

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E
TECNOLOGIA DA BAHIA - CAMPUS VALENÇA
LICENCIATURA EM MATEMÁTICA

Thaís Santos Santana

Uma Abordagem da Lei dos Grandes Números

Valença-BA
2022

Thaís Santos Santana

Uma Abordagem da Lei dos Grandes Números

Trabalho de Conclusão de Curso apresentado ao Instituto Federal de Educação, Ciência e Tecnologia da Bahia - Campus Valença como parte dos requisitos para a obtenção do título de Licenciado em Matemática.

Orientadora: Prof^ª. Ma. Renata de Moura Issa
Vianna

Valença-BA
2022

Thaís Santos Santana

Uma abordagem da Lei dos Grandes Números

Monografia apresentada a Coordenação do Curso de Licenciatura em Matemática do Instituto Federal de Educação, Ciência e Tecnologia da Bahia, *Campus* Valença, como requisito parcial para obtenção do título de Licenciada em Matemática.

Monografia aprovada em ____ / ____ / ____.

BANCA EXAMINADORA

Profa. Me. Renata de Moura Issa Vianna
Instituto Federal de Educação, Ciência e Tecnologia da Bahia

Prof. Dr. Diogo Soares Dórea da Silva
Instituto Federal de Educação, Ciência e Tecnologia da Bahia

Prof. Me. Roque da Silva Lyrio
Instituto Federal de Educação, Ciência e Tecnologia da Bahia

Valença - BA, ____ de ____ de 2022.

S232a Santana, Thaís Santos

Uma abordagem da lei dos grandes números. – Valença -
BA: IFBA, 2022.
38.;il.

Orientadora: Prof^ª. Ma. Renata de Moura Issa Vianna

Trabalho de conclusão de curso (Graduação) - Instituto
Federal de Educação, Ciência e Tecnologia da Bahia – Campus
Valença, 2022.

1. Probabilidade. 2. Lei dos Grandes Números. 3. Lei Fraca.
4. Lei Forte. I. Vianna, Renata de Moura Issa. II. Título.

CDD: 519.5

Ficha Catalográfica elaborada pela bibliotecária do IFBA campus Valença/
Cátia Almeida de Andrade CRB1403-5

Dedicatória

Dedico este trabalho a Deus, que me permitiu continuar a viver, mesmo quando não tinha mais chances, que me deu forças e coragem para alcançar meus objetivos.

Agradecimentos

Em primeiro lugar agradeço a Deus, pois não estaria aqui, me dando forças para superar as dificuldades, que foram muitas e sem ele não chegaria onde cheguei.

À minha família que sempre me incentivou, com palavras de apoio e perseverança, minha mãe Ana, meu pai Jose Carlos, minha avó Alice, minha tia Ailce, meu irmão Henrique e minhas primas Danila e Dayane, em especial quero agradecer a meu avô Humberto que se foi esse ano, mas me ensinou o sentido da palavra força.

Não posso deixar de mencionar as minhas amigas e parceiras, Ieda, Manuella e Naire, que sempre me colocaram pra cima, me ouviram, entenderam e me ergueram quando estava pra baixo e pensando em desistir, fizeram parte do meu alicerce para continuar.

Quero agradecer à instituição IFBA Valença, por ter me oferecido a oportunidade de ingressar no curso e ter aberto esse campo de possibilidades que vou levar para a vida.

Não posso deixar de demonstrar minha gratidão ao corpo docente do IFBA, em especial gostaria de destacar três professores pois sem eles nada disso seria possível, o professor Diogo Dórea, que quando tava quase desistindo me incentivou, ao professor Roque Lyrio, que quando realmente estava numa fase difícil e desisti do curso, me trouxe de volta, me deu todo o apoio possível, me entendeu e ajudou, e por último e mais importante a professora Renata Issa, sou grata por não ter desistido de mim, me deu todo o suporte pra tornar esse projeto possível, mesmo com toda a luta na sua vida pessoal, dispôs de tempo para me orientar.

A matemática pura é, à sua maneira, a poesia das ideias lógicas.

Albert Einstein (1879 - 1955)

Resumo

Este trabalho visa apresentar a Lei dos Grandes Números, que é um teorema fundamental da Probabilidade. Concisamente esta Lei nos diz que a média aritmética dos resultados da realização da mesma experiência repetidas vezes tende a se aproximar do valor esperado à medida que mais tentativas se sucederem. A princípio introduzimos o trabalho apresentando alguns conceitos iniciais de Probabilidade, como os Espaços de Probabilidade, as Variáveis Aleatórias, a Desigualdade de Tchebyshev e o Lema de Borel-Cantelli, que são assuntos base para a demonstração do nosso tema principal. Por fim, apresentamos a Lei dos Grandes Números, com suas duas versões que se separam pelo tipo de convergência utilizado. Elas são a Lei Fraca dos Grandes Números, que utiliza a Convergência em Probabilidade, e a Lei Forte dos Grandes Números, onde se aplica a Convergência Quase Certa.

Palavras-chave: Probabilidade. Lei dos Grandes Números. Lei Fraca. Lei Forte.

Abstract

This work aims to present the Law of Large Numbers, which is a fundamental theorem of probability. Concisely this Law tells us that the arithmetic mean of the results of the performing the same experiment over and over tends to approximate the expected value as more attempts are made. At first, we introduce the work by presenting some initial concepts of Probability, such as Probability Spaces, Random Variables, the Chebyshev Inequality and the Borel-Cantelli Lemma, which are basic subjects for the demonstration of our main theme. Finally, we present the Law of Large Numbers, with its two versions that are separated by the type of convergence used. They are the Weak Law of Large Numbers, which uses Convergence in Probability, and the Strong Law of Large Numbers, where Almost Certain Convergence applies.

Keywords: Probability. Law of Large Numbers. Weak Law. Strong Law.

Lista de Ilustrações

Figura 1.1 – Densidade da variável aleatória $X \sim U[c, d]$	12
Figura 1.2 – Densidade da variável aleatória $X \sim exp(\lambda)$	13
Figura 1.3 – Densidade da variável aleatória $X \sim N(\mu, \sigma)$	13

Sumário

0	Introdução	1
1	Conceitos Iniciais de Probabilidade	3
1.1	Espaços de Probabilidade	3
1.1.1	Probabilidade Condicional	7
1.1.2	Eventos Independentes	9
1.2	Variáveis Aleatórias	10
1.2.1	Valor Esperado, Esperança ou Média	14
1.2.2	Variância	15
1.3	Desigualdade de Tchebyshev	16
1.4	Lema de Borel-Cantelli	17
2	Lei Fraca dos Grandes Números	20
2.1	Convergência em Probabilidade	20
2.2	Lei Fraca dos Grandes Números	21
3	Lei Forte dos Grandes Números	23
3.1	Convergência Quase Certa	23
3.2	Lei Forte dos Grandes Números	24
4	Conclusões e Perspectivas	27
	Referências	28

Capítulo 0

Introdução

A Lei dos Grandes Números versa que a média aritmética dos resultados da realização da mesma experiência repetidas vezes tende a se aproximar do valor esperado à medida que mais tentativas se sucederem. Ela foi mencionada pela primeira vez pelo matemático italiano Gerolamo Cardano (1501 – 1576), quando afirmou que a precisão das estatísticas empíricas tende a melhorar à medida que o número de tentativas aumenta, embora sem qualquer prova rigorosa. Mais tarde, Jacob Bernoulli (1654 – 1705) levou mais de vinte anos para desenvolver uma demonstração completa em sua obra “Ars Conjectandi” (A Arte da Conjectura) em 1713, chamando-o de “Teorema Dourado”, apesar de se tornar mais conhecido como “Teorema de Bernoulli”. Em 1835, o matemático francês Siméon Denis Poisson (1781 – 1840) descreveu em detalhes A Lei dos Grandes Números, que veio aperfeiçoar a teoria.

Depois de Bernoulli e Poisson, outros autores também contribuíram para o aperfeiçoamento desta lei, incluindo Pafnuti Chebyshev, Andrei Markov, Émile Borel, Francesco Paolo Cantelli, Andrei Kolmogorov, Vladimir Vapnik, Alexey Chervonenkis e Aleksandr Khinchin. Com isso, surgiram duas formas da Lei dos Grandes Números. Uma é a chamada de “Lei Fraca dos Grandes Números” e a outra de “Lei Forte dos Grandes Números”. Estas não são leis diferentes, mas maneiras diferentes de representar a convergência da probabilidade medida para a probabilidade real. Em particular, a Lei Forte implica a Lei Fraca.

