UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

PEMAT-IM-UFRJ

Thiago Vasconcellos Batalha

CONTRIBUIÇÕES À CONSTRUÇÃO DO RACIOCÍNIO ESTATÍSTICO VIA LEI FORTE DOS GRANDES NÚMEROS E TEOREMA CENTRAL DO LIMITE

Orientadores:

Prof. Nei Carlos dos Santos Rocha

Prof. Gérard Émile Grimberg

Rio de Janeiro

2017

Thiago Vasconcellos Batalha

CONTRIBUIÇÕES À CONSTRUÇÃO DO RACIOCÍNIO ESTATÍSTICO VIA LEI FORTE DOS GRANDES NÚMEROS E TEOREMA CENTRAL DO LIMITE

Dissertação apresentada à Universidade Federal do Rio de Janeiro, como requisito para a conclusão do curso de Mestrado em Ensino da Matemática

Rio de Janeiro

2017

Thiago Vasconcellos Batalha

CONTRIBUIÇÕES À CONSTRUÇÃO DO RACIOCÍNIO ESTATÍSTICO VIA LEI FORTE DOS GRANDES NÚMEROS E TEOREMA CENTRAL DO LIMITE

ESTE TRABALHO DE CONCLUSÃO DE CURSO FOI JULGADO E APROVADO PARA OBTENÇÃO DO GRAU DE MESTRADO EM ENSINO DA MATEMÁTICA NA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO.

Aprovado em

____

Prof. Gérard Émile Grimberg (Orientador)
Instituto de Matemática - UFRJ

____

Prof. Nei Carlos dos Santos Rocha (Orientador)
Instituto de Matemática - UFRJ

____

Prof. Victor Giraldo
Instituto de Matemática – UFRJ

____

Prof.ª Flávia Landim
Instituto de Matemática - UFRJ

____

Prof.ª Ana Teresa de Oliveira
Faculdade de Educação - UFRJ

____

Prof. Wanderley Moura Rezende
Instituto de Matemática e Estatística - UFF

AGRADECIMENTOS

A confecção deste trabalho foi possível graças a nobre atitude de diversas pessoas que decidiram compartilhar seus respectivos conhecimentos.

Agradeço a Deus pela vida de cada uma dessas pessoas que dedicam seu tempo a transmitir seus saberes e valores contribuindo não apenas para a elaboração desta dissertação, mas também para a minha formação ética e profissional.

Agradeço à minha noiva Joane Alves Borges, por toda motivação e apoio para iniciar e concluir o curso de mestrado, bem como por compartilhar um pouco de seu conhecimento.

Agradeço aos meus pais Maria Alice e Walter, pelo incentivo ao estudo.

Agradeço aos professores Nei Carlos dos Santos Rocha e Gérard Émile Grimbert, pelas orientações, apoio, dedicação e incentivos para a confecção deste projeto.

Agradeço aos professores Victor Giraldo e Tatiana Roque, pelos ensinamentos e pela dedicação ao PEMAT.

Agradeço à banca examinadora, pelos valiosos conselhos e sugestões a respeito deste projeto.

Agradeço à equipe de professores da UFRJ, por contribuírem com minha formação de educador.

Agradeço aos amigos Valmir, Bianca, Rafael, Ana, Rodolfo e Julyanne, pelo apoio, incentivo e companheirismo.

RESUMO

A Estatística é frequentemente tratada no Ensino Médio de forma protocolar, se restringindo a um conjunto de fórmulas que devem ser aplicadas para a obtenção de um valor numérico. A partir de uma perspectiva histórica observa-se uma crescente necessidade no desenvolvimento de ferramentas para se estudar e compreender os diversos tipos de populações com vias a tomadas de decisão sob incerteza. Nesse sentido, a Estatística é um campo fundamental para se compreender os fenômenos aleatórios, cada vez mais perceptíveis aos membros de qualquer sociedade. A urgência em compreender fenômenos regidos pelo acaso estabelece que as teorias probabilísticas e estatísticas não sejam estudadas separadamente; pelo contrário, evidencia a necessidade de se utilizar a Probabilidade dentro da área de atuação da Estatística, pois aquela traz em sua estrutura elementos fundamentais para a construção desta. Certamente, dessa união derivam-se ferramentas bastante sofisticadas, sendo necessária a compreensão de conceitos advindos do Cálculo Diferencial e Integral, bem como da Análise. Contudo, as ideias principais de muitas dessas ferramentas estatísticas podem ser compreendidas mesmo por estudantes do Ensino Básico. Para isso é importante que aquele que pretende estudar Estatística compreenda as principais ideias dos teoremas e os conceitos que constituem a formação dessas ferramentas estatísticas, donde se destacam os teoremas limite da Probabilidade que, embora sua plena compreensão exija um conhecimento matemático avançado, seu núcleo pode ser tornado inteligível através de experimentações simuladas ou mesmo intuição. É a partir desses teoremas que pretendemos fornecer nossa contribuição para a promoção do raciocínio estatístico. A proposta desta dissertação é dual: discutir o manancial teórico que visa auxiliar a compreensão e o desenvolvimento do raciocínio estatístico através dos Teoremas Limite da Probabilidade e propor atividades por meio de tecnologias diversas (simulações reais ou computacionais) para que seja alcançada a especificidade do raciocínio estatístico.

Palavras-chave: Probabilidade, Estatística, teoremas limite da probabilidade, raciocínio estatístico, ensino de Estatística.

ABSTRACT

Statistics is often treated in High School in an operational way, often restricted to a set of formulas that must be applied to obtain a numerical value. From a historical perspective, we see a growing need to develop tools to study and understand the various types of populations, aiming at decision making. Statistics is a key field to understand random phenomena that are becoming increasingly ubiquitous to the members of any society. The urgency to understand phenomena governed by chance establishes that the Probability Theory and Statistics should not be studied separately. On the contrary, there is a strong need to use the Probability within the field of Statistics, since the interpretations of probability somehow govern the ulterior statistical reasoning. Certainly, this union promotes a lot ofquite sophisticated tools, requiring the understanding of concepts from the Differential and Integral Calculus and Analysis. However, even students of Basic Education can understand the main ideas of many of these statistical tools. In order to accomplish that, it is important that anyone eager to study Statistics be embodied with the main ideas behind the theorems and concepts that constitute the ground of these statistical tools, among which we highlight the limit theorems of probability, whose complete understanding, despite its advanced nature, can be made intelligible by means of simulated experiments and intuition. It is from these theorems that we intend to give our contribution to the enhancement of statistical reasoning. The purpose of this dissertation is twofold: to discuss the theoretical source that aims to help the understanding and development of statistical reasoning through Probability Limit Theorems and to propose activities through different technologies (real or computational simulations) to achieve the specificity of the statistical reasoning.

Keywords: Probability, Statistics, limit theorems of probability, statistical reasoning, statistical education.

Sumário

“Porque a sabedoria serve de defesa, como de defesa serve o dinheiro; mas a excelência da sabedoria é que ela preserva a vida de quem a possui”. (Eclesiastes 7: 12)

INTRODUÇÃO

Por que os estudantes do Ensino Básico apresentam tanta resistência à Matemática? Acreditamos que a falta de interesse em estudar a Matemática, dentre outras coisas, seja um reflexo da não compreensão da urgência histórica da criação de seus conceitos e da sua utilidade prática no cotidiano como ferramenta para dar inteligibilidade a estruturas fenomenológicas carentes de tradução para diversas relações, entre as quais aquelas modeladas por funções. No entanto, há que se admitir que a maior parte dos fenômenos da natureza não é *ipso facto* regida intrinsecamente por estruturas funcionais. São na verdade relações que, no mais das vezes, transgridem as condições que definem esse conceito tão caro à ciência como o de função. Como disse uma vez Albert Einstein: "*Na medida em que as leis da matemática se referem à realidade, elas não são certas; e na medida em que elas são certas, elas não se referem à realidade*".

A Estatística, por outro lado, é o campo científico por excelência na busca de modelos que visem a capturar estruturas funcionais para a parte previsível de fenômenos na sociedade, presentes em áreas tão diversas como Medicina, Economia, Engenharia, Astronomia, Finanças, Informática, etc. ao mesmo tempo em que deixa o desafio de modelizar a parcela não determinística (aleatória) do próprio fenômeno, a fim de dar pleno entendimento às suas previsões e tomadas de decisão ulteriores. Como preconizado por H. G. Wells no início do século XX: "*Pensar à maneira da Estatística será um dia tão necessário para o cidadão eficiente como a habilidade de ler e escrever*".

No entanto, embora estejamos vivendo um século XXI imerso em incertezas, pouco tem sido feito na Educação Matemática para a promoção de uma verdadeira Educação Estatística em sala de aula. Por "verdadeira", entende-se compreender a natureza singular do raciocínio estatístico em contraponto ao raciocínio matemático, dois processos que, embora se cruzem, têm suas especificidades.

Um dos desafios e obstáculos ao bom entendimento da natureza do raciocínio estatístico reside justamente na compreensão dos significados da principal entidade matemática na construção da matematização do acaso: a Probabilidade. Esse é um objeto desafiador na Educação e o fato de ter sido "marginal" na História da

Matemática já denuncia a delicadeza de sua natureza. Como bem expressou Piaget e Inhelder (1951) em seu famoso livro "*A Origem da Ideia do Acaso na Criança*": em contraste com as operações lógicas e aritméticas, a probabilidade é descoberta gradualmente.

O ensino hoje protocolar de probabilidade, quando oferecido, é quase sempre atrelado à análise combinatória e engessado ao conceito clássico que pouco ressoa nos problemas do cotidiano e mais causam desserviço à Literacia Estatística quando dado com a aparência de verdade final. Isso se dá, porque o próprio professor tem uma visão puramente matemática desse objeto singular e não consegue promover a construção de um raciocínio estatístico para uma tomada de decisão sob incerteza, algo que só pode ser alcançado se diversas formas de entendimento da probabilidade são discutidas e trabalhadas em sala de aula.

Os Parâmetros curriculares nacionais (PCN), da forma como são conduzidos, dão margem para que o ensino das Probabilidade e Estatística sejam deficitários. Os PCN possuem uma característica que é se referir a um dado assunto através de eixos temáticos, esse caráter abrangente é positivo na medida em que possibilita ao professor aprofundar o conceito tanto quanto julgar necessário para atender aos objetivos propostos, entretanto se torna negativo devido à dependência dos professores aos livros didáticos, e muitos destes livros por sua vez restringem a teoria
das probabilidades ao caso clássico e a Estatística a um conjunto de fórmulas para analisar medidas pontuais de uma dada amostra. Isto faz com que o ensino da Estatística seja bastante limitado no que diz respeito à tradução de fenômenos cotidianos e consequentemente inviabiliza a tomada de decisão frente a uma vasta quantidade de situações.

De fato, uma análise dos Parâmetros curriculares nacionais, tanto para o Ensino Médio quanto para o Ensino Fundamental, atesta que a interpretação de probabilidade via frequência relativa pode ser negligenciada, prevalecendo o modelo clássico. Infelizmente, isso mostra o papel marginal da Probabilidade no ensino, o que constitui uma incoerência. Tanto a Estatística quanto a Probabilidade são tratadas pelo PCN no bloco referente ao tratamento da informação, que tem sua importância ressaltada no próprio documento quando reconhece que esta é uma demanda social. Entretanto, ao mesmo tempo em que faz isso, afirma que este bloco de conteúdo (Tratamento da informação) pode ser incorporado aos demais blocos, a saber: grandezas e medidas, espaço e forma, números e operações.

Olhamos com bastante reserva essa incorporação pura e simples, uma vez que o campo Tratamento da Informação possui núcleos de conteúdos próprios como a Estatística e a Probabilidade que possuem características bastantes diferentes dos demais blocos, a saber, o aleatório e a incerteza nas conclusões. Obviamente isto não quer dizer que não deva haver interações entre esses blocos, e, sim, que o Tratamento da Informação deve constituir um núcleo específico para o aprendizado, tanto pela urgência em se compreender as diversas informações apresentadas cotidianamente, quanto pelas qualidades de seus conteúdos que possuem uma natureza completamente distinta dos demais blocos.

Desta forma, a incoerência a que nos referimos reside em reconhecer a importância do tema para o tratamento de fenômenos associados ao cotidiano, mas não exigir explicitamente uma ampliação do conhecimento sobre a probabilidade, que é ferramenta indispensável para a compreensão dos fenômenos regidos pelo acaso.

O Ensino dessas searas é tratado pelos PCN em um regime em espiral, sendo que uma grande parte dos conceitos relativos à análise exploratória de dados são discutidos no Ensino Fundamental. A Probabilidade como ferramenta da Estatística passa a ser assunto do Ensino Médio, porém sem que haja grande aprofundamento dos conceitos aprendidos no Ensino Fundamental.

Um fato merecedor de atenção é que, embora os PCN sejam um balizador para o Ensino Básico, na realidade, o que notamos quando dialogamos com alguns professores do Ensino Básico é que uma parcela significativa deles se atém apenas aos conceitos adotados no currículo mínimo, que em geral inclui Probabilidade e Estatística em apenas um dos três anos do Ensino Médio. Isso somado a um tratamento inadequado[1] dos temas pelos livros didáticos[2] produz como resultado um ensino bastante defasado desses campos.

---

[1] Segundo nossa compreensão, os conceitos, da forma como são abordados, pouco contribuem para a compreensão e análise de fenômenos estocásticos do nosso cotidiano.
[2] Foram analisados sete livros didáticos com data de publicação a partir do ano de 2010.

Uma das formas fundamentais de se entender como maximizar o ganho de nossas tomadas de decisão sob incerteza provém do conceito frequentista de probabilidade. Esse conceito guarda riquezas pedagógicas, cognitivas, teóricas e filosóficas, a partir do arcabouço teórico dos teoremas limite da Teoria das Probabilidades. A despeito de sua grande sofisticação matemática, seu núcleo é inteligível, intuitiva e empiricamente, e é fonte rica para um tratamento das probabilidades imbuídas de sua real problematização e urgências históricas.

A proposta desta dissertação é dual: discutir o manancial teórico que visa a auxiliar a compreensão e o desenvolvimento do raciocínio estatístico através dos Teoremas Limite da Probabilidade e propor atividades por meio de tecnologias diversas (simulações reais ou computacionais) para que seja alcançada a especificidade do raciocínio estatístico.

As contribuições serão realizadas a partir da abordagem frequentista de probabilidade. Essa abordagem se contrapõe ao princípio da razão insuficiente proposto por Laplace, em que se atribuía chances iguais à ocorrência de eventos, caso não se tenha nenhuma razão para suspeitar que um seja mais provável que o outro de ocorrer.

A visão frequentista baseia-se na frequência relativa com que um determinado evento ocorre, quando o experimento é realizado sob as mesmas condições, sendo a probabilidade de um dado evento descrita como o limite dessa frequência relativa quando o número de experimentações tende ao infinito.

O tratamento via frequência relativa depende de o evento poder ser reproduzido empiricamente, ou via simulação computacional, o que obstaculiza o tratamento de certos tipos de fenômenos, tais como as previsões do tempo, as previsões de regiões conterem petróleo, etc. Contudo, quando os eventos podem ser reproduzidos de forma mecânica ou via simulação, isso se mostra como uma ferramenta importante para o aprendizado, pois apoia a necessidade estatística da realização empírica dos experimentos, no tocante à Inferência Frequentista.

É dentro do contexto frequentista que enalteceremos as importantes contribuições advindas dos Teoremas Limite da Probabilidade, dentre os quais destacaremos a Lei dos Grandes Números e o Teorema Central do Limite para esta dissertação.

A Lei dos Grandes Números nos informa que se $X_1, X_2, \dots$ formam uma sequência de variáveis aleatórias independentes e identicamente distribuídas com média finita μ, então $\frac{S_n}{n}$ converge quase certamente para μ quando $n \to \infty$, onde $S_n = X_1 + X_2 + \cdots + X_n$.

Esse resultado é importante para o princípio da abordagem inferencial frequentista, pois é ele quem garante que a frequência relativa da ocorrência de um evento converge para a medida de probabilidade teórica, quando as variáveis X_i são indicadoras da ocorrência ou não do fenômeno aleatório (1 ou 0), assegurando que as probabilidades obtidas no contexto frequentista da Probabilidade sejam as mesmas daquelas obtidas na abordagem clássica, e, além disso, garantindo a eficácia de um método para o cálculo da probabilidade de fenômenos que não são cobertos pela interpretação Clássica de probabilidade.

Outro resultado que fornece contribuições valiosas para o raciocínio estatístico é o Teorema Central do Limite, cuja notoriedade se inicia na *Théorie Analytique des Probabilités* de Laplace e posteriormente se torna um dos principais teoremas da Matemática com as contribuições de Lyapunov e Lindeberg.

Cabe ressaltar que Laplace não propôs um teorema geral. Ele apenas resolveu um conjunto de problemas específicos que se referiam à aproximação da probabilidade das somas ou combinações lineares de um grande número de variáveis aleatórias. Muitos dos problemas estudados por Laplace necessitavam de uma grande quantidade ensaios de Bernoulli, fazendo com que as probabilidades fossem expressas por fórmulas muito complicadas que dificultavam ou mesmo impediam que pudesse ser feita uma avaliação numérica direta, daí a necessidade de buscar métodos algébricos para encontrar boas aproximações para as fórmulas. Um dos fenômenos de interesse que Laplace estudou versava sobre a Mecânica Celeste.

Com o auxílio do Teorema Central do Limite ele poderia calcular a probabilidade de que a média dos ângulos que os cometas formavam com um determinado referencial estivesse dentro de um dado intervalo.

Apesar de utilizarmos o termo Teorema Central do Limite, o próprio Laplace não utilizava essa nomenclatura, tampouco propôs o teorema da forma como o conhecemos hoje. Fisher (2011) com base no livro *Théorie Analytique des*

Probabilités afirma que, em notação moderna, o enunciado mais geral do Teorema Central do Limite de Laplace era:

Seja $\epsilon_1, \dots, \epsilon_n$ um grande número de observações de erros independentes, em que cada uma possui a mesma densidade com média μ e variância σ^2. Se $\lambda_1, \dots, \lambda_n$ são constantes multiplicativas e $a > 0$, então

$$P\left(\left|\sum_{j=1}^{n} \lambda_j(\epsilon_j - \mu)\right| \leq a\sqrt{\sum_{j=1}^{n} \lambda_j^2}\right) \approx \frac{2}{\sigma\sqrt{2\pi}} \int_0^a e^{-\frac{x^2}{2\sigma^2}}\,dx$$

Este enunciado denuncia o interesse de Laplace no estudo das medidas dos erros que posteriormente virá a contribuir com o cálculo das probabilidades aplicado às ciências morais, tais como a probabilidade dos julgamentos dos tribunais, estudos sobre mortalidade, médias de vida, médias de casamentos, etc.

Atualmente uma das formas possíveis de se enunciar este teorema é:

Seja $X_1, X_2, \dots$ uma sequência de variáveis aleatórias independentes e identicamente distribuídas, com média μ e variância σ^2 finita e positiva, e seja $S_n = X_1 + X_2 + \cdots + X_n$, então:

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{D} N(0,1)$$

A notação $\xrightarrow{D}$ indica a convergência em distribuição de que trataremos ao longo dessa dissertação.

Em outras palavras, isto significa que para todo $a \in \mathbb{R}$

$$\lim_{n\to\infty} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq a\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{x^2}{2}}\,dx$$

Este Teorema desempenha um papel de grande importância para o raciocínio estatístico. Essencialmente ele nos informa que se tomarmos a soma ou a média de uma sequência de variáveis aleatórias, que atendam às condições do teorema, estas convergirão, em distribuição, para uma variável aleatória Normal. Uma das principais utilizações desta ferramenta para a inferência clássica é a facilidade com que ela nos permite estudar as propriedades de aleatoriedade da média de uma amostra retirada de uma dada população de média desconhecida, independentemente de qual seja a distribuição de probabilidade da população.

As principais medidas que nos permitem tomar decisões a respeito de uma dada população são a Esperança Matemática e o Desvio Padrão e estes parâmetros por sua vez são, em geral, desconhecidos por dependerem de parâmetros populacionais. Realizar qualquer tipo de inferência sobre esses parâmetros sem o apoio dos Teoremas Limite, tanto na Teoria da Estimação quanto nos Testes de Hipóteses seria tarefa muito árdua e em alguns casos até mesmo inviável.

Além disso, o Teorema Central do Limite é o elo que une diversos tipos de distribuições probabilísticas, revelando sob quais condições tais distribuições se aproximam da distribuição Normal. Além disso, tal distribuição é utilizada para modelar uma vasta quantidade de fenômenos.

A sua importância para a Probabilidade e para a Estatística é tamanha, que, para alguns Matemáticos, ele é considerado um dos cem teoremas mais belos de toda Matemática.

Essa perspectiva sobre o Teorema Central do Limite conecta-se à afirmação de Shulman (1986) que diz que, além de definir quais são as verdades aceitas em determinado domínio do conhecimento, o professor deve ser capaz de explicar o porquê de uma proposição ser verdadeira, o porquê de essa informação ser importante e como ela se relaciona com outras proposições dentro e fora da disciplina, tanto na teoria quanto na prática.

No tocante ao segundo ponto de nossa proposta, sua importância é ressaltada na seguinte pergunta: Como, a partir de um ensino tão negligente com o raciocínio estatístico, poderíamos esperar que as pessoas fossem capazes de reconhecer e interpretar corretamente os dados que lhes são apresentados a fim de que possam tomar as melhores decisões?

Podemos afirmar que a ideia construída sobre o tema, apenas recorrendo à utilização de fórmulas, não é suficiente para construção de um saber de conteúdo, no sentido de Shulman. Portanto, com o objetivo de auxiliar a compreensão do tema, iremos, além de discutir os aspectos teóricos de forma já mencionada, propor algumas atividades que corroboram com o desenvolvimento da forma de pensar estatisticamente.

Em nossa proposta pretendemos discutir o manancial teórico, referente ao raciocínio estatístico, tendo como pilar principal os Teoremas Limite da Probabilidade, ao mesmo tempo em que nos ocuparemos das discussões referentes ao conteúdo, analisando cuidadosamente seus significados, suas interpretações e seus resultados (às vezes paradoxais, mas preciosos para o entendimento do acaso), a fim de desconstruir as concepções errôneas tão difundidas e presentes tanto na sociedade quando na própria sala de aula.

Por exemplo, quando estudamos a Probabilidade no Ensino Médio, aprendemos a fornecer uma medida de probabilidade a priori, mas como esta medida pode ser comprovada empiricamente? E mais precisamente, por que, para um grande número de testagens, os valores obtidos empiricamente se aproximam da medida calculada a priori? O que significa tomar melhores decisões, de uma perspectiva da Estatística, mesmo em contextos de não repetibilidade dos fenômenos estocásticos?

Esperamos assim com este trabalho refletir sobre um melhor tratamento tanto da Probabilidade quanto da Estatística em sala de aula, a fim de contribuir para a promoção do Letramento Estatístico, condição fundamental para uma boa tradução do mundo atual para as futuras gerações.

Estrutura da dissertação

Esse trabalho dissertativo se estrutura da seguinte forma:

No primeiro capítulo será descrita a metodologia de pesquisa adotada para essa dissertação.

No segundo capítulo discutiremos alguns pontos da história da Probabilidade e da Estatística, a fim de revelar de que forma os conceitos básicos para a promoção do raciocínio estatístico dependem de um desenvolvimento da Matemática como, por exemplo, o desenvolvimento das técnicas algébricas, da Análise e da Teoria da Medida.

O terceiro capítulo é destinado a uma reflexão sobre a natureza do raciocínio estatístico, e como alguns autores o concebem. A partir dos fatores de concordância e discordância dessas definições, pudemos identificar alguns elementos do raciocínio estatístico que são fundamentais, tanto para que se possa analisar os dados, quanto para que se possa tomar qualquer tipo de decisão sobre eles.

No quarto capítulo discutimos sobre como a Probabilidade contribui para o raciocínio estatístico. Inicialmente fornecemos uma visão geral sobre os diferentes tipos de entendimento da Probabilidade, bem como da unificação desta a partir da Teoria da Medida. Num segundo momento, centralizamos a discussão no papel da Probabilidade dentro da abordagem frequentista.

O quinto capítulo é dedicado a discutir a Lei Forte dos Grandes Números e o Teorema Central do Limite, bem como suas implicações para a Estatística sob a perspectiva frequentista.

No sexto capítulo, discutiremos qual é o papel desempenhado pelos Teoremas Limite abordados no capítulo anterior na construção de um raciocínio estatístico, valorizando assim o diálogo entre os saberes da graduação e o da sala de aula da Escola Básica.

Em seguida deixamos nossas conclusões, bem como uma lista bibliográfica com todos os livros e artigos que forneceram contribuições para este trabalho.

A fim de complementar alguns pontos desta dissertação, deixamos alguns apêndices que versam sobre: o critério de convergência quase certa, a demonstração da Lei Forte dos Grandes Números, o fator de correção ao aproximarmos uma distribuição discreta a uma distribuição contínua, a tabela Z, o planejamento das aulas que seriam aplicadas aos futuros professores e os procedimentos para a contrução dos diversos applets no Geogebra.

Há ainda um total de dez arquivos do Geogebra que foram construídos visando à realização das simulações discutidas no último capítulo.

METODOLOGIA

A metodologia adotada para este trabalho foi predominantemente de análise documental, visando à identificação e discussão de conceitos essenciais para a promoção do raciocínio estatístico. Além disso, embasados pelas contribuições de Shulman, foram planejadas três aulas que poderiam ser ministradas aos estudantes de Licenciatura em Matemática (futuros professores), por meio das quais alguns dados poderiam ser coletados e analisados. Contudo, por dificuldades de ordem operacional, não foi possível executar tais aulas.

Primeiramente, reunimos livros e artigos, indicados na bibliografia, que serviram de base para a elaboração deste texto, tanto no tocante à esfera conteudista quanto pedagógica.

Para tal, foi feito um levantamento de textos e artigos que apontam as principais dificuldades dos estudantes, tanto do Ensino Básico quanto do Ensino Superior, com relação à compreensão da Teoria das Probabilidades. A partir disso analisamos quais destas dificuldades estão de alguma forma relacionadas com os Teoremas Limite da Probabilidade para então podermos discutir os aspectos teóricos que visam a auxiliar a promoção do raciocínio estatístico através dos Teoremas Limite da Probabilidade.

Com base nessa discussão é que as aulas foram planejadas. No planejamento dessas aulas, enfatizamos a importância dos Teoremas Limite para auxiliar na compreensão dos principais conceitos da Probabilidade e da Estatística, a saber, probabilidade frequentista, média, medidas de dispersão, curva Gaussiana e vários outros que são indispensáveis para a Inferência Frequentista.

1 FUNDAMENTAÇÃO TEÓRICA

No dia a dia, somos confrontados com situações das mais diversas e, portanto, convidados a tomar decisões. Quando a mídia nos apresenta dados que dizem respeito a algum acontecimento pertinente à nossa vida, como procedemos para analisar esses dados? Segundo (SEDLMEIER, 1999), aquilo de que necessitamos para realizar a análise desses dados para fazermos alguma inferência razoável é o raciocínio estatístico. Mas o que vem a ser Estatística?

A Estatística é vista como a ciência que estuda fenômenos sujeitos à variabilidade, ou seja, fenômenos que, mesmo quando reproduzidos em condições iniciais similares, possam produzir resultados significativamente diferentes. A Estatística se dá, portanto, a partir da coleta, organização, síntese e interpretação de dados de fenômenos aleatórios, com o objetivo de tomar decisões ótimas (segundo um critério de otimalidade), tendo como grande aliada a Teoria das Probabilidades.

Tendo em vista a compreensão do que é Estatística, analisamos a natureza do raciocínio estatístico, sob a visão de diversos autores como COLADARCI & COBB (2014), GARFIELD & GAL (1997, 1999), CHERVANEY et al. (1977), a fim de evidenciar as principais características que o definem. Tomando por base esse conjunto de características, podemos identificar e discutir um conjunto de ferramentas e conceitos que são fundamentais para o raciocínio estatístico.

A importância do aprendizado da Estatística para os estudantes da Escola Básica é, atualmente, incontestável visto a sua aplicabilidade à vida cotidiana. Contudo, como afirmamos na introdução desse trabalho, este campo do conhecimento ainda é pouco explorado nas salas de aula.

Nossa proposta de contribuir para a construção de um raciocínio estatístico é um desafio que encontra sua primeira barreira na própria formação do professor de Matemática, a quem corresponde o papel de ensinar a Estatística.

Afirmamos que essa é uma barreira não apenas devido ao pouco tempo que um curso de graduação dedica a tratar a referida disciplina, mas também a forma como ela, geralmente, é tratada na Licenciatura. A preocupação em transmitir todo o conteúdo Matemático e Estatístico ocupa grande parte do tempo, senão todo o tempo, do curso. E ao final da graduação, espera-se que o professor possa abordar o tema com seus estudantes no Ensino Básico. Mas, como isto será feito?

Evidentemente que o professor não pedirá a seus alunos que realizem procedimentos tão complexos quanto os que foram vistos em sua graduação, tampouco tratará com eles grande parte do conteúdo visto. Como, a partir dos conceitos tratados na graduação, o professor deve pensar a matéria a ser ensinada? Quais os conhecimentos um professor precisa dominar?

Para refletir sobre estas questões adotaremos como referência as ideias propostas por Lee S. Shulman em suas importantes contribuições, (SHULMAN, 1986 e 1987), para os processos educacionais.

Shulman discute, nesses artigos, as concepções referentes ao conhecimento do professor. Para tal, ele realiza uma análise de documentos que dizem respeito à formação dos professores nas décadas de 1970 e 1980, o que incluiu a análise de testes de admissão, diários de professoras, políticas educacionais adotadas e o acompanhamento de professores iniciantes por um período de até dois anos.

O autor entende, nesses artigos, que os alicerces do saber necessário para a prática do ensino são distintos daqueles utilizados, por exemplo, por um pesquisador matemático, por um engenheiro químico, etc. Em SHULMAN (1986), é realizada uma discussão em torno do que o autor chama de saber de conteúdo, o qual se refere exclusivamente ao conhecimento sobre a matéria a ser ensinada.

Contudo, Shulman vai além e discute sobre uma das mais importantes contribuições para as pesquisas em educação, o saber pedagógico de conteúdo. Na visão do autor, o saber pedagógico de conteúdo não se restringe apenas ao conhecimento sobre a matéria. Trata-se de um tipo de saber diferenciado que promove uma junção dos conhecimentos sobre a matéria a ser ensinada e da pedagogia, sendo, portanto, um conhecimento sobre a matéria voltado para o ensino.

> “Eu ainda falo de conhecimento de conteúdo aqui, mas como uma forma particular de conhecimento de conteúdo que incorpora os aspectos de conteúdo mais relevante para a habilidade de ensinar”. (SHULMAN, 1986, p.09, tradução nossa).

Para além desses dois tipos de saberes evidenciados, Shulman chama a atenção para a existência de outros tipos de saberes que subjazem a compreensão que um professor possui a fim de promover a aprendizagem. Em SHULMAN (1987) o autor apresenta uma lista dos tipos de saberes que acredita que o professor deve ter.

> - "Saber de conteúdo.
> - Saber pedagógico, com especial referência a princípios gerais e estratégias de gestão e organização das salas de aula que parecem transcender o conteúdo específico.
> - Saber sobre o currículo, com especial compreensão dos materiais e programas que servem como "ferramentas de trabalho" para os professores.
> - Saber pedagógico de conteúdo, amálgama especial entre conteúdo e pedagogia que é exclusivo da docência, sua forma especial de entendimento profissional.
> - Saber sobre os estudantes e sobre suas características.
> - Saber sobre os contextos educacionais, abrangendo desde o funcionamento de grupos ou salas de aula, a administração e financiamento dos distritos escolares, as características de comunidades e culturas.
> - Saberes sobre os fins, propósitos e valores educacionais e suas bases filosóficas e históricas". (SHULMAN, 1987, p.08)

Essa, segundo o autor, compreende uma lista mínima de saberes, querendo dizer com isso que há ainda outras fontes de saber relevantes.

Contudo, como dissemos anteriormente, a contribuição de Shulman foca principalmente em dois desses saberes: o de conteúdo e o pedagógico de conteúdo.

A respeito do saber de conteúdo, temos por óbvio que saber o assunto a ser ensinado é fundamental para a prática docente. Entretanto, o mero conhecimento de teoremas, fórmulas e conceitos não são suficientes. "Pensar corretamente sobre o saber de conteúdo requer ir além do conhecimento dos fatos ou conceitos de um domínio." (SHULMAN, 1986, p.09, tradução nossa).

O saber de conteúdo abrange as várias formas em que os princípios e conceitos básicos de uma disciplina estão organizados, o conhecimento dos princípios que definem as verdades que podem ser aceitas dentro da disciplina a ser ensinada e a forma pela qual o assunto ensinado se relaciona com outros assuntos pertinentes ou não a disciplina.

Desta forma, o professor de Matemática que ensinará Estatística na Escola Básica deve compreender quais são os pilares que legitimam dizer, por exemplo, que uma boa medida de probabilidade pode ser obtida através da frequência relativa, para um grande número de amostras. Além disso, ele deve ser capaz de relacionar esse conceito a outros, tais como a esperança matemática, a variabilidade, aos conceitos da genética, aos resultados obtidos de fenômenos empíricos, etc.

Será que é legitimo dizer que se a probabilidade de uma empresa vender produtos com defeito é igual a $0{,}1$, então, ao se comprar cem produtos, exatamente dez apresentarão defeitos?

Para responder tal pergunta o professor deverá conhecer quais são os princípios que regem o raciocínio estatístico. Uma vez conhecido esse princípio, o professor será capaz de explicar ao seu aluno que a afirmação feita não é nem verdadeira nem falsa. Mas, possui um grau de confiabilidade a ela associado e que este grau de confiabilidade pode ser mensurado.

> "O professor não deve compreender, apenas, que algo é assim; o professor deve entender melhor o porquê de algo ser assim, baseado em que sua afirmação é válida e sob que circunstâncias nossa crença em sua justificação pode ser enfraquecida ou até mesmo negada." (SHULMAN, 1986, p. 09, tradução nossa).

Por fim, Shulman ressalta também que o professor deve ser capaz de reconhecer quais são e por que alguns pontos são mais relevantes do que outros dentro de determinado assunto.

Assim, neste trabalho, esperamos evidenciar a importância dos Teoremas Limite para o ensino dos conteúdos de Probabilidade e de Estatística na Escola Básica.

Como já pontuado, o saber pedagógico de conteúdo não se limita ao saber sobre o assunto. Ele é mais abrangente, na medida em que coaduna os aspectos pedagógicos ao conhecimento relativo ao conteúdo, constituindo, assim, um saber de conteúdo voltado para o Ensino.

> "Na categoria de saber pedagógico de conteúdo, eu incluo os temas mais regularmente ensinados em uma área específica, as formas mais úteis de representação dessas idéias, as mais poderosas analogias, ilustrações, exemplos, explicações e demonstrações - em uma palavra, as formas de representar e formular o assunto e torná-lo compreensível para os outros." (SHULMAN, 1986, p.09, tradução nossa).

Assim, não é suficiente que o professor saiba, por exemplo, a definição formal de Probabilidade tal como descrita por Kolmogorov. Ele precisa reconhecer os diferentes tipos de visões probabilísticas e quais são os contextos mais adequados para cada visão. Para isso, será necessário um conhecimento de conteúdo a respeito de cada visão a ser empregada, a fim de que possa reconhecer e utilizar corretamente as interpretações.

Contudo, é importante que o docente saiba, por exemplo, como a definição de Probabilidade vinculada à interpretação frequentista pode se relacionar com o conceito de média, e como estão relacionados as diferentes interpretações probabilísticas.

> “O saber pedagógico de conteúdo também inclui um entendimento do que torna a aprendizagem de tópicos específicos fácil ou difícil: as concepções e preconcepções que os alunos de diferentes idades e origens carregam para a aprendizagem dos tópicos e lições que são frequentemente ensinadas. Se essas preconcepções forem equivocadas, e muitas vezes são, os professores precisam conhecer as estratégias que têm maior probabilidade de sucesso na reorganização da compreensão dos alunos, porque os alunos, provavelmente, não se apresentarão a eles totalmente desprovidos de conhecimento.” (SHULMAN, 1986, p.09-10, tradução nossa).

Um dos pontos que torna o aprendizado das probabilidades difícil para os estudantes é, por exemplo, o conceito de fração, conceito esse de grande dificuldade aos alunos, como atestam diversos estudos. Além disso, há ainda concepções prévias relativas à própria Teoria das Probabilidades que os estudantes carregam desde o Ensino Básico.

Na obra de PIAGET & INHELDER (1951), os autores, ao estudarem como a ideia de Probabilidade é compreendida pelas crianças, apontam que há uma dificuldade por parte dessas de compreender a noção de mistura que, segundo os autores, é intrinseca à ideia de probabilidade.

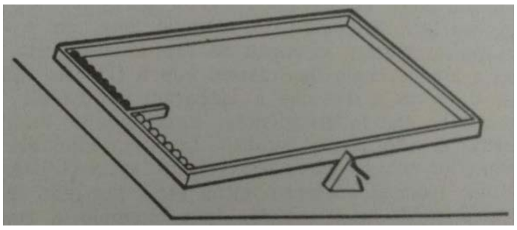


Figura 1: Caixa de bolinhas (PIAGET, INHELDER, 1951, p.16)

Como experiência, exibem a caixa, representada na figura 1, às crianças de três grupos de idades distintos, e perguntam como ficarão organizadas as bolinhas após oscilarem a caixinha (movimento em que as bolinhas vão até a direção opostas e retornam), ao que as crianças mais jovens respondem que retornarão à posição inicial, enquanto que as crianças que pertencem ao grupo de mais idade, já mostram alguma compreensão de mistura como um sistema de permutação devido às interferências que podem ocorrer na trajetória das bolinhas.

Ressaltamos aqui que a intenção de Piaget e Inhelder com essas experiências era, dentre outras coisas, mostrar que ao contrário das operações aritméticas, a probabilidade é descoberta gradualmente.

Outro ponto de dificuldade evidenciado por Piaget e Inhelder é a não compreensão da Lei dos Grandes Números relatada no capítulo 3 de sua obra aqui referida. Os autores iniciam um experimento no qual se coloca uma situação semelhante a acertar qual será o resultado obtido em uma roleta de cassino. Com esse experimento, pretendem verificar como as crianças reagem quando identificam que, a princípio, se trata de uma situação aleatória, porém, se veem contrariadas pela obtenção constante do mesmo resultado.

Em suma, nosso objetivo é fornecer um material capaz de promover uma profunda reflexão sobre os conceitos que são aceitos no domínio de conhecimento da Estatística, e, além disso, discutir o porquê desses conceitos serem importantes para a promoção do raciocínio estatístico. Isso envolve refletir sobre como eles se relacionam com os demais objetos e ferramentas da Estatística que são úteis tanto para compreender a própria Estatística quanto para a prática de se fazer Estatística.

Com isso esperamos auxiliar na construção de uma abordagem da Estatística, que não se limita a memorizações de fórmulas e mecanização de processos. Porém, não é nosso único objetivo auxiliar na construção de significado dos elementos que compõem o raciocínio estatístico. Pois, como apregoa Bachelard (2005, p.17): "O conhecimento do real é luz que sempre projeta algumas sombras. Nunca é imediato e pleno".

Além de, geralmente, o primeiro contato com um determinado tema ocorrer de maneira acrítica, o indivíduo carrega consigo experiências próprias que interagem com o novo conceito a ser aprendido, experiências tais que produzem bons resultados em casos específicos, mas que para além de situações particulares se mostram ineficientes. Por exemplo, um estudante que inicia seus estudos sobre a Teoria das Probabilidades, inicialmente é submetido a uma abordagem clássica, em que a probabilidade é descrita como uma razão de cardinalidades de conjuntos finitos. Entretanto, ao prosseguir em seus estudos sobre o tema ele pode ser confrontado com um experimento não equiprovável. Dessa forma seu conhecimento prévio pode ser um obstáculo para que a essência do novo conceito seja compreendida.

Segundo Bachelard (2005, p.17): "É no âmago do próprio ato de conhecer que aparecem, por uma espécie de imperativo funcional, lentidões e conflitos". As causas que levam à estagnação ou de regressão do processo de aprendizado, ou dito de outra forma, as causas que dificultam que o avanço da aprendizagem sobre um determinado conceito ocorra são chamados de obstáculos epistemológicos. É neste sentido que o termo obstáculo foi empregado anteriormente e é nesta linha de raciocínio que ocorrerá a nossa contribuição neste trabalho. A partir da identificação dos obstáculos para a aprendizagem, via análise documental, e discussão dos mesmos, esperamos fornecer material suficiente para que aqueles que pretendem ensinar possam elaborar estratégias e técnicas que auxiliem seus alunos a superarem os obstáculos que porventura venham a surgir, contribuindo assim para a promoção do raciocínio estatístico na Escola Básica. Obviamente não esgotaremos aqui a indetificação e discussão dos obstáculos que possam surgir.

2 EVOLUÇÃO HISTÓRICA DA PROBABILIDADE E ESTATÍSTICA

Neste capítulo objetivamos informar ao leitor alguns dos muitos obstáculos que tiveram que ser superados para que as teorias de Probabilidade e Estatística se desenvolvessem. O conhecimento das dificuldades encontradas pelos pensadores da época poderá nos auxiliar a compreender quais são os conhecimentos necessários para promover o raciocínio estatístico, bem como poderá auxiliar na identificação de conceitos que podem se tornar pontos de entraves para a compreensão dos temas pelos alunos, o que, segundo SHULMAN (1986), no tocante ao Saber Pedagógico de Conteúdo, é uma habilidade essencial para o professor.

É comum encontrarmos em livros e artigos pesquisadores afirmando que o início da Teoria das Probabilidades ocorre somente no século XVI, mesmo tendo plena consciência de que uma diversidade de jogos, em que o acaso se fazia presente, já existia há alguns milênios. Tal afirmação se embasa no fato de que as primeiras documentações sobre a Teoria das Probabilidades surgem a partirdo século XVI.

Mas, por que foi necessário tanto tempo para que começássemos a ver os registros sobre esta teoria? Nesta época já estávamos relativamente bem avançados quanto ao conhecimento da Geometria Euclidiana.Também já haviam sido dadas grandes contribuições à Álgebra. Porém, no que diz respeito à Probabilidade, pouco havia de substancial.

Segundo TABAK (2004), um dos principais fatores que dificultavam o estudo da aleatoriedade era, em primeiro lugar, a concepção da sociedade que, em sua maioria, compreendia que o resultado dos jogos não era aleatório e, sim, regido pela vontade de alguma divindade. Desta forma não havia motivos para que se estudassem os fenômenos cujos resultados não se podia prever. Outro fator apresentado diz respeito ao que seria o principal agente utilizado para promover a aleatoriedade, o astrágalo. Este é, na verdade, um osso de membros inferiores de mamíferos como carneiro, boi, etc, que os povos antigos usavam como uma espécie de dado para seus jogos. Porém, o astrágalo é um objeto bastante irregular e, sabemos hoje, que as chances de ocorrer cada uma de suas faces são diferentes, e mudam de um astrágalo para o outro, o que dificultou o desenvolvimento de uma teoria baseada em seu uso.

De acordo com TABAK (2004), a concepção da aleatoriedade como uma expressão da vontade divina foi uma barreira maior do que a provocada pela não uniformidade dos astrágalos, pois se tratava de uma barreira conceitual.

Apenas com o objetivo de situar o leitor sobre o momento histórico em que a Europa estava imersa no momento em que começam a surgir as primeiras teorizações da Probabilidade e Estatística, ressaltamos que ao final do século XIV inicia-se o movimento renascentista na Itália, posteriormente aceito por uma grande parte da Europa. A Renascença representa a ascensão dos ideais burgueses sobre o pensamento e a cultura medieval. Para isso foi necessário o rompimento com os paradigmas impostos pela Igreja Católica. No modelo de pensamento renascentista, o centro das pesquisas é o próprio ser humano e não mais a religião. Alguns historiadores apontam que essa ideologia, que questionava alguns dos paradigmas da Igreja Católica, quando somado à insatisfação de determinadas classes da população devido ao aumento de impostos, abriu espaço para que ocorresse a Reforma Protestante com Martinho Lutero no início do século XVI.

2.1. A PROBABILIDADE CLÁSSICA E O SURGIMENTO DA ESTATÍSTICA

É no século XVI, com o italiano Girolamo Cardano, que a compreensão do acaso começa a se distanciar dos aspectos da fé. Cardano é conhecido principalmente pela sua obra *Ars Magna* sobre Álgebra, mas suas contribuições foram além. Cardano fez diversas contribuições à ciência, dentre as quais consta um livro que versa sobre os jogos de azar, o qual chamou de *Liber de Ludo Aleae.* Nesse livro ele escreve sobre os jogos de dados, cartas, astrágalos e gamão. É nele que aparecem as primeiras noções básicas da probabilidade.

A compreensão do aleatório por Cardano não estava vinculada à vontade divina, porém não estava completamente livre de crenças. Para Cardano, um elemento fundamental nos jogos era a sorte, e esta era capaz de influenciar o resultado de um jogo, favorecendo o jogador que a possuía. TABAK (2004) cita um trecho do livro *Liber de Ludo Aleae* de Cardano que mostra como a crença na sorte influenciava os pensamentos de Cardano.

> "For this reason it is natural to wonder why those who throw the dice timidly are defeated. Does the mind itself have a presentiment of evil? But we must free men from error; for although this might be thought true, still we have a more manifest reason. For when anyone begins to succumb to adverse fortune, he is very often accustomed to throw the dice timidly; but if the adverse fortune persists, it will necessarily fall unfavorably. Then, since he threw it timidly, people think that it fell unfavorably for that very reason; but this is not so. It is because fortune is adverse that the die falls unfavorably, and because the die falls unfavorably he loses, and because he loses he throws the die timidly."(TABAK, 2004, p.18-19)

As crenças e preconceitos continuavam a ser um obstáculo para o avanço do raciocínio probabilístico. Entretanto, este não era o único fator. Devemos nos lembrar de que Cardano escreveu esta obra em 1526, muito antes que houvesse um bom sistema de notação algébrica, e isto possivelmente dificultou o desenvolvimento da teoria.

Tais dificuldades afetaram a compreensão do aleatório e levaram Cardano a cometer equívocos que, hoje, consideramos tolos, como, por exemplo, a afirmação de que a chance de se obter ao menos uma vez determinada face de um dado cúbico em três lançamentos era de 50%, o que também mostrava que Cardano tinha uma certa dificuldade com a multiplicação de probabilidades de eventos independentes, embora soubesse que a multiplicação poderia ser realizada.

Dentre os problemas que Cardano aborda em seu livro, consta um que intrigou matemáticos por muito tempo. Era o problema de determinar quantas vezes dois dados deveriam ser lançados a fim de que a chance de obter um duplo seis ao menos uma vez fosse a mesma de que não se obtivesse um duplo seis em nenhum desses lançamentos. Matemáticos como Luca Pacioli, Tartaglia e o próprio Cardano, não conseguiram responder corretamente a esta questão.

As versões do problema apresentadas por Pacioli e Tartaglia eram semelhantes. O problema consistia em como realizar uma divisão justa das apostas em um jogo entre dois jogadores. O vencedor era aquele que conseguisse chegar a seis pontos primeiro; porém a partida era interrompida quando o placar estava, por exemplo, 5 a 3.

Galileu Galilei, mais conhecido pelas suas contribuições à Física, também se interessou por problemas que envolviam o acaso. Ele estava intrigado em saber por que os números 10 e 11 apareciam mais frequentemente do que 9 e 12 como resultado da soma do lançamento de três dados cúbicos. Essencialmente, Galileu o tratou como um simples problema de contagem. Contou quantas combinações geravam a soma 10 e 11 e quantas delas geravam a soma 9 e 12 e concluiu que o resultado 10 poderia ser obtido de 27 formas diferentes enquanto o resultado 9 seria obtido de 25 formas. Por isso, a chance de se obter um 10 era maior do que a de se obter um 9, por exemplo. Embora, Galileu não faça uso de termos que possam ser associados à Probabilidade, tratando o problema apenas como uma questão de contagem e comparação, há uma sutil, mas importante, contribuição em sua forma de pensar. Primeiro, notamos que se começa a estudar eventos que concebemos hoje como não equiprováveis e em segundo lugar, no pensamento descrito por Galileu para resolver tal questão, os conflitos de crenças ou superstições não figuram, diferentemente do que ocorre com Cardano.

O problema da divisão das apostas atacado por Pacioli e Tartaglia permaneceu estagnado até meados do século XVII, quando o Chavalier de Méré, um homem fascinado por jogos, inicia uma discussão com o matemático Blaise Pascal, sobre qual seria a base matemática que estaria relacionada aos jogos. Eventualmente, Pascal recorreu ao matemático Pierre de Fermat, dando início assim, as famosas cartas sobre os jogos de azar de 1654.

Dentre os problemas abordados por Pascal e Fermat nessas cartas está o problema da divisão das apostas, bem como algumas variações deste. Pascal ainda publicou uma solução para a versão mais geral deste problema em sua obra intitulada *Traité du Triangle Arithmétique.* Para chegar a uma solução para a versão mais abrangente desta questão, Pascal precisou recorrer às propriedades do triângulo aritmético apresentadas nesse livro. Outro problema solucionado, de forma correta, que consta nas cartas, trata de responder quantas vezes é necessário lançar um dado até que uma determinada face ocorra, ao menos uma vez, a fim de que as chances desta face ocorrer ou não sejam as mesmas.

Ora, mas por que muitos consideram estas cartas como um marco importantíssimo para a Teoria das Probabilidades? Será por que eles foram capazes de fornecer soluções que realmente eram condizentes com a observação empírica do fenômeno? Na verdade, não. Ressaltamos que Cardano, Pacioli e Tartaglia em seus tempos também foram capazes de resolver algumas situações em que o acaso era presente e isto não lhes rendeu tanto crédito no que diz respeito à Teoria das Probabilidades quanto Pascal e Fermat obtiveram com algumas poucas cartas.

Pascal e Fermat não apenas forneceram respostas que correspondiam às observações empíricas dos fenômenos. Eles foram muito além, e mostraram soluções gerais para os problemas. Com a publicação da solução geral do problema em seu *Traité du Triangle Arithmétique,* Pascal foi capaz de fornecer uma fórmula geral o suficiente para dar conta de qualquer variação do problema da divisão das apostas. É importante notar que Pascal e Fermat já contavam com uma série de simbologias algébricas desenvolvidas por Descartes e outros matemáticos da época, o que facilitou a forma de representar suas ideias sobre a Probabilidade, embora nem ele nem Fermat, de fato, utilizem este termo. Outro ponto importante que outrora se apresentava nos discursos de Cardano e que não se faz presente nos discursos de Pascal e Fermat é a ideia da sorte que era capaz de exercer influência no jogo.

Uma vez que tinham um pensamento livre da influência de forças misteriosas atuando sobre o acaso, e de posse de um sistema de notações algébricas superior ao que Cardano, Tartaglia e outros tinham à sua disposição, Pascal e Fermat foram capazes de dar os primeiros passos em direção à compreensão da estabilidade das frequências em fenômenos regidos pela aleatoriedade.

Entretanto, ter o pensamento livre de crenças e superstições sobre os fenômenos regidos pelo acaso não impedia Pascal de manter a sua crença religiosa. Pascal chegou a empregar um argumento de decisão teórica, a fim de que pudesse decidir se acreditava ou não na existência de Deus.

Nesse contexto, uma aposta, na visão de Pascal, seria uma ação praticada pelo sujeito, isto é, ou a pessoa acreditaria na existência de Deus ou não acreditaria. Em suma, Pascal conclui que, se Deus não existe e o sujeito acredita na existência Dele, nada se perde, por outro lado se Deus existe e o indivíduo não acredita perde-se a salvação eterna. Como a salvação é um “prêmio” infinitamente mais desejável, Pascal afirma que mesmo se a chance de Deus existir for pequena, uma decisão razoável seria crer na existência de Deus. É neste contexto que começa a surgir uma visão probabilística sobre o grau de crença que se deve atribuir a um dado evento. Sendo considerado um marco para o desenvolvimento da abordagem subjetiva da Probabilidade.

Um ano após visitar Paris, em 1655, Christiaan Huygens escreve um ensaio chamado *De Ratiociniis in Ludo Aleae,* que é publicado em 1657. Nesta obra Huygens aborda os mesmos problemas que Pascal e Fermat trataram em suas cartas, porém com métodos diferentes de se alcançar a solução. Além disso, ele resolve problemas de sua própria autoria. Em contraponto com as cartas de Pascal e Fermat, Huygens além de abordar o motivo de algumas afirmações serem consideradas verdadeiras e como estas novas ideias poderiam ser utilizadas, ele introduz o conceito de esperança matemática associado aos jogos de azar. Huygens mostra em algumas proposições como o valor esperado de uma aposta pode ser obtido. Em outras palavras, se minha chance de vencer um jogo é p e eu aposto a, quanto eu posso esperar ganhar com esse jogo? A publicação dessa obra possui grande valor, pois foi responsável por disponibilizar este novo campo de pesquisa para um público mais amplo, segundo Tabak (2004). Além disso, trouxe a esperança matemática, um conceito que se tornará muito importante para a Estatística.

