UFRB - Universidade Federal do Recôncavo da Bahia
CETEC - Centro de Ciências Exatas e Tecnológicas
PROFMAT - Mestrado Profissional em Matemática

TEOREMA CENTRAL DO LIMITE: COMPREENDENDO E APLICANDO

JOSÉ EDUARDO COSTA ALVES

CRUZ DAS ALMAS
26 DE JULHO DE 2016

JOSÉ EDUARDO COSTA ALVES

TEOREMA CENTRAL DO LIMITE: COMPREENDENDO E APLICANDO

Dissertação apresentada ao Programa de Mestrado Profissional em Matemática do Centro de Ciências Exatas e Tecnológicas da Universidade Federal do Recôncavo da Bahia e da Sociedade Brasileira de Matemática como parte dos requisitos para a obtenção do título de Mestre .

Orientador: Juarez dos Santos Azevedo

Cruz das Almas
26 de julho de 2016

FICHA CATALOGRÁFICA

A474t

Alves, José Eduardo Costa.

Teorema central do limite: compreendendo e aplicando / José Eduardo Costa Alves._ Cruz das Almas, BA, 2016.

107f.; il.

Orientador: Juarez dos Santos Azevedo.

Dissertação (Mestrado) – Universidade Federal do Recôncavo da Bahia, Centro de Ciências Exatas e Tecnológicas.

1.Matemática – Problemas, exercícios etc. 2.Matemática – Estudo e ensino. I.Universidade Federal do Recôncavo da Bahia, Centro de Ciências Exatas e Tecnológicas. I.Título.

CDD: 510.7

Ficha elaborada pela Biblioteca Universitária de Cruz das Almas - UFRB.

Teorema Central do Limite: Compreendendo e aplicando.

José Eduardo Costa Alves

Dissertação apresentada ao Programa de Mestrado Profissional em Matemática em Rede Nacional (PROFMAT), coordenado pela Sociedade Brasileira de Matemática, ofertado pelo Centro de Ciências Exatas e Tecnológicas da Universidade Federal do Recôncavo da Bahia, como parte dos requisitos para obtenção do título de Mestre em Matemática.

Banca Examinadora:

Orientador: ______________________

Profº Dr. Juarez dos Santos Azevedo - UFRB

Membro: Mariana Pinheiro Gomes da Silva

Profª Dra. Mariana Pinheiro Gomes da Silva - UFRB

Membro: ______________________

Profª Ma. Silvia Patrícia Barreto Santana - UFRB

Cruz das Almas, 26 de julho de 2016.

Este trabalho é dedicado a minha querida família. A minha mãe Aidyl, a minha esposa Marta e aos meus filhos Igor e Ivan.

Agradecimentos

Quero registrar, aqui, minha gratidão a todos aqueles que fazem parte deste momento.

Primeiramente, quero agradecer a Deus pela saúde e disposição, pelo discernimento e por iluminar o meu caminho a fim de vencer esta jornada árdua porém necessária em minha vida.

Agradeço a minha mãe Aidyl pela compreensão em minhas ausências e pelo apoio dado.

Aos meus filhos Igor e Ivan, pela paciência e compreensão nas minhas ausências.

À minha querida esposa Marta que esteve sempre ao meu lado, apoiando-me e dando-me todo suporte necessário para concluir com exito esta etapa da minha caminhada.

Ao meu grande orientador, Juarez Azevedo, com quem tive o privilégio de estudar.

Ao Centro de Ciências Exatas e Tecnológicas pelo acolhimento nesses anos.

Aos professores do PROFMAT com quem tive o enorme prazer de conviver e estudar.

Aos meus colegas da turma de 2014.1 que, ao meu lado, venceram esta jornada.

A banca examinadora pela valiosa participação e acolhimento deste trabalho.

"A distância entre o sonho e a conquista, chama-se atitude."
(Autor Desconhecido.)

Resumo

Neste trabalho vamos abordar o Teorema Central do Limite desde o conceito mais intuitivo até as suas definições e características mais consistentes e rigorosas. Inicialmente, apresentaremos os conceitos fundamentais em probabilidade como variáveis aleatórias, esperança matemática e variância. Em seguida, teremos alguns teoremas como o Teorema da Unicidade, o Teorema de Helly-Bray e o Teorema da Continuidade de Paul-Levy servindo de apoio para as demonstrações dos teoremas limites inclusive o Teorema Central do Limite de Lindeberg. Depois, exibiremos algumas situações problemas mostrando o uso do Teorema Central do Limite na sociedade e nas ciências e, por fim, relatamos algumas propostas didáticas de aplicação de atividades em sala de aula no Ensino Médio.

Palavras-chave: Teorema Central do Limite, distribuição normal, situações problemas.

Abstract

In this paper we report the Central Limit Theorem from the most intuitive concept to your settings and more consistent and accurate characteristics. Initially, we present the fundamental concepts in probability as random variables, mathematical expectation and variance. Next, we will have some theorems as Theorem of the Oniness, Helly-Bray Theorem and Paul-Levy Continuity Theorem serving as support for the demonstrations of theorems boundaries including the central theorem of Lindeberg limit. Then, we'll show some problems situations showing the use of the central limit theorem in society and science finally, we report some didactic activities proposed application in the classroom in high school.

Keywords: Central Limit Theorem, normal distribution, problem situations .

Lista de ilustrações

Sumário

1 Introdução

A imprevisibilidade esteve presente na humanidade desde sempre, dia após dia, era após era, sempre cercando a vida de todos. A certeza absoluta, mesmo querendo despontar na frente, não impera na mente do ser humano, única e exclusivamente, porque não há certeza absoluta. Como seres humanos, somos relativos, nossos atos são relativos, o mundo que nós vivemos é relativo. Relativo no sentido de haver variáveis externas, fora de nosso controle, que interferem no resultado de nossas ações, sejam elas individuais ou coletivas, na tomada de decisões, até nos fenômenos naturais. Não conseguimos prever o que vem a seguir em nossas vidas ou na sociedade, no máximo, podemos inferir sobre um futuro de acordo com nosso interesse. Dado a essas e outras situações descritas é que o homem, com o intuito de estar no controle, de poder decidir ou até interferir nestes acontecimentos ditos aleatórios, desenvolveu e aperfeiçoou outros campos matemáticos que são a estatística e a teoria das probabilidades. O que veremos aqui neste trabalho é um elo fortemente estabelecido entre estes dois campos, ou seja, entre a estatística e a probabilidade que é o Teorema Central do Limite. No decorrer da introdução, vamos nos referir ao Teorema Central do Limite como TCL.

O TCL tem o termo central relacionado a palavra teorema e não a palavra limite. No sentido etimológico, é um teorema, logo é uma proposição passível de demonstração, é central porque tem a média e a variabilidade em torno da média como parâmetros e é um teorema limite porque suas propriedades derivam de convergências ou aproximações. Em termos gerais, o TCL afirma que uma sucessão de n variáveis independentes e identicamente distribuídas terá sua soma convergindo para uma distribuição normal reduzida ou normal padrão sempre que o tamanho n for suficientemente grande[1] ou estiver crescendo infinitamente. Isso quer dizer que se S_n é uma sucessão de n variáveis independentes e identicamente distribuídas com sua distribuição própria então, S_n terá uma distribuição aproximadamente normal com média nula e variância unitária, ou

[1] O termo suficientemente grande é muito relativo. Na distribuição de Poisson e na exponencial, o valor mínimo de seus parâmetros para a aplicação do TCL é para $\lambda \geq 5$. Para distribuições amostrais, o número n de elementos deve ser igual ou superior a 30.

seja, $S_n \sim N(0,1)$, através da variável normalmente reduzida Z, onde

$$Z = \frac{X - EX}{VarX}.$$

Nos referimos aqui a uma sucessão de variáveis aleatórias do tipo X_1, X_2, $\ldots$ X_n. Todas sendo partes de X pois S_n é um subconjunto de X.

Em outras palavras, variáveis aleatórias possuem distribuições específicas dentro de um estudo ou de uma análise de acontecimentos por exemplo, a massa e a altura de uma determinada população, geralmente, é normalmente distribuída. Calcular o tempo de espera até a ocorrência de algum evento segue, em geral, uma distribuição geométrica assim como a distribuição exponencial é muito utilizada na determinação do tempo de desintegração de uma partícula radioativa. Muitas destas distribuições possuem funções de densidades compostas por termos exponenciais ou termos combinatórios e, em estudos envolvendo grandes quantidades de elementos populacionais ou amostrais, como habitantes de um município ou grandes medidas de tempo, aplicar tais valores a estes termos, tornam o uso destas funções com pouca viabilidade. Através de uma aproximação normal, pela variável Z, não precisamos conhecer ou aplicar estas funções de densidades, bastando apenas conhecer a média ponderada da variável aleatória em questão. Essas características estão nas soluções das situações problemas presentes no Capítulo 5.

Na estatística inferencial, o TCL atua no campo amostral em situações bem específicas fazendo a distribuição amostral obedecer um comportamento muito próximo de uma distribuição normal padrão com o intuito de promover maior confiabilidade nas estimativas realizadas. Eis algumas dessas situações específicas[2]. Primeiramente se a distribuição populacional é normalmente distribuída com média conhecida e variância desconhecida, o TCL garante que a distribuição amostral terá distribuição normal padrão com média igual a zero e variância igual a 1, denominada também de distribuição Z. Para amostras de tamanho menor que trinta usa-se a distribuição T de Student, ou simplesmente a distribuição T. Se a distribuição populacional não for normal ou for desconhecida, com média e variância conhecida, o TCL aproxima a distribuição amostral em uma distribuição normal padrão. Se a população for normalmente distribuída ou sua distribuição for desconhecida, com sua variância também desconhecida então o TCL garante que a distribuição amostral tenderá a uma distribuição normal padrão através da variância amostral S. Ou seja, para variância desconhecido temos:

$$\frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} \xrightarrow{D} N(0,1).$$

[2] Vamos fixar, neste pequeno relato introdutório, o tamanho das amostras como sendo maior ou igual a trinta.

Continuando na inferência estatística, se o tamanho da população for finita, o TCL afirma que quanto mais próximo o tamanho da amostra estiver do tamanho real de uma população, mais próximos os estimadores amostrais estarão dos parâmetros reais seguindo o comportamento da distribuição normal, ou seja, mais concentrados em torno da média real os valores amostrais estarão e menor será a variabilidade destes valores. Com isso, quanto mais próximo da média real o valor estiver, maior é a probabilidade de ocorrência e quando um valor amostral estiver mais distante da média real, menor é a probabilidade de ocorrência deste valor. Para população de tamanho infinito, o TCL age da mesma maneira, ou seja, quanto maior a dimensão da amostra maior é a probabilidade de que a média amostral esteja bem próxima da média real e, consequentemente, a amplitude de variação dos valores dessa média amostral tende a diminuir.

O trabalho apresentado aqui tem como característica um enfoque básico do TCL porém este enfoque não é estritamente teórico, ou seja, não envolve abordagens históricas, nem apresentação dos conceitos em livros textos ou então sobre uma aplicação específica do TCL. Possui sim muita apresentação não textual como gráficos e fórmulas matemáticas nos teoremas e demonstrações já que é voltado para um público bem específico composto por estudantes de graduação ou graduados em área afins e professores licenciados em matemática.

Como dito acima, não houve uma fundamentação teórica baseada em relatos históricos, no entanto, podemos encontrar em (FISCHER, 2011) uma abordagem histórica sucinta do desenvolvimento do TCL, em (HALD, 2004) um desenvolvimento histórico da inferência estatística de Bernoulli a Fischer e, em (BELLHOUSE, 2007), uma biografia de Abraham DeMoivre, o precursor de tudo que foi abordado neste estudo.

A estrutura deste trabalho dar-se-à em forma de capítulos. Cada um dos capítulos contribui para a totalidade da obra de uma forma bem particular, ou seja, mesmo contendo abordagens distintas todos são indispensáveis uma vez que atendem a proposta de fazer o leitor compreender e conhecer as aplicações do Teorema Central do Limite. O Capítulo 2 aborda sobre os termos e notações utilizadas nos outros capítulos tais como a definição e os tipos de variáveis aleatórias, as relações e diferenças entre a função de distribuição e a função de densidade, a média e a variância de uma variável aleatória e as funções de distribuições utilizadas nos exemplos de aplicações, em especial para a distribuição normal e a sua curva normal. O Capítulo 3 relata acerca das funções características e da convergência em distribuição. Este capítulo é constituído por definições, teoremas e suas demonstrações com o intuito de dar suporte para o Capítulo 4 que apresenta formalmente o Teorema Central do Limite em três versões: o teorema limite para variáveis aleatórias independentes e identicamente distribuídas com sua versão inicial proposta por DeMoivre e Laplace e o teorema limite para variáveis independentes não identicamente distribuídas que é o Teorema Central do Limite de Lindeberg. Estes três capítulos nos fornecem confiabilidade e segurança para o desenvolvimento do trabalho.

Os três capítulos seguintes estão encarregados de apresentar as aplicações do TCL. O Capítulo 5 faz comparações entre a probabilidade através da função de densidade da variável e através da aproximação normal exemplificando toda a apresentação do Capítulo 5. O Capítulo 6 leva o TCL para o campo da inferência estatística mais precisamente para as estimativas pontuais e intervalares, apresentando também algumas noções básicas de inferência, distribuição amostral e intervalo de confiança para um melhor entendimento por parte dos leitores. O sexto capítulo conduz o TCL para a educação básica, mais especificamente para professores e alunos do Ensino Médio apresentando propostas de atividades em sala de aula, onde podemos introduzir o conceito do TCL em situações bastante práticas para os alunos. A conclusão finaliza este trabalho relatando de forma objetiva uma síntese de tudo que foi exibido nos capítulos anteriores. Faz recomendações de outra proposta de aplicação que é a tábua de Galton, ratificando assim toda a importância que o Teorema Central do Limite possui, tanto na teoria das probabilidades quanto na própria matemática como um todo.

2

Variáveis aleatórias, momentos centrais e distribuição.

Quando estudamos fenômenos aleatórios, ou seja, episódios com pouca ou nenhuma chance de previsibilidade, aplicamos métodos estatísticos e probabilísticos na intenção de quantificar dados que, muitas vezes, não são representados numericamente. Esses dados são reconhecidos como variáveis e a quantificação dessas variáveis podem ser realizadas por meio de funções. As variáveis são classificadas como qualitativas ou quantitativas. Neste trabalho, vamos utilizar apenas variáveis quantitativas [1].

Definição 2.1 *Seja (Ω, A, P) um espaço de probabilidades. Uma função real X: $\Omega \to \mathbb{R}$ é chamada de variável aleatória se, definida no espaço amostral Ω para todo $\omega \in \Omega$, associar um único elemento real $X(\omega)$, isto é, $\omega \mapsto X(\omega)$.*

Exemplo 2.1 *Dentro de uma urna há duas bolas brancas e duas vermelhas. Ao extrairmos duas bolas quaisquer, sem reposição, e observarmos a sua cor, temos que o espaço amostral Ω será definido como:*

$$\Omega = \{(B, B); (B, V); (V, B); (V, V)\},$$

onde B = bola branca e V = bola vermelha.

Seja X *a variável aleatória definida como o número de vezes que a bola de cor branca aparece nas duas bolas retiradas da urna. Assim para $\omega_1 = (B, B)$ temos que $X_1 = 2$. Para $\omega_2 = (B, V)$ e $\omega_3 = (V, B)$ temos que $X_2 = 1$ e, para $\omega_4 = (V, V)$ temos que $X_3 = 0$. Com isso, obtemos o conjunto X tal que* $\mathrm{X} = \{0, 1, 2\}$.

Exemplo 2.2 *No evento medir a vida útil de um carro, em anos, uma possibilidade para o espaço amostral (Ω) seria de todos os números reais não negativos, isto é, $\Omega = [0, \infty)$. Sendo X a variável aleatória definida como o número de anos de vida útil do carro, temos que X assume valores reais não negativos dentro do conjunto Ω.*

[1] As variáveis qualitativas podem ser nominais ou ordinais e as deixamos de fora pois não podemos atribuir valores numéricos.

No Exemplo 2.1, a variável aleatória em questão assume apenas valores inteiros, com um número definido de possíveis ocorrências. Estas características descrevem uma variável aleatória discreta. No Exemplo 2.2, a variável assume valores inteiros e não inteiros com um número infinito de ocorrências. Temos, então, uma variável aleatória contínua. Se observarmos a Definição 2.1, para a variável aleatória discreta temos o contradomínio da função X como um conjunto finito ou infinito enumerável de valores reais. Para a variável aleatória contínua, temos o contradomínio da função X como um conjunto infinito não enumerável ou intervalos de números reais.

2.1 Função de densidade e de distribuição

Nesta seção, vamos entender o que é uma função de densidade e o que é uma função de distribuição, ambas aplicadas a variáveis discretas ou contínuas. Uma função de densidade associa a cada valor atribuído pela variável aleatória a probabilidade do evento correspondente. Notemos pela seguinte definição,

Definição 2.2 *Seja X uma variável aleatória discreta. Uma função real f é dita função discreta de densidade de X se $f(a) = P(X = a)$, para algum a real. Caso $f(a) > 0$, podemos afirmar que a é um possível valor da variável X.*

Observaremos agora três propriedades de uma função discreta de densidade X.

Propriedade 2.1 *Para um x real, $f(x) \geq 0$.*

Prova : Pela Definição 2.2 $f(x) = P(X = x)$. Como característica, $0 \leq P(X = x) \leq 1$ então $f(x) \geq 0$.

■

Propriedade 2.2 *Se um subconjunto definido por $\{x|f(x) \neq 0\}$ for finito ou infinito enumerável dentro do contradomínio $\mathbb{R}$ então este subconjunto será representado por $\{x_1, x_2, \cdots\}$.*

Prova : Pela Propriedade 2.1, $f(x) \geq 0$. Logo, $f(x) = 0$ ou $f(x) > 0$. Para $f(x) > 0$, a Definição 2.2 afirma que x é um possível valor de X. Portanto, o conjunto $\{x_1, x_2, \cdots\}$ existe e representa o subconjunto $\{x|f(x) \neq 0\}$.

■

Propriedade 2.3 *Seja $\{x_i\}$ um subconjunto real com $i = 1, 2, \cdots$. Se $f(x_i) > 0$ então $\sum_i f(x_i) = 1$.*

Prova : Para todos os valores possíveis de i, os $f(x_i)$ são distintos entre si. Assim, a soma $f(x_1) + f(x_2) + \cdots$ compõe todo o espaço amostral em questão. Como a soma das probabilidades é a probabilidade da soma,

$$\sum_i f(x_i) = \sum_i P(X = x_i) = P\left(\sum_i (X = x_i)\right) = P\left(\bigcup_i \{X = x_i\}\right) = P(\Omega) = 1,$$

pois a união de todos os possíveis eventos forma todo o espaço amostral e a probabilidade de ocorrência de todo o espaço amostral é igual a 1. Portanto,

$$\sum_i f(x_i) = 1.$$

■

Se uma função real f goza das Propriedades 2.1, 2.2 e 2.3 então, podemos afirmar que f é uma função de densidade discreta.

Uma função de densidade contínua é também denominada de função densidade em relação a integração. Isso porque, para realizarmos uma soma dentro de um intervalo real não enumerável, utilizamos o cálculo integral. Em relação a Definição 2.2, afirmando que $f(a) = P(X = a)$, se a variável X for contínua, então $f(a) = 0$. Notemos que, dado $\epsilon > 0$,

$$f(a) = P(X = a) \leq P(a - \epsilon < X \leq a) = P(X \leq a) - P(X \leq a - \epsilon).$$

Para ϵ muito próximo de zero temos:

$$\lim_{\epsilon \to 0} P(X \leq a - \epsilon) = P(X \leq a).$$

Logo,

$$\begin{aligned}
\lim_{\epsilon \to 0} [P(X \leq a) - P(X \leq a - \epsilon)] &= \lim_{\epsilon \to 0} P(X \leq a) - \lim_{\epsilon \to 0} P(X \leq a - \epsilon) \\
&= \lim_{\epsilon \to 0} P(X \leq a) - \lim_{\epsilon \to 0} P(X \leq a) \\
&= 0.
\end{aligned}$$

Assim, para X contínua, $P(X = a) = 0$. Ou seja, a probabilidade de uma variável aleatória contínua assumir um determinado valor a é igual a zero. Podemos, com isso, concluir que a função contínua de densidade não trabalha com valores pontuais e sim com intervalos reais não enumeráveis. A Propriedade 2.3 pode, então, ser expressa como

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

A função de distribuição, denotada por F, nos dá uma maneira de descrever como as probabilidades são associadas aos valores ou aos intervalos de valores de uma variável aleatória. A probabilidade P, em função da variável X, definida por $P_X(I_{\mathbb{R}}) = P(X \in I_{\mathbb{R}})$, onde $I_{\mathbb{R}}$ é um intervalo real qualquer, é denominada de distribuição de X. Com isso, temos que

$$F(X) = P(X \leq x), \qquad -\infty < x < \infty.$$

Eis a definição de uma função de distribuição.

Definição 2.3 *Qualquer que seja uma função real F, um elemento x do domínio de F e $F(x^+)$ a imagem de F quando os valores do domínio de F tendem a x pela direita com as seguintes propriedades:*

(i) $\forall x,\ 0 \leq F(x) \leq 1$.

(ii) F é uma função não decrescente de x.

(iii) $F(-\infty) = 0$ *e* $F(+\infty) = 1$.

(iv) $F(x^+) = F(x)$ *para todo x.*

F é uma função de distribuição se satisfaz as quatro propriedades acima.

Se a variável X estiver entre dois valores reais teremos

$$P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a), \qquad \forall a,\ b \in \mathbb{R}. \tag{2}$$

A variável aleatória sendo discreta ou contínua, podemos representar sua função de distribuição através de sua função densidade. Esta relação está na proposição seguinte e sua demonstração encontra-se em (JAMES, 2010).

Proposição 2.1 *Seja X uma variável aleatória e f(x) sua função de densidade.*

(i) Se X for discreta com valores em $\{x_1, x_2, \cdots\}$ *temos que*

$$F(X) = P_X(I_{\mathbb{R}}) = \sum_{i;x_i \in I_{\mathbb{R}}} P(X = x_i) = \sum_{i;x_i \in I_{\mathbb{R}}} f(x_i),$$

onde qualquer $I_{\mathbb{R}}$ *é finito ou infinito enumerável.*

(ii) Se X for contínua temos que

$$F(X) = P_X(I_{\mathbb{R}}) = \int_{I_{\mathbb{R}}} f(x)dx,$$

onde qualquer $I_{\mathbb{R}}$ *é infinito não enumerável.*

O item *(ii)* da Proposição 2.1 estabelece uma importante relação entre a função de distribuição e a função densidade

$$F(X) = \int_{I_{\mathbb{R}}} f(x)dx \Longleftrightarrow F'(x) = f(x)$$

ou, de outra maneira,

$$\frac{dF(x)}{dx} = f(x) \Longleftrightarrow dF(x) = f(x)dx. \tag{3}$$

Ainda pelo item *(ii)* da Proposição 2.1, podemos reescrever a relação expressa pela igualdade (2) da seguinte maneira,

$$P(a < X \leq b) = \int_a^b f(x)dx = F(b) - F(a), \qquad \forall a,\ b\ \in \mathbb{R}.$$

Podemos perceber que, independentemente da situação da variável aleatória, a função de distribuição comporta-se semelhantemente, tanto através da probabilidade quanto pela integração.

2.2 Esperança matemática

Esta seção apresenta uma formalização para o conceito de esperança matemática. A esperança matemática ou valor esperado da variável X é representada por EX ou μ_X e refere-se a média ponderada dos possíveis valores de X. Quando sabemos que X é a variável em questão, utilizamos apenas μ ao invés de μ_X. Encontrar a média de uma variável aleatória é uma forma de tentar reduzir a distribuição de probabilidade a um único número que suponhamos representar o valor típico de X.

A definição geral de esperança matemática passa primeiro por um teorema no qual veremos agora.

Teorema 2.1 *Seja X uma variável aleatória. Uma aproximação para X será definida como X_ϵ tal que*

$$X_\epsilon = \epsilon k \quad se \quad \epsilon k \leq X < \epsilon(k+1) \quad k \in \mathbb{Z}.$$

Se, para algum $\epsilon > 0$, X_ϵ tem esperança finita então X_ϵ tem esperança finita para todo $\epsilon > 0$ e

$$\lim_{\epsilon \to \infty} EX_\epsilon \tag{4}$$

existe e é finito.

A partir do Teorema 2.1, temos a seguinte definição.

Definição 2.4 *X é uma variável aleatória e X_ϵ é uma aproximação de X apresentada no Teorema 2.1 com $\epsilon > 0$. Se, para algum $\epsilon > 0$, X_ϵ tem esperança finita então a variável X possui esperança finita que será denotada por*

$$EX = \lim_{\epsilon \to \infty} EX_\epsilon.$$

Para maiores esclarecimentos sobre a definição geral de esperança consultar o Capítulo 7 em (HOEL, 1978).

Para o caso da variável ser discreta, temos a seguinte definição.

Definição 2.5 *Seja X uma variável aleatória discreta qualquer, que assume um número finito de valores $x_1, x_2, \ldots, x_r$. O valor esperado ou a esperança de X, é o número*

$$EX = \sum_{i=1}^{\infty} x_i f(x_i). \tag{5}$$

Para

$$\sum_{i=1}^{\infty} |x_i| f(x_i) < \infty,$$

temos que X possui esperança finita e definimos sua esperança através da igualdade (5).

No caso da variável ser contínua, temos a seguinte definição.

Definição 2.6 *Seja X uma variável aleatória contínua qualquer, que assume um número infinito de valores reais não enumeráveis ou dentro de um intervalo real. O valor esperado ou a esperança de X, é o número:*

$$EX = \int_{-\infty}^{\infty} x f(x) dx \tag{6}$$

A esperança de uma variável contínua apresentada na igualdade (6) pode ser reescrita como:

$$EX = \int_{-\infty}^{\infty} x dF(x). \tag{7}$$

A integral apresentada na igualdade (7) é conhecida como a Integral de Riemann-Stieltjes.

O teorema seguinte é uma importante propriedade de esperança matemática, referindo-se ao fato que a soma de variáveis discretas com esperança finita também é finita. Para o caso de variáveis contínuas, o procedimento é bem semelhante porém utilizamos a integral contida na igualdade (7) ao invés da representação por somatórios.

Teorema 2.2 *Sejam X e Y duas variáveis aleatórias tendo esperanças finitas.Sendo assim, $X + Y$ tem esperança finita e $E(X + Y) = EX + EY$.*

Demonstração : Fazendo $W = X + Y$ e g, uma função de densidade em função de W onde $g(w) = g(x, y) = x + y$. Com isso, temos que g é uma função de densidade conjunta das variáveis X e Y. Vamos utilizar a desigualdade triangular $|x + y| \leq |x| + |y|$.

$$\begin{aligned}
E|W| &= \sum_{x,y} |g(w)|f(x, y) = \sum_{x,y} |g(x, y)|f(x, y) \\
&= \sum_{x}\sum_{y} |x + y|f(x, y) \\
&= \sum_{x}\sum_{y} |x + y|f(x, y) \leq \sum_{x}\sum_{y}(|x| + |y|)f(x, y) \\
&= \sum_{x}\sum_{y} |x|f(x, y) + \sum_{x}\sum_{y} |y|f(x, y) \\
&= \sum_{x} |x|\sum_{y} f(x, y) + \sum_{y} |y|\sum_{x} f(x, y) \\
&= \sum_{x} |x|f_X(x) + \sum_{y} |y|f_Y(y) \\
&= E|X| + E|Y| < \infty
\end{aligned}$$

pois, por hipótese, as variáveis X e Y possuem esperança finita. Com isso, $E|W| < \infty$, isto é, a variável W possui esperança finita.

■

2.3 Momentos Centrais

Nesta seção, vamos discorrer acerca de momentos centrais de uma variável aleatória discreta. Estar escrevendo sobre momentos é de fundamental importância pois estamos introduzindo a idéia e posterior definição de variância e desvio padrão que é uma medida estatística muito relevante para a utilização do Teorema Central do Limite. Os momentos podem nos dar informações parciais sobre uma determinada medida de probabilidade de uma variável aleatória. Os momentos principais apresentados nesta seção serão os momentos de primeira e de segunda ordem.

