Universidade Regional do Cariri - URCA
Departamento de Matemática
Programa de Mestrado Profissional em
Matemática em Rede Nacional

Algumas Funções Aplicadas à Estatística

João Lourenço da Cruz Sobrinho

Juazeiro do Norte - CE

2023

Algumas Funções Aplicadas à Estatística

João Lourenço da Cruz Sobrinho

Dissertação apresentada ao Departamento de Matemática Pura e Aplicada da Universidade Regional do Cariri como parte dos requisitos exigidos para a obtenção do título de Mestre em matemática.

Orientador

Prof. Dr. Jocel Faustino Norberto de Oliveira

Juazeiro do Norte - CE

2023

Ficha Catalográfica elaborada pelo autor através do sistema
de geração automático da Biblioteca Central da Universidade Regional do Cariri - URCA

Sobrinho, João Lourenço da Cruz

S677a Algumas Funções Aplicadas à Estatística / João Lourenço da Cruz Sobrinho. Juazeiro do Norte-CE, 2023.

129p. il.

Dissertação. Mestrado Profissional em Matemática em Rede Nacional da Universidade Regional do Cariri - URCA.

Orientador(a): Prof. Dr. Jocel Faustino Norberto de Oliveira

1.Estatística , 2.Funções Aplicadas, 3.Modelos Probabilísticos; I.Título.

CDD: 510

Algumas Funções Aplicadas à Estatística

João Lourenço da Cruz Sobrinho

Dissertação apresentada ao Departamento de Matemática Pura e Aplicada da Universidade Regional do Cariri como parte dos requisitos exigidos para a obtenção do título mestre em matemática.

BANCA EXAMINADORA

Jocel Faustino Norberto de Oliveira

Prof. Dr. Jocel Faustino Norberto de Oliveira (Orientador)

Universidade Regional do Cariri (URCA)

Dr. Flávio França Cruz

Universidade Regional do Cariri (URCA)

Cícero Carlos Félix de Oliveira

Prof. Dr. Cícero Carlos Félix de Oliveira

Instituto Federal de Educação do Ceará (IFCE)

Dedico aos meus pais, esposa, filha e irmão

Osmar Lourenço.

Agradecimentos

Agradeço a Deus por ser fonte de sabedoria, resiliência e força em todos os momentos da minha trajetória de vida.

Aos meus amados pais, Juvenal Lourenço e Maria Odete, por serem minha base educacional, meu porto seguro e exemplos de vida a seguir.

Aos meus irmãos, Osmar, Ossean e Jonas (in memorian), que comigo compuseram um núcleo familiar propício ao companheirismo e ao meu desenvolvimento como pessoa.

A minha esposa, Rosilânia, pelo amor e companheirismo expresso de tantas formas.

A minha filhinha, Maria Vitória, por representar a personificação do amor e o renascimento da alegria e da esperança no seio da nossa família.

Aos meus queridos tios, Mariana Donina e Juvêncio Lourenço(in memorian), que ao lado dos meus pais foram fonte contínua de apoio, amor e carinho.

A todos os professores da URCA que lecionaram na turma Profmat/Seduc pelo apoio, transmissão de conhecimentos, compreensão e sensibilidade nos momentos mais difíceis desta caminhada.

Em especial, ao Professor Dr. Jocel Faustino Norberto de Oliveira por seu profissionalismo, acessibilidade, incentivo e companheirismo como orientador deste trabalho.

Aos meus queridos colegas do Profmat pela partilha, companheirismo e amizade construída durante este percurso. A vocês, minha gratidão pelo acolhimento e apoio, sobretudo, nos momentos mais necessários.

A tantos outros amigos e colegas de trabalho, que através de uma palavra, de um gesto ou outra ação fizeram desta minha caminhada a busca de um sonho coletivo.

"A mente que se abre a uma nova ideia jamais voltará ao seu tamanho original." (Albert Einstein)

Resumo

Neste trabalho, contemplamos o conceito e a análise de algumas propriedades das funções exponencial de base natural, gama e beta como recursos matemáticos aplicados à Estatística. Sem adentrar no universo Inferencial da referida área, buscamos, através de pesquisa bibliográfica, evidenciar a importância que tais funções desempenham na estrutura de modelos probabilísticos amplamente utilizados. Neste sentido, mencionamos fenômenos aleatórios relacionados às Teorias de Confiabilidade, das Filas, a estudos meteorológicos e à análise de sobrevivência. Para tanto, revisamos fundamentos sobre Estatística descritiva, conceito e cálculo de probabilidade, variáveis aleatórias e modelos probabilísticos. Como desdobramento natural deste percurso e parte da reflexão proporcionada, buscamos também apresentar possíveis conexões que podem ser estabelecidas ou fortalecidas entre o currículo do ensino médio e temas abordados neste estudo, tais como variáveis aleatórias e alguns modelos probabilísticos discretos.