A nova área de pesquisa está imersa no estudo dos jogos de azar, e as soluções para os problemas estão essencialmente associadas a problemas de contagens simples. Então, em meados do século XVII, surge uma nova forma de se pensar esses jogos de azar, só que agora não se trata mais de apostas e sim de venda de anuidades e de seguros.

Os seguros marítimos se iniciam por volta do ano 1400, porém, segundo Howie (2004), ainda no início do século XVII, com a publicação sobre anuidades do inglês Richard Witt, a Teoria das Probabilidades não está presente. As taxas de empréstimos eram fixas, independentemente do serviço e condições oferecidas. A inclusão da probabilidade nesse campo se dá devido à publicação da primeira tabela de vidas na obra *Natural and Political Observations* de John Graunt em 1662. Esta tabela se baseava na causa das mortes ocorridas em Londres pela *Bills of Mortality* instituída em 1562 como aviso de pragas e epidemias. De posse desta tabela e com a suposição de que após os 6 anos de idade a chance de uma pessoa morrer permaneceria constante, a fração da população que sobreviveria após os 6 anos de idade foi estimada por Graunt. Além disso, ele também pôde estimar a população de Londres em um terço de milhão, um valor muito inferior ao que era assumido anteriormente. Pouco tempo depois desta publicação muitos estudiosos se interessaram pelo assunto, e logo uma grande quantidade de tabelas de mortalidade e cálculos de taxas de anuidade e seguro surgiram até o começo do século XVIII.

Assim, o que Tabak afirmava sobre os empecilhos para o desenvolvimento da Teoria das Probabilidades ocorria também entre as seguradoras. A ideia de que uma morte ocorria ou um navio naufragava por vontade divina, somada ao fato de que as seguradoras eram bastantes rentáveis, dificultava a aceitação e o investimento no estudo desse campo para que houvesse o desenvolvimento das teorias de Probabilidade e Estatística. Além disso, as seguradoras afirmavam que os julgamentos e experiências de seus profissionais não poderiam ser substituídos por equações matemáticas.

Mesmo que muitos matemáticos e pesquisadores da época tenham ultrapassado a barreira da crença e da superstição, parece claro que para muitas pessoas essa dificuldade conceitual permanecia. Outro fator que também chama nossa atenção é que os processos de estimação populacional de Graunt despertaram um grande interesse de líderes políticos; porém parece não ter surtido o mesmo efeito com as seguradoras. Então, além das questões de crença, outro fator que dificultava o desenvolvimento destas ciências era a sua aparente inutilidade para situações reais. Visto que a seguradora tinha boa faixa de lucro, não havia por que se preocupar com um assunto que estava tão intimamente conectado aos jogos de azar.

2.2. AS BASES DAS ABORDAGENS FREQUENTISTA E SUBJETIVA DA PROBABILIDADE

Fermat, além de colaborar para o desenvolvimento da Teoria das Probabilidades, também desenvolveu uma grande quantidade de ferramentas algébricas que auxiliaram no desenvolvimento do Cálculo Diferencial e Integral.Entretanto, foi através das contribuições de Isaac Newton e Gottfried Leibniz que o Cálculo se popularizou no século XVII.

O Cálculo amparou diversas áreas do conhecimento, dentre as quais a Teoria das Probabilidades. De fato, seria bastante complicado seguir para além das contribuições de Huygens sem o desenvolvimento do Cálculo.

O matemático suíço Jacob Bernoulli foi um dos primeiros a utilizar o Cálculo Diferencial e Integral em aplicações diferentes daquelas propostas por Leibniz e Newton. Bernoulli era fascinado pela Teoria das Probabilidades, tendo um grande apreço pela obra *De Ratiociniis in Ludo Aleae*de Huygens. O interesse que Bernoulli tinha pelo acaso era tão grande que dedicou boa parte de sua vida a escrever um dos livros sobre o tema e atribuiu a ele o nome de *Ars Conjectandi,* que em português significa “Arte da Conjectura”. Tal obra é considerada um divisor de águas na Teoria das Probabilidades, pois dentre outras coisas traz uma das contribuições mais valiosas para a abordagem frequentista, que foi chamada por Poisson, anos depois, de “A Lei Fraca dos Grandes Números”. Essa lei diz que em um conjunto de eventos aleatórios independentes (em Teoria das Probabilidades, dois eventos são ditos independentes quando a ocorrência de um não afeta a probabilidade de ocorrência do outro), se o número de experimentos for grande o suficiente, então a razão entre o número de experimentos ditos “sucessos” e o número total de experimentos ficará quase certamente dentro do intervalo $[p-\epsilon\ ,\ p+\epsilon]$, em que p é a probabilidade obtida a priori e ϵ é descrito como uma medida de aproximação.

Foi graças ao desenvolvimento do Cálculo que Bernoulli pôde fornecer uma prova para essa afirmação, que mais de 20 anos antes ele já suspeitava ser verdadeira.

Além da Lei dos Grandes Números, Bernoulli traz outras contribuições nessa obra. Ele utiliza a ideia dos jogos de azar para modelar situações cotidianas, como, por exemplo, qual é a chance de que um certo grupo de pessoas de uma dada população seja acometido por uma mesma doença? Recorrendo-se a uma urna com muitas bolas em que a quantidade de bolas brancas e de bolas pretas dentro da urna, respectivamente, reflete a proporção de pessoas que possuem a doença e que não possuem a doença apresentada nos registros médicos, pode-se, facilmente, recorrendo à Análise Combinatória, determinar qual a chance de que uma determinada quantidade de pessoas dessa população seja infectada pela doença.

Além disso, Bernoulli foi um dos primeiros a notar as duas vertentes da Probabilidade. A primeira vertente é, dada a probabilidade de um fenômeno ocorrer, descobrir com qual frequência ele ocorrerá. E na segunda (aqui ele não teve muito sucesso em desenvolvê-la) trata-se de, uma vez conhecida a frequência de ocorrência de um evento para uma certa quantidade de experimentos, utilizar esses dados para tentar estimar qual é a probabilidade de ocorrência do evento.

O uso da Probabilidade para tomar decisões racionais sobre situações quotidianas despertou o interesse de muitos matemáticos que tentaram aplicar a teoria para outras situações e, além disso, também tentaram generalizar os resultados obtidos por Bernoulli.

Um exemplo disso é a proposta feita pelo matemático Marquês de Condorcet no final do século XVIII. Ele considerava o modelo de julgamento inglês, em que a decisão do júri deveria ser unânime, inadequada. Condorcet, recorrendo à Teoria das Probabilidades, calculou o número ótimo de júri como sendo 30 e propôs que o resultado do julgamento fosse dado pela maioria dos votos e não mais por unanimidade.

O matemático Jacob Bernoulli não era o único que tinha um grande apreço pela obra de Huygens. O francês Abraham De Moivre também tinha grande fascínio pelo livro *De Ratiociniis in Ludo Aleae*, o que o motivou a escrever um livro chamado *The Doctrine of Chances or A Method of Calculating the Probabilities of Events in Play*. Nessa obra, De Moivre, recorrendo às ferramentas do Cálculo e da Álgebra, resolve uma grande quantidade de problemas, muitos dos quais foram propostos por Huygens e Bernoulli em suas respectivas obras.

De Moivre organizou essa obra de modo que a teoria fosse se desenvolvendo gradativamente, começando com exercícios simples. Ele introduz uma definição de probabilidade, e, conforme os exercícios se tornam mais complicados, outras ferramentas vão sendo incorporadas, tais como os teoremas de multiplicação e adição de probabilidades, o teorema de inclusão e exclusão, séries infinitas e algumas séries de sua própria autoria. Cabe ressaltar ainda que, com respeito ao teorema de multiplicação de probabilidade, De Moivre menciona a independência de eventos.

De Moivre dedica uma parte de sua obra a encontrar uma forma de aproximar a distribuição Binomial e é neste contexto que encontramos pela primeira vez a expressão algébrica da curva-sino. Porém, ressaltamos que De Moivre propôs apenas uma aproximação, as aplicações estavam ainda muito restritas. A curva-sino alcançará um significado mais amplo com Pierre-Simon Laplace, do qual falaremos posteriormente.

Outra contribuição extremamente importante foi feita através da republicação da obra *The Doctrine of Chances*em 1738. Nesta versão foi incluído um apêndice intitulado *A Treatise of Annuities on Lives,* em que De Moivre faz um estudo sobre seguros e anuidades e esse apêndice só foi incluído a partir da segunda edição porque é dependente de um artigo de Edmund Halley, que analisa taxas de nascimentos e mortes em Breslau, uma cidade da Europa Central.

Esse artigo de Halley é bem diferente daquele publicado por Graunt. Embora também contenha as tabelas de mortalidade, ele estava interessado em questões mais específicas do que estimar o número de pessoas de uma população. Halley queria saber, por exemplo, qual a chance de que um homem de 40 anos possa viver mais 7 anos? Quanto tempo se espera que um bebê recém-nascido viva hoje? Essas questões são aquelas que uma seguradora deve ser capaz de responder para poder oferecer um serviço adequado. Esse artigo intitulado *An Estimate of the Degrees of the Mortality of Mankind, Drawn from Curious Tables of the Births and Funerals at the City of Breslaw; with an Attempt to Ascertain the Price of Annuities upon Lives* é considerado o primeiro trabalho sério na área de Ciências Atuariais.

É importante notarmos que para Bernoulli, De Moivre, Halley e muitos outros estudiosos da época pudessem contribuir, sobretudo, para o desenvolvimento da

Probabilidade e da Estatística, foi necessário que houvesse, antes, um avanço no campo da Álgebra, bem como no desenvolvimento das ferramentas do Cálculo. Porém, a falta dessas ferramentas não era o único empecilho para o desenvolvimento das teorias de Probabilidade e Estatística. O outro fator está relacionado com o reconhecimento e a legitimação dessa área do conhecimento. Embora as searas supracitadas não estivessem ainda consolidadas como um campo do saber, é notório que o trabalho de Huygens começou a despertar o interesse de muitos matemáticos para o assunto, estimulando assim o desenvolvimento desses campos.

Como mencionamos anteriormente, Bernoulli percebeu outra forma de utilizar a Probabilidade, porém sem sucesso. Essa vertente é chamada de Probabilidade inversa. Essencialmente ela consiste em tentar deduzir causas desconhecidas a partir de efeitos conhecidos. Por exemplo, como poderíamos deduzir o número de bolas brancas em uma urna a partir de algumas extrações?

O reverendo Thomas Bayes foi o primeiro a conseguir algum sucesso nessa área, sendo conhecido hoje, principalmente, por um teorema que recebe seu nome. Esse teorema é conhecido pela seguinte fórmula: $P(A|B) = \frac{P(A \cap B)}{P(B)}$, mas Bayes jamais chegou a enunciá-lo de tal forma.

O artigo de Bayes intitulado *An Essay towards Solving a Problem in the Doctrine of Chances*, publicado na segunda metade do século XVIII, começa com um problema a partir do qual muitas proposições e definições vão sendo apresentadas.

Mencionaremos brevemente as ideias apresentadas nesse artigo que, embora não tenham recebido reconhecimento no tempo de sua publicação, hoje são parte integrante do estudo da Teoria das Probabilidades em qualquer curso sobre o tema. Para compreender as ideias publicadas nesse artigo, retomaremos o exemplo de urnas e bolas citado acima.

A partir de algumas extrações desejamos deduzir qual a quantidade de bolas brancas e pretas existentes na urna. Esta é uma tarefa complicada e para resolvê-la é preciso formular algumas hipóteses e confrontá-las. Por exemplo:

H_1: na urna existem somente bolas pretas.

H_2: há mais bolas brancas do que pretas dentro da urna.

H_3: há mais bolas pretas do que brancas dentro da urna.

Obviamente, as três hipóteses não podem ser verdadeiras ao mesmo tempo. A proposta de Bayes é que a probabilidade obtida a priori seja refinada a partir das amostras. Esse refinamento, nesse exemplo, significa que algumas hipóteses serão menos prováveis de ocorrer do que outras ou mesmo serão comprovadamente impossíveis e, consequentemente, deverão ser descartadas.

Imagine que dessa urna, em alguma extração, obtenhamos uma bola branca. Imediatamente poderemos descartar a hipótese H_1. E como poderemos decidir entre as hipóteses H_2 e H_3? Pela Lei fraca dos grandes Números, proposta por Bernoulli, sabemos que para um grande número de extrações, neste caso, com reposição, a chance de obtermos uma bola, digamos de cor branca, em qualquer uma das extrações se aproxima da razão entre o número de bolas brancas extraídas e o número total de extrações realizadas. Desta forma podemos decidir entre as duas hipóteses restantes. Entretanto, a grande contribuição de Bayes se refere às situações em que não podemos realizar um número grande de testagens. Ele descreve, em seu artigo, regras pelas quais os pesquisadores poderiam utilizar a sua intuição sobre a chance de um determinado evento ocorrer para estimar a probabilidade de que um dado fenômeno ocorra. As ideias contidas nesse artigo, possivelmente, foram as primeiras em direção ao assunto que conhecemos como Inferência Estatística.

O Teorema de Bayes, também conhecido como Teorema da Probabilidade Condicional, essencialmente reavalia a chance de um determinado evento ocorrer à luz de novos dados obtidos. Outro aspecto importante da Probabilidade Condicional é que ela nos fornece uma maneira de calcular a probabilidade de situações mais complexas a partir da probabilidade de outra situação já conhecida ou que, ao menos, tenhamos uma ideia sobre o valor dessa probabilidade. Por exemplo: considere os eventos D referente a um problema sanguíneo e S referente a um sintoma. O que desejamos saber é a chance de uma determinada pessoa ter problemas sanguíneos, uma vez que ela apresenta o sintoma S. Em notação atual escrevemos $P(D|S)$. Essa medida de probabilidade é, a princípio, complicada de se obter, porém a probabilidade de uma pessoa apresentar o sintoma S dado que possui problemas sanguíneos, denotado por $P(S|D)$, pode ser encontrada em registros médicos sobre pacientes com esse problema.

O Teorema da Probabilidade Condicional nos fornece uma maneira de calcular $P(D|S)$, se conhecermos $P(S|D)$, $P(S)$ e $P(D)$.

As ideias de Bayes demoraram para serem aceitas e mesmo assim não foram bem recebidas por todos os cientistas, uma vez que dependem dos valores atribuídos a cada hipótese, e esta irá variar de acordo com o conhecimento e intuição de cada pesquisador. Fato que lhe rendeu muitas críticas. Essa maneira de abordar a teoria das probabilidades ficou conhecida como Bayesiana enquanto a outra proposta à qual Bernoulli fez diversas contribuições recebe o nome de frequentista e é sobre esta última que daremos ênfase em nosso trabalho.

2.3 AS BASES DO TEOREMA CENTRAL DO LIMITE

No início do século XIX, o matemático francês Pierre-Simon Laplace fornece uma belíssima contribuição, em seu livro *Théorie Analytique des Probabilités*, para a Teoria das Probabilidades, a qual se consolidará, no século XX, como o Teorema Central do Limite (TCL). Nesta seção não nos ocuparemos em exibir demonstrações sobre o referido teorema, deixando para discutir as demonstrações dos Teoremas Limite em capítulos posteriores. Entretanto, aqui tentaremos expor as principais ideias que tornaram possível o desenvolvimento desse teorema.

Laplace possuía uma visão determinística da Probabilidade, muito possivelmente devido ao crescente sucesso, até o século XX, da mecânica que chamamos hoje de clássica. Tal como os matemáticos Jacob Bernoulli, Descartes e Pascal, Laplace também compartilhava desta ideia. Em defesa do determinismo, Laplace, em sua famosa obra *Essai Philosophique sur les Probabilités*, afirma sobre os eventos que:

> “Na ignorância dos elos que os unem ao sistema inteiro do Universo, fez-se com que eles dependessem de causas finais ou do acaso, conforme correspondessem e se sucedessem com regularidade ou sem ordem aparente. Mas essas causas imaginárias têm sido sucessivamente eliminadas pela extensão dos limites de nosso conhecimento e desaparecem completamente perante a sã filosofia, que vê nelas somente a

> expressão da nossa ignorância em relação às causas verdadeiras." (LAPLACE, 2010, p.42)

Através dessa citação, vemos que Laplace possui uma interpretação epistêmica da probabilidade. Para efeito de esclarecimento, chamaremos de uma abordagem epistêmica aquela cuja probabilidade é uma expressão de nossa ignorância frente ao evento a ser analisado; por exemplo, se estivermos em um estúdio de gravação, livres de sons e imagens do ambiente externo e perguntamos: qual a chance de estar chovendo? Esse é um fator que já está determinado. De fato, ao sairmos do estúdio poderemos constatar se está ou não chovendo. Porém, sem qualquer conhecimento sobre o ambiente externo, talvez fôssemos levados a atribuir 50%. Então, pode-se questionar sobre eventos futuros? Como Laplace lida com questões deste tipo?

> "Devemos considerar o estado presente do Universo como o efeito do seu estado anterior e como a causa do que vai se seguir. Uma inteligência que, em um dado instante, conhecesse todas as forças que animam a natureza e a situação respectiva dos seres que a compõem, e, além disso, fosse suficientemente ampla para submeter todos esses dados à análise, compreenderia na mesma fórmula os movimentos dos maiores corpos do Universo e aqueles do mais leve átomo; nada lhe seria incerto e o futuro bem como o passado estariam presentes em seus olhos". (LAPLACE, 2010, p.42-43)

Tal citação é conhecida como postulado do determinismo clássico universal. Se pensarmos a respeito dessa citação sobre o prisma da teoria das equações diferenciais, o modelo de universo defendido por Laplace é um problema em que se forem dadas as condições iniciais, todas as equações que regem os fenômenos e, além disso, as condições das variáveis em um dado instante, então todas as coisas poderiam ser calculadas com exatidão.Tudo estaria determinado.

Laplace não conduziu seu raciocínio apenas através da visão epistêmica. Ele utilizava livremente uma interpretação frequentista que se apoia fortemente nas contribuições de Bernoulli.

Este é o modelo de pensamento filosófico de Laplace quando escreve suas contribuições à Teoria das Probabilidades. Claramente, não é devido somente a esta corrente de pensamento que se deve a grande influência que as obras de Laplace, relacionadas às Probabilidades, tiveram no decorrer do século XIX, uma vez que muitos matemáticos contemporâneos e até mesmo anteriores a ele também possuíam uma visão determinística dos fenômenos. Então, porque as contribuições de Laplace foram tão influentes nesse século?

Um dos aspectos é que, segundo ROQUE (2012), a partir do século XIX inicia-se um processo de reflexão com relação à separação da Matemática e da Física, isto como consequência de uma necessidade de transformar o rigor matemático, uma vez que as crenças e técnicas dos matemáticos não eram suficientes para resolver os problemas que surgiam dentro da própria Matemática.

Apesar de Laplace tratar de fenômenos físicos utilizando a Probabilidade, ele pontua em seu livro *Théorie Analytique des Probabilités* que a Probabilidade também é importante para a Matemática. Muitos dos problemas tratados dependiam de uma grande quantidade de ensaios de Bernoulli e as probabilidades só poderiam ser obtidas através de fórmulas que eram complicadas demais para um cálculo direto. Desta forma, para que tais cálculos se tornassem realizáveis, algumas considerações deveriam ser feitas de modo a obter métodos de aproximação para essas fórmulas que levavam em consideração uma quantidade numerosa de ensaios.

Além das aplicações à Física e aos contextos moral e social, características marcantes do modelo clássico de probabilidade, surgia um aspecto puramente matemático associado aos métodos analíticos que eram necessários para a resolução de problemas probabilísticos, que o próprio Laplace chamava, em sua obra, de os mais delicados, os mais difíceis e os mais úteis de toda a teoria. Desta forma, as contribuições de Laplace caminhavam juntas com as novas noções de rigor matemático e começavam a legitimar a Teoria das Probabilidades como um campo do conhecimento que não se limita às aplicações clássicas, mas que necessita de um conteúdo específico para que se desenvolva.

Os problemas propostos por Laplace, essencialmente, poderiam ser divididos em dois grupos; o primeiro deles se referindo a problemas envolvendo soma de variáveis aleatórias[3] (basicamente para este grupo calculavam-se as probabilidades *a priori*); e o último sobre probabilidade inversa, o qual se ocupava das probabilidades *a posteriori*.

No final do século XVIII, Laplace já tinha desenvolvido alguns métodos de aproximação para o cálculo das probabilidades *a posteriori*. Porém, foi só em 1810, motivado pelo trabalho do matemático alemão Karl Friedrich Gauss, no qual constava uma rigorosa demonstração de que a observação dos erros seguia o modelo da curva-sino, que Laplace conseguiu obter aproximações para a probabilidade das somas de variáveis aleatórias independentes.Tal aproximação corresponde ao início do desenvolvimento do que chamamos hoje de Teorema Central do Limite. Ele generalizou a lei da curva normal de Gauss para qualquer situação em que uma grande quantidade de pequenas causas ocorriam.

Em 1812, Laplace publica sua obra mais famosa *Théorie Analytique des Probabilités* na qual consta uma síntese completa da Probabilidade das Causas, assim chamada devido à relação de causa e efeito que é uma característica da visão determinística, bem como um material sobre a Análise de Erros.

O TCL foi o teorema que, de fato, tornou famosas as obras de Laplace, pois através dele diversos problemas foram modelados durante o século XIX e foi a principal ferramenta utilizada para a teoria dos erros.

Devido à urgência do século XIX em transformar o conceito de rigor para responder a questões que permaneciam sem solução no interior da própria Matemática, o método de Laplace recebeu algumas críticas de matemáticos da época, relacionadas à confiabilidade do método de aproximação usado por Laplace na expansão das séries. Uma delas foi feita pelo matemático Adrien-Marie Legendre, sugerindo uma forma diferente de aproximação por expansão de séries e, além disso, afirmando que os métodos de Laplace funcionam no caso geral, apenas para expansões de séries semi-convergentes, isto é, as séries convergem, porém não absolutamente. Para Legendre isso levantava dúvidas sobre o tratamento de Laplace. Como seria justificada a negligência dos termos de ordem superior de uma série se ela fosse divergente?

[3] Não encontramos qualquer evidência de que Laplace tenha utilizado essa terminologia em seus trabalhos, entretanto aqui a adotamos a fim de não deixar quaisquer dúvidas a respeito do assunto que estamos nos referindo.

Críticas como essa sugerem que, para a época, os métodos de Laplace podem ter sido considerados não rigorosos por alguns matemáticos. De fato, os métodos de Análise Algébrica utilizados por ele poderiam ser traduzidos em manipulações algébricas para a expansão de séries e no desenvolvimento da teoria sobre as funções geradoras, excluindo o tratamento dos infinitesimais. Entretanto, a teoria que versa sobre os infinitesimais será fundamental dentro da nova concepção de rigor que começará a ganhar espaço na segunda metade do século XIX. Muitas questões em aberto, até então, só serão resolvidas com a inclusão dos infinitesimais na Análise.

Além de Legendre, outros Matemáticos como Siméon-Denis Poisson, Augustin-Louis Cauchy e Johann Peter Gustav Lejeune Dirichlet se interessaram pelo trabalho de Laplace sobre o TCL, porém, concordando que o rigor das demonstrações fornecidas por Laplace era inadequado para época. Eles deixaram também suas contribuições de modo a adequar o TCL à noção de rigor do século XIX.

Poisson compartilhava da visão de Laplace sobre o status da Probabilidade em um sentido clássico, isto é, para ele a Probabilidade ainda estava intimamente conectada com as aplicações à Física. Motivado pelos trabalhos de Laplace, ele produziu dois artigos que, devido às suas técnicas analíticas e às suas discussões sobre a validade da aproximação Normal, influenciaram Cauchy e Dirichlet, cujas contribuições, essencialmente, ajustam as técnicas de aproximação a um estilo diferente de Análise Algébrica. O que contribuiu para que se desenvolvessem novos padrões dentro da Análise e, também, para a modificação do status da Probabilidade.

Com as contribuições de Dirichlet e Cauchy, por volta de meados do século XIX, a Teoria das Probabilidades ganha um caráter mais abstrato, tornando-se parte da Matemática e não mais apenas uma ferramenta para aplicações às ciências Físicas, Sociais e Morais. Embora tivesse ocorrido um grande avanço no desenvolvimento do TCL, algumas das afirmações feitas a respeito dele ainda dependiam do contexto das aplicações. A autonomia, característica marcante da Matemática Moderna, só foi alcançada com a publicação de *Theorem of Laplace* do matemático russo Aleksandr

Lyapunov no início do século XX, segundo Mehrtens[4] (1990, apud Fisher, 2011, p. 68).

Para a consolidação do TCL, foi necessária uma grande mudança de pensamento. O movimento que buscava explicar certos entraves da Matemática existentes até o século XIX auxiliou também a desenvolver uma estrutura própria para a teoria das Probabilidades, trazendo esclarecimentos sobre determinados conceitos, como, por exemplo, sobre a Lei dos Grandes Números. Como definir o que é um grande número de ensaios? Seriam mil ensaios um número grande o suficiente para uma abordagem frequentista aplicada à Loteria? Determinar tal número só foi possível com a inclusão dos infinitesimais aos métodos da Análise. Essa foi uma das contribuições que possibilitou que a Teoria das Probabilidades caminhasse em direção a se tornar uma ciência independente, livre da necessidade de aplicações à Física ou às ciências Sociais. Desta forma para que a Teoria das Probabilidades possa ser compreendida de um modo mais amplo, fazem-se necessárias as ferramentas advindas da Análise moderna.

2.4 IMPACTOS DO TEOREMA CENTRAL DO LIMITE NA ESTATÍSTICA DO SÉCULO XIX

Em meados do século XVIII, ao menos na Alemanha, a Estatística estava intimamente relacionada ao clima ou aos aspectos geográficos de uma região. Já na virada do século, ela começa a ser compreendida como um tipo de sistema para medir a sociedade. Em outras palavras, a Estatística passa a ser o assunto que visa a reconhecer e compreender os aspectos da sociedade. Como exemplo disto podemos destacar o trabalho do político Sir John Sinclair que fez uso de sua publicação (SINCLAIR, 1791) para pressionar por reformas nas salas de aula e nos locais de trabalho. Muitas contribuições como essa foram feitas durante o século XIX. Essencialmente, elas tratavam de contar e de classificar uma grande variedade

[4]Mehrtens, Herbert 1990. *Moderne-Sprache-Mathematik*. Frankfurt: Suhrkamp.

de situações sociais, tais como os índices de natalidade e mortalidade, proliferação de doenças, estaturas dos cidadãos, etc.

O pensamento do século XIX enxerga a Estatística como um sistema a partir do qual a sociedade pode ser, pelo menos em parte, compreendida e, portanto, governada. Dados precisos eram importantes para a administração pública, principalmente no tocante à organização dos serviços militares e na fixação e coleta de impostos. A urgência em coletar dados era tanta que, em meados do século XIX, já haviam sido realizadas cerca de 400 publicações sobre os dados do governo.

Certamente, estudar uma ação individual era algo complicado do ponto de vista científico. Entretanto, quando se olhava para a sociedade e não mais para uma única pessoa pertencente a ela, algumas características próprias passaram a ser percebidas e compreendidas. Em um sentido probabilístico, isto quer dizer que a previsibilidade de uma situação, em um nível individual, era uma tarefa demasiadamente complexa e por muitas vezes até mesmo impossível. Porém, quando as diversas causas que atuavam sobre cada indivíduo eram agregadas, adquiriam certa estabilidade, isto é, em um sentido moderno, tornava-se mensurável a ocorrência de um dado fenômeno. É neste contexto que a Lei dos Grandes Números e o Teorema Central do Limite passam a ser ferramentas amplamente utilizadas.

Em 1844 o astrônomo Adolphe Quetelet fez uso da Teoria dos Erros em diversas áreas, tais como taxas de natalidade e de criminalidade, peso e tamanho de sapato de crianças, bem como em habilidades mentais e musicais. Em cada um desses casos ele encontrou uma distribuição que se aproximava muito da curva Normal de Gauss.

A nova forma de analisar a sociedade recorrendo a dados governamentais, juntamente com o crescente desenvolvimento da Teoria dos Erros, corroborou para que a visão Laplaceana determinística começasse a ser desacreditada e a abordagem frequentista, que se baseia na Lei dos Grandes Números e no TCL, emergisse. Uma das obras que marca o momento em que o tratamento frequentista começa a crescer se chama *Recherches sur la Probabilité des Jugements* publicada em 1837 de autoria de Poisson. Nela a palavra “probabilidade” fica restrita à interpretação epistêmica, enquanto que a palavra “acaso” é reservada para o tratamento frequentista.

Diferentemente da seção anterior, em que as críticas estavam relacionadas, principalmente, aos métodos que Laplace utilizava para calcular aproximações, aqui o que começava a ser questionado era a visão determinista com relação à Probabilidade. Tentaremos expor tal crítica com auxílio de um exemplo.

Considere o lançamento de uma moeda. Se admitirmos a visão Laplaceana, na qual nada é conhecido a respeito da ocorrência de qualquer uma das faces, possivelmente seriamos levados a atribuir chances iguais de ocorrência para ambos os lados da moeda ou, se soubermos algo a respeito da confecção dela, poderíamos atribuir uma probabilidade maior para uma das faces. Porém, segundo a visão frequentista nenhuma explicação seria suficiente para atribuir quaisquer valores de probabilidade para as faces da moeda. Qualquer hipótese que possa atribuir uma probabilidade a um dado fenômeno deve compor a base de dados para a abordagem frequentista, isto é, a probabilidade deve ser obtida através de um grande número de ensaios de Bernoulli.

Tais críticas corroboraram para desconstrução da visão determinística que era predominante na época, e, com o declínio desta concepção, a ciência Estatística caminha em direção à independência de crenças e posicionamentos políticos.

Obviamente que, com o crescimento da abordagem frequentista, muitas críticas pertinentes foram postas. Dentro da área médica, por exemplo, o principal argumento contrário a essa abordagem estava firmado na dificuldade de se obter dados para análise e, além disso, esta é uma área muito complexa, não seria adequado simplesmente dar o diagnóstico de um paciente com base no valor médio dos dados coletados. É necessária uma reavaliação constante, considerando a evolução do estado do paciente. As críticas não eram restritas apenas às áreas médicas, mas também eram feitas na esfera social. O sociólogo francês Auguste Comte, um dos fundadores do positivismo, desaprova o estudo da sociedade feito apenas pela quantificação e afirma que a utilização das médias obscurece o processo de evolução social, sendo necessário para compreender as interações um estudo histórico detalhado da sociedade.

Embora o TCL tenha desempenhado um grande papel em diversas áreas do conhecimento, há diversas situações em que ele não pode ser aplicado. Mas sua utilidade dentro e fora do campo da Estatística legitima sua importância e garante um lugar de destaque dentro do corpo de estudos.

Deste momento até meados do século XX, muitas contribuições foram feitas com relação aos Teoremas Limite de que trataremos posteriormente. Porém, gostaríamos de ressaltar que o matemático russo Andrei Kolmogorov, em 1933, traz um novo conceito de rigor para a Teoria das Probabilidades, em que a teoria adquire um caráter axiomático e, além disso, dependente da Teoria da Medida, sendo esta a forma atualmente aceita de rigor referente à Teoria e independente de quaisquer aplicações.

3 NATUREZA E ESPECIFICIDADE DO RACIOCÍNIO ESTATÍSTICO

No capítulo anterior vimos que a Estatística importou alguns elementos matemáticos e que, além disso, alguns obstáculos tiveram de ser superados, e isto foi crucial para o seu desenvolvimento, mas as formas de se pensar Estatística e Matemática são as mesmas? Sobre a Estatística, Moore (1992) afirma que, embora seja uma ciência Matemática, ela não é um subcampo da Matemática e, além disso, mesmo que a Estatística seja uma disciplina metodológica, ela não é simplesmente um conjunto de métodos que servem a outras áreas. A Estatística possui conteúdo e conceitos próprios e formas de raciocínio distintas.

Compreender as especificidades dessas searas é duplamente importante se pensarmos nos tipos de saberes que um professor deve ter. A respeito do Saber de Conteúdo, tal como descreve SHULMAN (1986), compreender as nuances entre as formas de se pensar Matemática e Estatística é fundamental para se compreender o que é ou não é legitimo dizer dentro de cada estrutura. Por outro lado, se pensarmos que normalmente os estudantes já trazem consigo algum conhecimento a respeito daquele conceito a ser tratado, e que em geral, não é incomum que tais conhecimentos apresentem equívocos, saber o que difere um modelo de raciocínio do outro será de extrema importância para auxiliar o estudante a desconstruir tais concepções errôneas, sendo este um dos elementos que compõem o que Shulman chama de Saber Pedagógico de Conteúdo.

Quais seriam as características que tornariam diferentes as formas de se pensar a Matemática e a Estatística? Neste capítulo tentaremos esclarecer quais são as sutilezas do raciocínio estatístico que o diferem da forma de raciocinar matematicamente.

3.1 A NATUREZA DO RACIOCÍNIO ESTATÍSTICO

Antes de começarmos a analisar os processos inerentes ao raciocínio estatístico, precisamos compreender o que ele é de fato. Encontramos diversos artigos que tratam do raciocínio estatístico, mas apenas alguns autores tentaram exprimir em palavras o que eles compreendem como sendo uma forma de raciocinar estatisticamente.

De acordo com Coladarci e Cobb (2014), o raciocínio estatístico se divide em duas áreas: a Estatística Descritiva e a Estatística Inferencial. A primeira se ocupa de organizar e sintetizar dados a fim de que eles sejam mais facilmente compreendidos, enquanto a segunda visa a conclusões sobre a população, baseando-se nas características de uma amostra dela.

Segundo Garfield e Gal (1999), é a forma como as pessoas raciocinam com as ideias estatísticas e dão sentido às informações estatísticas. Isso envolve interpretar tendo como base um conjunto de dados, representação gráfica e elaboração de resumos estatísticos. Para Garfield e Gal, o raciocínio estatístico combina ideias sobre dados e possibilidades, levando a inferências e interpretações de resultados estatísticos.

Chervaney et al.[5] (1977) e Chervaney et al.[6](1980 apud Garfield, 2002, p. 02) afirmam que o raciocínio estatístico é o que os estudantes são capazes de fazer com o conteúdo estatístico e as habilidades que eles demonstram ao utilizar conceitos estatísticos em problemas específicos. Para esses autores o raciocínio estatístico é um processo de três etapas. A primeira etapa se refere à compreensão, significando associar um problema particular a uma classe de problemas semelhantes; em seguida ocorrem o planejamento e a execução, que consistem em aplicações de métodos adequados para resolução de problemas; e a terceira etapa

[5]Chervaney et al., **A framework for the development of measurement instruments for evaluating the introductory statistics course**. The American Statistician 31, 17-23, 1977.

[6]Chervaney, et al., **The planning stage in statistical reasoning**. The American Statistician, 34, 222-226, 1980.

é onde ocorre a avaliação e a interpretação dos resultados referentes ao problema original.

Citaremos por fim Gal e Garfield (1997), que, ao descreverem os objetivos do ensino da estatística, evidenciam alguns dos elementos básicos que compõem o raciocínio estatístico. Para Gal e Garfiled, a partir do ensino da Estatística espera-se que os estudantes se tornem cidadãos bem informados capazes de compreender e lidar com a incerteza, variabilidade e informações estatísticas no mundo ao seu redor, além disso se pode esperar também que os estudantes contribuam e participem da produção, interpretação e comunicação de dados referentes aos problemas com que se depararem em suas vidas profissionais.

Para que tal objetivo seja alcançado Gal & Garfield (1997) descrevem um conjunto de metas básicas que devem ser atendidas no processo de ensino.

i) Compreender a finalidade e a lógica das investigações estatísticas.
Aqui os alunos devem compreender a necessidade das investigações estatísticas e as ferramentas importantes para a consulta ao banco de dados.

ii) Compreender o processo de investigação estatística.
Isto é, os alunos devem compreender a natureza dos processos envolvidos em uma investigação estatística e as considerações que afetam a elaboração de um plano para a coleta de dados. Além disso eles devem saber como, quando e porque as ferramentas estatísticas podem ser usadas para ajudar em um processo investigativo.

iii) Dominar as habilidades processuais.
Uma vez alcançado este objetivo espera-se que o estudante seja capaz de coletar dados, realizar alguns cálculos, tais como média, mediana e moda, intervalo de confiança, e construir tabelas, gráficos e diagramas, seja a mão ou com auxílio de computadores e calculadoras.

iv) Compreender as relações matemáticas.
Os alunos devem compreender de forma intuitiva e/ou formal as principais ideias matemáticas que servem de base para os conceitos e procedimentos estatísticos. Eles devem ser capazes de responder, por

exemplo, como valores extremos podem afetar o cálculo da média ou como a alteração dos dados afeta a mediana e a moda.

v) Compreender a Probabilidade e as possibilidades

Aqui a ideia defendida é de que a Probabilidade deva ser compreendida apenas informalmente e a partir dessa compreensão o raciocínio da inferência estatística prosseguiria. A proposta deles é de que o conhecimento de probabilidade a ser desenvolvido, para este fim, ocorra a partir de simulações de lançamentos de moedas, dados ou mesmo simulações computacionais.

vi) Desenvolver habilidades interpretativas e literacia estatística.

Aqui os estudantes precisam saber interpretar os resultados e estar cientes dos possíveis vieses e limitações sobre as generalizações que podem ser extraídas a partir dos dados. Isso envolve fazer perguntas críticas e reflexivas sobre os argumentos aos quais se referem as sínteses estatísticas ou o conjunto de dados fornecidos.

vii) Desenvolver a capacidade de se comunicar estatisticamente.

Os estudantes devem ser capazes de utilizar os termos probabilísticos e estatísticos para comunicar os resultados de maneira convincente; devem ser capazes de construir argumentos coerentes com base nos dados ou observações. Além disso, devem conseguir argumentar sobre a validade da interpretação de dados de outras pessoas, bem como levantar questões sobre a aceitação de generalizações feitas a partir de um único estudo ou de uma pequena amostra.

viii) Desenvolver uma maneira útil de utilizar as estatísticas.

Os estudantes devem apreciar o papel da aleatoriedade no mundo para os métodos estatísticos e experiências planejadas como ferramentas úteis e poderosas para o processo de tomada de decisão em face da incerteza.

Além disso, eles devem perceber que a utilização da estatística pode levar a melhores conclusões do que confiar apenas em suas experiências subjetivas, dados anedóticos ou em intuições, mas que isso não é garantido.

> Os alunos também devem aprender a adotar uma postura crítica quando são confrontados com um argumento que pretende basear-se em dados (por exemplo, "todas as pessoas são ...") ou um relatório de resultados ou conclusões de uma investigação estatística, estudo ou pesquisa empírica.

Não nos atreveremos neste trabalho a elaborar uma definição formal referente ao raciocínio estatístico, a fim de não pôr estacas que delimitem suas fronteiras. Porém, recorrendo às contribuições supracitadas, podemos evidenciar alguns dos elementos fundamentais que são inerentes ao raciocínio estatístico.

1. Compreensão do contexto no qual o problema está inserido. A qualidade predominante aqui está na associação do problema a uma classe de problemas semelhantes.
2. Interpretação do problema. Referimo-nos à forma como o problema será encarado. Isto envolve o planejamento de uma metodologia adequada para se resolver o problema.
3. Síntese e organização dos dados coletados. Este é um elemento importante, pois facilita a análise e a interpretação dos dados.
4. Compreensão de como a obtenção de uma amostra adequada é importante para se estudar uma população através da estimação de parâmetros ou testes de hipóteses.
5. Execução da metodologia selecionada. Neste elemento residem os cálculos probabilísticos e estatísticos e os procedimentos que devem ser seguidos a fim de que o método adotado seja satisfeito, o que pode incluir ou não a utilização de softwares. Embora entendamos como muito relevante a utilização de tecnologias digitais para a execução dos cálculos e procedimentos, nem todos os problemas de natureza estatística exigem necessariamente um conhecimento dos softwares.
6. Avaliação e interpretação dos resultados obtidos. Isto é, atribuir um contexto adequado para as respostas obtidas, munindo-as de um significado apropriado para o problema dentro da metodologia proposta. Especificamente, o aluno deverá estar ciente de que, ao contrário dos problemas gerais de Matemática

em que uma resposta dada é garantida com 100% de certeza de sua validade, na seara da Estatística todas as conclusões a respeito do fenômeno estudado são tomadas segundo algum nível de confiabilidade (nunca total).

7. Elaborar representações gráficas adequadas às respostas, a fim de evidenciar e apoiar a conclusão alcançada.
8. Comunicar as conclusões encontradas.

É claro que nem todos os itens apontados refletem características exclusivas da Estatística, como, por exemplo, a compreensão do contexto ou mesmo a interpretação do problema. Estas são características comuns a diversas áreas do conhecimento, tais como a Matemática, Física, Química, etc.

Porém, a partir deles, podemos ressaltar três pilares que regem o raciocínio Estatístico. O primeiro deles é a análise exploratória de dados. Este pilar é bastante prezado pelos PCN e pelos livros didáticos. Essencialmente, ele se refere a como extrair de uma situação (que pode ou não ser regida pelo acaso) informações que possam nos auxiliar a estudar um determinado fenômeno.

Ao tratarmos de fenômenos estocásticos, necessitaremos de ferramentas matemáticas que nos auxiliem a interpretar os dados obtidos, estruturando este fenômeno dentro de um modelo matemático. Neste segundo pilar, as ferramentas da Probabilidade são indispensáveis. Uma vez que a própria probabilidade é a instância da Matemática que permite modelar a incerteza.

O terceiro e último pilar é no nosso entender o “coração” da Estatística. É nele em que se inicia o processo de tomada de decisão sob incerteza, confluindo os dois primeiros pilares com a Teoria de Estimação. Para realizar tal tarefa é necessário coadunar elementos estatísticos e matemáticos, a fim de que seja possível tomar uma decisão ótima, em algum sentido estatístico, e tomar uma decisão com base na medida probabilística associada a uma dada afirmação. A seara da Estatística que trata desses processos de tomada de decisão é chamada Inferência.

3.2 DIFERENÇAS ENTRE O RACIOCÍNIO ESTATÍSTICO E O RACIOCÍNIO MATEMÁTICO

Para compreender a natureza do raciocínio estatístico, precisamos olhar para a forma como a Estatística se desenvolveu. No capítulo 2 desta dissertação, vimos que um dos primeiros estudos sobre a Estatística surge com John Graunt (que não era um matemático), com a publicação de uma tabela que se baseava na causa das mortes ocorridas em Londres. Fazendo uso dessa tabela e de ferramentas advindas da Probabilidade, Graunt pôde estimar, por exemplo, a população de Londres.

Outro campo que está na origem do desenvolvimento da Estatística é a Astronomia, cuja contribuição foi o desenvolvimento de métodos para analisar uma quantidade de medidas imperfeitas relacionadas aos planetas e suas órbitas. Houve uma apropriação dos métodos utilizados nessas searas, a fim de lidar com os dados variáveis que eram obtidos quando se estudavam as anuidades, seguros, epidemias, características populacionais, etc.

Observando o curso histórico da Estatística, concordamos com a argumentação de Moore (1992) de que a Estatística não se origina dentro da Matemática. A Estatística é o raciocínio feito a partir de dados incertos obtidos empiricamente que, por sua vez, pode fazer uso de ferramentas matemáticas. As Estatísticas não são apenas números em uma tabela. Estatísticas são números que carregam um contexto específico.

No contexto da inferência indutiva, são números referentes a um caso que nos possibilita fazer certas generalizações, como, por exemplo, as tabelas publicadas por Graunt. Podemos citar também o trabalho realizado pelo médico epidemiologista John Snow em 1849, que muito antes do conhecimento da teoria microbiana foi capaz de sobrepujar a teoria miasmática, que essencialmente dizia que as doenças como cólera eram causadas pela poluição do ar, sendo ele capaz de afirmar e mostrar que a doença era transmitida através da água contaminada.

Todos os exemplos que envolvem o raciocínio estatístico, que citamos tanto no capítulo 2 quanto neste, fazem uso de ferramentas matemáticas a fim de descrever modelos para analisar os dados. Uma metodologia que vem ganhando muito espaço para a análise de dados é a utilização de softwares estatísticos.

Analisar um conjunto de dados estatísticos é uma tarefa, muitas vezes, complexa e a utilização de uma ferramenta que projete gráficos de maneira dinâmica a partir de um conjunto de dados, pode tornar essa tarefa mais fácil. Forneceremos a seguir um exemplo simples de investigação que utiliza a projeção gráfica dinâmica.

No caso do epidemiologista John Snow, imagine que para descobrir a via pela qual a doença é transmitida ele pudesse fazer uso da tecnologia moderna. Ele veria o mapa da cidade de Londres em uma tela e a casa de cada cidadão atingido pela cólera apareceria como um ponto neste mapa, ocorrendo uma atualização constante nesse sistema. Com algumas dezenas de mortes, possivelmente ele já perceberia que havia uma ocorrência maior de mortes em um distrito de Londres chamado Soho. Quando o número de mortos computados pelo sistema chegasse a algumas centenas, Snow perceberia uma maior concentração de falecimentos próxima à Broad Street, o que sustentaria sua hipótese de que a doença estava sendo transmitida pela água, uma vez que os habitantes pegavam água de bombas públicas localizadas em algumas ruas e uma dessas bombas ficava nessa rua.

Obviamente Snow trabalhou com algumas hipóteses, como, por exemplo, o fato de os habitantes possuírem costume de buscar água nas bombas mais próximas à sua residência. Outro fato que Snow sabia era de que empresas diferentes forneciam água para Soho. Cada empresa era responsável por algumas bombas de água. O que Snow poderia perceber mais rapidamente com o uso da tecnologia era que não só a cólera estava sendo transmitida pela água, mas que era apenas uma das empresas fornecedoras que possuía a água contaminada e, uma vez que as bombas dessa empresa fossem interditadas, o surto de cólera diminuiria.

De fato, mesmo sem a tecnologia, Snow foi capaz de concluir isso, sem que houvesse a necessidade de utilizar as ferramentas advindas da Matemática para resolver um problema estatístico.

A Estatística muitas vezes recorre às ferramentas matemáticas para analisar dados, porém suas próprias ideias e ferramentas não podem ser totalmente acomodadas dentro da Matemática, como, por exemplo, o método da máxima verossimilhança ou as distribuições *a priori* e *a posteriori*, o que torna claro, segundo Moore (1992), a diferença que há entre Estatística e a Probabilidade, que é um campo da Matemática.

Além das ferramentas da Estatística que não podem ser exportadas para as áreas Matemáticas, podemos argumentar que o raciocínio estatístico é aplicado, como já mencionamos, sobre uma coleção de dados, enquanto o probabilístico pode ser aplicado sem que se tenha um dado coletado sequer. Por exemplo, quando lançamos uma moeda ao ar e questionamos a chance de sair coroa, que dados temos sobre essa moeda? Nenhum. E mesmo assim, podemos atribuir uma medida de Probabilidade à face desejada.

A discussão que temos feito até aqui mostra que a Estatística e a Matemática são campos diferentes, mas que podem interagir e essa interação nos ajuda a compreender uma grande quantidade de fenômenos. Contudo, isto não revela o que torna a Estatística diferente da Matemática.

Haveria prejuízo para os processos de ensino e de aprendizagem a não compreensão de suas naturezas diferentes? Se não houver prejuízo, então poderemos compreender uma afirmação Matemática da mesma forma como compreendemos uma afirmação Estatística. Entretanto, não podemos fazer tal coisa.

A Matemática é construída sobre as bases de uma lógica proposicional, a fim de que a partir de algumas afirmações seja possível deduzir novas afirmações. Desta forma, uma afirmação Matemática ou é verdadeira ou é falsa. Não existe uma terceira opção (princípio do terceiro excluído).

Em contraponto à Matemática, sobre uma afirmação Estatística, nunca será possível saber se ela é falsa ou verdadeira com total certeza. O que existe é apenas um grau de confiabilidade sobre a afirmação realizada. Por exemplo, nunca poderemos afirmar se um medicamento funciona. O que poderemos afirmar é que existe uma probabilidade p de que esse medicamento traga bons resultados e uma probabilidade $1 – p$ de que não traga.

Então, de fato, há prejuízo para o estudo da Estatística a não compreensão da sua natureza que a diferencia da Matemática. Uma vez que sem essa compreensão as afirmações estatísticas podem ser interpretadas erroneamente. Por exemplo, a pesquisa feita pelo instituto *O Círculo da Matemática no Brasil* afirma que 89% dos brasileiros com mais do que 25 anos dizem que não utilizam Matemática em situações do dia a dia.

Apesar de ser uma medida bastante significativa, isto não significa que, ao perguntarmos a um brasileiro com idade superior a 25 anos se ele utiliza Matemática no seu cotidiano, ele responderá que não. Apenas indica que se perguntarmos a um número suficientemente grande de brasileiros com idade superior a 25 anos, aproximadamente 89% deles afirmarão não utilizar, enquanto 11% afirmará o contrário.

Embora Estatística e Matemática sejam assuntos com conteúdo e raciocínios próprios, é bem verdade, como vimos, que há uma forte interação entre estes dois campos. A Estatística se serve de muitas ferramentas Matemáticas, como os elementos da Análise, do Cálculo Diferencial e Integral, da Álgebra e da Probabilidade, que foram, particularmente, úteis para o desenvolvimento de sua teoria (o Teorema Central do Limite, amplamente utilizado pela Estatística, sendo um bom exemplo a mencionar).

3.3 ALGUNS ELEMENTOS FUNDAMENTAIS PARA O RACIOCÍNIO ESTATÍSTICO

Na seção 3.1 deste trabalho discutimos alguns elementos do raciocínio estatístico. Agora, discorreremos um pouco mais sobre esses elementos.

Quando se pretende fazer o estudo de uma população, a primeira coisa a se fazer é obter informações (parciais) sobre ela a partir de uma amostra aleatória. Geralmente, a população a ser analisada possui um grande número de elementos, o que torna inviável estudar cada um deles individualmente, daí a necessidade de se obter uma amostra aleatória desta. Mas porque essa amostra extraída precisa ser aleatória?

O objetivo de tomar uma amostra aleatória é simplesmente evitar certas tendências que podem ocorrer localmente, influenciando a análise.

Considere, por exemplo, que desejamos fazer um estudo para verificar se as carnes produzidas em uma indústria alimentícia estão em boas condições para o consumo. Obviamente não poderíamos testar todas as peças de carne, uma vez que após a carne ser testada ela não pode ser mais consumida. Por outro lado, também não podemos selecionar uma amostra tão pequena, como por exemplo, de uma única peça, pois imagine se apenas 20% das carnes produzidas estivessem em boas condições e pegarmos justamente uma dessas peças. Certamente isso não refletiria a realidade e acarretaria uma série de problemas para a indústria e seus consumidores. Então, precisamos de uma amostra suficientemente grande para compor um conjunto de dados a ser analisado, o que é amparado pela Lei dos Grandes Números. Além disso, essa amostra deve ser aleatória para que seja garantida a honestidade da análise dos dados. Sem a aleatoriedade, por exemplo, um fiscal poderia pegar somente as carnes que ficaram em um frigorífico com a temperatura adequada, enquanto as outras poderiam ter ficado em frigoríficos que, por algum defeito, estavam em temperaturas muito superiores à apropriada e, consequentemente, se tornando impróprias para o consumo.

Dentre os elementos que destacamos a partir das contribuições de vários autores está a compreensão do contexto em que o problema está inserido. É a partir da compreensão do contexto que poderemos começar a pensar em quais são os métodos mais adequados que podem nos guiar no estudo do problema. É muito comum que nos deparemos com questões semelhantes a outras que são de nosso conhecimento. Perceber essas situações é uma habilidade bastante desejável, que poderá auxiliar na resolução de diversos problemas, nos poupando bastante tempo no planejamento de uma metodologia.

Outro elemento que evidenciamos é a interpretação do problema. Este é um elemento muito amplo, pois engloba diversos aspectos da questão a ser analisada, desde a escolha de uma metodologia até a análise dos elementos da população. Por exemplo, imagine que haverá daqui a uma semana uma grande manifestação em uma das principais avenidas da cidade. Após o evento, como saber quantas pessoas participaram dele? Existem diversos métodos para tratar dessa questão, porém os resultados obtidos por eles não serão necessariamente iguais ou próximos. Um método que utiliza algumas fotografias tiradas do alto da avenida durante a primeira hora da manifestação e recorre a uma estimativa via cálculo de área ocupada da avenida, quase certamente apresentará um resultado diferente de um método que envia uma equipe treinada ao local e que, de tempos em tempos, cada membro da equipe escolhe um ponto aleatório de um metro quadrado da avenida e conta quantas pessoas há naquele espaço.

É claro que a metodologia não é apenas a determinação da forma como os dados serão obtidos, mas também do refinamento dos dados coletados e da adequação do método. No primeiro caso, claramente o método não é adequado para contar o número de pessoas que estavam presentes na manifestação. Os dados coletados são apenas sobre a primeira hora, não há diferenciação dos que estavam presentes na manifestação daqueles que apenas estavam passando pelo local, etc. Existem diversos fatores que tornam essa metodologia inadequada. Já no segundo caso, por exemplo, poderia ser modificado incluindo uma pequena entrevista com os indivíduos contados, a fim de verificar quantos, de fato, pertencem à manifestação e quantos estão simplesmente passando pela avenida.

Além disso, é preciso estar atento à consistência dos dados obtidos.Seria um erro considerar que em determinado momento e local havia 56 pessoas em um metro quadrado.