Os momentos da variável X podem ser definidos pelas esperanças de potências de X. Vejamos a definição descrita a seguir.

Definição 2.7 *Seja X uma variável aleatória discreta e seja $r \geq 0$ um número inteiro. Dizemos que X tem um momento de ordem r se X^r tem esperança finita. Neste caso definimos o r-ésimo momento de X como EX^r.*

Pela Definição 2.5, podemos determinar o r-ésimo momento através da função de densidade f contida na igualdade abaixo

$$EX^r = \sum_x x^r f(x). \tag{8}$$

Teorema 2.3 *Se o r-ésimo momento de uma variável aleatória existir, então todos os momentos de ordem menores do que r também existem.*

Demonstração : Por definição, $r \geq 0$. Seja $k \geq 0$ um número inteiro com $k \leq r$. Por definição de potência, se $|x| \leq 1$ então $|x^k| = |x|^k$ e $|x|^r \leq |x|^k \leq 1$. Por outro lado, se $|x| > 1$, teremos $|x^k| = |x|^k \leq |x|^r$. Para os dois casos, podemos afirmar que $|x|^k \leq |x|^r + 1$.

Como $|x|^k \leq |x|^r + 1$, temos que $E(|x|^k) \leq E(|x|^r + 1)$. Vamos mostrar que $|x|^k$ tem esperança finita. Para isso, precisamos mostrar que $E(|x|^k) < \infty$.

$$E(|x|^k) \leq E(|x|^r + 1) = \sum_x |x^r + 1| f(x) \leq \sum_x |x^r| f(x) + \sum_x f(x).$$

Por hipótese, $E|x|^r < \infty$ e $\sum_x 1 f(x) = P(X = x) = 1$. Sendo assim,

$$\sum_x |x^r| f(x) + \sum_x 1 f(x) \leq \sum_x |x^r| f(x) + 1 < \infty.$$

Concluímos, então, que todos os momentos de ordem k, menores do que r, também existem.

■

Definição 2.8 *Seja X uma variável aleatória, com média aritmética $\mu = E(X)$ e $r \geq 0$ um número inteiro. Dizemos que $X - \mu$ tem um momento de ordem r se $(X - \mu)^r$ tem esperança finita. Neste caso, definimos o r-ésimo momento de $X - \mu$ como $E(X - \mu)^r$. Sendo assim, denominamos $(X - \mu)^r$ como o r-ésimo momento central de X, já que é o r-ésimo momento em torno da média.*

Pela igualdade (8), podemos determinar o r-ésimo momento central através da função de densidade f, onde

$$E(X - \mu)^r = \sum_x (x - \mu)^r f(x). \tag{9}$$

De agora até o final desta seção, vamos utilizar a notação EX ao invés de μ. Assim, escreveremos $(X - \mu)^r$ como $(X - EX)^r$ e, atribuiremos a r os valores 1 e 2.

Fazendo $r = 1$, no primeiro membro da igualdade (8) temos que

$$EX^r = EX^1 = EX.$$

Isto significa que o primeiro momento de X é a sua média ponderada μ.

Fazendo $r = 1$, no primeiro membro da igualdade (9) temos

$$E(X - EX)^r = E(X - EX) = EX - E(EX) = EX - EX = 0.$$

Podemos perceber que o primeiro momento central de X é igual a zero, ou seja, $\mu = 0$.

Ao atribuirmos $r = 2$ ao primeiro membro da igualdade (8) , vamos obter EX^2. Este segundo momento de X será denominado de variância pois $E(X) = 0$ e representamos por $VarX$ ou também por σ^2. Para os capítulos seguintes, vamos utilizar o segundo momento central ao invés apenas do segundo momento. Com isso, a partir da igualdade (9), temos

$$\begin{aligned} E(X - EX)^2 &= E[(X)^2 - 2XEX + (EX)^2] \\ &= E(X)^2 - 2EXEX + (EX)^2 \\ &= E(X)^2 - 2(EX)^2 + (EX)^2 \\ &= E(X)^2 - (EX)^2. \end{aligned}$$

Podemos então escrever a variância de X como

$$VarX = E(X)^2 - (EX)^2. \tag{10}$$

Consequentemente, temos que apresentar outra medida que surge a partir do segundo momento que é o desvio padrão. Ao representarmos a variância por σ^2 podemos obter o desvio padrão da seguinte maneira,

$$\sigma = \sqrt{VarX}. \tag{11}$$

Ou seja, o desvio padrão σ é a raiz quadrada da variância.

A seguir, temos algumas propriedades, muito uteis, para os capítulos seguintes.

Teorema 2.4 *Seja X uma variável aleatória com esperança finita.*

(i) $VarX \geq 0$

(ii) Se c é uma constante e $X = c$ *então* $VarX = 0$.

(iii) $Var(aX) = a^2 VarX$*, onde a é uma constante real.*

(iv) Se X e Y são variáveis aleatórias mutuamente independentes, então $Var(X + Y) = VarX + VarY$.

Demonstração :

(i) Por definição, $VarX = E(X - EX)^2$ como todo quadrado é sempre positivo então $VarX \geq 0$.

(ii) Como $Ec = c$, para todo c constante, se $X = c$ então

$$Var(c) = E(c - Ec)^2 = E(c - c)^2 = E0^2 = E0 = 0.$$

(iii) Seja Z' uma variável aleatória tal que $Z' = aX$. Se para uma constante a temos $E(aX) = aE(X)$ então

$$\begin{aligned} Var(aX) = VarZ' &= E(Z' - EZ')^2 \\ &= E(aX - E(aX))^2 \\ &= E(aX - aEX)^2 \\ &= E(a(X - EX))^2 \\ &= E(a^2(X - EX)^2) \\ &= a^2E(X - EX)^2 \\ &= a^2VarX. \end{aligned}$$

(iv) Seja W uma variável aleatória tal que $W = X + Y$. Vamos utilizar a fórmula $VarX = EX^2 - (EX)^2$.

$$\begin{aligned} Var(X + Y) = VarW &= EW^2 - (EW)^2 \\ &= E(X + Y)^2 - [E(X + Y)]^2 \\ &= E(X^2 + 2XY + Y^2) - (EX + EY)^2 \\ &= E(X^2 + 2XY + Y^2) - [(EX)^2 + 2EXEY + (EY)^2] \\ &= EX^2 + E2XEY + EY^2 - (EX)^2 - 2EXEY - (EY)^2 \\ &= EX^2 - (EX)^2 + EY^2 - (EY)^2 + 2EXEY - 2EXEY \\ &= EX^2 - (EX)^2 + EY^2 - (EY)^2 \\ &= VarX + VarY. \end{aligned}$$

■

Finalizando esta etapa, podemos entender porque a variância e o desvio padrão são medidas estatísticas de dispersão. Vejamos, pela igualdade (10), que eles representam a diferença em torno da média que é uma medida central. Na seção 2.5, com a apresentação da curva normal, esta observação torna-se mais clara e coerente.

2.4 Algumas distribuições de variáveis aleatórias

Vamos abordar, nesta seção, três distribuições de variáveis aleatórias discretas que é a distribuição uniforme discreta, a binomial e a distribuição de Poisson e duas distribuições de variáveis aleatórias contínuas que são a distribuição uniforme contínua e a distribuição exponencial. Basicamente, iremos expor suas funções de distribuição, a média e a variância de cada uma delas. Deixamos a distribuição normal para o próximo capítulo, sua importância é crucial para o desenvolvimento dos capítulos destinados a aplicações práticas do Teorema Central do Limite. As demonstrações das proposições apresentadas nesta seção que vão da Proposição 2.2 até a Proposição 2.6 encontram-se em (NATARIO, 2012).

1. **Distribuição uniforme discreta.** Uma variável aleatória X possui distribuição uniformemente discreta se a sua função de densidade é da forma

$$f(x) = \frac{1}{x} \qquad x = 1,\, 2,\, \cdots,\, n.$$

A distribuição uniforme discreta possui apenas o parâmetro n que significa o número de tentativas com a probabilidade de $\frac{1}{n}$ para cada tentativa, isto é, todas as tentativas possuem a mesma probabilidade de ocorrência. Se X segue uma distribuição uniforme discreta, representamos por $X \sim U(n)$.

Proposição 2.2 *Seja X uma variável aleatória com distribuição uniformemente discreta de parâmetro n. Então:*

(i) $EX = \dfrac{n+1}{2}.$

(ii) $VarX = \dfrac{n^2-1}{12}.$

(iii) $F(x) = \sum_{i=1}^{x} \dfrac{1}{n}.$

2. **Distribuição binomial.** Uma variável aleatória X possui distribuição binomial se a sua função de densidade é da forma

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x = 0,\, 1,\, 2,\, \cdots,\, n.$$

Os parâmetros de uma distribuição binomial são os números de tentativas independentes representada por n e a probabilidade p de sucesso em cada uma dessas tentativas. Importante destacar que o complementar de p é $1-p$, que significa a probabilidade de ocorrência de falha ou fracasso numa tentativa. Se X segue uma distribuição binomial representamos por $X \sim Bin(n, p)$.

Proposição 2.3 *Seja X uma variável aleatória com distribuição binomial de parâmetros n e p. Então:*

(i) $EX = np.$

(ii) $VarX = np(1\text{-}p)$.

(iii) $F(x) = \sum_{i=0}^{x} \binom{n}{x} p^x (1-p)^{n-x}$.

3. **Distribuição de Poisson.** Uma variável aleatória X possui distribuição de Poisson se a sua função de densidade é da forma

$$f(x) = e^{-\lambda}\frac{\lambda^x}{x!} \qquad x = 0, 1, 2, \cdots$$

A distribuição de Poisson possui apenas o parâmetro λ que indica a taxa de ocorrência por unidade medida. Por exemplo, λ é o número de ligações numa central telefônica ou o número de atendimentos numa agencia bancária. Se X segue uma distribuição de Poisson, representamos por $X \sim Poisson(\lambda)$.

Proposição 2.4 *Seja X uma variável aleatória com distribuição de Poisson de parâmetro λ. Então:*

(i) $EX = \lambda$.

(ii) $VarX = \lambda$.

(iii) $F(x) = \sum_{i=0}^{x} \dfrac{e^{-\lambda}\lambda^i}{i!}$.

4. **Distribuição uniforme contínua.** Uma variável aleatória X possui distribuição uniformemente contínua no intervalo $[a, b]$ se a sua função de densidade é da forma

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & \text{se} \quad a \leq x \leq b \\ 0, & \text{se} \quad \textit{para outros valores de } x. \end{cases}$$

Os parâmetros de uma distribuição uniformemente contínua são as extremidades do intervalo onde a variável aleatória está contida, ou seja, os valores de a e b em $[a, b]$ onde $-\infty < a \leq X < b < \infty$. Muito utilizado em programação já que grande parte das linguagens de programação ou planilhas de cálculo possuem um gerador de números aleatórios, que gera a partir de uma distribuição uniforme, com valores entre 0 e 1. Logo, se X segue uma distribuição uniformemente contínua, representamos por $X \sim U[a, b]$.

Proposição 2.5 *Seja X uma variável aleatória com distribuição uniforme contínua de parâmetros* a *e* b. *Então:*

(i) $EX = \dfrac{a+b}{2}$.

(ii) $VarX = \dfrac{(b-a)^2}{12}$.

$$(iii)\ \ F(x) = \begin{cases} 0, & se \quad x < a \\ \dfrac{x-a}{b-a}, & se \quad a \leq x < b. \\ 1, & se \quad x \geq b \end{cases}$$

.

5. **Distribuição exponencial.** Uma variável aleatória X possui distribuição exponencial se a sua função de densidade é da forma

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{se} \quad x > 0, \qquad \lambda > 0 \\ 0, & \text{se} \quad x \leq 0. \end{cases}$$

A distribuição exponencial possui apenas o parâmetro λ. Este tipo de distribuição serve de modelo para estudo de tempo de espera, tais como o tempo até a falha de um equipamento ou tempo necessário para completar uma tarefa. Neste caso, λ é o parâmetro de taxa de tempo como, por exemplo, o tempo médio de vida, enquanto que o x é o tempo de falha. Importante salientar que λ e x devem compartilhar da mesma unidade de medida. Se X segue uma distribuição exponencial, representamos por $X \sim Exp(\lambda)$.

Proposição 2.6 *Seja X uma variável aleatória, distribuída exponencialmente, de parâmetro λ. Então:*

(i) $EX = \dfrac{1}{\lambda}$.

(ii) $VarX = \dfrac{1}{\lambda^2}$.

$$(iii)\ \ F(x) = \begin{cases} 0, & se \quad x < 0 \\ 1 - e^{-\lambda x}, & se \quad x \geq 0. \end{cases}$$

2.5 A distribuição normal

A distribuição normal tem todo o mérito em ter uma seção dedicada exclusivamente ao seu estudo. Isso deve-se a grande relevância que ela tem na probabilidade e na estatística inferencial. Vamos observar esta importância nos capítulos que discorrem sobre as aplicações do Teorema Central do Limite. Esta é a última seção deste capítulo e faremos uma breve exposição sobre a distribuição normal, seus parâmetros, sua forma reduzida, a curva normal com suas propriedades e a tabela z. Para mais detalhes, vide (PINHEIRO, 2009).

Uma variável aleatória X é uma variável aleatória normalmente distribuída, se a função densidade de X for dada por:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \qquad x \in \mathbb{R},\ \mu \in \mathbb{R} \quad \text{e} \quad \sigma > 0. \tag{12}$$

A distribuição normal possui dois parâmetros que são μ e σ^2. Vimos, na Seção 2.2 e 2.3 que μ representa a esperança ou valor esperado de uma variável e σ^2 a sua variância. Muitas variáveis aleatórias seguem uma distribuição normal porém, o interessante é que outras distribuições derivam dela tais como a distribuição qui-quadrado, a lognormal e a de Cauchy[2]. Outras distribuições não normais podem se tornar aproximadamente normais a medida que o valor do parâmetro torna-se suficientemente grande. Esta aproximação se deve ao Teorema Central do Limite, no Capítulo 5. Se X segue uma distribuição normal, representamos por $X \sim N(\mu, \sigma^2)$.

Proposição 2.7 *Seja X uma variável aleatória, normalmente distribuída, de parâmetro μ e σ^2 . Então:*

(i) $EX = \mu$.

(ii) $VarX = \sigma^2$.

A demonstração da Proposição 2.7 encontra-se em (NATARIO, 2012).

A Figura 1, a seguir, representa graficamente a função de densidade normal.

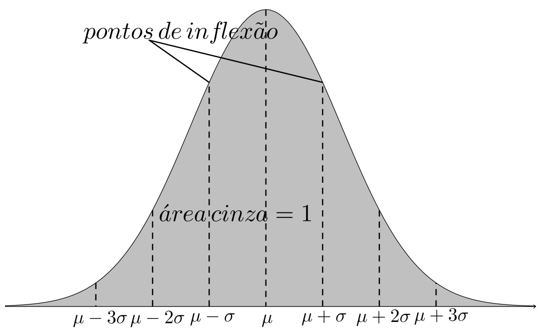


Figura 1 – Curva normal de média μ e desvio padrão σ.

Vejamos algumas características desta curva.

(1) A média, a mediana e a moda numa distribuição normal são iguais.

(2) A curva normal tem o formato de um sino e seu eixo de simetria é paralelo ao eixo das ordenadas e se localiza no valor da média.

(3) Quando a curva normal começa a se afastar da média (eixo de simetria), ela se aproxima cada vez mais do eixo das abscissas porém, a curva nunca o tocará (característica de função exponencial). Com isso dizemos que o eixo das abscissas é assintota da função.

[2] Consultar (FARIAS, 2009) para a qui-quadrado e lognormal e (GARCIA, 2008) para distribuição de Cauchy.

(4) No gráfico, os pontos $\mu - \sigma$ e $\mu + \sigma$ são pontos de inflexão da curva normal já que da esquerda até $\mu - \sigma$ e de $\mu + \sigma$ para a direita a curva tem concavidade voltada para cima enquanto que entre $\mu - \sigma$ e $\mu + \sigma$ a curva tem concavidade voltada para baixo.

(5) A área sob a curva normal determina a probabilidade de ocorrência do evento em questão. A área total sob a curva sendo igual a 1 quer dizer que a probabilidade de ocorrência é de 100%. Entre os pontos simétricos no eixo das abscissas temos o seguinte:

- Se a área da curva está entre $\mu - \sigma$ e $\mu + \sigma$, a probabilidade é igual a 68.26%.
- Se a área da curva está entre $\mu - 2\sigma$ e $\mu + 2\sigma$, a probabilidade é igual a 95.44%.
- Se a área da curva está entre $\mu - 3\sigma$ e $\mu + 3\sigma$, a probabilidade é igual a 99.73%.
- Se a área da curva está entre $\mu - 4\sigma$ e $\mu + 4\sigma$, a probabilidade é igual a 99.994%.

A igualdade (12) é a função de uma variável aleatória com distribuição normal. Fazendo uma substituição do tipo:

$$z = \frac{x - \mu}{\sigma}, \tag{13}$$

temos

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-z^2}{2}} dz.$$

Ou seja,

$$g(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}. \tag{14}$$

Ao compararmos as Funções (12) e (14) podemos observar que $g(z)$ possui todas as características de uma função com distribuição normal porém esta afirmação só é válida se $\mu = 0$ e $\sigma^2 = 1$. Com isso, temos uma variável aleatória Z, normalmente distribuída, com uma média fixada em zero e um desvio padrão unitário, isto é, $Z \sim N(0, 1)$. A Função (14) é denominada de função com distribuição normal padrão ou normal reduzida de abscissa z.

O gráfico de uma função com distribuição normal reduzida tem o mesmo comportamento e características que o gráfico da curva normal de parâmetros μ e σ, como podemos observar pela Figura 2.

A tabela z, apresentada na Figura 3, nos fornece a probabilidade de $Z \leq z$ que nada mais é que a área sob a curva normal de acordo com os valores de $z \geq 0$. Para valores negativos de z, podemos utilizar as duas igualdades abaixo que são originárias a partir da propriedade de simetria da curva normal em relação a média.

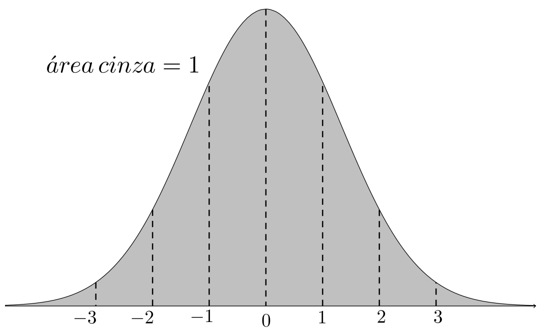


Figura 2 – Curva normal padrão com $\mu = 0$ e $\sigma = 1$.

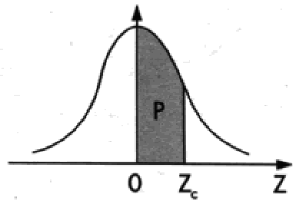
- ❏ $P(Z \leq -z) = 0.5 - P(Z \leq z)$.

As provas do Teorema Central do Limite necessitam de outros conceitos para o seu desenvolvimento. O conceito de convergência em distribuição e de funções características tem teoremas que o regem. No Capítulo 3, estes teoremas serão enunciados e demonstrados com o rigor que eles merecem.

Tabela III — Distribuição Normal Padrão

$Z \sim N(0, 1)$

Corpo da tabela dá a probabilidade p, tal que $p = P(0 < Z < Z_c)$



parte inteira e primeira decimal de Z_c	Segunda decimal de Z_c 0	1	2	3	4	5	6	7	8	9	parte inteira e primeira decimal de Z_c
	p = 0										
0,0	00000	00399	00798	01197	01595	01994	02392	02790	03188	03586	0,0
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535	0,1
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409	0,2
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173	0,3
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793	0,4
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240	0,5
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490	0,6
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524	0,7
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327	0,8
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891	0,9
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214	1,0
1,1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298	1,1
1,2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147	1,2
1,3	40320	40490	40658	40824	40988	41149	41309	41466	41621	41774	1,3
1,4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189	1,4
1,5	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408	1,5
1,6	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449	1,6
1,7	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327	1,7
1,8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062	1,8
1,9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670	1,9
2,0	47725	47778	47831	47882	47932	47982	48030	48077	48124	48169	2,0
2,1	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574	2,1
2,2	48610	48645	48679	48713	48745	48778	48809	48840	48870	48899	2,2
2,3	48928	48956	48983	49010	49036	49061	49086	49111	49134	49158	2,3
2,4	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361	2,4
2,5	49379	49396	49413	49430	49446	49461	49477	49492	49506	49520	2,5
2,6	49534	49547	49560	49573	49585	49598	49609	49621	49632	49643	2,6
2,7	49653	49664	49674	49683	49693	49702	49711	49720	49728	49736	2,7
2,8	49744	49752	49760	49767	49774	49781	49788	49795	49801	49807	2,8
2,9	49813	49819	49825	49831	49836	49841	49846	49851	49856	49861	2,9
3,0	49865	49869	49874	49878	49882	49886	49889	49893	49897	49900	3,0
3,1	49903	49906	49910	49913	49916	49918	49921	49924	49926	49929	3,1
3,2	49931	49934	49936	49938	49940	49942	49944	49946	49948	49950	3,2
3,3	49952	49953	49955	49957	49958	49960	49961	49962	49964	49965	3,3
3,4	49966	49968	49969	49970	49971	49972	49973	49974	49975	49976	3,4
3,5	49977	49978	49978	49979	49980	49981	49981	49982	49983	49983	3,5
3,6	49984	49985	49985	49986	49986	49987	49987	49988	49988	49989	3,6
3,7	49989	49990	49990	49990	49991	49991	49992	49992	49992	49992	3,7
3,8	49993	49993	49993	49994	49994	49994	49994	49995	49995	49995	3,8
3,9	49995	49995	49996	49996	49996	49996	49996	49996	49997	49997	3,9
4,0	49997	49997	49997	49997	49997	49997	49998	49998	49998	49998	4,0
4,5	49999	50000	50000	50000	50000	50000	50000	50000	50000	50000	4,5

Figura 3 – Tabela normal padrão para $z \geq 0$. Fonte: (MORETTIN, 2004).

3

Função característica e convergência em distribuição

No Capítulo 2, abordamos sobre a esperança e a variância de variáveis aleatórias com suas definições e propriedades muito bem definidas. Também vimos que se tratam de momentos de primeira e de segunda ordem respectivamente, isto é, para uma variável X, se a esperança existe então o primeiro e o segundo momento de X são representados por $\mathrm{E}(X)$ e $\mathrm{E}(X)^2$. Verificamos, também, onde estes momentos assumem papel fundamental dentro da estatística e probabilidade sendo esta na distribuição normal, um tipo de distribuição que abrange muitos fenômenos, dentre eles, fenômenos biológicos e sociais. Com isso, podemos apresentar um conceito muito importante para determinar distribuições de probabilidades de variáveis aleatórias, já que estas funções de distribuição muitas vezes, podem ser expressas através de seus momentos. Estamos falando de função geradora de momentos[1], a primeira parte do desenvolvimento deste capítulo. Em seguida, trataremos das funções características onde precisaremos enunciar e provar o Teorema da Unicidade e, por fim, sobre convergência em distribuição, onde será apresentado o Teorema da Continuidade de Levy fechando assim todo o aparato necessário para o sucesso no desenvolvimento do capítulo seguinte que trata-se do Teorema Central do Limite.

3.1 Função geradora de momentos

Muitos autores pesquisados não dão uma importância muito significativa para a função geradora de momentos devido ao fato que, em algumas distribuições, como por exemplo, a distribuição de Cauchy, ela não existe, isto é, não converge para um valor finito, fazendo-se necessário garantir primeiro sua existência. Verificamos que é necessário apresentar o conceito de função geradora de momentos pois dá suporte para os conceitos de função característica e pela sua importância na prova do Teorema Central do Limite de DeMoivre e Laplace.

[1] Para mais esclarecimentos, vide (MEYER, 1970).

A definição seguinte apresenta a função geradora de momentos para o caso discreto e o caso contínuo.

Definição 3.1 *Seja X uma variável aleatória com função de distribuição $p(x_j)$ para $j = 1, 2, 3, \cdots$ no caso discreto, ou função de densidade h(x) no caso contínuo. Representando a função geradora de momentos por $M_X(t)$, se a variável aleatória possui distribuição discreta então*

$$M_X(t) = \sum_{j=1}^{\infty} e^{tx_j} p(x_j).$$

Se a variável aleatória possui distribuição contínua então

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} h(x).$$

Porém, observe que em ambos os casos, $M_X(t)$ representa a esperança de e^{tx} logo, escreveremos a função geradora de momentos como

$$M_X(t) = Ee^{tX},$$

onde e^{tX} possui esperança finita.

Antes de continuarmos, vamos observar porque a função $M_X(t)$ possui o nome de função geradora de momentos.

Para garantirmos a existência de $M_X(t)$, vamos supor $-t_0 < t < t_0$ e que, para algum t_0 positivo, $M_X(t)$ seja finita.

A função e^x, escrita em Série de Maclaurin, é representada como

$$e^x = 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + \cdots . \tag{15}$$

Aplicando a igualdade (15) para a função e^{tx},

$$e^{tx} = 1 + tx + \frac{(tx)^2}{2!} + \cdots + \frac{(tx)^n}{n!} + \cdots = \sum_{n=0}^{\infty} \frac{t^n X^n}{n!}. \tag{16}$$

Com isso,

$$M_X(t) = Ee^{tx} = E\left(1 + tx + \frac{(tx)^2}{2!} + \cdots + \frac{(tx)^n}{n!} + \cdots\right), \tag{17}$$

ou seja,

$$M_X(t) = E\sum_{n=0}^{\infty} \frac{t^n X^n}{n!} = \sum_{n=0}^{\infty} \frac{EX^n}{n!} t^n, \tag{18}$$

onde $\frac{EX^n}{n!}$ representa o coeficiente de t^n. Escrevendo uma função definida por *f(x)* em Série de Maclaurin, teremos

$$f(x) = 1 + f'(0)t + \frac{f''(0)t^2}{2!} + \cdots + \frac{f^{(n)}(0)t^n}{n!} + \cdots .$$

Sendo assim, $M_X(t)$ será escrito como

$$M_X(t) = 1 + M'_X(0)t + \cdots + \frac{M_X^{(n)} t^n}{n!} + \cdots, \tag{19}$$

ou seja,

$$M_X(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} \frac{d^n}{dt^n} M_X(t)\,[t = 0], \tag{20}$$

onde $\frac{1}{n!}\frac{d^n}{dt^n} M_X(t)\,[t = 0]$ é o coeficiente de t^n.

Se igualarmos os coeficientes de t^n nas igualdades (18) e (20), podemos notar que

$$\frac{1}{n!}\frac{d^n}{dt^n} M_X(t)\,[t = 0] = \frac{EX^n}{n!} \Longleftrightarrow EX^n = \frac{d^n}{dt^n} M_X(t)\,[t = 0]. \tag{21}$$

Isso mostra que a função $M_X(t)$ para $t = 0$ gera momentos de X de acordo com a sua ordem de derivação.

Eis algumas propriedades relevantes para o desenvolvimento deste trabalho.

Propriedade 3.1 *Para $t = 0$, $M_X(0) = 1$.*

Prova : É imediato pois, da igualdade (19), ao atribuirmos $t = 0$, todas as parcelas, a partir da segunda, reduzir-se-à a zero, isto é,

$$M_X(0) = 1 + M'_X(0)0 + \frac{M_X^{(2)} 0^2}{2} + \cdots + \frac{M_X^{(n)} 0^n}{n!} + \cdots.$$

Assim, $M_X(0) = 1$.

■

Propriedade 3.2 *Sejam X e Y variáveis aleatórias independentes. As funções e^{tX} e e^{tY} também são independentes. Com isso,*

$$M_{X+Y}(t) = M_X(t)M_Y(t).$$

Prova :

$$\begin{aligned} M_{X+Y}(t) &= Ee^{t(X+Y)} = Ee^{tX+tY} \\ &= Ee^{tX}e^{tY} = Ee^{tX}Ee^{tY} \\ &= M_X(t)M_Y(t). \end{aligned}$$

■

A Propriedade 3.2 pode ser expandida para as variáveis $X_1, X_2, \cdots, X_n$, todas independentes, onde

$$M_{X_1+X_2+\cdots+X_n}(t) = M_{X_1}(t)M_{X_2}(t)\cdots M_{X_n}(t).$$

Em seguida, veremos dois exemplos envolvendo densidades normais. No primeiro exemplo, devemos encontrar a função geradora de momentos de uma variável com distribuição normal pois com este resultado vamos, no segundo exemplo, encontrar os momentos desta variável.