Palavras-chave: Estatística, Funções Aplicadas e Modelos Probabilísticos.

Abstract

In this work, we contemplate the concept and analysis of some properties of exponential functions of natural base, gama and beta as mathematical resources applied to statistics. Without going into the Inferential universe of that area, we seek through bibliographic research to highlight the importance that such functions play in the structure of widely used probabilistic models. In this sense, we mentioned random phenomena related to Queuing Reliability Theories, meteorological studies and survival analysis. To do so, we review the fundamentals of descriptive statistics, the concept and calculation of probability, random variables and probabilistic models. As a natural unfolding of this journey and part and part of the proportionate reflection, we also seek to present possible connections that can be established or strengthened between the high school curriculum and topics addressed in this study, such as random variables and some discrete probabilistic models.

Keywords: Statistics, Applied Functions and Probabilistic Models.

Lista de Figuras

Lista de Tabelas

Sumário

1 Introdução

É do conhecimento geral que as tecnologias da informação e da comunicação ocupam cada vez mais espaço na vida cotidiana. Em tarefas de lazer, estudo, trabalho ou de rotinas, o contato com informações diversas é contínuo. Em meio a essa realidade, o homem, enquanto ser social, torna-se cada vez mais carente de habilidades relacionadas ao saber decodificar e analisar informações que lhes chegam com rapidez por diversos meios.

Certamente, o pleno exercício da cidadania hoje exige muito mais pré-requisitos que outrora. Vivendo em um cenário onde a dinâmica das relações são influenciadas e modificadas constantemente, ser capaz de filtrar informações confiáveis e ponderar decisões, em meio a incertezas, tornaram-se necessidades nos diversos aspectos da vida.

Como ciência que faz uso da Matemática e, dentre outras atribuições, trata do estudo de comportamentos coletivos, talvez nunca antes a presença da Estatística se fez tão forte e necessária na vida das pessoas. Paradoxalmente, a educação básica brasileira ainda enfrenta problemas clássicos no tratamento destes e outros temas afins do currículo. Em parte, muitos ainda relacionados a formação de professores [34].

Embora maior espaço e regulamentação tenha ocorrido desde os PCNs(1997) até a BNCC(2018)[1], o aprendizado da Estatística e da Probabilidade são alguns dos mais superficiais dentro do currículo básico em Matemática. Nota-se ênfase a operações técnicas, em detrimento da compreensão e conexão de conceitos [34]. Afastando-se de habilidades preconizados na Educação Estatística e relacionadas ao letramento, ao pensamento e ao raciocínio estatísticos.

Nesse sentido, mesmo sendo impulsionada a se reinventar com ações como novo ensino médio(2023), a educação escolar brasileira ainda carrega problemas antigos como

[1]PCNs-Parâmetros Curriculares Nacionais, cujas idéias centrais estão pressntes na BNCC-Base Nacional Comum Curricular [30].

o baixo índice na aprendizagem de matemática. Em geral, agravada pela pouca empatia expressa por muitos discentes em relação a área. Os quais ainda caracterizam a Matemática como junção de conteúdos fragmentados, desprovidos de significados ou sem conexão com a realidade.

Motivado por tal constatação, esse trabalho parte da percepção de que a Estatística tem um caráter integrador tanto de temas matemáticos, quanto de outras áreas [34]. E assim, inspirado no capitulo 04 do livro um curso de cálculo de Hamilton Luiz Guidorizzi [09], tem por objetivo evidenciar a importância de algumas funções aplicadas à Estatística. Construindo a possibilidade de ressignificar conceitos e despertar em professores do ensino básico o desejo de ações semelhantes voltadas aos seus contextos de ensino.

Com tal propósito, ainda se apoia na visão que externava Elon Lages Lima ao conceber a aprendizagem matemática como algo que se concretiza nas dimensões da assimilação, da manipulação e da aplicação de conceitos. Buscando também instigar professores, do citado nível de ensino, a ampliarem suas visões sobre o alcance de temas que, por ventura, não tenham sido tão bem contemplados nas suas formações iniciais.