Inicialmente, a Lei dos Grandes Números baseou-se no conceito frequencista de probabilidade, em que as probabilidades são estabelecidas a posteriori, com base nos resultados observados pela realização de experiências aleatórias. Neste conceito, a probabilidade de ocorrência de um acontecimento A , de uma experiência aleatória, é associada à frequência relativa com que esse acontecimento é observado, ou seja,

$$\mathbb{P}(A) = f_r(A) = \frac{n_A}{n},$$

em que $f_r(A)$ representa a frequência relativa do acontecimento A , n_A representa o número de vezes que se observou o acontecimento A e n representa o número de vezes em que se

repetiu a experiência aleatória.

Foi através da Lei dos Grandes Números, formulada por Bernoulli, que se relacionou o conceito frequencista de probabilidade com o conceito clássico de probabilidade: para um grande número de experiências, tendo cada uma um resultado aleatório, a frequência relativa de cada um desses resultados tende a estabilizar, convergindo para um certo número que constitui a probabilidade desse resultado.

A Lei dos Grandes Números é muito utilizada em vários experimentos científicos e também em áreas como agricultura, economia, engenharia e diversas outras. Ela expõe a importância do uso de algumas técnicas estatísticas na realização de experimentos, como a amostragem, que calcula o número de vezes que um experimento deve ser repetido para que se obtenha um valor que realmente represente a população de estudo.

Capítulo 1

Conceitos Iniciais de Probabilidade

Neste capítulo, vamos analisar alguns conceitos introdutórios necessários para a compreensão do presente trabalho. Inicialmente, estudaremos a definição moderna de Probabilidade e suas propriedades básicas. Depois, definiremos as variáveis aleatórias, além de apresentarmos a Desigualdade de Tchebyshev e o Lema de Borel-Cantelli.

1.1 Espaços de Probabilidade

Definição 1.1 (Experimento Aleatório). *Os experimentos que repetidos sob as mesmas condições geram resultados que não podem ser previstos com certeza são chamados de experimentos aleatórios.*

Exemplo 1.1. *São experimentos aleatórios:*

- (i) *Lançar uma moeda honesta e observar a face superior;*
- (ii) *Lançar um dado não viciado e observar o número de sua face superior;*
- (iii) *Retirar uma bola ao acaso de uma urna com 20 bolas numeradas de 1 a 20 e observar o número retirado.*

Os experimentos listados acima possuem um conjunto de possíveis resultados. Esse conjunto é chamado de espaço amostral.

Definição 1.2 (Espaço Amostral). *Suponha que um experimento aleatório seja realizado sob certas condições fixas. Chamamos de espaço amostral o conjunto Ω , que é formado por todos resultados possíveis do experimento.*

Exemplo 1.2. *Considere o lançamento de uma moeda não viciada de forma aleatória. O espaço amostral deste experimento é*

$$\Omega = \{Cara, Coroa\}.$$

Se fizermos o lançamento de duas moedas, o espaço amostral será

$$\Omega = \{Cara, Coroa\} \times \{Cara, Coroa\}.$$

Exemplo 1.3. Considere o lançamento de um dado não viciado de forma aleatória. O espaço amostral deste experimento é

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

Se fizermos o lançamento de três dados, o espaço amostral será

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

Definição 1.3 (Evento). Um subconjunto $A \in \Omega$ é chamado evento.

Na Probabilidade, muitas notações são resumidas, omitindo alguns detalhes. No Exemplo 1.3, se o experimento fosse “considere n lançamentos de uma moeda honesta”, o espaço amostral seria

$$\begin{aligned} \Omega &= \{a_1, a_2, \dots, a_n : a_n \in \{Cara, Coroa\}\} \\ &= \{Cara, Coroa\} \times \{Cara, Coroa\} \times \dots \times \{Cara, Coroa\} \\ &= \{Cara, Coroa\}^n. \end{aligned}$$

Além disso, um conjunto pode ser representado descrevendo-o com palavras. Por exemplo, para representar o evento “Coroa no segundo de dois lançamentos de uma moeda”, a maneira formal seria

$$A = \{(a_1, a_2) : a_2 = Coroa\}.$$

Entretanto, para simplificar, denotaremos este evento simplesmente como

$$A = [\text{Coroa no } 2^\circ \text{ lançamento}].$$

Definição 1.4 (σ -álgebra). Uma coleção \mathbb{A} de subconjuntos de Ω é dita uma σ -álgebra se

- (i) $\Omega \in \mathbb{A}$;
- (ii) Se $A \in \mathbb{A}$, então $A^c \in \mathbb{A}$;
- (iii) Se $A_1, A_2, A_3, \dots \in \mathbb{A}$, então $\bigcup_{i=1}^{\infty} A_i \in \mathbb{A}$.

Uma coleção (ou família) de subconjuntos é um conjunto de subconjuntos. A palavra coleção (ou família) é usada na literatura por estilo, além de ressaltar que se trata de um conjunto cujos elementos também são conjuntos. Basicamente, a σ -álgebra é um conjunto de subconjuntos de Ω satisfazendo as três condições descritas acima.

Exemplo 1.4. $\mathbb{A} = \{\emptyset, \Omega\}$ é uma σ -álgebra, chamada de trivial. O conjunto das partes $\mathcal{P}(\Omega)$ também é uma σ -álgebra.

Exemplo 1.5. Se $\Omega = \{a, b, c\}$ então $\mathbb{A} = \{\emptyset, \{a\}, \{b, c\}, \{a, b, c\}\}$ é uma σ -álgebra, que não é nem a trivial, nem a das partes.

Definição 1.5 (Espaço Mensurável). Um par (Ω, \mathbb{A}) , em que Ω é um conjunto e \mathbb{A} é uma σ -álgebra de subconjuntos de Ω , é chamado de espaço mensurável.

Vejamos agora a definição axiomática de probabilidade. Foi o matemático russo Andrei Kolmogorov (1903-1987) que axiomatizou a probabilidade em sua obra “Foundations of the Theory of Probability”. Os Axiomas de Kolmogorov, como são chamados, foram estabelecidos em 1933 e estabelece que eventos são representados por conjuntos e probabilidade é apenas uma medida padronizada definida nesses conjuntos.

Definição 1.6 (Definição Axiomática de Probabilidade). Uma probabilidade \mathbb{P} em um espaço mensurável (Ω, \mathbb{A}) é uma função $\mathbb{P}: \mathbb{A} \rightarrow \mathbb{R}$ que satisfaz:

$$(i) \mathbb{P}(A) \geq 0, \forall A \in \mathbb{A};$$

(ii) Se $A_1, A_2, A_3, \dots \in \mathbb{A}$ são disjuntos, então

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i);$$

$$(iii) \mathbb{P}(\Omega) = 1.$$

Definição 1.7 (Espaço de Probabilidade). A uma terna $(\Omega, \mathbb{A}, \mathbb{P})$ chamamos de espaço de probabilidade.

Note que, pela Definição (1.6), uma probabilidade \mathbb{P} é uma função cujo domínio é \mathbb{A} , e não Ω . Isto quer dizer que atribuímos probabilidades a subconjuntos de Ω , não a elementos de Ω . É verdade que podemos atribuir probabilidade a um subconjunto unitário de Ω , que seria identificado com seu elemento, porém nem sempre isso é possível. No Exemplo (1.5), poderíamos atribuir uma probabilidade para o evento $\{a\}$, mas não é possível atribuir uma probabilidade ao evento $\{b\}$, pois este não pertence à σ -álgebra \mathbb{A} . Além disso, quando lidamos com uma σ -álgebra \mathbb{A} associada a uma probabilidade \mathbb{P} , é costume chamar os elementos de \mathbb{A} de eventos.