Exemplo 3.1 *Encontre a função geradora de momentos de uma variável aleatória X que possui distribuição normal de média μ e variância σ^2.*

Solução : Recordando a função densidade de uma variável com distribuição normal

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

Com isso podemos escrever $M_X(t)$ como

$$M_X(t) = Ee^{tX} = \int_{-\infty}^{\infty} e^{tx} f(x)dx = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx. \tag{22}$$

Fazendo $x - \mu = y$ temos que $x = y + \mu$ e $dy = dx$. Substituindo, na última integral da Igualdade (22), teremos

$$\begin{aligned}
\int_{-\infty}^{\infty} e^{tx} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx &= \int_{-\infty}^{\infty} e^{t(y+\mu)} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{y^2}{2\sigma^2}} dy \\
&= e^{t\mu} \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{ty} e^{-\frac{y^2}{2\sigma^2}} dy \\
&= e^{t\mu} \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{ty2\sigma^2}{2\sigma^2} - \frac{y^2}{2\sigma^2}} dy. \qquad (23)
\end{aligned}$$

Reescrevendo o expoente do integrando da integral dada em (23),

$$\begin{aligned}
\frac{ty2\sigma^2}{2\sigma^2} - \frac{y^2}{2\sigma^2} &= -\left[\frac{y^2}{2\sigma^2} - \frac{(2y)t\sigma^2}{2\sigma^2}\right] \\
&= -\left[\frac{y^2}{2\sigma^2} - \frac{(2y)t\sigma^2}{2\sigma^2} + \frac{(t\sigma^2)^2}{2\sigma^2} - \frac{(t\sigma^2)^2}{2\sigma^2}\right] \\
&= -\frac{(y-(t\sigma^2)^2)}{2\sigma^2} + \frac{(t\sigma^2)^2}{2\sigma^2} \\
&= -\frac{(y-(t\sigma^2)^2)}{2\sigma^2} + \frac{t^2\sigma^2}{2}.
\end{aligned}$$

Voltando para o desenvolvimento da integral em (23).

$$\begin{aligned}
e^{t\mu} \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{ty2\sigma^2}{2\sigma^2} - \frac{y^2}{2\sigma^2}} dy &= e^{t\mu} \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-(t\sigma^2)^2)}{2\sigma^2} + \frac{t^2\sigma^2}{2}} dy = \\
&= e^{t\mu} e^{\frac{t^2\sigma^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-(t\sigma^2)^2)}{2\sigma^2}} dy.
\end{aligned}$$

Logo,

$$M_X(t) = e^{t\mu} e^{\frac{t^2\sigma^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-(t\sigma^2)^2)}{2\sigma^2}} dy; \quad -\infty < t < \infty.$$

Observe que o integrando de $M_X(t)$ é uma função de densidade normal cujos parâmetros são $\mu = t\sigma^2$ e σ^2 e, pela igualdade (1), do Capítulo 2, esta integral é igual a 1. Com isso,

$$\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-(t\sigma^2)^2)}{2\sigma^2}} dy = 1$$

e

$$M_X(t) = e^{t\mu} e^{\frac{t^2\sigma^2}{2}}. \tag{24}$$

Note que a função geradora de momentos de X com distribuição normal também possui parâmetros μ e σ^2 assim como a função de distribuição normal.

No exemplo seguinte, podemos verificar como se comporta a função geradora de momentos para a variável aleatória $X \sim N(0, \sigma^2)$.

Exemplo 3.2 *Seja X uma variável aleatória com distribuição normal de média $\mu = 0$ e variância σ^2. Encontre os momentos da variável X através da função geradora de momentos.*

Solução : Pelo Exemplo 3.1 temos que

$$M_X(t) = e^{t\mu} e^{\frac{t^2\sigma^2}{2}}, \quad -\infty < t < \infty$$

Como $\mu = 0$,

$$M_X(t) = e^{\frac{t^2\sigma^2}{2}}, \quad -\infty < t < \infty.$$

Temos, pela igualdade (16), que

$$e^{tX} = \sum_{n=0}^{\infty} \frac{t^n X^n}{n!}.$$

Logo,

$$\begin{aligned} e^{\dfrac{t^2\sigma^2}{2}} &= \sum_{n=0}^{\infty} \left[\frac{t^2\sigma^2}{2}\right]^n \frac{1}{n!} \\ &= \sum_{n=0}^{\infty} \frac{t^{2n}\sigma^{2n}}{2^n n!} \\ &= \sum_{n=0}^{\infty} \frac{\sigma^{2n}}{2^n n!} t^{2n}. \end{aligned}$$

Pela segunda igualdade apresentada na equivalência em (21) onde

$$\frac{EX^n}{n!} = \frac{1}{n!}\frac{d^n}{dt^n} M_X(t)\,[t=0],$$

temos que,

$$\frac{EX^{2n}}{(2n)!} = \frac{\sigma^{2n}}{2^n n!}$$

e

$$EX^{2n} = \frac{\sigma^{2n}(2n)!}{2^n n!}. \tag{25}$$

Podemos observar, numa distribuição normal com N(0, σ^2), que os momentos de ordem par são dados pela igualdade (25) e os momentos de ordem ímpar são todos iguais a zero inclusive por hipótese, a média, que é o momento de primeira ordem, é igual a zero.

3.2 Função característica

Nesta seção, vamos apresentar algumas particularidades acerca da função característica, estabelecendo relações diretas com a função geradora de momentos, enunciar o Teorema da Fórmula de Inversão necessária para encontrar funções de densidades de uma variável a partir de sua função característica e o Teorema da Unicidade.

Eis a definição formal da função característica,

Definição 3.2 *A função característica φ, da variável aleatória X, é a função $\varphi : \mathbb{R} \to \mathbb{C}$ onde, para cada x real do domínio, temos $\varphi_X(t)$ como imagem, tal que,*

$$\varphi_X(t) = Ee^{itX}. \tag{26}$$

Observe que a função característica é definida no conjunto dos números complexos. Todos os autores cujas obras estão contidas na referência deste trabalho e que escreveram sobre funções características, apresentaram algumas propriedades de números complexos para mostrar que não haverá mudanças algébricas bruscas pois muitas propriedades são semelhantes aos dos números reais. Vide JAMES, (2010).

Na Seção 3.1, vimos que para um t do domínio, a função geradora de momentos pode não convergir para um valor finito. Por isso precisamos, antes de enunciar uma proposição, definir que a função geradora de momentos $M_X(t)$ seja finita. Vamos ver que a função característica de uma variável aleatória sempre será finita, ou seja, sempre está definida no domínio da função.

No Teorema 2.2 do Capítulo 2, vimos que se X e Y são duas variáveis aleatórias ambas com esperanças finitas então a variável resultante da soma $X+Y$ terá esperança finita. Sejam X_1 e X_2 duas variáveis aleatórias ambas com esperanças finitas e seja X' uma variável tal que $X' = X_1 + iX_2$, onde $i = \sqrt{-1}$. O número i é chamado de número imaginário ou coeficiente complexo. Denominamos X' de variável aleatória complexa que também terá esperança finita, já que X_1 e X_2 também os têm.

Se aplicarmos o fato explicitado no parágrafo anterior que a soma de variáveis com esperança finita possui esperança finita então, escrevendo e^{itx} através da Fórmula de Euler,

$$e^{itx} = cos(tx) + isen(tx). \tag{27}$$

A norma de e^{itx} é calculada como

$$\begin{aligned} |e^{itx}|^2 &= cos^2(tx) + sen^2(tx) \\ &= [cos^2(tx) + sen^2(tx)]^{\frac{1}{2}}. \end{aligned}$$

Pela Relação Fundamental da Trigonometria, $sen^2x + cos^2x = 1$. Logo,

$$[cos^2(tx) + sen^2(tx)]^{\frac{1}{2}} = [1]^{\frac{1}{2}} = 1,$$

ou seja, a função e^{itx} possui norma unitária. Como a esperança de uma constante é a própria constante, temos que

$$|\varphi_X(t)| = |Ee^{itx}| \leq E|e^{itx}| = E1 = 1, \quad \text{ou seja,} \quad |\varphi_X(t)| \leq 1.$$

Isso nos mostra que as funções características são limitadas. Logo, podemos afirmar que funções características possuem esperança finita $\forall t$ tal que $-\infty < t < \infty$.

Vamos enunciar agora algumas propriedades das funções características.

Propriedade 3.3 *Se em seu domínio $t = 0$ então a função característica é igual a 1, ou seja,* $\varphi(0) = 1$.

Prova : Para $t = 0$,

$$\varphi(0) = Ee^{i.0.x} = Ee^0 = E1 = 1.$$

■

Propriedade 3.4 *Sendo X e Y variáveis aleatórias independentes e, e^{itX} e e^{itY} também ambas independentes,*

$$\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t) \qquad \forall t \in \mathbb{R}.$$

Prova : Para X e Y independentes, temos,

$$\begin{aligned} \varphi_{X+Y}(t) &= Ee^{it(X+Y)} = Ee^{itX+itY} = \\ &= Ee^{itX}e^{itY} = Ee^{itX}Ee^{itY} = \\ &= \varphi_X(t)\varphi_Y(t). \end{aligned}$$

■

Da Propriedade 3.4, temos duas observações a fazer:

1. As propriedades com números reais utilizadas na prova desta propriedade estendem-se para os números complexos.

2. Por indução, podemos generalizar a propriedade para mais de duas variáveis, i.é,

$$\varphi_{X_1+X_2+X_3+\cdots+X_n}(t) = \prod_{j=1}^{n} \varphi_{X_j}(t).$$

Propriedade 3.5 *Seja Y uma variável aleatória em função de X, tal que $Y = aX + b$, com a e b constantes reais, assim,*

$$\varphi_Y(t) = e^{itb}\varphi_X(at).$$

Prova : Como $Y = aX + b$, temos que

$$\begin{aligned} \varphi_Y(t) = \varphi_{aX+b}(t) &= Ee^{it(aX+b)} = Ee^{itaX+itb} = \\ &= Ee^{itaX}e^{itb} = e^{itb}Ee^{i(at)X} = \\ &= e^{itb}\varphi_X(at). \end{aligned}$$

■

Não apresentaremos a prova desta propriedade, deixando ao leitor que tiver interesse consultar o sexto capítulo em (JAMES, 2010).

Propriedade 3.6 *Se o momento do módulo da variável X de ordem n existe então a função característica de X possui n derivadas contínuas, isto é,*

$$E|X|^n < \infty \Longleftrightarrow \varphi_x(t) = \int (ix)^k e^{itx} dF_k(X), \quad k = 1, 2, \cdots, n.$$

O exemplo seguinte nos dará a função característica da variável X quando esta se distribui normalmente.

Exemplo 3.3 *Seja X uma variável aleatória que possui distribuição normal de parâmetros $\mu = 0$ e variância σ^2. Vamos obter a função $\varphi_X(t)$.*

Solução : Através da expansão em Série de Maclaurin para a exponencial, apresentada na Igualdade (17), temos que

$$\begin{aligned} \varphi_X(t) = Ee^{itX} &= E\sum_{n=0}^{\infty} \frac{[itX]^n}{n!} \\ &= \sum_{n=0}^{\infty} \frac{i^n EX^n}{n!} t^n. \end{aligned}$$

Na Seção 3.1, Exemplo 3.2, vimos que uma variável que se distribui normalmente com média 0 e variância σ^2 possui apenas momentos de ordem $2n$, já que todos os momentos de ordem ímpar são iguais a zero e esses momentos são da forma

$$EX^{2n} = \frac{\sigma^{2n}(2n)!}{2^n n!}.$$

Sendo assim,

$$\begin{aligned}
\sum_{n=0}^{\infty} \frac{i^n EX^n}{n!} t^n &= \sum_{n=0}^{\infty} \frac{i^{2n} EX^{2n}}{(2n)!} t^{2n} = \sum_{n=0}^{\infty} \frac{i^{2n} \dfrac{\sigma^{2n}(2n)!}{2^n n!}}{(2n)!} t^{2n} = \\
&= \sum_{n=0}^{\infty} \frac{i^{2n}\sigma^{2n}(2n)!}{2^n n!(2n)!} t^{2n} = \sum_{n=0}^{\infty} \frac{i^{2n}\sigma^{2n}}{2^n n!} t^{2n} = \\
&= \sum_{n=0}^{\infty} \left[\frac{i^2\sigma^2 t^2}{2}\right]^n \frac{1}{n!} = \sum_{n=0}^{\infty} \left[\frac{(\dfrac{-t^2\sigma^2}{2})}{n!}\right]^n = \\
&= e^{\dfrac{-t^2\sigma^2}{2}}.
\end{aligned}$$

Portanto, a função característica de uma variável aleatória $X \sim \mathrm{N}(0, \sigma^2)$ é dada por

$$\varphi_X(t) = e^{\dfrac{-t^2\sigma^2}{2}}. \tag{28}$$

O Exemplo 3.4 mostrará uma relação interessante entre M_X e φ_X.

Exemplo 3.4 *Seja X uma variável aleatória tal que $X \sim N(\mu, \sigma^2)$, mostrar que*

$$M_X(it) = \varphi_X(t), \quad t \in \mathbb{R}.$$

Solução : Lembrando da igualdade (24) para variáveis com distribuições normais, onde

$$M_X(t) = e^{t\mu} e^{\frac{t^2\sigma^2}{2}}.$$

Podemos dizer que,

$$\begin{aligned}
M_X(it) &= e^{it\mu} e^{\frac{(it)^2\sigma^2}{2}} \\
&= e^{it\mu} e^{\frac{i^2 t^2\sigma^2}{2}} \\
&= e^{it\mu} e^{\frac{-t^2\sigma^2}{2}}.
\end{aligned}$$

Isto é,

$$M_X(it) = e^{it\mu} e^{\frac{-t^2\sigma^2}{2}}. \tag{29}$$

A igualdade (29) logo acima é a forma geral de $\varphi_X(t)$ na igualdade (28) para $X \sim N(\mu, \sigma^2)$. Portanto,

$$M_X(it) = e^{it\mu} e^{\frac{-t^2\sigma^2}{2}} = \varphi_X(t).$$

A propriedade enunciada a seguir não seguiu a ordem das propriedades desta seção pois possui uma particularidade muito interessante que é o fato de ser consequência da Fórmula de Inversão, um método capaz de obter a função de distribuição de uma variável aleatória através de sua função característica. Já sabemos que, através da função de distribuição de uma variável, podemos obter a sua função característica. Basta reescrever a Definição 3.2 como

$$\varphi_X(t) = \int_{-\infty}^{\infty} e^{itX} dF_X, \quad t \in \mathbb{R}.$$

Apenas como lembrete, para que uma função *f* seja contínua em um ponto a é necessário que a função esteja definida em a e que os valores de *f*, para x próximos de a, estejam próximos de *f(a)*. Assim dizemos que a é um ponto de continuidade da função *f*.

Antes de apresentar o Teorema da Fórmula de Inversão, vamos enunciar o Teorema da Convergência Dominada. Este teorema não será demonstrado já que não houve utilidade direta no desenvolvimento do Teorema Central do Limite embora o Teorema da Convergência Dominada, também conhecida como de Teorema da Convergência Dominada de Lebesgue seja um importante resultado da teoria da medida que nos permite garantir a integrabilidade de uma função mensurável.

Teorema 3.1 (Teorema da Convergência Dominada) *Seja $f_n : E \to \mathbb{R}$ uma sequência de funções integráveis que converge, quase sempre em E, para uma função real mensurável f. Se existe uma função integrável g tal que $|f_n| \leq g$ para todo n natural, então f é integrável e*

$$\int_E f_n(x)dx \longrightarrow \int_E f(x)dx.$$

A demonstração encontra-se em (JAMES, 2010).

O teorema seguinte enuncia a Fórmula de Inversão que é conhecido também como a transformada inversa de Fourrier. Antes de enunciá-lo, vamos ressaltar alguns tópicos essenciais para o seu entendimento.

(i) Uma função F tem um limite L à direita ou à esquerda no ponto x se $F(x + h) \to L$ quando $h \to 0$. Quando estes limites existem, representamos o limite à direita como $F(x^+)$ e o limite à esquerda como $F(x^-)$.

(ii) Seja $\tilde{F}(c)$ tal que

$$\tilde{F}(c) = \frac{1}{2}(F(c^+) + F(c^-)), \forall c \in \mathbb{R}$$

com x, y e u sendo números reais e $u > 0$ e $x < y$.

(iii) Caso F seja contínua em c, dizemos que $\tilde{F}(c) = F(c)$.

(iv) Fazendo

$$h(t) = \begin{cases} \dfrac{e^{-itx} - e^{-ity}}{it}, & \text{quando} \quad x \neq 0 \\ y - x, & \text{quando} \quad x = 0 \end{cases} \tag{30}$$

Temos que *h(t)* é uma função contínua e limitada para todo t real. Vejamos:

$$|h(t)| = \left| \frac{e^{-itx} - e^{-ity}}{it} \right|.$$

Como

$$\left| e^{\frac{i(x+y)t}{2}} \right| = 1,$$

temos que

$$\begin{aligned} |h(t)| &= \left| e^{\frac{i(x+y)t}{2}} \right| \left| \frac{e^{-itx} - e^{-ity}}{it} \right| \\ &= \left| \frac{e^{\frac{1}{2}i(y-x)t} - e^{\frac{1}{2}i(x-y)t}}{it} \right|. \end{aligned}$$

Pela Fórmula de Euler,

$$\left| \frac{e^{\frac{1}{2}i(y-x)t} - e^{\frac{1}{2}i(x-y)t}}{it} \right| = \left| \frac{2sen\left[\frac{(y-x)t}{2}\right]}{t} \right|.$$

Como $|sen\alpha| \leq \alpha$, $\forall \alpha \in \mathbb{R}$, então

$$\left| \frac{2sen\left[\frac{(y-x)t}{2}\right]}{t} \right| = \left|\frac{2}{t}\right| \left| sen\left[\frac{(y-x)t}{2}\right] \right| \leq \frac{2}{t}\frac{(y-x)t}{2} \leq y - x.$$

Teorema 3.2 (Fórmula de Inversão) *Dado uma variável aleatória X que possui uma função de distribuição F_X e função característica φ_X, para x e y, pontos de continuidade de F_X, com $x < y$, temos que*

$$\frac{1}{2\pi} \lim_{u \to \infty} \int_{-u}^{u} \frac{e^{-itx} - e^{-ity}}{it} \varphi(t) dt = \tilde{F}(y) - \tilde{F}(x) \tag{31}$$

A demonstração encontra-se em (JAMES, 2010).

O próximo teorema é o Teorema da Unicidade. Este nos mostra que a função de distribuição de uma variável aleatória arbitrária é determinada por sua função característica correspondente.

Teorema 3.3 (Teorema da Unicidade) *Sejam X e Y duas variáveis cujas funções características são φ_X e φ_Y e de funções de distribuição igual a F_X e F_Y, respectivamente. Se $\varphi_X = \varphi_Y$, então $F_X = F_Y$.*

Demonstração : Considere $\tilde{F}(\omega) = \frac{1}{2}(F(\omega^+) + F(\omega^-)), \forall \omega \in \mathbb{R}$. Caso F seja contínua em ω, dizemos que $\tilde{F}(\omega) = F(\omega)$.

Pelo Teorema 3.2, onde a e $b \in \mathbb{R}$ com $a < b$,

$$\tilde{F}_X(b) - \tilde{F}_X(a) = \tilde{F}_Y(b) - \tilde{F}_Y(a) \tag{32}$$

e, fazendo a tender a menos infinito, temos $\lim_{a \to -\infty} \tilde{F}_X(a) \longrightarrow 0$ e $\lim_{a \to -\infty} \tilde{F}_Y(a) \longrightarrow 0$.

Com isso, a igualdade (32) reduz-se a $\tilde{F}_X(b) = \tilde{F}_Y(b), \forall b \in \mathbb{R}$.

Seja $c < b$, pela monotonicidade de F_X e F_Y e pela definição de $\tilde{F}$,

$$F_X(c) \leq \tilde{F}_X(b) \leq F_X(b) \quad \text{e} \quad F_Y(c) \leq \tilde{F}_Y(b) \leq F_Y(b).$$

Pela continuidade à direita da função de distribuição[2],

$$\lim_{b \downarrow c} \tilde{F}_X(b) = F_X(c) \quad \text{e} \quad \lim_{b \downarrow c} \tilde{F}_Y(b) = F_Y(c).$$

Então, $F_X(c) = F_Y(c) \quad \forall c \in \mathbb{R}$.

Na próxima seção, trataremos da convergência em distribuição, conceito muito importante, indispensável para o sucesso das provas apresentadas no Capítulo 4.

3.3 Convergência em distribuição

Esta seção é de fundamental importância para os estudos deste trabalho já que as demonstrações apresentadas no capítulo seguinte se baseiam fortemente neste conceito e suas consequências. O principal objetivo desta seção é a prova do Teorema da Continuidade de Paul Levy.

O termo convergência refere-se a tendência de vários aspectos de se identificarem num ponto, mais especificamente, uma sequência ou uma série é dita convergente quando a quantidade n de seus termos cresce indefinidamente e a sequência ou a série se aproxima cada vez mais de um número real (ou complexo). Este conceito está estreitamente relacionado ao conceito de limite do Cálculo Diferencial. Por isso para maiores esclarecimentos consultar (GUIDORIZZI, 2006).

Referindo-se a convergência de variáveis aleatórias há três principais tipos de convergências que é a convergência quase certa, a convergência em probabilidade e a convergência em distribuição. Esta última como ferramenta de estudos. Os outros dois tipos: a quase certa e em probabilidade, serão apresentados apenas como informes complementares de maneira bem breve. Para mais informação, vide (COLETTI, 2008).

Vamos considerar $X_1, X_2, \cdots$ e X como variáveis aleatórias contidas num espaço de probabilidade (ω, A, P).

[2] $b \downarrow c$ significa que o valor de b decresce até o valor de c, já que $c < b$.

1. **Convergência quase certa.**

   X_n converge para X quase certamente, denotado por $X_n \xrightarrow{q.c} X$, se o evento

   $$\{\omega \in \Omega : X_n(\omega) \xrightarrow{n\to\infty} X(\omega)\}$$

   tem probabilidade 1.

   Este modo de convergência é a base para o Teorema da Lei dos Grandes Números juntamente com a Lei Forte de Kolmogorov e o Lema de Borel-Cantelli.

2. **Convergência em probabilidade.**

   X_n converge para X em probabilidade, denotado por $X_n \xrightarrow{P} X$, se para qualquer $\epsilon > 0$,

   $$P(|X_n - X| > \epsilon) \xrightarrow{n\to\infty} 0.$$

   A convergência em probabilidade está relacionada com a Lei Fraca dos Grandes Números, onde o principais conceitos utilizados são as desigualdades de Markov e de Chebyshev [3].

O interessante na convergência quase certa não é dizer que

$$\forall\omega \in \Omega, X_n(\omega) \xrightarrow{n\to\infty} X(\omega)$$

e sim que a probabilidade do evento $D = \{\omega \in \Omega : |Z(\omega)| \geq 1\}$, denominado conjunto de exceção, seja igual a zero.

Na convergência em probabilidade, a idéia da definição nos retrata que a distância entre X_n e X quanto menor, maior é a probabilidade de acontecer e quando tiver muito próximas, a probabilidade é bem alta.

A definição de convergência em distribuição será apresentada com maior rigor pela sua importância dentro do nosso estudo. Vejamos formalmente a sua definição.

Definição 3.3 (Convergência em Distribuição) *As variáveis aleatórias $X_1, X_2, \cdots, X_n$ da sequência $(X_n)_{n\in\mathbb{N}}$ converge em distribuição para a variável aleatória X se*

$$\lim_{n\to\infty} F_n(X) = F(X).$$

Reescrevendo a Definição 3.3 a fim de reforçar a relação direta da variável aleatória e sua função de distribuição:

$$\text{Para } n \in \mathbb{N} \quad, F_n(X) \xrightarrow{D} F(X) \quad \Longrightarrow \quad X_n \xrightarrow{D} X, \quad \text{quando } n \to \infty.$$

[3] Para os lema de Borel-Cantelli e a Desigualdade de Chebyshev consultar (FELLER, 1968).

A convergência em distribuição relaciona as funções de distribuições das variáveis ao invés da variável em si, ela nos mostra que quanto maior o n, mais a sequência das funções de distribuições tende a função de distribuição propriamente dita. Pelo fato da condição de convergência se dar através das funções de distribuições de X_n e X, este modo de convergência é considerado o mais fraco entre os três citados nesta seção. Por esse motivo, quando uma sequência de variáveis converge em distribuição para uma variável dizemos que converge fracamente.

A seguir, vamos apresentar a definição de limite de sequência. (GUIDORIZZI, 2006).

Definição 3.4 *Sejam (a_n) uma sequência e a um número real. Se $\lim\limits_{n\to\infty} a_n = a$, então dizemos que (a_n) converge para a. Assim, definimos que para todo $\epsilon > 0$, existe um $n_0 \in \mathbb{N}$ onde,*

$$\lim_{n\to\infty} a_n = a \Longleftrightarrow n > n_0 \Longrightarrow |a_n - a| < \epsilon. \tag{33}$$

O ponto principal desta seção é o Teorema da Continuidade de Paul Levy. Para obtermos a prova com o rigor matemático que ele merece, vamos precisar do resultado de um outro teorema que é o Teorema de Helly-Bray. Este nos mostra que a convergência de funções de distribuições de uma sequência de suas respectivas variáveis leva a convergência das esperanças de determinadas funções contínuas e limitadas.

Teorema 3.4 (Teorema de Helly-Bray) *Dadas as variáveis aleatórias $X_1, X_2, \cdots$, com suas respectivas funções de distribuições $F_1, F_2, \cdots$ e X com sua função de distribuição F. Se F_n converge em distribuição para F então*

$$\int g(x)dF_n(x) \xrightarrow{n\to\infty} \int g(x)dF(x) \tag{34}$$

onde g é uma função contínua, limitada e definida no conjunto dos reais.

Demonstração : Vamos aplicar aqui o conceito de limite de uma sequência no cálculo diferencial apresentado na Definição 3.4. Sendo assim, para todo n natural com $n \geq n_0$ e $\epsilon > 0$, devemos mostrar que

$$\left| \int g(x)dF_n(x) - \int g(x)dF(x) \right| < \epsilon. \tag{35}$$

Sejam a e b pontos de continuidade de F, onde x está definido no intervalo $[a, b]$, tal que $-\infty < a < b < \infty$.

$$
\begin{aligned}
\left|\int g(x)dF_n(x) - \int g(x)dF(x)\right| &= \left|\int g(x)dF_n(x) - \int_a^b g(x)dF_n(x) + \int_a^b g(x)dF_n(x)\right. \\
&- \left.\int_a^b g(x)dF(x) + \int_a^b g(x)dF(x) - \int g(x)dF(x)\right| \\
&\leq \left|\int g(x)dF_n(x) - \int_a^b g(x)dF_n(x)\right| + \left|\int_a^b g(x)dF_n(x)\right. \\
&- \left.\int_a^b g(x)dF(x)\right| + \left|\int_a^b g(x)dF(x) - \int g(x)dF(x)\right| \quad (36)
\end{aligned}
$$

Por hipótese, seja $c = sup|g(x)| < \infty$, com $x \in \mathbb{R}$ e considere $\epsilon > 0$.