Para tanto, este trabalho parte da revisão sobre conceitos básicos em estatístico no capitulo preliminar; Prossegue com apresentação dos conceitos de probabilidade e variável aleatória no capítulo 03; Discorre sobre modelos probabilísticos como temas que fazem conexão da estatística com a probabilidade no capítulo 04. E por fim, tem no último capitulo, a apresentação de algumas funções, uma análise de suas propriedades e a exposição de evidências da importância que desempenham para alguns modelos probabilísticos de ampla utilização pela Estatística. Como fecho, ainda apresenta algumas percepções sobre temas abordados e suas conexões com o currículo do ensino médio.

2 Preliminares

2.1 Estatística: Sobre a origem e conceitos elementares

A etimologia da palavra Estatística faz referência a "Status"(Estado em Latim) e sua origem se relaciona a técnicas que serviam aos propósitos dos governantes, sobretudo para contagem de pessoas, conhecimentos de seus domínios e melhor organização na arrecadação de impostos. Tais práticas são evidenciadas por registros históricos sobre civilizações antigas como as da Babilônia e a do Egito. Como tais vidências datam de cerca de 3000 anos a.C [01], precisar a origem da Estatística é algo difícil. Já que evidências mais antigas podem existir.

Aos poucos, apontamentos simples, caracterizados por números absolutos, dispostos em tábuas e|ou tabelas, foram dando espaço a presença de números relativos e a organizações mais elaboradas. Neste sentido, consta que no século XVII, através das Tábuas de mortalidade de Jonh Graunt(1620 - 1674) e William Petty (1623 - 1687) [01]foi possível perceber que, em termos percentuais, nasciam mais meninas que meninos. Algo que pode ser considerado uma das primeiras manifestações do caráter analítico e interpretativo na Estatística. E, assim, o despertar do teor científico.

Todavia, foi somente em 1708 que ocorreu a criação do primeiro curso de Estatística na Universidade de Yena na Alemanha. E apenas em 1740 a palavra estatística foi empregada pela primeira vez por Godofredo Achenwall [01]. O qual também é considerado o primeiro a fazer conexões desta com outras áreas. Nesse sentido, com base na mesma fonte, algo que permitiu a Estatística ampliar seu campo de ação foi o emprego, em suas técnicas, do cálculo de probabilidades em meados do século XIX.

Neste posto de ciência, construída em paralelo a história da humanidade e atrelada a suas necessidades, como tantas outras áreas do saber, a Estatística figura, em dias atuais, como ciência que trata da **coleta**, da **organização**, da **representação**, da **análise** e da **interpretação** de dados tendo como propósito a obtenção de informações de maneira **indutiva**. As quais auxiliam em diversas tomadas de decisões sob a presença de incerteza.

Em essência, a Estatística é responsável pelo estudo de comportamentos coletivos influenciados por diversos fatores, cujo o controle ou isolamentos individuais não são possíveis para estudo através do método experimental, próprio de outras ciências como a Química, Física e a Biologia, por exemplo [02].

Em essência, a pesquisa estatística compreende duas dimensões: A descrição dos dados, que fica a cargo da **Estatística Descritiva**; E a construção de inferências, de incumbência da **Estatística Inferencial**. Respectivamente, temos nelas um trabalho de compactação de dados com vistas a extrair informações mais importantes; Seguida de procedimentos que visam estender conclusões a um todo(população), a partir do que se obteve de uma parte representativa (amostra).

A pesquisa estatística busca estudar comportamentos ou características comuns a um determinado conjunto denominado **População ou Universo Estatístico**. Este pode ser finito ou não e ser formado por pessoas, regiões, objetos etc. Em função do tamanho da população, do tempo necessário a conclusão do estudo, e sobretudo das despesas que pode demandar uma pesquisa, essa pode ser dois tipos **censitária** ou **amostral**. Respectivamente, são pesquisas que envolvem toda população ou uma parte que seja representativa da população (**amostra**).

A maioria das pesquisas são amostrais e para tanto demandam uso de técnicas adequadas para garantir que a amostra escolhida seja representativa da população. Cada elemento da amostra é uma unidade de observação. E as características comuns investigadas são denominadas **Variáveis** . Essas são passíveis de mensuração ou classificação segundo alguma "escala"que não, necessariamente, tem o significado usual da palavra.

A variável pode ser **quantitativa**, se é de natureza numérica; ou **qualitativa**, caso contrário. O primeiro tipo subclassifica-se em **quantitativa discreta** e **quantitativa contínua**, caso represente contagem ou medida respectivamente. Já o segundo, pode ser **qualitativa ordinal**, se os conceitos ou atributos podem ser ordenados; ou ser **qualitativa nominal**, em caso contrário.