Definição 1.8. Sejam A e $A_n, n \in \mathbb{N}$, conjuntos. Denotamos $A_n \nearrow A$ e dizemos que A_n cresce para A , se

$$A_n \subset A_{n+1}, \forall n \in \mathbb{N}, \text{ e } A = \bigcup_{i=1}^{+\infty} A_i.$$

Analogamente, definimos $A_n \searrow A$ (A_n decresce para A) se

$$A_n \supset A_{n+1}, \forall n \in \mathbb{N}, \text{ e } A = \bigcap_{i=1}^{+\infty} A_i.$$

A seguir, apresentaremos algumas propriedades importantes da probabilidade.

Proposição 1.1. *Seja $(\Omega, \mathbb{A}, \mathbb{P})$ um espaço de probabilidade e sejam $A, B, A_1, A_2, \dots \in \mathbb{A}$. Então,*

(i) $\mathbb{P}(\emptyset) = 0$;

(ii) *Se $A_1, A_2, \dots, A_n \in \mathbb{A}$ são disjuntos, então*

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

(iii) *Se $A \subset B$, então $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B - A)$;*

(iv) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$;

(v) $\mathbb{P}(A) \leq 1, \forall A \in \mathbb{A}$;

(vi) *(Monotonicidade da Probabilidade) Se $A \subset B$, então $\mathbb{P}(A) \leq \mathbb{P}(B)$;*

(vii) *(Continuidade da Probabilidade) Se $A_i \nearrow A$, então $\mathbb{P}(A_i) \nearrow \mathbb{P}(A)$;*

(viii) *(Continuidade da Probabilidade) Se $A_i \searrow A$, então $\mathbb{P}(A_i) \searrow \mathbb{P}(A)$.*

Omitiremos as demonstrações das propriedades acima. Elas poderão ser encontradas em (FRANCO, 2020).

Proposição 1.2 (Subaditividade). *Seja $(\Omega, \mathbb{A}, \mathbb{P})$ um espaço de probabilidade. Se $A_1, A_2, \dots \in \mathbb{A}$ são eventos quaisquer, então,*

$$\mathbb{P}\left(\bigcup_{i=1}^{+\infty} A_i\right) \leq \sum_{i=1}^{+\infty} \mathbb{P}(A_i);$$

Demonstração. Para demonstrar esta proposição, vamos definir conjuntos que sejam disjuntos e cuja união seja igual à união dos conjuntos A_i . Sejam

$$\begin{aligned} B_1 &= A_1 \\ B_2 &= A_2 - A_1 \\ B_3 &= A_3 - (A_1 \cup A_2) \\ &\vdots \\ B_n &= A_n - (A_1 \cup A_2 \cup \dots \cup A_{n-1}) \\ &\vdots \end{aligned}$$

Observe que $B_i = A_i - \bigcup_{j=1}^{i-1} A_j$, para todo $i = \{1, 2, 3, \dots\}$. Logo, os conjuntos B_i são disjuntos e a união é igual à união dos conjuntos A_i .

Como $B_i \subset A_i$, pelo item (vi) da Proposição (1.4) temos que $\mathbb{P}(B_i) \leq \mathbb{P}(A_i)$, $\forall i$. Além disso, $\bigcup_{i=1}^{+\infty} A_i = \bigcup_{i=1}^{+\infty} B_i$. Portanto,

$$\mathbb{P}\left(\bigcup_{i=1}^{+\infty} A_i\right) = \mathbb{P}\left(\bigcup_{i=1}^{+\infty} B_i\right) = \sum_{i=1}^{+\infty} \mathbb{P}(B_i) \leq \sum_{i=1}^{+\infty} \mathbb{P}(A_i),$$

em que na segunda igualdade usamos o item (ii) da Definição (1.6), já que os conjuntos B_i são disjuntos. Na primeira igualdade e na desigualdade usamos as duas observações dadas acima. \square

1.1.1 Probabilidade Condicional

Definição 1.9. *Seja $(\Omega, \mathbb{A}, \mathbb{P})$ um espaço de probabilidade. Dado dois eventos $A, B \in \mathbb{A}$, sendo $\mathbb{P}(B) > 0$, definimos a probabilidade condicional de A dado B por*

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

Intuitivamente, condicionar o evento A ao evento B significa restringir o espaço amostral ao evento B .

Vejamos um teorema que mostra a relação entre a probabilidade de A acontecer dado que B aconteceu e a probabilidade de B acontecer dado que A aconteceu.

Teorema 1.1 (Teorema de Bayes). *Sejam B_1, B_2, \dots eventos, todos de probabilidade positiva, que particionem o espaço Ω , ou seja, são disjuntos e sua união é igual a Ω . Dado um evento A , vale que*

$$(a) \quad \mathbb{P}(A) = \sum_{k=1}^{\infty} \mathbb{P}(A|B_k) \cdot \mathbb{P}(B_k);$$

(b) *Suponha que $\mathbb{P}(A) > 0$. Então, para qualquer índice $i \in \mathbb{N}$,*

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i) \cdot \mathbb{P}(B_i)}{\sum_{k=1}^{\infty} \mathbb{P}(A|B_k) \cdot \mathbb{P}(B_k)}.$$

Demonstração. Sejam B_1, B_2, \dots eventos, todos de probabilidade positiva, que particionem o espaço amostral Ω .

(a) Como B_1, B_2, \dots particionam Ω , temos que $\Omega = \bigcup_{k=1}^{\infty} B_k$.

Considere um evento qualquer $A \in \mathbb{A}$. Logo, $\mathbb{A} = A \cap \Omega = A \cap \left(\bigcup_{k=1}^{\infty} B_k\right) = \bigcup_{k=1}^{\infty} (A \cap B_k)$.

Como os B_k são disjuntos, temos que $A \cap B_k$ também são disjuntos. Daí, $\mathbb{P}(A) = \mathbb{P}(\bigcup_{k=1}^{\infty} (A \cap B_k))$. Pelo item (ii) da Definição (1.6) dos Axiomas de Kolmogorov, segue que

$$\mathbb{P}(A) = \sum_{k=1}^{\infty} \mathbb{P}(A \cap B_k).$$

Pela Definição (1.9), temos que

$$\mathbb{P}(A) = \sum_{k=1}^{\infty} \mathbb{P}(A|B_k) \cdot \mathbb{P}(B_k).$$

(b) Temos que

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

Pela Definição (1.9) e pelo item (a) da demonstração anterior, temos que

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i) \cdot \mathbb{P}(B_i)}{\sum_{k=1}^{\infty} \mathbb{P}(A|B_k) \cdot \mathbb{P}(B_k)}.$$

□

Exemplo 1.6. Duas máquinas, M_1 e M_2 , são responsáveis por 60% e 40%, respectivamente, da produção de uma empresa. Os índices de peças defeituosas na produção destas máquinas valem 3% e 7% respectivamente. Se uma peça defeituosa foi selecionada da produção desta empresa, qual é a probabilidade de que tenha sido produzida pela máquina M_2 ?

Inicialmente, vamos definir os eventos:

$$M_1 = [\text{peça produzida por } M_1];$$

$$M_2 = [\text{peça produzida por } M_2];$$

$$D = [\text{peça defeituosa}].$$

Queremos encontrar $\mathbb{P}(M_2|D)$.

Pelo Teorema de Bayes, segue que

$$\mathbb{P}(M_2|D) = \frac{\mathbb{P}(D|M_2)\mathbb{P}(M_2)}{\mathbb{P}(D|M_1)\mathbb{P}(M_1) + \mathbb{P}(D|M_2)\mathbb{P}(M_2)}.$$

Pelo enunciado da questão, temos que

$$\mathbb{P}(M_1) = 0,60,$$

$$\mathbb{P}(M_2) = 0,40,$$

$$\mathbb{P}(D|M_1) = 0,03,$$

$$\mathbb{P}(D|M_2) = 0,07.$$

Daí,

$$\mathbb{P}(M_2|D) = \frac{0,07(0,40)}{0,03(0,60) + 0,07(0,40)} = \frac{0,028}{0,046} \approx 0,6087 = 60,87\%.$$

Portanto, a probabilidade de uma peça escolhida ao acaso ser produzida pela máquina M_2 dado que ela é defeituosa é de aproximadamente 60,87%.