Precisaremos transformar todos os três módulos do segundo membro da desigualdade (36). Pegando inicialmente o último módulo, temos:

$$
\begin{aligned}
\left|\int_a^b g(x)dF(x) - \int g(x)dF(x)\right| &= \left|\int_{-\infty}^a g(x)dF(x) + \int_b^\infty g(x)dF_n(x)\right| \\
&\leq \left|\int_{-\infty}^a g(x)dF(x)\right| + \left|\int_b^\infty g(x)dF(x)\right| \\
&\leq \int_{-\infty}^a |g(x)|dF(x) + \int_b^\infty |g(x)|dF(x) \\
&< \int_{-\infty}^a cdF(x) + \int_b^\infty cdF(x) \\
&< c\left(\int_{-\infty}^a dF(x) + \int_b^\infty dF(x)\right) \\
&= c\left(F(a) - F(-\infty) + F(+\infty) - F(b)\right),
\end{aligned}
$$

pelo item (*iii*) da Definição 2.3,

$$c\left(F(a) - F(-\infty) + F(+\infty) - F(b)\right) = c\left(F(a) + 1 - F(b)\right).$$

Como pontos de continuidade são densos, se pegarmos, no intervalo $[a, b]$, um valor de a suficientemente pequeno e um valor de b suficientemente grande, ou melhor, se fizermos $a \to -\infty$ e $b \to +\infty$ teremos que

$$F(a) + 1 - F(b) \to 0.$$

Sendo assim,

$$\left|\int_a^b g(x)dF(x) - \int g(x)dF(x)\right| \leq c(F(a) + 1 - F(b)) < \epsilon.$$

Continuando com a desigualdade (36),

$$\begin{aligned}
\left|\int g(x)dF_n(x) - \int_a^b g(x)dF_n(x)\right| &= \left|\int_{-\infty}^a g(x)dF_n(x) + \int_b^\infty g(x)dF_n(x)\right| \\
&\leq \left|\int_{-\infty}^a g(x)dF_n(x)\right| + \left|\int_b^\infty g(x)dF_n(x)\right| \\
&\leq \int_{-\infty}^a |g(x)|dF_n(x) + \int_b^\infty |g(x)|dF_n(x) \\
&< \int_{-\infty}^a cdF_n(x) + \int_b^\infty cdF_n(x) \\
&< c\left(\int_{-\infty}^a dF_n(x) + \int_b^\infty dF_n(x)\right) \\
&= c(F_n(a) + 1 - F_n(b)).
\end{aligned}$$

Como, por hipótese, $F_n(X) \xrightarrow{D} F(x)$ então,

$$c(F_n(a) + 1 - F_n(b)) \xrightarrow{D} c(F(a) + 1 - F(b)) < \epsilon.$$

Para finalizar a demonstração, precisamos que mostrar que, para valores suficientemente grandes de n,

$$\left|\int_a^b g(x)dF_n(x) - \int_a^b g(x)dF(x)\right| < \epsilon. \tag{37}$$

Ao tomarmos pontos no intervalo $[a,\ b]$ denotados por $x_0, x_1, x_2, \cdots, x_N$ onde $a = x_0 < x_1 < x_2 < \cdots < x_N = b$, podemos considerar $x_0, x_1, x_2, \cdots, x_N$ como pontos de continuidade de F. Sendo *g(x)* uma função contínua com $x,\ x_j \in [a, b]$, temos que

$$|g(x) - g(x_j)| < \epsilon, \tag{38}$$

para todo $x \in [x_j, x_{j+1}]$ onde $j \in \{0, 1, 2, \cdots, N-1\}$.

Aplicando integral definida no intervalo $[x_j, x_{j+1}]$ em cada termo da desigualdade (38),

$$\begin{aligned}
|g(x) - g(x_j)| < \epsilon \iff & \left|\int_{x_j}^{x_{j+1}} g(x)dF_n(x) - \int_{x_j}^{x_{j+1}} g(x_j)dF_n(x)\right| < \int_{x_j}^{x_{j+1}} \epsilon dF_n(x) \\
\iff & \left|\int_{x_j}^{x_{j+1}} g(x)dF_n(x) - g(x_j)\int_{x_j}^{x_{j+1}} dF_n(x)\right| < \epsilon\int_{x_j}^{x_{j+1}} dF_n(x) \\
\iff & -\epsilon\int_{x_j}^{x_{j+1}} dF_n(x) < \int_{x_j}^{x_{j+1}} g(x)dF_n(x) - g(x_j)\int_{x_j}^{x_{j+1}} dF_n(x) \\
< & \ \epsilon\int_{x_j}^{x_{j+1}} dF_n(x).
\end{aligned} \tag{39}$$

$$\begin{aligned}
&\iff -\epsilon\int_{x_j}^{x_{j+1}} dF_n(x) + g(x_j)\int_{x_j}^{x_{j+1}} dF_n(x) < \int_{x_j}^{x_{j+1}} g(x)dF_n(x) \\
&< \epsilon\int_{x_j}^{x_{j+1}} dF_n(x) + g(x_j)\int_{x_j}^{x_{j+1}} dF_n(x) \\
&\iff -\epsilon[F_n(x_{j+1}) - F_n(x_j)] + g(x_j)[F_n(x_{j+1}) - F_n(x_j)] < \int_{x_j}^{x_{j+1}} g(x)dF_n(x) \\
&< \epsilon[F_n(x_{j+1} - F_n(x_j)] + g(x_j)[F_n(x_{j+1} - F_n(x_j)] \\
&\iff [g(x_j) - \epsilon][F_n(x_{j+1}) - F_n(x_j)] < \int_{x_j}^{x_{j+1}} g(x)dF_n(x) \\
&< [g(x_j) + \epsilon][F_n(x_{j+1}) - F_n(x_j)]. \qquad (40)
\end{aligned}$$

Fazendo

$$m_{nj} = [g(x_j) - \epsilon][F_n(x_{j+1}) - F_n(x_j)]$$

e

$$M_{nj} = [g(x_j) + \epsilon][F_n(x_{j+1}) - F_n(x_j)]$$

teremos,

$$m_{nj} < \int_{x_j}^{x_{j+1}} g(x)dF_n(x) < M_{nj}. \qquad (41)$$

Por hipótese, $F_N \xrightarrow{D} F$. Logo, a última equivalência em (39) pode ser reescrita como

$$[g(x_j) - \epsilon][F(x_{j+1}) - F(x_j)] < \int_{x_j}^{x_{j+1}} g(x)dF(x) < [g(x_j) + \epsilon][F(x_{j+1}) - F(x_j)] \qquad (42)$$

Fazendo

$$m_j = [g(x_j) - \epsilon][F(x_{j+1}) - F(x_j)]$$

e

$$M_j = [g(x_j) + \epsilon][F(x_{j+1}) - F(x_j)]$$

teremos,

$$m_j < \int_{x_j}^{x_{j+1}} g(x)dF(x) < M_j. \qquad (43)$$

A diferença entre as desigualdades (41) e (43) é igual a

$$m_{nj} - M_j \leq \int_{x}^{x_{j+1}} g(x)dF_n(x) - \int_{x_j}^{x_{j+1}} g(x)dF(x) \leq M_{nj} - m_j. \qquad (44)$$

Para j variando de 0 a $N-1$,

$$\sum_{i=0}^{N-1}(m_{nj}-M_j) \leq \int_a^b g(x)dF_n(x) - \int_a^b g(x)dF(x) \leq \sum_{i=0}^{N-1}(M_{nj}-m_j). \tag{45}$$

Como $m_{nj} \xrightarrow{n\to\infty} m_j$ e $M_{nj} \xrightarrow{n\to\infty} M_j$ temos que

$$\sum_{j=0}^{N-1}(m_{nj}-M_j) \longrightarrow \sum_{j=0}^{N-1}(m_j-M_j) = -2\epsilon(F(b)-F(a)) \geq -2\epsilon \tag{46}$$

e

$$\sum_{j=0}^{N-1}(M_{nj}-m_j) \longrightarrow \sum_{j=0}^{N-1}(M_j-m_j) = 2\epsilon(F(b)-F(a)) \leq 2\epsilon. \tag{47}$$

De (46) e (47), e para n suficientemente grande,

$$\left|\sum_{j=0}^{N-1}(m_{nj}-M_j) - \sum_{j=0}^{N-1}(m_j-M_j)\right| < \epsilon \quad e \quad \left|\sum_{j=0}^{N-1}(M_{nj}-m_j) - \sum_{i=0}^{N-1}(M_j-m_j)\right| < \epsilon$$

assim,

$$\sum_{j=0}^{N-1}(m_{nj}-M_j) \geq -3\epsilon \quad e \quad \sum_{j=0}^{N-1}(M_{nj}-m_j) \leq 3\epsilon.$$

Com isso, a desigualdade (45) será reescrita como

$$-3\epsilon \leq \int_a^b g(x)dF_n(x) - \int_a^b g(x)dF(x) \leq 3\epsilon \tag{48}$$

e, consequentemente,

$$\left|\int_a^b g(x)dF_n(x) - \int_a^b g(x)dF(x)\right| \leq 3\epsilon.$$

Pela desigualdade (36),

$$\begin{aligned}\left|\int g(x)dF_n(x) - \int g(x)dF(x)\right| \;&\leq\; \left|\int g(x)dF_n(x) - \int_a^b g(x)dF_n(x)\right| + \left|\int_a^b g(x)dF_n(x)\right.\\ &-\; \left.\int_a^b g(x)dF(x)\right| + \left|\int_a^b g(x)dF(x) - \int g(x)dF(x)\right|\\ &\leq\; \epsilon+\epsilon+3\epsilon = 5\epsilon.\end{aligned}$$

Portanto,

$$\left|\int g(x)dF_n(x) - \int g(x)dF(x)\right| \leq \epsilon.$$

■

O resultado imediato desta aplicação e consequentemente do Teorema de Helly-Bray é que uma sequência de n funções características de uma sequência de n variáveis aleatórias converge para a sua função característica, ou seja,

$$\varphi_{X_N}(t) \xrightarrow{n\to\infty} \varphi_X(t). \tag{49}$$

Chegamos ao teorema mais importante dentro do conceito de convergência em distribuição que é o teorema da continuidade pois o teorema nos diz que o limite de uma sequência de funções características também é uma função característica com a condição de ser contínua no ponto zero.

Teorema 3.5 (Teorema da Continuidade de Paul-Levy) *Sejam $F_1, F_2, \cdots$, funções de distribuição e sejam $\varphi_1, \varphi_2, \cdots$, suas respectivas funções características. Se φ_n converge pontualmente para um limite φ e se φ é contínua no ponto zero, então,*

(a) existe uma função de distribuição F tal que $F_n \xrightarrow{D} F$ e,

(b) φ é a função característica de F.

Demonstração : Para provar que F_n converge fracamente para alguma função de distribuição, primeiro vamos provar que para toda sequência de funções de distribuição atendendo as condições do Teorema 3.5, existem uma subsequência $F_{n_1}, F_{n_2}, \cdots$ e uma função de distribuição F tais que $F_{nj} \xrightarrow{D} F$, quando $j \to \infty$. Para isso vamos dividir a demonstração em duas etapas:

(i) Existem uma subsequência $F_{n_1}, F_{n_2}, \cdots$ e uma função $F : \mathbb{R} \to [0, 1]$ tais que F é não-decrescente e contínua à direita e

$$F_{nj} \xrightarrow{D} F, \quad \text{quando} \quad j \longrightarrow \infty,$$

para todo x como ponto de continuidade de F e

(ii) a função F, apresentada no item (i), é uma função de distribuição, onde $F(\infty) = 1$ e $F(-\infty) = 0$.

Para provarmos o item (i), descrito logo acima, vamos usar o método da diagonalização. Sejam $r_1, r_2, \cdots$ uma enumeração dos racionais da reta. Seja M a seguinte matriz:

$$M = \begin{bmatrix} F_1 & F_2 & F_3 & F_4 & \cdots \\ F_1^1 & F_2^1 & F_3^1 & F_4^1 & \cdots \\ F_1^2 & F_2^2 & F_3^2 & F_4^2 & \cdots \\ F_1^3 & F_2^3 & F_3^3 & F_4^3 & \cdots \\ F_1^4 & F_2^4 & F_3^4 & F_4^4 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Na matriz M temos que a sequência $\left(F_1^j, F_2^j, \cdots\right)$ contida na $(j+1)$-ésima linha é uma subsequência da sequência contida na j-ésima linha que converge no racional r_j, para $j \leq 1$. Podemos notar que, como $\left(F_1^{j-1}(r_j), F_2^{j-1}(r_j), F_3^{j-1}(r_j), \cdots\right)$ é uma sequência limitada de números reais, ela possui uma subsequência convergente logo, podemos escolher a sequência $\left(F_1^j, F_2^j, F_3^j, \cdots\right)$ conforme descrito acima. Considerando $F_{n_j} = F_j^j$, para $j \geq 1$, temos que a subsequência $(F_{n_j})_j$ converge em todos os racionais da reta. Chamemos o limite de $F(r_k)$, de modo que $F_{n_j}(r_k) \longrightarrow F(r_k)$ para todo k. Como F é uma função de distribuição então $0 \leq F(r_k) \leq 1$ e é não decrescente no conjunto dos racionais. Vamos definir F em função de um x irracional como

$$F(x) = \lim_{r \downarrow x;\, r \in \mathbb{Q}} F(r)$$

dessa forma, F é definida como uma função não-decrescente, mas não é necessariamente contínua à direita.

Vamos provar que $F_{n_j}(x) \longrightarrow F(x)$ para todo x como ponto de continuidade de F, isto é, mostrar apenas que há a convergência, não garantindo que esta convergência seja em distribuição.

Supondo x um ponto de continuidade de F e r' e r'' racionais tais que $r' < x < r''$ e $F(r'') - \epsilon < F(x) < F(r'') + \epsilon$ então,

$$\begin{aligned} F(x) - \epsilon < F(r') &= \lim_{j\to\infty} F_{n_j}(r') \leq \lim_{j\to\infty} inf F_{n_j}(x) \\ &\leq \lim_{j\to\infty} sup F_{n_j}(x) \leq \lim_{j\to\infty} F_{n_j}(r'') \\ &= F(r'') < F(x) + \epsilon. \end{aligned}$$

Para ϵ arbitrário, $F_{n_j}(x) \xrightarrow{j\to\infty} F(x)$. Mostrando assim a convergência de F_n para F. Vamos, agora, mostrar que F é uma função de distribuição, ou seja, provar o item (ii).

Iniciaremos observando que

$$\int_0^t \varphi_{n_j}(s)ds = \int_0^t \int_{-\infty}^{\infty} e^{isx} dF_{n_j}(x)ds. \tag{50}$$

Sendo a função e^{isx} uma função limitada, a ordem das integrais pode ser alterada sem qualquer prejuízo, isto é,

$$\begin{aligned} \int_0^t \varphi_{n_j}(s)ds &= \int_0^t \int_{-\infty}^{\infty} e^{isx} dF_{n_j}(x)ds \\ &= \int_{-\infty}^{\infty} \left[\int_0^t e^{isx} ds\right] dF_{n_j}(x) \\ &= \int_{-\infty}^{\infty} \left[\left.\frac{e^{isx}}{ix}\right|_0^t\right] dF_{n_j}(x) \\ &= \int_{-\infty}^{\infty} \frac{e^{itx}-1}{ix} dF_{n_j}(x). \end{aligned}$$

Seja $g'(x)$ uma função tal que

$$g'(x) = \begin{cases} \dfrac{e^{itx}-1}{ix}, & \text{quando} \quad x \neq 0 \\ t, & \text{quando} \quad x = 0. \end{cases} \tag{51}$$

Observe que $g'(x)$ é uma função contínua em $x = 0$ e limitada. Assim, pelo Teorema de Helly-Bray,

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{e^{itx}-1}{ix} dF_{n_j}(x) &= \int_{-\infty}^{\infty} g'(x) dF_{n_j}(x) \longrightarrow \int_{-\infty}^{\infty} g'(x) dF(x) = \\ &= \int_{-\infty}^{\infty} \frac{e^{itx}-1}{ix} dF(x) = \int_0^t \int_{-\infty}^{\infty} e^{isx} dF_{n_j}(x) ds. \end{aligned}$$

O processo acima nos mostra que há uma convergência entre φ_{n_j} e φ. Como φ é contínua no ponto zero, podemos dizer que φ é limitada e mensurável e, pelo Teorema da Convergência Dominada, temos que

$$\int_0^t \varphi_{n_j}(s) ds \longrightarrow \int_0^t \varphi(s) ds. \tag{52}$$

Através da segunda integral temos,

$$\int_0^t \varphi(s) ds = \int_0^t \int_{-\infty}^{\infty} e^{isx} dF(x) ds \Longleftrightarrow \frac{1}{t} \int_0^t \varphi(s) ds = \frac{1}{t} \int_0^t \left[\int_{-\infty}^{\infty} e^{isx} dF(x) \right] ds, \quad t = 0$$

Como, pela Propriedade 3.3 , $\varphi(0) = 1$, com $s = 0$ e fazendo t tender a zero,

$$\frac{1}{t} \int_0^t \varphi(s) ds \longrightarrow \varphi(0) \quad \text{e} \quad \frac{1}{t} \int_0^t \int_{-\infty}^{\infty} e^{isx} dF(x) ds \longrightarrow \int_{-\infty}^{\infty} 1 dF(x). \tag{53}$$

Com isso podemos afirmar que

$$\varphi(0) = \int_{-\infty}^{\infty} 1 dF(x) = F(\infty) - F(-\infty) = 1.$$

Como $F(\infty) - F(-\infty) = 1$ mostramos que $F(\infty) = 1$ e $F(-\infty) = 0$

Agora, mostraremos que a convergência do item *(i)*, do início desta demonstração, é uma convergência em distribuição.

Para isso, precisamos de sequências de funções de distribuição atendendo as condições do teorema onde, em toda sequência, haverá uma subsequência $F_{n_1}, F_{n_2} \cdots$ e uma função de distribuição F tal que, para todo j tendendo ao infinito teremos que F_{nj} convergirá fracamente para F.

Aplicando o método de redução ao absurdo, vamos supor que F_n não convirja em distribuição para F porém, a subsequência F_{nj} convirja em distribuição para F.

Sendo assim, seja x um ponto de continuidade em F e escreveremos uma subsequência $F_{1'}, F_{2'}, \cdots$, atendendo as condições do Teorema 3.5, onde esta subsequência convirja em distribuição para um ponto a em F, diferente de $F(x)$. Por hipótese, haverá uma subsequência $F_{1''}, F_{2''}, \cdots$ da subsequência $F_{1'}, F_{2'}, \cdots$ e outra função de distribuição G onde $F_{n''} \xrightarrow{D} G$. Porém, como a subsequência $F_{n''} \xrightarrow{D} G$ é uma subsequência de $F_{nj} \xrightarrow{D} F$ então as funções F e G possuem a mesma função característica φ e consequentemente, são iguais. Ora, como $F_{n''} \xrightarrow{D} G(x) = F(x)$ e $F_{n''} \xrightarrow{D} a \neq F(x)$, temos uma contradição em relação a subsequência $F_{n''}$, logo, $F_n \xrightarrow{D} F$.

■

Ao concluir tal demonstração, observamos a grande relação de uma função característica e sua função de distribuição. As funções de distribuição "dependem continuamente"de suas funções características. Por esta razão este teorema é conhecido comumente como "Teorema da Continuidade"(HOEL, 1978).

Ao final deste capítulo, carregado de definições e teoremas, construímos um sólido alicerce, fundamentando conceitos muito importantes para o desenvolvimento do próximo capítulo. Agora, chegamos a um dos pontos principais que é apresentar algumas versões de um conceito considerado um dos mais importante dentro da estatística e probabilidade que é o Teorema Central do Limite.

4

Teorema Central do Limite

Vamos iniciar este capítulo com um exemplo prático, retratando uma situação normal do dia a dia, com o intuito de apresentar uma idéia do que trata o Teorema Central do Limite.

Numa empresa de *Call Center* ou, melhor dizendo, de telemarketing, constatou-se que, em média, a cada hora atendem-se 4 ligações. A proposta é analisar ao longo de um dia de trabalho o comportamento gráfico e tirar algumas conclusões.

A variável X representa o número de ligações a cada hora de trabalho e possui distribuição discreta de Poisson de parâmetro $\lambda = 4$.

Analisaremos o seu gráfico de quatro maneiras diferentes, mantendo a média de ligações por hora.

(1) durante uma hora de trabalho.

(2) durante um turno de quatro horas de trabalho.

(3) durante uma jornada de oito horas de trabalho.

(4) durante uma jornada de dez horas de trabalho, adicionando duas horas extras.

Temos assim outra variável representada por Y, que está em função das variáveis $X_1, X_2, \ldots, X_k$ onde k possui variações específicas em cada um dos itens acima e

$$Y = \sum_{k=1}^{n} X_k,$$

onde Y também possui Distribuição de Poisson, com média $\mu = n\lambda$ e variância $\sigma^2 = n\lambda$.

(1) Para uma hora de trabalho, temos $\lambda = 4$ e $n = 1$ logo, $\mu = 4$.

(2) Para um turno de trabalho, temos $\lambda = 4$ e $n = 4$ logo, $\mu = 16$.

(3) Para uma jornada de oito horas de trabalho, temos $\lambda = 4$ e $n = 8$ logo, $\mu = 32$.

(4) Para uma jornada de dez horas de trabalho, temos $\lambda = 4$ e $n = 10$ logo, $\mu = 40$.

Vejamos os gráficos dos respectivos itens.

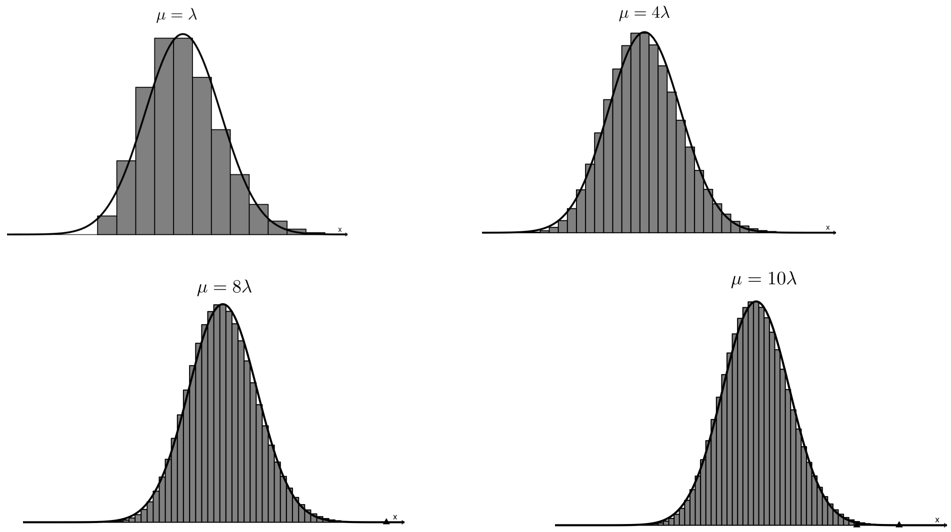


Figura 4 – Comparação entre o histograma de Poisson e a curva normal

Podemos observar, na Figura 4, que, quando o ampliamos o tempo de trabalho nos itens de (1) a (4) descritos logo acima consequentemente, o número de telefonemas também aumenta e o histograma de Poisson vai se aproximando cada vez mais da curva normal. Se contabilizássemos o período de uma semana ou de um mês, o histograma seria muito mais aproximado da curva normal. Isso deve-se ao Teorema Central do Limite.

O Teorema Central do Limite nos diz que para uma quantidade n de variáveis independentes, enquanto o valor de n cresce infinitamente, a distribuição da variável tenderá a uma distribuição normal de parâmetros $\mu = 0$ e $\sigma^2 = 1$.

Considere uma sequência de variáveis aleatórias independentes, $X_1, X_2, \cdots$ definidas no mesmo espaço de probabilidade (Ω, A, P), e seja $S_1, S_2, \cdots$ a sequência de somas parciais, definidas por $S_n = X_1 + X_2 + \cdots + X_n$. A diferença $\dfrac{S_n}{n} - \mu$ multiplicada pela raiz quadrada de n converge fracamente para uma distribuição normal de média nula, ou seja,

$$\sqrt{n}\left(\frac{S_n}{n} - \mu\right) \xrightarrow{D} N(0, \sigma^2).$$

Este capítulo está dividido em duas seções. Na primeira seção, vamos enunciar e provar o teorema limite de DeMoivre-Laplace para variáveis com distribuição binomial como um caso particular do teorema limite para variáveis com qualquer distribuição porém sempre com a condição de que as variáveis sejam iid (independentes e identicamente distribuídas). Na segunda seção, abordaremos sobre a Condição de Lindeberg e, consequentemente, o teorema limite, cujas variáveis são independentes mas não são identicamente distribuídas, finalizando toda abordagem dedutiva, necessária para fundamentar os próximos capítulos onde aplicaremos todo este aparato em situações práticas e concretas.

4.1 O Teorema Central do Limite para variáveis independentes e identicamente distribuídas.

Na situação descrita no início deste capítulo, as variáveis são iid, onde todas possuem a distribuição de Poisson. Logo, se encaixa diretamente na abordagem desta seção.

O teorema abaixo é a forma geral do teorema limite para variáveis iid, que seguiu soberana até o aparecimento da Condição de Lindeberg e sua irrefutável veracidade.

Teorema 4.1 *$X_1, X_2, \cdots$, são variáveis aleatórias independentes e identicamente distribuídas com média e variância de X_i igual a μ e σ^2, respectivamente. Se $S_n = X_1 + X_2 + \cdots + X_n$, com média e variância igual a $n\mu$ e $n\sigma^2$ respectivamente, então:*

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{D} N(0,1). \tag{54}$$

Demonstração : No Capítulo 2, vimos que se $Z = \sigma X + \mu$ então $Z^* = \frac{X-\mu}{\sigma}$ e que $E(Z^*) = 0$ e $Var(Z^*) = 1$. Com isso, iniciamos supondo que $\mu = 0$.

Pelo Teorema da Continuidade de Paul Levy, esta demostração consiste em mostrar que,

$$\varphi_{\frac{S_n}{\sigma\sqrt{n}}}(t) \longrightarrow e^{\frac{-t^2}{2}}, \quad \forall t \in \mathbb{R}.$$

Pela definição de função característica, na qual $\varphi_X(t) = Ee^{itX}$, podemos escrever

$$\varphi_{\frac{S_n}{\sigma\sqrt{n}}}(t) = Ee^{it\frac{S_n}{\sigma\sqrt{n}}}.$$

$$\varphi_{\frac{S_n}{\sigma\sqrt{n}}}(t) = Ee^{it\frac{S_n}{\sigma\sqrt{n}}} = Ee^{\frac{itS_n}{\sigma\sqrt{n}}} = Ee^{i\frac{t}{\sigma\sqrt{n}}S_n}$$

Fazendo uma mudança de variável, $v = \frac{t}{\sigma\sqrt{n}}$,

$$Ee^{i\frac{t}{\sigma\sqrt{n}}S_n} = Ee^{ivS_n} = \varphi_{S_n}(v) = \varphi_{S_n}\left(\frac{t}{\sigma\sqrt{n}}\right)$$

Como, por hipótese, $S_n = X_1 + X_2 + \cdots + X_n$ e os X_i, $i = 1, 2, \cdots, n$ são identicamente distribuídos então,

$$\begin{aligned}
\varphi_{S_n}\left(\frac{t}{\sigma\sqrt{n}}\right) &= \varphi_{X_1}\left(\frac{t}{\sigma\sqrt{n}}\right)\varphi_{X_2}\left(\frac{t}{\sigma\sqrt{n}}\right)\cdots\varphi_{X_n}\left(\frac{t}{\sigma\sqrt{n}}\right)\\
&= \prod_{i=1}^{n}\varphi_{X_i}\left(\frac{t}{\sigma\sqrt{n}}\right)\\
&= \varphi_{X_i}^{n}\left(\frac{t}{\sigma\sqrt{n}}\right)
\end{aligned}$$

Ou seja, $\varphi_{\frac{S_n}{\sigma\sqrt{n}}}(t) = \varphi_{X_i}^{n}\left(\frac{t}{\sigma\sqrt{n}}\right)$.

Lembremos que $\varphi(t) = \varphi_X(t) = \int e^{itx} dF_X(x)$.