A síntese e a organização dos dados coletados nos auxiliarão a identificar possíveis erros na coleta de dados e facilitarão o refinamento dos dados obtidos. Suponha que a manifestação não ocorra em toda a extensão da avenida, mas somente em uma parte dela. Podemos refinar os nossos dados retirando aqueles que foram obtidos em locais distantes de onde ocorreu a manifestação. A organização dos dados também servirá para analisá-los, facilitando a inserção deles em algoritmos apropriados, cujos resultados precisam ser avaliados e interpretados.

O que o resultado obtido representa? Será que ele está coerente com as observações? Muitas das vezes para que possamos fazer uma boa avaliação dos resultados recorremos a gráficos. A escolha de um modelo gráfico apropriado pode ser bastante reveladora, como, por exemplo, os gráficos construídos por John Snow em seu estudo sobre a cólera. Ter o conhecimento de diferentes tipos de gráficos e como eles podem ser utilizados para enaltecer determinada característica se torna fundamental para o desenvolvimento do raciocínio estatístico.

Há diversas maneiras pelas quais os resultados podem ser comunicados, porém uma escolha inadequada de termos ou imagens pode obscurecer os resultados obtidos e até mesmo encobrir a importância da pesquisa realizada.

Além disso, ressaltamos a importância das ferramentas probabilísticas que permeiam esses elementos, sendo úteis tanto para interpretação dos dados quanto para a sua análise, além de tais ferramentas desempenharem um papel de grande relevância para que se possa ter um posicionamento crítico frente aos resultados obtidos. Por exemplo, considere que desejamos fazer um estudo sobre obesidade infantil. Neste caso, dificilmente teremos acesso a toda população. Então necessitaremos recorrer a uma amostra. A partir dessa amostra gostaríamos de realizar determinados tipos de afirmações sobre o peso médio das crianças dessa população, a variância desses pesos, o percentual de crianças com sobrepeso ou com o grau de elevado de obesidade, etc.

Entretanto, para que possamos fazer afirmações é necessária uma base probabilística sólida para tratar da teoria da estimação, dos testes de hipóteses, das leis que regem o comportamento dos estimadores, etc.

Tomando por base esses elementos, podemos notar claramente a importância que o contexto no qual o problema está inserido possui para a Estatística. A compreensão do contexto se torna indispensável desde o início do raciocínio estatístico até o final, quando os resultados são comunicados. Com isso, deixamos aqui expressa a nossa reserva quanto à forma atual como a Estatística é ensinada no Ensino Básico. Saturar o estudante com uma enorme quantidade de contas de médias, desvio padrão, mediana, etc, totalmente desprovidas de um significado, não contribui *ipso facto* para que ocorra o desenvolvimento de um raciocínio estatístico. Esses conceitos se tornam apenas fórmulas matemáticas, tais como a fórmula para resolução de equações do segundo grau ou para obter os vértices de uma parábola. A mera aplicação de fórmulas mais presta um desserviço ao raciocínio estatístico do que colabora com seu aprendizado.

4 O PAPEL DA PROBABILIDADE NA CONSTRUÇÃO DO RACIOCÍNIO ESTATÍSTICO

A Teoria das Probabilidades desempenha um papel fundamental na construção do raciocínio estatístico e compreender como essas duas teorias se relacionam, segundo SHULMAN (1986), no que se refere ao Saber de Conteúdo, é importantíssimo para que o professor possa estruturar sua aula de maneira a permitir que o estudante compreenda o assunto tratado. A importância da Probabilidade para a Estatística é notória quando analisamos o desenvolvimento de sua história; e isso não se restringe apenas aos seus conceitos e fórmulas. Pascal e Fermat, quando discutem o problema proposto por de Méré, já utilizam a probabilidade como uma ferramenta para uma tomada de decisão, ao indagarem se, ao realizar um grande número de apostas, de Méré terá lucro ou prejuízo na empreitada.

Ora, a questão proposta por de Méré foi motivada pelo fato de suas observações empíricas não estarem condizentes com seu cálculo probabilístico. Desta forma, a probabilidade tornou-se uma ferramenta utilizada para compreender o fenômeno empírico, e contribuiu à tomada de decisão sobre apostar ou não.

Ainda à guisa de exemplo, sobre a importância histórica da probabilidade para o desenvolvimento da estatística, podemos citar o conceito de esperança matemática desenvolvido por Huygens em termos de valor esperado de uma aposta, ou mesmo as contribuições de Graunt na estimaçãodo tamanho de uma população a partir de uma tabela de vidas. Nesses processos, a probabilidade aparece como uma ferramenta indispensável para a tomada de decisão, ganhando, inclusive, interesse político a partir das contribuições de Graunt. Desta forma afirmamos que a compreensão da Probabilidade é essencial para que o raciocínio Estatístico ocorra.

A Probabilidade possui um caráter dual, isto é, ela tanto pode ser vista como um objeto matemático como uma medida de conjuntos dentro da Teoria da Medida, quanto pode ser vista como uma ferramenta para a Estatística. Enquanto objeto matemático a Probabilidade é um tipo de medida que diz respeito à chance de ocorrência de um evento não determinístico, como por exemplo, a chance de que em um lançamento de um dado cúbico ocorra uma determinada face. O valor atribuído à chance de ocorrência desse evento é uma medida de probabilidade. Por outro lado, a probabilidade desempenha, dentro da Estatística, um papel crucial para a tomada de decisão. Considere o caso em que um juiz deve julgar um caso de paternidade cujo homem alega não ser o pai da criança. Para o juiz as semelhanças físicas entre o homem e a criança não são um bom indício de que há, de fato, algum grau de parentesco entre os dois, porém o resultado de um exame de DNA é suficiente para que o juiz tome sua decisão contra ou a favor do homem. Neste exemplo, a probabilidade é uma ferramenta estatística que auxilia o juiz a concluir o caso. Pois, para um juiz, duas pessoas possuírem algumas características em comum é bastante natural, no sentido de a probabilidade de que o homem seja pai da criança, tomando apenas isto como base, é muito pouco confiável. Em outras palavras, a probabilidade de que o juiz tome uma decisão errada com base nessa evidência é bastante alta. Por outro lado, quando o resultado do exame de DNA é apresentado a chance de que o juiz tome uma decisão errada se torna muito pequena, pois a probabilidade de que um exame de DNA esteja incorreto é ínfima.

Neste capítulo iremos discutir em maiores detalhes como as probabilidades são importantes para a Estatística e daremos atenção especial à abordagem frequentista devido à sua forte conexão com os Teoremas Limite que apresentaremos no capítulo 5, bem como a sua ampla quantidade de aplicações, que podem ser trabalhadas em nível de Ensino Básico.

4.1 INTERPRETAÇÕES

Uma boa compreensão da Probabilidade se dá a partir de seus muitos aspectos, dentre eles consideramos importante a diversidade de interpretações. Aqui destacaremos apenas as interpretações clássica, frequentista e subjetiva, por serem, geralmente, os primeiros modelos utilizados no processo de ensino-aprendizagem de Probabilidade e por possuírem grande relevância para o estudo da Estatística. A abordagem clássica é aquela que entende todos os resultados de um dado experimento como equiprováveis; a abordagem frequentista se refere aos eventos estatisticamente testáveis através da frequência relativa; e a abordagem subjetiva se refere aos métodos estatísticos bayesianos que recorrem à medida de crença que um indivíduo possui a priori sobre a ocorrência de um determinado evento.

Além desses três modelos de interpretação de probabilidade há ainda outros como: lógica, propensão, intersubjetiva, etc. Cada modelo citado possui seu próprio conjunto de ideias e definições que servem para compreender e analisar uma gama de fenômenos não determinísticos. Porém, esses métodos ainda carecem de uma formalização matemática, pois em suas definições encontra-se um grande apelo à intuição. Neste sentido, faz-se necessário uma teoria unificadora, abrangente o suficiente que incorpore as diversas interpretações de probabilidade, e que, além disso, defina com o rigor matemático o que seja Probabilidade.

Em 1933, Kolmogorov apresenta um conjunto de axiomas matemáticos que definem Probabilidade. Através desse conjunto de axiomas é possível concluir cada uma das interpretações mencionadas como um caso particular dessa definição. Entretanto, ressaltamos que, neste trabalho, não realizaremos tais demonstrações por se distanciarem do objetivo principal.

Após as discussões sobre as interpretações, iremos apresentar brevemente uma abordagem axiomática relacionada à Teoria da Medida, a fim de termos uma definição matematicamente rigorosa de Probabilidade. Desta forma, as interpretações de probabilidade serão parte de uma estrutura cognitiva que subjazem ao conceito de Probabilidade.

Assim, as diferentes interpretações probabilísticas são tão úteis para a análise e compreensão dos fenômenos quanto para a compreensão dos próprios conceitos de Probabilidade e Estatística.

4.1.1 Interpretação clássica

A interpretação clássica de probabilidade é tratada por diversos matemáticos. Dentre eles, destaca-se Laplace que no início de sua obra *Ensaio filosófico sobre as* *probabilidades* define a forma de calcular as probabilidades, e é esta definição que conduz o modelo de interpretação clássico.

> “A teoria dos acasos consiste em reduzir todos os eventos do mesmo gênero a um certo número de casos igualmente possíveis, de forma tal que estejamos igualmente indecisos sobre sua existência, e em determinar o número de casos favoráveis ao evento cuja probabilidade é desejada. A relação entre esse número e aquele de todos os casos possíveis é a medida dessa probabilidade, que corresponde assim a uma fração cujo numerador é o número dos casos favoráveis e o denominador é o número de todos os casos possíveis.” (LAPLACE, 2010, p.46)

O que Laplace nos informa com essa definição? Ela é suficiente para responder aos diversos fenômenos no qual o acaso se faz presente? De acordo com esta forma de interpretar podemos perceber que a própria realização do experimento não é necessária para que se possa atribuir uma medida de probabilidade a um dado evento.

O ponto central do modelo clássico está intimamente conectado ao que ficou conhecido como *princípio da razão insuficiente*, esse princípio afirma que dado uma quantidade $n > 1$ de eventos, se não sabemos nada a respeito desses eventos que se relacione com a chance de ocorrência, então a eles serão atribuídas as mesmas medidas de probabilidade.

Há além desse, alguns outros aspectos da probabilidade clássica que podemos notar a partir da definição de Laplace, como, por exemplo, a quantidade finita de eventos elementares e o conhecimento de todos os eventos possíveis de ocorrer. Naturalmente, existem críticas a essa interpretação mesmo se nos restringirmos às aplicações dos jogos de azar, nicho ao qual esse modelo é mais comumente aplicado.

Suponha que duas pessoas disputem um jogo de dardos, em que o vencedor será o primeiro que acertar o alvo. Nesse jogo os dois competidores revezam o lançamento, então qual será a probabilidade de vitória daquele que começar o jogo?
Ora, mesmo se aplicarmos o princípio da razão insuficiente e fixarmos a chance de cada jogador acertar o alvo a cada rodada em 50%, ainda assim não poderíamos calcular a probabilidade do primeiro jogador vencer, pois a quantidade lançamentos

que cada jogador deve fazer até que o jogo termine é desconhecida, podendo ainda ocorrer de nenhum dos jogadores conseguir acertar o alvo num número finito e pré-fixado de vezes. A consideração de todos os casos possíveis implicaria um número infinito de possibilidades e sobre estes casos a interpretação clássica é insuficiente.

Mas, a principal crítica, de fato, é sobre o princípio da razão insuficiente. Como atribuir medidas iguais de probabilidades a eventos sobre os quais somos ignorantes a respeito das chances de ocorrência? Parece tão temerário quanto atribuir uma probabilidade de 30% e outra de 70% para as faces de uma moeda sobre a qual não temos qualquer informação. Atribuir probabilidades iguais não faz com que a chance de cometer esse erro seja menor. Apenas se propõe a realizar, em certo sentido, um tipo de justiça que não atribui qualquer favorecimento aos eventos cujas chances de ocorrência nada sabemos.

Para os defensores da interpretação clássica, isso não era apenas uma atribuição de medidas iguais de probabilidade e sim uma forma pela qual a ignorância deve ser epistemologicamente conduzida. Isto é, nenhum evento deve ser favorecido, se não há qualquer evidência de que sua chance de ocorrência seja maior ou menor do que a chance de ocorrência dos demais eventos. A forma pela qual o não favorecimento de qualquer evento é realizado se dá através de atribuições equivalentes de medidas de probabilidade.

4.1.2 Interpretação frequentista

Este modelo de interpretação, diferentemente do modelo clássico, está intimamente relacionado aos experimentos empíricos. A definição de probabilidade adotada para esta abordagem é descrita pelo Matemático e Engenheiro Mecânico *Richard Von Mises.*

> Probabilidade de um evento é tomada como limite da frequência relativa, considerando repetições infinitas de experimentos, sob as mesmas condições (daí o nome frequentista que esta teoria também possui). No entanto, observação empírica é finita e esta teoria seria aplicável àquelas

> sequências finitas cujo comportamento se aproxima da idealização da sequência infinita. [...]. (MISES[7], 1928 apud ARAUJO; IGLIORI, 2013, p.64)

Segundo esta definição, a medida de probabilidade é obtida através de uma sequência de eventos aleatórios, tal como a sequência de lançamentos de uma moeda. Segundo Popper (1993), a sequência de eventos à qual Mises se refere deve obedecer a duas condições, que são conhecidas como o axioma de convergência e o axioma de aleatoriedade. O axioma da convergência garante que a sequência de eventos observados tende para um limite definido, enquanto que o axioma de aleatoriedade afirma que, para qualquer subconjunto infinito dessa sequência de eventos, o limite permanece o mesmo daquele encontrado para o conjunto de todos os eventos. Qualquer sequência infinita de eventos que satisfaça a esses dois axiomas, Mises chama de *kollektiv* ou *coletivo* em tradução livre.

Esse modelo de interpretação é bastante útil para as ciências Estatísticas, porém é restrito aos eventos que podem ser reproduzidos empiricamente diversas vezes e é neste ponto que recebe duras críticas, pois existe uma grande quantidade de fenômenos aleatórios que não podem ser reproduzidos empiricamente. Por exemplo, qual a probabilidade de que na próxima semana faça sol? Qual a probabilidade de que um determinado indivíduo venha a falecer em no máximo um ano?

A interpretação frequentista de Mises foi alvo de outras críticas devido à imprecisão da definição. O termo limite empregado por ele na definição gera uma má compreensão, uma vez que esse limite não poderia ser entendido como um limite matemático, pois dados n e ϵ ambos maiores que 0 poderá existir um $n_0 > n$ tal que $\left|P(A) - \frac{n(A)}{n_0}\right| > \epsilon$. Embora isto seja improvável, não é impossível. Nesta notação $P(A)$ corresponderia ao limite da frequência relativa do evento A e $\frac{n(A)}{n_0}$ é a frequência relativa definida para um n_0 tão grande quanto se queira. Além disso, como se trata de um experimento empírico não faz sentido considerar o número de ensaios n tendendo para o infinito. Portanto, não há garantia de que tal limite exista.

Outro ponto que destacamos é que a Lei dos Grandes Números, amplamente utilizada nesta visão, é restrita a eventos independentes, isto é, cada elemento de

[7] Mises, Von. Probability, Statistic and Truth, 1928.

uma coleção de Mises, não exerce influência que altere a medida de probabilidade de qualquer outro elemento da coleção. Desta forma, exclui-se o refinamento da probabilidade a partir da experiência. Falaremos mais sobre isto na seção 4.1.3.

Apesar de a interpretação frequentista estar restrita a fenômenos que podem ser reproduzidos empiricamente, em certo sentido podemos afirmar que ela expande os conceitos probabilísticos do modelo clássico. Mesmo que o modelo clássico não tenha essa característica de reprodutibilidade do fenômeno e o seu cálculo esteja mais associado a medidas combinatórias, através de um modelo apropriado, podemos interpretar esses mesmos fenômenos à luz da interpretação frequentista. Por exemplo, se queremos saber qual a chance de obter a soma 11 como resultado do lançamento de dois dados cúbicos cujas faces equiprováveis são numeradas de 1 a 6, basta analisar, sob a visão clássica, quantas são as combinações diferentes de resultados que somam 11, isto é, em pares ordenados, como (6,5) e (5,6) e depois estabelecer a razão da quantidade de resultados distintos que somam 11 pela quantidade total de resultados possíveis; já na visão frequentista, essa medida de probabilidade seria confirmada através de um número suficientemente grande de lançamentos realizados.

Assim, as medidas de probabilidade obtidas para a visão clássica podem ser confirmadas através da visão frequentista.

Mesmo com todas as críticas e apontamentos sobre sua limitação, a interpretação frequentista detém um alto grau de importância para o desenvolvimento da Estatística Inferencial de caráter frequentista, pois há uma grande quantidade de fenômenos relevantes para o desenvolvimento científico e social que podem ser tratados por este modelo. Como exemplo, podemos citar os já mencionados estudos de Graunt, de Snow ou ainda as contribuições para a teoria dos erros no qual o Teorema Central do Limite é um elemento fundamental.

Quanto ao rigor da definição proposta por Mises, Kolmogorov, ao descrever uma base axiomática para a teoria das Probabilidades, fundamentou os conceitos intuitivos que eram aceitos.

4.1.3 Interpretação subjetiva

Os primeiros Matemáticos a defenderem este modelo foram *Bruno de Finetti* e *Frank Ramsey*. Ambos fizeram isso no início do século XX. O grande foco da interpretação subjetiva está na possibilidade de refinar a probabilidade levando em consideração os diversos aspectos conhecidos sobre um determinado evento. Para realizar esta tarefa, o modelo subjetivo se apoia nas contribuições de Bayes, mais especificamente no teorema da probabilidade condicional.

A interpretação subjetiva, diferentemente da frequentista, não necessita da realização empírica do evento, pois ela se baseia no grau de crença subjetivo que um sujeito possui sobre determinado evento. O problema disto é que as medidas de probabilidade atribuídas podem variar de indivíduo para indivíduo. Por exemplo, se perguntarmos a um sujeito, que não conhece nada a respeito do Brasil, sobre qual é a chance de que no mês de janeiro ocorra um surto de dengue, este poderá, pela falta de maiores informações, atribuir a probabilidade de 50%. Enquanto outro que possui maiores informações sobre o clima do país e as condições que favorecem a proliferação do mosquito pode considerar uma probabilidade de 70%. E ainda um terceiro indivíduo que, além disso, reside no país e sabe que é comum ocorrer surtos de dengue em janeiro pode atribuir uma probabilidade de 100%.

Porém, esse é um problema que pode ser, de certa forma, solucionado. Uma vez que as diferentes visões dos observadores podem ser inseridas e consideradas dentro de um determinado modelo probabilístico. Entretanto, uma das principais críticas a esse modelo está justamente no agente que atribui a medida de probabilidade. Quais devem ser as características respeitadas pelos agentes que pretendem atribuir uma medida de probabilidade a fim de que ela possa ser útil?

Em primeiro lugar, os axiomas da probabilidade descritos por Kolmogorov devem ser atendidos, ao menos para uma quantidade finita de eventos, para que a medida de probabilidade seja considerada coerente. Esta é a essência do Teorema de Ramsey-De Finetti. Por exemplo, se A e B são dois eventos disjuntos, não pode ocorrer que $P(A \cup B) < P(A) + P(B)$. Mas o que é coerência? Segundo Gillies, em seu livro *Philosofical Theories of Probability* (2006), coerência está relacionada com o comportamento racional de um indivíduo.

> If Mr B has to bet on a number of events E1,..., En, his betting quotients are said to be coherent if and only if Ms A cannot choose stakes S1, ..., Sn such that she wins whatever happens. If Ms A can choose stakes so that she wins whatever happens, she is said to have made a Dutch book[8] against Mr B. It is taken as obvious that Mr B will want his bets to be coherent, that is to say he will want to avoid the possibility of his losing whatever happens. Surprisingly, this condition is both necessary and sufficient for betting quotients to satisfy the axioms of probability. (GILLIES, 2006, p. 58-59)

Por essa definição, o grau de crença subjetivo de um sujeito está relacionado com o seu comportamento no que diz respeito à importância que ele está disposto a apostar na ocorrência de um determinado evento. Na definição apresentada, o termo que se refere a isto é o *betting quotient*, que em tradução livre significa quociente de aposta. Sendo mais preciso, o quociente de aposta é o grau de crença subjetivo que um determinado sujeito possui sobre um certo evento, e de acordo com Gillies (2006), um conjunto de quocientes de apostas é dito coerente *(coherent),* se e somente se, os quocientes de apostas obedecem aos axiomas da probabilidade.

Com isso, não se espera que sujeitos com conhecimentos diferentes a respeito de um dado evento concordem com uma mesma medida de probabilidade, mas são definidas as bases para que o grau de crença seja obtido.

É importante ressaltar que a interpretação subjetiva não está restrita a eventos que não podem ser reproduzidos empiricamente. A frequência relativa de um evento, quando conhecida, pode ser incorporada ao modelo probabilístico subjetivo através do teorema de Bayes, que para eventos A e B mutuamente independentes podem ser obtidos da seguinte forma:

$$P(A \cap B) = P(A).P(B)$$

Ou seja, a medida de probabilidade de que dois eventos independentes ocorram é o produto das medidas de probabilidade de que cada um ocorra.

Sobre essa interpretação, pesam as críticas de que nem sempre é possível atribuir um valor numérico para a medida de probabilidade e de que um agente

[8] Um *Dutch book* consiste em uma série de apostas que são favoráveis ao agente, mas que quando olhadas em conjunto garantem a sua perda, segundo Zalta et al, 2007

poderia ignorar algumas situações de seu conhecimento, a fim de que a probabilidade obtida seja vantajosa para ele.

Mesmo assim, a utilização da interpretação subjetiva pode ser bastante vantajosa para o cálculo estatístico, em muitos casos. Por exemplo, quando perguntamos qual a probabilidade de que uma determinada mulher grávida conceba um menino, podemos simplesmente utilizar uma interpretação frequentista a partir dos dados do censo, o que nos fará atribuir uma probabilidade de aproximadamente 50%. Porém, se soubermos que essa mulher reside em uma pequena cidade, em que, o nascimento de mulheres é muito superior ao de homens, poderíamos atribuir uma medida de probabilidade maior para que a criança a ser concebida seja do sexo feminino.

4.1.4 Tratamento axiomático

Nesta seção, apresentaremos as bases axiomáticas descritas por Kolmogorov em 1933. Esse conjunto de axiomas e definições tem o objetivo de atribuir à Probabilidade um rigor Matemático tão forte quanto o da Geometria Euclidiana ou da Álgebra. Além disso o desenvolvimento dessa teoria culmina na unificação das múltiplas interpretações probabilísticas, isto é, através dos axiomas descritos por Kolmogorov, todas as interpretações mencionadas aqui podem ser obtidas como um caso particular. Porém, isso não minimiza e tampouco invalida as diferentes interpretações. Essas interpretações estão conectadas a certos tipos de fenômenos e seus respectivos contextos. Já a abordagem axiomática que apresentaremos tem caráter puramente matemático, em que o contexto no qual os fenômenos estão inseridos é abstraído, e tem o objetivo de colocar em bases sólidas as teorias probabilísticas desenvolvidas até o primeiro terço do século XX.

Para esta seção recorreremos à obra de James (2011) ao invés de utilizarmos uma tradução inglesa da obra *Grundbegriffe der Wahrscheinlichkeitrechnung* de Kolmogorov simplesmente pela familiaridade com as notações e formas de escrita. Porém, ressaltamos que não há qualquer prejuízo a respeito do conteúdo.

Antes de darmos início ao modelo axiomático, que tem por base a teoria da Medida, será necessário introduzir algumas definições básicas. A Probabilidade está relacionada a *experimentos não determinísticos*, isto é, os resultados desses experimentos são completamente imprevisíveis, mesmo quando são reproduzidos em condições iguais. Porém, a imprevisibilidade do resultado não nos impede de conhecer quais são os possíveis resultados que podem ocorrer para um determinado tipo de experimento. Por exemplo, ao jogar uma moeda, não sabemos qual a face ficará voltada para cima, mas temos a certeza de que, a menos de alguma extravagância, esta face será cara ou coroa.

O conjunto de todos os resultados elementares e indissociáveis de um experimento é chamado *espaço amostral* e denotaremos por Ω. Definir os elementos de um espaço amostral não é uma tarefa necessariamente simples. Por exemplo, se formos solicitados a tomar um alimento aleatoriamente, e a determinar o tempo em que este alimento permanece adequado para o consumo, como escrever o espaço amostral desse experimento? O espaço amostral não precisa ser caracterizado como um conjunto minimal dos resultados possíveis de ocorrer.

O que é necessário é apenas que todos os resultados possíveis estejam contemplados pelo conjunto amostral. Por exemplo, para o experimento citado acima, podemos tomar $\Omega = [0, \infty)$, em que os números do intervalo representam o tempo (em horas) em que o alimento permanece adequado para o consumo, embora nos pareça óbvio que nenhum alimento durará mil anos. No entanto, a análise exploratória de dados nos fornecerá informações valiosas para modelar distribuições de probabilidade para a variável tempo de tal maneira que as probabilidades convirjam para zero conforme o tempo seja grande.

Desta forma o *espaço amostral* é definido como um conjunto que obedece às seguintes condições:

1. Contém todos os resultados possíveis de ocorrer no experimento.
2. Todo resultado possível corresponde a um, e somente um, ponto $\omega \in \Omega$.
3. ω não pode representar mais do que um resultado.

Além disso, chamaremos de *evento* todo subconjunto $A \subset \Omega$, ao qual seja possível atribuir uma medida de probabilidade, ou seja, todo conjunto mensurável pela medida de probabilidade. Na Teoria das Probabilidades, Ω é dito um *evento certo*, o conjunto vazio $\emptyset$ é denominado de *evento impossível*, e, para todo $\omega \in \Omega$, o evento $\{\omega\}$ é chamado *evento elementar*.

Aos eventos mensuráveis podemos atribuir uma medida de probabilidade, como visto anteriormente nas subseções referentes às diferentes interpretações de probabilidades. Há diferentes formas pelas quais podemos atribuir uma probabilidade a um *evento aleatório*.

Estas definições serão suficientes para que possamos introduzir as contribuições da Teoria da Medida à Probabilidade.

Consideremos $\mathcal{A}$ uma classe de eventos aleatórios que satisfazem às seguintes propriedades:

A1. $\Omega \in \mathcal{A}$

A2. Se $A \in \mathcal{A}$, então $A^c \in \mathcal{A}$

A3. Se $A_n \in \mathcal{A}$ para $n = 1,2,3, \dots$, então $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$ (diremos apenas que $\mathcal{A}$ é fechada com relação a união infinita enumerável).

Uma classe $\mathcal{A}$ de subconjuntos de um conjunto não vazio Ω é tal que o espaço amostral é um elemento dessa classe, todo evento na classe tem o seu complementar nessa classe e, além disso, a união infinita enumerável de eventos pertencentes a classe $\mathcal{A}$ é um elemento da classe $\mathcal{A}$. Uma tal classe é chamada de σ-*álgebra* de subconjuntos de Ω.

A partir dessas três propriedades da σ-*álgebra* podemos concluir outras bastante relevantes. Por exemplo, podemos concluir de A1 e A2 que o conjunto vazio $\emptyset$ é um elemento de $\mathcal{A}$ e de A2 e das contribuições de De Morgan para teoria dos conjuntos pode-se concluir que a σ-*álgebra* $\mathcal{A}$ é também fechada com relação a interseções infinitas enumeráveis.

Como mencionamos em seções anteriores, há diferentes formas de se interpretar e de se definir uma probabilidade. Porém, nesta seção, apenas assumiremos que medidas deprobabilidade são atribuídas a todos os membros de uma σ-*álgebra*.

Seja $A \in \mathcal{A}$ um evento, associaremos a ele um número real $P(A)$, chamado *probabilidade* de A que satisfaz os seguintes axiomas:

1. A probabilidade é sempre não negativa, isto é, $P(A) \geq 0$;
2. A probabilidade aplicada ao espaço amostral é sempre igual a 1, em notação matemática, isto equivale a escrever $P(\Omega) = 1$;
3. Se $A_1, A_2, \dots \in \mathcal{A}$ são disjuntos, ou seja, a interseção entre quaisquer dois conjuntos distintos é $\emptyset$, então $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$ (axioma conhecido

como σ-aditividade). Em outras palavras, a probabilidade da união infinita enumerável de eventos é igual a soma infinita enumerável das probabilidades dos eventos.

Pode-se mostrar que a propriedade de σ-aditividade vale também para coleções finitas de eventos, bastando para isso construir sequências infinitas de conjuntos tais que $A_i = \emptyset$ para *i* = *n*+1, *n*+2, ... e usar o fato de que $P(A_{n+1}) = P(A_{n+2}) = \cdots = 0$.

Uma função P definida numa σ-álgebra $\mathcal{A}$ e satisfazendo os axiomas listados acima é denominada uma *probabilidade* em $\mathcal{A}$.

Dessa forma, se desejamos verificar se uma determinada maneira de medir um conjunto é uma probabilidade, como por exemplo, as discutidas em 4.1.1, 4.1.2 e 4.1.3, basta verificarmos se os três axiomas são satisfeitos.

Como consequência desses axiomas, podemos deduzir uma sequência de propriedades importantes para o cálculo das probabilidades. Estas propriedades podem ser encontradas em James (2011, p. 12-13).

Com isso podemos concluir que um modelo probabilístico é constituído de:

1. Ω, tal que $\Omega \neq \emptyset$;
2. De uma *σ-álgebra* $\mathcal{A}$ de subconjuntos de Ω; e
3. De uma probabilidade P em $\mathcal{A}$.

A terna $(\Omega, \mathcal{A}, P)$ é denominada *espaço de probabilidade*.

Esses são os alicerces que fundamentam a Teoria das Probabilidades sobre as bases sólidas da Matemática, concedendo-lhe o mesmo grau de rigor que possui a Geometria ou a Álgebra.

4.2 FOCO: FREQUENTISTA

Alguns conceitos fundamentais para este trabalho vinham sendo explorados ou de forma intuitiva ou admitindo-se que o leitor os soubesse, sem que isso prejudicasse a leitura. Contudo, a partir deste ponto, ter em mente a definição precisa desses conceitos é crucial para o entendimento de tudo que será discutido adiante.

Variável aleatória: uma variável aleatória $X(\omega)$ é uma função mensurável que associa cada elemento ω do espaço amostral Ω a um valor real.

População: o conjunto formado por todos os elementos que possuem um atributo em um estudo, em outras palavras é o universo de pesquisa de um estudo. O conceito de população em Estatística tem paridade com o conceito de espaço amostral na Teoria das Probabilidades, daí o fato de funções de probabilidade e funções de densidade de probabilidade serem encaradas como populações na modelagem probabilística.

Amostra aleatória: qualquer subconjunto formado por elementos escolhidos aleatoriamente de uma população. Na condição de observação parcial da população, a amostra induzirá incertezas sobre quantidades obtidas através dela e ensejará a modelização probabilística de funções de resultados amostrais.

Parâmetro: qualquer função dos elementos populacionais, ou seja, qualquer descrição numérica de uma característica da população, por exemplo, sua média, variância, proporção, etc. No contexto da Inferência Frequentista, os parâmetros são vistos como quantidades desconhecidas e não aleatórias. No contexto Bayesiano, os parâmetros serão também vistos como variáveis aleatórias passíveis, portanto, de modelagem probabilística *a priori* dada pelo especialista.

Estimador: qualquer função dos elementos da amostra; como, por exemplo, a média amostral, a proporção amostral, etc. Os estimadores são, portanto, vistos dentro da Estatística como variáveis aleatórias, cujas leis de probabilidade devem ser construídas a partir de hipóteses tecidas sobre a população alvo.

Estimativa: o valor numérico obtido do estimador, quando obtida uma amostra específica.

Vimos na seção anterior que existem diversas formas de interpretar a probabilidade, e procuramos ressaltar a importância das diferentes abordagens para a compreensão tanto da probabilidade quanto da estatística. Porém, neste trabalho, nos restringiremos à abordagem frequentista de probabilidade e através dela interpretaremos os Teoremas Limite e sua importância para a promoção do raciocínio estatístico.

Como dissemos na seção anterior, a interpretação frequentista tem como característica a repetibilidade de um experimento para que se possa estimar a chance de ocorrência de determinados eventos. O coração, por assim dizer, da interpretação frequentista está na estabilidade empírica das frequências relativas da ocorrência dos eventos, sendo essa estabilidade alcançada após numerosas repetições do experimento. A garantia dessa estabilidade pode ser explicada pela Lei dos Grandes Números. Assim, pode ser observado que, à medida que o número de realizações experimentais aumenta, menor tende a ser a variação da frequência relativa encontrada, ao ponto de se estabilizar em um valor próximo da probabilidade real.

Como esta abordagem pode ser útil à compreensão da Estatística?

Essencialmente, as probabilidades serão utilizadas como elementos para estabelecer decisões ótimas ou auxiliar na construção de inferências a respeito de parâmetros de uma população. Nesse modelo de inferência, conhecido como Inferência Clássica ou Frequentista, o valor do parâmetro é, geralmente, desconhecido e fixo, cuja estimação (sujeita a erros) é construída através das propriedades probabilísticas dos estimadores propostos para ele.

À guisa de exemplo sobre como se dá de forma emblemática a interação entre os conceitos chave mencionados anteriormente, suponha uma população de apenas quatro artefatos explosivos, e sejam $X_1 = 100\,m$, $X_2 = 110\,m$, $X_3 = 140\,m$ e $X_4 = 105\,m$ os tamanhos do raio de atuação de cada explosivo, respectivamente. Neste caso, o parâmetro dado pela média populacional é dado por $\mu = 113{,}75\,m$. Porém, ao invés de testarmos cada um dos quatro artefatos, testaremos apenas dois. Para obter uma estimativa do valor de μ, baseado numa amostra de tamanho 2, calcularemos a média dos valores observados da amostra. Note que existem, ao todo, seis amostras distintas possíveis (sem ordem). A fim de ilustrar o processo colocaremos em uma tabela todas as seis médias possíveis.

Elemento 1	Elemento 2	Média
100	110	105,0
100	140	120,0

100	105	102,5
110	140	125,0
110	105	107,5
140	105	122,5

Tabela 1: Médias

Podemos observar que nenhum dos valores apresentados neste exemplo é igual ao valor do parâmetro μ, e todos possuem algum grau de erro. Além disso, as diferentes estimativas para a média populacional são realizações da variável aleatória dada pela média amostral, e como tal, faz jus a um modelo de probabilidade (nesse caso, cada média amostral tem probabilidade de 1/6, pois não há valores repetidos e cada amostra tem probabilidade de 1/6 de ser tomada).

Assim, se obtivermos a esperança matemática do estimador, ou seja, o valor médio das médias amostrais, verificaremos que este equivale ao valor do parâmetro μ. Na linguagem da Teoria da Estimação da Estatística, isso equivale a dizer que o estimador proposto para o parâmetro de interesse é não tendencioso, ou não viesado, ou não viciado. Essencialmente, isto indica que a média amostral é um bom estimador para a média populacional, no sentido de que, em média, a média amostral coincide com o parâmetro.

Outro aspecto importante a se ressaltar é que quanto maior for o tamanho amostral, menor será a variabilidade das estimativas possíveis do parâmetro. Assim se a amostra fosse de tamanho 3 nesse exemplo, verificaríamos que a média de todas as estimativas continuaria a coincidir com o parâmetro, mas os possíveis resultados apresentariam menor variabilidade do que os obtidos quando a amostra tinha tamanho 2. No jargão da Estatística, isto significa dizer que o estimador proposto é não somente não tendencioso, mas também consistente. Esse fato importante também nos garantirá, como veremos posteriormente nesse trabalho, que a média amostral se aproximará de uma variável normal conforme o tamanho da amostra aumente, independentemente de qual seja a distribuição da população alvo.

Esse pequeno exemplo, que pode ser trabalhado em sala de aula, ilustra de forma bastante didática a problematização central da Estatística, quer a população seja finita de qualquer tamanho ou mesmo infinita, ou seja, a de se estimar quantidades desconhecidas e não acessíveis da população por meio de observações parciais desta (amostras). Dado o caráter de incerteza vindo de uma única amostra obtida da população, o que se deseja, em última análise, é estimar o parâmetro μ, com certo grau de confiabilidade, ou seja, estabelecer um intervalo I_0 que contivesse o parâmetro μ com probabilidade $1-\alpha$ (α representando nossa probabilidade de erro na estimativa, denominado na Estatística de nível de significância). Isto é importante porque cada amostra gerará um intervalo de confiança diferente e este pode ou não conter o parâmetro de interesse. Contudo, a probabilidade de que qualquer um dos possíveis intervalos formados contenha o parâmetro é $1-\alpha$.

Dito de outra forma, se estimarmos uma grande quantidade de intervalos de confiança, todos seguindo o mesmo método, a proporção dos intervalos que conterá o parâmetro é $1-\alpha$.

Nesse pequeno exemplo, muitos conceitos importantes para o raciocínio estatístico foram utilizados direta ou indiretamente, tais como:

a) A Lei dos Grandes Números que diz que, com um alto valor de probabilidade,a média amostral pertencerá a um intervalo arbitrariamente pequeno centrado em torno do parâmetro μ, conforme o tamanho amostral cresce. Além disso, garante, como veremos posteriormente, que a média amostral convergirá para a média populacional com probabilidade 1.
b) O Teorema Central do Limite que diz que a distribuição de probabilidade do estimador dado pela média amostral converge para a distribuição Normal, conforme o tamanho da amostra aumenta.

Esses dois teoremas são fundamentais para uma primeira compreensão da Teoria da Estimação, no que se refere à Inferência Frequentista. Naturalmente, a Teoria da Estimação é por demais ampla e rica para ser reduzida a apenas esses dois teoremas, objetos dessa dissertação, mas estamos convencidos de que o apelo intuitivo e empírico desses dois teoremas os torna fundamentais para a construção do raciocínio estatístico.

Essas teorias e conceitos auxiliam no processo de tomada de decisão, porém, tomar decisão não se trata de apenas realizar cálculos, mas também de interpretar o significado dos resultados obtidos. A abordagem frequentista nos ajuda a dar significado a nossos cálculos quando for razoável supor que o fenômeno possa ser reproduzido, em especial no auxílio aos testes de hipóteses sobre determinados fenômenos aleatórios. Por exemplo, suponha que uma empresa tenha realizado melhorias em suas máquinas de perfuração a fim de aumentar, em média, a profundidade alcançada. O valor médio da profundidade alcançada antes das melhorias era de 120 metros com um desvio padrão de 6 metros. Considerando que o desvio padrão continua o mesmo após ter sido realizada a melhoria das máquinas, um agente responsável por supervisionar a qualidade das máquinas inspecionou 20 máquinas e constatou que em média as máquinas dessa amostra perfuravam uma profundidade de 122 metros. Nessas circunstâncias seria razoável supor que houve melhoria?

A partir da hipótese inicial de que não houve alteração na média o agente precisa estipular um nível de significância para a verificação, digamos 0,05. Isto é, se este procedimento de verificação for utilizado diversas vezes o agente rejeitará, equivocadamente, essa hipótese em 5% das vezes.

Pelo Teorema Central do Limite, sabemos que a distribuição da média pode ser aproximada por uma distribuição Normal. A ideia fundamental do teste de hipótese é avaliar se a média amostral de 122 pode ser considerada como superior à média de 120 (antes da melhoria) por cair numa região de probabilidade inferior ou igual a 5% da cauda superior da distribuição Normal. Ao tomarmos essa decisão, estamos dizendo que preferimos acreditar que a média populacional tenha aumentado após a intervenção, pelo fato de a média amostral ter caído numa região de baixa probabilidade, quando estabelecemos que a média populacional permanece como 120, daí o nível de significância de 5% nos informar que se trata da probabilidade de se rejeitar a hipótese de não melhoria, quando de fato não houve melhoria após a intervenção.

A possibilidade de reproduzir o experimento, embora, algumas vezes possa parecer inadequada, pode ajudar na compreensão do porque é preferível tomar uma decisão à outra. Isso será visto, posteriormente, nesse trabalho, na seção 6.2, quando nos debruçarmos sobre o clássico problema de Monty Hall (exemplo 2) e diagnosticarmos que é preferível optar pela estratégia de troca de porta a manter a escolha inicial, após o apresentador abrir uma das duas portas não escolhidas.

Embora seja estranho pensar na repetibilidade desse experimento, uma vez que só teremos uma realização para a escolha da porta, a escolha da estratégia ótima depende de se pensar que, se esse experimento fosse realizado um grande número de vezes e em todas essas vezes trocássemos de porta, ganharíamos o prêmio cerca de $\frac{2}{3}$ das vezes, enquanto que se mantivéssemos a escolha inicial da porta ganharíamos o prêmio apenas em $\frac{1}{3}$ das vezes. Dessa forma seria mais sensato trocar de porta, mesmo que tenhamos uma única oportunidade de realização do jogo.

Em resumo, nos concentramos nesta seção em revelar as ideias centrais da Inferência Frequentista à luz dos teoremas de que trataremos no próximo capítulo, bem como a importância de uma abordagem concreta destes por experimentações simples no Ensino Básico.

Mais uma vez ressaltamos que estamos cientes de que a abordagem frequentista de probabilidade não é suficiente para tratar todos os aspectos da Estatística. Entretanto, ela dá inteligibilidade a uma grande quantidade de conceitos e teoremas, bem como às respostas encontradas através dos cálculos.

5 OS TEOREMAS LIMITE

O objetivo desse capítulo é discutir dois dos Teoremas Limite que fornecem grande contribuição à Estatística e à Matemática. A importância de ambos os teoremas que trataremos aqui é notória para a construção de um raciocínio estatístico, pois, são eles que orientam sobre como conduzir afirmações e tomadas de decisão dentro da Inferência Frequentista ou Clássica. Já mencionamos anteriormente que um dos principais objetivos da Estatística é a obtenção de informações sobre parâmetros populacionais à luz de amostras cuja parcialidade induz a erros e incertezas nas afirmações.

Os dois teoremas, objetos de nosso estudo (a Lei dos Grandes Números e o Teorema Central do Limite), são resultados poderosos no processo de estimação, na tomada de decisão e até na solução de problemas de Análise Matemática no cálculo de séries e integrais delicadas.

Não pretendemos aqui esgotar todo o desdobramento desses dois teoremas, uma vez que eles são ferramentas bastante sofisticadas que, quase sempre, exigem o conhecimento de teorias que não são tratadas em cursos de graduação.

O que faremos é explorar, em detalhes, as nuances da Lei Forte dos Grandes Números e do Teorema Central do Limite, em suas respectivas versões que são comumente tratadas na graduação, e naturalmente faremos uma breve extensão dessas versões, a fim de ampliar aquilo que já é sabido, dos cursos de graduação, sobre eles.

Ao explorar tais nuances estaremos tratando de duas questões tidas como fundamentais para Shulman, que são os saberes de Conteúdo e Pedagógico de Conteúdo. Por um lado, quando analisamos o valor das hipóteses para os teoremas, bem como suas fundamentações, contribuímos para a contrução do primeiro; já quando tratarmos das formas pelas quais as ideias desses teoremas podem ser compreendidas, estaremos contribuindo para a construção do último.

Contudo, antes de iniciarmos um tratamento tanto intuitivo quanto formal desses teoremas, é importante compreendermos alguns conceitos e teoremas que são os alicerces destes.

Apresentaremos inicialmente três importantes definições sobre a convergência de variáveis aleatórias em um espaço de probabilidade, a saber, convergência em probabilidade, convergência quase certa e convergência em distribuição. O primeiro tipo de convergência é a base sobre a qual repousa a Lei Fraca dos Grandes Números, já o segundo e o terceiro tipos de convergência fundamentam a Lei Forte dos Grandes Números e o Teorema Central do Limite, respectivamente.

Em seguida trataremos de um teorema que relaciona os tipos de convergências tratados até aqui.

Convergência em Probabilidade: Seja $\{X_n\}_{n\geq 1}$ uma sequência de variáveis aleatórias e X uma variável aleatória definidas sobre um mesmo espaço de probabilidade $(\Omega, \mathcal{A}, P)$.

Dizemos que X_n *converge em probabilidade* para X se

$$\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0, \forall \epsilon > 0,$$

ou, equivalentemente,

$$\lim_{n\to\infty} P(|X_n - X| \leq \epsilon) = 1, \forall \epsilon > 0.$$

Adotaremos a seguinte notação para este caso: $X_n \xrightarrow{P} X$.

Desta forma, o que a convergência fraca afirma é que a sequência a_n = $P(|X_n - X| \leq \epsilon)$ de números reais em [0,1] converge para 1 para todo $\epsilon > 0$. Nada é dito aqui sobre a convergência pontual de $X_n\ (\omega)$ para $X\ (\omega)$.

Convergência Quase Certa: Seja $\{X_n\}_{n\geq 1}$ uma sequência de variáveis aleatórias e X uma variável aleatória definidas sobre um mesmo espaço de probabilidade $(\Omega, \mathcal{A}, P)$.

Dizemos que X_n*converge quase certamente* para X se

$$P\left(\lim_{n\to\infty} X_n = X\right) = 1$$

e denotamos por $X_n \xrightarrow{q.c} X$.

A convergência quase certa nos afirma que X_n irá convergir para X para quase todos os pontos $\omega \in \Omega$, de modo que o conjunto formado pelos elementos do espaço amostral em que a convergência pontual não ocorre tem medida nula. A convergência quase certa tem, portanto, paridade com a convergência pontual exceto em um conjunto de medida nula.

A fim de auxiliar na compreensão dos conceitos de convergência em probabilidade e de convergência quase certa apresentaremos alguns exemplos. Porém, será antes necessário apresentar dois resultados auxiliares para a averiguação da convergência forte (quase certa), a saber,o critério para a convergência quase certa e o Lema de Borel-Cantelli.

Critério para a convergência quase certa: Seja $\{X_n\}_{n\geq 1}$ uma sequência de variáveis aleatórias e X uma variável aleatória. Se $P(|X_n - X| > \epsilon \; infinitas \; vezes) = 0$, para todo $\epsilon > 0$, então $X_n \xrightarrow{q.c} X$.

Em outras palavras, se a probabilidade de que X_n se afaste de X, por mais do que ϵ unidades infinitamente vezes é 0, então isso equivale a dizer que X_n converge quase certamente para X.

Demonstração: Vide Apêndice A.

Lema de Borel-Cantelli: Sejam $A_1, A_2, \ldots$ eventos aleatórios em um espaço de probabilidade $(\Omega, \mathcal{A}, P)$.

(i) Se $\sum_{n=1}^{\infty} P(A_n) < \infty$, então $P(A_n \; infinitas \; vezes) = 0$;

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

O termo $A_n \; infinitas \; vezes$, representando a ocorrência de uma infinidade dos eventos A_n, tem equivalência matemática do $\limsup A_n := \bigcap_{n=1}^{\infty} \bigcup_{n=k}^{\infty} A_k$.

Demonstração:

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

$$[A_n\ infinitas\ vezes] \subset \bigcup_{k=n}^{\infty} A_k\ \ \forall n,$$

logo

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

Portanto, $P(A_n\ infinitas\ vezes) = 0$.

(ii) Lembre-se que $[A_n\ infinitas\ vezes] = \bigcap_{n=1}^{\infty} \bigcup_{n=k}^{\infty} A_k$ e como a interseção de um número enumerável de eventos de probabilidade 1, é também de probabilidade 1. Então, basta provar que

$$P\left(\bigcup_{k=n}^{\infty} A_k\right) = 1\ \forall n$$

Para tal, seja $B_n = \bigcup_{k=n}^{\infty} A_k$. Então B_n contém $\bigcup_{k=n}^{n+m} A_k\ \forall m$, e

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

Logo, $\forall m$ tem-se que

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

Pela independência dos A_n segue que

$$\prod_{k=n}^{n+m} P(A_k^c) = \prod_{k=n}^{n+m} (1 - P(A_k)).$$

Como $1 - p \leq e^{-p}$ para $0 \leq p \leq 1$, temos

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

Quando $m \to \infty$, pois $\sum_{k=n}^{n+m} P(A_k) \to +\infty$ quando $m \to \infty$.

Logo $P(B_n) = 1\ \forall n$.

■

A fim de ilustrar o que diz o lema acima considere o exemplo clássico a seguir: ao colocar um macaco diante de uma máquina de escrever, supondo que este possa apertar as teclas da máquina durante um tempo infinito, é razoável admitirmos que haja uma chance positiva, embora pequeníssima, de que ele digite as obras completas de Shakespeare sem erro.

Nesse experimento, chamamos de sucesso o ensaio no qual o macaco digita com perfeição as obras e de fracasso o ensaio em que ele não consegue realizar tal façanha. Supondo ainda que o macaco não se fadigue ou ocorra qualquer outra coisa que faça com que os ensaios não sejam independentes. Então, pelo Lema de Borel-Cantelli a probabilidade de que o macaco seja capaz de datilografar as obras de Shakespeare infinitas vezes é 1.

Retomemos agora os conceitos de convergência em probabilidade e de convergência quase certa como alguns exemplos que evidenciam as suas diferenças.

Para ambos os exemplos que forneceremos a seguir considere $\{X_n\}_{n\geq 1}$ uma sequência de variáveis aleatórias independentes.

Exemplo 1: Considere a seguinte função de probabilidade

$$P(X_n = k) = \begin{cases} 1 - \dfrac{1}{n^2}, & se\ k = 0 \\ \dfrac{1}{n^2}, & se\ k = 1 \end{cases}$$

para $n \geq 1$.

Mostraremos que $X_n \xrightarrow{P} 0$ e que $X_n \xrightarrow{q.c} 0$.

a) Convergência em probabilidade

$$P(|X_n - 0| > \epsilon) = P(|X_n| > \epsilon) = \begin{cases} P(X_n = 1) = \dfrac{1}{n^2}, & se\ 0 < \epsilon < 1 \\ 0 & , se\ \epsilon \geq 1 \end{cases}$$

daí segue que $\lim_{n\to\infty} P(|X_n - 0| > \epsilon) = \lim_{n\to\infty} \frac{1}{n^2} = 0$. Portanto, $X_n \xrightarrow{P} 0$.

b) Convergência quase certa

Se $\epsilon \geq 1$, $A_n = [|X_n - 0| > \epsilon] = 0$.

Se $0 < \epsilon < 1$, $A_n = [|X_n - 0| > \epsilon] = [X_n = 1]$.

Como$\sum_{n=1}^{\infty} P(A_n) \leq \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty$, pelo Lema de Borel-Cantelli segue-se que $P(A_n\ infinitas\ vezes) = 0$ e, pelo critério de convergência, temos que $X_n \xrightarrow{q.c} 0$.

Exemplo 2: Tome a seguinte função de probabilidade

$$P(X_n = k) = \begin{cases} 1 - \dfrac{1}{n}, & se\ k = 0 \\ \dfrac{1}{n}, & se\ k = 1 \end{cases}$$

para $n \geq 1$.

Diferentemente do exemplo anterior, temos que $X_n \xrightarrow{P} 0$, mas X_n não converge quase certamente para 0.

a) Convergência em probabilidade

$$P(|X_n - 0| > \epsilon) = P(|X_n| > \epsilon) = \begin{cases} P(X_n = 1) = \dfrac{1}{n}, & se\ 0 < \epsilon < 1 \\ 0 \qquad\qquad\qquad , & se\ \epsilon \geq 1 \end{cases}$$

Daí segue que $\lim_{n\to\infty} P(|X_n - 0| > \epsilon) = \lim_{n\to\infty} \frac{1}{n} = 0$. Portanto, $X_n \xrightarrow{P} 0$.

b) Convergência quase certa

Se $0 < \epsilon < 1$, $A_n = [|X_n - 0| > \epsilon] = [X_n = 1]$.

Como $\sum_{n=1}^{\infty} P(A_n) = \sum_{n=1}^{\infty} \frac{1}{n} = +\infty$, para $0 < \epsilon < 1$, temos, pela independência dos A_n e pelo Lema de Borel-Cantelli, que $P(A_n\, infinitas\, vezes) = 1$ e, portanto, pelo critério de convergência temos que X_n não converge quase certamente a 0.

Além desses exemplos, poderíamos nos perguntar se existe uma sequência de variáveis aleatórias convergente quase certamente para um valor X, porém não convergente em probabilidade para este mesmo valor. A resposta a esta pergunta é não. De fato, toda sequência de variáveis aleatórias convergente quase certamente é também convergente em probabilidade. A demonstração desta afirmação será vista ainda nesta seção.

Convergência em Distribuição: Sejam $X, X_1, X_2, \ldots$ variáveis aleatórias com, respectivamente, função de distribuição $F, F_1, F_2, \ldots$. Dizemos que X_n converge em distribuição para X quando $n \to \infty$, se $F_n(x)$ converge para $F(x)$ em todo x ponto de continuidade de F, e denotamos por $X_n \xrightarrow{D} X$.

Atente para o fato de que este é um tipo de convergência que não se refere a um ponto, tal como ocorre com a convergência quase certa. Lembre-se dos capítulos anteriores quando tratamos da história da Estatística e da Probabilidade. Em certo momento, começou-se a busca por métodos que fornecessem aproximações para a distribuição Binomial.

O Método proposto por De Moivre, que tempos depois veio a ser chamado de versão do Teorema Central do Limite de De Moivre e Laplace, trata de estipular condições sob as quais a distribuição Binomial se aproxima da distribuição Normal, ou, dito de outra forma, trata de estipular condições sob as quais a distribuição Binomial converge para a distribuição Normal.