2.2 Representação de dados

Na sequência de etapas do método estatístico, como competência da Estatística Descritiva, ocorre a representação dos dados. Nesse estágio, busca-se condensar as observações de forma objetiva através de recursos que facilitem a leitura e a interpretação do que se deseja apresentar [01]. Para tal fim, faz-se o uso de **tabelas** e **gráficos**. A escolha do tipo e as configurações desses recursos seguem normatizações estabelecidas por órgãos competentes (como ABNT- Associação Brasileira de Normas Técnicas, e IBGE-Instituto Brasileiro de Geografia e Estatística).

Tabelas

Por tabela compreende-se toda estrutura na qual as informações estão dispostas em linhas e colunas. De modo geral, a normatização sobre esse recurso exige que o mesmo contenha **título**, **cabeçalho**, **coluna indicadora** e **fonte** como itens fundamentais [01]. De forma respectiva, esses itens informam: O que se representa, onde e quando foi realizado o estudo; O que está presente em cada coluna; Como se ler as informações das linhas; E de quem é a responsabilidade pelo estudo ou onde as informações foram coletadas.

De um modo geral, a importância das tabelas reside na possibilidade de agregar informações de maneira sucinta com detalhamento necessário a uma leitura objetiva.

Com este propósito, um conjunto de dados pode ser apresentado em função de fatores como **tempo**, **local**, **espécie** ou **frequência** com que os dados aparecem [02]. Sendo denominadas, respectivamente, séries estatísticas **Temporais**, **Geográficas**, **Específicas** ou **Distribuições de Frequências**.

As séries podem ser simples, se usam apenas um fator, ou mista em caco contrário. Já uma distribuição de frequência pode ser simples, se contempla uma única variável ou de classificação cruzada, caso contrário. Na tabela 2.1, temos um exemplo de série específica cronológica. Séries que fazem uso de dois fatores também são conhecidas por tabelas de dupla entrada [02].

Tabela 2.1: Premiações obtidas por alunos do Ceará na OBMEP, nas edições(2017-2019)

Premiações	2017	2018	2019
Medalha de ouro	28	32	31
Medalha de prata	119	82	110
Medalha de Bronze	260	258	318
Menção honrosa	2653	2579	3191

Fonte: OBMEP[03]

Nestas tabelas, os números indicativos das ocorrências dos dados chamam-se frequências e podem ser números absolutos ou relativos. Já as classes podem ser números reais (inteiros ou não), intervalos reais ou categorias de uma variável qualitativa. De modo geral, o processo de construção de uma distribuição de frequência passa por uma análise centrada na variável ou nas variáveis cujas observações deseja-se representar.

A literatura recomenda que o número de classes não seja menor que cinco nem maior que quinze [01]. E para definição dessa quantidade pode-se usar $k = 1 + 3,32 \log n$, chamada Fórmula de Sturges, ou $k = \sqrt{n}$. Nessas relações, K representa a quantidade de classes, e n o total de dados da amostra. Na tabela 2.2, que exemplificamos a seguir, optamos por k=6 que é raiz quadrada de n=36.

Tabela 2.2: Estatura em metro de alunos de Porteiras-Ce na EEMTI Aristarco Cardoso de em 2022.

j	Classes	F_j	F'_j	f_j	f'_j
1	$1,59 \vdash 1,62$	6	6	0,1667	0,1667
2	$1,62 \vdash 1,65$	8	14	0,2222	0,3889
3	$1,65 \vdash 1,68$	5	19	0,1389	0,5278
4	$1,68 \vdash 1,71$	6	25	0,1667	0,6945
5	$1,71 \vdash 1,74$	6	31	0,1667	0,8612
6	$1,74 \vdash 1,77$	5	36	0,1389	1
	Σ	36	-	1	-

Fonte: Ilustração do autor

A distribuição de frequência com intervalos de classe agrega bastante poder de síntese na representação de dados. E, em geral, é utilizada para representar variáveis contínuas, porque essas, na maioria dos casos, assumem muitos valores distintos o que torna inviável tomar cada valor como classe. Na ilustração, F_j, F'_j, f_j e f'_j, indicam, respectivamente, as frequências absoluta, acumulada, relativa e relativa acumulada da classe j a qual se refere. Em síntese, tais frequências denotam o número de observações registrado em cada classe ou até cada classe de forma absoluta ou relativa.