Por hipótese, $E^2(X_1)$ existe então a função φ possui duas derivadas contínuas, ou seja,

$$\varphi^{(n)}(t) = \int (ix)^n e^{itx} dF_X(x), \quad n = 0, 1, 2$$

Fazendo $t = 0$ temos,

$$\begin{aligned}
\varphi^{(n)}(0) &= \int (ix)^n e^0 dF_X(x)\\
&= \int (ix)^n dF_X(x)\\
&= i^n \int x^n dF_X(x)\\
&= i^n E X^n.
\end{aligned}$$

Aplicando a Expansão de Taylor[1] de ordem 1 em volta de 0 na função φ, com Resto de Lagrange com $0 < \alpha(t) \leq t$,

$$\begin{aligned}
\varphi(t) &= \sum_{n=0}^{1}\frac{\varphi^{(n)}(0)}{n!}(t-0)^n + \frac{\varphi^{(2)}(\alpha(t))}{2}\\
&= \frac{\varphi^{(0)}(0)}{0!}(t-0)^0 + \frac{\varphi^{(1)}(0)}{1!}(t-0) + \frac{\varphi^{(2)}(\alpha(t))}{2!}(t-0)^2\\
&= \varphi(0) + \varphi'(0)t + \frac{\varphi''(\alpha(t))}{2}t^2
\end{aligned}$$

Precisamos obter os valores de $\varphi^n(0)$ com $n = 0, 1, 2$. Para isso, vamos fazer o seguinte:

[1] Para maiores esclarecimentos, consultar (GUIDORIZZI, 2006).

$$\begin{aligned}
\varphi(t) &= \varphi(0) + \varphi'(0)t + \frac{\varphi''(\alpha(t))}{2}t^2 \\
&= \varphi(0) + \varphi'(0)t + \frac{\varphi''(\alpha(t))}{2}t^2 + \frac{\varphi''(0)}{2}t^2 - \frac{\varphi''(0)}{2}t^2 \\
&= \varphi(0) + \varphi'(0)t + \frac{\varphi''(0)}{2}t^2 + \frac{t^2}{2}[\varphi''(\alpha(t)) - \varphi''(0)]
\end{aligned}$$

Como $\varphi^{(n)}(0) = i^n EX^n$, podemos calcular $\varphi(0)$, $\varphi'(0)$ e $\varphi''(0)$. Com isso,

$$\begin{aligned}
\varphi(0) &= i^0 EX^0 = 1, \\
\varphi'(0) &= iEX = 0 \quad \text{e} \\
\varphi''(0) &= i^2 EX^2 = -1\sigma^2 = -\sigma^2.
\end{aligned}$$

Sendo assim, podemos escrever $\varphi(t)$ como

$$\varphi(t) = 1 - \frac{-t^2\sigma^2}{2} + \frac{t^2}{2}e(t),$$

onde e(t) é uma função originária de $\varphi''(\alpha(t)) - \varphi''(0)$.

Vamos agora utilizar $\varphi\left(\frac{t}{\sigma\sqrt{n}}\right)$ ao invés de $\varphi(t)$.

$$\begin{aligned}
\varphi\left(\frac{t}{\sigma\sqrt{n}}\right) &= 1 + \frac{-\left(\frac{t}{\sigma\sqrt{n}}\right)^2\sigma^2}{2} + \frac{\left(\frac{t}{\sigma\sqrt{n}}\right)^2}{2}e\left(\frac{t}{\sigma\sqrt{n}}\right) \\
&= 1 + \frac{-\frac{t^2}{\sigma^2\sqrt{n}^2}\sigma^2}{2} + \frac{\frac{t^2}{\sigma^2\sqrt{n}^2}}{2}e\left(\frac{t}{\sigma\sqrt{n}}\right) \\
&= 1 + \frac{-\frac{t^2}{n}}{2} + \frac{\frac{t^2}{n\sigma^2}}{2}e\left(\frac{t}{\sigma\sqrt{n}}\right) \\
&= 1 - \frac{t^2}{2n} + \frac{t^2}{2n\sigma^2}e\left(\frac{t}{\sigma\sqrt{n}}\right) \\
&= 1 - \frac{t^2}{2n}\left[1 + \frac{1}{\sigma^2}e\left(\frac{t}{\sigma\sqrt{n}}\right)\right]
\end{aligned}$$

Elevando ambos os membros a n,

$$\left[\varphi\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n = \left[1 - \frac{t^2}{2n}\left[1 + \frac{1}{\sigma^2}e\left(\frac{t}{\sigma\sqrt{n}}\right)\right]\right]^n$$

Precisamos analisar o comportamento da expressão $1 + \dfrac{1}{\sigma^2}e\left(\frac{t}{\sigma\sqrt{n}}\right)$ quando n tende ao infinito, ou seja, calcular o seu limite. Assim:

$$\lim_{n\to\infty}\left[1+\frac{1}{\sigma^2}e\left(\frac{t}{\sigma\sqrt{n}}\right)\right]=1+\frac{1}{\sigma^2}\lim_{n\to\infty}e\left(\frac{t}{\sigma\sqrt{n}}\right)=1+\frac{1}{\sigma^2}0=1.$$

Isso significa que quanto mais o valor de n cresce, a expressão $1+\frac{1}{\sigma^2}e\left(\frac{t}{\sigma\sqrt{n}}\right)$ tende a 1 e, consequentemente, a expressão $\left[1-\frac{t^2}{2n}\left[1+\frac{1}{\sigma^2}e\left(\frac{t}{\sigma\sqrt{n}}\right)\right]\right]^n$ tende a $1-\frac{t^2}{2n}$. Isto é,

$$\left[1-\frac{t^2}{2n}\left[1+\frac{1}{\sigma^2}e\left(\frac{t}{\sigma\sqrt{n}}\right)\right]\right]^n \xrightarrow{D} 1-\frac{t^2}{2n}.$$

No cálculo diferencial, o limite de Euler, no conjunto dos números complexos, nos mostra que se uma sequência c_n converge para c então

$$\lim_{n\to\infty}\left[1+\frac{c_n}{n}\right]^n=e^c.$$

Assim, para n tendendo ao infinito,

$$\lim_{n\to\infty}\left[1-\frac{t^2}{2n}\right]^n=\lim_{n\to\infty}\left[1+\frac{-\frac{t^2}{2}}{n}\right]^n=e^{-\frac{t^2}{2}}.$$

Isto significa que

$$\varphi^n\left(\frac{t}{\sigma\sqrt{n}}\right)=\varphi_{\frac{S_n}{\sigma\sqrt{n}}}(t)\to e^{\frac{-t^2}{2}},\quad \forall t\in\mathbb{R},$$

e que a convergência dada em (54) é válida. Portanto,

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

■

Com essa demonstração, o Teorema 4.1 está disponível para ser aplicado em situações que envolvam variáveis independentes e identicamente distribuídas para diferentes tipos de distribuições.

O próximo teorema é um caso particular do Teorema Central do Limite para variáveis independentes e identicamente distribuídas já que as variáveis possuem distribuição de Bernoulli. Este teorema tem uma grande importância para o desenvolvimento da probabilidade pois constitui o marco inicial para o grandioso estudo feito posteriormente por outros.

Teorema 4.2 (Teorema Central do Limite de DeMoivre e Laplace) *$X_1, X_2, \cdots$, são variáveis aleatórias independentes e identicamente distribuídas que possui distribuição binomial de parâmetros n e p, com $0 < p < 1$. Se $S_n = X_1 + X_2 + \cdots + X_n$, então:*

$$\frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow{D} N(0,1) \tag{55}$$

Demonstração: Vamos seguir o mesmo procedimento utilizado na demonstração do Teorema 4.1, atentando ao fato que X_i possui média igual a p e variança igual a *p(1 - p)*, ou seja, $E(X_i) = p$ e $Var(X_i) = p(1-p)$ já que X_i possui distribuição binomial. Sendo assim, a média de S_n é igual a *np*, e a variância de S_n é igual a *np(1 - p)*.

Vamos considerar Y como sendo a variável padronizada de X. Sendo assim, podemos escrever a variável Y como

$$Y = \frac{X - np}{\sqrt{np(1-p)}}$$

A função característica de X é $\varphi_Y(t) = Ee^{tY}$. Vamos mostrar que a variável X possui função de distribuição normal padrão.

$$\begin{aligned} Ee^{tY} = Ee^{\frac{X-np}{\sqrt{np(1-p)}}} &= Ee^{\frac{tx}{\sqrt{np(1-p)}} - \frac{tnp}{\sqrt{np(1-p)}}} \\ &= e^{-\frac{tnp}{\sqrt{np(1-p)}}} Ee^{\frac{tx}{\sqrt{np(1-p)}}} \end{aligned}$$

Lembremos que, sendo X uma variável discreta então sua média é dada por

$$EX = \sum_{j=0}^{n} x_j f(x_j),$$

onde $f(x_j)$ é a função de distribuição de X. Com isso,

$$\begin{aligned} e^{-\frac{tnp}{\sqrt{np(1-p)}}} Ee^{\frac{tx}{\sqrt{np(1-p)}}} &= e^{-\frac{tnp}{\sqrt{np(1-p)}}} \sum_{x=o}^{n} e^{\frac{tx}{\sqrt{np(1-p)}}} \binom{n}{x} p^x (1-p)^{n-x} \\ &= e^{-\frac{tnp}{\sqrt{np(1-p)}}} \sum_{x=o}^{n} \binom{n}{x} \left(pe^{\frac{t}{\sqrt{np(1-p)}}} \right)^x (1-p)^{n-x} \end{aligned}$$

$$= e^{-\frac{tnp}{\sqrt{np(1-p)}}}\left(1-p+pe^{\frac{t}{\sqrt{np(1-p)}}}\right)^n \tag{56}$$

$$= \left[e^{-\frac{tp}{\sqrt{np(1-p)}}}\left(1-p+pe^{\frac{t}{\sqrt{np(1-p)}}}\right)\right]^n \tag{57}$$

$$= \left[(1-p)e^{-\frac{tp}{\sqrt{np(1-p)}}}+pe^{\frac{t(1-p)}{\sqrt{np(1-p)}}}\right]^n \tag{58}$$

Escrevendo e^x como uma soma infinita teremos

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

Nesse sentido podemos afirmar que

$$e^{-\frac{tp}{\sqrt{np(1-p)}}} = \left(1-\frac{tp}{\sqrt{np(1-p)}}+\frac{t^2p^2}{2np(1-p)}+\cdots\right) \tag{59}$$

e

$$e^{\frac{t(1-p)}{\sqrt{np(1-p)}}} = \left(1+\frac{t(1-p)}{\sqrt{np(1-p)}}+\frac{t^2(1-p)^2}{2np(1-p)}+\cdots\right) \tag{60}$$

Utilizando as igualdades (59) e (60) temos que,

$$\begin{aligned}
(1-p)e^{-\frac{tp}{\sqrt{np(1-p)}}}+pe^{\frac{t(1-p)}{\sqrt{np(1-p)}}} &= (1-p)\left(1-\frac{tp}{\sqrt{np(1-p)}}+\frac{t^2p^2}{2np(1-p)}+\cdots\right)\\
&+ p\left(1+\frac{t(1-p)}{\sqrt{np(1-p)}}+\frac{t^2(1-p)^2}{2np(1-p)}+\cdots\right)\\
&= (1-p)+p-\frac{tp(1-p)}{\sqrt{np(1-p)}}+\frac{tp(1-p)}{\sqrt{np(1-p)}}\\
&+ \frac{t^2p^2(1-p)}{2np(1-p)}+\frac{t^2p(1-p)^2}{2np(1-p)}+\cdots\\
&= 1+\frac{p(1-p)t^2}{2np(1-p)}+\cdots\\
&= 1+\frac{t^2}{2n}+\cdots
\end{aligned}$$

Elevando esta última igualdade a n-ésima potência, poderemos escrever Ee^{tY} como

$$\begin{aligned} Ee^{tY} &= (1 + \frac{t^2}{2n} + \cdots)^n \\ &= (1 + \frac{\frac{t^2}{2}}{n} + \cdots)^n \xrightarrow{n\to\infty} e^{\frac{t^2}{2}}. \end{aligned}$$

Ora, as variáveis discretas X e Y possuem a mesma distribuição, ambas são binomiais, e também a função característica da variável Y se aproxima da função de distribuição normal padrão. Através do Teorema da Unicidade, citada no Capítulo 3, podemos concluir que a distribuição da variável X também se aproxima de uma distribuição normal padrão finalizando assim a demonstração.

$\blacksquare$

4.2 O Teorema Central do Limite de Lindeberg

Um matemático finlandês, por volta de 1920, desenvolve um método capaz de generalizar a aproximação normal a partir de qualquer distribuição de probabilidade e não só com variáveis aleatórias identicamente distribuídas. Seu nome é Jarl Waldemar Lindeberg (1876 - 1932) e este método é o teorema limite sob uma condição muito forte conhecida por Condição de Lindeberg, descrita logo abaixo na igualdade (61).

$$\forall \epsilon > 0, \lim_{n\to\infty} \left[\frac{1}{s_n^2} \sum_{k=1}^{n} \int_{|x-\mu_k|>\epsilon s_n} (x-\mu_k)^2 dF_k(x) \right] = 0. \tag{61}$$

Sendo s_n^2, a soma das variâncias de X_k, com $k = 1, 2, \cdots$, as parcelas de s_n^2 são do tipo σ_k^2, ou seja,

$$s_n^2 = \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2.$$

A Condição de Lindeberg mostra que, para valores de n suficientemente grandes, a influência de uma parcela σ_k^2 é muito pequena não afetando a característica da soma s_n^2. Como as variâncias não possuem distribuições idênticas então utilizamos a parcela de maior variância para mostrar este fato. Em outras palavras,

$$\max_{1\le k\le n} \frac{\sigma_k^2}{s_n^2} \xrightarrow{n\to\infty} 0. \tag{62}$$

Antes de apresentarmos a prova do Teorema Central do Limite de Lindeberg, precisamos enunciar dois lemas e realizar a verificação da Condição de Lindeberg para a convergência (63) descrita logo abaixo, onde

$$\frac{S_n - ES_n}{s_n} \xrightarrow{D} N(0,1). \tag{63}$$

O primeiro lema trata da ordem de séries do tipo $\sum_k k^\lambda$.

Lema 4.1 *Para $\lambda > 0$,*

$$\frac{1}{n^{\lambda+1}}\sum_{k=1}^{n} k^{\lambda} \xrightarrow{n\to\infty} \frac{1}{\lambda+1},$$

de maneira que $\sum_{k=1}^{n} k^{\lambda}$ é da ordem de $n^{\lambda+1}$.

Demonstração:

Sejam x^{λ} e k^{λ} duas potências atendendo a seguinte condição,

- ❏ $x^{\lambda} \leq k^{\lambda}$, para $k-1 \leq x \leq k$ e

$$\int_{k-1}^{k} x^{\lambda}dx \leq \int_{k-1}^{k} k^{\lambda}dx = k^{\lambda}x|_{k}^{k-1} = k^{\lambda}.$$

- ❏ $x^{\lambda} \geq k^{\lambda}$, para $k \leq x \leq k+1$ e

$$\int_{k}^{k+1} x^{\lambda}dx \geq \int_{k}^{k+1} -k^{\lambda}dx = k^{\lambda}x|_{k+1}^{k} = k^{\lambda}.$$

Com isso, podemos escrever a seguinte desigualdade,

$$\int_{k-1}^{k} x^{\lambda}dx \leq k^{\lambda} \leq \int_{k}^{k+1} x^{\lambda}dx. \tag{64}$$

Fazendo o valor de k variar de 1 até n, a desigualdade (64) pode ser reescrita da seguinte forma

$$\int_{0}^{n} x^{\lambda}dx \leq \sum_{k=1}^{n} k^{\lambda} \leq \int_{1}^{n+1} x^{\lambda}dx. \tag{65}$$

A integral da esquerda em (65) é igual a $\dfrac{n^{\lambda+1}}{\lambda+1}$. Isto é,

$$\int_{0}^{n} x^{\lambda}dx = \left.\frac{x^{\lambda+1}}{\lambda+1}\right|_{0}^{n} = \frac{n^{\lambda+1}}{\lambda+1}.$$

A integral da direita em (65) é igual a $\dfrac{(n+1)^{\lambda+1}-1}{\lambda+1}$. Ou seja,

$$\int_{1}^{n+1} x^{\lambda}dx = \left.\frac{x^{\lambda+1}}{\lambda+1}\right|_{1}^{n+1} = \frac{(n+1)^{\lambda+1}-1}{\lambda+1}.$$

Temos ainda que

$$\frac{(n+1)^{\lambda+1}-1}{\lambda+1} \leq \frac{(n+1)^{\lambda+1}}{\lambda+1}.$$

Assim, a desigualdade (65) será reescrita como

$$\frac{n^{\lambda+1}}{\lambda+1} \leq \sum_{k=1}^{n} k^{\lambda} \leq \frac{(n+1)^{\lambda+1}}{\lambda+1}.$$

Dividindo-se todos os membros da desigualdade por $n^{\lambda+1}$, o resultado será

$$\frac{1}{\lambda+1} \leq \frac{1}{n^{\lambda+1}} \sum_{k=1}^{n} k^{\lambda} \leq \frac{1}{\lambda+1}\left(\frac{n+1}{n}\right)^{\lambda+1}.$$

Para concluirmos, precisamos mostrar que

$$\lim_{n\to\infty} \frac{1}{\lambda+1}\left(\frac{n+1}{n}\right)^{\lambda+1} = \frac{1}{\lambda+1}.$$

Sendo assim,

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{\lambda+1}\left(\frac{n+1}{n}\right)^{\lambda+1} &= \frac{1}{\lambda+1}\lim_{n\to\infty}\left(\frac{n+1}{n}\right)^{\lambda+1} \\
&= \frac{1}{\lambda+1}\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^{\lambda+1} \\
&= \frac{1}{\lambda+1}.
\end{aligned}$$

Após a prova do Lema 4.1, podemos verificar a Condição de Lindeberg para a convergência normal dada em (63).

Sejam $X_1, X_2, \cdots$, variáveis independentes onde X_n possui distribuição uniforme de parâmetros -n e n.

Admitindo-se que $\mu_k = 0$, a variância de X_k é dada por

$$\sigma_k^2 = \int (x-\mu_k)^2 dF_k(x).$$

Como a variável X é uniformemente distribuída então $dF_k(x) = f_k(x)dx = \frac{1}{2k}dx$ e, para $\mu_k = 0$, podemos escrever

$$\begin{aligned}
\int (x-\mu_k)^2 dF_k(x) &= \int_{-k}^{k} (x-0)^2 \frac{1}{2k} dx \\
&= \frac{1}{2k}\int_{-k}^{k} x^2 dx \\
&= \frac{1}{2k}\left.\frac{x^3}{3}\right|_{-k}^{k} \\
&= \frac{k^2}{3}
\end{aligned}$$

e

$$s_n^2 = \sum_{k=1}^{n} \sigma_k^2 = \sum_{k=1}^{n} \frac{k^2}{3}. \tag{66}$$

Para $1 \leq k \leq n$ e $\epsilon > 0$,

$$\int_{|x|>\epsilon s_n} x^2 dF_k(x) = \int x^2 I_{|x|>\epsilon s_n}(x) dF_k(x) = \frac{1}{2k} \int_{-k}^{k} x^2 I_{|x|>\epsilon s_n}(x) dx. \tag{67}$$

Pela definição de função de densidade uniforme, *f(x)* assume valor zero para valores fora do intervalo fechado [-*k*, *k*] logo, a integral da direita na igualdade (67) tende a zero quando n tende ao infinito. Pela igualdade (66),

$$\frac{s_n^2}{n^3} = \frac{1}{n^{2+1}} \sum_{k=1}^{n} \frac{k^2}{3} = \frac{1}{3}\left(\frac{1}{n^{2+1}} \sum_{k=1}^{n} k^2\right).$$

O termo entre parênteses atende a condição do Lema 4.1. Ou seja,

$$\frac{1}{n^{2+1}} \sum_{k=1}^{n} k^2 \longrightarrow \frac{1}{3},$$

e consequentemente

$$\frac{s_n^2}{n^3} \longrightarrow \frac{1}{9}.$$

Para concluirmos, vamos mostrar que

$$\forall \epsilon > 0,\ \frac{1}{s_n^2} \sum_{k=1}^{n} \int_{-k}^{k} x^2 I_{|x|>\epsilon s_n}(x) dx \longrightarrow 0.$$

Note que $\dfrac{s_n^2}{n^2} = \dfrac{s_n^2}{n^3} n$. Com isso,

$$\begin{aligned}
\lim_{n\to\infty} \frac{s_n^2}{n^2} &= \lim_{n\to\infty} \frac{s_n^2}{n^3} n \\
&= \lim_{n\to\infty} \frac{s_n^2}{n^3} \lim_{n\to\infty} n \\
&= \infty.
\end{aligned}$$

Ora, se $\dfrac{s_n^2}{n^2}$ tende ao infinito então $\dfrac{1}{\frac{s_n^2}{n^2}}$ tende a zero. Consequentemente,

$$\frac{1}{s_n^2} \sum_{k=1}^{n} \int_{-k}^{k} x^2 I_{|x|>\epsilon s_n}(x) dx \to 0.$$

■

O próximo lema trata-se de uma generalização do Limite de Euler para números complexos utilizado na demonstração do Teorema 4.1, onde diz que se $c_n \to c$ então $\left[1+\frac{c_n}{n}\right]^n \to e^c$.

Lema 4.2 *Seja $c_{n,k}$ uma sequência de números complexos onde $\sum_{k=1}^{n} c_{n,k} \xrightarrow{n\to\infty} c$. Se*

$$\max_{1\leq k\leq n} |c_{n,k}| \xrightarrow{n\to\infty} 0$$

e, sendo M uma constante que não está em função de n,

$$\sum_{k=1}^{n} |c_{n,k}| \leq M < \infty$$

então,

$$\prod_{k=1}^{n}(1+c_{n,k}) \xrightarrow{n\to\infty} e^c.$$

Bem, agora conseguimos reunir todos os requisitos necessários para podermos provar o Teorema Central do Limite de Lindeberg. O processo de demonstração segue alguns caminhos semelhantes à demonstração do Teorema 4.1 além da utilização do Lema 4.1 e 4.2.

Teorema 4.3 (Teorema Central do Limite de Lindeberg) *$X_1, X_2, \cdots$, são variáveis aleatórias independentes que não possuem distribuições idênticas. Por isso, a média de X_n é igual a μ_n e a variância de X_n é igual a σ_n^2. X_n possui variância finita e para algum n, $\sigma_n^2 > 0$.*

Sendo $S_n = X_1 + X_2 + \cdots + X_n$ com variância s_n^2 tal que $s_n = \sqrt{VarS_n}$, a convergência normal

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

é válida se a Condição de Lindeberg for satisfeita:

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

Demonstração : Como dito anteriormente, iremos proceder com o mesmo objetivo apresentado no início da demonstração do Teorema 4.1 , ou seja, vamos concluir a demonstração quando mostrarmos que :

$$\varphi_{\frac{S_n - ES_n}{s_n}}(t) \xrightarrow{n\to\infty} e^{\frac{-t^2}{2}}. \tag{68}$$

Por hipótese, as variáveis são independentes e $S_n = X_1 + X_2 + \cdots + X_n$. Como $EX_k = \mu_k$,

$$\begin{aligned}
\varphi_{\frac{S_n - ES_n}{s_n}}(t) &= \varphi_{\sum_{k=1}^{n} \frac{X_k - EX_k}{s_n}}(t) \\
&= \varphi_{\frac{X_1-\mu_1}{s_n}}(t)\varphi_{\frac{X_2-\mu_2}{s_n}}(t)\cdots\varphi_{\frac{X_n-\mu_n}{s_n}}(t) \\
&= \prod_{k=1}^{n} \varphi_{\frac{X_k-\mu_k}{s_n}}(t) \\
&= \prod_{k=1}^{n} Ee^{it\frac{X_k-\mu_k}{s_n}}
\end{aligned}$$

Assim, precisamos mostrar que

$$\prod_{k=1}^{n} Ee^{it\frac{X_k-\mu_k}{s_n}} \xrightarrow{n\to\infty} e^{\frac{-t^2}{2}}. \tag{69}$$

Primeiramente, vamos utilizar a fórmula de expansão em série de Taylor dada por:

$$e^{itx} = \sum_{n=0}^{\infty} \frac{(itx)^n}{n!}.$$

A expansão em Série de Taylor de segunda ordem é

$$\begin{aligned}
e^{itx} \approx \sum_{n=0}^{2} \frac{(itx)^n}{n!} &= \frac{(itx)^0}{0!} + \frac{itx}{1!} + \frac{(it\bar{x})^2}{2!} \\
&= 1 + itx + \theta_1(x)\frac{t^2x^2}{2}, \quad \text{com} \quad |\theta_1(x)| < 1,
\end{aligned}$$

e a de terceira ordem,

$$\begin{aligned}
e^{itx} \approx \sum_{n=0}^{3} \frac{(itx)^n}{n!} &= \frac{(itx)^0}{0!} + \frac{itx}{1!} + \frac{(itx)^2}{2!} + \frac{(it\bar{x})^3}{3!} \\
&= 1 + itx - \frac{t^2x^2}{2} + \theta_2(x)\frac{t^3x^3}{6}, \quad \text{com} \quad |\theta_2(x)| < 1.
\end{aligned}$$

Dado um $\epsilon > 0$ e $|x|$ variando em torno de ϵ, vamos reescrever e^{itx} como uma única expressão,

$$e^{itx} \approx 1 + itx - \frac{t^2x^2}{2} + r_\epsilon(x), \tag{70}$$

onde $r_\epsilon(x)$ é definida por duas sentenças em função de x com parâmetro ϵ, ou seja,

$$r_\epsilon(x) = \begin{cases} (1+\theta_1(x))\frac{t^2x^2}{2} & \text{se} \quad |x| > \epsilon \\ \theta_2(x)\frac{t^3x^3}{6} & \text{se} \quad |x| \le \epsilon \end{cases}.$$

Vamos escrever $Ee^{it\frac{X_k-\mu_k}{s_n}}$ através da Expressão (70),

$$\begin{aligned}
Ee^{it\frac{X_k-\mu_k}{s_n}} = \int e^{it\frac{x-\mu_k}{s_n}}\, dF_k(x) &= \int \left[1 + it\left(\frac{x-\mu_k}{s_n}\right) - \frac{t^2}{2}\left(\frac{x-\mu_k}{s_n}\right) + r_\epsilon\left(\frac{x-\mu_k}{s_n}\right)\right] dF_k(x) \\
&= \int dF_k(x) + \int it\left(\frac{x-\mu_k}{s_n}\right) dF_k(x) - \int \frac{t^2}{2}\left(\frac{x-\mu_k}{s_n}\right) dF_k(x) \\
&+ \int r_\epsilon\left(\frac{x-\mu_k}{s_n}\right) dF_k(x).
\end{aligned}$$

Pela definição de função densidade,

$$\int dF_k(x) = 1$$

e, por linearidade,

$$r_\epsilon\left(\frac{x-\mu_k}{s_n}\right) = \left[1 + \theta_1\left(\frac{x-\mu_k}{s_n}\right)\right]\frac{t^2}{2}\left(\frac{x-\mu_k}{s_n}\right)^2 + \theta_2\left(\frac{x-\mu_k}{s_n}\right)\frac{t^3}{6}\left(\frac{x-\mu_k}{s_n}\right)^3, \tag{71}$$

onde $|x-\mu_k| > \epsilon s_n$ na primeira parcela e $|x-\mu_k| \leq \epsilon s_n$ na segunda parcela ambas no segundo membro da igualdade (71).