Forneceremos a seguir um exemplo para tentar ilustrar a diferença entre os tipos de convergência, bem como para tentar ressaltar a importância de definir a convergência apenas em pontos de continuidade.

Exemplo3: Seja $\{X_n\}_{n\geq 1}$ uma sequência de variáveis aleatórias, definidas em um espaço da probabilidade qualquer, tal que $X_n(\omega) = \frac{1}{n}, \forall\omega \in \Omega$ e seja $X = 0$.

Das convergências tratadas na seção anterior, vimos que X_n não converge quase certamente à X, porém, podemos afirmar que $X_n \xrightarrow{P} X$. Considere as funções de distribuição de X_n e de X, respectivamente

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

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

Portanto, $x > 0$ implica que $F_n(x) \to F(x) = 1$ e $x < 0$ implica que $F_n(x) \to F(x) = 0$, desta forma, $F_n(x) \to F(x)$ nos pontos de continuidade. Embora em $x = 0$, $F_n(0) \to 0$ e $F(0) = 1$, temos que a convergência em distribuição se verifica, pois $x = 0$ é ponto de descontinuidade.

Antes de prosseguir, introduziremos aqui um teorema que relaciona os três tipos de convergência enunciados.

Teorema: Seja $\{X_n\}_{n\geq 1}$ uma sequência de variáveis aleatórias e X uma variável aleatória, então tem-se que

a) se $X_n \xrightarrow{q.c} X$, então $X_n \xrightarrow{p} X$;

b) se $X_n \xrightarrow{p} X$, então $X_n \xrightarrow{D} X$.

Demonstração:

a) Suponha que $X_n \xrightarrow{q.c} X$ e seja $\epsilon > 0$. Devemos provar que $P(|X_n - X| \geq \epsilon) \to 0$.

Seja $A_0 = \{\omega: X_n(\omega) \to X(\omega)\}$. Por hipótese, $P(A_0) = 1$.

$\forall \omega \in A_0, |X_n(\omega) \to X(\omega)| < \epsilon$, para todo n suficientemente grande. Seja A_n o evento "$\forall k \geq n, |X_k - X| < \epsilon$", isto é,

$$A_n = \bigcap_{k=n}^{\infty} [|X_n - X| < \epsilon]$$

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

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

Portanto,

$$1 = P(A_0) \leq P\left(\bigcup_{n=1}^{\infty} A_n\right)$$

E, pela continuidade da medida de probabilidade,$P(A_n) \to 1$, de forma crescente.

Mas $A_n \subset [|X_n - X| < \epsilon]$, logo $P(|X_n - X| < \epsilon) \to 1$ e

$$P(|X_n - X| \geq \epsilon) = 1 - P(|X_n - X| < \epsilon) \to 0.$$

■

b) Seja x um ponto de continuidade de F_X e, por hipótese, temos que $X_n \xrightarrow{p} X$. Precisamos provar que $F_{X_n}(x) \to F_X(x)$ quando $n \to \infty$.

Como para $\epsilon > 0$, $X_n \leq x \Longrightarrow X \leq x + \epsilon$ ou $X - X_n > \epsilon$, temos $[X_n \leq x] \subset [X \leq x + \epsilon] \cup [|X_n - X| > \epsilon]$. Logo,

$$F_{X_n}(x) = P(X_n \leq x) \leq F_X(x + \epsilon) + P(|X_n - X| > \epsilon)$$

Por outro lado, $X \leq x - \epsilon \Longrightarrow X_n \leq x$ ou $X_n - X > \epsilon$, de modo que

$$F_X(x - \epsilon) \leq F_{X_n}(x) + P(|X_n - X| > \epsilon)$$

Fazendo inicialmente $n \to \infty$ e depois $\epsilon \to 0$, temos primeiro

$$F_X(x - \epsilon) \leq \lim_{n\to\infty} \inf F_{X_n}(x) \leq \lim_{n\to\infty} \sup F_{X_n}(x) \leq F_X(x + \epsilon)$$

Pois, $X_n \xrightarrow{p} X$. E, portanto, $F_X(x) = \lim_{n\to\infty} F_{X_n}(x)$, onde x é um ponto de continuidade, por hipótese.

■

Embora não tenhamos mostrado aqui, deixamos de alerta ao leitor que as recíprocas dos itens (a) e (b) não valem de modo geral. Considerando isto podemos afirmar que dos tipos de convergência apresentadas, a convergência quase certa é a “mais forte” delas e por outro lado, a convergência em distribuição é a “mais fraca”.

5.1 LEI DOS GRANDES NÚMEROS

A teoria das probabilidades possui interesse na busca por certos tipos de regularidades em fenômenos regidos pela incerteza. Essas regularidades, em uma abordagem frequentista, podem ser traduzidas em termos de frequência relativa da ocorrência de um evento de interesse.

Na busca dessas regularidades, quando nos deparamos com eventos cujas probabilidades de ocorrência não sejam nem muito altas nem muito baixas, não ficamos surpresos com o resultado e tendemos a aceitar, com certa facilidade, a ocorrência ou não daquele evento.

Por outro lado, se as probabilidades de um determinado evento ocorrer forem altíssimas, próximas de 1, nos inclinamos a aceitar que aquele evento ocorrerá recorrentemente, independentemente do número realizações do experimento.

De igual maneira, quando as chances são ínfimas, muito próximas de 0, admitimos que tal evento dificilmente vá ocorrer.

Contudo, um estudo das probabilidades sobre os eventos que possuem frequências relativas muito próximas a 1 ou muito próximas a 0 pode ser bastante revelador.

A fim de ilustrar o que foi dito, considere que uma aposta vencedora, em uma certa loteria, tem probabilidade p de ocorrer (p é um valor muito pequeno). Suponha que exista uma pessoa que sempre joga nessa loteria. A probabilidade p^n de que essa pessoa faça n apostas vencedoras consecutivamente é um evento raríssimo de ocorrer de tal modo que é natural acreditarmos na sua improbabilidade.

Contudo, do Lema de Borel-Cantelli decorre que, com probabilidade 1, esse evento ocorrerá e além disso, ele garante que com probabilidade 1 esse evento ocorrerá infinitas vezes, pois $\sum_{k=1}^{\infty} p^n = \infty$ e os eventos são independentes (considerando o evento sucesso se a pessoa faz *n* apostas consecutivas vencedoras e fracasso caso contrário). Isto é, no campo platônico das ideias, se existisse uma tal pessoa que sempre jogasse na loteria eternamente, poderíamos dizer que a probabilidade dessa pessoa realizar n apostas vencedoras infinitas vezes é 1.

É claro que, para este exemplo, estamos admitindo uma série de hipóteses adicionais, tais como, a loteria nunca deixará de existir, apostas sempre serão permitidas, etc.

Em suma, esse exemplo nos mostra que mesmo um evento raríssimo, tais como uma mulher dar origem a óctuplos, ocorrerá, e, mais do que isso, a probabilidade de que esse tipo de evento ocorra infinitas vezes é 1.

Porém, isso só é possível se for razoável admitirmos algumas hipóteses que satisfaçam às condições para utilização do Lema de Borel-Cantelli, tais como a probabilidade de dar origem a óctuplos ser um valor ou uma expressão p tal que $\sum_{n=1}^{\infty} p = \infty$ e os eventos serem independentes.

O Lema de Borel-Cantelli é sobretudo importante para a aferição da Lei Forte dos Grandes Números, como já mencionamos. Porém, antes de tratarmos especificamente da Lei Forte, que é o nosso principal objeto de estudo nesta seção, abordaremos a Lei Fraca. Em ambos os casos, queremos analisar o comportamento de uma sequência de variáveis aleatórias X_n quando n tende o infinito.

Tanto a Lei Fraca quanto a Lei Forte, nos auxiliam a realizar estimativas pontuais a respeito de um parâmetro de uma dada população. Aqui discutiremos sobre os teoremas em si, deixando para o capítulo seguinte a discussão de como a Lei Forte pode nos auxiliar no processo de tomada de decisão sobre incerteza.

A Lei Fraca dos Grandes Números, como ficou conhecida, surge pela primeira vez entre o final do século XVII e início do século XVIII, sendo publicada em 1713 por Bernoulli em seu livro *Ars Conjectandi.* O seu enunciado pode ser entendido, em termos atuais, da seguinte forma:

Lei Fraca dos Grandes Números (Bernoulli): Seja X_n uma sequência de ensaios binomiais independentes, com probabilidade p de "sucesso" em cada ensaio. Seja S_n o número de sucessos nos n primeiros ensaios, então

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

Demonstração: Em JAMES (2011, p.198).

Uma outra forma de enunciar a Lei Fraca dos Grandes Números é: sejam n_o o número de ocorrências de um evento A em n ensaios independentes e p a probabilidade de ocorrência do evento A em cada um dos ensaios. Então, para todo$\epsilon > 0$,

$$\lim_{n\to\infty} P\left(\left|\frac{n_o}{n} - p\right| \leq \epsilon\right) = 1$$

Note que a segunda forma de enunciar a Lei Fraca é, na verdade, a aplicação direta do conceito de convergência em probabilidade discutido no início deste capítulo. Sendo, $X_n = \frac{n_0}{n}$ e $X = p$.

Após esta, várias outras versões da Lei dos Grandes Números foram desenvolvidas na tentativa de formular um conjunto de hipóteses que fossem necessárias e suficientes para a sua aplicação. Não nos estenderemos mais na Lei Fraca, pois como veremos a Lei Forte é a aplicação do conceito de convergência quase certa e, como já visto anteriormente, neste capitulo, a convergência quase certa implica a convergência em probabilidade.

Abordaremos nesta seção a versão da Lei Forte dos Grandes Números, que é comumente tratada nos cursos de graduação e procuraremos explicitar sua importância. Para tal, faremos uso de teoremas, cujas demonstrações serão omitidas, oferecendo, no entanto, as devidas referências onde estas podem ser encontradas.

Na primeira década do século XX, Émile Borel, propõe uma versão da Lei Forte dos Grandes Números que pode ser enunciada da seguinte maneira:

Seja $\{X_n\}_{n\geq 1}$ uma sequência de variáveis aleatórias independentes e identicamente distribuídas tal que $P(X_n = 1) = p$ e $P(X_n = 0) = 1 - p$. Então,

$$\frac{S_n}{n} \xrightarrow{q.c} p$$

onde, $S_n = X_1 + X_2 + \cdots X_n$.

Essa versão da Lei dos Grandes Números nos diz que, à medida que o número de experimentos aumenta, a frequência relativa empírica do evento tende, em certo sentido probabilístico, para a medida de probabilidade teórica do evento em estudo, legitimando assim a própria abordagem frequentista de probabilidade.

Essencialmente, a única diferença desta versão para aquela que foi proposta por Bernoulli é o tipo de convergência adotada. Se fizermos $X_n = \frac{S_n}{n}$ e $X = p$, poderemos, desde que as condições para a aplicabilidade do teorema sejam atendidas, reescrever essa versão da Lei Forte, tal como definimos a convergência quase certa $P\left(\lim_{n\to\infty} \frac{S_n}{n} = p\right) = 1$.

Uma outra formulação é estabelecida por Kolmogorov em 1933 e é sobre esta que repousa a Lei Forte do Grandes Números, atualmente adotada nos cursos de graduação. A seguir, enunciaremos a Lei, conforme descrita e demonstrada em James (2011, p.212).

Lei Forte dos Grandes Números (Kolmogorov): Sejam $X_1, X_2, \dots$ variáveis aleatórias independentes, identicamente distribuídas e integráveis, com $E(X_n) = \mu$. Então,

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

Demonstração: Vide Apêndice B.

Uma das modificações para esta versão é que ela diz respeito à esperança matemática da ocorrência dos eventos. Contudo, existem versões da Lei dos Grandes Números anteriores às contribuições de Kolmogorov, que convergem para a esperança.

É importante ressaltar que a convergência quase certa, neste caso, está vinculada às hipóteses de independência, distribuição idêntica e integrabilidade das variáveis aleatórias.

Tal como a versão da Lei dos Grandes Números de Borel, esta também diz respeito ao comportamento da média dos resultados experimentais quando n cresce. Porém, nesta, está contemplada a importante contribuição de Huygens, a Esperança Matemática, valor para o qual a média dos resultados experimentais converge quase certamente.

Um método de se verificar a validade desse resultado experimentalmente, até mesmo em sala de aula, seria executar muitas vezes o experimento: lançar um dado cúbico e registrar o valor da face que está voltada para cima.

Se a cada experimento for calculado o novo valor médio dos resultados obtidos de todos os experimentos já realizados, então para um grande número de repetições, se verificará que a média dos resultados experimentais obtidos se aproxima da esperança matemática da variável aleatória representando a face superior do dado, ou seja, do valor de 3,5.

Há outras formas de enunciar a Lei dos Grandes Números sem supor, por exemplo, que as variáveis aleatórias sejam identicamente distribuídas. Porém, será necessário que a seguinte condição seja satisfeita:

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

onde $V(X_n)$ representa a variância.

A condição apresentada é conhecida como a primeira Lei Forte dos Grandes Números ou também como a condição de Kolmogorov, e a demonstração da validade de tal condição pode ser encontrada em James (2011, p.208). Ela é suficiente para que uma sequência de variáveis aleatórias independentes e integráveis satisfaçam a Lei Forte dos Grandes Números. Entretanto, ela não é necessária.

Com isso queremos dizer que toda sequência de variáveis aleatórias independentes e integráveis que atenda a condição de Kolmogorov satisfará à Lei

Forte dos Grandes Números. Contudo, existem sequências de variáveis aleatórias que não atendem à condição de Kolmogorov e que satisfazem a esta Lei.

Mostraremos isto com um exemplo extraído de Stoyanov (1997):

Exemplo 4: Seja $\{X_n\}_{n\geq 1}$ uma sequência de variáveis aleatórias independentes tal que:

$$P(X_n = 1) = P(X_n = -1) = \frac{1}{2}\left(1 - \frac{1}{2^n}\right)$$
$$P(X_n = 2^n) = P(X_n = -2^n) = \frac{1}{2^{n+1}}$$

Daí temos que $E(X_n) = 0$ e $V(X_n) = 1 + 2^n - \left(\frac{1}{2^n}\right)$. Então $\frac{\sum_{n=1}^{\infty} V^2(X)}{n^2}$ diverge, e, portanto, a sequência não satisfaz à condição de Kolmogorov. Entretanto, mostraremos que ela satisfaz à Lei dos Grandes Números.

Antes de prosseguir com esse exemplo, necessitaremos da seguinte definição:

Sequências equivalentes no sentido de Khintchine: Duas sequências de variáveis aleatórias $\{Z_n\}_{n\geq 1}$ e $\{W_n\}_{n\geq 1}$ são ditas equivalentes, no sentido de Khintchine, se $\sum_{n=1}^{\infty} P(Z_n \neq W_n) < \infty$.

Assim, do lema de Borel-Cantelli segue que a probabilidade de que o evento $(Z_n \neq W_n)$ ocorra infinitas vezes é 0. Portanto, se $S_n = \frac{Z_1 + \cdots Z_n}{n}$ obedece a Lei Forte dos Grandes Números, então $S_n^* = \frac{W_1 + \cdots W_n}{n}$ também obedece. Assim, concluímos que ou ambas as sequências satisfazem a Lei forte dos Grandes Números ou ambas não satisfazem.

Voltando ao nosso exemplo, considere a sequência de variáveis aleatórias $\{Y_n\}_{n\geq 1}$, onde $P(Y_n = 1) = P(Y_n = -1) = \frac{1}{2}(1 - \frac{1}{2^n})$ e $P(Y_n = 0) = \frac{1}{2^n}$.

Note que $E(X_n) = E(Y_n) = 0$ e, além disso, $P(X_n \neq Y_n) = \frac{1}{2^n}, \forall n \geq 1$. Desta forma, a série $\sum_{n=1}^{\infty} P(X_n \neq Y_n) = \sum_{n=1}^{\infty} \frac{1}{2^n}$ é convergente. Portanto, $\{X_n\}$ e $\{Y_n\}$ são equivalentes no sentido de Khintchine. Assim, se mostrarmos que $\{Y_n\}$ satisfaz à Lei Forte dos Grandes Números, então $\{X_n\}$ também satisfará.

Como $V(Y_n) = 1 - \frac{1}{2^n}$, segue que $\frac{\sum_{n=1}^{\infty} V^2(Y)}{n^2} < \infty$. Logo, $\{Y_n\}$ satisfaz a Lei Forte dos Grandes Números e portanto $\{X_n\}$ também.

Isso mostra que a condição de Kolmogorov não é necessária para que uma sequência de variáveis aleatórias independentes e integráveis obedeçam à Lei Forte dos Grandes Números. Contudo, se a condição for atendida, então esta sequência certamente satisfará à Lei Forte. A demonstração desta última parte pode ser encontrada em JAMES (2011, p.208).

Voltando ao enunciado da Lei Forte dos Grandes Números que citamos anteriormente, podemos perguntar: será que a integrabilidade das variáveis aleatórias é uma condição necessária e suficiente para que todas as sequências formadas por variáveis aleatórias independentes e identicamente distribuídas atendam à Lei Forte dos Grandes Números?

Primeiro devemos entender o que significa dizer que X é integrável. Sabemos que a esperança matemática é definida como $E(X) = \int_{-\infty}^{+\infty} x dF(x)$, em que $F(x)$ é a função distribuição de X. Podemos reescrever a esperança matemática da seguinte forma:

$$E(X) = \int_{-\infty}^{0} x dF(x) + \int_{0}^{+\infty} x dF(x) = I + II$$

Temos que $I \leq 0$ e $II \geq 0$. Dessa forma temos que:

(i) Se I e II são finitas, então X é integrável;

(ii) Se I é finita, mas II não é, então $E(X) = +\infty$;

(iii) Se I não é finita, mas II é, então $E(X) = -\infty$;

(iv) Se I e II são ambas divergentes, então $E(X)$ não está bem definida.

Assim, uma variável aleatória X é dita integrável se, e somente se, $E(|X|) < \infty$, isto é, a esperança matemática é finita.

Demonstraremos agora que a esperança matemática ser finita é uma condição necessária e suficiente para que a Lei Forte dos Grandes Números seja satisfeita por uma sequência de variáveis aleatórias $\{X_n\}_{n\geq 1}$ independentes e identicamente distribuídas.

Da esperança ser finita segue que $\int_{-\infty}^{\infty} |x| dF_X(x) < \infty$.

Considere a seguinte sequência $Y_n = \begin{cases} X_n, & se\ |X_n| \leq n \\ 0, & se\ |X_n| > n \end{cases}$. Daí temos que

$$V(Y_n) = E(Y_n^2) - E^2(Y_n) \leq E(Y_n^2) = \int_{-n}^{n} x^2 dF_X(x) \leq \sum_{k=1}^{n} (k)^2 P(-k < |x| \leq k) =$$

$$= \sum_{k=0}^{n} (k+1)^2 P(1-k < |x| \leq k+1)$$

Além disso,

$$\sum_{n=1}^{\infty} \frac{V(Y_n)}{n^2} \leq \sum_{n=1}^{\infty} \sum_{k=0}^{n} \frac{(k+1)^2}{n^2} P(1-k < |x| \leq k+1) \leq$$

$$\leq \sum_{k=0}^{n} (k+1)^2\, P(1-k < |x| \leq k+1) \sum_{n \geq k}^{\infty} \frac{1}{n^2}$$

Como $\sum_{n \geq k}^{\infty} \frac{1}{n^2}$ converge e $\sum_{k=0}^{n} (k+1)^2\, P(1-k < |x| \leq k+1)$ converge pelo teste de comparação por limite e por $\int_{-\infty}^{\infty} |x| dF_X(x) < \infty$, concluímos que $\sum_{n=1}^{\infty} \frac{V(Y_n)}{n^2} < \infty$, ou seja, $\{Y_n\}_{n \geq 1}$ satisfaz a Lei Forte dos Grandes Números. Contudo, ainda resta mostrar que $\{X_n\}_{n \geq 1}$ também satisfaz. Para isso utilizaremos a equivalência de Khintchine, isto é, vamos mostrar que $\sum_{n=1}^{\infty} P(X_n \neq Y_n) < \infty$.

Note que

$$\sum_{n=1}^{\infty} P(X_n \neq Y_n) = \sum_{n=1}^{\infty} P(|X_n| > n) \leq \sum_{n=1}^{\infty} n\, P(n \leq |X_n| < n+1) \leq \sum_{n=1}^{\infty} n \int_{n \leq |x| < n+1} dF_X(x)$$

$$\leq \sum_{n=1}^{\infty} \int_{n \leq |x| < n+1} |x| dF_X(x) = \int_{|x| \geq n} |x| dF_X(x) < \infty$$

Portanto, $\{X_n\}_{n \geq 1}$ e $\{Y_n\}_{n \geq 1}$ são equivalentes no sentido de Khintchine, logo $\{X_n\}_{n \geq 1}$ satisfaz à Lei Forte dos Grandes Números.

Com isso mostramos que, com a hipótese da integrabilidade, toda sequência de variáveis aleatórias independentes e identicamente distribuídas satisfazem à Lei Forte dos Grandes Números.

Mesmo que a versão apresentada aqui da Lei Forte dos Grandes Números não possa ser aplicada a diversos fenômenos, esta continua sendo uma ferramenta poderosa que propicia o tratamento de uma grande quantidade de situações, bastando para isso supor, além da integrabilidade, que a distribuição não sofrerá modificação ao longo do trabalho.

5.2 TEOREMA CENTRAL DO LIMITE

Vimos na seção anterior que uma sequência de variáveis aleatórias, sob certas circunstâncias, obedece à Lei Forte dos Grandes Números, isto é, a média $\frac{S_n}{n}$ converge quase certamente para μ, a esperança matemática das variáveis aleatórias independentes e identicamente distribuídas.

Existe uma vasta quantidade de fenômenos aleatórios cujo o principal interesse está no comportamento da variável aleatória $S_n = X_1 + \cdots + X_n$. Contudo, analisar o comportamento dessas variáveis aleatórias, também vistas como convoluções das variáveis aleatórias X_i, *i* = 1, 2, ..., *n*, pode ser uma tarefa bastante complexa.

Nesse sentido, o TCL permite a construção de leis de probabilidade aproximadas para somas e médias de variáveis aleatórias, e contribui ao mesmo tempo para o raciocínio estatístico no momento da tomada de decisão.

Além de podermos obter as estimativas pontuais das quais comentamos brevemente na seção anterior, através da Lei Forte, poderemos obter estimativas intervalares, com o TCL, o que nos permitirá aferir mais a respeito do parâmetro de interesse, como a formação de intervalos de confiança.

Deixaremos a discussão da contribuição do TCL para uma tomada de decisão sob incerteza para o capítulo seguinte. Nesta seção mostraremos, sob certas condições, que as somas parciais normalizadas convergem em distribuição para a distribuição Normal padrão. Em termos matemáticos, significa garantir que

$$\frac{S_n - E(S_n)}{\sqrt{V(S_n)}} \xrightarrow{D} N(0,1)$$

Esse resultado, quando possível de ser aplicado, é poderoso, pois nos permite acessar de forma bem aproximada o comportamento da variável aleatória S_n, qualquer que seja a distribuição das variáveis aleatórias X_i, *i* = 1, 2, ..., *n.*

Há diversas versões do Teorema Central do Limite, aqui discutiremos apenas aquelas que normalmente são vistas em cursos de graduação para licenciandos de Matemática.

Apresentaremos agora a versão do Teorema Central do Limite para variáveis aleatórias independentes e identicamente distribuídas com média μ e variância σ^2, onde $0 < \sigma^2 < \infty$. Se as variáveis aleatórias atenderem a essas condições, então

$$\frac{S_n - E(S_n)}{\sqrt{V(S_n)}} \xrightarrow{D} N(0,1)$$

Recorde que $E(S_n) = E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n)$ e como as variáveis são identicamente distribuídas $E(X_n) = E(X_1) = \mu,\ \forall n \in \mathbb{N}$. Além disso, $V(S_n) = V(X_1) + V(X_2) + \cdots + V(X_n)$ e $V(X_n) = V(X_1) = \sigma^2$. Assim, podemos escrever

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{D} N(0,1)$$

Dentre outras coisas, o Teorema Central do Limite nos permite analisar o comportamento de somas e médias de variáveis aleatórias. Assim, sob certas hipóteses, o TCL nos diz que a distribuição da média amostral padronizada tende a se aproximar da curva normal padrão conforme n cresce.

Um experimento simples que pode auxiliar na compreensão dessa ideia é lançar dois dados cúbicos e anotar o valor médio dos resultados obtidos. A cada novo experimento calcula-se a frequência relativa com que cada média possível ocorreu e constrói-se um gráfico de barras exibindo as frequências relativas. Se o experimento for executado com mais dados, a média dos resultados terá um comportamento ainda mais próximo na normal. Conforme o número de dados cresce é possível notar que o gráfico gerado a partir das frequências relativas tende a se aproximar do gráfico da curva Normal, cada vez mais densamente.

Observamos que a Lei Forte dos Grandes Números nos informa que $\frac{S_n}{n} - \mu \xrightarrow{q.c} 0$. Por outro lado, o Teorema Central do Limite afirma que ao multiplicarmos o membro da direita por $\frac{\sqrt{n}}{\sigma}$ este resultado convergirá em distribuição para a Normal padrão, ou seja,

$$\frac{\sqrt{n}}{\sigma}\left(\frac{S_n}{n} - \mu\right) \xrightarrow{D} N(0,1)$$

Antes de demonstrar essa versão do Teorema Central do Limite será mister definir alguns elementos.

Sejam X e Y variáveis aleatórias reais definidas em um mesmo espaço de probabilidade. Dizemos que $Z = X + iY$ é variável aleatórias complexa. (Recorde que $i = \sqrt{-1}$.) Além disso, se $E(X)$ e $E(Y)$ são finitas, então $E(Z) = E(X) + iE(Y)$.

Como consequência desta definição temos, tendo em mente que $e^{ix} = \cos(x) + isen(x)$ e que ambas as funções trigonométricas, seno e cosseno, são limitadas, que

$$E\left(e^{itX}\right) = E(\cos(tX)) + iE\big(sen(tX)\big), \qquad t \in \mathbb{R}$$

Denotaremos por ϕ a função característica de uma variável aleatória X definida por $\phi: \mathbb{R} \to \mathbb{C}$, onde $\phi(t) = \phi_X(t) = E(e^{itX})$, ou, simplesmente

$$\phi(t) = E(\cos(tX)) + iE\big(sen(tX)\big), t \in \mathbb{R}.$$

Além disso, a função característica da distribuição Normal padrão é dada por

$$E\left(e^{itX}\right) = \int_{-\infty}^{\infty} e^{itx} \times \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2-2itx}{2}} dx = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \int_{-\infty}^{\infty} e^{-\frac{(x-it)^2}{2}} dx = e^{-\frac{t^2}{2}}$$

Omitimos aqui a explicação da última integração nos complexos. Contudo, esta pode ser vista em JAMES (2011, p.231).

A função característica, tal como a função de distribuição e a função de densidade, é uma forma pela qual podemos representar uma medida de probabilidade. Cada representação possui propriedades específicas que, em certos momentos, podem ser bastante úteis. Por exemplo, é geralmente mais fácil calcular os momentos de uma variável aleatória utilizando a função característica, pois estes são calculados pelo processo de diferenciação da função característica. Além disso, também são mais simples os cálculos para encontrar a distribuição da soma de variáveis aleatórias independentes. Por fim, é ela que nos auxiliará na demonstração do Teorema Central do Limite para variáveis aleatórias independentes e identicamente distribuídas.

Apenas para facilitar a leitura enunciaremos novamente o Teorema Central do Limite e em seguida iniciaremos a sua demonstração. No decorrer da demonstração utilizaremos alguns teoremas e proposições que não serão demonstrados aqui, a fim de que este trabalho não se torne demasiadamente extenso. Contudo, forneceremos uma referência onde essas demonstrações podem ser verificadas.

Teorema Central do Limite para Variáveis Aleatórias i.i.d.: Sejam $X_1, X_2, \dots$ variáveis aleatórias independentes e identicamente distribuídas com $E(X_n) = \mu$ e $V(X_n) = \sigma^2$ estritamente positiva. Então,

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{D} N(0,1)$$

Podemos considerar, sem prejuízos para a demonstração, que $\mu = 0$, pois caso não seja basta provar o resultado para $Y_n = X_n - \mu$, já que $E(Y_n) = 0$.

Como dissemos anteriormente, as funções características nos auxiliarão a demonstrar a validade deste teorema. Para tal, precisaremos de um importante resultado conhecido como Teorema da Continuidade de Paul Lévy. Este teorema pode ser enunciado da seguinte forma:

Teorema da Continuidade de Paul Lévy: Sejam $F_1, F_2, \dots$ funções de distribuição e $\phi_1, \phi_2, \dots$ suas respectivas funções características. Se ϕ_n converge pontualmente para um limite ϕ e se, além disso, ϕ for contínua no ponto 0, então existe uma função F tal que $F_n \xrightarrow{D} F$ e ϕ é a função característica de F. A demonstração deste teorema pode ser vista em JAMES (2011, p.237).

Assim, para que $\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{D} N(0,1)$, basta mostrar que $\phi_{\frac{S_n}{\sigma\sqrt{n}}}(t) \to e^{-\frac{t^2}{2}}$, $\forall t \in \mathbb{R}$, visto que $e^{-\frac{t^2}{2}}$ é a função característica da distribuição Normal padrão.

Além do Teorema da Continuidade de Paul Lévy, precisaremos também de algumas propriedades, facilmente verificadas sob a hipótese de independência:

(a) Se X e Y são variáveis aleatórias independentes, então $\phi_{X+Y}(t) = \phi_X(t).\phi_Y(t),\ \forall t \in \mathbb{R}$.

(b) Se Y é uma variável aleatória tal que $Y = aX + b$, então $\phi_Y(t) = e^{itb}\phi_X(at)$.

(c) Se $E(|X|^n) < \infty$, então ϕ_X possui n derivadas contínuas e

$$\phi_X^{(k)}(t) = \int (ix)^k e^{itx} dF_X(x)\,, k = 1,2, \dots, n$$

Em particular, $\phi_X^{(k)}(0) = i^k E(X^k)$.

As demonstrações de cada uma das propriedades expostas acima poderão ser verificadas em JAMES (2011, cap. 6).

Demonstração do Teorema Central do Limite: Da propriedade (b), segue-se que $\phi_{\frac{S_n}{\sigma\sqrt{n}}}(t) = \phi_{S_n}(\frac{t}{\sigma\sqrt{n}})$. Como as variáveis aleatórias são independentes, temos de (a) que

$$\phi_{S_n}\left(\frac{t}{\sigma\sqrt{n}}\right) = \prod_{k=1}^{n} \phi_{X_k}\left(\frac{t}{\sigma\sqrt{n}}\right)$$

Além disso,como as variáveis aleatórias são identicamente distribuídas, obtemos a seguinte igualdade

$$\prod_{k=1}^{n} \phi_{X_k}\left(\frac{t}{\sigma\sqrt{n}}\right) = \phi^n\left(\frac{t}{\sigma\sqrt{n}}\right)$$

onde, $\phi = \phi_{X_1}$.

Como estamos supondo $\mu = 0$, segue que $V(X_n) = E(X_n^2) < \infty$, por hipótese. Da propriedade (c) segue que ϕ possui duas derivadas contínuas. Então, utilizando a expansão de Taylor de ϕ e o fato de que $\phi^{(k)}(0) = i^k E(X_1^k)$, temos que

$$\phi(t) = \phi(0) + \phi'(0)t + \frac{t^2}{2}\phi''\big(\theta(t)\big)$$

onde, $|\theta(t)| \leq |t|$.

Podemos reescrever a expressão acima da seguinte forma

$$\phi(t) = \phi(0) + \phi'(0)t + \frac{t^2}{2}\phi''(0) + \frac{t^2}{2}\left[\phi''\big(\theta(t)\big) - \phi''(0)\right]$$

Com $\phi''\big(\theta(t)\big) - \phi''(0) \to 0$, quando $t \to 0$.

Do fato de $\phi(0) = 1$, $\phi'(0) = i\mu = 0$ e $\phi''(0) = -E(X_1^2) = -\sigma^2$, onde os dois últimos resultados decorrem da propriedade (c), segue-se que

$$\phi(t) = 1 - \frac{\sigma^2 t^2}{2} + \frac{t^2}{2}\left[\phi''\big(\theta(t)\big) - \phi''(0)\right]$$

Então, para t fixo tem-se

$$\begin{aligned}\phi^n\left(\frac{t}{\sigma\sqrt{n}}\right) &= \left(1 - \frac{t^2}{2n} + \frac{t^2}{2\sigma^2 n}\left[\phi''\left(\theta\left(\frac{t}{\sigma\sqrt{n}}\right)\right) - \phi''(0)\right]\right)^n \\ &= \left(1 - \frac{t^2}{2n}\left[1 - \frac{1}{\sigma^2}\left(\phi''\left(\theta\left(\frac{t}{\sigma\sqrt{n}}\right)\right) - \phi''(0)\right)\right]\right)^n\end{aligned}$$

Mas $1 - \frac{1}{\sigma^2}\left(\phi''\left(\theta\left(\frac{t}{\sigma\sqrt{n}}\right)\right) - \phi''(0)\right) \to 1$, quando $n \to \infty$. Além disso, pelo limite de Euler para números complexos, temos que se $c_n \to c$ então $\left(1 + \frac{c_n}{n}\right)^n \to e^c$. Assim

$$\left(1 - \frac{t^2}{2n}\left[1 - \frac{1}{\sigma^2}\left(\phi''\left(\theta\left(\frac{t}{\sigma\sqrt{n}}\right)\right) - \phi''(0)\right)\right]\right)^n \to e^{-\frac{t^2}{2}},$$

o que conclui a demonstração do teorema.

■

Esta versão do Teorema Central do Limite apresentada aqui é mais geral do que a já mencionada versão atribuída a De Moivre e Laplace, de modo que a versão de De Moivre e Laplace é, na verdade, um caso particular da versão que acabamos de expor.

Teorema Central do Limite de De Moivre-Laplace: Seja S_n o número de sucessos em n ensaios de Bernoulli independentes, com probabilidade p de sucesso em cada ensaio, onde $0 < p < 1$. Então,

$$\frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow{D} N(0,1)$$

A demonstração dessa versão do teorema segue imediatamente do resultado anterior, basta para isso notar que as variáveis aleatórias envolvidas são independentes e identicamente distribuídas com distribuição de Bernoulli com média p e variância $p(1-p)$. Como os ensaios são de Bernoulli, S_n tem distribuição Binomial.

A notória vantagem de se utilizar o primeiro resultado ao invés deste último é que no primeiro não precisaremos nos restringir à distribuição binomial.

Contudo, há ainda versões mais gerais do que esta que apresentamos. A seguir, apenas a título de curiosidade, enunciaremos a versão do Teorema Central do Limite de Lindeberg, uma vez que esta versão do teorema, em geral, não é vista nos cursos de licenciatura em Matemática. Sua demonstração poderá ser verificada em JAMES (2011, p.273). Esta versão possui a vantagem de as variáveis aleatórias não serem identicamente distribuídas.

Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes tais que $E(X_n) = \mu_n$ e $V(X_n) = \sigma_n^2$, onde $\sigma_n^2 < \infty$ e pelo menos um $\sigma_n^2 > 0$. Sejam $F_n = F_{X_n}$, $S_n = X_1 + \cdots + X_n$ e $s_n = \sqrt{\sigma_1^2 + \cdots + \sigma_n^2}$. Então,

$$\frac{S_n - E(S_n)}{s_n} \xrightarrow{D} N(0,1)$$

quando $n \to \infty$, se a seguinte condição de Lindeberg for satisfeita:

$$\forall \epsilon > 0, \lim_{n\to\infty} \frac{1}{s_n^2} \sum_{k=1}^{n} \int_{|x-\mu_k|>\epsilon s_n} (x-\mu_k)^2 dF_k(x) = 0.$$

Note que

$$\sigma_k^2 = \int (x-\mu_k)^2 dF_k(x) = \int_{|x-\mu_k|>\epsilon s_n} (x-\mu_k)^2 dF_k(x) + \int_{|x-\mu_k|\leq\epsilon s_n} (x-\mu_k)^2 dF_k(x)$$

Ou seja, a condição de Lindeberg significa que a variância da soma é muito pouco influenciada pelas “caudas” das variáveis aleatórias situadas a mais de ϵ desvios padrão de suas respectivas médias.

O Teorema Central do Limite de Lindeberg não será útil para as atividades propostas em sala de aula para que os alunos possam apreender a utilidade e a força do TCL, pois os experimentos realizados estarão dentro do quadro teórico de variáveis aleatórias independentes e identicamente distribuídas. Esse é o papel do capítulo seguinte dessa dissertação.

6 O PAPEL DA LEI DOS GRANDES NÚMEROS E DO TEOREMA CENTRAL DO LIMITE NA CONSTRUÇÃO DE UM RACIOCÍNIO ESTATÍSTICO

Defendemos, neste trabalho, a importância de uma construção do raciocínio estatístico baseado nos Teoremas Limite para um bom entendimento do que seja uma tomada de decisão num ambiente de incerteza. Isto vai além dos conceitos da Estatística Descritiva, que são unicamente tratados ao longo do Ensino Básico. Neste capítulo refletiremos como a Lei dos Grandes Números e o Teorema Central do Limite podem contribuir para a construção de um raciocínio estatístico inferencial e ao final proporemos um planejamento de uma oficina com futuros professores, com o objetivo de verificar como os Teoremas Limite podem contribuir para o desenvolvimento de um raciocínio estatístico. Contudo, como destacamos anteriormente, por questões adversas não foi possível dar início à oficina.

Por intermédio das diversas atividades de naturezas distintas (simulações braçais e digitais) esperamos contribuir para o desenvolvimento do que SHULMAN (1986) chama de Saber Pedagógico de Conteúdo do professor.

Antes de iniciarmos a discussão sobre qual é o papel desempenhado por esses Teoremas Limite para o ensino da Estatística, vamos retomar algumas reflexões feitas anteriormente sobre a natureza e a importância da Inferência Estatística, tanto como um dos pilares da Estatística, quanto para a formação estatística do cidadão.

6.1 INFERÊNCIA ESTATÍSTICA

A Estatística Descritiva é essencialmente a organização, resumo e a representação dos dados. Para tal são utilizadas as medidas de tendência central, os quartis, as medidas de dispersão, gráficos e medidas de assimetria e curtose, entre outros. Contudo, ao coletar dados para realizar uma pesquisa qualquer, nem sempre é possível obter os dados de todos os elementos pertencentes a uma população de interesse.

Assim, como realizar uma pesquisa a respeito de uma população cuja informação obtida é parcial, visto que não é possível vistoriar todos os elementos? Esse é o papel da Inferência Estatística. Quando não pudermos, e geralmente não podemos, coletar os dados de todos os elementos de uma determinada população para obter seus parâmetros utilizamos os dados de uma amostra para realizar estimativas sobre esses parâmetros.

Em geral, sobre os dados coletados de uma amostra populacional aplicamos um raciocínio estatístico para deduzir, com algum grau de confiabilidade, um intervalo em que se situa o valor do parâmetro de interesse. A esse raciocínio estatístico chamamos de Inferência.

Como pontuado na Introdução desse trabalho, nos ateremos apenas à metodologia da Inferência Frequentista, também dita Clássica, que entende os parâmetros populacionais como quantidades fixas (não aleatórias) e desconhecidas; e não nos dedicaremos a discutir a metodologia da Inferência Bayesiana, que entende o próprio parâmetro populacional como uma quantidade aleatória, passível, portanto, ele próprio, de um modelo de probabilidade.

Na seção 4.2, discutimos a importância da abordagem frequentista de probabilidade para a Estatística e enfatizamos justamente o tratamento inferencial. Contudo, antes de iniciarmos a discussão sobre as contribuições dos Teoremas Limite, tratados neste trabalho, para a Inferência Estatística precisaremos entender um pouco mais sobre os estimadores, em particular, precisaremos estudar um pouco mais sobre as características de $\bar{X}$ (a média dos resultados dos experimentos), uma vez que ela possui um papel importante tanto para a Lei Forte quanto para o Teorema central do Limite. Para tal, consideremos o exemplo a seguir, passível de ser trabalhado no Ensino Básico.

Exemplo 1: Considere uma microempresa em que quatro funcionários compõem a população de trabalhadores. Na tabela 1 são apresentados os salários de cada um desses trabalhadores.

Funcionário	Salário em R$
T1	1000,00
T2	800,00
T3	850,00
T4	950,00

Tabela 1

Nesse caso, os parâmetros média, variância e desvio padrão populacionais são, respectivamente, $\mu = 900{,}00$, $\sigma^2 = 6250$ e $\sigma = 79{,}06$. Num problema em que não tenhamos acesso à informação da população, estes parâmetros serão desconhecidos e fixos e uma questão de interesse da Inferência Estatística aqui, seria como estimar a média salarial dos funcionários dessa empresa se, só pudéssemos tomar uma amostra de tamanho 2, por exemplo.

A estrutura mais comum da Estatística é criar estimadores para o parâmetro em estudo, quase sempre clonando a forma funcional do parâmetro nos elementos da amostra, desde que garantamos que em média o estimador "acerta" o parâmetro desejado. No caso da média populacional μ nosso estimador $\bar{x}$ poderia ser descrito por $\bar{x} = \frac{x_1+x_2}{2}$, com x_1 e x_2 elementos de $\{800, 850, 950, 1000\}$, pois, como veremos adiante, esse estimador tem média coincidente com o valor do parâmetro, ou seja, é um estimador não tendencioso.

No entanto, não existe apenas um estimador não tendencioso possível. Poderíamos descrever diversos outros estimadores para a média μ, com essa mesma propriedade. Mas, ao contrário da média amostral, os outros apresentariam mais variabilidade de resultados (variância) e não seriam preferíveis à média amostral.

Em particular, neste capítulo, estaremos interessados em discutir duas características essenciais dos estimadores, a saber, não-tendenciosidade e precisão.

A primeira característica nos informa que a variável aleatória dada pelo estimador tem média equivalente ao valor do parâmetro populacional (desconhecido) e a segunda característica nos informa sobre a variabilidade do estimador em torno de seu valor esperado.

Estimador não tendencioso (ou não viesado, ou não viciado): Seja $\{X_1, X_2, \dots, X_n\}$ uma amostra aleatória. Um estimador $T = T(X_1, X_2, \dots, X_n)$ é dito não tendencioso para um determinado parâmetro θ da população se, $E(T) = \theta$. Em outras palavras, a média de todas as possíveis estimativas obtidas coincide com o valor do parâmetro.

Seguindo o exemplo do salário dos trabalhadores da tabela 1, apresentaremos a seguir, na tabela 2, quatro estimadores distintos e os analisaremos sob a perspectiva da não tendenciosidade, numa amostra aleatória simples, ou seja, com reposição dos elementos sorteados.

Amostras de tamanho 2	Estimador 1 $\bar{X} = \frac{X_1 + X_2}{2}$	Estimador 2 $\bar{Y} = \frac{2X_1 + 3X_2}{5}$	Estimador 3 $\bar{Z} = \frac{X_1 + 2X_2}{4}$	Estimador 4 $\bar{W} = \max\{X_1, X_2\}$
{1000, 1000}	1000	1000	750	1000
{1000, 800}	900	880	650	1000
{1000, 850}	925	910	675	1000
{1000, 950}	975	970	725	1000
{800, 1000}	900	920	700	1000
{800, 800}	800	800	600	800
{800,850}	825	830	625	850
{800,950}	875	890	675	950
{850,1000}	925	940	712,5	1000
{850,800}	825	820	612,5	850
{850,850}	850	850	637,5	850
{850,950}	900	910	687,5	950
{950,1000}	975	980	737,5	1000
{950,800}	875	860	637,5	950
{950,850}	900	890	662,5	950
{950,950}	950	950	712,5	950
Médias	900	900	675	943,75

Tabela 2

Dentre os estimadores apresentados apenas $\bar{X}$ e $\bar{Y}$ são estimadores não tendenciosos (não viesados), isto é, $E(\bar{X}) = \mu = E(\bar{Y})$, os outros dois estimadores apresentam vícios, o que pode ser verificado tanto do ponto de vista probabilístico, destituído de valores empíricos, quanto pelos resultados obtidos empiricamente.

a) $E(\bar{X}) = E\left(\frac{X_1+X_2}{2}\right) = \frac{1}{2}E(X_1 + X_2) = \frac{1}{2}[E(X_1) + E(X_2)] = \frac{1}{2}[\mu + \mu] = \mu = 900$

b) $E(\bar{Y}) = E\left(\frac{3X_1+2X_2}{5}\right) = \frac{1}{5}E(3X_1 + 2X_2) = \frac{3}{5}E(X_1) + \frac{2}{5}E(X_2) = \frac{3}{5}\mu + \frac{2}{5}\mu = \mu = 900$

c) $E(\bar{Z}) = E\left(\frac{X_1+2X_2}{4}\right) = \frac{1}{4}E(X_1) + \frac{2}{4}E(X_2) = \frac{1}{4}\mu + \frac{2}{4}\mu = \frac{3}{4}\mu = 675$

d) $E(\bar{W}) = E(\max\{X_1, X_2\}) = \frac{1}{16}800 + \frac{3}{16}850 + \frac{5}{16}950 + \frac{7}{16}1000 = 943{,}75$

Já a segunda característica (precisão) versa sobre a variabilidade das estimativas produzidas.

Um estimador $T = T(X_1, X_2, \ldots, X_n)$ é dito preciso se $V(T)$ é pequena. Além disso, dados dois estimadores $T = T(X_1, X_2, \ldots, X_n)$ e $U = U(X_1, X_2, \ldots, X_n)$ não tendenciosos para um parâmetro θ, diremos que T é mais eficiente que U se $V(T) < V(U)$.

Dentre os estimadores não tendenciosos apresentados, no exemplo anterior, é possível demonstrar também que a variância de $\bar{X}$ é menor do que a variância de $\bar{Y}$.

a) $V(\bar{X}) = V\left(\frac{X_1+X_2}{2}\right) = \frac{1}{4}V(X_1 + X_2) = \frac{V(X_1)+V(X_2)}{4} = \frac{\sigma^2}{2} = \frac{6250}{2} = 3125$

b) $V(\bar{Y}) = V\left(\frac{3X_1+2X_2}{5}\right) = \frac{1}{25}V(3X_1 + 2X_2) = \frac{1}{25}[9V(X_1) + 4V(X_1)] = \frac{13}{25}\sigma^2 = \frac{\sigma^2}{\frac{25}{13}} = \frac{6250}{\frac{25}{13}} = 3250$

Assim, $V(\bar{X}) < V(\bar{Y})$. Isso nos diz que o estimador $\bar{X}$ é mais preciso, e, portanto, mais eficiente, que o estimador $\bar{Y}$.

Portanto, o estimador $\bar{X}$ é mais adequado que os demais para estimar a média salarial dos trabalhadores.

No caso em que desejamos estimar a variância populacional, temos uma pequena complicação ao tentarmos simplesmente tomar um estimador que “copia” a forma pela qual σ^2 foi obtida. A seguir analisaremos dois estimadores distintos, ambos adotados em livros didáticos.

Amostras de tamanho 2	Média salarial da amostra $\bar{X} = \frac{X_1+X_2}{2}$ em R$	Estimador 1 $\hat{\sigma}^2 = \frac{\sum(x_i - \bar{X})^2}{n}$	Estimador 2 $s^2 = \frac{\sum(x_i - \bar{X})^2}{n-1}$
{1000, 1000}	1000	0	0
{1000, 800}	900	10000	20000
{1000, 850}	925	5625	11250
{1000, 950}	975	625	1250
{800, 1000}	900	10000	20000
{800, 800}	800	0	0
{800,850}	825	625	1250
{800,950}	875	5625	11250
{850,1000}	925	5625	11250
{850,800}	825	625	1250
{850,850}	850	0	0
{850,950}	900	2500	5000
{950,1000}	975	625	1250
{950,800}	875	5625	11250
{950,850}	900	2500	5000
{950,950}	950	0	0
Médias	900	3125	6250

Tabela 3

A tabela 3 nos mostra que o estimador $\hat{\sigma}^2$, que é uma “cópia” de σ^2, é tendencioso, pois sua média 3125 não coincide com o valor da variância populacional 6250, enquanto o estimador s^2 se mostra não tendencioso para esse mesmo parâmetro. Esse é um bom exemplo prático para se mostrar aos alunos do Ensino Básico por que os livros didáticos prescrevem a divisão na fórmula da variância amostral por *n* -1 ao invés de *n* como na fórmula da variância populacional.

Se desejássemos mostrar o comportamento dos estimadores com o aumento da amostra para tamanho 3, teríamos o seguinte quadro:

Amostras			$\bar{X} = \frac{X_1 + X_2 + X_3}{3}$	Estimador 1 $\hat{\sigma}^2 = \frac{\sum(x_i - \bar{X})^2}{3}$	Estimador 2 $s^2 = \frac{\sum(x_i - \bar{X})^2}{3-1}$
1000	1000	1000	1000	0	0
1000	1000	800	933,33	8888,89	13333,33
1000	1000	850	950	5000	7500
1000	1000	950	983,33	555,56	833,33
1000	800	1000	933,33	8888,89	13333,33
1000	800	800	866,67	8888,89	13333,33
1000	800	850	883,33	7222,22	10833,33
1000	800	950	916,67	7222,22	10833,33
1000	850	1000	950	5000	7500
1000	850	800	883,33	7222,22	10833,33
1000	850	850	900	5000	7500
1000	850	950	933,33	3888,89	5833,33
1000	950	1000	983,33	555,56	833,33
1000	950	800	916,67	7222,22	10833,33
1000	950	850	933,33	3888,89	5833,33
1000	950	950	966,67	555,56	833,33
800	1000	1000	933,33	8888,89	13333,33
800	1000	800	866,67	8888,89	13333,33
800	1000	850	883,33	7222,22	10833,33
800	1000	950	916,67	7222,22	10833,33
800	800	1000	866,67	8888,89	13333,33
800	800	800	800	0	0
800	800	850	816,67	555,56	833,33
800	800	950	850	5000	7500
800	850	1000	883,33	7222,22	10833,33
800	850	800	816,67	555,56	833,33
800	850	850	833,33	555,56	833,33
800	850	950	866,67	3888,89	5833,33

800	950	1000	916,67	7222,22	10833,33
800	950	800	850	5000	7500
800	950	850	866,67	3888,89	5833,33
800	950	950	900	5000	7500
850	1000	1000	950	5000	7500
850	1000	800	883,33	7222,22	10833,33
850	1000	850	900	5000	7500
850	1000	950	933,33	3888,89	5833,33
850	800	1000	883,33	7222,22	10833,33
850	800	800	816,67	555,56	833,33
850	800	850	833,33	555,56	833,33
850	800	950	866,67	3888,89	5833,33
850	850	1000	900	5000	7500
850	850	800	833,33	555,56	833,33
850	850	850	850	0	0
850	850	950	883,33	2222,22	3333,33
850	950	1000	933,33	3888,89	5833,33
850	950	800	866,67	3888,89	5833,33
850	950	850	883,33	2222,22	3333,33
850	950	950	916,67	2222,22	3333,33
950	1000	1000	983,33	555,56	833,33
950	1000	800	916,67	7222,22	10833,33
950	1000	850	933,33	3888,89	5833,33
950	1000	950	966,67	555,56	833,33
950	800	1000	916,67	7222,22	10833,33
950	800	800	850	5000	7500
950	800	850	866,67	3888,89	5833,33
950	800	950	900	5000	7500
950	850	1000	933,33	3888,89	5833,33
950	850	800	866,67	3888,89	5833,33
950	850	850	883,33	2222,22	3333,33
950	850	950	916,67	2222,22	3333,33
950	950	1000	966,67	555,56	833,33
950	950	800	900	5000	7500

950	950	850	916,67	2222,22	3333,33
950	950	950	950	0	0
Média			900	4166,67	6250

Tabela 4

Observe que novamente $E(s^2) = \sigma^2$, algo que pode ser mostrado no contexto da Teoria das Probabilidades. Nessa medida, poderíamos nos perguntar sobre o valor de $E(\hat{\sigma}^2)$.

$$E(s^2) = \sigma^2$$

$$E\left(\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{X})^2\right) = \sigma^2$$

$$\frac{1}{n-1}E\left(\sum_{i=1}^{n}(x_i - \bar{X})^2\right) = \sigma^2$$

$$E\left(\sum_{i=1}^{n}(x_i - \bar{X})^2\right) = (n-1)\sigma^2$$

$$\frac{1}{n}E\left(\sum_{i=1}^{n}(x_i - \bar{X})^2\right) = \frac{1}{n}(n-1)\sigma^2$$

$$E\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^2\right) = \frac{n-1}{n}\sigma^2$$

$$E(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2$$

Isso explica o porquê de a média de $\hat{\sigma}^2$ ser metade do valor do parâmetro na tabela 3 e $\frac{2}{3}$ do valor do parâmetro na tabela 4.

Abaixo são apresentados dois gráficos que ilustram a distribuição de $\bar{X}$ para amostras de tamanho 2 e 3, respectivamente.