Gráficos

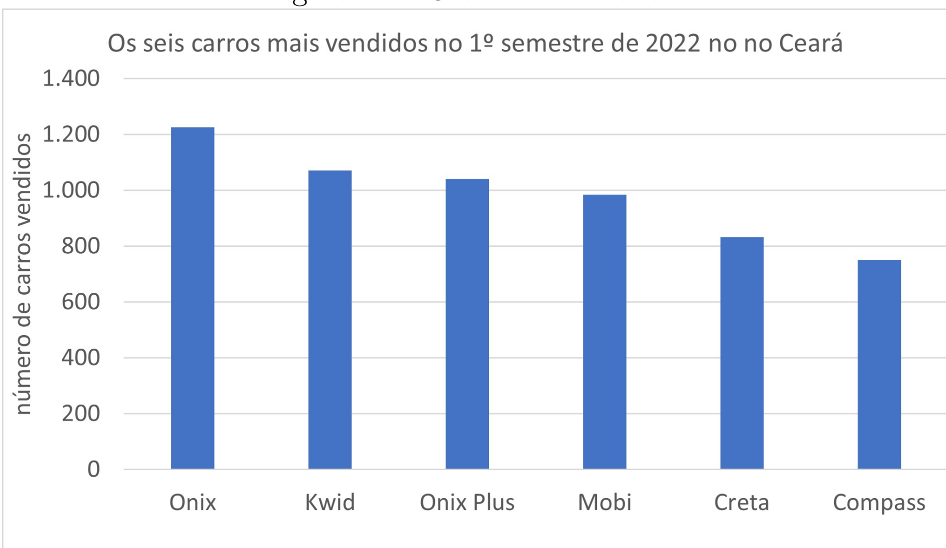
Os gráficos constituem recursos de representação de dados cuja finalidade é também compactar informações, possibilitando leitura e interpretação rápida. Todavia, sua aplicabilidade não substitui o uso das tabelas. Em essência, tais recursos cumpre com o propósito descrito, mas o fazem de formas distintas: Enquanto as tabelas seguem mais normatizações e oferecem maiores condições de análise [01]; Os gráficos, por sua vez, são mais diretos sem pormenorizar detalhes que fundamentem a leitura. Tais como as tabelas existem em variedade, cuja escolha deve se adequar ao tipo de variável e aos objetivos da representação.

Tipos de gráficos

De modo geral, os gráficos são construídos a partir de sistemas ortogonais cartesianos. E, na maioria das vezes, usa-se o sistema bidimensional como referencial em tais construções [01]. Em função de particularidades e|ou objetivos atribuídos para as diversas representações, os gráficos podem ser classificados nas seguintes categorias: **Diagramas**, **Estereogramas**, **Pictogramas** e **Cartogramas**. Dentre as quais, pelo maior uso e simplicidade de construção, destaca-se a de diagramas. Compõe tal categoria os gráficos de barras, gráficos de colunas, gráficos de linhas, gráficos de setores e Histogramas.

Gráfico de colunas

Figura 2.1: Gráfico de colunas



Fonte: Victor Ximenes[04]

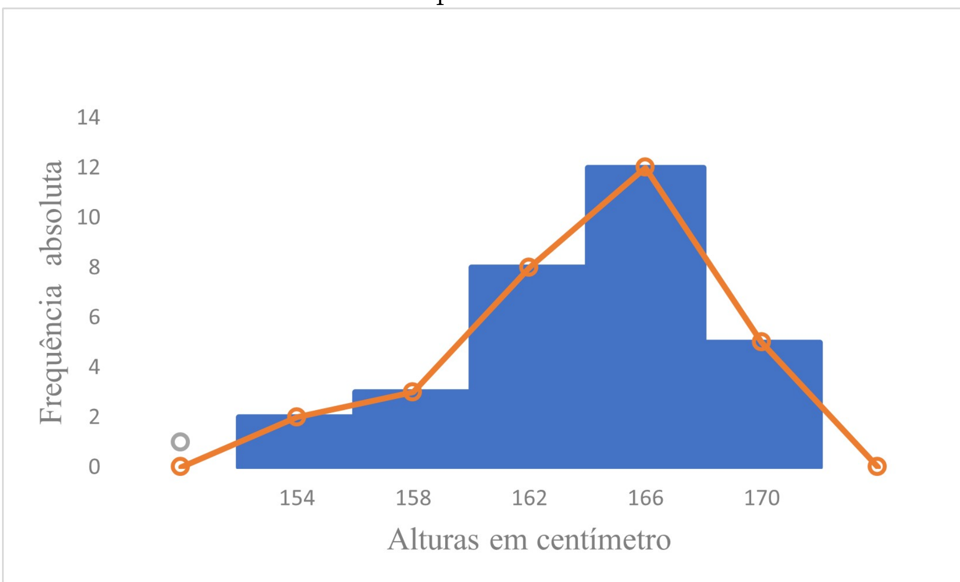
Conforme mostra a figura 2.1, nesse tipo de gráfico usa-se retângulos de mesma base dispostos na vertical. No que diz respeito a altura, essa é proporcional ao número que mensura a informação disposta na horizontal.