Portanto,

$$\begin{aligned}
\int dF_k(x) &+ \int it\left(\frac{x-\mu_k}{s_n}\right) dF_k(x) - \int \frac{t^2}{2}\left(\frac{x-\mu_k}{s_n}\right)^2 dF_k(x) + \int r_\epsilon\left(\frac{x-\mu_k}{s_n}\right) dF_k(x) \\
&= 1 + itE\left(\frac{X_k-\mu_k}{s_n}\right) - \frac{t^2}{2}E\left(\frac{X_k-\mu_k}{s_n}\right)^2 \\
&+ \frac{t^2}{2}\int_{|x-\mu_k|>\epsilon s_n}\left[1 + \theta_1\left(\frac{x-\mu_k}{s_n}\right)\right]\left(\frac{x-\mu_k}{s_n}\right)^2 dF_k(x) \\
&+ \frac{t^3}{6}\int_{|x-\mu_k|\leq\epsilon s_n}\theta_2\left(\frac{x-\mu_k}{s_n}\right)\left(\frac{x-\mu_k}{s_n}\right)^3 dF_k(x).
\end{aligned}$$

Admitindo-se que $EX_k = \mu_k = 0$ e $EX_k^2 = \text{Var}X_k = \sigma^2$, temos que:

$$\begin{aligned}
1 &+ itE\left(\frac{X_k-\mu_k}{s_n}\right) - \frac{t^2}{2}E\left(\frac{X_k-\mu_k}{s_n}\right)^2 \\
&+ \frac{t^2}{2}\int_{|x-\mu_k|>\epsilon s_n}\left[1+\theta_1\left(\frac{x-\mu_k}{s_n}\right)\right]\left(\frac{x-\mu_k}{s_n}\right)^2 dF_k(x) \\
&+ \frac{t^3}{6}\int_{|x-\mu_k|\leq\epsilon s_n}\theta_2\left(\frac{x-\mu_k}{s_n}\right)\left(\frac{x-\mu_k}{s_n}\right)^3 dF_k(x) \\
&= 1 + it\left(\frac{EX_k-\mu_k}{s_n}\right) - \frac{t^2}{2s_n^2}(EX_k^2 - 2EX_k\mu_k + \mu_k^2) \\
&+ \frac{t^2}{2}\int_{|x-\mu_k|>\epsilon s_n}\left[1+\theta_1\left(\frac{x-\mu_k}{s_n}\right)\right]\left(\frac{x-\mu_k}{s_n}\right)^2 dF_k(x) \\
&+ \frac{t^3}{6}\int_{|x-\mu_k|\leq\epsilon s_n}\theta_2\left(\frac{x-\mu_k}{s_n}\right)\left(\frac{x-\mu_k}{s_n}\right)^3 dF_k(x) \\
&= 1 - \frac{t^2\sigma_k^2}{2s_n^2} + \frac{t^2}{2s_n^2}\int_{|x-\mu_k|>\epsilon s_n}\left[1+\theta_1\left(\frac{x-\mu_k}{s_n}\right)\right](x-\mu_k)^2 dF_k(x) \qquad (72) \\
&+ \frac{t^3}{6s_n^2}\int_{|x-\mu_k|\leq\epsilon s_n}\left(\frac{x-\mu_k}{s_n}\right)\theta_2\left(\frac{x-\mu_k}{s_n}\right)(x-\mu_k)^2 dF_k(x) \\
&= 1 - \frac{t^2\sigma_k^2}{2s_n^2} + e_{n,k}.
\end{aligned}$$

Vamos analisar o resto $e_{n,k}$ e reescrevê-la de forma a estar aplicando a Condição de Lindeberg e, com isso, simplificar a igualdade (72).

Observe que $|\theta_1(x)| < 1$, logo $0 < 1 + \theta_1(x) < 2$ e que $|X_k - \mu_k| \leq \epsilon s_n$. Sendo assim, $\left|\frac{X_k-\mu_k}{s_n}\right| \leq \epsilon$ e que,

$$|\theta_2(x)| < 1 \iff -\epsilon < \epsilon\theta_2(x) < \epsilon, \quad \epsilon > 0.$$

Com isso,

$$\begin{aligned}
|e_{n,k}| &= \frac{t^2}{2s_n^2}\int_{|x-\mu_k|>\epsilon s_n}\left|1+\theta_1\left(\frac{x-\mu_k}{s_n}\right)\right|(x-\mu_k)^2 dF_k(x) \\
&+ \frac{|t|^3}{6s_n^2}\int_{|x-\mu_k|\leq\epsilon s_n}\left|\frac{x-\mu_k}{s_n}\right|\left|\theta_2\left(\frac{x-\mu_k}{s_n}\right)\right|(x-\mu_k)^2 dF_k(x) \\
&\leq \frac{t^2}{2s_n^2}\int_{|x-\mu_k|>\epsilon s_n}2(x-\mu_k)^2 dF_k(x) + \frac{|t|^3}{6s_n^2}\int_{|x-\mu_k|\leq\epsilon s_n}\epsilon(x-\mu_k)^2 dF_k(x) \\
&\leq \frac{t^2}{s_n^2}\int_{|x-\mu_k|>\epsilon s_n}(x-\mu_k)^2 dF_k(x) + \frac{\epsilon|t|^3}{6s_n^2}\int_{|x-\mu_k|\leq\epsilon s_n}(x-\mu_k)^2 dF_k(x).
\end{aligned}$$

Continuando,

$$\sum_{k=1}^n |e_{n,k}| \leq t^2\frac{1}{s_n^2}\sum_{k=1}^n\int_{|x-\mu_k|>\epsilon s_n}(x-\mu_k)^2 dF_k(x) + \frac{\epsilon|t|^3}{6}\frac{1}{s_n^2}\sum_{k=1}^n\int_{|x-\mu_k|\leq\epsilon s_n}(x-\mu_k)^2 dF_k(x).$$

A Condição de Lindeberg nos diz que

$$\lim_{n\to\infty}\left[\frac{1}{s_n^2}\sum_{k=1}^{n}\int_{|x-\mu_k|>\epsilon s_n}(x-\mu_k)^2 dF_k(x)\right]=0$$

onde,

$$\frac{1}{s_n^2}\sum_{k=1}^{n}\int_{|x-\mu_k|\le\epsilon s_n}(x-\mu_k)^2 dF_k(x)=1.$$

Sendo assim, quando n for suficientemente grande,

$$\sum_{k=1}^{n}|e_{n,k}|\le\frac{\epsilon|t|^3}{6}.$$

Precisamos escolher uma sequência convergente de epsilons, do tipo ϵ_j, com $j = 1,2,3,\cdots$, tal que esta sequência seja simples de manipular e convirja para zero. Sendo assim, tomemos $\epsilon=\frac{1}{m}$.

Se $\epsilon=\frac{1}{m}$ então, para $m>n$, existe um n_m tal que

$$\sum_{k=1}^{n}|e_{n,k}|\le\frac{|t|^3}{6m},$$

onde os restos $e_{n,k}$ com $n_m\le n\le n_{m+1}$ são baseados no termo geral $\frac{1}{m}$.

Como a sequência $\epsilon=\frac{1}{m}$ converge para zero e $\sum_{k=1}^{n}|e_{n,k}|\le\frac{|t|^3}{6m}$ então, podemos afirmar que:

$$\sum_{k=1}^{n}|e_{n,k}|\xrightarrow{n\to\infty}0. \tag{73}$$

Dando uma olhada na Expressão (69), podemos escrever que

$$\prod_{k=1}^{n}Ee^{it\frac{X_k-\mu_k}{s_n}}=\prod_{k=1}^{n}\left(1-\frac{t^2\sigma_k^2}{2s_n^2}+e_{n,k}\right).$$

Para provarmos a convergência (69), precisamos validar o Lema 4.2. Para isso, vamos fazer $c_{n,k}=-\frac{t^2\sigma_k^2}{2s_n^2}+e_{n,k}$ e $c=-\frac{t^2}{2}$.

Primeiramente, veremos se $\sum_{k=1}^{n}c_{n,k}\le M$ com $M<\infty$. Então,

$$\begin{aligned}\sum_{k=1}^{n}c_{n,k} &\le \sum_{k=1}^{n}\left(-\frac{t^2\sigma_k^2}{2s_n^2}+e_{n,k}\right)\\ &\le \sum_{k=1}^{n}-\frac{t^2\sigma_k^2}{2s_n^2}+\sum_{k=1}^{n}e_{n,k}\\ &\le -\frac{t^2}{2}\sum_{k=1}^{n}\frac{\sigma_k^2}{s_n^2}+\sum_{k=1}^{n}e_{n,k}\longrightarrow-\frac{t^2}{2},\end{aligned}$$

pois $\sum_{k=1}^{n} \frac{\sigma_k^2}{s_n^2} = 1$, já que $s_n^2 = \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2$, e $\sum_{k=1}^{n} e_{n,k} \to 0$.

Com isso, $\sum_{k=1}^{n} c_{n,k}$ é limitado de maneira uniforme, o que quer dizer que existe, pelo menos um $M < \infty$ onde

$$\forall n, \sum_{k=1}^{n} c_{n,k} \leq M.$$

Agora, basta provarmos que o máximo de $c_{n,k}$ tende a zero quando n tende ao infinito. Logo,

$$\begin{aligned} \max_{1\leq k\leq n} |c_{n,k}| &\leq \max_{1\leq k\leq n} \left| -\frac{t^2\sigma_k^2}{2s_n^2} + e_{n,k} \right| \\ &\leq \max_{1\leq k\leq n} \frac{t^2\sigma_k^2}{2s_n^2} + \max_{1\leq k\leq n} |e_{n,k}| \\ &\leq \frac{t^2}{2} \max_{1\leq k\leq n} \frac{\sigma_k^2}{s_n^2} + \max_{1\leq k\leq n} |e_{n,k}|. \end{aligned} \tag{74}$$

A primeira parcela da Expressão (74) tende a zero pelo princípio da Condição de Lindeberg que diz que para n tendendo ao infinito o máximo dentre as parcelas $\frac{\sigma_k^2}{s_n^2}$ tende a zero. A segunda parcela também tende a zero pela convergência (73).

Portanto, pelo Lema 4.2, podemos concluir que

$$\prod_{k=1}^{n} \left(1 - \frac{t^2\sigma_k^2}{2s_n^2} + e_{n,k}\right) \xrightarrow{n\to\infty} e^{\frac{-t^2}{2}},$$

concluindo assim a demonstração do teorema limite de Lindeberg.

■

O Teorema Central do Limite enunciado e demonstrado por Lindeberg é considerada até hoje o Teorema Central do Limite mais genérico produzido. Isso porque a condição de Lindeberg é satisfeita em todos os outros modelos de TCL[2].

O próximo capítulo abordará a utilização do Teorema Central do Limite através de várias situações problemas.

[2] Considerando que as variáveis são independentes entre si

5

Aplicando o Teorema Central do Limite

Vamos abordar aqui situações práticas onde o Teorema Central do Limite faz jus a sua importância em diferentes áreas do conhecimento humano como nas engenharias, na área de saúde, economia, dentre outros tantos que surge a partir da necessidade humana de colher informações para realizar planejamentos estratégicos e na tomada de decisões. Outro cenário em que o TCL atua são nas situações cotidianas ou corriqueiras pelo qual pessoas comuns podem se deparar tal como uma criança jogando um jogo de trilha com os amigos e um deles com o seu próprio dado, através de seus lançamentos, começa a sair apenas números altos como cinco ou seis. Numa única partida as outras crianças podem não perceber. Porém, com o decorrer das partidas, eles com certeza notarão algo de estranho naquele dado que sai mais a face cinco ou a seis comparado aos outros dados, ou seja, um número maior de jogadas em algumas partidas mostrou-se suficiente para apresentar desconfiança em relação ao dado do coleguinha. Vejam o Teorema Central do Limite em ação, na sua forma mais intuitiva, mostrando-se presente e "dizendo": "estou aqui, conheçam-me pois sou muito útil!".

Este capítulo será dividido em duas seções distintas. A primeira seção tratará de situações problemas envolvendo diferentes distribuições de probabilidade atentando-se a justificativa que um dos objetivos do uso do TCL é aproximar diferentes distribuições para a distribuição normal inclusive propondo, em alguns casos a inviabilidade de se utilizar função de densidade de probabilidade característica da distribuição em questão. A segunda seção apresentará outras situações problemas resolvida apenas pela aproximação normal, a fim de enriquecer o capítulo, abrangendo a área de atuação do TCL.

5.1 Primeiras aplicações

Nesta primeira seção, o principal objetivo é mostrar que, através do Teorema Central do Limite, podemos aproximar uma distribuição discreta ou uma distribuição contínua para uma distribuição normal padrão utilizando a variável Z, descrita na

Seção 2.5. Não haverá detalhamento nas distribuições apresentadas pois, todos estes conceitos estão no Capítulo 2. Vamos apresentar aqui quatro situações práticas, as duas primeiras envolvem variáveis aleatórias discretas e as duas últimas, variáveis aleatórias contínuas. O intuito é resolver as questões de duas maneiras distintas uma é através da função de distribuição encontrando as probabilidades reais e a outra é aplicando o Teorema Central do Limite e aproximando normalmente tais distribuições. Em seguida, ao compararmos estes resultados devemos verificar quão próximos eles são.

Vamos ao primeiro exemplo. Aqui temos uma situação envolvendo uma variável aleatória discreta de distribuição binomial.

Exemplo 5.1 *Baseado em dados locais, a empresa WXY constatou que o número de mulheres grávidas correspondiam a, mais ou menos, 28% da população de mulheres daquele local. Baseado nesta informação, se a empresa coletasse uma amostra aleatória com 120 mulheres, qual seria a probabilidade de que, pelo menos, 45 delas estivessem grávidas?*

Solução : Vamos denominar de X a variável aleatória referente ao número de mulheres grávidas dentro da amostra coletada. Podemos notar que as características de X segue uma distribuição binomial de parâmetros $n = 120$ e $p = 0.28$, cuja função de densidade é dada por

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x = 0, 1, 2, \cdots, n.$$

Como visto no Capítulo 2, $f(x) = P(X = x)$ porém, precisamos encontrar a probabilidade de, no mínimo, 45 mulheres estarem grávidas, ou seja, $P(X \geq 45)$.

Para calcularmos $P(X \geq 45)$ temos duas opções:

1. $P(X \geq 45) = P(X = 45) + P(X = 46) + \cdots + P(X = 120)$
2. $P(X \geq 45) = 1 - P(X < 45) = 1 - [P(X = 44) + P(X = 43) + \cdots + P(X = 0)]$

Neste caso a resolução será feita de forma computacional. Com isso, $P(X \geq 45) = 0.0151$.

Com a sua média sendo $\mu_X = 33.6$ e sua variância, $\sigma_X^2 = 24.2$ com desvio padrão $\sigma_X = 4.92$, vamos aplicar o Teorema Central do Limite já que 120 mulheres foram pesquisadas e esse valor é suficientemente grande para tal aplicação.

Assim, vamos calcular $P(X \geq 45)$ através da aproximação normal pela variável Z. Lembrando que a variável aleatória Z, para uma aproximação binomial, é escrita como

$$Z = \frac{X - np}{\sqrt{np(1-p)}}.$$

Com isso, temos que

$$Z = \frac{X - np}{\sqrt{np(1-p)}} = \frac{45 - 33.6}{4.92} = 2.32$$

e

$$P(X \geq 45) = 1 - P(X \leq 45) = 1 - P(Z \leq 2.32).$$

Pela tabela z da Figura 3,

$$1 - P(Z \leq 2.32) = 1 - 0.9898 = 0.0102.$$

Portanto, a probabilidade é que, pelo menos, 1.02% das mulheres estejam grávidas. A Figura 5 ilustra graficamente como a probabilidade refere-se a área cinza sob a curva normal.

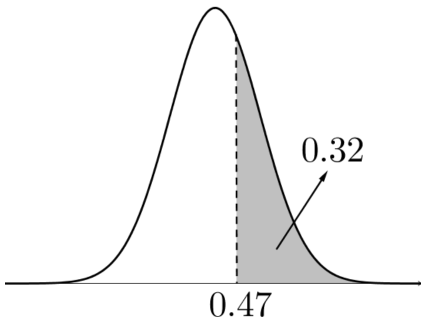


Figura 5 – Probabilidade de 1.02% de mulheres grávidas

O próximo exemplo retrata uma aplicação da distribuição de Poisson. Adaptação de (MARTINS, 2010).

Exemplo 5.2 *Um consultório médico tem um atendimento médio diário de 10 pacientes. Num dia normal de funcionamento, determine a probabilidade de atender, no máximo, 6 pacientes.*

Solução : Seja a variável X a quantidade de atendimentos em um dia de funcionamento na clínica. A probabilidade de atender, no máximo, 6 pacientes é representado por $P(X \leq 6)$.

Primeiramente, vamos calcular a probabilidade através de sua função de probabilidade encontrada no Capítulo 2, $f(x) = \dfrac{e^{-\lambda}\lambda^x}{x!}$, de parâmetro $\lambda = 10$. Atender, no máximo, 6 pessoas quer dizer que $x \leq 6$.

Encontrar a probabilidade de, no máximo, 6 atendimentos é o mesmo que encontrar a soma das probabilidades para x menores ou iguais a 6, $x = 0, 1, 2, 3, 4, 5, 6$, ou seja,

$$P(X \leq 6) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) + P(X = 6)$$

ou,

$$P(X \leq 6) = \sum_{k=0}^{6} P(X = k). \tag{75}$$

Tais probabilidades serão calculadas individualmente, como veremos abaixo. Para isso, temos que $e^{-10} = 0,00004539$.

Para $x = 0$,

$$P(X = 0) = \frac{e^{-10}10^0}{0!} = 0,000045.$$

Para $x = 1$,

$$P(X = 1) = \frac{e^{-10}10^1}{1!} = 0,00045.$$

Para $x = 2$,

$$P(X = 2) = \frac{e^{-10}10^2}{2!} = 0,002269.$$

Para $x = 3$,

$$P(X = 3) = \frac{e^{-10}10^3}{3!} = 0,007565.$$

Para $x = 4$,

$$P(X = 4) = \frac{e^{-10}10^4}{4!} = 0,018912.$$

Para $x = 5$,

$$P(X = 5) = \frac{e^{-10}10^5}{5!} = 0,037825.$$

Para $x = 6$,

$$P(X = 6) = \frac{e^{-10}10^6}{6!} = 0,063041.$$

Agora vamos aplicar estes valores na igualdade (75).

$$\begin{aligned} P(X \leq 6) &= 0,000045 + 0,000453 + 0,002269 \\ &+ 0,007565 + 0,018912 + 0,037825 \\ &+ 0,063041 = 0,13011. \end{aligned}$$

Aproximadamente, há 13% de probabilidade de haver, no máximo, 6 atendimentos num dia normal de funcionamento.

Em seguida, vamos usar a aproximação da distribuição de Poisson por uma normal através do Teorema Central do Limite. Lembremos que, para podermos aproximar normalmente a distribuição de Poisson, devemos ter $\lambda \geq 5$. A média e a variância da distribuição de Poisson neste caso são

$$\mu = \lambda = 10$$

e

$$\sigma = \sqrt{\lambda} = \sqrt{10}.$$

O valor de x será aproximado para um valor decimal pois estamos no campo dos reais através da transformação de uma variável aleatória discreta para uma contínua. Com isso, vamos fazer $x = 6.2$.

Seja Z a variável aleatória normalizada de X. Assim,

$$P(X \leq 6.2) = P\left(Z \leq \frac{x-\mu}{\sigma}\right) = P\left(Z \leq \frac{6.2-10}{\sqrt{10}}\right) = P(Z \leq -1.20).$$

Através da propriedade $P(Z \leq -z) = 0.5 - P(Z \leq z)$ aplicada na tabela z,

$$P(Z \leq -1.20) = 0.1151.$$

A Figura 6 ilustra a área encontrada que é a probabilidade calculada pela aproximação normal.

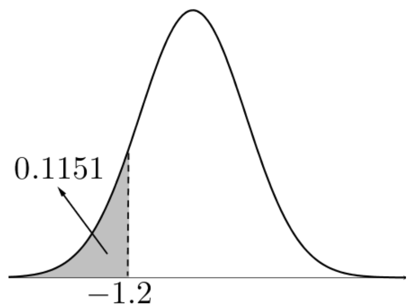


Figura 6 – Probabilidade de ocorrer até 6 atendimentos.

Assim, há uma probabilidade de 11.51% de ocorrer, no máximo, 6 atendimentos num dia normal de funcionamento.

Neste próximo exemplo, apresentamos uma situação envolvendo uma variável aleatória com distribuição uniforme contínua. Exemplo adaptado de (FARIAS, 2009).

Exemplo 5.3 *O senhor Antônio é um investidor e, dentre os diferentes ramos de atuação, vai investir em lotes de terra para revenda ou construção. Ele está participando de um leilão e está interessado em dar um lance de um lote de terra. Há um outro licitante com os mesmos interesses que o senhor Antônio. O leiloeiro comunicou que, pelas regras estabelecidas para este leilão, o lance mais alto acima de R\$ 100 000.00 será aceito. Vamos supor que o lance dado pelo outro licitante interessado nesse lote de terra seja uma variável aleatória uniformemente distribuída entre R\$ 100 000.00 e R\$ 150 000.00. Sabendo disso,*

(a) se o senhor Antônio der um lance de R\$ 120 000.00, qual é a probabilidade dele ficar com o lote?

(b) *se o senhor Antônio der um lance de R\$ 140 000.00, qual é a probabilidade dele ficar com o lote?*

Solução :

(a) Seja X uma variável aleatória que representa o lance dado pelo senhor Antônio. Podemos observar que esta variável possui distribuição uniforme contínua de parâmetros $a = R\$ 100\,000.00$ e $b = R\$ 150\,000.00$. No Capítulo 2 encontram-se a função de distribuição, a média e a variância de X. Como a função de distribuição é dada por:

$$F(x) = \frac{x-a}{b-a},$$

vamos encontrar $P(X = x)$, com $x = 120\,000$.

Como X é uma variável aleatória contínua, $P(X = x) = 0$. Por definição de distribuição uniforme contínua, $a \leq x \leq b$ logo, vamos encontrar a probabilidade de um lance estar entre R\$ 100 000.00 e R\$ 120 000.00, ou seja, $P(X \leq 120\,000)$.

$$P(X \leq 120\,000) = \frac{120\,000 - 100\,000}{150\,000 - 100\,000} = \frac{20\,000}{50\,000} = 0.4.$$

Assim, com um lance de R\$ 120 000.00 você tem uma chance de 40% em ficar com o lote.

Agora, a resolução dar-se-á através da aproximação normal pelo Teorema Central do Limite. Vamos encontrar o valor da média e do desvio padrão de X.

$$\mu = \frac{a+b}{2} = \frac{100\,000 + 150\,000}{2} = 125\,000$$

e

$$\sigma = \sqrt{\frac{(b-a)^2}{12}} = \sqrt{\frac{(150\,000 - 100\,000)^2}{12}} = \frac{50\,000}{\sqrt{12}} = 14\,433.7567.$$

Assim,

$$P(X \leq 120\,000) = P\left(Z \leq \frac{120\,000 - 125\,000}{14\,433.7567}\right) = P(Z \leq -0.35) = 0.3632.$$

A Figura 7 ilustra a transformação realizada na variável X e a probabilidade em forma de área sob a curva normal.

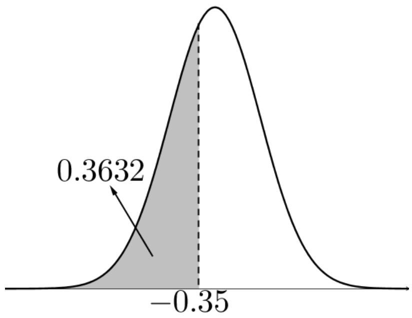


Figura 7 – Probabilidade no lance entre R$ 100 000.00 e R$ 120 000.00

Através da aproximação normal, a chance em ficar com o lote é de 36.32%.

(b) Seguindo os mesmos procedimentos apresentados no item (a), vamos aplicar a função de distribuição de X para $x \leq 140\,000$.

$$P(X \leq 140\,000) = \frac{140\,000 - 100\,000}{150\,000 - 100\,000} = \frac{40\,000}{50\,000} = 0.8.$$

Assim, com um lance de R$ 140 000.00 você tem uma chance de 80% em ficar com o lote.

Agora, a resolução dar-se-á através da aproximação normal pelo Teorema Central do Limite. Conhecendo o valor da média e do desvio padrão de X, calculado no item (a), vamos calcular $P(X \leq 140\,000)$. Assim,

$$P(X \leq 140\,000) = P\left(Z \leq \frac{140\,000 - 125\,000}{14\,433.7567}\right) = P(Z \leq 1.04) = 0.8508.$$

Através da aproximação normal, a chance em ficar com o lote é de 85.08%.

A seguir, temos o último exemplo desta seção. Aqui, a situação problema aborda uma distribuição exponencial e uma característica específica desta distribuição é a propriedade de falta de memória[1].

Exemplo 5.4 *Segundo uma lei conhecida como a lei dos quinze minutos, todas as agências bancárias tem que atender o cliente em até 15 minutos sob a condição de sofrer sanções judiciais como advertências ou até multas. Uma agência bancária de grande movimentação, tanto de pessoas quanto de dinheiro tem o tempo médio de atendimento por cliente de 8 minutos e esse tempo é uma variável aleatória exponencial.*

(a) Determine a função de distribuição de X.

(b) Encontre a probabilidade de um cliente, ao receber sua senha, demorar mais que 15 minutos para ser atendido.

[1] Para consultas, vide (ROSS, 2010).

(c) Encontre a probabilidade de um cliente, ao receber sua senha, demorar mais que 8 minutos para ser atendido.

(d) Um cliente teve sua senha chamada e já está sendo atendido há 7 minutos. Encontre a probabilidade de um cliente demorar mais que 15 minutos para ser atendido.

Solução :

(a) A função de distribuição de uma variável aleatória exponencial é dada por

$$F(x) = \begin{cases} 0 & \text{se} \quad x < 0 \\ 1 - e^{-\lambda x} & \text{se} \quad x \geq 0 \end{cases}$$

O parâmetro de λ será extraído da média fornecida pela questão. Ou seja,

$$\mu = \frac{1}{\lambda} = 8 \Leftrightarrow \lambda = \frac{1}{8}$$

Logo, a função de distribuição que atende a situação apresentada na questão é

$$F(x) = \begin{cases} 0 & \text{se} \quad x < 0 \\ 1 - e^{-x/8} & \text{se} \quad x \geq 0 \end{cases}$$

(b) Primeiramente, resolver pela função de distribuição encontrada no item (a). Como $x = 15$, vamos calcular $P(X > 15)$. Assim,

$$P(X > 15) = 1 - P(X \leq 15) = 1 - (1 - e^{-15/8}) = e^{-15/8} = 0.1533.$$

Logo, a probabilidade de um cliente demorar mais que 15 minutos nesta agência é de 15.33%.

Agora, através do Teorema Central do Limite, vamos utilizar a variável Z e realizar uma aproximação normal. Conhecendo a média de atendimento $\mu = 8$ e o seu desvio padrão

$$\sigma = \sqrt{\frac{1}{\lambda^2}} = \sqrt{\frac{1}{\frac{1}{8^2}}} = 8,$$

vamos calcular $P(X > 15)$. Sendo assim,

$$P(X > 15) = 1 - P(X \leq 15) = 1 - P(Z \leq \frac{15 - 8}{8})1 - P(Z \leq 0.875) = 1 - 0.8092 = 0.1908.$$

Assim, a probabilidade de um cliente demorar mais que 15 minutos nesta agência, através da aproximação normal, é de 19.08%. A Figura 8 ilustra claramente a área sob a curva normal como a probabilidade encontrada com a aproximação normal.

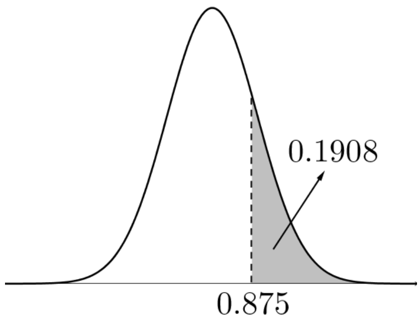


Figura 8 – Probabilidade de 19.08% de demorar mais que 15 minutos

(c) Seguindo o mesmo procedimento apresentado no item (b). Primeiramente resolver pela sua função de distribuição. Como $x = 8$, vamos calcular $P(X > 8)$. Logo,

$$P(X > 8) = 1 - P(X \leq 8) = 1 - (1 - e^{-8/8}) = e^{-1} = 0.3679.$$

Assim, a probabilidade de um cliente demorar mais que 8 minutos nesta agência é de 36.79%.

Agora, através do Teorema Central do Limite, vamos utilizar a variável Z e realizar uma aproximação normal. Conhecendo a média de atendimento $\mu = 8$ e o seu desvio padrão $\sigma = 8$, percebemos que o desvio padrão, numa distribuição exponencial, possui o mesmo valor que a sua média, vamos calcular $P(X > 8)$. Com isso,

$$\begin{aligned} P(X > 8) &= P(X \geq 8.5) = 1 - P(X < 8.5) \\ &= 1 - P\left(Z < \frac{8.5 - 8}{8}\right) \\ &= 1 - P(Z < 0.06) = 1 - 0.5239 \\ &= 0.4761. \end{aligned}$$

Assim, a probabilidade de um cliente demorar mais que 15 minutos nesta agência, através da aproximação normal, é de 47.61%.