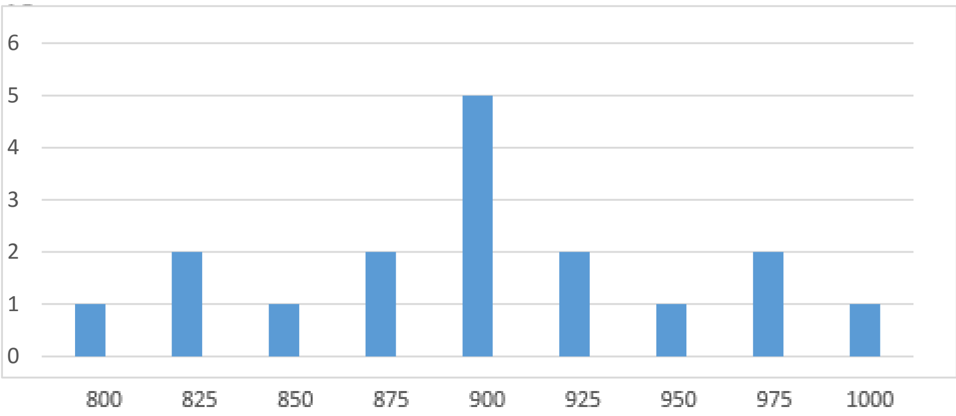


Gráfico 1: Amostras de tamanho 2

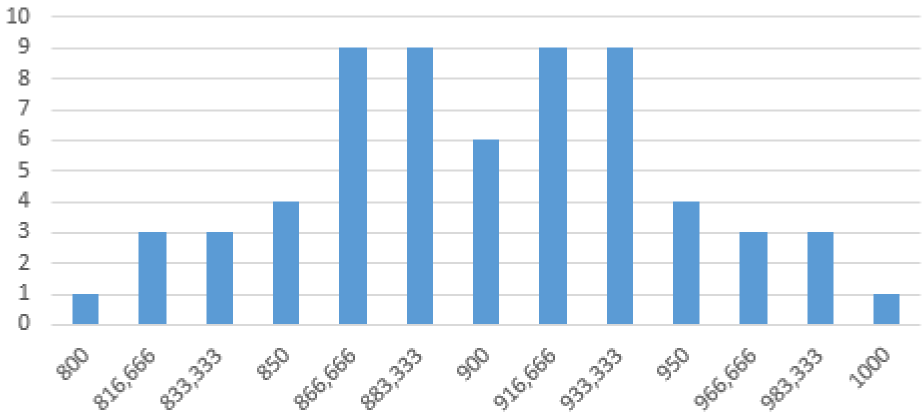


Gráfico 2: Amostras de tamanho 3

Apresentaremos agora uma terceira característica desejável em um estimador.

Estimador consistente: Dizemos que um estimador $T = T(X_1, X_2, \dots, X_n)$ para um parâmetro θ desconhecido da população é consistente se as duas condições a seguir são satisfeitas.

(i) $\lim_{n\to\infty} V(T) = 0$ e

(ii) $\lim_{n\to\infty} E(T) = \theta$.

Assim, estimadores consistentes podem ser tendenciosos, desde que seu viés convirja para zero. No caso de estimadores não tendenciosos, para que estes sejam consistentes, basta garantir que $\lim_{n\to\infty} V(T) = 0$.

Com base nos gráficos 1 e 2 apresentados podemos notar que a variância de $\bar{X}$ diminui conforme o tamanho da amostra aumenta. De fato, se tomarmos uma amostra de tamanho n temos:

$$V(\bar{X}) = V\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{1}{n^2}V(X_1 + \cdots + X_n) = \frac{V(X_1) + \cdots + V(X_n)}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Assim, se tomarmos o limite de $V(\bar{X})$ quando o número n de elementos da amostra tende ao infinito, temos que a variância do estimador tende a 0.

$$\lim_{n\to\infty} V(\bar{X}) = \lim_{n\to\infty} \frac{\sigma^2}{n} = 0$$

Obviamente, como $\bar{X}$ é não tendencioso para μ, a segunda condição é naturalmente satisfeita, pois

$$\lim_{n\to\infty} E(\bar{X}) = \lim_{n\to\infty} E\left(\frac{X_1+\cdots+X_n}{n}\right) = \lim_{n\to\infty} \frac{1}{n}E(X_1 + \cdots + X_n) = \lim_{n\to\infty} \frac{1}{n}nE(X_1) = \mu.$$

Assim, o estimador $\bar{X}$ é um estimador consistente para a média populacional. Essa característica é bastante desejável em processos de estimação em que o tamanho da população é desconhecido. Pois, quanto maior for o tamanho da amostra tomada, maior será a probabilidade de que a estimativa obtida esteja próxima do valor do parâmetro de interesse, algo bastante intuitivo.

A média amostral $\bar{X}$ não é apenas um estimador não tendencioso e consistente, é também o de variância mínima, dentre todos os estimadores para a média populacional, daí sua importância na tomada de decisão em geral.

A partir dessa breve discussão a respeito de alguns conceitos relevantes para a compreensão da Estatística Inferencial podemos continuar a discussão dos teoremas limite que motivaram esta dissertação e que trazem grandes contribuições para a construção do raciocínio estatístico.

6.2 LEI FORTE DOS GRANDES NÚMEROS

Na seção 6.1 vimos, através de exemplos de populações pequenas, para viabilizar os cálculos em sala de aula, a importância de determinadas propriedades de estimadores, como não tendenciosidade, precisão e consistência. Contudo, em geral, a população de interesse é demasiadamente grande ou até mesmo infinita.

Nesses casos poderíamos nos perguntar: como poderíamos ter certeza de que a partir de uma amostra dessa população, os resultados alcançados na seção anterior seriam mantidos? Isto é, tomando o exemplo dos funcionários (tabela 1), como poderíamos ter certeza de que, se a população de funcionários fosse muito grande, ou mesmo considerada infinita, o valor de $\bar{X}$ seria adequado para estimar o parâmetro μ (média salarial da população de funcionários)?

É a Lei Forte dos Grandes Números que garante esse resultado. Enunciaremos novamente esse teorema limite apenas com a finalidade de facilitar a leitura.

Sejam $X_1, X_2, \dots$ variáveis aleatórias independentes e identicamente distribuídas com $E(X_n) = \mu$. Então,

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow{q.c} E(X)$$

onde $\xrightarrow{q.c}$ indica a convergência quase certa.

Assim, se $X_1, X_2, \dots, X_n$ são variáveis aleatórias que constituem uma amostra aleatória do salário dos funcionários , então o estimador $\bar{X}$ (já visto como não viciado e consistente) para a média populacional μ tende ao valor do parâmetro conforme n aumenta com probabilidade 1.

Para auxiliar na percepção da importância da Lei Forte nos processos de tomada de decisão em situações regidas pela incerteza, proporemos uma série problemas ricos a serem tratados em aula via experimentações e simulações no Geogebra. Naturalmente os exemplos abordados não esgotam todas as situações possíveis de serem trabalhadas com fins ao Letramento Estatístico.

A fim de facilitar a leitura, adotaremos a seguinte estrutura para discutir os exemplos: apresentação da questão; discussão sobre os objetivos e aplicabilidade; estratégia de resolução; proposta de simulação via utilização de software (Geogebra) ou via braçagem, quando possível.

A ideia de utilizar as simulações tem o apelo pedagógico da verificação experimental do resultado obtido, facilitando os cálculos, ao mesmo tempo que ajuda na percepção do estudante para a importância Lei Forte na Ciência.

Os passos para as simulações via software (Geogebra) serão descritos no Apêndice.

Exemplo 1: Estimando a proporção de bolas vermelhas em uma piscina

Considere que em uma piscina há apenas bolas vermelhas e bolas brancas compondo um total de N bolas, de maneira que há tantas bolas dentro dessa piscina que é inviável contá-las. Como proceder para estimar pontualmente a proporção de bolas vermelhas na piscina?

Discussão sobre a questão

O objetivo desta questão é revelar de que forma a Lei Forte nos auxilia a realizar deduções a respeito de um parâmetro de interesse.

Esse problema trata de estimar a proporção de elementos em uma população que têm uma determinada característica. Alguns problemas reais que podem ser modelados por essa questão são: pesquisas de opinião pública, pesquisas sobre a proporção de pessoas que sofrem de uma determinada doença, pesquisas sobre proporção de peças defeituosas em uma produção, etc.

Uma abordagem de questões como essas em sala de aula traria um ganho substancial para a formação do estudante, não apenas pelo seu valor matemático, mas também pela possibilidade de dialogar com situações da realidade do aluno.

Pode-se sugerir que desenvolvam uma pesquisa simples que estime a quantidade de estudantes que almoçam na escola em cada dia da semana, prestando assim um serviço que auxilia na gestão da compra de alimentos da escola. Pode-se também sugerir questões mais complexas como por exemplo: estimar a proporção de estudantes que consomem drogas ilícitas, ou que sofreram um tipo específico de violência, entre outras coisas.

Todas essas questões prestam um serviço à comunidade escolar no que diz respeito ao seu planejamento de atividades pedagógicas para abordar esses mesmos temas transversais ao ensino. Assim, se for verificado uma que há uma grande proporção de estudantes que fumam, por exemplo, o gestor da escola pode promover palestras e debates sobre o tema. A forma pela qual os estudantes podem realizar esses trabalhos é muito parecida com a que utilizaremos para estimar a proporção de bolas vermelhas. A principal diferença se dará apenas na escolha da estratégia para obter uma amostra. Vejamos agora como estimar tal proporção.

Naturalmente que apenas estimar pontualmente a média neste caso não seria suficiente. Porém, como já discutido, é possível também, a partir de uma amostra, estimarmos a variância e o desvio-padrão, e assim construir intervalos de confiança para o parâmetro, por meio do Teorema Central do Limite, situação que será retomada na próxima seção, quando nos debruçarmos sobre as atividades em sala de aula com o TCL.

Estratégia

Para estimar a proporção de bolas vermelhas, podemos tomar, por exemplo, várias amostras aleatórias de mesmo tamanho $n < N$, em que n é pequeno, e calcular a média de cada amostra. Neste caso, a variável aleatória X assumirá o valor 1 quando a bola for vermelha e 0 quando for branca, e assim $\bar{X} = \hat{p}$, onde $\hat{p}$ é o estimador representando a proporção de bolas vermelhas na amostra. Além disso, como a proporção amostral é uma média, seu valor esperado coincidirá com a proporção populacional de bolas vermelhas na piscina. Assim, uma boa estratégia, seria por exemplo, tomar cinco amostras de tamanhos viáveis para a aferição da proporção amostral de bolas vermelhas e em seguida obter a média dessas médias, lançando mão assim, tanto da Lei Forte dos Grandes Números, quanto da propriedade de não-tendenciosidade do estimador proposto. Suponha, portanto, que tenhamos obtido cinco amostras de tamanho 100 com $\bar{X}_1 = 0{,}4$, $\bar{X}_2 = 0{,}3$, $\bar{X}_3 = 0{,}5$, $\bar{X}_4 = 0{,}35$ e $\bar{X}_5 = 0{,}45$. Uma boa estimativa pontual para a proporção de bolas vermelhas na piscina seria, portanto, $\bar{X} = \frac{\bar{X}_1+\bar{X}_2+\bar{X}_3+\bar{X}_4+\bar{X}_5}{5} = 0{,}4$.

A seguir apresentaremos formas pelas quais essa situação pode ser simulada, facilitando o entendimento, não apenas da questão, mas também da própria Lei Forte.

Simulação via braçagem

A simulação proposta é a seguinte: Em um saco cujo interior não possa ser visto coloque 80 bolinhas vermelhas e 120 brancas (as bolinhas precisam ter aproximadamente os mesmos tamanho e peso). Em seguida execute o procedimento abaixo:

1) Misture bem as bolinhas no interior do saco e peça para que um aluno tire 10 bolinhas do saco. Registre a proporção de bolinhas vermelhas obtida.
2) Refaça o passo 1 e além de registrar a nova proporção, registre também a média das duas proporções.
3) Repita o procedimento registrando sempre a média de todas as proporções já obtidas. Após muitas braçagens a média das proporções provavelmente estará próxima da proporção real de bolas vermelhas no saco, ou seja, 40%.

Exemplo dos registros

Amostra de tamanho 10	Proporção p	Média das proporções $\bar{X}$
Ensaio 1	0,2	0,2
Ensaio 2	0,7	0,45
$\vdots$	$\vdots$	$\vdots$
Ensaio n	0,1	0,38

Simulação via software

Para esta e para todas as demais simulações via software nos valemos do Geogebra. Tentaremos na medida do possível explicitar os procedimentos, exigindo apenas o conhecimento sobre a interface do programa, isto é, a localização de cada objeto citado e sua função.

Para criar um modelo de simulação adequado às nossas propostas no Geogebra, precisaremos desenvolver uma estrutura que executa a mesma tarefa sucessivas vezes. Para tal, adotaremos um **Procedimento para reprodução sucessiva de um experimento** como padrão para todos os applets que faremos, comentando as modificações necessárias para cada simulação proposta. Esse procedimento pode ser lido no Apêndice F deste trabalho.

Já as modificações que deverão ser realizadas para a construção do applet que simulará este exemplo se encontram em **Exemplo 1: Estimando a proporção de bolas vermelhas em uma piscina – Lei Forte**, também no Apêndice F.

Por meio da simulação é possível perceber que com poucas extrações de amostras de tamanho 10, a proporção de bolas vermelhas estimada já está bem próxima da proporção real.

O mais importante aqui é evidenciar o papel da Lei Forte para os processos de estimação.

O applet permite definir o número total de bolas em uma piscina e quantas serão vermelhas, bastando apenas modificar os valores nos campos de entrada. Além disso, por meio do controle deslizante n é possível alterar o tamanho da amostra, com repetição, que será tomada. Já o controle deslizante v permite controlar a velocidade com que as amostras são obtidas.

A fim de evidenciar o papel da Lei Forte para os processos de estimação, o applet foi programado para fornecer a média das proporções encontradas em todas as amostras de tamanho n já obtidas, bem como para exibir esse resultado dinamicamente.

Assim, é possível perceber que quanto maior for o número de amostras obtidas, mais próximo da proporção real de bolas a média das proporções encontradas nas amostras estará.

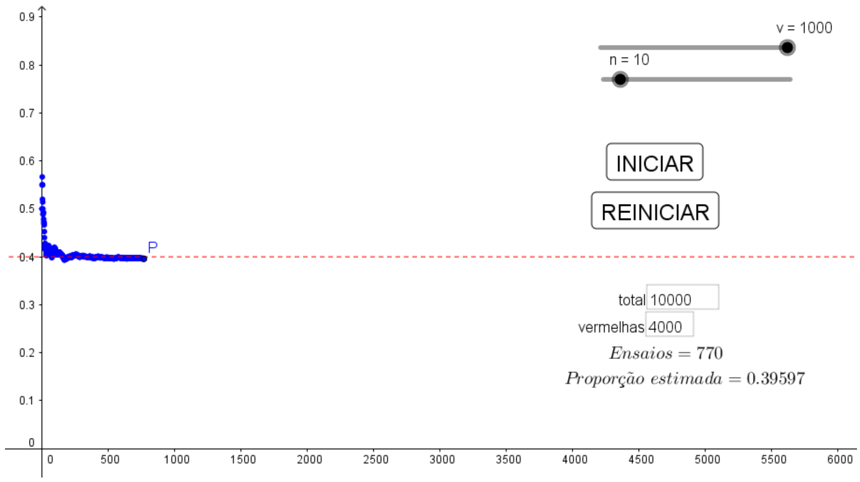


Figura 2: Simulação do exemplo 1 – Lei Forte

O download do arquivo pode ser feito pelo link: https://www.geogebra.org/m/XS9eAhW5

Exemplo 2: O problema clássico de Monty Hall

Em um programa de auditório o apresentador solicita que o convidado escolha uma dentre três portas. Atrás de uma das portas há um carro e atrás das outras duas há um bode. Após o convidado escolher uma das portas o apresentador abre uma das outras duas portas não escolhidas (sabendo que a porta que abrirá não terá o carro) e pergunta ao convidado se ele deseja trocar de porta ou se deseja manter a porta escolhida inicialmente. Qual será a melhor estratégia para ganhar o carro: trocar de porta, manter a escolha inicial ou é indiferente?

Discussão sobre a questão

O objetivo de discutir essa questão aqui é, além de reforçar a ideia de que a Lei Forte nos auxilia no processo de tomada de decisão ótima, mostrar também a possibilidade de resolver essa questão a partir de uma abordagem frequentista de probabilidade, uma vez que, geralmente, essa questão é abordada para o tratamento da probabilidade subjetiva, via teorema de Bayes.

Vejamos a seguir como essa questão pode ser abordada do ponto de vista frequentista.

Estratégia:

Nossa abordagem ao problema será dividi-lo em duas estratégias: na primeira o convidado adota a estratégia de nunca mudar de porta e na segunda o convidado sempre mudará de porta. Considere a variável aleatória X que assume o valor 1 toda vez que o carro é obtido e que assume o valor 0 caso contrário.

Situação 1: O jogador adota a estratégia de nunca trocar de porta

Nesta situação, se o jogador escolher a porta que contém o carro ele vencerá certamente, pois não irá trocar de porta. Assim, $P(X_i = 1) = \frac{1}{3}$. Por outro lado, se ele escolher a porta que contém o bode 1 ou o bode 2, ele certamente não vencerá, desta forma, para a escolha das outras duas portas teremos $P(X_i = 0) = \frac{2}{3}$.

Assim, supondo que pudéssemos realizar esse experimento infinitas vezes e nunca trocássemos de porta, pela Lei Forte, teríamos que:

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \xrightarrow{q.c} E(X_i) = \frac{1}{3}$$

Pois $E(X_i) = 1 \times \frac{1}{3} + 0 \times \frac{2}{3} = \frac{1}{3}$, o que nos leva a concluir que, adotando a estratégia de não mudar de porta, o jogador ganharia o carro em apenas um terço das vezes.

Situação 2: O jogador adota a estratégia de sempre trocar de porta

Aqui, se o jogador escolher inicialmente a porta com o carro, certamente a perderá, pois trocará de porta. Logo, $P(X_i = 0) = \frac{1}{3}$. Porém, se escolher inicialmente qualquer uma das duas portas que contém o bode, certamente ganhará o prêmio. Desta forma tem-se $P(X_i = 1) = \frac{2}{3}$.

Assim, em raciocínio análogo ao aplicado na situação 1, teremos pela Lei Forte dos Grandes Números

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \xrightarrow{q.c} E(X_i) = \frac{2}{3}$$

pois $E(X_i) = 0 \times \frac{1}{3} + 1 \times \frac{2}{3} = \frac{2}{3}$

Assim, se fosse adotada a estratégia de sempre trocar de porta, o jogador ganharia o carro em dois terços das vezes.

Portanto, a melhor estratégia para aumentar as chances de ganhar o carro é sempre trocar de porta. Abaixo exibimos uma imagem que ilustra passo a passo o raciocínio discutido aqui.

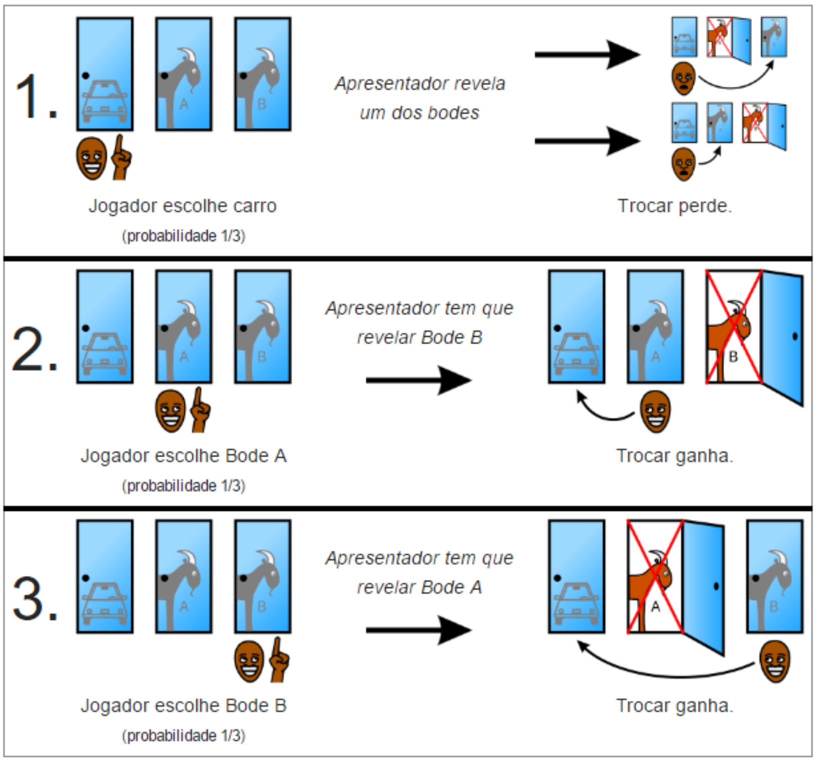


Figura 3: Problema de Monty Hall
(https://pt.wikipedia.org/wiki/Problema_de_Monty_Hall#A_solu.C3.A7.C3.A3o)

A seguir veremos as formas de simular essa questão em sala de aula, auxiliando na compreensão desta questão, aparentemente paradoxal, e na percepção do papel da Lei Forte no processo de tomada de decisão.

Simulação via braçagem

Simular esta situação em sala pode ser bastante divertido, principalmente se seguirmos o estilo de um programa de auditório. Porém, ao invés de utilizarmos portas, carros e bodes, utilizaremos, por exemplo, três cartas de um baralho todas de mesma cor e forma para representar as portas, em que duas são um número qualquer (representando o bode) e a outra é, por exemplo, um ás de ouro (representando o carro).

Divida a turma em dois grupos: o primeiro adotando a estratégia de nunca trocar de porta e o segundo com a estratégia de sempre trocar de porta.

Em seguida peça para que os estudantes sigam as instruções da questão para a escolha da carta (porta). Anotando o valor 1 sempre que o ás for encontrado e o valor 0 caso contrário (isso vale para ambos os grupos).

Tal como na questão anterior, peça para que ao final de cada experimento, além dos estudantes registrarem os 0's e 1's, registrem também a média de todos os resultados. Assim, teríamos (no caso da estratégia de não mudar de porta) uma tabela como a seguir:

Ensaios	Encontra ou não o Ás	Média $\bar{X}$
Ensaio1	0	0
Ensaio 2	1	$\frac{1}{2}$
⋮		
Ensaio n	0	$\frac{31}{100} = 0{,}31$

No exemplo acima, os estudantes deverão perceber que quanto mais ensaios forem realizados mais próximo o valor de $\bar{X}$ ficará de $\frac{1}{3}$, que, na tabela referente à outra estratégia, o valor se aproximaria de $\frac{2}{3}$.

Simulação via Software

Para simular essa questão no Geogebra, siga os passos descritos em **Exemplo 2: O problema clássico de Monty Hall – Lei Forte**, no Apêndice F.

Esse applet simula repetições sucessivas do experimento e exibe a proporção de vezes em que o carro é e não é encontrado em cada estratégia, a saber: a estratégia 1 é aquela em que a porta escolhida inicialmente é sempre mantida, já na estratégia 2 sempre ocorre a troca de porta.

O applet permite simular, além da versão clássica do problema, com três portas, situações com n portas, com n variando de 3 a 100, bastando para isso modificar o valor de n por meio do controle deslizante.

Em qualquer dessas variações o participante deverá escolher uma dentre n portas e o apresentador, após isso, abrirá n – 2 portas. Deixando o participante com a opção de manter a porta escolhida ou trocá-la por apenas uma outra.

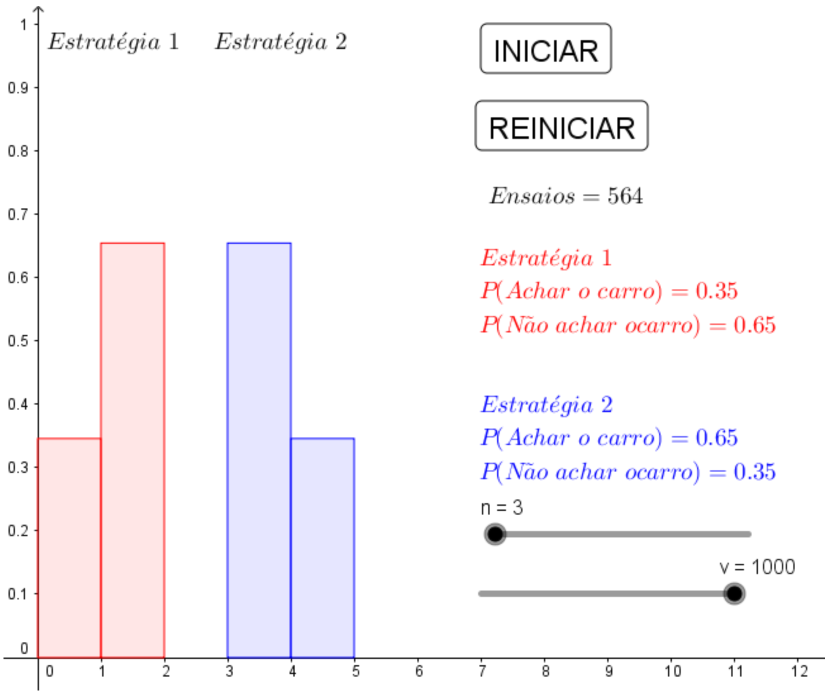


Figura 4: simulação do exemplo 2 – Lei Forte

Com essa simulação podemos perceber que se, pudéssemos realizar esse experimento infinitas vezes a estratégia 2 nos permitiria ter uma percentual maior de ganhos. E isto pode ficar ainda mais evidente, se refizermos a simulação adotando

100 para o valor de n. Isto é, simularemos o mesmo problema com 100 portas ao invés de 3 e serão abertas 98 portas após a escolha ao invés de 1. Evidenciando que, de fato, a melhor estratégia é trocar de porta.

O download do arquivo pode ser feito pelo link: https://www.geogebra.org/m/TZcGqTXu.

Exemplo 3: Problema de Otimização Estocástica: Venda de flores

Um floricultor constatou, depois de muito tempo, que a cada dia ele vende zero, uma, duas, três ou quatro flores de determinado tipo. Nunca vendeu mais do que quatro flores desse tipo em um dia. Ele organizou em uma tabela o número de flores compradas por dia e a frequência relativa com que esse número de flores é comprada, tal como a tabela abaixo

Nº de flores	0	1	2	3	4
Freq. relativa	p_0	p_1	p_2	p_3	p_4

Sabendo-se que o floricultor vende a flor por um valor R_1 e a compra do fornecedor por um valor R_2 e toda flor não vendida no dia não pode ser vendida no dia seguinte e deve ser jogada fora, qual seria a estratégia de compra diária de flores que o floricultor deveria adotar para maximizar seu lucro?

Discussão sobre a questão

Tal como no exemplo anterior, a Lei Forte, aqui, nos auxilia na tomada de decisão ótima. A decisão sujeita à incerteza a ser tomada aqui é a escolha do número de flores a se comprar diariamente, a fim de otimizar o lucro da empresa. Este exemplo representa uma vasta coleção de situações relacionadas ao comércio e à otimização em ambiente de incerteza (otimização estocástica).

Vejamos a seguir como a Lei Forte nos auxilia na escolha da melhor estratégia.

Estratégia:

Tal como no exemplo anterior aqui subdividiremos o problema em alguns casos. Considere que o lucro do floricultor é dado por $L = nR_1 - mR_2$, em que n e m representam os números de vezes que as flores foram vendidas aos clientes e compradas do fornecedor, respectivamente. Admitamos que o cliente que quer comprar três flores, por exemplo, comprará apenas uma se, só houver uma disponível ou duas se este for o número máximo de flores à venda.

Situação 1: O floricultor decide comprar sempre 0 flores deste tipo.

Neste caso, o lucro do floricultor é zero com probabilidade 1. Pois, mesmo que chegue alguém querendo comprar a flor o vendedor não poderá fazer a venda.

Assim, pela Lei Forte dos Grandes Números, temos que, com essa estratégia, após muitos dias, o floricultor pode esperar lucrar em média 0 reais.

Situação 2: O floricultor decide comprar diariamente sempre 1 flor deste tipo.

Se ninguém aparece para comprar esta flor, então o lucro é dado por: $L = -R_2$ com probabilidade p_0.

Por outro lado, se aparece alguém querendo comprar a flor, então o lucro é dado por $L = R_1 - R_2$ com probabilidade $p_1 + p_2 + p_3 + p_4$.

Assim, pela Lei Forte dos Grandes Números, temos que, com essa estratégia, após muitos dias, o floricultor pode esperar lucrar em média

$E(L) = (p_1 + p_2 + p_3 + p_4)(R_1 - R_2) - p_0R_2$ reais.

Situação 3: O floricultor decide comprar sempre 2 flores deste tipo diariamente.

Caso ninguém apareça para comprar essas flores, então $L = -2R_2$ com probabilidade p_0.

Se aparecer alguém querendo comprar exatamente uma flor nesse dia, então o lucro do dia é dado por: $L = R_1 - 2R_2$ com probabilidade p_1.

Já se forem vendidas as duas flores, então o lucro será $L_3 = 2(R_1 - R_2)$ com probabilidade $p_2 + p_3 + p_4$.

Assim, pela Lei Forte dos Grandes Números, temos que, com essa estratégia, após muitos dias, o floricultor pode esperar lucrar em média

$E(L) = 2(p_2 + p_3 + p_4)(R_1 - R_2) + p_1(R_1 - 2R_2) - 2p_0R_2\ reais.$

Situação 4: O floricultor decide comprar sempre 3 flores deste tipo diariamente.

Caso ninguém apareça para comprar essas flores, então $L = -3R_2$ com probabilidade p_0.

Se houver a demanda de uma flor, então o lucro do dia é dado por: $L = R_1 - 3R_2$ com probabilidade p_1.

Por outro lado, se houver a demanda de duas flores, então o lucro do dia é dado por: $L = 2R_1 - 3R_2$ com probabilidade p_2.

Se as três flores são vendidas, então $L = 3(R_1 - R_2)$ com probabilidade $p_3 + p_4$.

Assim, pela Lei Forte dos Grandes Números, temos que, com essa estratégia, após muitos dias, o floricultor pode esperar lucrar em média

$E(L) = 3(p_3 + p_4)(R_1 - R_2) + p_2(2R_1 - 3R_2) + p_1(R_1 - 3R_2) - 3p_0R_2$ reais.

Situação 5: O floricultor decide comprar sempre 4 flores deste tipo diariamente.

Caso ninguém apareça para comprar essas flores, então $L = -4R_2$ com probabilidade p_0.

Se houver a demanda de uma flor, então o lucro do dia é dado por: $L = R_1 - 4R_2$ com probabilidade p_1.

Se houver a demanda de duas flores, então o lucro do dia é dado por: $L = 2R_1 - 4R_2$ com probabilidade p_2.

Se as três flores são vendidas, então o lucro será: $L = 3R_1 - 4R_2$ com probabilidade p_3.

Por fim, se todas as quatro flores são vendidas, então o lucro é dado por $L = 4(R_1 - R_2)$ com probabilidade p_4.

Assim, pela Lei Forte dos Grandes Números, temos que, com essa estratégia, após muitos dias, o floricultor pode esperar lucrar em média

$E(L) = 4p_4(R_1 - R_2) + p_3(3R_1 - 4R_2) + p_2(2R_1 - 4R_2) + p_1(R_1 - 4R_2) - 4p_0R_2\ reais.$

Portanto, a estratégia ótima será aquela cujo lucro médio diário seja maior dentre os cinco cenários apresentados.

A partir das simulações propostas a seguir, esperamos facilitar a compreensão do papel da Lei Forte na escolha da estratégia ótima.

Simulação via braçagem

Para simular essa questão vamos adotar os seguintes valores

Nº de flores	0	1	2	3	4
Freq. relativa	0,1	0,2	0,4	0,1	0,2

$R_1 = 20$ reais e $R_2 = 10$ reais

Como gerador de aleatoriedade para essa situação utilizaremos uma urna contendo 10 bolas, sendo 1 branca, 2 azuis, 4 vermelhas, 1 verde e 2 amarelas, de tal maneira que o sorteio da bola branca equivale à demanda de 0 flores, o sorteio da bola azul à demanda de 1 flor, o sorteio de bola vermelha à demanda de duas flores, o sorteio de verde à demanda de 3 flores e, finalmente, o sorteio de bola amarela à demanda de 4 flores.

Separe a turma em quatro grupos, um para cada uma das estratégias possíveis (excluindo é claro a estratégia de comprar 0 flores, pois naturalmente o lucro esperado será 0).

Os grupos deverão proceder da seguinte forma:

Grupo 1: comprar sempre 1 flor

As bolinhas são misturadas na urna e uma bolinha é extraída, se essa bolinha for da cor branca, então é registrado um lucro de -R$10,00, representando que nesse dia nenhuma flor foi vendida.

Caso seja extraída uma bolinha de qualquer outra cor, então é registrado um lucro de R$10,00, representando que a única flor comprada foi vendida.

Ao fim de cada realização do experimento pede-se que os estudantes calculem a média dos valores registrados. No caso desse grupo, os estudantes provavelmente perceberão que conforme o número de realizações aumenta mais a média tende a ficar próxima de R$8,00.

Esse mesmo procedimento será adotado nos outros grupos, apenas modificando os valores registrados de acordo com o resultado dos dados.

Assim, no grupo 2, em que são compradas sempre duas flores, os estudantes deverão registrar R$-20,00 cada vez que o resultado da extração for uma bola branca; R$0,00 se for extraída uma bola azul e R$20,00 no demais casos. Este grupo deverá notar que a média, após muitas realizações, provavelmente ficará próxima de R$12,00.

Já o grupo 3, em que são compradas sempre três flores, registrará R$-30,00, toda vez que o resultado for uma bola branca; R$-10,00, quando o resultado for uma bola azul; R$10,00 quando uma bola vermelha for retirada e R$30,00 nos demais casos. Ao fim de muitos experimentos os estudantes notarão que a esperança do lucro converge quase certamente à R$8,00.

Por fim, o grupo 4, em que são compradas sempre quatro flores, registrará R$-40,00 quando o resultado for uma bola branca; R$-20,00 sempre que for obtida uma bola azul; R$0,00 quando forem obtidas bolas vermelhas; R$20,00 caso o resultado seja uma bola verde e R$40,00 caso a bola seja amarela. Neste caso, após muitas repetições do experimento, a média do lucro observado será, provavelmente um valor próximo a R$2,00.

Simulação via Software

Para desenvolver uma simulação para esta questão no Geogebra execute os códigos descritos em **Exemplo 3: Problema de Otimização Estocástica: Venda de flores – Lei Forte**, no Apêndice F.

Esse applet permite que o usuário modifique a questão inicial determinando valores de compra e venda, bem como as frequências relativas com que o floricultor vende zero, uma, duas, três ou quatro flores.

Com base nesses valores, ele simula sucessivas demandas diárias de flores e calcula o lucro obtido para cada uma das cinco estratégias. Possibilitando o reconhecimento da melhor estratégia em qualquer que seja a situação definida pelo usuário.

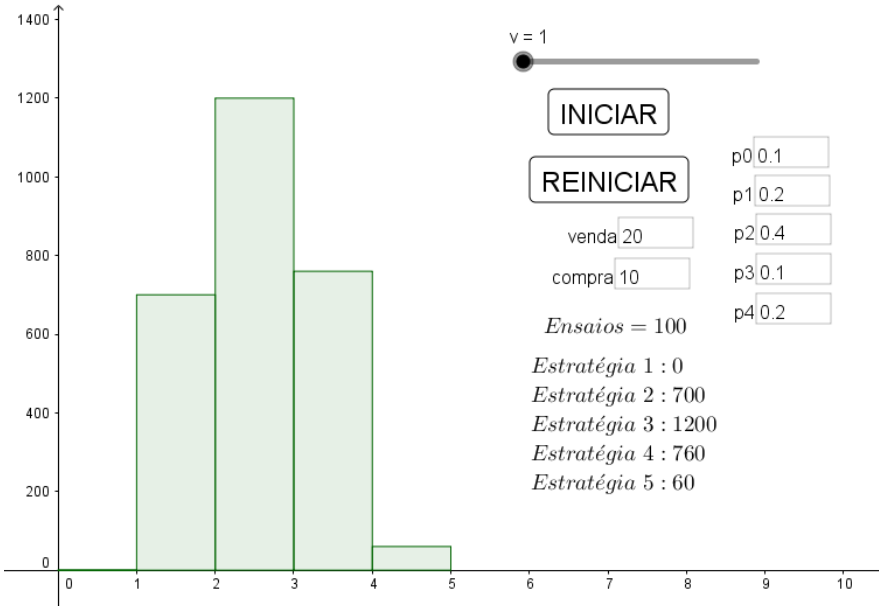


Figura 5: simulação do exemplo 3 – Lei Forte

Com as simulações propostas, tanto a braçal quanto esta, esperamos que o papel da Lei Forte para os processos de tomada de decisão seja compreendido e apreciado pelos alunos, a partir da clarividência de que quanto mais vezes o experimento for reproduzido, maior será o lucro, se adotarmos a estratégia 3: comprar sempre duas flores. A título de observação, inserimos ao lado de cada estratégia o lucro esperado após *n* realizações do experimento.

O download do arquivo pode ser feito pelo link: https://www.geogebra.org/m/tFXGB8Nr.

Exemplo 4: Probabilidade Geométrica

Duas pessoas marcaram de se encontrar entre 12h e 13:40h (100 minutos), para que uma não fique aguardando pela outra por muito tempo, estipulou-se que o tempo de espera pela outra pessoa será de exatamente 15 minutos. Se uma pessoa não aparecer nesse período de tempo ou até às 13:40h, o que ocorrer primeiro, a pessoa vai embora. Qual é a probabilidade que essas pessoas se encontrem, dado que as duas compareceram no local marcado dentro do horário previsto?

Discussão sobre a questão

O objetivo dessa questão é mostrar que não há uma necessidade de que o espaço amostral seja discreto, tal como ocorre nos contextos de Probabilidade Clássica. O papel da Lei Forte aqui é o de mostrar que a definição probabilidade geométrica pode ser obtida como consequência de sua aplicação, visando à convergência da taxa de ocupação de uma figura, a partir de objetos lançados sobre ela.

A partir disso é possível estimar qualquer tipo de área de figuras planas por mais complexas que elas sejam.

Naturalmente que, neste trabalho, explicaremos o desenvolvimento desta questão citando cada um dos conceitos matemáticos utilizados. Contudo, ao tratar de questões desse tipo com estudantes do Ensino Básico, sugerimos a utilização de softwares que auxiliam a modelar a situação proposta, a fim de que possam perceber, além do papel desempenhado pela Lei Forte, como o exercício está sendo modelado, não os sobrecarregando com uma série de notações e conceitos matemáticos.

Deixaremos como sugestão ao leitor o software Geogebra, a partir do qual realizamos a modelagem deste problema e plotamos aqui algumas imagens para auxiliar na compreensão.

Da forma como construímos o applet no Geogebra, pontos aleatórios $P(X,Y)$ são plotados em um plano cartesiano, em que cada coordenada do ponto representa a hora de chegada de um dos amigos. Se a diferença entre X e Y é menor que 15 um ponto vermelho é exibido nestas coordenadas, caso contrário é exibido um ponto azul.

Abaixo discutiremos a estratégia adotada para modelar e solucionar a situação proposta.

Estratégia:

Para tratar dessa situação considere que X e Y são as variáveis aleatórias que medem o tempo que o primeiro amigo e o segundo amigo, respectivamente, demoram a chegar, a partir de 12h.

Assim, temos que $X \sim U[0,1]$ e $Y \sim U[0,1]$, variáveis aleatórias independentes.

Para resolver a questão, devemos calcular $P(|X - Y| \leq 0{,}15)$.

Seja $P(X, Y)$ um ponto no plano cartesiano, de forma que se o primeiro amigo chega às 12:30h e o segundo às 12:46h, $P(0{,}30;\ 0{,}46)$.

Assim, o espaço amostral pode ser descrito como a área de um quadrado de lado 1, tal como ilustrado na figura abaixo:

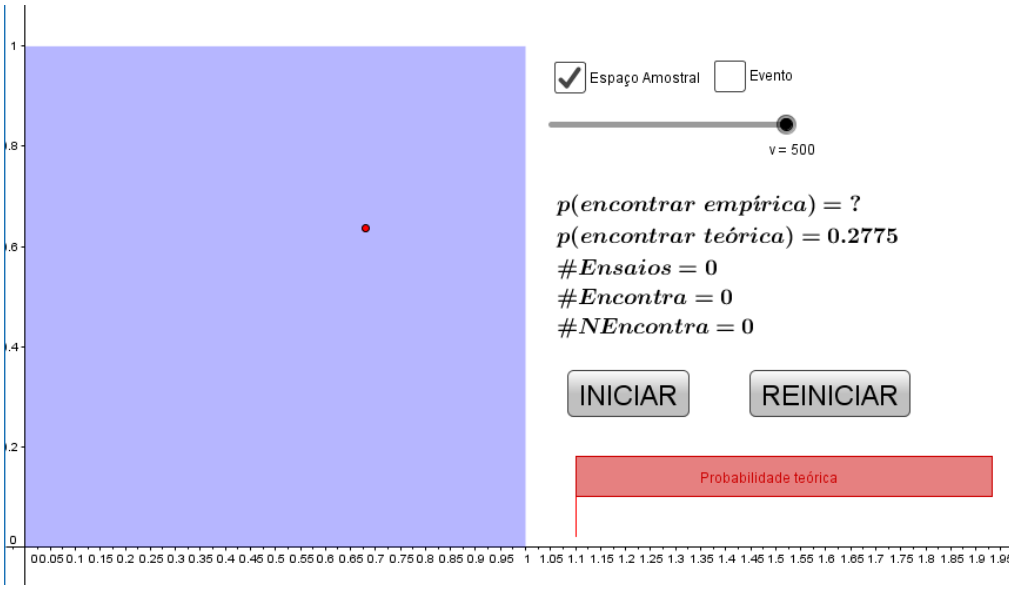


Figura 6: Probabilidade no contínuo

O evento desejado corresponde à área da figura delimitada pelo quadrado e pelas retas $y = x - 0{,}15$ e $y = x + 0{,}15$. Pois, como $|X - Y| \leq 0{,}15$, então

$X \geq Y - 0{,}15$ e $X \leq Y + 0{,}15$, tal como na figura 4 a seguir.

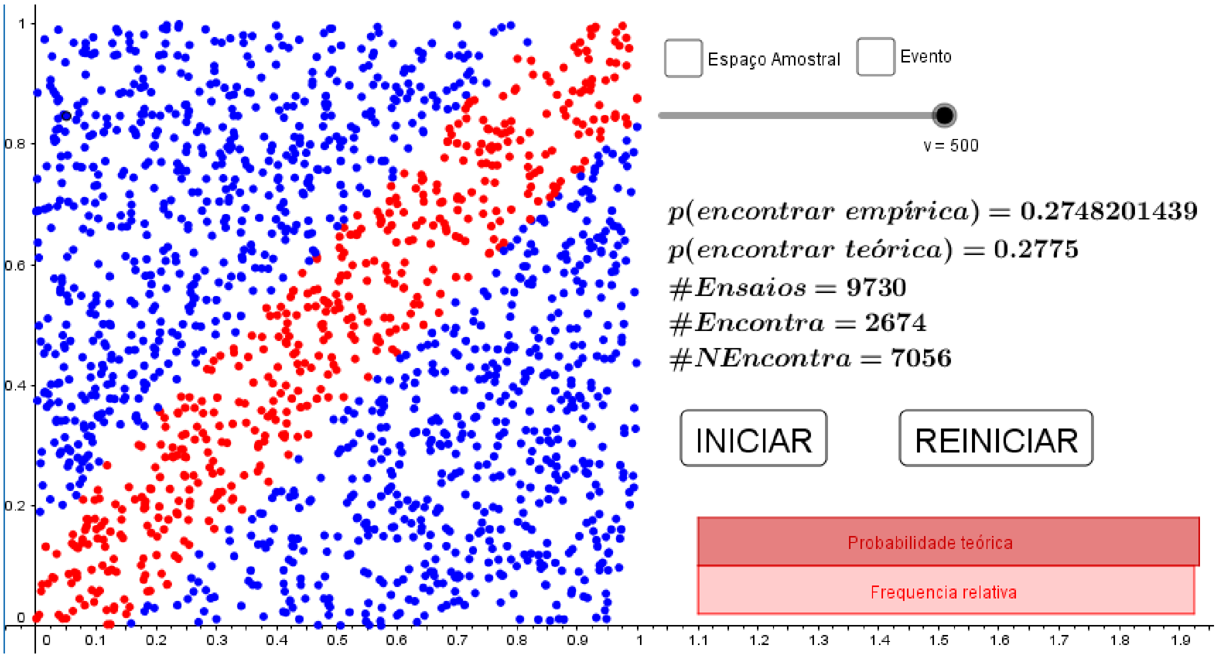


Figura 7: Probabilidade no contínuo - plotagem de pontos

Assim, realizamos um experimento em que uma grande quantidade de pontos $P(X,Y)$ são plotados aleatoriamente no gráfico, com $X \sim U[0,1]$ e $Y \sim U[0,1]$, tal como na figura acima. Considere a sequência de variáveis aleatórias W_1, W_2, ... tal que

$$W_i = \begin{cases} 1, & se\ P\ pertence\ a\ região\ correspondente\ ao\ evento \\ 0, & caso\ contrário \end{cases}$$

Note que $W_1, W_2, \ldots$ são variáveis aleatórias independentes e identicamente distribuídas com distribuição de Bernoulli.

Pela Lei Forte dos Grandes Números temos que

$$\frac{W_1 + W_2 + \cdots + W_n}{n} \xrightarrow{q.c} E(W)$$

Mas, $E(W) = 1 \times p + 0 \times (1 - p) = p$, onde p é a probabilidade de que as duas pessoas se encontrem.

Assim, podemos escrever

$$\frac{W_1 + W_2 + \cdots + W_n}{n} \xrightarrow{q.c} p$$

A abordagem frequentista de probabilidade exige um posicionamento ativo dos estudantes, uma vez que é necessário que realizem as experiências diversas vezes, seja por meio de execuções manuais ou de simulações via software.

Assim, nessa questão, o estudante seria levado a perceber que a média das variáveis $W_1,\ W_2, \dots, W_n$ converge para o valor de p (a razão entre as áreas do quadrilátero definido pela região $|x - y| \leq 0{,}15$ e área do quadrado, representando o espaço amostral).

Como mencionamos anteriormente, o desenvolvimento desta questão não é trivial. Ele exige muitos conceitos matemáticos que em geral não são tratados no Ensino Básico. Contudo, isso não impede que abordemos a Probabilidade Geométrica sob um ponto de vista da Lei Forte. A seguir propomos uma maneira de simular essa e outras questões de Probabilidade Geométrica, que podem ser realizadas em sala de aula.

Simulação via Software

Na explicação deste exemplo, mostramos cada raciocínio empregado para modelar o exercício no software Geogebra. No Apêndice F, em **Exemplo 4: Probabilidade Geométrica – Lei Forte**, descrevemos o passo-a-passo para desenvolver este applet.

Esse applet realiza sucessivos ensaios relativos a hora de chegada dos dois amigos e representa isso como um ponto do plano cartesiano. Além disso, o applet é programado para contar quantas vezes eles se encontram e não se encontram, para fornecer as respectivas probabilidades desses acontecimentos e para representar graficamente se eles se econtram (ponto vermelho) ou não (ponto azul).

Esperamos que este applet auxilie a compreender o papel da Lei Forte na definição da Probabilidade Geométrica.

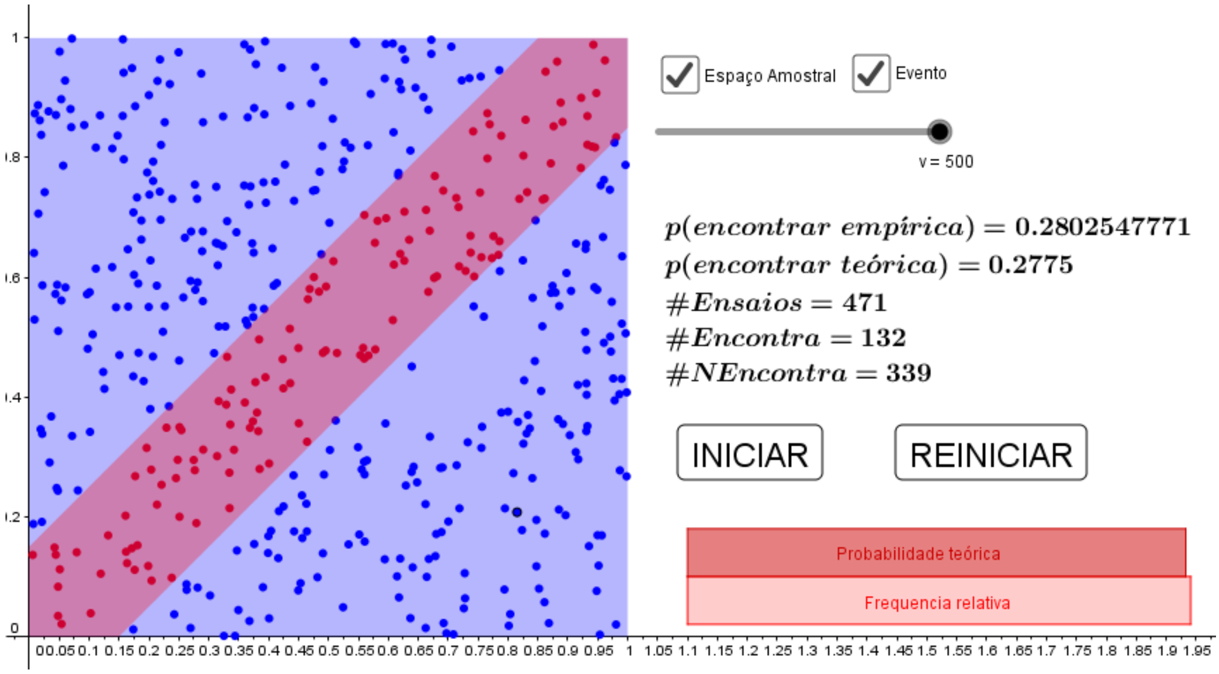


Figura 8: simulação do exemplo 4 – Lei Forte

O download do arquivo pode ser feito pelo link: https://www.geogebra.org/m/BxrhZMWc

Exemplo 5: Integral de Monte Carlo

Calcule a área entre a curva $f(x) = x^2$ e o eixo x, no intervalo $[0,1]$.

Discussão sobre a questão

Nesta questão o objetivo é mostrar que o papel da Lei Forte dos Grandes Números não se restringe a situações aleatórias, mas pode, também, auxiliar em questões determinísticas como o cálculo de integrais. Embora esse problema esteja além do conteúdo programático do Ensino Básico, ele pode ser utilizado nos cursos de Estatística da graduação como forma de ilustrar a importância da Lei Forte no cálculo de integrais complicadas.

Como no exemplo anterior, a utilização do Geogebra se mostrou igualmente eficiente para modelar essa questão via abordagem frequentista, valorizando o papel da Lei Forte.

A seguir apresentaremos a estratégia para solucionar tal questão.

Estratégia:

Esse método de cálculo de áreas é conhecido como Integral de Monte Carlo e a ideia central é que, a partir de muitas plotagens aleatórias de pontos $(x, f(x))$, sejamos capazes de estimar a integral de qualquer função integrável em um determinado intervalo.

Suponha desconhecido o valor da integral de $f(x) = x^2$ no intervalo $[0,1]$.

$$c = \int_0^1 f(x)\, dx$$

Se tomarmos um ponto $P = (x, f(x))$ tal que $X \sim U[0,1]$. A função densidade da variável aleatória X é dada por

$$h(x) = \begin{cases} 1, & 0 \leq x \leq 1 \\ 0, & caso\ contrário \end{cases}$$

Como $\int_0^1 f(x)\, dx = \int_0^1 f(x)h(x)\, dx = E\big(f(X)\big)$, a intergral desejada pode ser vista como a média da variável aleatória $Y = f(X)$.

Logo, $Y = f(X)$, com $X \sim U[0,1]$, é variável aleatória com média c (o valor da integral desejada). Portanto, pela Lei Forte dos Grandes Números, temos que

$$\frac{Y_1 + Y_2 + \cdots Y_n}{n} \xrightarrow{q.c} E(Y)$$

ou seja,

$$\frac{f(X_1) + f(X_2) + \cdots f(X_n)}{n} \xrightarrow{q.c} E\big(f(X)\big) = c$$

Assim, para estimar a integral de uma função $f(x)$ em um intervalo $[0,1]$ basta tomar uma amostra aleatória de pontos P e calcular o valor médio dos valores da imagem.