Dentro desse contexto, consideramos ainda a representação gráfica das distribuições de frequências. Dentre as quais, mencionamos o **Histograma** e o **polígono de frequência**. O primeiro consiste de retângulos justapostos, dispostos na vertical e centrados sobre os pontos médios das classes (caso sejam intervalos) [02]. A largura de um retângulo representa a amplitude desta classe, enquanto a altura indica a frequência absoluta (ou relativa). Já o segundo, consiste de segmentos que ligam os pontos de coordenadas iguais ao valor médio e a frequência de cada classe.

Na ilustração que segue, estamos tomando como exemplo uma amostra de estatura de 30 pessoas, e ainda considerando uma distribuição de frequência com cinco intervalos de classe de amplitude 4.

Histograma e Polígono de frequência

Figura 2.2: Histograma e Polígono de frequência
Estatura de 30 pessoas em centímetro



Fonte:Ilustração do autor

Do conceito de **polígono de frequência** decorre outros como **curva de frequência**, distribuições em forma de **cino**, em forma de **jota** ou de **u** como uma ideia tendencial do fenômeno estudado, aplicando um procedimento chamadado de polimento do polígono [02].

Além dos recursos citados, figuram neste contexto o **gráfico de dispersão**, também chamado de diagrama de pontos; o **diagrama de ramos e folhas** e o **diagrama em caixa** ou boxplot que se destacam por fundamentar técnicas simples de exploração de dados [01].

2.3 Medidas descritivas

Essencialmente, as medidas descritivas são valores obtidos em função dos dados coletados numa pesquisa e que contém as informações principais [01]. Disponíveis em grande variedade, possuem propriedades comuns e outras específicas que são buscadas ou preteridas a depender dos objetivos de algum estudo. Mediante particularidades e diferenças, tais medidas podem ser dispostas nos seguintes grupos: Medidas de Tendência Central; Medidas Separatrizes; Medidas de Dispersão e Medidas de Formato.

2.3.1 Medidas de tendência central

Compreendem tal grupo, medidas utilizadas para servir de referência na localização e descrição da posição de valores de uma variável numa dada distribuição. Por meio destas, avalia-se em torno de qual ponto se concentra os valores da variável. Compõe tal grupo as seguintes medidas : *Média aritmética*; *Mediana* e *Moda*.

Dada uma variável estatística X, cujos os n valores assumidos sejam : x_1; x_2; $x_3; \cdots ; x_n$ definiremos essas medidas como segue :

Média Aritmética

Definição 1. *Chama-se média aritmética simples dos n valores dados o valor indicado por $\overline{x}$, e obtido conforme expressão*

$$\overline{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n} \Leftrightarrow \overline{x} = \frac{\sum_{i=1}^{i=n} x_i}{n}.$$

Observamos que em algumas situações no cálculo de média aritmética, podemos ter valores repetidos ou valores com pesos distintos mediante condições dadas. Nesses casos, costuma-se designar tal medida por média aritmética ponderada e indicá-la por $\overline{x}_p$.

Definição 2. *Chama-se média aritmética ponderada dos n valores $x_1; x_2; x_3; \cdots; x_n$, com seus respectivos pesos $p_1; p_2; p_3; \cdots; p_n$, o valor indicado por $\overline{x}_p$ e obtido conforme expressão*

$$\overline{x}_p = \frac{x_1.p_1 + x_2.p_2 + x_3.p_3 + \cdots + x_n.p_n}{p_1 + p_2 + p_3 + \cdots + p_n} \Leftrightarrow \overline{x}_p = \frac{\sum_{i=1}^{i=n} x_i.p_i}{\sum_{i=1}^{i=n} p_i}.$$

De um modo geral, a média aritmética é uma medida descritiva que se destaca pela simplicidade do cálculo e por propriedades importantes tais como as que seguem com base em [01]: Ao somar uma constante c a todos os valores x_{is} de um conjunto com média $\overline{x}$, a média do conjunto passa a ser $\overline{x} + c$; Ao multiplicar por uma constante c todos os valores x_{is} de um conjunto com média $\overline{x}$, a média passa a ser $c\overline{x}$; A soma dos desvios de um conjunto de valores x_{is} em relação à média é nula $(\sum(x_i - \overline{x}) = 0)$; A soma dos quadrados dos desvios de um conjunto de valores x_{is} em relação a uma constante c $(\sum (x_i - c)^2)$ é mínima, quando c é a média do conjunto.