(d) Neste caso vamos calcular a probabilidade condicional do atendimento ser superior a quinze minutos mesmo já tendo passado-se sete minutos de atendimento, ou seja, vamos calcular $P(X > 15|X > 7)$. Para isso precisamos calcular a probabilidade do tempo do atendimento ser maior do que sete minutos, ou seja, $P(X > 7)$. Assim,

$$P(X > 7) = e^{-7/8} = 0.4169.$$

Com este dado e o resultado do item (b) temos,

$$P(X > 15|X > 7) = \frac{P(X > 15)}{P(X > 7)} = \frac{0.1533}{0.4169} = 0.3677.$$

Comparando este resultado com o do item (c) podemos observar que as probabilidades são iguais. Em outras palavras, o tempo de duração do atendimento de oito minutos passado ou não algum tempo de atendimento caracteriza probabilidades iguais. Este fato refere-se a uma propriedade conhecida como a falta de memória da distribuição exponencial. Esta propriedade, resumidamente falando, afirma que o tempo restante para concluir o evento independe do tempo passado desde o inicio do evento. Mais especificamente, o tempo restante para concluir o atendimento deste cliente na agência não está ligado ao tempo decorrido desde o início do atendimento, ou seja, a probabilidade do cliente ser atendido num tempo superior a quinze minutos não depende dos sete minutos de atendimento já decorridos.

Eis alguns comentários a respeito dos exemplos desta seção e suas formas de resolução. Podemos observar que os resultados obtidos através da função de distribuição e da aproximação normal são próximos e essa proximidade tende a aumentar quando o valor do parâmetro populacional cresce, resumindo, quanto maior o valor do parâmetro mais próximo os resultados serão. No Exemplo 5.2, ao aumentarmos o valor do parâmetro λ por exemplo, para $\lambda = 20$, vamos obter, aproximadamente, 0.026% através da função de densidade e 0.09% pela aproximação normal, ou seja, os valores tendem a ficar mais próximos, graças ao Teorema Central do Limite pois proporciona um alto grau de confiabilidade nos resultados.

Outro ponto a ser considerado é praticidade da aproximação normal sobre a função de probabilidade, isto é, com poucas operações é possível se chegar ao resultado esperado. Para exemplos que envolvam funções exponenciais, como os Exemplos 5.2 e 5.4, para uma resolução realizada manualmente, mesmo com o uso de uma calculadora científica por exemplo, quanto maior o valor de λ, mais inviável vai ficando o cálculo a ponto de ser resolvido apenas de forma computacional com software específicos.

Estes comentários expressam a importância do Teorema Central do Limite nas aplicações acima, nas aplicações da próxima seção e do próximo capítulo. Na verdade, em todas as situações onde podemos aplicá-lo.

5.2 Outras aplicações

As situações problemas apresentadas aqui não serão resolvidas pela função de distribuição de suas respectivas variáveis e pela aproximação normal como ocorreu na Seção 5.1. Aqui, a forma de resolução será através da aproximação normal, ampliando e reforçando o emprego do Teorema Central do Limite. Esta seção contém três exemplos e cada um deles faz referências a diferentes área de atuação humana.

Neste primeiro exemplo, temos uma questão onde não se conhece a função de distribuição da variável porém a variância é conhecida e como $n \geq 30$, a utilização do

Teorema Central do Limite é indispensável.

Exemplo 5.5 *Uma empresa vende caixas com biscoitos e, quando lhe é solicitado, envia-as pelo correio. Para as evitar pesar, cobra sempre o valor de portes de correio correspondente a admitir que qualquer caixa pesa 1445g. Cada caixa leva 80 biscoitos e o peso da embalagem plástica é desprezável. Se soubermos que o peso de cada biscoito é variável mas que em média pesa 18g com um desvio padrão de 5g, determine a probabilidade do valor pago em portes de correio com o envio de uma caixa ser inferior ao valor que pagaria, caso a caixa fosse pesada.*

Solução : Vamos considerar duas variáveis aleatórias X e Y. X é uma variável com distribuição desconhecida que representa o peso, em gramas, de cada biscoito com $\mu_X = 18$ e $\sigma_X = 5$. Y é uma variável dependente de X, ou seja,

$$Y = \sum_{k=1}^{80} X_k, \quad \text{com} \quad n = 80,$$

onde Y possui a mesma distribuição de X e $\mu_Y = n\mu_X$ e $\sigma_Y = \sigma_X \sqrt{n}$.

O valor pago em portes de correio com o envio de uma caixa ser inferior ao valor que pagaria, caso a caixa fosse pesada significa que $Y \geq 1445$. A probabilidade de acontecer é $P(Y \geq 1445)$.

Para $n = 80$, o Teorema Central do Limite aproxima a distribuição de Y para uma distribuição normal reduzida. Sendo $\mu_Y = 1440$ e $\sigma_Y = 20\sqrt{5}$ temos,

$$\begin{aligned} P(Y \geq 1445) &= P\left(z \geq \frac{1445 - \mu_Y}{\sigma_Y}\right) \\ &= P\left(z \geq \frac{1445 - 1440}{20\sqrt{5}}\right) \\ &= P\left(z \geq \frac{1}{4\sqrt{5}}\right) \\ &= P(z \geq 0.11). \end{aligned}$$

Observando a tabela z de distribuição normal reduzida, temos que

$$P(z \geq 0.11) = 0.4562.$$

Portanto, a probabilidade de pagar menos em portes do correio que na pesagem da caixa é de 45.62%. Graficamente, através da Figura 9, podemos observar a área cinza sob a curva normal referente a probabilidade calculada pela aproximação normal.

O exemplo seguinte trabalha diretamente com duas distribuições, uma é a distribuição normal e a outra é desconhecida. Como a quantidade analisada é o questionamento do problema, assumimos essa quantidade igual ou superior a 30 para

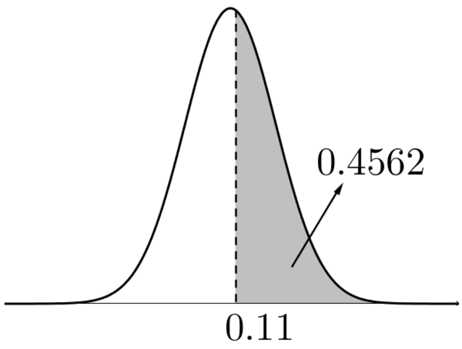


Figura 9 – Probabilidade em se pagar menos em portes do correio que através do peso da caixa.

aplicarmos o Teorema Central do Limite. Este exemplo é uma adaptação de (ROSS, 2010).

Exemplo 5.6 *Numa firma de construção, alguns engenheiros analisam um vão de uma ponte e acreditam que o peso suportado, sem afetar a estrutura, é uma variável normalmente distribuída de média 400 e variância 1600. Analisando que um carro possua peso com média 3 e variância 0.09 eis a questão: que quantidade destes carros é necessária para que a probabilidade de haver algum dano na estrutura desse vão da ponte seja maior que 0.1?*

Solução : Vamos considerar duas variáveis aleatórias X e Y. X é uma variável com distribuição normal que representa o peso, em toneladas, de um vão da ponte. Possui média 400 ($\mu_X = 400$) e desvio padrão 40 ($\sigma_X = 40$). A variável Y, com distribuição desconhecida, representa o peso, em toneladas, de um carro que passa pela ponte. Possui média 3 e desvio padrão 0.3, ou seja, $\mu_Y = 3$ e $\sigma_Y = 0.3$.

Uma terceira variável V que representa o peso, em toneladas, de uma quantidade n de carros sobre o vão da ponte em questão é tal que

$$V = \sum_{k=1}^{n} Y_k,$$

com a mesma distribuição de n com média igual a $n\mu_Y$ e desvio padrão, $n\sigma_Y$.

O número de carros tal que a probabilidade de haver algum dano na estrutura desse vão da ponte seja maior que 0.1 é dada por

$$P(X \leq V) > 0.1.$$

Pelas médias das variáveis n e Y percebemos que o número de carros será grande o suficiente para utilizarmos a aproximação normal através do Teorema Central do

Limite. Ou seja, através da aproximação a uma distribuição normal reduzida temos que

$$P(X \leq V) = P(z \leq \frac{\sum_{k=1}^{n} Y_k - 400}{40}) > 0.1.$$

Na tabela z, de distribuição normal, o valor da normal reduzida, cuja probabilidade é de 0.1, é igual a -1.28. Com isso,

$$\begin{aligned} P(z \leq \frac{\sum_{k=1}^{n} Y_k - 400}{40}) > 0.1. \quad &\Longleftrightarrow \quad \frac{\sum_{k=1}^{n} Y_k - 400}{40} > -1.28 \\ &\Longleftrightarrow \quad \sum_{k=1}^{n} Y_k = -1.28 \text{ x } 40 + 400 \\ &\Longleftrightarrow \quad \sum_{k=1}^{n} Y_k = 348.8. \end{aligned}$$

Encontramos o somatório do peso dos n carros. Para encontrarmos o total de carros vamos utilizar a média da variável V, isto é,

$$\mu_V = 348.8 \Longleftrightarrow 3n = 348.8 \Longleftrightarrow n = 116.3 \Longleftrightarrow n \approx 117.$$

Assim podemos dizer que o número mínimo é de 117 carros.

Este último exemplo traz uma situação de variáveis com distribuição desconhecida. Aqui, o Teorema Central do Limite se faz presente pela quantidade ser maior ou igual a 30. Traduzido de (DEKKING, 2005).

Exemplo 5.7 *Uma indústria fabrica elos de metais para a confecção de correntes. O laboratório de pesquisa da indústria modela o comprimento, em cm, de um elo pela variável aleatória X, com o valor esperado E[X] = 5 e variância Var[X] = 0.04. O comprimento de um elo é definido de tal maneira que o comprimento de uma corrente é igual à soma dos comprimentos de seus elos. A indústria vende correntes com 50 metros de comprimento. Para maior garantia, 1002 elos são usados para essas correntes. A indústria garante que a corrente não é menor do que 50 metros. Se, por um acaso, uma corrente for muito curta, o cliente é reembolsado, e uma nova corrente lhe é dado gratuitamente.*

(a) Determinar uma estimativa de probabilidade de que, para uma corrente de pelo menos 50 metros, mais de 1002 elos são necessários.

(b) O departamento de vendas da indústria percebe que ele tem que entregar um monte de correntes grátis e pede ao laboratório de pesquisa que está errado. Após mais investigações, os relatórios do laboratório de pesquisa para o departamento de vendas informam que o valor esperado 5 (cm) é incorreto, e que o valor correto é 4.99 (cm). Você acha que era necessário comunicar uma mudança tão pequena deste valor?

Solução : Para resolver esta situação primeiramente devemos entender o papel das variáveis aleatórias participantes. Para isso, temos X como a primeira variável representando o comprimento, em centímetros, de cada elo da corrente. O enunciado da questão nos diz que o comprimento de um elo é definido de tal maneira que o comprimento de uma corrente é igual à soma dos comprimentos de seus elos. Assim, temos outra variável, em função de X, representada por Y designando a soma dos comprimentos de seus elos na formação de uma corrente, ou seja,

$$Y = \sum_{k=1}^{n} X_k,$$

onde n é o números de elos necessários numa corrente de 50 metros.

(a) Devemos encontrar a probabilidade de que, numa corrente de 50 metros, mais de 1002 elos sejam necessários. Reescrevendo esta afirmação, podemos dizer que devemos encontrar a probabilidade de que, ao usarmos 1002 elos, o comprimento da corrente será menor que 50 metros. Representando 50 metros como 5000 centímetros para igualar as unidades de medida temos,

$$Y = \sum_{k=1}^{1002} X_k < 5\,000.$$

Sendo assim,

$$\begin{aligned} \sum_{k=1}^{1002} X_k < 5000 \quad &\Leftrightarrow \quad \frac{1}{1002}\sum_{k=1}^{1002} X_k < \frac{5000}{1002} \\ &\Leftrightarrow \quad \frac{1}{1002}\sum_{k=1}^{1002} X_k < 4.99 \\ &\Leftrightarrow \quad \mu_Y < 4.99. \end{aligned}$$

Vamos encontrar a probabilidade de que para mais de 1002 elos formarem uma corrente de 50 metros a média de comprimento de cada elo deve ser menor que 4.99 centímetros, isto é,

$$P(\mu_Y < 4.99).$$

Como $n = 1002$, o Teorema Central do Limite nos permite utilizar a aproximação normal através da variável Z.

$$Z = \frac{\mu_Y - \mu_X}{\sqrt{\sigma_Y}} = \frac{4.99 - 5}{\sqrt{d\frac{0.04}{1002}}} = \frac{-0.01}{0.00632} = -1.58.$$

Com isso, $P(\mu_Y < 4.99) = P(Z < -1.58) = 0.0571$.

Portanto, 5.71% é a probabilidade para uma corrente de pelo menos 50 metros, serão necessários mais de 1002 elos. A Figura 10 ilustra claramente a probabilidade como sendo a área cinza sob a curva normal referente aos 5.71%.

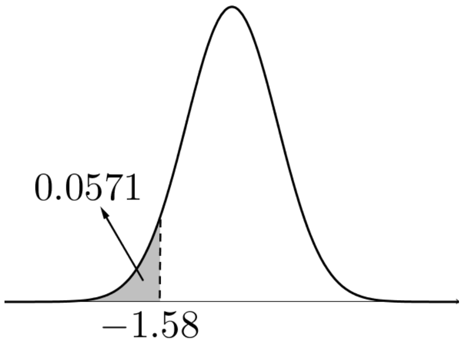


Figura 10 – Probabilidade de mais de 1002 elos numa corrente de 50 metros.

(b) Na solução do item (a) vimos que a média de comprimento para cada elo era de 4.99 centímetros e como o laboratório de pesquisa concluiu este mesmo valor em suas investigações então realmente é necessário informar ao departamento de vendas a mudança mesmo esta sendo tão pequena.

O capítulo seguinte continuará abordando situações onde podemos aplicar o Teorema Central do Limite porém de maneira mais específica que é na inferência estatística.

6 Aplicações do Teorema Central do Limite na inferência estatística.

Na inferência estatística, o TCL atinge um alto grau de relevância pois há condições extremamente necessárias e favoráveis para o cumprimento de suas propriedades. Seus objetivos tornam-se completos já que em situações em que a distribuição da variável é desconhecida, com amostras iguais ou superiores a trinta, as estimativas acontecem mediante a aproximação normal utilizando-se de técnicas simples e rápidas através da curva normal e da tabela z.

O desenvolvimento deste capítulo dar-se-á através de três seções. Na primeira seção haverá uma abordagem inicial do método inferencial tais como estimadores, parâmetros e outros tópicos de relevância. Na segunda, veremos situações problemas envolvendo distribuições amostrais com estimativas pontuais. Na terceira e última, veremos situações envolvendo as estimativas intervalares, comumente chamadas de intervalos de confiança. Para maiores estudos, consultar (DEVORE, 2006) e (FARBER, 2010).

Torna-se essencial o tratamento de algumas definições e características da estatística inferencial para uma melhor compreensão das situações problemas propostas pois não foi discorrido em nenhum capítulo anterior. Este é o propósito da seção a seguir.

6.1 Noções gerais de inferência estatística

O termo inferência nos remete a um tipo de raciocínio chamado de indutivo no qual, através de casos particulares chegamos a uma conclusão geral. Na estatística, a inferência exerce a mesma função de generalizar conclusões porém as premissas são chamadas de amostras dentro de um geral denominado de população. Assim, os termos população e amostra são a base inicial na estatística inferencial.

Ao incluirmos a probabilidade, estamos criando uma ligação entre os métodos da estatística descritiva e da estatística inferencial levando a uma melhor compreensão de como as técnicas inferenciais são desenvolvidas e aplicadas, como as conclusões estatísticas podem ser traduzidas para a linguagem cotidiana e interpretadas,transformando-se em tomada de decisões e na estimativas de erros que podem ocorrer durante o processo inferencial. Enquanto que a probabilidade faz suas considerações da população para a amostra caracterizando o método do raciocínio dedutivo, a estatística inferencial faz suas considerações da amostra para a população distinguindo-se assim como raciocínio indutivo.

Na inferência estatística lidamos com conceitos específicos que precisam estar bem definidos a fim de não gerar dúvidas nas questões propostas deste capítulo. Um parâmetro é uma característica da população. Como exemplos, podemos citar a média aritmética, a mediana ou a variância. Representamos os parâmetros através de letras gregas. Em muitas populações, mesmo de tamanho finito, encontrar os valores numéricos precisos de seus parâmetros é muito complicado ja que nem todas as possibilidades existentes podem ser estudadas. Dado a esse fato, precisamos de estimadores ou estatísticas que cumpram esse papel de fornecer dados importantes num estudo inferencial. Estes dados numéricos que um estimador fornece é denominado de estimativas. Representamos os estimadores através de letras do nosso alfabeto em formato maiúsculo. Eis alguns dos estimadores utilizados neste trabalho.

- ❏ $\overline{X} = \dfrac{\sum X_i}{n}$ para o parâmetro μ (média populacional).
- ❏ $S^2 = \dfrac{\sum(X_i - X)^2 F_i}{n-1}$ para o parâmetro σ^2 (variância populacional).
- ❏ $S = \sqrt{S^2}$ para o parâmetro σ (desvio padrão populacional).

Há outros modelos de estimadores tais como: a frequência relativa, utilizado para a proporção amostral de um evento populacional e a soma de duas médias populacionais, citados aqui apenas a título de informação.

Concluindo esta etapa podemos observar que a inferência estatística se apodera de amostras coletadas e, através de estimadores e estimativas, generalizam características fundamentais como os parâmetros, realizando assim, estudos de uma população. Os dois pontos fortes do estudo de uma população pela inferência estatística é estimar parâmetros de forma pontual ou intervalar e realizar testes de hipótese. Estes procedimentos serão relatados nos tópicos abaixo. A seguir, vamos entender onde o Teorema Central do Limite se encaixa na estatística inferencial.

6.2 Distribuição amostral

O primeiro passo na inferência estatística consiste em selecionar uma amostra para análise. Definindo formalmente uma amostra aleatória , temos

Definição 6.1 *Dadas uma população e uma variável X que se pretende estudar, uma amostra aleatória de X é um conjunto de dimensão n representado por $(X_1, X_2, \ldots, X_n)$ onde cada termo da amostra possua a mesma característica de X.*

A Definição 6.1 afirma que se a variável X de uma população possuir, por exemplo, uma distribuição uniforme contínua então cada termo de uma amostra também possuirá distribuição uniforme contínua. Em estatística, a distribuição da variável representa uma de suas características.

Para uma variável aleatória populacional X, tomemos uma amostra aleatória denominada de $\overline{X}$. Quando X possuir distribuição normal ou aproximadamente normal, através do Teorema Central do Limite, temos $X \sim N(\mu, \frac{\sigma^2}{n})$. Isso significa que a média de $\overline{X}$ é a mesma média de X,

$$E(\overline{X}) = E(X) = \mu$$

e, a variância de $\overline{X}$ é a variância populacional sobre o número de elementos de $\overline{X}$, ou seja,

$$Var(\overline{X}) = \frac{Var(X)}{n} = \frac{\sigma^2}{n}.$$

O enfoque principal, neste capítulo, foi mostrar a utilização do Teorema Central do Limite em situações onde a distribuição da população fosse desconhecida[1]. Assim, temos duas situações distintas a considerar: uma é onde a média e a variância populacional são conhecidas. Neste caso, a igualdade (13) do Capítulo 2 é reescrita como

$$z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}. \tag{76}$$

A outra é quando conhecemos a média populacional mas a variância populacional é desconhecida. Neste caso, estimamos σ^2 através da variância amostral S^2. Com isso, a igualdade (76) será

$$z = \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}}. \tag{77}$$

[1] Está subentendido que o número de elementos das amostras n é igual ou superior a 30 evitando assim redundâncias desnecessárias. Lembrando que $n = 30$ é o limite inferior, o ideal são amostras de tamanhos bem maiores que 30.

Quando o tamanho de uma amostra for inferior a 30, estimamos σ^2 por S^2 e utilizamos a distribuição T de Student com $n - 1$ graus de liberdade. Para maiores consultas vide (NATARIO, 2012).

Vamos apresentar aqui dois exemplos representando situações problemas com distribuições amostrais. O primeiro exemplo apresenta uma situação que aborda o conceito de proporção populacional e amostral, a relação da proporção com a distribuição binomial e a influência do Teorema Central do Limite na solução do problema. Este exemplo foi adaptado de (MORETTIN, 2004).

Exemplo 6.1 *Uma vacina contra a gripe será utilizada e o fabricante confirma a eficácia em 80% dos casos. Para admitir como verdade o que foi dito pelo fabricante, serão realizado testes para a verificação da imunidade ou não destes indivíduos. Com isso, 36 indivíduos que tomaram a vacina foram selecionados para estes testes. Baseado nestas informações e admitindo que o fabricante está correto em sua afirmação, encontre a probabilidade, na amostra selecionada, da proporção de imunizados ser inferior a 0.75.*

Solução : Vamos admitir uma variável aleatória X representando o número de imunizados na amostra selecionada. Não conhecemos a média e nem a variância desta distribuição porém, podemos perceber que nestes testes só há duas possibilidades, estarem imunizados ou não e estes resultados retrata o modelo de distribuição binomial. Pelo tamanho da amostra, o Teorema Central do Limite nos permite realizar uma aproximação normal, através da variável Z. Seja P a proporção amostral com $P < 0.75$, devemos encontrar $\mathrm{P}(P < 0.75)$. Com isso,

$$Z = \frac{P - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{0.75 - 0.8}{\sqrt{\frac{0.8(0.2)}{36}}} = \frac{-0.05}{0.067} = -0.75$$

Assim,

$$\mathrm{P}(P < 0.75) = \mathrm{P}(Z < -0.75) = 0.2266$$

Portanto a probabilidade de imunizados, na amostra, ser inferior a 0.75 é de 22.66%. A Figura 11 ilustra graficamente a probabilidade encontrada através da aproximação normal.

Podemos observar na resolução do Exemplo 6.1 que a proporção é equivalente a média, seja ela amostral ou não.

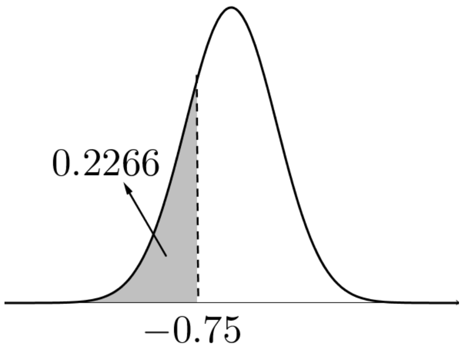


Figura 11 – Representação gráfica da probabilidade de imunizados

O próximo exemplo nos traz uma variável com distribuição desconhecida e a população possui um número finito de elementos.

Exemplo 6.2 *A fim de ter uma referência numérica em torno de um total de 300 contas sob seus cuidados, um analista financeiro toma uma amostra de 15% destas contas e acha um saldo médio amostral* $\overline{X}$ *e um desvio padrão de* $S = R\$\,35\,750.00$. *Admitindo que o saldo médio das 300 contas seja de R\$ 138 000.00, encontre um valor estimado da probabilidade de se obter* $\overline{X}$ *igual ou superior a* $R\$\,148\,500.00$.

Solução : Este exemplo retrata uma situação onde a distribuição da variável aleatória X é desconhecida. Mesmo a questão nos informando o valor do desvio padrão, precisamos estimar o valor do erro padrão da média amostral, necessário no cálculo da probabilidade exigida. Como a população possui um valor finito de contas, vamos aplicar a seguinte fórmula:

$$S_{\overline{X}} = \frac{S}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

Assim, o erro padrão de $\overline{X}$,

$$S_{\overline{X}} = \frac{35\,750}{\sqrt{45}} \sqrt{\frac{300-45}{300-1}} = 5329.2953\,\sqrt{0.8528} = 4921.4534$$

Temos que o erro padrão é igual a R\$ 4921.45.

Com este valor podemos calcular a probabilidade proposta no enunciado da questão, isto é, $P(X \geq 148500)$. Como o número da amostra coletada é superior a 30, o Teorema Central do Limite nos permite utilizar a variável Z aproximando a distribuição para uma normal padrão, ou seja $X \sim N(0,1)$.

Sendo assim temos,

$$\begin{aligned}
P(X \geq 148500) &= P\left(Z \geq \frac{\overline{X} - \mu}{S_{\overline{X}}}\right) = P\left(Z \geq \frac{148500 - 138000}{4921.45}\right) \\
&= P(Z \geq 2.13) = 0.5 - P(Z \leq 2.13) \\
&= 0.5 - 0,4834 = 0.0166.
\end{aligned}$$

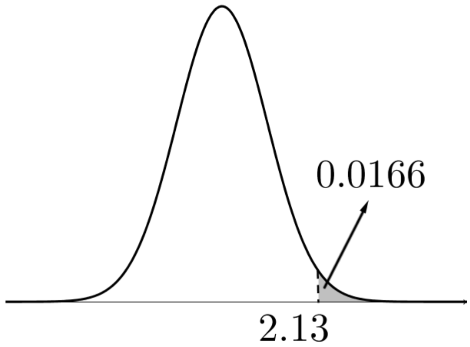


Figura 12 – Representação gráfica da probabilidade do saldo médio amostral

Portanto, a probabilidade do saldo médio da amostra coletada das contas feita pelo analista financeiro ser igual ou superior a R$ 148 500.00 é de 1.66%.

Vejam que na aproximação normal padrão, a probabilidade representada pela área sombreada da Figura 12 compreende a cauda do lado direito da curva normal. E para valores próximos às caudas, a área é muito pequena, explicando assim o resultado encontrado na questão.

6.3 Intervalos de confiança

Uma estimativa pontual, como vimos na Seção 6.2, é representado por um único valor e, em muitos casos, não reflete a verdadeira característica do parâmetro exato de uma população. Um intervalo possui poder de abrangência maior que um valor único. Por isso, uma estimativa no formato intervalar reflete mais significativamente um parâmetro populacional exprimindo a probabilidade de confiança que este intervalo contenha o parâmetro em questão. A este intervalo denominamos de intervalo de confiança.

Podemos definir o intervalo de confiança como sendo um intervalo aberto, com números racionais, utilizado para estimar parâmetros populacionais. Geralmente os parâmetros mais estimados são a média, a variância, o desvio padrão, a diferença de médias e a proporção populacional. Para um melhor entendimento das resoluções apresentadas nesta seção precisamos abordar algumas nomenclaturas específicas. A probabilidade que o intervalo vai conter o parâmetro estimado é denominado de nível de confiança ou coeficiente de confiança. Representamos por $(1-\alpha)$x100% onde α é a probabilidade complementar do nível de confiança. A extremidade à esquerda é denominado de limite inferior de confiança e a extremidade à direita, de limite superior de confiança. Por fim, a margem de erro num intervalo de confiança é a diferença entre os limites de confiança e sua referente estimativa pontual.

Para o tamanho n de uma amostra suficientemente grande, $n \geq 30$, o Teorema Central do Limite aproxima a função de distribuição amostral para uma distribuição normal padrão e estudamos o comportamento do intervalo a partir desta aproximação. Através da Figura 13, podemos perceber melhor os conceitos apresentados acima, isto é, o nível de confiança $(1-\alpha)$x100% é a área cinza. Quanto maior o nível de confiança, maior é a área na curva. A área restante é denominada de α. As extremidades da curva são chamadas de caudas e como há duas caudas, temos que cada cauda tem área $\alpha/2$. Os valores $\pm z_{1-\frac{\alpha}{2}}$ são os pontos críticos da curva.

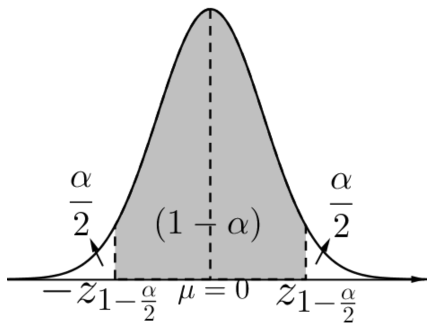


Figura 13 – Representação gráfica do intervalo de confiança numa curva normal.