1.1.2 Eventos Independentes

Definição 1.10. Dizemos que dois eventos A e B são independentes se

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

Logo, dois eventos são independentes quando a ocorrência de um não afeta a probabilidade de ocorrência do outro. Observe que, se B tiver probabilidade positiva, temos que $\mathbb{P}(A|B) = \mathbb{P}(A)$. Analogamente, se A tiver probabilidade positiva, então $\mathbb{P}(B|A) = \mathbb{P}(B)$.

Definição 1.11. Dizemos que os eventos $A_1, A_2, \dots \in \mathbb{A}$ são independentes dois a dois se, para quaisquer $i \neq j$, vale

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i) \cdot \mathbb{P}(A_j),$$

e dizemos que $A_1, A_2, \dots \in \mathbb{A}$ são coletivamente independentes (ou simplesmente independentes) se para qualquer família finita de índices $1 \leq i_1 < i_2 < \dots < i_l$ vale

$$\mathbb{P}\left(\bigcap_{k=1}^l A_{i_k}\right) = \prod_{k=1}^l \mathbb{P}(A_{i_k}).$$

Exemplo 1.7. Uma experiência consiste em lançar simultaneamente um dado e uma moeda. Qual a probabilidade de obter a face 4 no dado e uma cara na moeda?

A princípio definiremos os eventos:

$$A = [\text{Face 4 no dado}];$$

$$B = [\text{Cara na moeda}].$$

Como são eventos independentes, temos que

$$\begin{aligned} \mathbb{P}(A \cap B) &= \mathbb{P}(A) \cdot \mathbb{P}(B) \\ \mathbb{P}(A \cap B) &= \frac{1}{6} \cdot \frac{1}{2} = \frac{1}{12}. \end{aligned}$$

1.2 Variáveis Aleatórias

Dado um fenômeno aleatório qualquer, com um certo espaço amostral, desejamos estudar a estrutura probabilística de quantidades associadas a esse fenômeno.

Por exemplo, ao descrever uma peça manufaturada podemos empregar duas classificações: defeituosa ou não defeituosa. Para facilitar a análise, vamos atribuir um número real a cada resultado do experimento. Assim podemos atribuir o valor zero às peças não defeituosas e o valor um às defeituosas.

Este característico numérico do resultado de um experimento é chamado de variável aleatória. Nós podemos entender por variável aleatória uma função que associa a cada elemento do espaço amostral (no exemplo anterior os elementos são defeituosa e não defeituosa) um número real. Assim, uma variável aleatória é um número real associado a um resultado aleatório.

Definição 1.12. *Seja $(\Omega, \mathbb{A}, \mathbb{P})$ um espaço de probabilidade. Dizemos que uma função $X : \Omega \rightarrow \mathbb{R}$ é uma variável aleatória se, para todo intervalo $I \in \mathbb{R}$, vale*

$$\{\omega \in \Omega : X(\omega) \in I\} \in \mathbb{A}.$$

As variáveis aleatórias são fundamentais para as aplicações, pois elas representam as características de interesse em uma população.

Exemplo 1.8. *Em uma linha de usinagem de peças estamos interessados em controlar o diâmetro das peças produzidas. Neste caso, o resultado da medição do diâmetro é a variável aleatória de interesse.*

Exemplo 1.9. *Em um ensaio clínico, estamos interessados em avaliar o tempo de vida dos pacientes. Neste caso, o tempo de vida corresponde à variável aleatória.*

Definição 1.13. *(a) Uma variável aleatória X é dita discreta se existe um conjunto enumerável $B = \{x_1, x_2, \dots\} \subset \mathbb{R}$ tal que*

$$\mathbb{P}(X \in B) = 1.$$

(b) Uma variável aleatória X é dita absolutamente contínua se existe uma função $f : \mathbb{R} \rightarrow \mathbb{R}_+$ digamos, contínua por partes, tal que

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f(x)dx, \text{ para quaisquer } a, b \in \mathbb{R}.$$

Esta função f é chamada de densidade da variável aleatória X . Note que sempre vale

$$\int_{-\infty}^{+\infty} f(x)dx = 1.$$

Vejamos alguns exemplos de variáveis aleatórias. Os primeiros quatro exemplos a seguir são de variáveis aleatórias discretas. Os demais são de variáveis aleatórias absolutamente contínuas.

Exemplo 1.10. *Uma variável aleatória X é dita ter Distribuição Bernoulli de parâmetro $p \in (0, 1)$ se*

$$\mathbb{P}(X = 1) = p \text{ e } \mathbb{P}(X = 0) = 1 - p.$$

Notação: $X \sim \text{Bernoulli}(p)$.

Observe que uma variável aleatória de Bernoulli é aquela que assume apenas dois valores, 1 para sucesso e 0 para fracasso.

Um exemplo clássico de uma variável aleatória de Bernoulli é uma jogada única de uma moeda. A moeda pode dar coroa com probabilidade p e cara com probabilidade $1 - p$.

Exemplo 1.11. *Uma variável aleatória X é dita ter Distribuição Geométrica de parâmetro $p \in (0, 1)$ se*

$$\mathbb{P}(X = k) = (1 - p)^{k-1}p.$$

Notação: $X \sim \text{geom}(p)$.

Uma variável aleatória geométrica representa o número de tentativas até obter o primeiro sucesso repetindo um experimento cujo resultado é sucesso com probabilidade p e fracasso com probabilidade $1 - p$.

Exemplo 1.12. *Uma variável aleatória X é dita ter Distribuição Binomial de parâmetros $n \in \mathbb{N}$ e $p \in (0, 1)$ se*

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

Notação: $X \sim \text{binom}(n, p)$.

Uma variável aleatória binomial representa o número de sucessos quando repetimos n vezes, de maneira independente, um experimento cujo resultado é sucesso com probabilidade p e fracasso com probabilidade $1 - p$.

Exemplo 1.13. *Uma variável aleatória X é dita ter Distribuição Poisson de parâmetro $\lambda > 0$ se*

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

Notação: $X \sim \text{Poisson}(\lambda)$.

Uma variável aleatória de Poisson é aplicável quando o número de possíveis ocorrências discretas é muito maior do que o número médio de ocorrências em um determinado intervalo de tempo ou espaço.

Algumas aplicações são: acidentes envolvendo automóveis, número de usuários de computador ligados à internet, números de carros que chegam ao posto de gasolina, quantidade de clientes que passam no caixa de um supermercado e erros de digitação que ocorrem em determinado período de tempo.

Exemplo 1.14. Uma variável aleatória X é dita ter Distribuição Uniforme no intervalo $[c, d]$ se tem densidade

$$f(x) = \begin{cases} \frac{1}{d-c}, & x \in [c, d], \\ 0, & x \notin [c, d]. \end{cases}$$

Notação: $X \sim U[c, d]$.

Na Figura (□1) há uma ilustração da densidade desta distribuição.

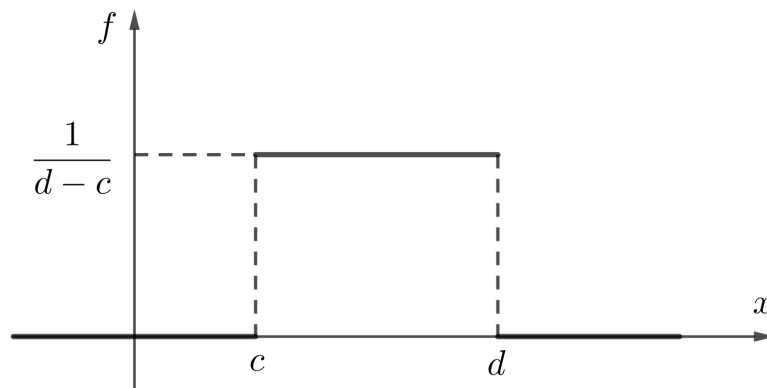


Figura 1.1: Densidade da variável aleatória $X \sim U[c, d]$.

Um caso particular importante é $X \sim U[0, 1]$. Assim,

$$\mathbb{P}(a < X < b) = \int_a^b 1 \, dx = b - a,$$

para quaisquer $0 \leq a < b \leq 1$.

Na variável aleatória uniforme, a probabilidade de gerar qualquer ponto em um intervalo contido no espaço amostral é proporcional ao tamanho do intervalo.

Exemplo 1.15. Uma variável aleatória X é dita ter Distribuição Exponencial de parâmetro $\lambda > 0$ se tem densidade

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Notação: $X \sim \exp(\lambda)$.