Tais propriedades, dentre outras razões, justificam o maior emprego dessa medida comparada a outras. Contudo, sua desvantagem é ser bastante influenciada por valores discrepantes [01]. Ou seja, a presença de um valor grande ou pequeno, em relação aos demais, altera, significativamente, a média do conjunto.

Mediana

Definição 3. *Dado um conjunto com n valores ordenados de uma variável X, entende-se por mediana, e indica-se por M_d, a medida descritiva que divide tal conjunto em duas partes com a mesma quantidade de valores.*

A partir da definição, subentende-se que a determinação de tal medida não envolve cálculo, bem como, sua existência não é influenciada por valores discrepantes. Entretanto, precisamos considerar duas situações na sua determinação:

Primeira : Se n é um número impar, então basta procurarmos o dado de posição central dada por $p = \dfrac{n+1}{2}$.

Segunda : Se n é um número par, então teremos dois dados centrais de posições dadas por $\dfrac{n}{2}$ e $\dfrac{n}{2} + 1$. Neste caso, a mediana é dada pela média aritmética entre tais dados. Observamos ainda, que, nessa situação, a mediana pode ou não coincidir com algum dos dados.

Moda

Definição 4. *Dado um conjunto com n valores de uma variável X, compreende-se por moda, e indica-se por M_o, o valor que aparece com mais frequência.*

Desta definição, decorre que a moda não é uma medida descritiva que exige cálculo. Além disso, que dado um conjunto de valores esse pode ou não conter uma moda. E caso contenha, essa pode não ser única [01]. Quando não existe, dizemos que essa amostra é amodal.

Convém considerar que estando os dados agrupados em tabelas de distribuição de frequências com intervalo de classe, a determinação de tais medidas passa por alteração.

E de modo geral, o processo baseia-se, sobretudo, na determinação dos centros das classes como ponto médio entre seus extremos. Maiores detalhes podem ser encontrados nas referências [01] e [02].

2.3.2 Medidas Separatrizes

Compõe tal grupo medidas que determinam posições dentro de variáveis ordenadas. E, assim, que possuem finalidade análoga a desempenhada pela Mediana(M_d). Enquanto essa divide um conjunto de dados ordenados em duas partes com a mesma quantidade de elementos, chama-se separatrizes medidas que dividem um conjunto ordenado em quatro ou mais partes com a mesma quantidade de elementos. Os principais representantes desse grupo são: *Quartis, Decis e Percentis.*

Quartis

Definição 5. *Chama-se quartis três medidas intituladas primeiro quartil*(Q_1)*, segundo quartil*(Q_2) *e terceiro quartil*(Q_3) *que dividem um conjunto ordenado de valores em quatro partes com a mesma quantidade de elementos.*

Decorre dessa definição que os quartis dividem um conjunto ordenado de valores de modo que 25% sejam menores que o primeiro quartil, 50% sejam menores que o segundo quartil e 75% sejam menores que o terceiro quartil. Raciocínio análogo se aplica a qualquer quantil [01].

Decis

Definição 6. *Chama-se decis nove medidas intituladas primeiro decil*(D_1)*, segundo decil*(D_2)*, ... e nono decil*(D_9) *que dividem um conjunto ordenado de valores em dez partes com a mesma quantidade de elementos.*

Percentis

Definição 7. *Chama-se percentis noventa e nove medidas intituladas primeiro percentil (P_1), segundo percentil (P_2), ... e nonagésimo nono percentil (P_{99}) que dividem um conjunto ordenado de valores em 100 partes com a mesma quantidade de elementos.*

2.3.3 Medidas de Dispersão

Neste grupo, estão medidas descritivas cuja finalidade é aferir o quanto os dados diferem entre si e, sobretudo, da média aritmética do conjunto que, em geral, é a medida de posição central mais utilizada como referência para situar qualquer dado numa distribuição.

Como exemplo da importância, podemos pensar numa situação clássica em que dois participantes de um concurso obtiveram a mesma média em relação a testes aplicados e se faz necessário encontrar um critério justo para o desempate. Nesta situação, o uso de medidas de dispersão podem apontar qual candidato teve desempenho mais regular, e portanto que é merecedor da vaga.
As principais representantes desse grupo são: *Amplitude total, Variância e Desvio Padrão.*

Considerando os n valores (x_i) não agrupados de uma variável X, definiremos tais medidas como segue :

Amplitude Total

Definição 8. *Amplitude total indicada por a_t é a diferença entre o maior(ES) e o menor (EI) valor assumidos pela variável dentre as observações.*

Nesta definição, (ES) indica extremo superior, e (EI) extremo inferior do conjunto de valores [01]. Convém observar que tal medida não tem eficácia para uma análise mais rigorosa, em virtude do seu cálculo levar em conta apenas os valores extremos.