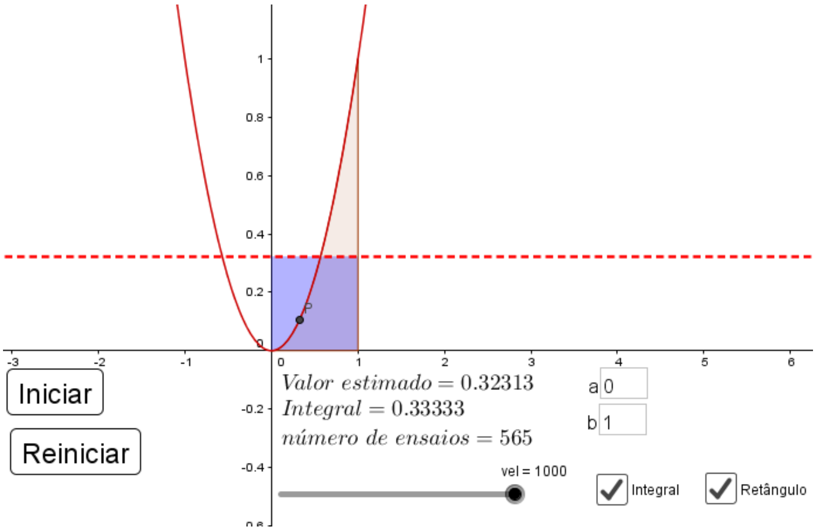


Figura 9: Estimar valor da integral de$f(x)$ em [0, 1]

Se quiséssemos estender esse problema para o caso de estimar a integral de um função para qualquer intervalo $[a, b]$, como poderíamos proceder? A estratégia para isto é a seguinte:

Seja $c = \int_a^b f(x)\,dx$, se tomarmos a variável $y = \frac{x-a}{b-a}$, teremos que $dy = \frac{1}{b-a}dx$ e, além disso, quando $x = a$, tem-se que $y = 0$ e quando $x = b$, obtém-se $y = 1$. Note também que, $x = y(b-a) + a$ e $dx = (b-a)dy$. Assim, temos que

$$c = \int_a^b f(x)\,dx = \int_0^1 f(y(b-a)+a)(b-a)\,dy$$

Portanto, se tomarmos o ponto $P(X, f(X(b-a)+a)(b-a))$, onde $X \sim U[0,1]$ e simularmos com o software diversas plotagens do ponto P, a integral da função f definida no intervalo $[a, b]$ será igual a área de um retângulo de base 1 e altura igual a média de todas as imagens$f(x(b-a)+a)(b-a)$, obtidas pelas simulações. Isso ocorre pois, se $X \sim U[0,1]$, então sua função densidade é dada por

$$h(x) = \begin{cases} 1, & 0 \leq x \leq 1 \\ 0, & caso\ contrário \end{cases}$$

Logo, $g(X) = f(X(b-a)+a)(b-a)$ é variável aleatória, e das propriedades de Esperança Matemática temos que

$$E\big(g(X)\big) = \int_0^1 g(x)h(x)\,dx = \int_0^1 g(x)\,dx = c.$$

Assim, $g(X_1)$, $g(X_2)$, ..., $g(X_n)$, com $X_i \sim U[0,1]$ para *i* = 1, 2, 3,... são variáveis aleatórias independentes e identicamente distribuídas com média c. Portanto, pela Lei Forte dos Grandes Números, temos

$$\frac{1}{n}\sum_{i=1}^{n} g(X_i) \xrightarrow{q.c} E\big(g(X)\big) = c$$

Na imagem abaixo representamos o cálculo da integral da função $f(x) = \sqrt{senh^2(x)}$ no intervalo [0, 2].

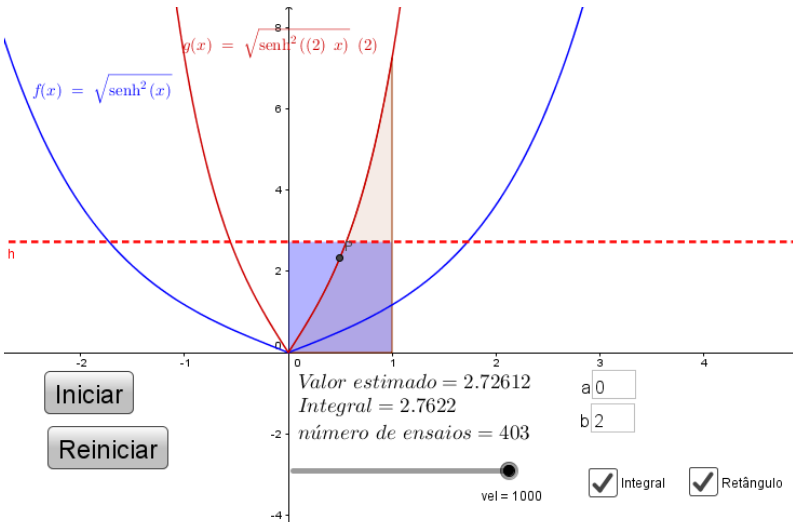


Figura 10: Estimar valor da integral def(x) em [0, 2]

Apesar de o Geogebra ser um bom software para modelar esse tipo de situação, ele pode apresentar lentidão nos cálculos a depender do número de simulações. Há, no entanto, vários outros softwares mais adequados a problemas desse tipo no trabalho científico.

Simulação via Software

Para desenvolver o applet no Geogebra, siga os passos descritos em **Exemplo 5: Integral de Monte Carlo – Lei Forte**, no Apêndice F.

O applet foi programado para seguir fielmente a ideia que descrevemos sobre a Integral de Monte Carlo. Contudo, ele permite, além de estimar o valor da integral, calcular o valor real da integral em um intervalo que pode ser definido pelo usuário, bastando para isso modificar os valores nos campos de entrada a e b.

:

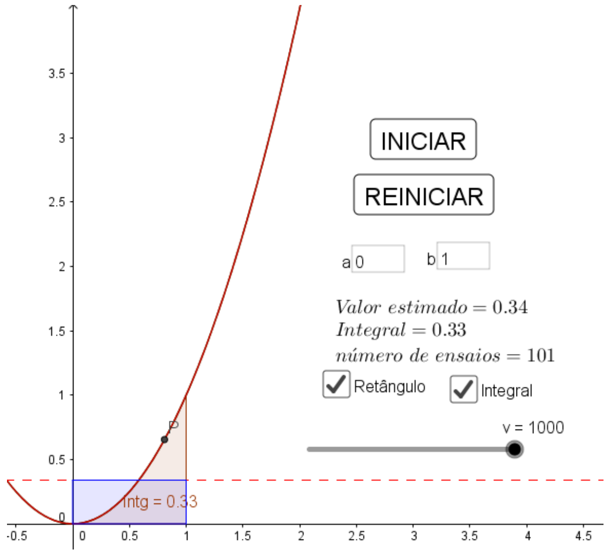


Figura 11: Simulação do exemplo 5 – Lei Forte

Esperamos que este applet permita que o leitor perceba a importância da Lei Forte também no tratamento de fenômenos determinísticos.

O download do arquivo pode ser feito pelo link: https://www.geogebra.org/m/BxrhZMWc.

Na seção anterior, vimos que o estimador $\bar{X}$, amplamente utilizado nesta seção por meio da Lei Forte, possui excelentes características estatísticas na estimação pontual da média populacional. Contudo, por ser uma estimativa pontual, é muito pouco provável que, à luz de uma única amostra, a estimativa produzida por ele seja igual ao valor do parâmetro de interesse. Para que possamos obter mais informação sobre a incerteza a respeito dessa estimativa, faz-se necessária a construção de intervalos de confiança para o parâmetro à luz da amostra selecionada. Na seção seguinte, com auxílio do Teorema Central do Limite, veremos como elaborar estimativas intervalares e como medir a probabilidade de que o valor do parâmetro pertença ao intervalo estimado.

6.3 TEOREMA CENTRAL DO LIMITE

Na seção anterior vimos como a Lei Forte pode contribuir para a tomada de decisão em uma situação regida pela incerteza. Assim como ocorreu lá, aqui, também, não temos, em geral, acesso a todas as informações da população e, portanto, precisamos recorrer a uma amostra dela. Porém, existem situações que exigem de nós um pouco mais de informação a fim de que tomemos decisões "mais acertadas".

Para tal, ao invés fazermos uso de uma estimativa pontual, precisaremos estimar um intervalo que tenha uma determinada probabilidade (que pode ser mensurada) de conter o parâmetro de interesse.

Uma outra grande contribuição do TCL é a possibilidade de aproximar a soma ou a média de uma sequência de variáveis aleatórias, por vezes de distribuição complicada ou desconhecida, à uma curva normal, desde que as condições do TCL que discutimos anteriormente sejam satisfeitas.

Contudo, antes de iniciarmos a discussão das valiosas contribuições do Teorema Central do Limite, através de exemplos, é preciso que nos aprofundemos um pouco mais no estudo sobre o estimador $\bar{X}$. Pois, como veremos nesta seção, ele possui uma característica bastante desejável para os processos de estimação intervalar, além, é claro, das características já debatidas neste trabalho.

Considere que um dado de seis faces "honesto" é lançado 2000 vezes. Pela Lei Forte, podemos esperar que a média dos resultados obtidos seja muito próximo a

$$E(\bar{X}) = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + \cdots + 6 \times \frac{1}{6} = 3{,}5$$

A figura a seguir exibe a distribuição de $\bar{X}$ após dois mil lançamentos de um único dado, simulados em um software (Geogebra). No eixo horizontal encontram-se os valores obtidos para a média amostral de uma amostra de tamanho 1 e o eixo vertical representa a probabilidade (frequência relativa) de cada média amostral possível.

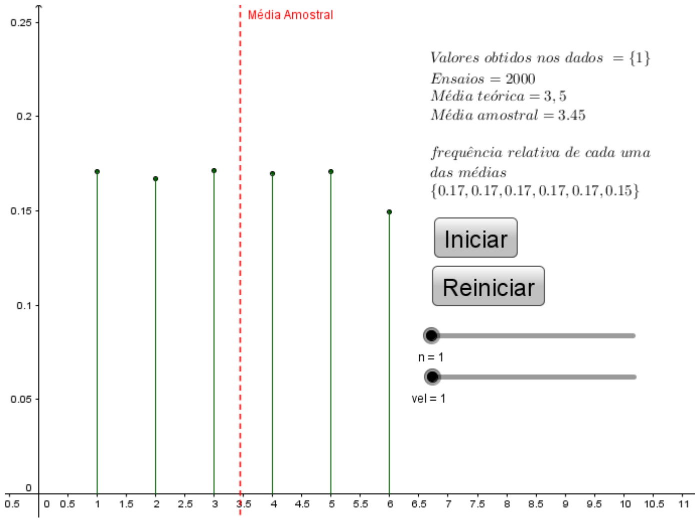


Figura 12: 2000 ensaios com amostras de tamanho 1

As próximas figuras exibem a distribuição de $\bar{X}$ para amostras de tamanhos 2, 5 e 10, em que para cada amostra de tamanho diferente foram realizados, igualmente, dois mil ensaios.

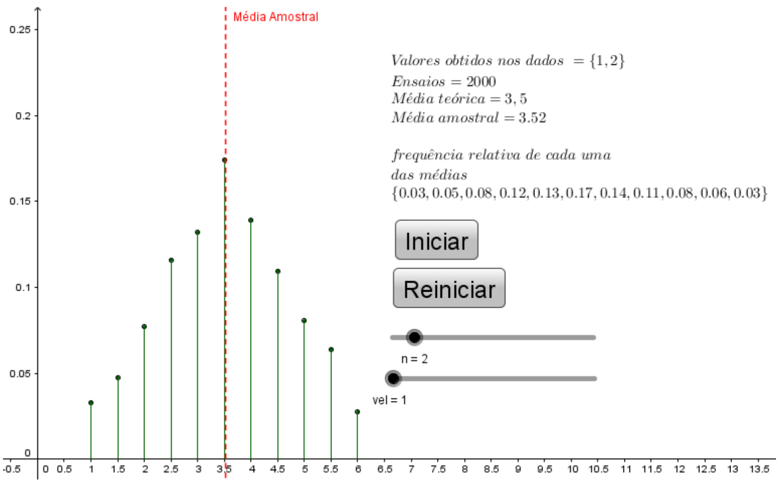


Figura 13: 2000 ensaios com amostras de tamanho 2

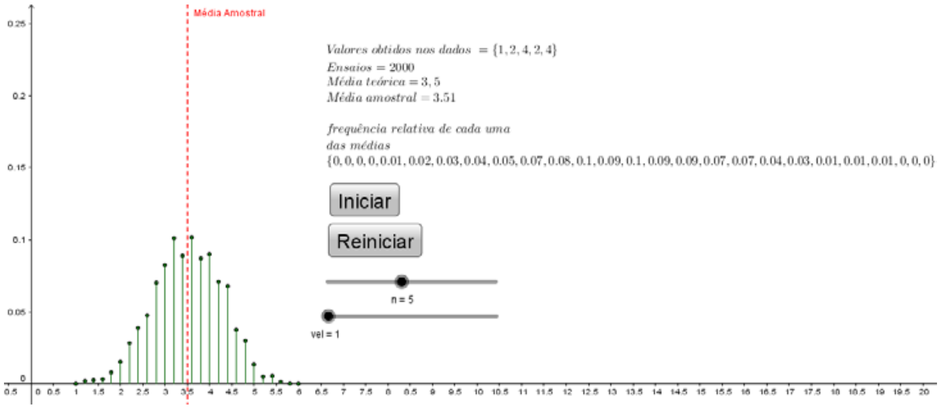


Figura 14: 2000 ensaios com amostras de tamanho 5

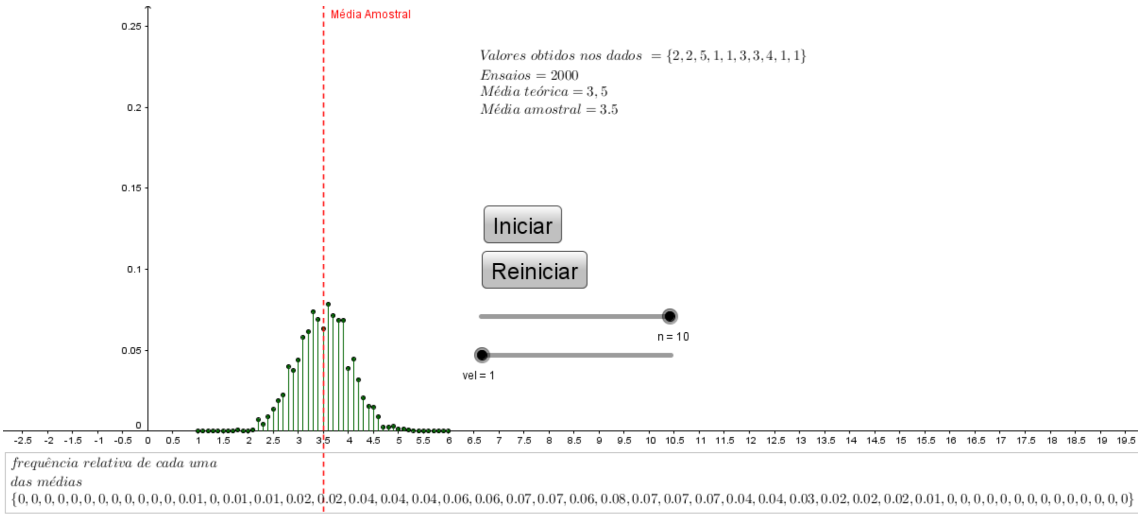


Figura 15: 2000 ensaios com amostras de tamanho 10

Notamos que, apesar da variável aleatória $\bar{X}$ ser discreta, quando o tamanho da amostra aumenta, a distribuição se torna cada vez mais "densa". Percebemos também que, quanto maior o número de elementos n da amostra, mais a distribuição de $\bar{X}$ se aproxima de uma curva Normal com média μ e variância $\frac{\sigma^2}{n}$. Isso é importante, pois, em muitos casos, é menos trabalhoso realizar cálculos e deduções a partir de uma curva Normal do que em uma distribuição discreta.

Além disso, note que quando n aumenta, a variabilidade da variável aleatória $\bar{X}$ diminui. De modo que, quando $n \to \infty$, a distribuição de $\bar{X}$ tende a se degenerar na reta $x = 3{,}5$. Isto é uma consequência do fato de $\bar{X}$ ser um estimador consistente.

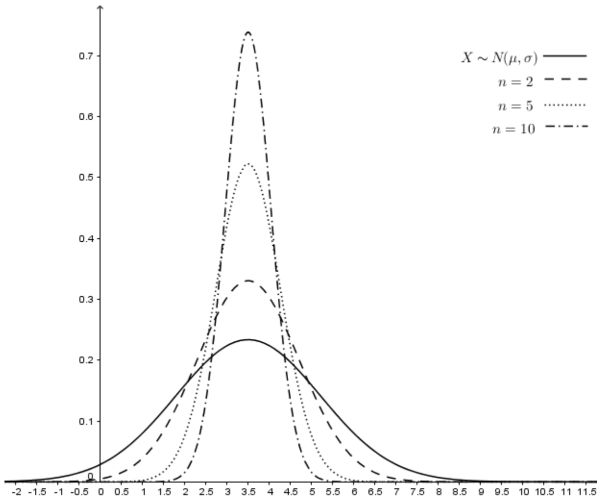


Figura 16: Distribuição da média amostral

De fato, $\bar{X} \approx N\left(\mu, \frac{\sigma^2}{n}\right)$, resultado este garantido pelo TCL, desde que $\mu < +\infty$ e $0 < \sigma^2 < \infty$ e que as variáveis aleatórias sejam independentes e identicamente distribuídas.

Como já vimos, pela Lei Forte dos Grandes Números, quando $n \to \infty$, $\bar{X}$ converge quase certamente para μ. Como $\sigma^2 = Var(X)$, e como as variáveis aleatórias são independentes e identicamente distribuídas, segue-se que

$$V(\bar{X}) = V\left[\frac{1}{n}(X_1 + X_2 + \cdots + X_n)\right] = \frac{1}{n^2}[V(X_1) + V(X_2) + \cdots + V(X_n)]$$
$$= \frac{1}{n^2}[\sigma^2 + \sigma^2 + \cdots + \sigma^2] = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Assim,o desvio-padrão da variável aleatória $\bar{X}$ é dado por $\sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$.

Esse fato sobre a variável aleatória $\bar{X}$ de uma certa maneira antecipa nossa intuição para o entendimento do Teorema Central do Limite de que conforme o tamanho da amostra aumenta, mais a distribuição da média amostral se aproxima de uma distribuição Normal. Apenas para facilitar a leitura enunciaremos novamente o TCL, dessa vez em termos de $\bar{X}$.

Se $X_1, X_2, \dots$ são variáveis aleatórias independentes e identicamente distribuídas com média $\mu < \infty$ e variância σ^2, tal que $0 < \sigma^2 < \infty$. Então,

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{D} N(0,1)$$

onde $\xrightarrow{D}$ indica a convergência em distribuição.

Quanto maior o valor de n melhor é a aproximação, mas, em geral, para $n \geq 30$, na maioria dos casos, a aproximação é muito boa.

Uma das primeiras aplicações do TCL foi apresentada por De Moivre e esta consistia em aproximar a distribuição binomial à distribuição Normal Padrão. A ideia era viabilizar o cálculo das probabilidades, uma vez que se torna inviável calcular a probabilidade de variáveis aleatórias binomiais com *n* muito grande. Porém, com auxílio da tabela da Normal padrão, calcular a probabilidade de variáveis aleatórias com distribuição Normal é uma tarefa bem simples.

Para mostrar experimentalmente a aproximação da distribuição Binomial à distribuição Normal, Galton propôs a utilização de um aparato conhecido por Galton Board (também chamado de Quincunx, Bean Machine, Galton Box, etc).

Conseguimos programar o Geogebra de modo a executar virtualmente o que esse aparato faz mecanicamente. Deixaremos ao leitor, muitas contribuições sobre simulações no Geogebra, porém, devido à grande quantidade de códigos que essa simulação exige, não descreveremos nesta dissertação os passos para sua construção. Em contrapartida, o leitor poderá obter essa simulação pelo seguinte link: https://www.geogebra.org/m/Am22GFJG.

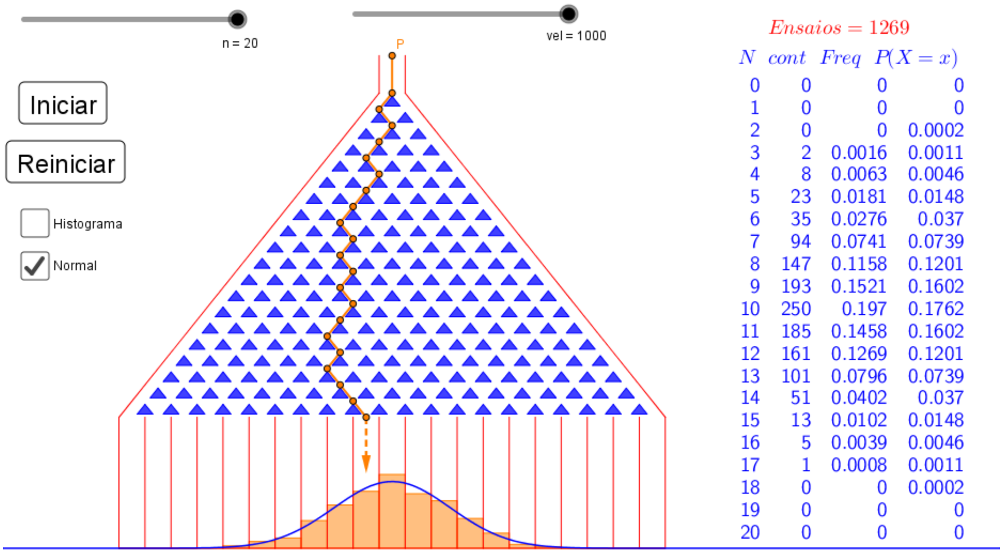


Figura 17: Aparato de Galton

Como mencionamos acima, o TCL desempenha uma função importante para o raciocínio estatístico tanto pelo seu resultado em si que permite que aproximemos sequências de variáveis aleatórias à uma curva normal padronizada, sob certas circunstâncias, quanto pela possibilidade de realização de estimativas intervalares.

Para que possamos nos aprofundar no mérito dessa segunda contribuição que citamos, será necessário retomar alguns conceitos que já prenunciamos anteriormente, tais como:

Nível de significância: a probabilidade α de que o intervalo a estimado não contenha o parâmetro θ de interesse.

Nível de confiança: a probabilidade $1-\alpha$ de que o intervalo a ser estimado contenha o parâmetro θ de interesse.

Note que, uma vez definido o nível de significância, o nível de confiança fica determinado e vice-versa.

Margem de erro: a estimativa E do erro da amostra.

Intervalo de confiança: o intervalo construído, para este trabalho, em torno de $\overline{X}$ que considera o erro da amostra e possui probabilidade $1-\alpha$ de conter o parâmetro θ.

Assim, o intervalo I é tal que $I=[\overline{X}-E, \overline{X}+E]$ tem probabilidade $1-\alpha$ de conter θ.

A partir disso, podemos agora estudar as contribuições do TCL para a promoção do raciocínio estatístico, por intermédio de exemplos, que seguem a mesma estruturação da seção anterior.

Exemplo 1: Aproximação da Binomial à Normal

Uma moeda honesta é lançada 100 vezes. Qual a probabilidade de se obter no máximo 40 caras?

Discussão sobre a questão

O objetivo de tratarmos essa situação aqui é justamente mostrar o papel desempenhado pelo TCL na aproximação de uma distribuição qualquer, neste caso Binomial, à curva Normal, desde que atendidas as condições de aplicabilidade do teorema; e em como isso minimiza e facilita as contas que devem ser realizadas, a fim de encontrarmos a probabilidade do evento desejado.

A questão proposta corresponde a um modelo de questão que, como já mencionamos, motivou De Moivre a desenvolver a expressão algébrica da curva Normal. Esta é, portanto, uma aplicação do Teorema Central do Limite a questões do próprio conteúdo de Probabilidade. Questões desse tipo podem ser tratadas em sala de aula, desde que os estudante tenham em mãos a Tabela Z (Apêndice D).

Alguns exemplos de questões que podem ser tratadas em sala e que seguem um modelo igual ou similar a este são: Qual a probabilidade de que em uma população de N pessoas, tenhamos no máximo n pessoas infectadas por um vírus, sabendo-se que a probabilidade de que qualquer pessoa seja infectada por esse vírus é p. Sabendo-se que há uma probabilidade p de que uma peça produzida não seja defeituosa, qual a probabilidade de que de um total de N peças produzidas, pelo menos, 98% delas não apresente defeito?

Apresentaremos agora a estratégia para resolver a questão proposta.

Estratégia:

Nesse exemplo, a variável aleatória X conta o número de caras obtidas em 100 lançamentos. Então para responder essa questão teríamos que calcular

$$P(X=0)+P(X=1)+\cdots+P(X=40)$$

com X tendo distribuição Binomial de parâmetros 100 e ½, o que envolve diversos cálculos combinatórios, já que a distribuição exata de X é dada por

$$P(X=k)=\binom{100}{k}\left(\frac{1}{2}\right)^{k}\left(\frac{1}{2}\right)^{100-k}, k=0,1,2,\dots,100.$$

Por outro lado, sabemos que

$E(X)=np=100\times\frac{1}{2}=50$, e que $DP(X)=\sqrt{np(1-p)}=\sqrt{100\times\frac{1}{4}}=5$.

Assim, se tomarmos a variável aleatória $Z=\frac{X-E(X)}{DP(X)}$, temos que $Z\sim N(0{,}1)$, e tendo em mente a necessidade do fator de correção de continuidade, cuja justificativa pode ser vista no Apêndice C, temos

$$P(X\leq 40{,}5)\approx P\left(\frac{X-E(X)}{DP(X)}\leq\frac{40{,}5-50}{5}\right)=P(Z\leq -1{,}9)=0{,}0287$$

A última igualdade é obtida através da tabela Z, que se encontra no Apêndice D.

Se calculássemos o valor exato de$P(X\leq 40)$, via distribuição Binomial, obteríamos como resultado 0,02844. Assim, o resultado obtido via aproximação pela curva Normal é bastante satisfatório.

A seguir propomos uma forma de simular essa questão que nos ajude a compreender como atua o TCL neste caso.

Simulação via Software

Para desenvolver um applet no Geogebra a fim de simular essa situação, execute os passos descritos em **Exemplo 1: Aproximação da Binomial à Normal – TCL**, no Apêndice F.

Esse applet permite que o usuário estime a probabilidade da ocorrência de, no máximo, n caras, em que o valor de n pode ser modificado por meio do controle deslizante. Além disso, exibe graficamente a frequência relativa com que ocorreram k caras (com k variando de 0 a 100).

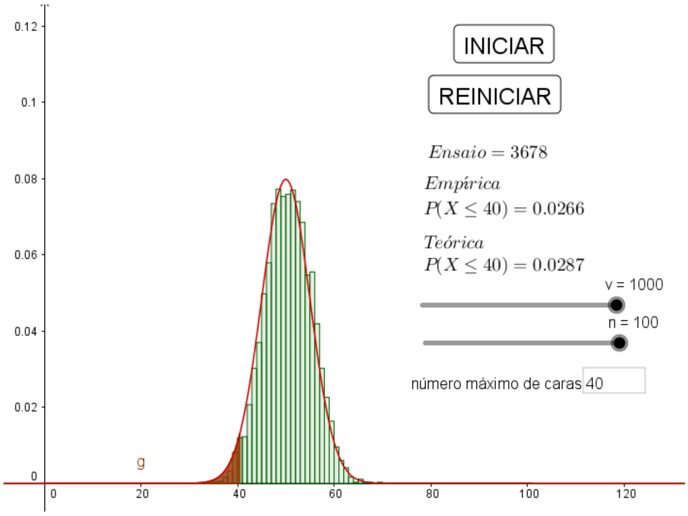


Figura 18: simulação do exemplo 1 – TCL

Esperamos, através dessa simulação, que o estudante de Estatística possa perceber o que significa o conceito de convergir em distribuição e como o TCL pode nos auxiliar a resolver cálculos probabilísticos mais complicados.

O download do arquivo pode ser feito pelo link: https://www.geogebra.org/m/WKNTe6bZ.

Exemplo 2: Fazer uma estimativa intervalar do número de bolinhas em uma piscina

Considere que em uma piscina há apenas bolas vermelhas e bolas brancas compondo um total de $1.000.000$ de bolas. De maneira que são tantas as bolas dentro dessa piscina que é inviável contá-las. Como proceder para fazer uma estimativa intervalar do número de bolas vermelhas que estão dentro da piscina?

Discussão sobre a questão

O objetivo aqui é tratar justamente da outra contribuição do TCL para o raciocínio estatístico, que é o processo de estimação intervalar, fornecendo assim mais informações do que aquelas veiculadas pontualmente pela Lei Forte, como visto na seção 6.2.

Aqui, além de obtermos mais informações a respeito do parâmetro de interesse, delimitando um intervalo para o possível número de bolas, poderemos medir a probabilidade de que o número de bolas vermelhas esteja dentro do intervalo estimado.

Não nos estenderemos aqui discutindo a aplicabilidade dessa questão a situações reais e relevantes, pois já o fizemos na seção anterior, e tal como esta, o TCL também poderia ser aplicado as demais questões que lá mencionamos. Trataremos aqui de “aprimorar” aquele estudo, a fim de permitir que tomemos decisões mais embasadas.

Contudo, não poderíamos deixar de ressaltar que seria uma excelente oportunidade trabalhar Estatística com os estudantes por meio de softwares, sejam eles voltados totalmente para Estatística ou simplesmente planilhas em que podem ser inseridas funções que agilizem os cálculos, tais como o Calc da Libre Office ou o Excel da Microsoft Office.

Passemos agora para a discussão de como procederemos para fazer a estimativa intervalar.

Estratégia:

A parte inicial é a mesma da que foi feita na seção 6.2. Coleta-se a amostra e verifica-se a proporção de bolas vermelhas. Vamos supor que da nossa amostra de tamanho $n = 900$ obtenhamos $\overline{X} = 0{,}3$.

O que precisamos fazer agora é estimar um intervalo de confiança em torno de $\overline{X}$ que tenha um determinado nível de confiança $1-\alpha$. Consideremos $1-\alpha = 0{,}95$. Assim, determinamos que o intervalo que estimaremos terá uma probabilidade de 95% de conter o parâmetro p (proporção de bolas vermelhas na piscina). Em linguagem matemática isto equivale a escrever

$$P(p \in [\bar{X} - E, \bar{X} + E]) = 0{,}95 \quad (\text{ I })$$

em que E é o erro da amostra.

Como X é uma variável aleatória independente e identicamente distribuída com média $\mu < \infty$ e variância $0 < \sigma^2 < \infty$, o TCL pode ser aplicado. Assim,

$$Z = \frac{\bar{X} - p}{\frac{\sigma}{\sqrt{n}}} \sim N(0{,}1)$$

Reescrevendo (I) em termos da variável aleatória Z, temos

$$P\left(-Z_{\frac{\alpha}{2}} \leq Z \leq Z_{\frac{\alpha}{2}}\right) = 0{,}95$$

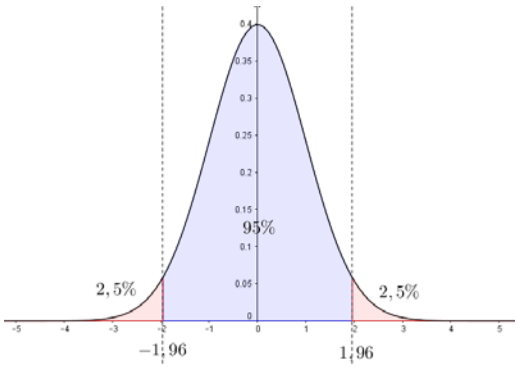


Figura 19: Curva Normal $N(0{,}1)$ para formação do intervalo de confiança

em que $Z_{\frac{\alpha}{2}}$ é tal que

$$P\left(Z \leq -Z_{\frac{\alpha}{2}}\right) = P\left(Z \geq Z_{\frac{\alpha}{2}}\right) = 0{,}025$$

Então, pela tabela Z do Apêndice D, temos $Z_{0,025} = 1{,}96$ e como $Z = \frac{\bar{X}-p}{\frac{\sigma}{\sqrt{n}}}$, segue que

$$P\left(-1{,}96 \leq \frac{\bar{X}-p}{\frac{\sigma}{\sqrt{n}}} \leq 1{,}96\right) = P\left(-1{,}96\frac{\sigma}{\sqrt{n}} \leq \bar{X}-p \leq 1{,}96\frac{\sigma}{\sqrt{n}}\right) =$$
$$= P\left(\bar{X} - 1{,}96\frac{\sigma}{\sqrt{n}} \leq p \leq \bar{X} + 1{,}96\frac{\sigma}{\sqrt{n}}\right) = 0{,}95$$

Como $\sigma = \sqrt{p \times q} = \sqrt{0{,}3 \times 0{,}7} = 0{,}458$, $\sqrt{n} = \sqrt{900} = 30$ e $\bar{X} = 0{,}3$, então

$$= P\left(0{,}3 - 1{,}96\frac{0{,}458}{30} \leq p \leq 0{,}3 + 1{,}96\frac{0{,}458}{30}\right) = 0{,}95$$

Assim, há 95% de chance do intervalo $\left[0{,}3 - 1{,}96\frac{0{,}458}{30}\ ;\ 0{,}3 + 1{,}96\frac{0{,}458}{30}\right] = [0{,}270\ ;\ 0{,}330]$ conter o parâmetro p estimado. Para esse intervalo de confiança, a margem de erro amostral é dada por $E = Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$.

Com isso, estimamos a proporção de bolinhas vermelhas na piscina. Para estimar a quantidade de bolinhas vermelhas basta aplicar as proporções dos extremos do intervalo ao número total de bolinhas que é de 1.000.000. Assim temos,

$$[0{,}270 \times 10^6, 0{,}330 \times 10^6] = [270.000, 330.000]$$

Portanto, podemos afirmar com um grau de confiabilidade de 95% que há entre 270.000 e 330.000 bolinhas vermelhas na piscina.

A seguir daremos uma sugestão de como poderia ser realizada uma simulação que auxiliasse a evidenciar a importância do TCL para a promoção do raciocínio estatístico.

Simulação via Software

Nosso objetivo com essa simulação é evidenciar que ao fazer uma estimativa intervalar para um determinado parâmetro, de fato, após muitas realizações do experimento a probabilidade de que o intervalo estimado contenha o parâmetro é $1-\alpha$.

Os passos para desenvolver o applet para esta simulação pode ser lido em **Exemplo 2: Fazer uma estimativa intervalar do número de bolinhas em uma piscina – TCL**.

Esse applet é similar ao realizado no exemplo 1 da seção anterior, porém, neste, são gerados intervalos de confiança à 95% para cada amostra de tamanho n obtida (a quantidade de elementos na amostra pode ser redefinida pelo usuário).

O applet é programado para verificar a proporção de vezes que o intervalo de confiança estimado contém o parâmetro, bem como a proporção de vezes que o intervalo não contém o parâmetro.

Observe que se o tamanho n da amostra for menor que 30, a aproximação à curva Normal, como dissemos no início desse capítulo, não será tão boa. Consequentemente, a proporção de vezes que o intervalo estimado conterá o parâmetro não será de 95%.

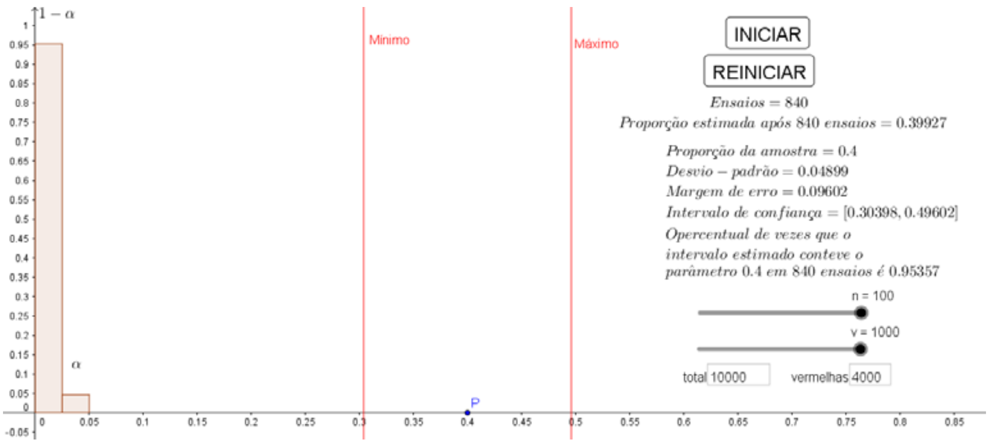


Figura 20: simulação do exemplo 2– TCL

O download do arquivo pode ser feito pelo link: https://www.geogebra.org/m/c7egrjhQ.

Exemplo 3: Estimando o lucro da venda de flores

Retomemos o problema e otimização do floricultor cuja distribuição da demanda se encontra na tabela reproduzida a seguir:

Nº de flores	0	1	2	3	4
Freq. relativa	0,1	0,2	0,4	0,1	0,2

Sabendo-se que o floricultor vende a flor por R$ 20,00 e a compra do fornecedor por R$10,00 qual seria a estratégia que o floricultor deveria adotar para maximizar seus lucros? Estime o intervalo de confiança para o lucro do vendedor após 100 dias, à 95% de confiança.

Discussão sobre a questão

Tal como na questão anterior, o objetivo aqui é produzir estimativas intervalares, ressaltando a importância dessas estimativas para o processo de tomada de decisão. A partir do TCL, expandiremos a discussão, feita na seção anterior, sobre o lucro da venda. Já que agora poderemos estimar, com algum grau de confiabilidade (aqui definido como 95%) qual seria a faixa de lucro que conteria o parâmetro lucro esperado na estratégia ótima.

Estratégia:

No exemplo discutido na seção anterior vimos como encontrar a melhor estratégia, utilizando a Lei Forte, a fim de que o floricultor tenha o maior ganho possível. Assim, apenas aplicaremos os resultados obtidos anteriormente para verificar qual das situações fornece o maior lucro e em seguida estimaremos o intervalo de confiança para esse lucro.

Situação 1: O floricultor compra sempre 0 flores deste tipo. $E(L) = 0$.

Situação 2: O floricultor compra sempre 1 flor deste tipo.

$$E(L) = (0{,}2 + 0{,}1 + 0{,}4 + 0{,}2)(20 - 10) - 0{,}1 \times 10 = 8.$$

Situação 3: O floricultor compra sempre 2 flores deste tipo.

$$E(L) = 2(0{,}2 + 0{,}1 + 0{,}4)(20 - 10) + 0{,}2(20 - 2 \times 10) - 2 \times 0{,}1 \times 10 = 12.$$

Situação 4: O floricultor compra sempre 3 flores deste tipo.

$$E(L) = 3(0{,}2 + 0{,}1)(20 - 10) + 0{,}4(2 \times 20 - 3 \times 10) + 0{,}2(20 - 3 \times 10) +$$
$$-3 \times 0{,}1 \times 10$$
$$E(L) = 8.$$

Situação 5:O floricultor compra sempre 4 flores deste tipo.

$$E(L) = 4 \times 0{,}2(20 - 10) + 0{,}1(3 \times 20 - 4 \times 10) +$$
$$+0{,}4(2 \times 20 - 4 \times 10) + 0{,}2(20 - 4 \times 10) - 4 \times 0{,}1 \times 10$$
$$E(L) = 2.$$

Assim, a estratégia que fornece o lucro máximo para o vendedor é aquela descrita na situação 3, em que o vendedor compra sempre 2 flores.

Considerando a sequência de variáveis aleatórias $\{L_n\}_{n\geq 1}$ que correspondem ao lucro obtido pelo floricultor no dia n. Queremos estimar o valor de $S_n = L_1 + \cdots + L_n$, isto é, da soma dos lucros de $n = 100$ dias. Como $\{L_n\}_{n\geq 1}$ é uma sequência de variáveis aleatórias independentes e identicamente distribuídas, com $E(L_i) = 12$, em que $1 \leq i \leq n$, então $E(S_n) = nE(L_i) = n \times 12 = 1200$ e variância $V(S_n) = nV(L_i)$, em que $V(L_i) = E(L_i^2) - E^2(L_i) = 320 - 144 = 176$, lembrando que $P(L_i = 20) = 0{,}7$, $P(L_i = 0) = 0{,}2$ e $P(L_i = -20) = 0{,}1$ e que $E(L_i^2) = (-20)^2\, P(L_i = 20) +\ (0)^2 P(L_i = 0) + (20)^2\, P(L_i = 20)$.

Logo, $V(S_n) = 17.600$.

Portanto, podemos aplicar o TCL.

$$Z_n = \frac{S_n - E(S_n)}{\sqrt{V(S_n)}} \approx N(0,1)$$

Como queremos estimar o somatório do lucro em 100 dias com um grau de confiabilidade de 95%, então $\alpha = 0{,}05$ e portanto queremos calcular $P\left(-Z_{\frac{\alpha}{2}} \leq Z_n \leq Z_{\frac{\alpha}{2}}\right) = 0{,}95$. Pela tabela Z, $Z_{\frac{\alpha}{2}} = Z_{0{,}025} = 1{,}96$. Assim, temos que

$$P\left(-1{,}96 \leq \frac{S_n - E(S_n)}{\sqrt{V(S_n)}} \leq 1{,}96\right) = 0{,}95$$

$$P\left(E(S_n) - 1{,}96 \times \sqrt{V(S_n)} \leq\ S_n \leq E(S_n) + 1{,}96 \times \sqrt{V(S_n)}\right) = 0{,}95$$

$$P(1200 - 1{,}96 \times 132{,}66 \leq\ S_n \leq 1200 + 1{,}96 \times 132{,}66) = 0{,}95$$

$$P(939{,}99 \leq\ S_n \leq 1460{,}01) = 0{,}95$$

Assim, mais do que saber que o florista poderia esperar ganhar R$1200,00 com essa estratégia, ao longo de 100 dias. Sabemos que, com um grau de confiabilidade de 95%, o intervalo $[939{,}99;\ 1460{,}01]$ conterá o valor do lucro do florista no intervalo de tempo determinado.

Naturalmente que em situações como a que é ilustrada por este exemplo, não teremos a oportunidade de verificar na prática do dia a dia qual seria a melhor estratégia. Contudo, nós propomos a seguir uma forma de simular essa questão a fim de revelar a importância do TCL para uma tomada de decisão.

Simulação via Software

O desenvolvimento da simulação dessa questão pode ser lido em **Exemplo 3: Estimando o lucro da venda de flores – TCL**, no Apêndice F.

O applet criado para simular esta questão é bastante similar ao do exemplo 3 da seção anterior, porém, neste applet são gerados intervalos de confiança, à 95%, em torno do lucro estimado em n dias para a estratégia definida pelo usuário por meio do controle deslizante. Além disso, o Applet mostra a proporção de vezes que o intervalo estimado conteve o parâmetro.

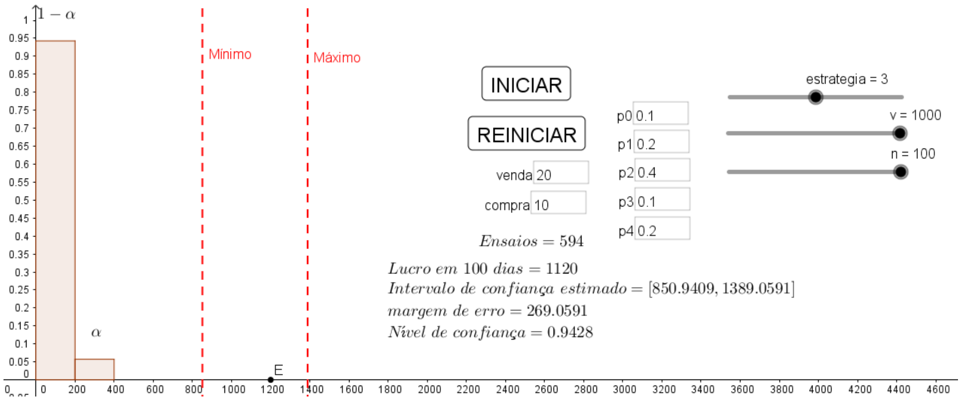


Figura 21: simulação do exemplo 3 – TCL

O download do arquivo pode ser feito pelo link: https://www.geogebra.org/m/WhXwuezG.

Exemplo 4: Estimação intervalar para área de figura planas

Estime um intervalo de confiança, à 95% para a área do quadrilátero interno ao quadrado na figura abaixo:

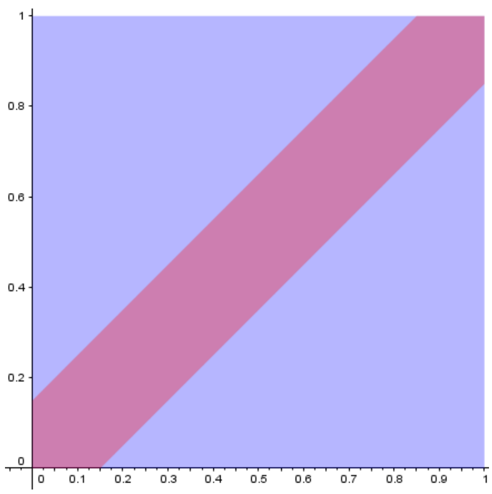


Figura 22: Estimativa para a área do quadrilátero

Discussão sobre a questão

O objetivo dessa questão é mostrar como a partir de uma abordagem frequentista de probabilidade pode-se, com auxílio dos Teoremas Limite, tirar conclusões a respeito não apenas de fenômenos aleatórios, mas também de fenômenos determinísticos. Isto é, podemos empregar o raciocínio estatístico também em situações não estocásticas.

O motivo de fazermos isso é bastante simples: em uma questão como essa do cálculo de áreas, às vezes a forma da figura a qual desejamos mensurar a área é tão complexa que não tenhamos, por alguma razão, condições de calculá-la. Contudo, independentemente da forma, o TCL nos mostra que podemos estimar a medida dessa área com algum grau de confiabilidade.

Uma estimativa pontual para essa questão já foi discutida em detalhes na seção 6.2, inclusive valorizando um raciocínio para a programação do software que nos auxilia realizando as plotagens aleatórias. A ideia aqui é fornecer aos estudantes um método alternativo para o cálculo de áreas de figuras planas quaisquer que sejam essas figuras.

Ressaltamos que, ao menos no software que utilizamos (Geogebra), quanto mais complexa for a figura que se deseja estimar a área, mais complicado será para programar o software para contar se o ponto plotado pertence ou não ao interior da figura desejada.

Porém, mesmo com essas “limitações” ainda sim é um software bastante interessante para se trabalhar o assunto, pois ele fornece a possibilidade de utilizar uma linguagem de programação bastante intuitiva e acessível.

A seguir daremos início a discussão de como o TCL pode ser empregado a fim de nos auxiliar a solucionar tal questão.

Estratégia

Como em todas as questões que temos discutido neste capítulo, aqui, também será necessário fazer uma estimativa pontual, fundamentada pela Lei Forte. Diferentemente do que fizemos em 6.2, onde discutimos a construção das funções que delimitavam a área que estávamos interessados em estimar, nesta seção, falaremos apenas da ideia geral para realizar a estimativa pontual.

Seja X a variável aleatória que conta quantas vezes o ponto P, plotado aleatoriamente pelo software dentro de um conjunto Universo descrito (neste caso, um quadrado de lado 1), “cai” dentro da região A que se quer estimar a área. Isto é,

$$X = \begin{cases} 1, se\ P\ pertence\ a\ região\ descrita\ por\ A \\ 0, caso\ contrário \end{cases}$$

A seguir com auxílio do software tomaremos uma amostra e verificaremos o percentual de vezes em que o ponto P “caiu” sobre a região A.

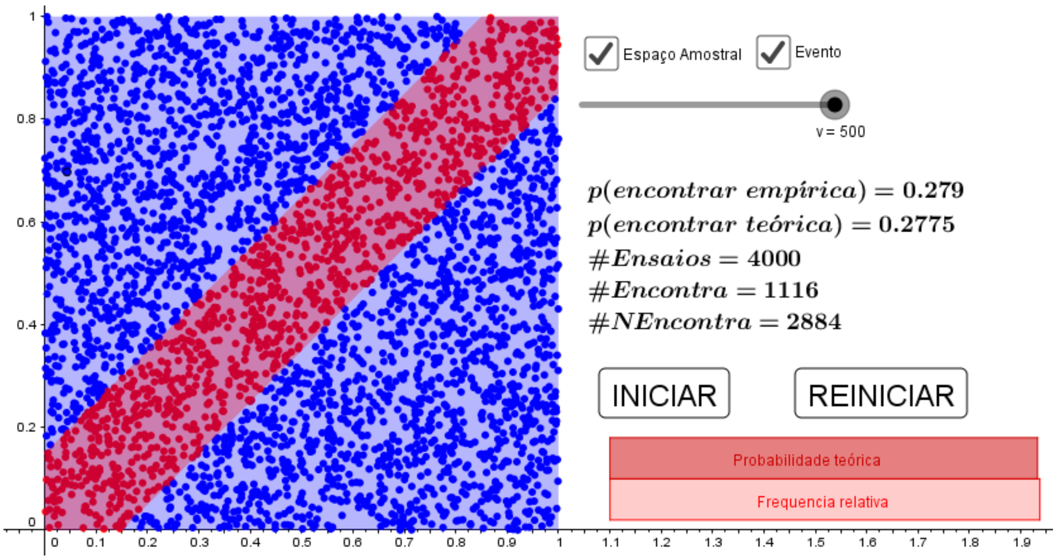


Figura 23: Estimativa para a área do quadrilátero (plotagens)

Na figura acima foram plotados 4000 pontos, dentre os quais 1116 estão sobre a região desejada. Assim,

$$\overline{X} = \frac{1116}{4000} = 0{,}279$$

Logo, temos que $\sigma = \frac{\sqrt{0{,}279 \times 0{,}721}}{\sqrt{4000}} \approx 0{,}022$ e $E = 1{,}96 \times 0{,}022 \approx 0{,}043$. Portanto, uma estimativa intervalar para o percentual de vezes que os pontos P "caem" sobre a região A é dada por

$$[0{,}279 - 0{,}043, 0{,}279 + 0{,}043] = [0{,}236, 0{,}322]$$

e como a área da região que corresponde ao conjunto Universo mede 1 unidade quadrada, então segue que uma estimativa intervalar para a área da região A é

$$[0{,}236, 0{,}322] \times 1 = [0{,}236, 0{,}322]$$

Assim, podemos afirmar com um grau de confiabilidade de 95% que a área da região A é uma medida compreendida entre 0,236 e 0,322 unidades quadradas.

Da facilidade e da velocidade com que o software pode plotar esses pontos, poderíamos facilmente tomar uma amostra maior, reduzindo assim a margem de erro, e consequentemente fazendo estimativas intervalares cada vez mais próximas de $\overline{X}$ (vide figura 4).

Em ambos os casos (figura 4 e figura 14), nossa estimativa intervalar contém o parâmetro de interesse que é a medida da região $A = 0{,}2775$, facilmente calculada pela Geometria. Porém, ressaltamos aqui que isto poderia não ter ocorrido, não sendo isto um erro de cálculo. Lembre-se que, como estipulamos um nível de confiança de 95%, há uma probabilidade de 5% de que o intervalo estimado não contenha o parâmetro de interesse.

Veremos agora como simular essa situação para realizar o tratamento em uma sala de aula.

Simulação via Software

Para fazer a simulação deste exemplo, siga os passo descritos em **Exemplo 4: Estimação intervalar para área de figura planas – TCL**, no Apêndice F.

Esse applet é similar ao construído para o exemplo 4 da seção anterior. Entretanto, neste são geradas amostras de tamanho n (definida pelo usuário), em que cada amostra é um conjunto de n pontos que representam se os amigos se encontram ou não. A partir dessa amostra o programa calcula a proporção de elementos na amostra que representam o encontro dos amigos e gera um intervalo de confiança, à 95%.

Por fim o applet exibe a proporção de vezes que o intervalo estimado conteve o parâmetro de interesse.

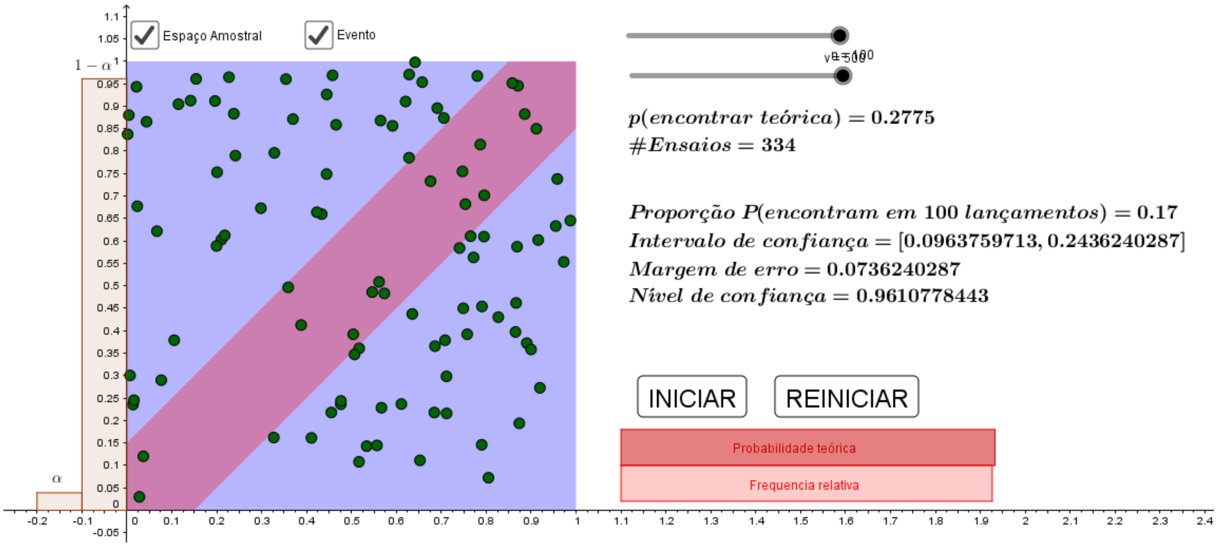


Figura 24: simulação do exemplo 4 – TCL

O download do arquivo pode ser feito pelo link: https://www.geogebra.org/m/m7ud6ec8.