Na Figura (□2) há uma ilustração da densidade desta distribuição.

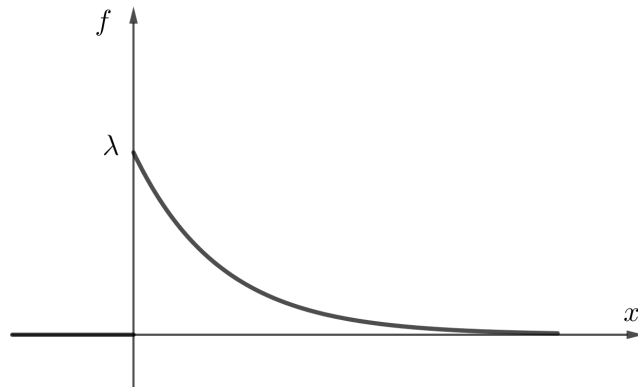


Figura 1.2: Densidade da variável aleatória $X \sim \exp(\lambda)$.

A variável aleatória exponencial pode ser utilizada para descrever as probabilidades envolvidas no tempo que decorre para que um determinado evento aconteça.

Há uma conexão muito próxima entre a distribuição Exponencial e a de Poisson. As relações entre essas duas distribuições podem ser associadas a um processo estocástico, chamado de processo de Poisson. Para exemplificar: se queremos caracterizar o número de gols por partida de futebol, essa variável aleatória é uma Poisson. Ademais, podemos ainda caracterizar o tempo entre essas ocorrências, sendo que essa variável aleatória é uma Exponencial.

Exemplo 1.16. Uma variável aleatória X é dita ter Distribuição Normal ou Gaussiana de parâmetros μ e σ^2 , sendo μ a média e σ^2 a varância da distribuição, se tem densidade

$$f(x) = \frac{e^{-\frac{x-\mu}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}.$$

Notação: $X \sim N(\mu, \sigma)$.

A Figura (1.3) ilustra a densidade da variável aleatória gaussiana.

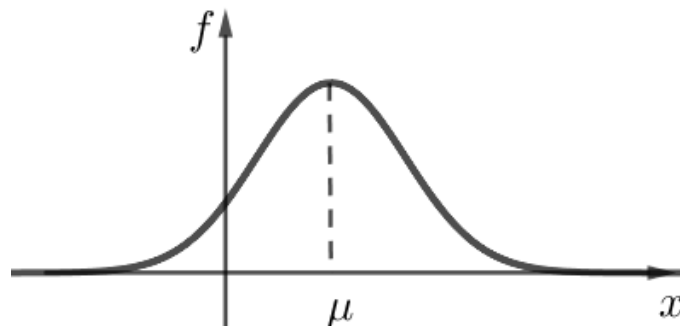


Figura 1.3: Densidade da variável aleatória $X \sim N(\mu, \sigma)$.

A distribuição normal é a mais importante distribuição contínua. Sua importância se deve a vários fatores, entre eles podemos citar o Teorema Central do Limite, que é um resultado fundamental em aplicações práticas e teóricas, pois ele garante que, mesmo que

os dados não sejam distribuídos segundo uma normal, a média dos dados converge para uma distribuição normal conforme o número de dados aumenta.

Além disso, diversos estudos práticos têm como resultado uma distribuição normal, sendo uma das mais utilizadas para modelar fenômenos naturais.

Podemos citar como exemplos a altura ou o peso de uma determinada população em geral, a pressão sanguínea de um grupo de pessoas, o tempo que um grupo de estudantes gasta para realizar uma prova e o número de casos novos da COVID-19 no Brasil.

Um caso particular muito importante é quando temos a distribuição normal padrão, que é a distribuição normal com $\mu = 0$ e $\sigma^2 = 1$.

1.2.1 Valor Esperado, Esperança ou Média

Exemplo 1.17. Em um restaurante quando pedimos nossa comida e perguntamos para o garçom quanto tempo leva para ficar pronto, ele vai nos fornecer um valor esperado, ou seja, o tempo médio em que a comida deve demorar para ficar pronta.

Exemplo 1.18. Quando estamos em um ponto de ônibus e perguntamos para a pessoa ao lado quanto tempo leva até que o próximo ônibus venha, ela prontamente vai nos dar o valor esperado, o qual ela conseguiu constatar depois de algum tempo de experiência.

Nos exemplos acima, o garçom e a pessoa que esperava no ponto de ônibus resumiram toda a informação de um modelo em um único número, o valor esperado. A esperança de uma variável aleatória é o valor médio esperado como resultado da respectiva variável aleatória, o qual não precisa ser um valor possível de ser observado, conforme definição a seguir.

Definição 1.14. Seja X uma variável aleatória.

(a) Se X é discreta assumindo os valores $\{x_1, x_2, \dots\}$, então definimos a esperança de X (caso exista) como

$$\mathbb{E}X = \sum_{k=1}^{+\infty} x_k \mathbb{P}(X = x_k).$$

(b) Se X é absolutamente contínua com densidade $f : \mathbb{R} \rightarrow \mathbb{R}_+$, então definimos a esperança de X (caso exista) como

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

Para evitar ambiguidades, é comum escrever $\mathbb{E}[X]$ em vez de $\mathbb{E}X$ para enfatizar sobre qual variável aleatória (ou quais variáveis aleatórias) a esperança está sendo tomada.

Exemplo 1.19. *Considere que em um certo jogo o apostador ganhe 1 real com probabilidade $\frac{1}{4}$ e perca 1 com probabilidade $\frac{3}{4}$. Desta forma, o valor esperado da variável aleatória X , que representa o seu ganho, será*

$$\mathbb{E}X = 1 \cdot \left(\frac{1}{4}\right) + (-1) \cdot \left(\frac{3}{4}\right) = -\frac{1}{2}.$$

Note que $-\frac{1}{2}$ não é um valor possível de ser observado.

Proposição 1.3. (a) *(Monotonicidade) Se $X \leq Y$ variáveis aleatórias, então $\mathbb{E}X \leq \mathbb{E}Y$;*

(b) *(Linearidade) Se $Z = aX + Y$, então $\mathbb{E}Z = a\mathbb{E}X + \mathbb{E}Y$;*

(c) *(Esperança do Produto) Se X e Y são variáveis aleatórias independentes, então $\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$;*

(d) *(Mudança de Variáveis) Dada $h : \mathbb{R} \rightarrow \mathbb{R}$ e X variável aleatória, então*

$$\mathbb{E}[h(X)] = \sum_{k=1}^{+\infty} h(x_k) \mathbb{P}(X = x_k),$$

se X é discreta, ou

$$\mathbb{E}[h(X)] = \int_{-\infty}^{+\infty} h(x) f(x) dx,$$

se X for absolutamente contínua.

1.2.2 Variância

Suponha que X seja a duração de vida de lâmpadas que estão sendo recebidas de um fabricante e que $\mathbb{E}(X) = 1000$ horas. Isto pode significar que a maioria das lâmpadas devem durar um período de tempo compreendido entre 900 horas e 1100 horas. Poderia significar também que as lâmpadas são formadas por dois tipos muito diferentes: cerca da metade é de muita boa qualidade e durará aproximadamente 1400 horas, enquanto que a outra metade é de muito má qualidade e durará aproximadamente 600 horas.

Assim, existe uma necessidade óbvia de se introduzir uma medida que possa distinguir entre essas duas situações.

Na Teoria da Probabilidade, a variância de uma variável aleatória é uma medida da sua dispersão, indicando o “quão longe” em geral seus valores se encontram do valor esperado.

Definição 1.15. *A variância de uma variável aleatória X é definida por*

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

caso a esperança acima exista. Em palavras, a variância de uma variável aleatória representa uma medida de dispersão da variável aleatória X em torno da média.

Observação: Esta fórmula serve tanto para variáveis discretas, quanto para variáveis contínuas.