Nesta seção haverá dois exemplos envolvendo parâmetros distintos. Mais especificamente, um parâmetro é a diferença de duas médias e o outro é a proporção populacional. Poderemos observar a funcionalidade do Teorema Central do Limite estimando intervalos de confiança, amplitude de intervalos e margem de erros. O primeiro exemplo apresenta uma situação envolvendo o cálculo de intervalos de confiança da diferença de duas médias populacionais, onde a variável aleatória possui distribuição desconhecida mas as variâncias σ^2 são conhecidas.

Exemplo 6.3 *Os níveis de monóxido de carbono (CO) no ar é medido em ppm, que quer dizer, partes por milhão.*[2] *A fim de conhecer os níveis de poluição atmosférica de uma capital brasileira, foi solicitado um estudo comparando tais níveis de poluição durante o dia e durante a noite. Duas amostras foram coletadas. Para compor a primeira amostra, foram realizadas medições de concentrações de CO no ar, em ppm, durante 32 dias não consecutivos, sempre às 15h da tarde, registrando-se uma média de 0.25 ppm. Para compor a segunda amostra, foram realizadas*

2 A concentração em ppm indica quantas partes do soluto existem em um milhão (10^6) de partes da solução (em volume ou em massa). Para soluções gasosas, a concentração em ppm é expressa em volume.

medições de concentrações de CO no ar, em ppm, durante 30 dias, aleatoriamente, sempre às 2h da manhã, registrando-se uma média de 0.15 ppm. Supondo que concentração de CO no ar seja uma variável aleatória de desvio padrão 0.05ppm durante a noite e 0.12ppm durante o dia, construa um intervalo de confiança, a um nível de significância de 90%, para a diferença das concentrações médias de CO de dia e à noite.

Solução : Seja X_1 a variável aleatória referente ao nível de CO, em ppm, no ar durante o dia nesta capital brasileira e X_2 referente ao nível de CO, em ppm, durante a noite. A distribuição de X_1 e X_2 são desconhecidas porém a variância é conhecida e o número de amostras é igual ou superior a 30. Sendo assim, podemos aplicar o Teorema Central do Limite e fazer uma aproximação normal, através da variável aleatória Z.

Para encontrarmos a solução do problema, devemos aplicar a fórmula do intervalo de confiança para a diferença de médias, $\mu_1 - \mu_2$, descrita abaixo,

$$IC_{(1-\alpha)100\%} = \left((\overline{X_1} - \overline{X_2}) - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} ; (\overline{X_1} - \overline{X_2}) + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right).$$

O enunciado do problema nos fornece o seguinte, $n_1 = 32$, $n_2 = 30$, $\overline{X_1} = 0.25$, $\overline{X_2} = 0.15$, $\sigma_1^2 = 0.05$ e $\sigma_2^2 = 0.12$. Precisamos encontrar $z_{1-\dfrac{\alpha}{2}}$, através da tabela z da Figura 3, onde $P(Z \leq z_{1-\frac{\alpha}{2}}) = 0.9$. Logo, $z_{1-\frac{\alpha}{2}} = 1.645$

Com isso,

$$\begin{aligned}
IC_{90\%} &= \left((\overline{X_1} - \overline{X_2}) - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} ; (\overline{X_1} - \overline{X_2}) + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right) \\
&= \left((0.25 - 0.15) - 1.645 \sqrt{\frac{0.05^2}{32} + \frac{0.12^2}{30}} ; (0.25 - 0.15) + 1.645 \sqrt{\frac{0.05^2}{32} + \frac{0.12^2}{30}} \right) \\
&= \left(0.1 - 1.645\sqrt{0.000558} ; 0.1 + 1.645\sqrt{0.000558} \right) \\
&= (0.061; 0.138).
\end{aligned}$$

Portanto, P($0.061 \leq \mu_1 - \mu_2 \leq 0.138$) = 0.9. Isso quer dizer que há 90% de confiança de que o intervalo entre 0.061 e 0.138 contenha a diferença das médias reais de concentração de CO no ar.

Observe que o número zero não pertence ao intervalo encontrado. Logo, as médias μ_1 e μ_2 não são iguais e, como as extremidades do intervalo são positivas então $\mu_1 > \mu_2$, ou seja, a média de concentração de CO no ar durante o dia é maior que a concentração de CO durante a noite.

O próximo exemplo trata-se de uma situação envolvendo estimativa intervalar para proporção populacional p, cuja variável aleatória não possui distribuição normal e a variância σ^2 é desconhecida.

Exemplo 6.4 *Devido ao aumento da inadimplência por parte da população, o gerente financeiro de uma rede de lojas solicitou uma pesquisa de seus clientes acerca do uso de cartão de crédito mencionando o seguinte aspecto: quitação total ou parcial das faturas mensais do ano anterior. O resultado da pesquisa foi o seguinte, numa amostra, selecionada ao acaso, de 180 clientes, verificou-se que 115 pagaram juros devido a pagamentos parciais de suas faturas.*

(a) Encontre o intervalo de confiança, a um nível de 97.5%, para a proporção real dos clientes com cartão de crédito que pagaram juros no ano anterior.

(b) Qual o tamanho de uma amostra para que, a um nível de confiança de 97.5%, o intervalo possua uma amplitude de 0.05?

Solução :

(a) Seja X a variável aleatória que representa o número de clientes dessa rede de lojas que usam o cartão de crédito. Dos clientes usuários de cartão de crédito, interessa apenas aqueles que pagaram juros no ano anterior ou seja, uma proporção da amostra. Seja p a proporção populacional. Como p é desconhecida, devemos estimá-lo por P, a proporção amostral. Encontrando o valor de P,

$$P = \frac{115}{180} = 0.639.$$

A variável X possui distribuição binomial de parâmetros $n = 180$ e $P = 0.639$. Como $nP \geq 5$ e $n(1-P) \geq 5$, vamos realizar uma aproximação normal pela variável Z, através do Teorema Central do Limite. Para encontrarmos o intervalo de confiança devemos utilizar a fórmula:

$$IC_{(1-\alpha)100\%} = \left(P - z_{1-\frac{\alpha}{2}}\sqrt{\frac{P(1-P)}{n}}; P + z_{1-\frac{\alpha}{2}}\sqrt{\frac{P(1-P)}{n}}\right).$$

Através da tabela z da Figura 3, temos que $z_{1-\frac{\alpha}{2}} = 2.24$. Com isso, temos todos os dados necessários para compor a fórmula acima com $n = 180$ e $P = 0.639$. Assim,

$$\begin{aligned} IC_{97.5\%} &= \left(P - z_{1-\frac{\alpha}{2}}\sqrt{\frac{P(1-P)}{n}}; P + z_{1-\frac{\alpha}{2}}\sqrt{\frac{P(1-P)}{n}}\right) \\ &= \left(0.639 - 2.24\sqrt{\frac{0.639(1-0.639)}{180}}; 0.639 + 2.24\sqrt{\frac{0.639(1-0.639)}{180}}\right) \\ &= \left(0.639 - 2.24\sqrt{0.00128}; 0.639 + 2.24\sqrt{0.00128}\right) \\ &= (0.559; 0.719) \end{aligned}$$

Portanto, P$(0.559 \leq p \leq 0.719) = 0.975$. Isso quer dizer que há 97.5% de confiança que o intervalo entre 0.559 e 0.719 contenha a proporção real de clientes que pagaram juros no cartão de crédito no ano anterior.

(b) A amplitude de um intervalo é dado pela diferença de seus valores extremos. No intervalo de confiança, dizemos que a amplitude é a diferença dos limites de confiança inferior e superior, expressa por:

$$2z_{1-\frac{\alpha}{2}}\sqrt{\frac{P(1-P)}{n}}$$

onde $z_{1-\frac{\alpha}{2}}\sqrt{\frac{P(1-P)}{n}}$ é conhecido como margem de erro ou erro máximo de estimativa. Vamos agora calcular o tamanho n de uma amostra cuja amplitude seja de 0.05. Os valores necessários para compor a fórmula abaixo são os mesmos do item (a).

$$\begin{aligned}
2z_{1-\frac{\alpha}{2}}\sqrt{\frac{P(1-P)}{n}} &= 0.05 \\
2.24\sqrt{\frac{0.639(1-0.639)}{n}} &= \frac{0.05}{2} \\
\frac{\sqrt{0.2307}}{\sqrt{n}} &= \frac{0.025}{2.24} \\
\sqrt{n} &= \frac{2.24\sqrt{0.2307}}{0.025} \\
n &= \left(\frac{2.24\sqrt{0.2307}}{0.025}\right)^2 \\
n &= 1852.096.
\end{aligned}$$

Como o tamanho da amostra é representado por um número natural, temos que $n \approx 1853$ clientes selecionados para a pesquisa.

O próximo capítulo trata-se do conceito do Teorema Central do Limite aplicado em sala de aula no Ensino Médio. As atividades apresentadas correspondem ao nível que um aluno do Ensino Médio possa acompanhar sem problemas em relação a pré-requisitos conceituais.

7

Proposta de atividades em sala de aula

Este capítulo destina-se a apresentar algumas propostas de atividades a serem aplicadas no Ensino Médio. Estas atividades vão buscar o conceito intuitivo que os alunos possuem acerca das características do Teorema Central do Limite utilizando, para isso, os conhecimentos matemáticos já estudados, a experiência individual de mundo e o conteúdo atual. O principal objetivo deste capítulo é mostrar que o conceito do Teorema Central do Limite não está apenas nos estudantes ou profissionais das áreas afins como matemática, estatística ou as engenharias, encontra-se nas pessoas, independentemente de seu grau de escolaridade ou área de atuação, isso porque estes conceitos estão no mundo em que vivemos e na sociedade onde somos parte integrante. O que acontece é que muitas pessoas não se deparam ou não são estimuladas a refletir nestas situações que exigem tal compreensão. No âmbito escolar, estas situações são colocadas em prova para que os alunos reflitam, questionem e construam conceitos próprios encontrando técnicas para solucionar as situações problemas existentes.

Segundo a proposta curricular da Secretaria de Educação para o Ensino Médio, em conjunto com o PCNEM (BRASIL, 1999), o aluno, no dever de suas atribuições, deve ser capaz de utilizar conhecimentos de estatística e probabilidade como recurso para a construção de argumentação com o intuito de avaliar propostas de intervenção na realidade que o cerca. A necessidade do professor estar desenvolvendo tais competências perpassa por atividades como as que serão apresentadas aqui.

Vamos discorrer sobre três atividades em sala de aula onde os alunos colocarão em prática seus conhecimentos, mostrando o quanto aprenderam acerca do conteúdo que estão estudando agora. A primeira atividade relaciona conceitos de estatística e de probabilidade para chegarmos a noção do Teorema Central do Limite.

Atividade 01 : (**Experimento, em sala de aula, com o dado.**)

- ❏ Objetivo da atividade:
  Reconhecer o Teorema Central do Limite como fundamento probabilístico/estatístico

necessário para aproximar características reais com base nas características amostrais.

- ❏ Conteúdos prévios necessários:
  Leitura e interpretação de gráficos estatísticos e medida de tendência central e de dispersão.

- ❏ Recursos necessários:
  Um dado, caderno, lápis ou caneta, uma calculadora básica, data show, quadro e piloto ou giz branco.

- ❏ Desenvolvimento da atividade:

Cada aluno vai trazer de casa um dado comum. A atividade pode ser de individual porém, o professor pode pedir que os alunos formem grupos de acordo com a afinidade de cada um. Os alunos escolherão por conta própria o número de jogadas por rodada, contanto que o número de jogadas estejam em ordem crescente, por exemplo, duas jogadas na 1ª rodada, 5 jogadas na 2ª rodada e assim por diante. Em cada rodada, os alunos registrarão quantas jogadas foram realizadas e o número da face do dado voltado para cima em cada jogada. Dez rodadas são suficientes para podermos tirar as conclusões necessárias.

Finalizado esta etapa onde todos os alunos da turma fizeram suas dez rodadas e registraram no caderno, a próxima etapa é calcular a média ponderada por rodada com a ajuda de uma calculadora, caso necessite.

O professor calcula a média real no quadro e os alunos vão comparar seus resultados com o resultado encontrado pelo professor.

$$\frac{1+2+3+4+5+6}{6} = \frac{21}{6} = 3.5. \tag{78}$$

O professor lança a seguinte pergunta: O que acontece com as médias encontradas por vocês enquanto o número de jogadas aumentam?

O professor apresenta uma ilustração composta por histogramas de algumas possíveis jogadas. Analisando a ilustração, o professor esclarece que a medida que o números de jogadas aumentam, as médias encontradas vão aproximando-se de um valor central e a dispersão diminui, ou seja, os valores na abscissa não se distanciam muito do centro. Outro questionamento é lançado para os alunos: Se o números de jogadas numa rodada for muito grande, algo em torno de mil jogadas, o que acontece com essa média? E com o histograma que o representa?

Após algumas opiniões, o professor esclarece que para uma rodada com jogadas cada vez maiores, a média dessa rodada se torna muito próxima da média real calculada no quadro e que a coluna central do histograma estará cada vez maior. A explicação matemática para esta observação é o Teorema Central do Limite que afirma que para rodadas com um número de jogadas indefinidamente crescente, a sua média coincidirá coma média real do dado.

A segunda proposta consiste numa atividade onde os alunos terão a possibilidade de interferir no acaso a partir de um experimento aleatório bastante comum nos estudos de probabilidade que são retirada de bolas de uma urna. Vejamos a proposta e como essa intervenção pode ser feita.

Atividade 02 : (**Amostras a partir de uma urna.**)

- ❏ Objetivo da atividade:
  Compreender que a probabilidade de ocorrência de um evento dentro de uma espaço amostral aumentar é diretamente proporcional ao número de amostras coletadas e relacionar este fato como uma propriedade do Teorema Central do Limite.
- ❏ Conteúdos prévios necessários:
  noções básicas de estatística e cálculo de probabilidades.
- ❏ Recursos necessários:
  Uma caixa, oito bolinhas de mesmo tamanho e textura, caderno, lápis ou caneta, quadro e piloto ou giz branco.
- ❏ Desenvolvimento da atividade:

Uma urna contendo oito bolas de duas cores distintas, quatro azuis e quatro marrons. As bolas azuis tem valor um e as marrons, valor dois. O aluno tem a opção de retirar três ou quatro bolas sem reposição. Após as retiradas, as bolas voltam para a urna e o processo recomeça com o próximo aluno. Todos os alunos daquela turma participarão desta atividade. Os alunos devem conseguir média aritmética igual ou superior a média populacional ou real das bolas na urna para saírem vencedores. Com a ajuda do professor, calculam-se a média real das oito bolas contidas na urna,

$$\frac{1+1+1+1+2+2+2+2}{8} = \frac{12}{8} = 1.5. \tag{79}$$

Logo, a média populacional é de 1.5.

Antes de cada aluno ir a urna realizar sua tentativa, o professor vai coletar a primeira informação importante. No quadro, o professor coloca três opções: 3 tentativas, 4 tentativas ou tanto faz, cada aluno faz a sua escolha e diz o porque da escolha. A partir daí, cada um faz sua tentativa, calcula a sua média e compara com a média populacional obtida anteriormente, $\mu = 1.5$. O processo de retirar as bolas da urna é finalizado por todos os alunos.

Após sorrisos, tristezas e a angustias, o professor revela que nem tudo teve a sorte como fator preponderante. Na opção de realizar três ou quatro tentativas, houve-se um mecanismo fazendo a sorte tender para o nosso lado que é através da probabilidade. E, retomando a discussão da escolha de 3, 4 tentativas ou tanto faz, mostra-se que quanto maior o número de tentativas ou quanto mais próximo o número de tentativas estiver do total de uma população, mais próximo da média real, os resultados estarão. O professor questiona quem sabia ou imaginava a afirmação acima, cria-se um ambiente de discussão, com mediação do professor, para introduzir o conceito intuitivo do Teorema Central do Limite.

Esta última proposta é uma atividade que simula um jogo de loteria. Ela representa um experimento, que através de uma atividade dinâmica, os alunos terão a oportunidade de conhecer o Teorema Central do Limite e também estarão construindo conceitos importantes dos conhecimentos necessários na aplicação desta atividade. A proposta foi criada a partir de uma atividade contida em (FRAGA, 2013).

Atividade 03 : (**Loteria em sala de aula**)

- ❏ Objetivo da atividade:
  Identificar, nas situações cotidianas, condições favoráveis para a atuação do Teorema Central do Limite encurtando distâncias entre estimativas amostrais e reais.

- ❏ Conteúdos prévios necessários:
  Soma dos termos de uma progressão aritmética e medidas estatísticas: média aritmética, variância e desvio padrão.

- ❏ Recursos necessários:
  Uma caixa, não transparente, para sortear os números, 25 bolinhas numeradas de 1 a 25, fichas impressas contendo seis cartelas cada uma, lápis ou caneta, quadro, piloto ou giz branco.

- ❏ Desenvolvimento da atividade:

Nesta atividade o professor vai dividir a turma em grupos com quatro ou cinco alunos. A cada aluno será entregue uma ficha com seis cartelas idênticas enumeradas de 1 a 25. Os alunos irão marcar cinco números por cartela da maneira que quiserem. A atividade vai ser desenvolvida individualmente, mesmo os alunos estando em grupos. A proposta consiste em sortear cinco números desses 25 números completando assim uma rodada e realizar seis rodadas completando todas as cartelas da ficha que foi entregue. A cada rodada, o professor observa se algum aluno acertou os cinco números sorteados numa mesma cartela anotando no quadro os números que saíram. Ao final das seis rodadas, o professor tem em mãos as seis cartelas preenchidas em cada ficha e as quantidades que cada número foi sorteado. Caso algum aluno complete uma ou mais cartelas de sua ficha, o que é pouco provável, ótimo, o professor e a turma parabeniza-o mais afirma que o objetivo principal está em construção.

O passo seguinte é direcionar a atenção dos alunos para o número de vezes que os números da cartela foram sorteados. O professor reescreve os dados agrupados numa tabela em ordem decrescente e pergunta aos alunos quais números mais apareceram ao final das seis rodadas. Apos as observações perante a tabela no quadro, o professor lança outro questionamento: estes valores fariam vocês reverem a escolha dos números para uma próxima rodada, ou não interferem na opção de vocês? Os alunos comentarão acerca das questionamento feito das mais diversas formas possíveis. Muitas opiniões levantadas pelos alunos vão gerar novas discussões aí o professor deve mediar este momento para não perder o direcionamento e o objetivo da atividade.

Em seguida, os alunos vão obter a média aritmética dos 25 números contidos numa cartela com o seu desvio padrão e também a média aritmética e o desvio padrão por rodada comparando os valores encontrados em cada um a fim de construírem uma conclusão fundamentada na proposta da atividade. A média aritmética μ dos números contidos na cartela é dada por:

$$\mu = \frac{1}{25}\sum_{n=1}^{25} n = \frac{325}{25} = 13.$$

A variância σ^2 dos números da cartela é:

$$\begin{aligned}\sigma^2 = \frac{1}{25}\sum_{n=1}^{25}(x_n - \mu)^2 &= \frac{1}{25}\left[(1-13)^2 + (2-13)^2 + \cdots + (24-13)^2 + (25-13)^2\right] \\ &= \frac{1300}{25} = 52\end{aligned}$$

e o desvio padrão é $\sigma = \sqrt{52} = 7.2$. Com isso, o intervalo numérico onde os números mais sorteados deveriam estar é entre 13 - 7.2 e 13 + 7.2, ou seja, entre 5.8 e 20.2.

De posse destes dados, o professor solicita aos alunos que façam comparação com a tabela dos sorteios contida no quadro. Provavelmente a maioria dos números estarão contidos no intervalo formado pelo desvio padrão em torno da média ou simplesmente não. O professor deve lembrar aos alunos que trata-se de um experimento aleatório e consequentemente imprevisível. O questionamento final é: Se as rodadas com cinco sorteios cada uma estendesse mais e mais vezes numa quantidade muito maior que seis, os dados seriam mais confiáveis? Ou seja, a probabilidade daqueles números serem sorteados numa determinada rodada eram maiores? Neste momento de exposição de ideias, o professor apresenta o Teorema Central do Limite, como noção conceitual, mostrando aos alunos que a partir de uma quantidade n de rodadas maior ou igual a trinta, o Teorema Central do Limite entra em ação fazendo a distribuição amostral das rodadas de sorteio aproximar-se de uma forma de distribuição conhecida como Distribuição Normal. O professor deve mostrar graficamente o comportamento de uma distribuição normal através da curva normal para os alunos possam observar a importância do Teorema Central do Limite, ressaltando que só é possível fazer tal aproximação quando $n \geq 30$. Quanto maior for o número n de rodadas, mais próxima a média amostral das médias obtidas por rodada estará da média real $\mu = 13$. Asim, os alunos perceberão que suas ideias intuitivas tem um rigor matemático muito bem fundamentado.

8

Conclusão

A elaboração deste trabalho atendeu aos requisitos mínimos de compreensão acerca de um tópico tão relevante e tão amplo na matemática que é o Teorema Central do Limite. Todas as notações, terminologias e definições básicas em probabilidade e estatística, utilizadas no trabalho, tais como variáveis aleatórias, esperança matemática, variância e função de distribuição foram atendidas, de forma clara e sucinta, no primeiro capítulo. No segundo capítulo, as proposições necessárias para desenvolver os teoremas limites como o Teorema da Unicidade e o Teorema da Continuidade de Paul-Levy, foram apresentadas de forma coerente e concisa, comprovando seu papel de oferecer os pré requisitos necessários para o desenvolvimento do Capítulo 4 . Os teoremas limites foram enunciados e demonstrados no terceiro capítulo, obedecendo todo o rigor matemático que eles exigem porém, mostrando-se compreensível em todo seu raciocínio dedutivo. Os Capítulos 5 e 6, dedicados a aplicabilidade do TCL, apresentaram, de forma correta e descomplicada, várias situações problemas no intuito de romper toda a rigidez matemática dos capítulos anteriores, abrindo espaço para a compreensão de utilidade prática e concreta essencial para o Capítulo 7, que descreve propostas de atividades que levam para os alunos do Ensino Médio o conceito do Teorema Central do Limite de forma dinâmica e prazerosa.

Muitos alunos, essencialmente nas escolas públicas, concluem a Educação Básica sem serem apresentados aos conceitos de estatística e/ou probabilidade e, quando tem acesso a estes conhecimentos, é de forma tão superficial que não dá espaço para construção de conceitos importantes e necessários para seu desenvolvimento conceitual e atitudinal. O Teorema Central do Limite vem como um mediador, fazendo com que esses alunos entendam que estes processos de intuição, que eles possuem, acerca de fatos imprevisíveis podem ser esclarecidos matematicamente, quantificados e postos em prática de uma forma compreensível e de fácil manipulação formando, assim, agentes ativos, participativos e transformadores no processo de construção do próprio conhecimento, abrindo portas para futuros profissionais na área da estatística e probabilidade.

Um ponto que não foi abordado aqui mas merece toda a atenção, principalmente, no que diz respeito a disseminação do Teorema Central do Limite, é a criação de materiais manipuláveis que torne as pessoas como agente ativo e participativo no processo de ensino e aprendizagem e, neste caso, temos a tábua de Galton. Muito interessante como este material é de fácil manipulação e consegue entregar realmente o conceito proposto por ele que é o processo de aleatoriedade, a forma de distribuição dos elementos e a noção do Teorema Central do Limite. Para maiores consultas vide (AQUINO, 2004) que ensina a confeccionar uma tábua de Galton e (SILVEIRA, 2011) aborda sobre erros e incertezas na metrologia a partir da tábua de Galton onde o Teorema Central do Limite faz-se presente na discussão.

Com tudo que foi exposto, os objetivos de compreensão e aplicação do Teorema Central do Limite foram alcançados com êxito. Esperamos que este trabalho contribua na divulgação da teoria das probabilidades em nossa sociedade, que a linguagem utilizada esteja acessível a alunos e professores da Educação Básica, graduandos, graduados e profissionais de áreas afins que estejam ávidos por novos conhecimentos ou necessitam destes conhecimentos. Os reflexos do Teorema Central do Limite estão ao nosso redor, acompanhado-nos em todos os lugares, não podemos simplesmente ignorar, pelo contrário, precisamos conhecê-lo e tirar bons proveitos a nosso favor.

Referências

AQUINO, P. M. de. **Relatório Final F-809: O Estudo da Distribuição Normal por Galton**. Campinas, 2004.

BELLHOUSE, D. R. **Maty's Biography of Abraham De Moivre, Translated, Annotated and Augumented**. [S.l.]: Institute of Mathematical Statístics, 2007. v. 22.

BRASIL. **Parâmetros Curriculares Nacionais: Ensino Médio**. [S.l.]: Ministério da Educação. Secretaria de Educação Média e Tecnológica, 1999.

COLETTI Élcio Lebensztayn e C. F. **Probabilidade: Teoria e Exercícios**. São Paulo: IME-USP, 2008. Desenvolvimento de material didático ou instrucional - Notas de Aula.

DEKKING, F. M. **A Modern Introduction to Probability and Statistics. Understanding Why and How**. [S.l.]: Springer, 2005.

DEVORE, J. L. **Probabilidade e Estatística para Engenharia e Ciências. Tradução de Joaquim Pinheiro Nunes da Silva**. São Paulo: Editora Thompson, 2006.

FARBER, R. L. e B. **Estatística Aplicada. Tradução por Luciane Ferreira Pauleti Vianna**. São Paulo: Pearson Prentice Hall, 2010.

FARIAS, A. M. L. de. **Variáveis aleatórias Contínuas**. Rio de Janeiro: Universidade Federal Fluminense, 2009. Desenvolvimento de material didático ou instrucional - Notas de Aula.

FELLER, W. **An Introduction to Probability Theory and its Applications, Volume 1**. [S.l.: s.n.], 1968.

FISCHER, H. **A History of the Central Limit Theorem From Classical to Modern**. New York: Springer, 2011.

FRAGA, R. R. **O ensino das loterias: Uma Abordagem Motivadora e Facilitadora para a aprendizagem de Probabilidade no Ensino Médio**. Rio de Janeiro-RJ: IMPA, 2013. Dissertação de Mestrado.

GARCIA, A. L. **Probability, Statístics, and Random Processes for Electrical Engineering**. Toronto: Pearson Prentice Hall, 2008.

GUIDORIZZI, H. L. **Um curso de cálculo, Volume I**. Rio de Janeiro: LTC Editora, 2006.

HALD, A. **A History of Parametric Statistical Inference from Bernoulli to Fisher, 1713 to 1935**. Copenhagen: Department of Aplied Mathematics and Statistics - University of Copenhagen, 2004.

HOEL, S. C. P. e. C. J. S. P. G. **Introdução à teoria da probabilidade. Tradução de Fernando Yassou Chiyoshi**. Rio de Janeiro: Editora Interciência, 1978.

JAMES, B. R. **Probabilidade: um curso em nível intermediário**. Rio de Janeiro: LTC Editora, 2010.

MARTINS, J. S. F. e Gilberto de A. **Curso de Estatística**. São Paulo: Editora Atlas, 2010.

MEYER, P. L. Probabilidade: Aplicações à estatística. In: **Probabilidade: Aplicações à Estatística**. [S.l.]: Livro Técnico, 1970.

MORETTIN, W. de O. Bussab e P. A. **Estatística Básica**. São Paulo: Editora Saraiva, 2004.

NATARIO, I. **Probabilidade e Estatística**. Caparica, Portugal: Universidade Nova Lisboa, 2012. Desenvolvimento de material didático ou instrucional-Notas de aula.

PINHEIRO, J. I. D. **Estatística Básica: A arte de trabalhar com dados**. Rio de Janeiro: Elsevier Editora, 2009.

ROSS, S. **Probabilidade: um curso moderno com aplicações. Tradução de Alberto Resende De Conti**. Porto Alegre: Editora Bookman, 2010.

SILVEIRA, P. L. J. e Fernando Lang da. **Discutindo os conceitos de erros e incerteza a partir da tábua de galton com estudantes de graduação: uma contribuição para incorporação de novas abordagens da metrologia ao ensino da física superior**. Porto Alegre-RS: Caderno Brasileiro de Ensino de Física, 2011. v. 28.