Como se pode perceber, a Lei Forte dos Grandes Números em conjunção com o Teorema Central do Limite possibilitam o estudo de situações cujo interesse está na soma de uma sequência de variáveis aleatórias ou na média dessa sequência, desde que as condições do teorema sejam atendidas. Isso torna possível tratar uma grande quantidade de questões cotidianas e de aplicações dentro da própria matemática como discutimos nos exemplos acima.

Embora seja possível ensinar a Estatística Inferencial, por meios convencionais (quadro e giz somente), acreditamos que esta seria uma boa oportunidade para que os estudantes estabeleçam uma interatividade maior com a matéria, façam experiências, sejam por testagens braçais, pesquisas de campo ou por simulações computacionais, deixando de lado aquela postura passiva, tradicional das salas de aula.

Ambos os Teoremas Limite tratados aqui são desenvolvimentos matemáticos muito sofisticados. Por isso, não se espera que eles compreendam todos os detalhes da Lei Forte ou do TCL, apresentados neste trabalho. O que se espera é que eles possam compreender a ideia principal de cada um dos teoremas e ter em mente a importância deles no estudo de diversos fenômenos.

Tal ideia pode ser alcançada por diversos meios; em particular, acreditamos ser importante que ocorra o processo experimental. A aleatoriedade do processo precisa ser conscientizada pelo estudante, do contrário soará para o estudante como mais um dos algoritmos da matemática que produzem resultados sempre exatos e que em nada contribuem para o desenvolvimento de um raciocínio estatístico.

Destacamos também a importância que os conceitos estatísticos, atualmente ensinados nas Escolas Básicas, adquirem, quando trabalhados em conjunto com esses dois teoremas. A média, a variância e o desvio-padrão que são tratados exaustivamente de forma protocolar, em que praticamente todo o enfoque está no cálculo dessas medidas, passam a ter significado como balizadores no tratamento de problemas imersos em incerteza. Assim sua importância é ressaltada dentro de estudos sobre situações reais.

6.4 PLANEJAMENTO PARA DISCUSSÃO DOS TEOREMAS LIMITE COM FUTUROS PROFESSORES

A partir das discussões realizadas neste trabalho, elaboramos uma pequena oficina que visa a discutir com os futuros professores a relevância da Lei Forte e do Teorema Central do Limite para a construção de um raciocínio estatístico que aprofunde a discussão a respeito de uma tomada de decisão sob incerteza. A seguir apresentaremos o planejamento de cada um dos encontros para a realização da oficina.

Planejamento do primeiro encontro

Conteúdos: População, amostra, parâmetros e estimadores
Habilidades: Compreender a problematização central da Estatística por meio da necessidade de obtenção de informação sobre parâmetros da população através de uma informação incompleta veiculada pelo estimador obtido pela amostra.
Tempo: 03:00h

Etapa 1: Discutir o papel da Estatística para o estudo de fenômenos estocásticos.

Proposta de Problema para discussão do poder da Estatística como ferramenta para uma estimação sujeita à incerteza:

Suponha uma piscina com N bolinhas brancas ou vermelhas. Como você procederia para estimar o número de bolinhas vermelhas (ou brancas) com base numa amostra de tamanho n bem menor que o número total de bolas? Essa amostra única nos fornece uma estimativa pontual, a partir da qual podemos ter uma boa noção a respeito da verdadeira quantidade de bolas vermelhas no tanque? Como faremos para ter estimativas intervalares do número de bolinhas vermelhas?

A proposta acima ensejará a discussão emblemática dos quatro fatores fundamentais da Teoria Estatística, a saber, população, amostra, parâmetro e estimador, bem como a contemplação (dentro da teoria clássica da inferência) de quais quantidades são aleatórias e quais são determinísticas. Além disso, a proposta fornece também a possibilidade de estudar os casos limites (casos em que se tem acesso a todas as unidades populacionais) em que a Estatística já não tem mais tanto a sua primazia, uma vez que o conhecimento total da população não acarreta mais a variabilidade das estimativas.

Etapa 2: Discutir a importância dos estimadores, bem como a importância de que sejam precisos e não viesados para os processos de inferência Estatística.
Proposta de problema para discussão sobre os estimadores e sobre as estimativas:

Atividade 1: Considere uma microempresa em que quatro funcionários compõem a população de trabalhadores. Na tabela 1 são apresentados os salários de cada um desses trabalhadores.

Funcionário	Salário em R$
T1	1000,00
T2	800,00
T3	850,00
T4	950,00

Tabela 1

Essa população tem média salarial $\mu = 900{,}00$, variância $\sigma^2 = 6250$ e desvio-padrão aproximado $\sigma = 79{,}06$. Note que μ, σ^2 e σ são características da população de interesse, portanto, μ, σ^2 e σ são chamados parâmetros dessa população.

Lembrando que $\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$, em que N é o número de elementos na população.

De acordo com o que foi discutido na situação problema anterior, como poderíamos estimar a média salarial se, só pudermos tomar uma amostra de tamanho 2, em que as variáveis aleatórias são independentes? E a variância? Quais as características que são desejáveis em um estimador?

O objetivo dessa proposta é debater a relevância de se escolher conscientemente um estimador para os parâmetros levando em consideração a sua precisão e o viés.

Na tabela 2 são exibidos 4 estimadores para a média populacional.

Amostras de tamanho 2	Estimador 1 $\bar{X} = \frac{X_1+X_2}{2}$	Estimador 2 $\bar{Y} = \frac{2X_1+3X_2}{5}$	Estimador 3 $\bar{Z} = \frac{X_1+2X_2}{4}$	Estimador 4 $\bar{W} = \max\{X_1, X_2\}$
{1000, 1000}	1000	1000	750	1000
{1000, 800}	900	880	650	1000
{1000, 850}	925	910	675	1000
{1000, 950}	975	970	725	1000
{800, 1000}	900	920	700	1000
{800, 800}	800	800	600	800
{800,850}	825	830	625	850
{800,950}	875	890	675	950
{850,1000}	925	940	712,5	1000
{850,800}	825	820	612,5	850
{850,850}	850	850	637,5	850
{850,950}	900	910	687,5	950
{950,1000}	975	980	737,5	1000
{950,800}	875	860	637,5	950
{950,850}	900	890	662,5	950
{950,950}	950	950	712,5	950
Médias	900	900	675	943,75

Tabela 2

Dentre os estimadores apresentados apenas $\bar{X}$ e $\bar{Y}$são estimadores não tendenciosos (não viesados), isto é, $E(\bar{X}) = \mu = E(\bar{Y})$, os outros dois estimadores apresentam vícios. O que pode ser verificado matematicamente.

e) $E(\bar{X}) = E\left(\frac{X_1+X_2}{2}\right) = \frac{1}{2}E(X_1 + X_2) = \frac{1}{2}[E(X_1) + E(X_2)] = \frac{1}{2}[\mu + \mu] = \mu = 900$

f) $E(\bar{Y}) = E\left(\frac{3X_1+2X_2}{5}\right) = \frac{1}{5}E(3X_1 + 2X_2) = \frac{3}{5}E(X_1) + \frac{2}{5}E(X_2) = \frac{3}{5}\mu + \frac{2}{5}\mu = \mu = 900$

g) $E(\bar{Z}) = E\left(\frac{X_1+2X_2}{4}\right) = \frac{1}{4}E(X_1) + \frac{2}{4}E(X_2) = \frac{1}{4}\mu + \frac{2}{4}\mu = \frac{3}{4}\mu = 675$

h) $E(\bar{W}) = E(\max\{X_1, X_2\}) = \frac{1}{16}800 + \frac{3}{16}850 + \frac{5}{16}950 + \frac{7}{16}1000 = 943{,}75$

Dentre os estimadores não tendenciosos apresentados é possível demonstrar também que a variância de $\bar{X}$ é menor do que a variância de $\bar{Y}$.

c) $V(\bar{X}) = V\left(\frac{X_1+X_2}{2}\right) = \frac{1}{4}V(X_1 + X_2) = \frac{V(X_1)+V(X_2)}{4} = \frac{\sigma^2}{2} = \frac{6250}{2} = 3125$

d) $V(\bar{Y}) = V\left(\frac{3X_1+2X_2}{5}\right) = \frac{1}{25}V(3X_1 + 2X_2) = \frac{1}{25}[9V(X_1) + 4V(X_1)] = \frac{13}{25}\sigma^2 = \frac{\sigma^2}{\frac{25}{13}} = \frac{6250}{\frac{25}{13}} = 3250$

Como $2 > \frac{25}{13}$, então $V(\bar{X}) < V(\bar{Y})$. Isso nos diz que o estimador $\bar{X}$ é mais preciso que o estimador $\bar{Y}$.

Na tabela 3 são apresentados dois estimadores para a variância amostral.

Amostras de tamanho 2	Média salarial da amostra $\bar{X} = \frac{X_1+X_2}{2}$ em R$	Estimador 1 $\hat{\sigma}^2 = \frac{\sum(x_i - \bar{X})^2}{n}$	Estimador 2 $s^2 = \frac{\sum(x_i - \bar{X})^2}{n-1}$
{1000, 1000}	1000	0	0
{1000, 800}	900	10000	20000
{1000, 850}	925	5625	11250
{1000, 950}	975	625	1250
{800, 1000}	900	10000	20000
{800, 800}	800	0	0
{800,850}	825	625	1250
{800,950}	875	5625	11250
{850,1000}	925	5625	11250
{850,800}	825	625	1250
{850,850}	850	0	0
{850,950}	900	2500	5000
{950,1000}	975	625	1250
{950,800}	875	5625	11250
{950,850}	900	2500	5000
{950,950}	950	0	0
Médias	900	3125	6250

Tabela 3

A tabela 3 nos mostra que o estimador $\hat{\sigma}^2$, que é uma "cópia" de σ^2, é tendencioso, enquanto o estimador s^2 não é tendencioso.

Observe a tabela 4 que exibe os mesmos cálculos realizados para uma amostra de tamanho 3.

Amostras			$\bar{X} = \frac{X_1 + X_2}{2}$	Estimador 1 $\hat{\sigma}^2 = \frac{\sum(x_i - \bar{X})^2}{n}$	Estimador 2 $s^2 = \frac{\sum(x_i - \bar{X})^2}{n-1}$
1000	1000	1000	1000	0	0
1000	1000	800	933,3333	8888,889	13333,33
1000	1000	850	950	5000	7500
1000	1000	950	983,3333	555,5556	833,3333
1000	800	1000	933,3333	8888,889	13333,33
1000	800	800	866,6667	8888,889	13333,33
1000	800	850	883,3333	7222,222	10833,33
1000	800	950	916,6667	7222,222	10833,33
1000	850	1000	950	5000	7500

1000	850	800	883,3333	7222,222	10833,33
1000	850	850	900	5000	7500
1000	850	950	933,3333	3888,889	5833,333
1000	950	1000	983,3333	555,5556	833,3333
1000	950	800	916,6667	7222,222	10833,33
1000	950	850	933,3333	3888,889	5833,333
1000	950	950	966,6667	555,5556	833,3333
800	1000	1000	933,3333	8888,889	13333,33
800	1000	800	866,6667	8888,889	13333,33
800	1000	850	883,3333	7222,222	10833,33
800	1000	950	916,6667	7222,222	10833,33
800	800	1000	866,6667	8888,889	13333,33
800	800	800	800	0	0
800	800	850	816,6667	555,5556	833,3333
800	800	950	850	5000	7500
800	850	1000	883,3333	7222,222	10833,33
800	850	800	816,6667	555,5556	833,3333
800	850	850	833,3333	555,5556	833,3333
800	850	950	866,6667	3888,889	5833,333
800	950	1000	916,6667	7222,222	10833,33
800	950	800	850	5000	7500
800	950	850	866,6667	3888,889	5833,333
800	950	950	900	5000	7500
850	1000	1000	950	5000	7500
850	1000	800	883,3333	7222,222	10833,33
850	1000	850	900	5000	7500
850	1000	950	933,3333	3888,889	5833,333
850	800	1000	883,3333	7222,222	10833,33
850	800	800	816,6667	555,5556	833,3333
850	800	850	833,3333	555,5556	833,3333
850	800	950	866,6667	3888,889	5833,333
850	850	1000	900	5000	7500
850	850	800	833,3333	555,5556	833,3333
850	850	850	850	0	0
850	850	950	883,3333	2222,222	3333,333
850	950	1000	933,3333	3888,889	5833,333
850	950	800	866,6667	3888,889	5833,333
850	950	850	883,3333	2222,222	3333,333
850	950	950	916,6667	2222,222	3333,333
950	1000	1000	983,3333	555,5556	833,3333
950	1000	800	916,6667	7222,222	10833,33
950	1000	850	933,3333	3888,889	5833,333
950	1000	950	966,6667	555,5556	833,3333

950	800	1000	916,6667	7222,222	10833,33
950	800	800	850	5000	7500
950	800	850	866,6667	3888,889	5833,333
950	800	950	900	5000	7500
950	850	1000	933,3333	3888,889	5833,333
950	850	800	866,6667	3888,889	5833,333
950	850	850	883,3333	2222,222	3333,333
950	850	950	916,6667	2222,222	3333,333
950	950	1000	966,6667	555,5556	833,3333
950	950	800	900	5000	7500
950	950	850	916,6667	2222,222	3333,333
950	950	950	950	0	0
Média			900	4166,667	6250

Tabela 4

Observe que $E(s^2) = \sigma^2$ para as amostras de tamanho 2 e 3, então qual é o valor de $E(\hat{\sigma}^2)$?

$$E(s^2) = \sigma^2$$
$$E\left(\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{X})^2\right) = \sigma^2$$
$$\frac{1}{n-1}E\left(\sum_{i=1}^{n}(x_i - \bar{X})^2\right) = \sigma^2$$
$$E\left(\sum_{i=1}^{n}(x_i - \bar{X})^2\right) = (n-1)\sigma^2$$
$$\frac{1}{n}E\left(\sum_{i=1}^{n}(x_i - \bar{X})^2\right) = \frac{1}{n}(n-1)\sigma^2$$
$$E\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^2\right) = \frac{n-1}{n}\sigma^2$$
$$E(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2$$

Isso explica o porquê de $\hat{\sigma}^2$ ser metade do valor do parâmetro na tabela 3 e $\frac{2}{3}$ do parâmetro na tabela 4.

Abaixo são apresentados dois gráficos que ilustram a distribuição de $\bar{X}$ para amostras de tamanho 2 e 3, respectivamente

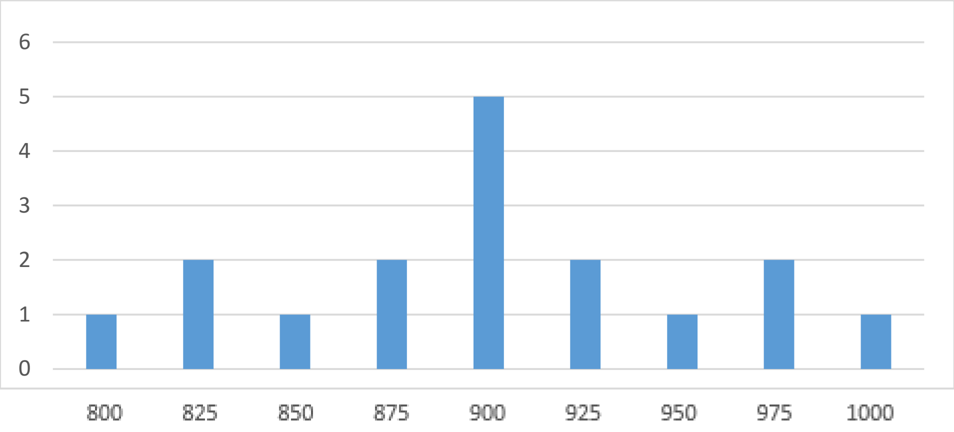


Gráfico 1: Amostras de tamanho 2

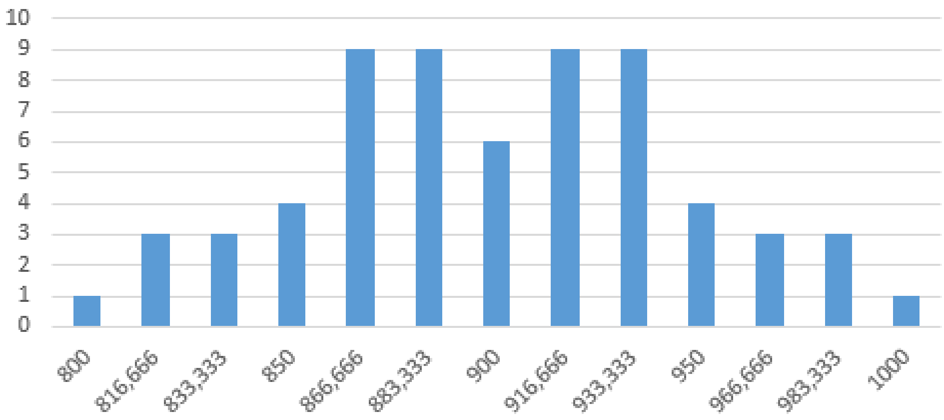


Gráfico 2: Amostras de tamanho 3

Com base nos gráficos o que podemos afirmar que acontece com a variância amostral quando o número n de elementos da amostra aumenta?

Etapa 3: Discutir a importância da Lei Forte dos Grandes Números para estimar parâmetros no contexto da Inferência Clássica.

Considerando ainda o problema de estimar o valor médio do salário dos trabalhadores. Na grande maioria dos casos teremos apenas uma amostra e não todas as possíveis conforme foi exibido na etapa anterior. Assim, devemos pensar: o que acontece com o valor de $\bar{X}$ quando aumentamos o número de elementos da amostra?

Se $X_1, X_2, \ldots, X_n$ são variáveis aleatórias que constituem uma amostra do salário dos trabalhadores (podendo haver reposição), então o estimador $\bar{X}$ que como visto é não viciado e preciso que estima o parâmetro média populacional μ tende a se aproximar do valor do parâmetro conforme n aumenta.

Este resultado é sustentado pela Lei Forte dos Grandes Números que diz:

Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes e identicamente distribuídas com

$E(X_n) = \mu$. Então,

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow{q.c} E(X)$$

Onde $\xrightarrow{q.c}$ indica a convergência quase certa. Isto é, quando $n \to \infty$, $\frac{X_1+\cdots+X_n}{n} \to E(X)$ com probabilidade 1. Isso não quer dizer que $\frac{X_1+\cdots+X_n}{n}$ convergirá para $E(X)$ sempre que $n \to \infty$. De fato, tal convergência pode não ocorrer, mas a probabilidade de que não ocorra a convergência é 0.

Assim, no problema proposto em que devemos estimar o número de bolinhas vermelhas na piscina, podemos tomar, por exemplo, várias amostras de mesmo tamanho $n < N$ e calcular a média amostral (Neste caso, a variável aleatória X assumirá o valor 1 quando a bola for vermelha e 0 quando for branca, assim $\bar{X} = \hat{p}$, onde $\hat{p}$ é um estimador para a proporção de bolas vermelhas). Pois, cada média amostral é uma variável aleatória e pela Lei Forte temos que quando o número de amostras aumenta a média das médias amostrais converge quase certamente para o parâmetro μ.

Planejamento do segundo encontro

Conteúdos: Lei Forte dos Grandes Números, Probabilidade no contínuo e Integral de Monte Carlo.
Habilidades: Compreender o papel da Lei Forte dos Grandes Números em uma abordagem frequentista de probabilidade.
Tempo: 03:00h

Etapa 1: Probabilidade no contínuo

Atividade 1: Duas pessoas marcaram de se encontrar entre 12h e 13:40h (100 minutos), para que uma não fique aguardando pela outra por muito tempo, estipulou-se que o tempo de espera pela outra pessoa será de exatamente 15 minutos, uma vez que a outra pessoa não apareça nesse período de tempo ou até as 13:40h, o que ocorrer primeiro, a pessoa vai embora. Qual é a probabilidade que essas pessoas se encontrem, dado que as duas compareceram no local marcado dentro do horário previsto?

Solução

Sejam X e Y as variáveis aleatórias que medem o tempo que o primeiro amigo e o segundo amigo, respectivamente, demoram a chegar, a partir de 12h.

Assim, temos que

$$X \sim U[0,1] \text{ e } Y \sim U[0,1]$$

Para resolver a questão, basta calcular $P(|X - Y| \leq 0{,}15)$.

Seja P um ponto no plano cartesiano, tal que $P = (X, Y)$, desta forma se o primeiro amigo chega às 12:30h e o segundo as 12:46h, $P = (0{,}30, 0{,}46)$

Assim, o espaço amostral pode ser descrito como a área de um quadrado de lado 1, tal como ilustrado na figura abaixo

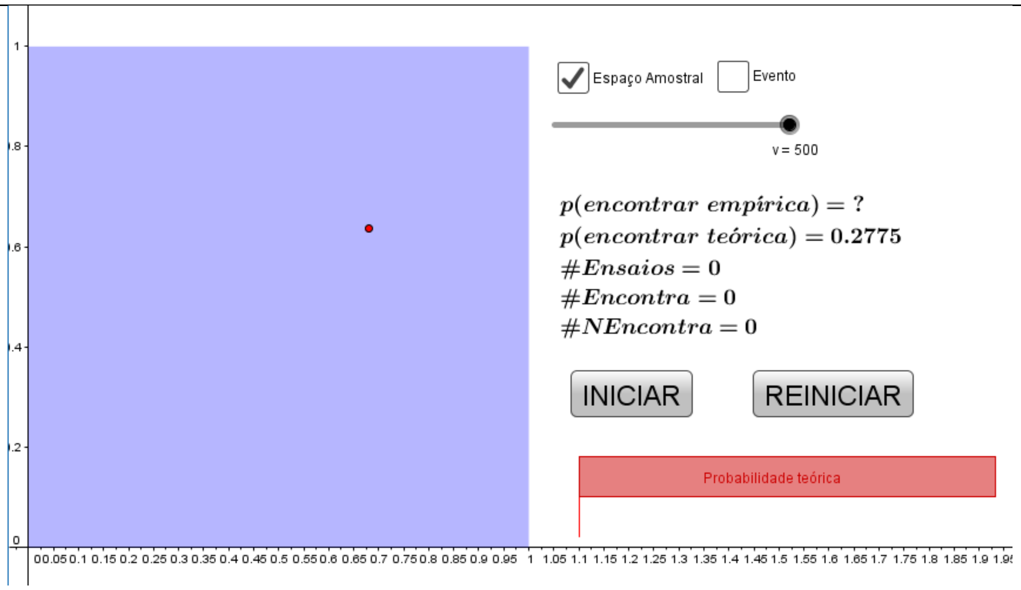


Figura 1: Probabilidade no contínuo

e o evento desejado corresponde a área da figura delimitada pelo quadrado e pelas retas $y = x - 0{,}15$ e $y = x + 0{,}15$. Pois, como $|X - Y| \leq 0{,}15$, então

$$X \geq Y - 0{,}15 \text{ e } X \leq Y + 0{,}15$$

Assim, se realizarmos um experimento em que uma grande quantidade de pontos $P = (X, Y)$ são plotados aleatoriamente no gráfico, com $X \sim U[0,1]$ e $Y \sim U[0,1]$. Considerando a variável aleatória W tal que

$$W = \begin{cases} 1, se\ P\ pertence\ a\ região\ correspondente\ ao\ evento \\ 0, caso\ contrário \end{cases}$$

Pela Lei Forte dos Grandes Números temos que

$$\frac{W_1 + W_2 + \cdots + W_n}{n} \xrightarrow{q.c} E(W)$$

Mas, $E(W) = 1 \times p + 0 \times (1 - p) = p$

onde p é a probabilidade de que as duas pessoas se encontrem.

Assim, podemos escrever

$$\frac{W_1 + W_2 + \cdots + W_n}{n} \xrightarrow{q.c} p$$

Note que W é variável aleatória independente pois a ocorrência de um ponto não afeta a probabilidade de ocorrência do outro ponto, além disso, W é identicamente distribuída com distribuição Binomial.

Etapa 2: Integral de Monte Carlo

Nesta etapa, apresentaremos um método para cálculo de integrais utilizando a Lei Forte dos Grandes Números, tal método é conhecido como Integral de Monte Carlo e a ideia central é que, a partir de muitas plotagens aleatórias de pontos $(x, f(x))$, sejamos capazes de estimar a integral de qualquer função integrável em um determinado intervalo.

Suponha desconhecido o valor da integral de $f(x) = x^2$ no intervalo $[0,1]$.

$$c = \int_0^1 f(x)\, dx$$

Se tomarmos um ponto $P = (x, f(x))$ tal que $X \sim U[0,1]$, então $Y = f(x)$ será também uma variável aleatória com função densidade

$$h(x) = \begin{cases} 1, 0 \leq x \leq 1 \\ 0, caso\ contrário \end{cases}$$

como $E(x) = \int_{-\infty}^{\infty} x f(x)\, dx$, então $E(Y) = E\big(f(x)\big) = \int_{-\infty}^{\infty} f(x) h(x)\, dx = \int_0^1 f(x)\, dx = c$

Logo, $f(x)$ é variável aleatória independente, identicamente distribuída e com média c. Portanto, pela Lei Forte dos Grandes Números

$$\frac{Y_1 + Y_2 + \cdots Y_n}{n} \overset{q.c}{\longrightarrow} E(Y)$$

Ou seja,

$$\frac{f(x_1) + f(x_2) + \cdots f(x_n)}{n} \overset{q.c}{\longrightarrow} E\big(f(x)\big) = c$$

Assim, para estimar a integral de uma função $f(x)$ em um intervalo $[0,1]$ basta tomar uma amostra aleatória de pontos P e calcular o valor médio dos valores da imagem.

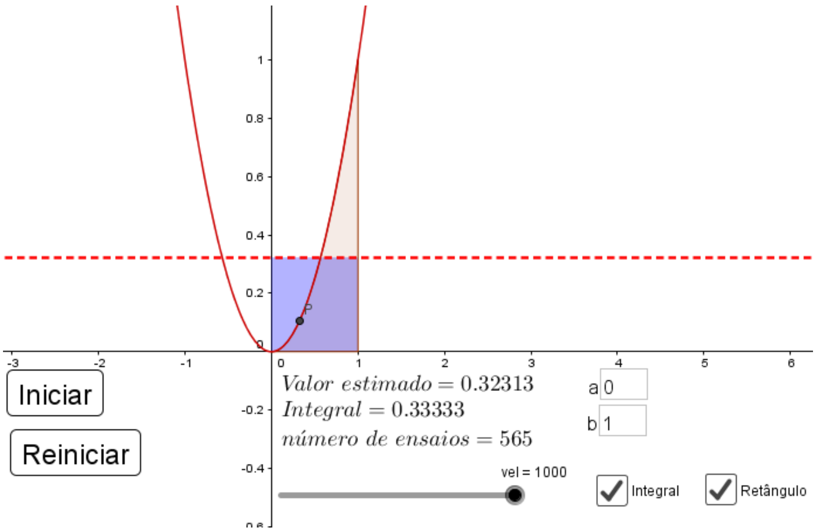


Figura 2: Estimar valor de f(x)

Atividade 2: Estime o valor da integral da função $f(x) = x^2$ no intervalo $[0,1]$.

Atividade 3: Estime o valor da integral da função $f(x) = \sqrt{senh^2(x)}$ no intervalo $[0,1]$.

Etapa 3: Estendendo o método para qualquer intervalo $[a, b], a \in \mathbb{R}$ e $b \in \mathbb{R}$.

Atividade 3: Estime o valor da integral da função $f(x)$ no intervalo $[a, b]$.

Seja $c = \int_a^b f(x)\, dx$, se tomarmos a variável $y = \frac{x-a}{b-a}$, teremos que $dy = \frac{1}{b-a} dx$ e, além disso, quando $x = a$, tem-se que $y = 0$ e quando $x = b$, obtém-se $y = 1$. Note também que, $x = y(b-a) + a$ e $dx = (b-a)dy$. Assim, temos que

$$c = \int_a^b f(x)\, dx = \int_0^1 f(y(b-a) + a)(b-a)\, dy$$

Portanto, se tomarmos o ponto $P = (x, f(x(b-a) + a)(b-a))$, onde $x \sim U[0,1]$ e simularmos com o software diversas plotagens do ponto P, a área da integral da função f definida no intervalo $[a, b]$ será igual a área de um retângulo de base 1 e altura igual a média de $f(x(b-a) + a)(b-a)$.

Isso ocorre pois, se $x \sim U[0,1]$, então a função densidade de $f(x(b-a) + a)(b-a)$ é dada por

$$h(x) = \begin{cases} 1, 0 \leq x \leq 1 \\ 0, caso\ contrário \end{cases}$$

Logo, $g(x) = f(x(b-a) + a)(b-a)$ é variável aleatória, e das propriedades de Esperança Matemática temos que

$$E\big(g(x)\big) = \int_0^1 g(x)h(x)\, dx = \int_0^1 g(x)\, dx = c$$

Assim, $g(x_1), g(x_2), \dots, g(x_n)$ são variáveis aleatórias independentes e identicamente distribuídas com média c. Portanto, pela Lei Forte dos Grandes Números

$$\frac{1}{n}\sum_{i=1}^{n} g(x_i) \xrightarrow{q.c} E\big(g(x)\big) = c$$

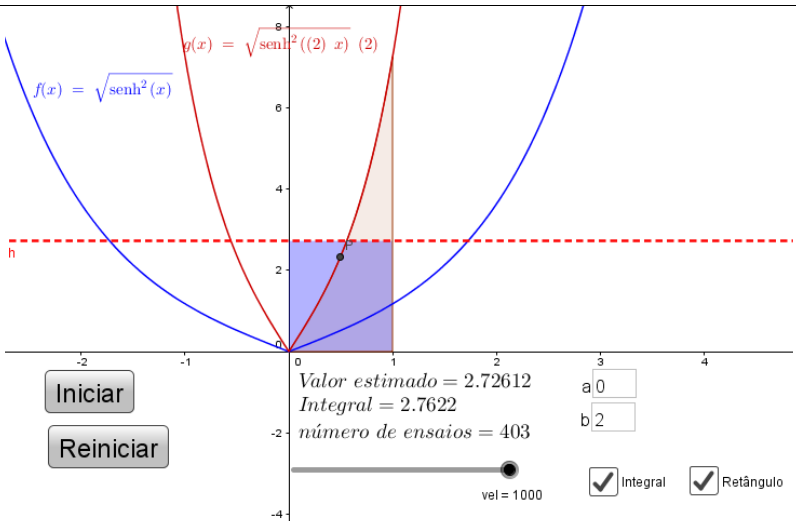


Figura 3: Estimar valor de g(x)

Na imagem a função $g(x) = f(x(b-a)+a)(b-a)$.

Verifique com auxílio do applet *Monte Carlo – Integral 3* que quanto maior for o número de pontos plotados mais o valor estimado se aproximará do valor real da integral. Assim, a teoria da estimação se mostra útil não apenas em áreas regidas pela incerteza, mas também em campos determinísticos.

Atividade 4: Estime o valor da integral da função $f(x) = \sqrt{senh^2(x)}$ no intervalo $[1,3]$.

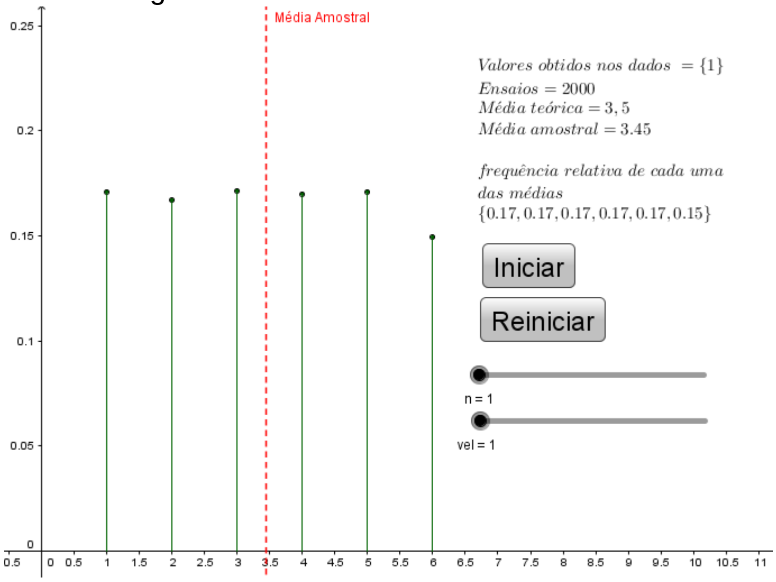
Planejamento do terceiro encontro

Conteúdos: Teorema Central do Limite, Intervalos de Confiança, Nível de significância e Margem de erro.
Habilidades: Compreender o papel do Teorema Central no processo de estimação de parâmetros.
Tempo: 03:00h

Etapa 1: Observar empiricamente que a distribuição de $\bar{X}$ se aproxima da distribuição Normal, quando o tamanho da amostra aumenta.

Atividade 1: Utilizando o applet *Lançamento de dados observar média*, aumente o número de elementos na amostra e observe o aspecto da distribuição de $\bar{X}$.

Figura 1: 2000 ensaios com amostras de tamanho 1



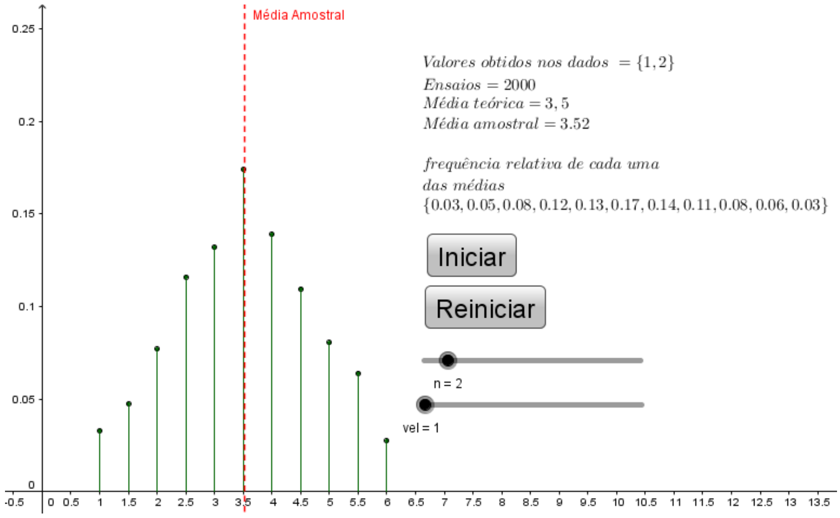


Figura 2: 2000 ensaios com amostras de tamanho 2

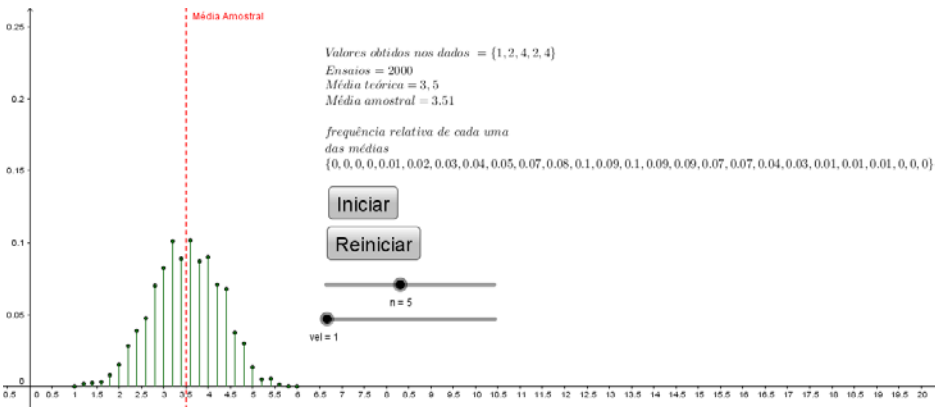


Figura 3: 2000 ensaios com amostras de tamanho 5

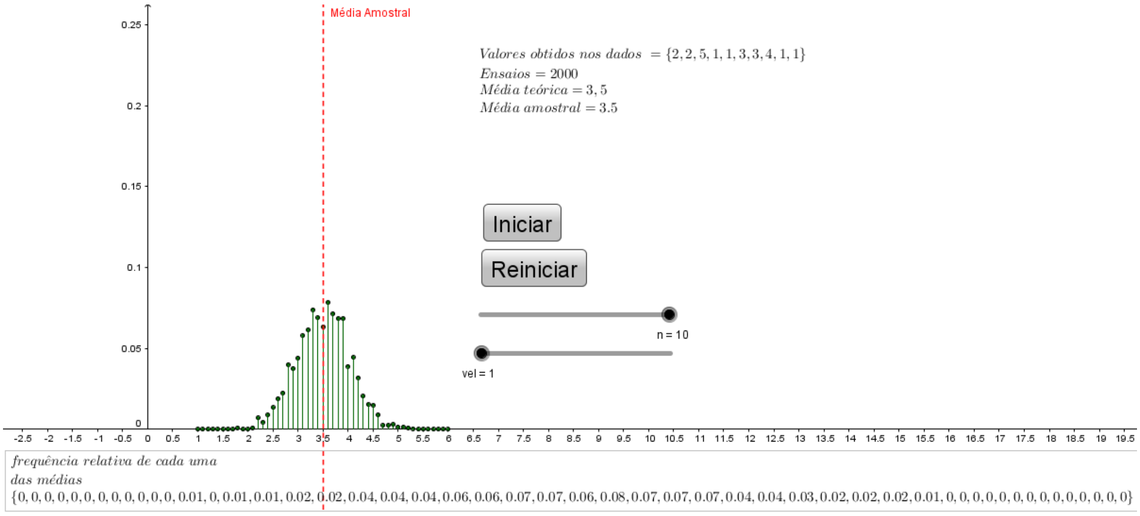


Figura 4: 2000 ensaios com amostras de tamanho 10

A distribuição de $\bar{X}$ se assemelha a alguma função que você conheça? Qual? Quando aumentamos o tamanho da amostra o que ocorre com os possíveis valores de $\bar{X}$? Se o número n de elementos da amostra tender ao infinito, o que deve ocorrer com os valores possíveis de $\bar{X}$? A variabilidade aumenta ou diminui em torno da média quando n aumenta?

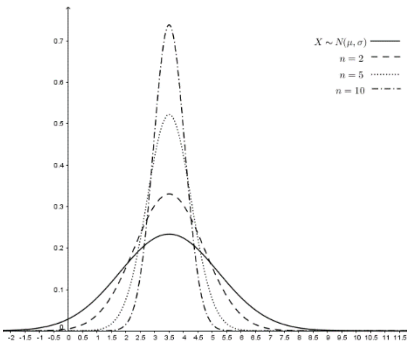


Figura 5: Distribuição da média amostral

Note que, apesar da variável aleatória $\bar{X}$ ser discreta, quando o número de elementos da amostra aumenta, a distribuição se torna cada vez "mais densa". Assim, quanto maior o número de elementos n da amostra, mais a distribuição de $\bar{X}$ se aproxima de uma curva Normal. Isso é importante, pois, em muitos casos, é menos trabalhoso realizar cálculos e deduções a partir de uma curva Normal do que em uma distribuição discreta.

Além disso, note que quando n aumenta a variabilidade da distribuição de $\bar{X}$ diminui. De modo que, quando $n \to \infty$, a distribuição de $\bar{X}$ tende a se parecer cada vez mais com a reta $x = 3{,}5$.

De fato, $\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$. Como já vimos, pela Lei Forte dos Grandes Números, quando $n \to \infty$, $\bar{X}$ converge quase certamente para μ. Além disso, também podemos demonstrar que o desvio-padrão de $\bar{X}$ é $\frac{\sigma}{\sqrt{n}}$.

Lembre-se que $\sigma^2 = E(X^2) - E(X)^2$. Assim, como a amostra aleatória é independente e identicamente distribuída segue que

$$E(\bar{X}^2) - E(\bar{X})^2 = \frac{n}{n^2}E(X_1^2) - \frac{n}{n^2}E(X_1)^2 = \frac{1}{n}(E(X_1^2) - E(X_1)^2) = \frac{\sigma^2}{n}$$

dessa forma temos que o parâmetro desvio-padrão da distribuição de $\bar{X}$ é

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

Etapa 2: Discutir o papel do Teorema Central do Limite para o cálculo das probabilidades.

Quando desenvolvemos um modelo matemático para estudar estruturas fenomenológicas regidas pela incerteza é comum que nesse modelo apareçam sequências de variáveis aleatórias, tais como as que vimos nas aulas anteriores.

No estudo do comportamento dessa sequência de variável podemos identificar qual é a distribuição de probabilidade que melhor se adéqua a sequência de variáveis aleatórias fornecidas.

No entanto, há uma ferramenta valiosíssima que pode ser utilizada para tratar da soma ou da média dessa sequência de variáveis aleatórias, se as variáveis aleatórias atenderem a determinadas condições. Tornando uma tarefa muito menos árdua estudar o comportamento de sequencias de variáveis aleatórias como essas. Essa ferramenta se chama Teorema Central do Limite e possui várias versões, cada qual com uma série de benefícios e restrições para serem aplicadas. A seguir apresentaremos uma versão desse teorema que atende a muitos aspectos de situações cotidianas.

Teorema Central do Limite (TCL)

Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes e identicamente distribuídas com média $\mu < \infty$ e variância σ^2, tal que $0 < \sigma^2 < \infty$. Então, considerando $S_n = X_1 + \cdots + X_n$ temos que

$$Z = \frac{S_n - E(S_n)}{DP(S_n)} \xrightarrow{D} N(0,1)$$

ou

$$Z = \frac{S_n - n\mu}{\sqrt{n}\sigma} \xrightarrow{D} N(0,1)$$

Onde $\xrightarrow{D}$ indica a convergência em distribuição. Isto é, a distribuição dada por $\frac{S_n - n\mu}{\sigma\sqrt{n}}$ tende para a distribuição $N(0,1)$ quando $n \to \infty$.

Assim, multiplicando Z por $\frac{\frac{1}{n}}{\frac{1}{n}}$, temos

$$Z = \frac{\frac{S_n}{n} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{D} N(0,1)$$

Dessa forma podemos reescrever Z de maneira conveniente, o que nos permite aproximar a soma ou a médias de variáveis aleatórias que pertencem a uma mesma distribuição à distribuição Normal. Em geral, para $n \geq 30$ essa é uma boa aproximação. Contudo, quanto maior o valor de n melhor é a aproximação.

Lembre que a independência de variáveis aleatórias ocorre quando a ocorrência de uma não altera a probabilidade de ocorrência da outra, além disso, atente também para o fato de que para aplicar esta versão do TCL, será necessário que todas as variáveis aleatórias pertençam a uma mesma distribuição.

Atividade 2: Considere que uma moeda honesta seja lançada 100 vezes. Qual a probabilidade de se obter no máximo 40 caras?

Nesse exemplo, a variável aleatórias X conta o número de caras obtidas em 100 lançamentos. Então para responder essa questão teríamos que calcular

$$P(X=0)+P(X=1)+\cdots+P(X=40)$$

O que envolve diversos cálculos combinatórios.
Por outro lado, sabemos que
$E(X)=np=100\times\frac{1}{2}=50$, e que
$DP(X)=\sqrt{np(1-p)}=\sqrt{100\times\frac{1}{4}}=5$
Assim, se tomarmos a variável aleatória $Z=\frac{X-E(X)}{DP(X)}$, temos que $Z \sim N(0{,}1)$, e tendo em mente a necessidade do fator de correção de continuidade, temos

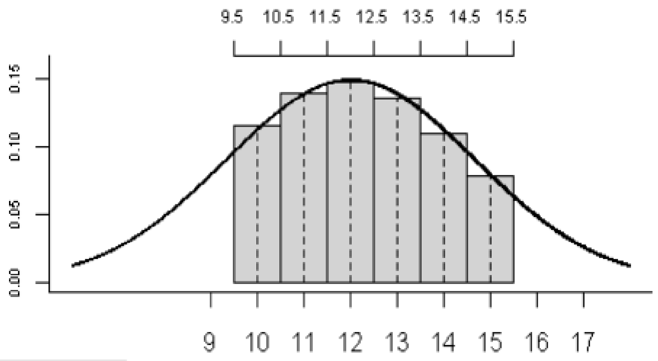
$$P(X\leq 40{,}5)\approx P\left(\frac{X-E(X)}{DP(X)}\leq\frac{40{,}5-50}{5}\right)=P(Z\leq -1{,}9)=0{,}0287$$

A última igualdade é obtida através da tabela Z (anexada na última página).
Se calculássemos $P(X\leq 40)$ obteríamos como resultado 0,02844. Um valor razoavelmente próximo.

Fator de correção de continuidade

O fator de correção de continuidade é utilizado quando aproximamos uma distribuição discreta (Binomial) à uma distribuição contínua (Normal). Observe na figura a seguir uma representação da aproximação de uma distribuição Binomial à uma distribuição Normal

Figura 6: Importância do fator de continuidade
Fonte: PINHEIRO, J. et. al, 2011, p.131



Note que se desejarmos calcular $P(10\leq X\leq 15)$, com $a\in\mathbb{Z}$ e $b\in\mathbb{Z}$, pela distribuição

Binomial, deveremos calcular

$$P(X = 10) + P(X = 11) + \cdots + P(X = 15)$$

Pela figura 1 apresentada, isso equivale a calcular a soma das áreas de todos os retângulos de base 1 e altura $P(X = x)$. Por outro lado, se optarmos por aproximar a Binomial à curva Normal, o cálculo deverá ser feito através da integral

$$\int_{10}^{15} f(x)dx$$

Contudo, veja na figura 1 que tanto a metade (da esquerda) do retângulo cujo o ponto médio da base é 10 quanto a metade (da direita) do retângulo cujo o ponto médio da base é 15, não são calculados pela integral definida acima. A fim de obter uma aproximação melhor para o valor da integral, basta subtrairmos 0,5 do ponto a extrema esquerda do intervalo de integração e somarmos 0,5 ao ponto a extrema direita desse intervalo. Assim, para uma melhor aproximação da Binomial à Normal, poderemos calcular

$$\int_{9,5}^{15,5} f(x)dx$$

Etapa 3: Discutir o papel do Teorema Central do Limite para obtenção de uma estimativa intervalar.

Vimos nas aulas anteriores o que é um estimador e sua variabilidade. O que nos fornece uma ideia sobre qual é o valor do parâmetro de interesse. Porém, é desejável que consigamos construir um intervalo em torno do valor estimado que possua uma determinada probabilidade de conter o parâmetro. Para construir esse intervalo, conhecido como Intervalo de confiança, utilizaremos o TCL.

Atividade 3: Suponha que para estimar a média dos valores obtidos no lançamento de um dado equiprovável tenham sido realizados 90 ensaios, a média dessa amostra é 3,31 e o desvio-padrão é 2,92.

Queremos construir um intervalor entorno da média amostral que possua 95% de chance de conter o parâmetro estimado. A probabilidade do intervalo conter o parâmetro é chamada de Nível de Confiança e é denotada por $(1 - \alpha)$, e α é chamado Nível de Significância. Isto é, queremos calcular o valor E para o qual temos

$$P(\mu \in [\bar{X} - E, \bar{X} + E]) = 0,95$$

Como X é uma variável aleatória independente e identicamente distribuída, temos que

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

Então, queremos determinar

$$P\left(-Z_{0,025} \leq Z \leq Z_{0,025}\right) = 0,95$$

Pela Tabela Z anexada ao final do texto, temos que $Z_{0,025} = 1,96$

Mas, $Z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$. Assim, podemos escrever

$$P\left(-1{,}96 \le \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \le 1{,}96\right) = P\left(-1{,}96\frac{\sigma}{\sqrt{n}} \le \bar{X}-\mu \le 1{,}96\frac{\sigma}{\sqrt{n}}\right) =$$
$$= P\left(\bar{X} - 1{,}96\frac{2{,}92}{\sqrt{90}} \le \mu \le \bar{X} + 1{,}96\frac{2{,}92}{\sqrt{90}}\right) = 0{,}95$$
$$P\left(3{,}31 - 1{,}96\frac{2{,}92}{\sqrt{90}} \le \mu \le 3{,}31 + 1{,}96\frac{2{,}92}{\sqrt{90}}\right) = 0{,}95$$

Assim, há 95% de chance do intervalo $\left[3{,}31 - 1{,}96\frac{2{,}92}{\sqrt{90}}, 3{,}31 + 1{,}96\frac{2{,}92}{\sqrt{90}}\right] = [-2{,}70, 3{,}91]$ conter o parâmetro estimado.

Para ajudar a compreender essa afirmação utilize o applet *Amostras aleatórias*. Esse applet simula 100 amostras aleatórias de tamanho 10 e verifica quantos intervalos de confiança (95%) contém o parâmetro média estimado. Neste applet a variável aleatória assume valores de 1 a 6, como no lançamento de um dado.

Denotaremos por $E = Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$ a margem de erro amostral.

Atividade 4: Considere um jogo em que serão extraídas duas bolas de uma urna, sem reposição. As bolas estão numeradas de 1 a 10 e são produzidos tantos bilhetes de dois números quantos forem os compradores, de maneira que todas as combinações possíveis sejam contempladas em igual proporção. Assim, se 90 pessoas comprarem os bilhetes, qual a melhor estimativa pontual do número de ganhadores do sorteio? Construa um intervalo de confiança, com nível de confiança de 95% para o número de ganhadores do sorteio.

Solução

A probabilidade de obter um bilhete vencedor é dada por $P = \frac{1}{\binom{10}{2}} = \frac{1}{45}$.
Como o número de compradores é 90 e X tem distribuição binomial, podemos esperar que $E(X) = Np = 90.\frac{1}{45} = 2$ pessoas ganhem.
A variância é dada por $V(X) = Np(1-p) = 90.\frac{1}{45}.\frac{44}{45} = 1{,}956$.
Com isso temos que,

$$Z = \frac{X-2}{\sqrt{1{,}956}} \sim N(0{,}1)$$

Como o intervalo de confiança tem nível de confiança de 95%, então $Z_{\frac{\alpha}{2}} = Z_{0{,}025} = 1{,}96$.

$$P(Z_{0{,}025} \le Z \le Z_{0{,}025}) = 0{,}95$$
$$P(-1{,}96 \le Z \le 1{,}96) = 0{,}95$$

Substituindo Z, temos

$$P\left(-1{,}96 \leq \frac{X-2}{\sqrt{1{,}956}} \leq 1{,}96\right) = 0{,}95$$

Daí segue que

$$P\left(2 - 1{,}96\sqrt{1{,}956} \leq X \leq 2 + 1{,}96\sqrt{1{,}956}\right) = 0{,}95$$

Substituindo os valores e fazendo os cálculos, obtemos

$$P(-0{,}741 \leq X \leq 4{,}741) = 0{,}95$$

Logo, há 95% de chance de que o número de vencedores desse jogo pertença ao intervalo $[-0{,}741, 4{,}741]$, ou seja, existe 95% de chance que tenhamos algo entre 0 e 4 vencedores, inclusive.

Atividade 5: Suponha uma piscina com 100.000 bolinhas brancas ou vermelhas. Estime o número de bolinhas vermelhas com base numa amostra de tamanho 1000 bolas, dentre as quais 400 são brancas e 600 são vermelhas. Estime um intervalo de confiança de 95% para a quantidade de bolas vermelhas.

Solução:

Como das 1000 bolas da amostra 600 são vermelhas, então a proporção amostral de bolas vermelhas é $\hat{p} = 0{,}6$.

Seja X a variável aleatória que conta o número de bolas vermelhas, então $X \sim B(100000, p)$, em que p é o parâmetro que indica a proporção de bolas vermelhas.

Aproximando a distribuição Binomial pela Normal temos do TCL que

$$Z = \frac{p - \hat{p}}{\frac{\sqrt{pq}}{n}}$$

Assim, segue que $Z \sim N(0{,}1)$. Logo, se quisermos construir um intervalo de confiança com um nível de confiança de 95% temos

$$P(-Z_{0{,}025} \leq Z \leq Z_{0{,}025}) = 0{,}95$$

$$\left(-1{,}96 \leq \frac{p-\hat{p}}{\frac{\sqrt{pq}}{n}} \leq 1{,}96\right) = P\left(-1{,}96\frac{\sqrt{pq}}{n} \leq X - \hat{p} \leq 1{,}96\frac{\sqrt{pq}}{n}\right) =$$

$$= P\left(\hat{p} - 1{,}96 \times \frac{\sqrt{\frac{6}{10} \times \frac{4}{10}}}{\sqrt{1000}} \leq X \leq \hat{p} + 1{,}96 \times \frac{\sqrt{\frac{6}{10} \times \frac{4}{10}}}{\sqrt{1000}}\right) = 0{,}95$$

$$P(0{,}6 - 0{,}03 \leq X \leq 0{,}6 + 0{,}03) = 0{,}95$$

Dessa forma, há 95% de chance de que o intervalo $[0{,}57, 0{,}63]$, contenha o parâmetro desejado. Isto é, se tomarmos n amostras nessas mesmas condições, pela Lei Forte dos Grandes Números, podemos esperar que 95% dos intervalos construídos sob as condições mencionadas no exercícios contenham o parâmetro de interesse.