Proposição 1.4. *A variância satisfaz*

(a) $Var(X + c) = Var(X)$, para qualquer $c \in \mathbb{R}$;

(b) $Var(aX) = a^2 Var(X)$, para qualquer $a \in \mathbb{R}$;

(c) $Var(X) = \mathbb{E}[X^2] - (\mathbb{E}X)^2$;

(d) Se X e Y são independentes, então $Var(X + Y) = Var(X) + Var(Y)$;

(e) Se X_1, \dots, X_n são independentes, então $Var\left(\sum_{k=1}^n X_k\right) = \sum_{k=1}^n Var(X_k)$.

Demonstração. Os itens (a) e (b) são consequência imediata da linearidade da esperança.

Para o item (c),

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

Para o item (d),

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

Para o item (e), basta fazer indução. □

1.3 Desigualdade de Tchebyshev

Veremos agora a Desigualdade de Tchebyshev, um importante resultado envolvendo esperança de variáveis aleatórias. Faremos uso desta desigualdade nas demonstrações da Lei Fraca e da Lei Forte dos Grandes Números. Antes, precisamos definir o que é a Função Indicadora.

Definição 1.16 (Função Indicadora). *Dado um conjunto A , definimos a chamada função indicadora $\mathbb{1}_A$ por*

$$\mathbb{1}_A(x) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A. \end{cases}$$

Proposição 1.5 (Desigualdade de Tchebyshev). *Sejam $X \geq 0$ variável aleatória e $\lambda > 0$ um número real. Então*

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

Demonstração. Considere a decomposição de X utilizando a função indicadora da seguinte maneira:

$$X = X\mathbf{1}_{[X \geq \lambda]} + X\mathbf{1}_{[X < \lambda]}.$$

Pela linearidade da esperança, temos que

$$\mathbb{E}X = \mathbb{E}[X\mathbf{1}_{[X \geq \lambda]}] + \mathbb{E}[X\mathbf{1}_{[X < \lambda]}].$$

Como $X \geq 0$, temos que as três esperanças acima são não negativas. Por linearidade e monotonicidade, segue que

$$\mathbb{E}X \geq \mathbb{E}[X\mathbf{1}_{[X \geq \lambda]}] \geq \mathbb{E}[\lambda\mathbf{1}_{[X \geq \lambda]}] = \lambda\mathbb{E}[\mathbf{1}_{[X \geq \lambda]}] = \lambda\mathbb{P}(X \geq \lambda).$$

Portanto,

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

□

1.4 Lema de Borel-Cantelli

O Lema de Borel-Cantelli é um célebre lema da Teoria das Probabilidades denominado desta forma em homenagem a dois matemáticos probabilistas: Émile Borel e Francesco Paolo Cantelli. Ele garante que se a probabilidade de algo acontecer for finita, então as chances são zero de acontecer infinitas vezes. Além disso, ele afirma que se algo possui uma chance de acontecer e for tentado um número grande de vezes, de maneira independente, isto acontecerá com 100% de chance, mesmo que esta chance seja ínfima.

Utilizaremos o Lema de Borel-Cantelli para demonstrarmos a Lei Forte dos Grandes Números.

Antes de enunciarmos o lema, vamos definir uma operação entre conjuntos. Dados os conjuntos A_1, A_2, \dots , definimos

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

Observe que $\limsup_n A_n$ é o conjunto de resultados que ocorrem infinitas vezes dentro da sequência de eventos infinita A_n . Logo, ele pode ser descrito como o conjunto

$$\{\omega : \omega \text{ pertence a } A_n \text{ para infinitos índices } n \in \mathbb{N}\}.$$

Assim, é costume denotar $\limsup_n A_n$ por $[A_n \text{ infinitas vezes}]$ ou simplesmente por $[A_n \text{ i. v.}]$

Proposição 1.6 (Lema de Borel-Cantelli). *Seja $(\Omega, \mathbb{A}, \mathbb{P})$ um espaço de probabilidade e considere os eventos $A_1, A_2, \dots \in \mathbb{A}$.*

(a) *Se $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, então $\mathbb{P}(A_n \text{ i.v.}) = 0$;*

(b) *Se A_1, A_2, \dots são independentes e $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$, então $\mathbb{P}(A_n \text{ i.v.}) = 1$.*

Demonstração. Considere os eventos $A_1, A_2, \dots \in \mathbb{A}$.

(a) Como $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, temos que esta série é convergente. Desta forma, $\lim_{k \rightarrow \infty} \sum_{n=k}^{\infty} \mathbb{P}(A_n) = 0$. Como para qualquer $n \in \mathbb{N}$,

$$\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k \subset \bigcup_{k=n}^{\infty} A_k,$$

Pela monotonicidade da probabilidade, detalhada no item (vi) da Proposição (□□), temos que

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k\right) \leq \mathbb{P}\left(\bigcup_{k=n}^{\infty} A_k\right) \leq \sum_{k=n}^{\infty} \mathbb{P}(A_k),$$

em que na última desigualdade utilizamos a subaditividade descrita na Proposição (□□). Como a série é convergente, deduzimos que o lado esquerdo no começo das desigualdades é nulo. Portanto,

$$\mathbb{P}(A_n \text{ i.v.}) = 0;$$

(b) Para provarmos este item, precisamos mostrar que $\mathbb{P}(\bigcup_{n=k}^{\infty} A_n) = 1$, para todo natural k , pois a interseção enumerável de eventos de probabilidade um também tem probabilidade igual a um. Para isso definimos

$$B_n := \bigcup_{k=n}^{\infty} A_k \supset \bigcup_{k=n}^m A_k.$$

Aqui, é mais conveniente trabalhar com eventos complementares. Logo, usando as Leis de De Morgan e independência de eventos (Definição □□), obtemos

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

Aplicando a desigualdade $1 - x \leq e^{-x}$, que vale para todo x real, $0 \leq x \leq 1$, segue que

$$\mathbb{P}(B_n^c) \leq \exp\left(-\sum_{k=n}^m \mathbb{P}(A_k)\right).$$

Ademais, como a soma das probabilidades de A_k diverge, deduzimos que $\mathbb{P}(B_n^c) = 0$, ao tomar $n \rightarrow \infty$, ou seja, $\mathbb{P}(B_n) = 1$

□

Exemplo 1.20. *Um exemplo clássico do Lema de Borel-Cantelli é o do macaco que escreve Shakespeare. Ao colocarmos um macaco em frente de uma máquina de escrever, em um intervalo de tempo infinito ele irá, com probabilidade 1, escrever a obra completa de William Shakespeare. Supondo que as realizações sejam independentes e as repetições infinitas, ele não escreverá apenas uma vez, mas infinitas vezes.*

Capítulo 2

Lei Fraca dos Grandes Números

Iremos apresentar neste capítulo a Lei Fraca dos Grandes Números, onde salienta que a média amostral de um determinado experimento converge em probabilidade para a média populacional, ou seja, a probabilidade da média dessas variáveis aleatórias se afastarem da média populacional tende a zero quando o tamanho dessas amostras crescem.

Exemplo 2.1. *Considere o lançamento de um dado não viciado de forma aleatória. Seja o evento em que obtemos o número 2. Logo, a probabilidade de que o número 2 apareça é $\frac{1}{6}$ (o dado tem 6 faces, uma delas é 2).*

A Lei dos Grandes Números nos diz que a medida que aumentamos o número de repetições da nossa experiência (fazemos mais lançamentos do dado), a frequência relativa com que o evento se repetirá (obtermos 2) ficará mais próxima de uma constante, que terá um valor igual à sua probabilidade ($\frac{1}{6}$ ou 16,67%).

É possível que nos primeiros 10 ou até 20 lançamentos, a frequência com que conseguiremos 2 não será de 16%, mas de outro percentual. Porém, à medida que fazemos mais e mais lançamentos, a frequência com que o 2 aparece será muito próxima de 16,67%.

A seguir veremos o tipo de convergência que utilizaremos para demonstrar a Lei Fraca dos Grandes Números.

2.1 Convergência em Probabilidade

Definição 2.1. *Dadas as variáveis aleatórias X e X_n , com $n \in \mathbb{N}$, dizemos que a sequência X_n converge a X em probabilidade se, para todo $\varepsilon > 0$,*

$$\lim_{n \rightarrow +\infty} \mathbb{P}(|X - X_n| > \varepsilon) = 0.$$

Notação: $X_n \xrightarrow{P} X$.

