure.** John Wiley & Sons, 1995.

[3] BOREL, E.; **Les probabilités dénombrales et leurs applications arithmétiques.** Supplemento di rend. circ. Mat. Palermo 27, 1909, p.247-271.

[4] BREIMAN, L.; **Probability.** Addison-Wesley, Reading, Massachusetts, 1968.

[5] CHAMPERNOWNE, D.G.; **The construction of decimals normal in the sacle of ten.** J. London Math. Soc. s1-8 (4), 1933, p. 254-260.

[6] JAMES, B.R.; **Probabilidade: um curso em nível intermediário.** Projeto Euclides, IMPA, 1981.

[7] LIMA, E.L.; **Curso de Análise, vol. 1.** 7ª edição, Projeto Euclides, Instituto de Matemática Pura e Aplicada, CNPq, Rio de Janeiro, 1976.

[8] MAZHDRAKOV, M., BENOV, D., VALKANOV, N.; **The Monte Carlo Method.** Engineering Applications. ACMO Academic Press, 2018.

[9] METROPOLIS, N.; **The beginning of the Monte Carlo method.** Los Alamos Science (1987 Special Issue dedicated to Stanislaw Ulam), 1987, p. 125-130.

[10] RUDDIN, W.; **Princípios de Análise Matemática.** Ao Livro Técnico, Rio de Janeiro, 1971. Traduzido por Eliana Rocha Henriques de Brito da 2ª edição americana (1964).

[11] SILVA, M.N.P.; **História da Probabilidade.** Brasil Escola. Disponível em: https://brasilescola.uol.com.br/matematica/historia-probabilidade.htm. Acesso em: 04 de outubro de 2021.