7 CONCLUSÃO

A partir de nosso estudo a respeito de como a Lei Forte dos Grandes Números e o Teorema Central do Limite podem contribuir para a construção de um raciocínio estatístico, percebemos que há uma urgência em incluir no ensino de Estatística ao menos alguns conceitos fundamentais para o estudo da Estatística Inferencial.

Desde o século XIX, a importância do raciocínio inferencial tem sido ressaltada no estudo do comportamento da sociedade, das políticas públicas, do comércio e de outras áreas. O crescimento da Estatística foi tão grande a partir daí que, atualmente, no século XXI, podemos afirmar que praticamente todas as decisões importantes, que afetam um grande número de pessoas, são tomadas considerando as contribuições advindas das análises estatísticas.

Constatada a importância da Estatística para o desenvolvimento da sociedade, nos questionamos: Por que a Estatística permanece negligenciada no Ensino Básico? Ora, um dos objetivos da Escola é promover a formação do cidadão, ensinando-lhe ferramentas que o auxiliem no seu dia a dia.

Vivemos atualmente em um século caracterizado pelo historiador Eric Hobsbawm de era das incertezas, devido a uma grande diversidade de questões que se colocam a respeito da humanidade e que permanecem sem resposta. Embora a Estatística não seja a única ciência a fornecer contribuições às grandes problematizações de nosso novo milênio, ela tem certamente muito a iluminar várias searas do conhecimento humano, pois, como vimos, recorrendo apenas a dois teoremas, muita coisa já pode ser dita a respeito das inquietações da sociedade atual, seja no campo econômico, político, gestão de recursos, científico, etc.

Acreditamos que o processo de formação educacional do cidadão precisa contemplar ferramentas que lhe possibilitem “ler” o mundo e consequentemente que aumentem as chances de que ele tome as melhores decisões.

Não incluir o estudo da Estatística Inferencial, ainda que de forma incipiente no Ensino Básico, é, ao nosso ver, negar ao estudante a possibilidade de compreender e interpretar o mundo ao seu redor.

Além disso, se desejamos debater nas escolas as teorias estatísticas que objetivam a tomada de decisão sob incerteza, fica evidente com o que foi discutido nesta dissertação a importância de se trabalhar a Probabilidade em conjunto com a Estatística, uma vez que a construção do raciocínio a respeito de um dos aspectos mais relevantes da Estatística para este fim, a Inferência, depende da compreensão e das ferramentas advindas da teoria das Probabilidades e seu diálogo com a Estatística.

Como procuramos evidenciar, é importante não se deter apenas no conceito de Probabilidade Clássica, atrelada ao conceito de equiprobabilidade e à Análise Combinatória, já que poucos são os problemas reais que recaem nessa estrutura. A abordagem frequentista, por outro lado, possibilita estender o conhecimento a uma grande variedade de situações e, além disso, favorece o desenvolvimento de um raciocínio inferencial, que tem como seus principais pilares a Lei Forte e o Teorema Central do Limite.

Esses dois Teoremas Limite contribuem para a construção de um raciocínio estatístico em que as decisões a serem tomadas não são completamente corretas, tal como ocorre em um raciocínio matemático. Pelo contrário, o raciocínio estatístico evidencia justamente a impossibilidade de se afirmar que uma determinada proposição é verdadeira ou falsa, mas a Estatística fornece ferramentas que possibilitam mensurar o grau de incerteza de uma afirmação. Isso por si só já auxilia o estudante a compreender que há uma grande diferença entre o pensamento estatístico e o pensamento matemático.

Contrariamente a isso, o modelo atual de ensino de Estatística, em nossa opinião, pouco ou nada contribui para que a essência do raciocínio estatístico seja compreendida, uma vez que esse se baseia apenas em uma pequena parte da Estatística Descritiva, cujo tratamento quase sempre aritmético e destituído de significação dos objetos quase nada contribui para o Letramento Estatístico.

Nosso estudo pretendeu evidenciar que os teoremas limite não apenas auxiliam na construção de um raciocínio estatístico que visa à tomada de decisão em uma situação de incerteza, como também viabilizam o estudo de diversas situações cotidianas. Para tal, necessitam da compreensão e da utilização de conceitos básicos que fazem parte da Estatística, tal como população, amostra, parâmetros e estimador. Estes por sua vez abrem espaço para uma discussão sobre as medidas de tendência central e de dispersão, dentro de um contexto no qual cada uma dessas medidas torna-se indispensável para o estudo do fenômeno.

Assim, a aplicação de conceitos como média, variância, etc, ganha um sentido real, deixando de pertencer apenas a uma lista de fórmulas e métodos que devem ser seguidos para se determinar um número. Afinal, o principal interesse, para o raciocínio estatístico, não está em determinar números e, sim, em estimar os valores que não podem ser calculados.

Apesar da complexidade inerente aos Teoremas Limite abordados aqui, esperamos ter deixado aos interessados no assunto sugestões não dogmáticas e não exaustivas que permitam ao estudante alcançar a compreensão das ideias mais fundamentais e importantes dos referidos teoremas, por meio das simulações braçais ou computacionais. Confiamos, portanto, na sensibilidade do professor para desenvolver questões outras de relevância a serem discutidas e tratadas em sala de aula, com a mesma finalidade de promover o Letramento Estatístico.

Finalmente, esperamos que a leitura dessa dissertação tenha possibilitado ao leitor refletir sobre o papel crucial que os Teoremas Limite têm para a Estatística, bem como o relevante papel que desempenhariam caso fossem tratados de forma introdutória no âmbito da Educação Estatística nas Escolas Básicas.

REFERÊNCIAS BIBLIOGRÁFICAS

ALVARADO, Hugo; BATANERO, Carmen. *Significado Del Teorema Central Del Limite En Textos Universitarios De Probabilidad Y Estadistica.* Estudios Pedagógicos XXXIV, Nº 2: 7-28, 2008.

ALVARADO, Hugo; BATANERO, Carmen. *El Significado Del Teorema Central Del Límite: Evolución Histórica A Partir De Sus Campos De Problemas.* Investigación en Didáctica de las Matemáticas (pp. 13-36), 2005.

ARAUJO, Péricles; IGLIORI, Sonia; *O Problema Epistemológico da Probabilidade.* Caderno de Física da UEFS 11, (01 e 02): 57-75, 2013.

BATANERO, Carmen; TAUBER, Liliana M.; SÁNCHEZ, Victória. *Students' Reasoning About The Normal Distribution.*

BACHELARD, Gaston. *A formação do espírito científico: Contribuições para uma psicanálise do conhecimento*. Rio de Janeiro: Contraponto, 2005.

BENNETT, Deborah J. *Aleatoriedade.* São Paulo: Martins Fontes, 2003.

BERNOULLI, Jacobi. *Ars Conjectandi : Opus Posthumum. Accedit Tractatus De Seriebus Infinitis, Et Epistola Gallicè Scripta De Ludo Pilae Reticularis,* 1713.

BRANCO, João; MARTINS, Maria E. Graça. *Literacia Estatística.* Educação e Matemática, número 69, 2002.

CAZORLA, Irene Mauricio; CASTRO, Franciana Carneiro. *O Papel da Estatística Na Leitura Do Mundo: O Letramento Estatístico,* 2008.

CHANCE, Beth. *Components of Statistical Thinking and Implications for Instruction and Assessment.* Journal of Statistics Education, Volume 10, Number 3, 2002.

COLADARCI, Theodore; COBB, Casey D. *Fundamentals of Statistical Reasoning in Education.* Ed. 4. EstadosUnidos: Wiley, 2014.

DELMAS, Robert. *Statistical Literacy, Reasoning and Learning: A Commentary.* Journal of Statistics Education, Volume 10, Number 3, 2002.

DOWDY, Shirley; WEARDEN, Stanley; CHILKO, Daniel. *Statistics for Research.* Ed. 3. New Jersey: Wiley, 2004.

FISHER, Hans.*A History of Limit Central Theorem: From Classical to Modern Probability Theory.* Springer, 2011.

FREUND, John. *Estatística aplicada: Economia, administração e contabilidade.* Ed. 11. Porto Alegre: Bookman, 2006.

GAL, Iddo; GARFIELD, Joan. *The Assessment Chalenges in Statistical Education.* International Statistical Institute, 1997.

GARFIELD, Joan. *The Challenge of Developing Statistical Reasoning.* Journal of Statistics Education, Volume 10, Number 3, 2002.

GARFIELD, Joan; GAL, Iddo. *Teaching and Assessing Statistical Reasoning.* National Council of Teachers of Mathematics, 1999.

GEOVANNY, José Ruy; BONJORNO, José Roberto. *Matemática uma nova abordagem: progressões vol 2*. Ed. 2. São Paulo: FTD, 2011.

GEOVANNY, José Ruy; GEOVANNY Jr, José Ruy; BONJORNO, José Roberto. *Matemática fundamental: uma nova abordagem, vol. único*, Ed 2. São Paulo: FTD, 2011.

GILLIES, Donald. *Philosophical Theories of Probability.* New York: Routledge, 2006.

GNEDENKO, B. V. *The Theory of Probability.* Moscow: Mir Publishers, 1978.

GOMES, Paulo; CAMPOS, Pedro; BACELAR, Sérgio. *ALEA: Um Contributo Para A Promoção Da Literacia Estatística (Análise De Dados e Ensino Da Estatística Nas Escolas Secundárias).*

HOWIE, David. *Interpreting Probability: Controversies and Developments in the Early Twentieth Century*. Cambridge University Press, 2004.

HUGENII, Christiani. *Libellus De Ratiociniis In Ludo Aleae or, The Value Of All Chances.* London, 1714.

IEZZI, Gelson; DOLCE, Osvaldo; DEGENSZAJN, David; PÉRIGO, Roberto; ALMEIDA, Nilze. *Matemática: ciência e aplicações vol. 2.* Ed. 5. São Paulo: Atual, 2010.

JAMES, Barry R. *Probabilidade: Um curso em nível intermediário.* Ed. 3. Rio de Janeiro: IMPA, 2011.

JENDRASZEK, Patricia. *An Analysis of a Misconception of Probability among Future Mathematics Teachers.* Journal of Mathematics Educations at Teachers College vol. 1, Spring-Summer, 2010.

KERMOND, John. *Counterexamples in Probability and Statistics*. Haileybury College Senior School.

KOLMOGOROV, Andrey. *Foundations of the Theory of Probability.* Ed. 2. New York: Chelsea, 1956.

LAPLACE, Pierre-Simon. *Ensaio filosófico sobre as probabilidades.*Rio de Janeiro: PUC – RJ, 2010.

LAPLACE, Pierre-Simon. *Théorie Analytique des Probabilités.* Ed. 2. Paris, 1814.

LEI DE DIRETRIZES E BASES DA EDUCAÇÃO NACIONAL, Ed. 5, 2010.
LESIGNE, Emmanuel. *Heads or Tails: An Introduction to Limit Theorems in Probability.* Vol. 28. American MathematicalSociety, 2005.

MACHADO, Antônio dos Santos. *Matemática machado, vol. Único*. Ed 1. São Paulo: Atual, 2012.

MEYER, Paul. *Probabilidade: Aplicações à Estatística.* Ed. 2. Rio de Janeiro: LTC, 2009.

MLODINOW, Leonard. *O Andar do Bêbado.* Rio de Janeiro: Jorge Zahar Editora Ltda, 2009

MOIVRE, Abraham. *The Doctrine of Chances: or, A Method of Calculating the Probabilities of Events in Play.* Ed. 3. London, 1756.

MOOD, Alexander. *Introduction to the theory of statistics.* Ed. 3. United States of American: McGraw-Hill, 1974.

MOORE, David. *Teaching Statistics as a Respectable Subject.* Purdue University, 1992.

PAIVA, Manoel Rodrigues. *Matemática: Paiva Moderna Plus Vol. 2*. Ed. 2. São Paulo: Moderna, 2010.

PARÂMENTROS CURRICULARES NACIONAIS TERCEIRO E QUARTO CICLOS DO ENSINO FUNDAMENTAL: MATEMÁTICA,1998

PARÂMENTROS CURRICULARES NACIONAIS, 2000.

PARÂMENTROS CURRICULARES NACIONAIS ENSINO MÉDIO +: ORIENTAÇÕES EDUCACIONAIS COMPLEMENTARES AOS PARAMETROS CURRICULARES NACIONAIS, 2000.

PARKE, Carol. *Reasoning and Communicating in the Language of Statistics.* Journal of Statistics Education, Volume 16, Number 1, 2008.

PIAGET, Jean; INHELDER, Bärbel. *A ideia da origem do acaso na criança.* Editora Record, 1951.

PINHEIRO, J. et al. *Princípios de Estatística.* Editora Elsevier-Campus, versão preliminar para uso da UFRJ, 2011.

POPPER, Karl. *A Lógica da Pesquisa Científica.* São Paulo: Cultrix, 1972.

RÉVÉSZ, Pál. *The Laws of Large Numbers*. New York: Academic Press Inc, 1968.

ROQUE, Tatiana. *História da matemática: Uma visão crítica, desfazendo mitos e lendas.* Rio de Janeiro: Zahar, 2012.

RUMSEY, Deborah. *Statistical Literacy as a Goal for Introductory Statistics Courses.* Journal of Statistics Education, Volume 10, Number 3, 2002.

SANCHES, Paulo S. B; SAMPAIO, Fausto A; RANGEL, Cristiano M; RIBEIRO, Flávio E. *Matematikós vol. Único.* São Paulo: Saraiva, 2010.

SEDLMEIER, Peter. *Improving Statistical Reasoning: Theoretical Models and Praticals Implications.* New Jersey London: Lawrence Erlbaum Associates, 1999.

SHEYNIN, Oscar; *Theory of Probability. A Historical Essay.* Ed. 2. Berlin: Sheynin, 2009.

SHULMAN, Lee S. *Knowledge and Teaching: Foundations Of The New Reform,* 1987.

SHULMAN, Lee S. *Those Who Understand: Knowledge Growth in Teaching*, 1986.

SINCLAIR, John. *The Statistical Account of Scotland: Draw Up From The Communication of The Ministers of The Different Parishes.* London: William Creech, 1791.

SOUZA, Joamir R. *Novo olhar matemática: versão com progressões vol. 2.*Ed 1. São Paulo: FTD, 2011.

STOYANOV, Jordam M. *Conterexamples in Probability.* Ed 2. Chichester: John Wiley & Sons Ltda, 1997.

TABAK, John. *Probability and statistics: the science of uncertainty.* New York: Facts On File*,*2004

TALL, David; VINNER, Shlomo. *Concept Image and Concept Definitionin Mathematicswith particular reference to Limits and Continuity*, Published in *Educational Studies in Mathematics*, 12, 151–169, (1981).

APÊNDICE A: CRITÉRIO DE CONVERGÊNCIA QUASE CERTA

Sejam $\{X_n\}_{n\geq 1}$ uma sequência de variáveis aleatórias e X uma variável aleatória. Dizemos que se $P(|X_n - X| > \epsilon \; infinitas \; vezes) = 0$, para todo $\epsilon > 0$, então$X_n \xrightarrow{q.c} X$.

Demonstração:

Se $P(|X_n - X| > \epsilon \; infinitas \; vezes) = 0$ para todo $\epsilon > 0$, então

$$P\left(\lim_{n\to\infty} X_n \neq X\right) \leq \sum_{n=1}^{\infty} P(|X_n - X| > \epsilon \; infinitas \; vezes) = 0$$

Portanto,

$$P\left(\lim_{n\to\infty} X_n = X\right) = 1$$

■

APÊNDICE B: LEI FORTE DOS GRANDES NÚMEROS

Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes, identicamente distribuídas e integráveis, com $E(X_n) = \mu$. Então

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

Demonstração

Seja U_n uma variável aleatória tal que $U_n = X_n - \mu$, assim as variáveis aleatórias U_n são independentes e identicamente distribuídas, além disso, note que

$$U_1 + \cdots + U_n = X_1 + .. + X_n - n\mu$$

Assim, se $\frac{U_1+\cdots+U_n}{n} \xrightarrow{q.c} 0$, então $\frac{X_1+\cdots+X_n}{n} \xrightarrow{q.c} \mu$

Dessa forma, podemos supor $\mu = 0$ sem qualquer prejuízos para a demonstração.

Definamos $Y_n = \begin{cases} X_n, se -n < X_n \leq n \\ 0, caso\ contrário \end{cases}$ e $Z_n = X_n - Y_n$, de modo que $X_n = Y_n - Z_n$ e

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

Para demonstrar o teorema basta mostrar que:

a) $\frac{Z_1+\cdots+Z_n}{n} \xrightarrow{q.c} 0$;

b) $\frac{Y_1+\cdots+Y_n}{n} - \frac{E(Y_1)+\cdots+E(Y_n)}{n} \xrightarrow{q.c} 0$; e

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

Demonstração de (a)

Da definição de Z_n temos que $Z_n \neq 0 \Leftrightarrow Y_n \neq X_n \Leftrightarrow X_n \notin (-n, n]$. Então, segue que $P(Z_n \neq 0) = P(X_n \notin (-n, n]) \leq P(|X_n| \geq n)$. Considerando os eventos $A_n = [Z_n \neq 0]$, temos que

$$\sum_{n=1}^{\infty} P(A_n) \leq \sum_{n=1}^{\infty} P(|X_n| \geq n)$$

do fato de que as variáveis aleatórias X_n são identicamente distribuídas segue que

$$\sum_{n=1}^{\infty} P(|X_n| \geq n) = \sum_{n=1}^{\infty} P(|X_1| \geq n)$$

e como X_1 é integrável por hipótese, tem-se

$$\sum_{n=1}^{\infty} P(|X_1| \geq n) < \infty$$

Pelo lema de Borel-Cantelli (enunciado na seção 5.1) decorre que

$$P(Z_n \neq 0 \; infinitas \; vezes) = 0$$

isto é, $Z_n \neq 0$ para apenas um número finito de n, ou dito de outra forma, Z_n assume o valor 0 para infinitos valores de n. Com isso, temos que $Z_n \to 0$ e $\frac{Z_1+\cdots+Z_n}{n} \to 0$, logo $P\left(\frac{Z_1+\cdots+Z_n}{n} \to 0\right) = 1$.

Demonstração de (b)

Seja F a função de distribuição comum, $F = F_{X_n}$. Vamos verificar que as variáveis aleatórias Y_n satisfazem a condição de Kolmogorov (enunciada na seção 5.1).

Como $Y_n = X_n$, se $-n < X \leq n$, tem-se que

$$V(Y_n) \leq E(Y_n^2) = E(X_n^2, com \; X_n \in (-n, n]) = \int \left[x^2, com \; X_n \in (-n, n]\right] dF(x) =$$

$$= \int_{-n}^{n} x^2 dF(x) < \infty$$

Utilizando o Lema (1) enunciado a seguir, segue da condição de Kolmogorov que as variáveis aleatórias Y_n satisfazem a Lei Forte dos Grandes Números, e assim, (b) está provado.

Lema (1): Seja X uma variável aleatória integrável com função de distribuição F. Então, $\sum_{n=1}^{\infty}\left[\frac{1}{n^2}\int_{-n}^{n} x^2 dF(x)\right] < \infty$. (A demonstração desse lema pode ser lida em (JAMES, 2011, p.211))

Demonstração de (c)

Basta demonstrar que $E(Y_n) \to 0$, como $E(Y_n) = E(X_n, com - n < X_n \leq n)$, do fato de as variáveis aleatórias X_n serem identicamente distribuídas segue que $E(X_n, com - n < X_n \leq n) = E(X_1, com - n < X_1 \leq n) \to E(X_1) = 0$, pelo Teorema da Convergência Dominada que diz que se $Y, X, X_1, X_2, \ldots$ são variáveis aleatórias em $(\Omega, \mathcal{A}, P)$ tais que Y é integrável, $|X_n| \leq Y$ $\forall n$, e $X_n \to X$. Então X e X_n são integráveis e $E(X_n) \to E(X)$. (A demonstração desse teorema pode ser lida em (JAMES, 2011, p.137)).

APÊNDICE C: FATOR DE CORREÇÃO DE CONTINUIDADE

O fator de correção de continuidade é utilizado quando aproximamos uma distribuição discreta (por exemplo, Binomial) à uma distribuição contínua (Neste caso, Normal). Observe na figura a seguir uma representação da aproximação de uma distribuição Binomial à uma distribuição Normal

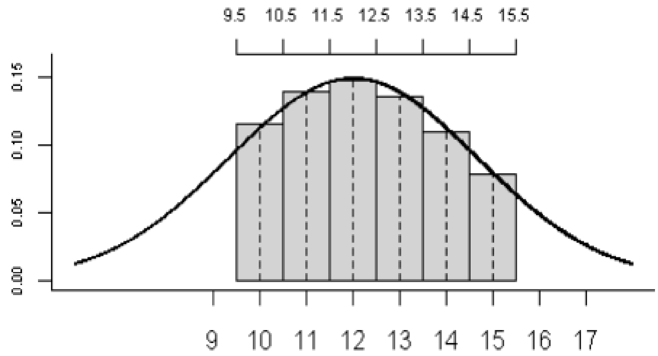


Figura 6: Importância do fator de continuidade

Fonte: PINHEIRO, J. et. al, 2011, p.131

Note que se desejarmos calcular $P(10 \leq X \leq 15)$, com $a \in \mathbb{Z}$ e $b \in \mathbb{Z}$, pela distribuição Binomial, deveremos calcular

$$P(X = 10) + P(X = 11) + \cdots + P(X = 15)$$

Pela figura 6 apresentada, isso equivale a calcular a soma das áreas de todos os retângulos de base 1 e altura $P(X = x)$. Por outro lado, se optarmos por aproximar a Binomial à curva Normal, o cálculo deverá ser feito através da integral

$$\int_{10}^{15} f(x)dx$$

Contudo, veja na figura 1 que tanto a metade (da esquerda) do retângulo cujo o ponto médio da base é 10 quanto a metade (da direita) do retângulo cujo o ponto médio da base é 15, não são calculados pela integral definida acima. A fim de obter uma aproximação melhor para o valor da integral, basta subtrairmos 0,5 do ponto a extrema esquerda do intervalo de integração e somarmos 0,5 ao ponto a extrema direita desse intervalo. Assim, para uma melhor aproximação da Binomial à Normal, poderemos calcular

$$\int_{9,5}^{15,5} f(x)dx$$

APÊNDICE D: TABELA Z (CURVA NORMAL)

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

APÊNDICE E: PROCEDIMENTOS PARA CONSTRUÇÃO DOS APPLETS NO GEOGEBRA

Procedimento para reprodução sucessiva de um experimento

1) Crie um controle deslizante “a”, com um intervalo mínimo de -5, máximo de 5 e incremento 1.
   Esse controle fará com que o experimento seja reproduzido a cada vez que seu valor for modificado.
2) Crie uma variável chamada *play* e atribua a ela o valor *true*. Para isso basta inserir no campo de entrada a expressão *play* = *true*.
   Essa variável controlará o início da reprodução do experimento e a sua pausa.
3) Crie um botão e dê o nome de INICIAR. No campo “Código Geogebra” insira a linha de programação abaixo, da forma como se apresenta:

   ```
   play = play+1
   IniciarAnimação[a,play]
   DefinirLegenda[bt1, Se[play=true, "PARAR", "INICIAR"]]
   ```

   Os códigos digitados acima são executados todas as vezes que o botão for clicado. A linha 1 é responsável pela alteração da variável play de true para false e vice-versa; a linha 2 envia um comando para que o controle deslizante “a” seja animado, caso play = true, isto é, ele ficará alterando os valores dentro do intervalo de -5 a 5 continuamente e parará caso play = false; e a linha 3 é responsável apenas por modificar a legenda do botão de INICIAR para PARAR e vice-versa.
   Criamos esse botão para facilitar a animação do controle deslizante e para controlá-la melhor, como veremos mais a frente.
4) Crie um controle deslizante “n” com intervalo de 1 a 100 e incremento 1.
   Esse controle será responsável pelo tamanho da amostra. Deixe por enquanto em 100.

5) Crie um botão o nomeie de REINICIAR e no campo “Código Geogebra” insira o comando abaixo:

```
play = false
IniciarAnimação[a,play]
contador = Sequência[0, i, 0,n]
contador1 = CopiarObjetoLivre[contador]
DefinirLegenda[bt1, Se[play==true, "Parar", "Iniciar"]]
Ampliar[1]
```

A primeira e a segunda linha são responsáveis pela pausa da reprodutibilidade do experimento; a terceira linha cria uma sequência de n+1 elementos, todos iguais a zero ou a redefine para este formato, caso ela já exista; a quarta linha cria uma cópia da sequência definida anteriormente; a quinta linha apenas redefine a legenda do botão 1 (bt1) “INICIAR”; e a última linha apenas limpa a tela, no caso de habilitarmos o rastro de algum ponto.

A primeira sequência “contador” será uma sequência para podermos reiniciar todo o processo, já a sequência “contador1” será responsável pela contagem da ocorrência de cada evento ocorrido.

6) Crie um controle deslizante “v” com intervalo de 0 a 1000 e incremento 1; vá em propriedades do controle deslizante “a”, em animação, no campo velocidade insira a letra v (representando o controle deslizante recém criado). Este controle nos permitirá determinar com que velocidade queremos que o experimento seja reproduzido.

Exemplo 1: Estimando a proporção de bolas vermelhas em uma piscina – Lei Forte

1) Modifique a terceira linha de programação do botão reiniciar para contador = Sequência[0, i, 1,1] e adicione após todas as linhas de comando, o código: cont = 0.

2) Insira no campo entrada as variáveis total = 0 e vermelhas = 0. Em seguida crie dois campos de entrada, cada um vinculado a uma dessas variáveis. Atribua os valores 10000 para a variável total e 4000 para a variável vermelhas.
3) Para criar o conjunto Universo vamos precisar de três passos. Crie a sequência u1 = Sequência[1, i, 1, vermelhas, 1](esta é uma sequência de números 1 com a quantidade definida em “vermelhas” de elementos). Crie a sequência u2 = Sequência[0, i, 1, total - vermelhas] (este comando cria uma sequência com (total-vermelhas) elementos, todos iguais a 0). Crie uma sequência u = Embaralhar[Concatenar[{u1, u2}]] (este comando é responsável pela construção do nosso conjunto Universo). Ele une aleatoriamente os elementos de u1 e u2 que representam as bolas vermelhas “1” e as brancas “0”.
4) Para a nossa amostra usaremos o seguinte comando:
   Amostra = Sequência[Elemento[u, NúmeroAleatório[1, total] + 0a], i, 1, n](este comando cria uma sequência de n elementos aleatórios a partir do conjunto Universo u). (Há o comando Amostra[lista, tamanho] que faz a mesma coisa, porém, por alguma razão, ele exige mais processamento o que torna a simulação mais lenta.)
5) Crie a variável Sn = Soma[amostra], que conta o número de bolas vermelhas na amostra e crie também a variável prop = Sn/n que mede a proporção de bolas vermelhas em cada amostra.
6) Vá em propriedade do objeto contador1, na guia programação, na aba Ao atualizar e digite o seguinte código:

```
cont = cont + 1
DefinirValor[contador1,1, (Elemento[contador1,1]+prop)]
```

   A primeira linha é responsável por contar quantos ensaios foram realizados e a segunda linha calcula a soma das proporções de bolas vermelhas na amostra, obtidas em cada ensaio, no objeto contador1.

7) Crie a variável est = Elemento[contador1, 1] / cont, responsável por calcular a média das proporções obtidas em cada ensaio.

Com isso concluímos todos os passos para a construção de um applet que modela a situação. Quanto aos aspectos gráficos, deixaremos isso a cargo do leitor, pois são tarefas relativamente simples que podem ser facilmente encontradas no próprio site do desenvolvedor do software.

Exemplo 2: O problema clássico de Monty Hall – Lei Forte

1) Siga o **Procedimento para reprodução sucessiva de um experimento**. Fazendo as seguintes alterações:altere o código da terceira linha do botão REINICIAR para contador = Sequência[0, i, 1,2]; adicione a este mesmo botão o código contador2 = CopiarObjetoLivre[contador] e modifique o valor mínimo do intervalo do controle deslizante n para 3.

   Os contadores 1 e 2 serão responsáveis pelas estratégias de trocar ou não de porta, respectivamente. Cada contador terá apenas dois elementos, sendo o primeiro um contador para as vezes em que a porta for encontrada e o outro para as vezes que não for. O valor do controle deslizante n representará o número de portas que podem ser escolhidas. Defina, a princípio, como 3.
2) Crie as variáveis pcarro = NúmeroAleatório[1, n] + a*0 e pescolha = NúmeroAleatório[1, n] + a*0.

   A primeira variável definirá em qual porta estará o carro, em cada ensaio. Já a segunda definirá qual a porta escolhida pelo participante, também em cada ensaio.
3) No objeto controle deslizante “a”, vá em propriedades, em seguida na guia programação, na aba Ao atualizar e digite o seguinte código:

   Se[pcarro == pescolha, DefinirValor[contador1,1, Elemento[contador1,1]+1], DefinirValor[contador1,2, Elemento[contador1,2]+1]]

Se[pcarro != pescolha, DefinirValor[contador2,1, Elemento[contador2,1]+1], DefinirValor[contador2,2, Elemento[contador2,2]+1]]

A primeira linha será responsável pela contagem da primeira estratégia, isto é, a estratégia é não trocar de porta. Assim, se a porta escolhida (pescolha) for igual a porta que está o carro (pcarro), ela adiciona o valor 1 ao primeiro elemento do objeto contador1 e adiciona igualmente o valor 1 ao segundo elemento desse objeto caso contrário.

A segunda linha faz o mesmo para a estratégia de sempre trocar de porta. Neste caso, sempre que pescolha for diferente de pcarro o participante vencerá.

4) Crie a variável Ensaio = Soma[contador1]. Responsável pela contagem de quantos ensaios foram realizados.
5) Por fim crie as sequências abaixo:
   frelativa1 = Sequência[Elemento[contador1, i]/Ensaio, i, 1, 2]; e
   frelativa2 = Sequência[Elemento[contador2, i]/Ensaio, i, 1, 2]

A frelativa1 é uma sequência de dois elementos, sendo o primeiro deles referente à frequência relativa com que o carro é encontrado com a estratégia 1 e o segundo referente à frequência relativa com que o carro não é encontrado. A frelativa2 faz exatamente o mesmo, só que para a estratégia 2.

Tal como fizemos no exemplo 1, deixaremos a construção dos aspectos visuais para o professor.

Exemplo 3: Proplema de Otimização Estocástica: Venda de flores – Lei Forte

1) Execute o **Procedimento para reprodução sucessiva de um experimento**, com as seguintes modificações: altere, a terceira linha de programação do botão REINICIAR para contador = Sequência[0, i, 0,4] e não crie o controle deslizante n.

2) Crie as variáveis compra = 0 e venda = 0, crie também dois campos de entrada, um para compra e outro para venda vinculando-os as respectivas variáveis criadas. Atribua, a princípio, o valor de compra como 10 e o de venda 20.
3) Crie as variáveis p0 = 0, p1 = 0, p2 = 0, p3 = 0, p4 = 0, crie em seguida cinco campos de entrada, um para cada variável descrita acima. O campo de entrada deverá ser vinculado às respectivas variáveis. Atribua, tal como no exemplo feito na simulação braçal, os valores 0.1, 0.2, 0.4, 0.1 e 0.2, para as variáveis p0, p1, p2, p3 e p4, respectivamente.
4) Crie a variável sort = NúmeroAleatório[0, 99] + 0*a.
5) No controle deslizante “a”, vá em propriedades, na guia programação, na aba Ao atualizar e digite o código abaixo:

   Se[sort < p0*100, DefinirValor[contador1,1, Elemento[contador1,1]+1], Se[sort < p0*100+ p1*100, DefinirValor[contador1,2, Elemento[contador1,2]+1], Se[sort < p0*100+ p1*100+p2*100, DefinirValor[contador1,3, Elemento[contador1,3]+1], Se[sort < p0*100+ p1*100+p2*100+p3*100, DefinirValor[contador1,4, Elemento[contador1,4]+1],DefinirValor[contador1,5, Elemento[contador1,5]+1]]]]]

   Esse código é responsável pela contagem dos dias em que entrou alguém na floricultura querendo comprar 0, 1, 2, 3 ou 4 flores. Sendo essa contagem feita em cada um dos cinco elementos da sequência contador1, o primeiro elemento correspondendo a 0 flores, o segundo a 1 flor e assim por adiante.

6) Crie a variável ensaio = Soma[contador1]. Responsável pela contagem do número de vezes que o experimento foi realizado.
7) Crie a sequência frelativa = Sequência[Elemento[contador1, i] / ensaio, i, 1, 5]. Responsável por calcular a frequência relativa com que alguém entre na floricultura querendo comprar 0, 1, 2, 3 ou 4 flores.
8) Crie as variáveis L0, L1, L2, L3 e L4, tal como descrito abaixo:
   L0 = 0;

L1 = (venda - compra)*(1 - Elemento[frelativa, 1]) – compra*Elemento[frelativa, 1];

L2 = 2*(venda - compra)*(1 - Elemento[frelativa, 2] - Elemento[frelativa, 1]) + (venda - 2compra)*Elemento[frelativa, 2] – 2*compra*Elemento[frelativa, 1];

L3 = 3*(venda - compra)*(1 - Elemento[frelativa, 1] - Elemento[frelativa, 2] - Elemento[frelativa, 3]) + (2*venda – 3*compra)*Elemento[frelativa, 3] + (venda – 3*compra)*Elemento[frelativa, 2] – 3*compra*Elemento[frelativa, 1]; e

L4 = 4*(venda - compra)*Elemento[frelativa, 5] + (3*venda – 4*compra) Elemento[frelativa, 4] + (2*venda – 4*compra)*Elemento[frelativa, 3] + (venda – 4*compra)*Elemento[frelativa, 2] – 4*compra*Elemento[frelativa, 1]

Essas são as variáveis responsáveis por medir o lucro médio diário, se adotarmos a estratégia 1, 2, 3, 4 ou 5, respectivamente, que foram adotadas na solução teórica do problema.

9) Crie a sequência lucro = {L0, L1, L2, L3, L4}*ensaio. Essa sequência fornece o valor do lucro, para cada estratégia, após n ensaios executados.

Embora, deixemos a cargo do leitor a organização do ambiente no Geogebra, bem como o desenvolvimento de gráficos, deixaremos também uma sugestão de código que pode ser útil. Esse código redefine a escala do eixo y de acordo com o provável maior valor de lucro após n ensaios.

Clique com o botão direito do mouse na tela, acesse a opção Janela de Visualização..., vá em preferências – Janela de Visualização. Na seção Dimensões, altere o campo y Mín para -0.2*L2*ensaio e o campo y Máx para L2*ensaio+(L2* ensaio/5).

Dependendo dos valores de compra, venda, p0, p1, p2, p3, e p4, pode ser interessante utilizar uma função que tome o valor do lucro máximo (Máximo[lista]), ao invés de L2.

Exemplo 4: Probabilidade Geométrica – Lei Forte

1) Execute o **Procedimento para reprodução infinita de um experimento**, com as seguintes modificações: altere, a terceira linha de programação do botão REINICIAR para contador = Sequência[0, i, 1,2] e não crie o controle deslizante n.
2) Abra a planilha do Geogebra (atalho: ctrl + shift +s) e digite os códigos e textos abaixo nas células correspondentes:

   Na coluna A vamos inserir apenas textos para indicar o que faremos.

   A1: pessoa1

   A2: pessoa2

   A3: Diferença

   A4: Ponto

   A5: Encontra

   A6: Não Encontra

   Na coluna B vamos inserir a maior parte dos códigos. Para isso precisaremos exibir o campo de entrada da planilha clicando em f_x no canto superior esquerdo da tela. Digite os códigos a seguir sem as aspas

   B1: “=random() + 0*a”. Este comando gera, com igual probabilidade, um número aleatório entre 0 e 1, diferentemente do comando NúmeroAleatório[a,b] que gera um número aleatório inteiro, com igual probabilidade, entre a e b.

   B2: “=random() + 0*a”

   B3: “=B1 - B2”. Este comando é responsável por verificar a diferença de tempo com que os amigos chegaram ao local.

   B4: “=(B1,B2)”. Este comando plota um ponto de coordenadas B1 e B2 no plano cartesiano.

3) Vá em propriedades da célula B3, na guia Programação, na aba Ao atualizar e digite o seguinte código:

```
Se[-0.150<=B3<=0.150, DefinirCor[B4, "Vermelho"],DefinirCor[B4, "Azul"]]
```

Se[-0.150<=B3<=0.150,DefinirValor[contador1, 1, Elemento[contador1,1]+1], DefinirValor[contador1, 2, Elemento[contador1,2]+1]]

A primeira linha de comando, apenas define a cor do ponto a ser plotado como vermelho ou azul. O ponto será vermelho, caso os amigos se encontrem e azul caso contrário. Já a segunda linha é responsável por contar quantas vezes os amigos se encontram (ponto vermelho) ou não (ponto azul). Sendo armazenado no primeiro elemento da sequência contador1 o número de vezes que se encontram e no segundo elemento o número de vezes que não se encontram.

4) Digite os seguintes códigos nas células indicadas:

B5: "=Elemento[contador1, 1]". Este comando apenas exibe o primeiro elemento da sequência contador 1 na célula B5.
B6: "=Elemento[contador1, 2]". Este é análogo ao comando de B5.
B7: "=Soma[B5:B6]". Esta célula servirá como contador do número de ensaios.
C4: "Frequência". É apenas um texto de referência, pois nas células abaixo escreveremos as frequências relativas dos amigos se encontrarem e de não se encontrarem.
C5: "=B5/B7"
C6:"=B6/B7"

Esses são os códigos utilizados para a simulação. Como temos sinalizado, a organização do espaço e dos gráficos é deixada a cargo do implementador. Porém, destacamos que pode ser útil deixar os pontos plotados marcados na tela, para isso basta clicar com o botão direito do mouse sobre B4 (na janela de álgebra) e selecionar a opção habilitar rastro.

Exemplo 5: Integral de Monte Carlo – Lei Forte

1) Execute o **Procedimento para reprodução sucessiva de um experimento**, com as seguintes modificações: altere, a terceira linha de programação do botão REINICIAR para contador = Sequência[0, i, 1,1], não crie o controle deslizante n, modifique o nome do controle deslizante "a" para "r", a fim de mantermos a notação que é geralmente adotada para a integral definida (lembre de alterar o nome do controle em todas as linhas de comando) e adicione uma linha com o comando cont = 0, que servirá para contar o número de ensaios realizados.
2) Insira as variáveis a = 0 e b = 0, em seguida crie um campo de entrada vinculado a cada uma delas. Atribua b = 1.
3) Insira as funções $f(x) = x^2$e $g(x) = f(x(b-a)+a)(b-a)$, no campo de entrada, da seguinte maneira:

   f(x) = x^2
   g(x) = f(x*(b - a) + a)*(b - a)

4) Crie a variável va = NúmeroAleatórioUniforme[0, 1] + r*0, o comando NúmeroAleatórioUniforme[0, 1], possui a mesma função do comando random() quando aplicado ao intervalo [0,1]. Crie também a variável va1 = g(va).
5) Crie o ponto P = (va, va1), este é um ponto aleatório da curva g(x), todos os pontos P gerados dessa forma possuem a mesma probabilidade de ocorrer.
6) Na linha de programação do controle deslizante "r" (na guia Ao atualizar) digite o seguinte código:

   cont = cont + 1
   Se[play == true, DefinirValor[contador1, 1, Elemento[contador1, 1] + va1]]

   A primeira linha do código faz a contagem do número de ensaios realizados e a segunda linha garante que o contador1 some todos os valores obtidos em va1, isto é ela soma todos os valores assumidos pela função g(x) nos ensaios já executados.

7) Crie a variável me = Elemento[contador1, 1] / cont. Esta variável representa o valor estimado da integral da função f(x) no intervalo [a,b]. Crie também a variável intg = Integral[g(x), 0, 1], que calcula a integra da função g(x) no intervalo [0,1], para fins de comparação da medida de área teórica com o valor estimado por pela variável me.

Como sempre, deixaremos os aspectos gráficos por conta do implementador.

Exemplo 1: Aproximação da Binomial à Normal - TCL

1) Execute o **Procedimento para reprodução sucessiva de um experimento**.
2) Crie a sequência moeda = Sequência[NúmeroAleatório[0, 1] + a 0, i, 1, n]. Essa sequência representará o lançamento de n moedas sendo lançadas consecutivamente.
3) Crie a variável cont = Soma[moeda], que conta, por exemplo, quantas caras ocorreram no lançamento de n moedas consecutivas.
4) Crie a variável c = 0 e crie o campo de entrada número máximo de caras, vinculado a variável c. Crie também a variável ensaio = Soma[contador1], que conta o número de ensaios realizados.
5) Adicione o seguinte código no campo de programação (Ao atualizar) do controle deslizante a:

DefinirValor[contador1, cont+1, Elemento[contador1, cont+1]+1]

Esse comando conta quantas vezes ocorreram c caras e registra no elemento c+1 da sequência contador1.

6) Crie a sequência frelativa = Sequência[Elemento[contador1, i] / ensaio, i, 1, n + 1], cujos elementos representa a frequência relativa com que ocorreram 0 caras, 1 cara, 2 caras, etc., respectivamente.
7) Crie a sequência nmax = Sequência[Elemento[frelativa, i], i, 1, c + 1]. Essa é a sequência de frequências relativas de todas as vezes que ocorreram c ou

menos caras. Em seguida crie a variável P = Soma[nmax], que mede a probabilidade de ocorrer no máx c caras.

Em geral, temos deixado a cargo do leitor construir os gráficos associados. Porém, neste caso, como é importantíssimo que o estudante perceba que a distribuição Binomial converge para a distribuição Normal, conforme o número de ensaios aumenta, deixaremos um código que constrói o histograma das frequências relativas dos números de caras obtidas. O código que deve ser inserido no campo de entrada é:

pol = Sequência[Polígono[(i,0), (i+1,0), (i+1, Elemento[frelativa, i+1]), (i,Elemento[frelativa, i+1])], i, 0, n].

De modo semelhante, pode-se criar o histograma que contém apenas as frequências relativas que desejamos mensurar. Para isso escreva no campo de entrada:

pol = Sequência[Polígono[(i,0), (i+1,0), (i+1, Elemento[frelativa, i+1]), (i,Elemento[frelativa, i+1])], i, 0, c+1]

Além disso, é recomendável também plotar a função densidade da curva Normal. Para isso, digite:

f(x) = DistribuiçãoNormal[n 0.5, (0.5*0.5*n)^0.5, x]

Note que, aqui, para efeitos de simulação, estamos adotando a média $\mu = 0{,}5 \times n$ e o desvio-padrão $\sigma = \sqrt{n \times 0.5 \times 0{,}5}$.

Para obter o cálculo teórico da probabilidade de se obter no máximo c caras, escreva o código g = Integral[f, 0, c +0.5]. Embora, o correto fosse considerar a integral de $-\infty$ até (c+0,5), o software não realiza esse cálculo. Porém, como ele também trabalha com truncamentos, não haverá uma perda significativa, se considerarmos a integração no intervalo [0,c+0,5]. Lembrando que utilizamos c+0,5 ao invés de somente c, devido ao fator de correção que devemos considerar nas aproximações das distribuições discretas às contínuas.

Exemplo 2: Fazer uma estimativa intervalar do número de bolinhas em uma piscina – TCL

Faça uma cópia do arquivo criado em **Exemplo 1: Estimando a proporção de bolas vermelhas em uma piscina – Lei Forte**, descrito neste apêndice e abra-o, em seguida.

1) Na programação (Ao clicar) do botão REINICIAR, insira o seguinte código no final: Intconf = {0,0}. Esta lista contará o número de vezes que o intervalo estimado conteve ou não o parâmetro.
2) Crie a variável prob = vermelhas/total.
3) Criaremos agora as variáveis responsáveis pelo cálculo do desvio-padrão, margem de erro (para um nível de confiança de 95%), estimativa do valor mínimo do intervalo e estimativa do valor máximo.
   dp = sqrt(prop*(1 - prop) / n);
   erro = 1.96*dp;
   Min = prop – erro; e
   Max = prop + erro.
4) Vamos adicionar um comando à programação (Ao atualizar) do controle deslizante “a”. Para isso insira após a última linha de comando, o código:

   Se[Min<=prob<=Max, DefinirValor[Intconf, 1, Elemento[Intconf, 1]+1], DefinirValor[Intconf, 2, Elemento[Intconf, 2]+1]]

   Esse comando é responsável por realizar a contagem do número de vezes que o parâmetro estava contido no intervalo estimado e armazenar na lista “Intconf”.

5) Por fim crie a variável conf = Elemento[Intconf, 1] / cont. Essa variável é responsável por calcular o percentual de vezes que o intervalo estimado conteve o parâmetro.

Sobre os aspectos de exposição gráfica, deixamos a cargo do implementador. Porém, deixaremos aqui uma sugestão que pode ser útil na organização do espaço:Em Janela de visualização, na guia Preferências e aba Básico, altere os valores de mínimo de x para “prob – 0.5” e o de máximo para “prob+0.5”.

A ideia é que ao plotar o intervalo (utilizando retas ou segmentos de retas perpendiculares ao eixo x), e o ponto P representando o parâmetro de interesse, fique sempre visível o intervalo estimado, com um zoom razoavelmente bom para identificar se o intervalo contém ou não o parâmetro.

Exemplo 3: Estimando o lucro da venda de flores - TCL

Crie uma cópia do arquivo do **Exemplo 3: Proplema de Otimização Estocástica: Venda de flores – Lei Forte**, descrito neste apêndice e abra esse arquivo.

Faça as seguintes modificações no arquivo copiado:

Insira no botão REINICIAR, na guia programação (Ao clicar), o seguinte código: Intconf = {0,0}. Tal como no exercício anterior, aqui contaremos o número de vezes que o intervalo estimado a partir de uma amostra conterá o parâmetro.

Crie o controle deslizante “n” com intervalo de 1 a 100 e incremento de 1. Com este controle definiremos o tamanho da amostra selecionada.

Crie a variável esperanca = 0. Aqui armazenaremos a esperança matemática para um estratégia escolhida pelo usuário. Para que o usuário defina qual estratégia ele quer verificar, proceda da seguinte maneira:

Crie o controle deslizante “estrategia”, com intervalo de 1 a 5 e incremento de 1. Em programação (Ao atualizar), digite o seguinte código:

Se[estrategia==1, DefinirValor[esperanca, 0], Se[estrategia==2, DefinirValor[esperanca, (venda-compra)*(1-p0)-compra*p0], Se[estrategia==3, DefinirValor[esperanca, (2*venda - 2*compra)*(1-p1-p0)+(venda-2*compra)*p1 - 2*compra*p0], Se[estrategia==4, DefinirValor[esperanca, (3*venda-3*compra)*(1-p2-p1-p0)+(2*venda-3*compra)*p2 +(venda-3*compra)*p1-3*compra*p0], Se[estrategia==5, DefinirValor[esperanca, (4*venda-4*compra)*p4+(3*venda-4*compra)*p3+(2*venda-4*compra)*p2+(venda-4*compra)*p1-4*compra*p0]]]]]

Com isso, os valores que o controle deslizante “estrategia” pode assumir, atribuem a variável “esperanca”, a esperança matemática correspondente.

A seguir crie a variável esperancan = esperança*n. Que é o valor esperado de lucro para a estratégia selecionada, em uma amostra de tamanho n.

Para selecionar uma amostra a partir de um conjunto que não poderemos escrever, procederemos da seguinte forma:

Crie a lista

amostra0 = Sequência[NúmeroAleatório[0, 99] + 0a, i, 1, n].

Essa lista gerará uma amostra de *n* números aleatórios que variam de 0 a 99.

Agora vamos reescrever essa sequência de modo que cada número represente, de acordo com as probabilidade indicadas pelas variáveis p0, p1,p2, p3 e p4, a quantidade de flores vendidas em cada um dos n dias que compõem a amostra.

Crie a lista

amostra1=Sequência[Se[Elemento[amostra0,i]<p0*n,0,Se[Elemento[amostra0, i]<p1*n+p0*n, 1, Se[Elemento[amostra0, i] < p2*n + p1*n + p0*n, 2, Se[Elemento[amostra0, i] < p3*n + p2*n + p1*n + p0*n, 3, 4]]]], i, 1, n]

Com uma nova sequência adequamos a amostra à estratégia escolhida.

amostra2=Se[estrategia == 1, Sequência[0, i, 1, n], Se[estrategia == 2, Sequência[Se[Elemento[amostra1, i] > 0, 1, 0], i, 1, n], Se[estrategia == 3, Sequência[Se[Elemento[amostra1, i] > 1, 2, Se[Elemento[amostra1, i] > 0, 1, 0]], i, 1, n], Se[estrategia == 4, Sequência[Se[Elemento[amostra1, i] > 2, 3, Se[Elemento[amostra1, i] > 1, 2, Se[Elemento[amostra1, i] > 0, 1, 0]]], i, 1, n], Se[estrategia == 5, Sequência[Se[Elemento[amostra1, i] > 3, 4, Se[Elemento[amostra1, i] > 2, 3, Se[Elemento[amostra1, i] > 1, 2, Se[Elemento[amostra1, i] > 0, 1, 0]]]], i, 1, n]]]]]

Por fim, criaremos a sequência que será de fato a nossa amostra de tamanho n.

dinheiroamostra = Sequência[-(estrategia - 1)*compra + Elemento[amostra2, i]* venda, i, 1, n]

Isto é, a lista “dinheiroamostra” é uma amostra de tamanho n com a estratégia definida pelo usuário, cujos elementos correspondem ao lucro diário.

Agora, definiremos as variáveis que representarão as somas das variáveis aleatórias contidas na amostra, o desvio-padrão da amostra, a margem de erro e os valores mínimo e máximo do intervalo de confiança, à 95%.

Para tal, insira no campo de entrada:

S = Soma[dinheiroamostra];

dpamostra = DesvioPadrãoAmostral[dinheiroamostra]*sqrt(n);

erro = dpamostra*1.96;

Minimo = S – erro; e

Maximo = S + erro

Para realizar a contagem de quantos intervalos estimados, a partir de uma amostra de tamanho n, contém o parâmetro “esperancan”, insira a seguinte linha de código no controle deslizante “a”, em programação (Ao atualizar):

Se[Minimo<=esperancan<=Maximo, DefinirValor[Intconf, 1, Elemento[Intconf, 1]+1], DefinirValor[Intconf, 2, Elemento[Intconf, 2]+1]]

Por fim crie a variável conf = Elemento[Intconf, 1]/ensaio. Que calcula o percentual de vezes que o intervalo estimado conteve o parâmetro de interesse.

Deixamos as apresentações gráficas a cargo do leitor.

Exemplo 4: Estimação intervalar para área de figura planas - TCL

Crie uma cópia do arquivo construído em **Exemplo 4: Probabilidade Geométrica – Lei Forte**, deste apêndice e abra esse arquivo.

1) No botão REINICIAR, na linha de programação (Ao clicar), adicione o seguinte código: Intconf = {0,0} e crie o controle deslizante "n" com intervalo de 1 a 100 e incremento de 1.
2) Crie o polígono pol1=Polígono[(0, 0), (1, 0), (1, 1), (0, 1)]. Este polígono representa o conjunto Universo. Crie também o polígono pol2 = Polígono[(0, 0), (0, 0.15), (0.85, 1), (1, 1), (1, 0.85), (0.15, 0)], que representa o conjunto evento desejado.
3) Para criar nosso conjunto amostra, crie a lista amostra0 = Sequência[(random() + 0*a, random() + 0*a), i, 1, n]. Essa é uma lista de pontos com coordenadas aleatórias entre 0 e 1. Em seguida crie a sequência amostra1 = Sequência[abs(x(Elemento[amostra0, i]) – y(Elemento[amostra0, i])), i, 1, n], que é uma lista de n elementos, em que todos eles são o modulo da diferença das coordenadas dos pontos de amostra0.
4) Agora vamos contar quantos elementos da amostra representam o evento "os amigos se encontram" e quantos elementos estão em seu complementar. Para tal, crie as variáveis encontraamostra = ContarSe[x ≤ 0.15, amostra1] e nencontraamostra = n – encontraamostra. A primeira conta quantas vezes eles

se encontram em uma amostra de tamanho n e outra quantas vezes eles não se encontram.

5) Faremos aqui a construção das variáveis responsáveis pela proporção da amostra dos amigos que se encotram, o desvio-padrão amostral, a margem de erro e os pontos extremos do intervalo de confiança. Essas variáveis podem ser escritas da seguinte forma:

   propencontraamostra=encontraamostra/(encontraamostra+nencontraamostra);
   dpamostra = sqrt(propencontraamostra *(1 – propencontraamostra) / n);
   erro = 1.96*dpamostra;
   Min = 1.96-dpamostra; e
   Max = 1.96+dpamostra.

6) Ao controle deslizante "a" insira o seguinte código de programação na guia programação (Ao atualizar)

Se[Min<=pol2<=Max,DefinirValor[Intconf,1,Elemento[Intconf,1]+1],
DefinirValor[Intconf,2,Elemento[Intconf,2]+1]]

Pol2 é a área da região que corresponde ao evento desejado.

7) Por fim, criaremos a variável aleatória que mede a proporção de intervalos de confiança que contém o parâmetro desejado.

   conf = Elemento[Intconf, 1] / Soma[Intconf]

A organização dos aspectos gráficos ficará a cargo do implementador interessado.