A convergência em probabilidade de X_n para X afirma que, para qualquer $\varepsilon > 0$, a probabilidade de X_n estar a uma distância maior do que ε de X se torna arbitrariamente pequena quando n cresce.

2.2 Lei Fraca dos Grandes Números

Vejamos o enunciado e a demonstração da Lei Fraca dos Grandes Números.

Teorema 2.1 (Lei Fraca dos Grandes Números). *Sejam X_1, X_2, \dots variáveis aleatórias independentes tais que $\mathbb{E}[X_k] = \mu$ e $\mathbb{E}[X_k^2] \leq c < \infty, \forall k \in \mathbb{N}$. Denote $S_n = X_1 + X_2 + \dots + X_n$. Então,*

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

Demonstração. Queremos mostrar que para todo $\varepsilon > 0$

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) \rightarrow 0,$$

quando $n \rightarrow \infty$.

Note que, multiplicando toda a inequação por n e elevando ao quadrado toda a inequação, temos

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) = \mathbb{P}(|S_n - n\mu| > n\varepsilon) = \mathbb{P}(|S_n - n\mu|^2 > n^2\varepsilon^2).$$

Pela Desigualdade de Tchebyshev, temos que

$$\mathbb{P}(|S_n - n\mu|^2 > n^2\varepsilon^2) \leq \frac{\mathbb{E}[(S_n - n\mu)^2]}{n^2 \cdot \varepsilon^2}. \quad (2.1)$$

Vamos calcular $\mathbb{E}[(S_n - n\mu)^2]$ separadamente. Observe que, usando a linearidade da esperança, temos que

$$\begin{aligned} \mathbb{E}[S_n] &= \mathbb{E}[X_1 + X_2 + \dots + X_n] \\ &= \mathbb{E}[X_1] + \mathbb{E}[X_2] + \dots + \mathbb{E}[X_n] \\ &= n\mu. \end{aligned}$$

Assim,

$$\begin{aligned} \mathbb{E}[(S_n - n\mu)^2] &= \text{Var}(S_n) \\ &= \text{Var}\left(\sum_{i=1}^n X_i\right) \\ &= \sum_{i=1}^n \text{Var}(X_i), \end{aligned} \quad (2.3)$$

em que na última igualdade usamos o item (e) da Proposição (□4).

Além disso, pelo item (c) da Proposição (□4), segue que

$$\begin{aligned}\mathbb{E}[(S_n - n\mu)^2] &= \sum_{i=1}^n (\mathbb{E}[X_i^2] - (\mathbb{E}[X_i])^2) \\ &\leq nc - n\mu^2 \\ &\leq nc.\end{aligned}$$

Portanto, na expressão (□1) temos que

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) \leq \frac{\mathbb{E}[(S_n - n\mu)^2]}{n^2\varepsilon^2} \leq \frac{nc}{n^2\varepsilon^2} = \frac{c}{n\varepsilon^2},$$

que tende a zero quando $n \rightarrow \infty$, concluindo a demonstração. \square

Exemplo 2.2. 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:

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

De fato, seja

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

Então, X_1, X_2, \dots são independentes, identicamente distribuídas e integráveis com média $\mu = p$. Além disso, $\mathbb{E}(X_i^2) < \infty$. Portanto, a Lei Fraca dos Grandes Números implica que

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

ou, equivalentemente,

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

Capítulo 3

Lei Forte dos Grandes Números

Neste capítulo, veremos a Lei Forte dos Grandes Números. Ela é similar à Lei Fraca dos Grandes Números, porém utilizando uma noção mais forte de convergência. Enquanto a Lei Fraca assegura que para um valor grande de n , a média $\frac{S_n}{n}$ é próxima de μ com alta probabilidade, a lei não informa que, uma vez estando próxima de μ , a sequência de médias permanecerá próxima de μ , que é o que garante a Lei Forte dos Grandes Números.

Inicialmente, vejamos a definição da convergência utilizada nesta lei e sua relação com a convergência em probabilidade.

3.1 Convergência Quase Certa

Definição 3.1 (Convergência Quase Certa). *Sejam X, X_1, X_2, \dots variáveis aleatórias definidas no espaço de probabilidade $(\Omega, \mathbb{A}, \mathbb{P})$. Dizemos que a sequência X_n converge quase certamente para X se existe um evento $A \in \mathbb{A}$ tal que $\mathbb{P}(A) = 1$ e, para todo $\omega \in A$, vale*

$$\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega).$$

Notação: $X_n \rightarrow X$ q.c.

A convergência quase certa diz que, com probabilidade 1, a sequência dos resultados X_n converge para o resultado X . Isto significa que os valores de X_n se aproximam do valor de X no sentido de que os eventos para os quais X_n não converge a X têm probabilidade zero.

Vejamos algumas propriedades da convergência quase certa.

Proposição 3.1. *Se $X_n \rightarrow X$ q.c. e $Y_n \rightarrow Y$ q.c., então:*

- (a) $X_n + Y_n \rightarrow X + Y$ q.c.;
- (b) Se $c \in \mathbb{R}$, então $X_n + c \rightarrow X + c$ q.c.;
- (c) Para todo $\alpha \in \mathbb{R}$, vale que $\alpha X_n \rightarrow \alpha X$ q.c.;

(d) $X_n - Y_n \rightarrow X - Y$ q.c.;

(e) Se $f : \mathbb{R} \rightarrow \mathbb{R}$ é contínua, então $f(X_n) \rightarrow f(X)$ q.c.;

(f) Se $\mathbb{P}[Y = 0] = \mathbb{P}[Y_n = 0] = 0$, para todo $n \in \mathbb{N}$, então $X_n/Y_n \rightarrow X/Y$ q.c.;

Proposição 3.2. Se $X_n \rightarrow X$ q.c., então $X_n \xrightarrow{P} X$.

Demonstração. Defina o conjunto

$$B = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} [|X_k - X| > \varepsilon].$$

Afirmção: $\mathbb{P}(B) = 0$.

De fato, como $X_n \rightarrow X$ q.c., temos que existe um evento $A \in \mathbb{A}$ tal que $\mathbb{P}(A) = 1$ e, para todo $w \in \mathbb{A}$,

$$\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega).$$

Observe que $B \subset A^c$. Logo, pela monotonicidade da probabilidade (item (vi) da Proposição [3.1](#)), temos que $\mathbb{P}(B) \leq \mathbb{P}(A^c)$. Além disso, como $\mathbb{P}(A) = 1$, temos que $\mathbb{P}(A^c) = 0$. Daí, $\mathbb{P}(B) \leq \mathbb{P}(A^c) = 0$. Já que $\mathbb{P}(B) \geq 0$, concluimos que $\mathbb{P}(B) = 0$.

Agora, defina

$$B_n = \bigcup_{k=n}^{\infty} [|X_k - X| > \varepsilon].$$

Como $B_n \searrow B$, pela continuidade da probabilidade (item (viii) da Proposição [3.1](#)), temos que $\mathbb{P}(B_n) \searrow \mathbb{P}(B) = 0$. Ademais, note que $[|X_n - X| > \varepsilon] \subset B_n$. Logo, novamente pela monotonicidade da probabilidade, $\mathbb{P}[|X_n - X| > \varepsilon] \leq \mathbb{P}(B_n)$. Portanto, $X_n \xrightarrow{P} X$. \square

3.2 Lei Forte dos Grandes Números

Vejamos agora o enunciado e a demonstração da Lei Forte dos Grandes Números.

Teorema 3.1 (Lei Forte dos Grandes Números). *Sejam X_1, X_2, \dots variáveis aleatórias independentes identicamente distribuídas com $\mathbb{E}(X_i) = \mu$ e $\mathbb{E}X_i^4 < \infty$. Seja $S_n = X_1 + X_2 + \dots + X_n$. Então,*

$$\frac{S_n}{n} \longrightarrow \mu \text{ q.c.}$$

Demonstração. Primeiramente, suponha que a média de X_i é igual a 0. Logo, iremos mostrar que $\frac{S_n}{n}$ converge para 0 quase certamente. Como $\mathbb{E}X_i^4 < \infty$ é finito, temos que

$$\mathbb{E}S_n^4 = \mathbb{E}[(X_1 + X_2 + \dots + X_n)^4].$$

$$\mathbb{E}S_n^4 = \mathbb{E}[(X_1 + X_2 + \dots + X_n)(X_1 + X_2 + \dots + X_n)(X_1 + X_2 + \dots + X_n)(X_1 + X_2 + \dots + X_n)].$$

Aplicando a distributividade nos fatores do valor esperado, obtemos uma expansão que resulta em termos $X_i^4, X_i^3 X_j, X_i^2 X_j^2, X_i^2 X_j X_k$ e $X_i X_j X_k X_l$, em que i, j, k e l são todos diferentes.

Pela linearidade da esperança, temos que

$$\mathbb{E}S_n^4 = \sum_{1 \leq i, j, k, l \leq n} \mathbb{E}[X_i X_j X_k X_l].$$

Como as variáveis aleatórias são independentes, o produto da esperança é a esperança do produto. Além disso, como por suposição todas as variáveis aleatórias X_i têm média 0, resulta da independência destas variáveis que

$$\mathbb{E}[X_i^3 X_j] = \mathbb{E}[X_i^3] \mathbb{E}[X_j] = 0;$$

$$\mathbb{E}[X_i^2 X_j X_k] = \mathbb{E}[X_i^2] \mathbb{E}[X_j] \mathbb{E}[X_k] = 0;$$

$$\mathbb{E}[X_i X_j X_k X_l] = 0.$$

Assim, para que um termo $\mathbb{E}[X_i X_j X_k X_l]$ seja não nulo, temos apenas duas possibilidades:

- (i) Os quatro índices são iguais, da forma $\mathbb{E}X_i^4$; ou
- (ii) São dois pares de índices iguais, da forma $\mathbb{E}[X_i^2 X_j^2] = (\mathbb{E}X_i^2)^2$, onde esta última igualdade é devido às variáveis X_i possuírem a mesma distribuição.

Vamos calcular quantos de cada um destes termos nós temos. Para quatro índices iguais, temos n possibilidades. Já para escolher dois pares de índices iguais (sendo os pares diferentes), temos $\binom{n}{2} = \frac{n!}{2!(n-2)!} = \frac{n(n-1)}{2}$ possibilidades. Ademais, estes dois pares podem aparecer de $\binom{4}{2} = \frac{4!}{2!(4-2)!} = 6$ maneiras diferentes dentre i, j, k, l . Logo,

$$\mathbb{E}S_n^4 = n \cdot \mathbb{E}X_i^4 + 6 \frac{n(n-1)}{2} \cdot (\mathbb{E}X_i^2)^2.$$

Pela hipótese de independência, como

$$\text{Var}(X_i^2) = \mathbb{E}[X_i^4] - (\mathbb{E}[X_i^2])^2 \geq 0,$$

temos que

$$(\mathbb{E}[X_i^2])^2 \leq \mathbb{E}[X_i^4] < \infty.$$

Daí, do desenvolvimento anterior, obtemos que

$$\begin{aligned} \mathbb{E}S_n^4 &\leq n \cdot \mathbb{E}X_i^4 + 3n(n-1) \cdot \mathbb{E}X_i^4 \\ &= (3n^2 - 2n) \cdot \mathbb{E}X_i^4 \\ &\leq 3n^2 \cdot \mathbb{E}X_i^4 \\ &= cn^2, \end{aligned}$$

em que $c = 3\mathbb{E}X_i^4$, já que $\mathbb{E}X_i^4$ é finita.

Utilizando a Desigualdade de Tchebyshev, temos que

$$\mathbb{P}(|S_n| > n\varepsilon) \leq \frac{\mathbb{E}S_n^4}{(n\varepsilon)^4} \leq \frac{cn^2}{n^4\varepsilon^4} = \frac{c}{n^2\varepsilon^4}.$$

Logo, pelo Lema de Borel-Cantelli, temos que

$$\mathbb{P}\left(\frac{|S_n|}{n} > \varepsilon \text{ i.v.}\right) = 0,$$

o que implica que, a partir de certo natural n_0 aleatório, $\frac{|S_n|}{n}$ é sempre menor do que $\varepsilon > 0$. Como $\varepsilon > 0$ é arbitrário, temos que $\frac{S_n}{n} \rightarrow \mu$ q.c.

Agora, suponha que $\mu \neq 0$. Daí, aplicamos o argumento anterior às variáveis aleatórias $X'_i = X_i - \mu$, que possuem média zero. Portanto,

$$\frac{(X_1 - \mu) + (X_2 - \mu) + \cdots + (X_n - \mu)}{n} \rightarrow 0 \text{ q.c.},$$

o que implica que

$$\frac{S_n}{n} \rightarrow \mu \text{ q.c.}$$

□

Exemplo 3.1. *Uma aplicação da Lei Forte dos Grandes Números que está relacionada à Teoria da Renovação pode ser encontrada no livro (DURRETTI, 2005). A Teoria da Renovação é um campo especial da Teoria da Probabilidade que lida com processos que se comportam da maneira como se comportavam quando o experimento começou, após cada vez que o estado inicial foi alcançado.*

Sejam X_1, X_2, \dots variáveis aleatórias independentes identicamente distribuídas com $0 < X_i < \infty$. Considere $S_n = X_1 + \dots + X_n$ e pense em S_n como o tempo da n -ésima ocorrência de algum evento.

Pensando em uma situação concreta, considere um diligente zelador que substitui uma lâmpada no instante em que ela queima, ou seja, admitimos que o tempo que gastamos para identificar o problema e trocar de lâmpada é muito pequeno e pode ser assumido como instantâneo. Suponha que a primeira lâmpada é colocada no tempo 0 e seja X_i o tempo de vida da i -ésima lâmpada. Nesta interpretação, S_n é o tempo que a n -ésima lâmpada queima e $N_t = \sup\{n : S_n \leq t\}$ é o número de lâmpadas que queimaram até o tempo t .

Pela Lei Forte dos Grandes Números,

$$\frac{S_n}{n} \rightarrow \mathbb{E}X_i \text{ q.c.}$$

Outros exemplos seriam para trabalhos de manutenção mais gerais ou para tempos de atendimento ao cliente, que aparecem em um ponto de expedição de acordo com uma distribuição predeterminada e formam uma fila. Nestes casos, a Teoria de Renovação fornece informações para intervalos de manutenção ideais ou pessoal ideal nos pontos de serviço.

Capítulo 4

Conclusões e Perspectivas

Neste trabalho foi apresentada uma abordagem da Lei dos Grandes Números. Para tanto, foi utilizado o livro (FRANCO, 2020) como principal referência bibliográfica, além de (DURRETT, 2003), (JAMES, 2013), (CARVALHO et al., 2004) e (LIMA, 1976, Volume 1).

Primeiro, foram analisados alguns conceitos iniciais de probabilidade relevantes para o entendimento dos resultados principais. Neste processo, foram apresentados a Desigualdade de Tchebyshev e o Lema de Borel-Cantelli, ferramentas essenciais para as demonstrações da Lei Fraca e da Lei Forte dos Grandes Números.

Depois, dedicamos um capítulo para cada uma das leis, a Lei Fraca e a Lei Forte. Neles, definimos os tipos de convergência a serem utilizados e enunciamos e demonstramos os resultados.

Concluimos destacando que este trabalho provoque um interesse por parte dos leitores, que contemple situações em que acenda a necessidade de continuar com a pesquisa, levando-os a necessidade de investigar, visando que esse assunto possui diversos tópicos interessantes que podem ser explorados, abrindo um vasto campo de novos projetos.

Ademais, temos como foco facilitar a compreensão e acessibilidade dos alunos, visto que através do tema abordado e das demonstrações apresentadas, torna possível o acesso deste conteúdo que não é muito explorado.

Referências

CARVALHO, P. et al. Análise Combinatória e Probabilidade. 6. ed. Rio de Janeiro: SBM, 2004.

DURRETT, R. Probability: theory and examples. Cambridge University Press: Thomson-Brooks/Cole, 2005.

FRANCO, T. Princípios de Combinatória e Probabilidade. Rio de Janeiro: IMPA, 2020.

JAMES, B. R. Probabilidade: Um curso em nível intermedio. [S.l.]: IMPA, Rio de Janeiro, 2013.

LIMA, E. L. Curso de Análise. 8. ed. São Paulo: IMPA, 1976, Volume 1.