Variância

Por envolver todos os valores de um conjunto assumidos por uma variável, essa medida revela-se mais eficiente e, por essa razão, é mais usada.

Definição 9. *Considerando os n valores (x_i) de uma variável cuja média aritmética no conjunto considerado é $\overline{x}$, chama-se de variância, e denota-se por S^2, a média dos quadrados dos desvios em relação a média. E calcula-se conforme expressão:*

$$S^2 = \frac{\sum(x_i - \overline{x})^2}{n-1}.$$

Sobre tal expressão, ao menos dois esclarecimentos faz-se necessário:

Primeiro: Que as diferenças $(x_i - \overline{x})$ constitui o que está sendo designado por desvio em relação à média. E que se utiliza os quadrados destes desvios, pois conforme propriedades da média apresentadas, tal soma é mínima e dessa forma pode-se se encontrar a variabilidade.

Segundo: Que na expressão dividi-se por $n-1$, embora seja n desvios, pelo fato de que sob esse formato a medida constitui um melhor estimador para variância populacional. Além disso, como a soma dos n desvios é nula, a escolha de $n-1$ desvios determina o último. Para efeito apenas descritivo (sem intuito de inferência), podemos substituir $(n-1)$ por n.

Desvio Padrão

Definido a partir da variância, tal medida se destaca por apresentar maior facilidade na interpretação dos resultados de variabilidade ou dispersão. Enquanto a variância apresenta um resultado com a unidade de medida da variável ao quadrado, o desvio padrão fornece informação na mesma unidade como sendo um valor da variável.

Definição 10. *Chama-se desvio padrão e indica-se por S a medida descritiva obtida como raiz quadrada da variância, conforme expressão*

$$S = \sqrt{\frac{\sum (x_i - \overline{x})^2}{n-1}}. \tag{1}$$

Observamos que para efeito de cálculo, com intuito de simplificar o processo, podemos usar a seguinte igualdade na obtenção da variância e do desvio padrão [02]:

$$\sum (x_i - \overline{x})^2 = \sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}.$$

Demonstração. $\sum (x_i - \overline{x})^2 = \sum (x_i^2 - 2x_i\overline{x} + \overline{x}^2)$. Aplicando propriedades do somatório, obtemos que

$$\sum (x_i - \overline{x})^2 = \sum x_i^2 - 2\overline{x}\sum x_i + \sum \overline{x}^2 = \sum x_i^2 - 2\overline{x}n\frac{\sum x_i}{n} + n\overline{x}^2.$$

Como

$$\sum x_i^2 - 2\overline{x}n\frac{\sum x_i}{n} + n\overline{x}^2 = \sum x_i^2 - 2n\overline{x}^2 + n\overline{x}^2 = \sum x_i^2 - n\overline{x}^2.$$

Visto que

$$\sum x_i^2 - n\overline{x}^2 = \sum x_i^2 - n\frac{\left(\sum x_i\right)^2}{n^2} = \sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}.$$

Conclui-se que

$$\sum (x_i - \overline{x})^2 = \sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}.$$

□

Observa-se ainda que, em algumas situações, a expressão do desvio padrão com percentual de variação em relação à média torna mais significativa a ideia do quanto os valores se dispersaram. Essa opção consiste em calcular o que se chama de coeficiente de variação $CV = \dfrac{s}{\overline{x}} \cdot 100\%$.

2.3.4 Medidas de Formato

Sobre esse último grupo, nos limitaremos a uma breve descrição da sua composição e importância. Tendo como principais representantes o **coeficiente de assimetria** (a_3), e o **coeficiente de curtose**(a_4), sua utilidade diz respeito a uma análise de como os valores se distribuem numa amostra ou população e como essas informações se relacionam a sua interpretação gráfica.

Nesse contexto, enquanto o coeficiente de assimetria busca informar se os valores se concentram mais no centro ou nas extremidades da curva; O coeficiente de curtose retrata o grau de achatamento dessa curva.

Como principais medidas de assimetria e de curtose, os coeficientes citados são obtidos a partir de momentos da distribuição.

Definição 11. *Chama-se de momentos de ordem r centrados numa constante a, as medidas descritivas de cunho geral que se prestam a diversos fins no estudo de uma distribuição e que são dadas por*

$$\frac{\sum (x_i - a)^r}{n}.$$

Os momentos mais relevantes são do tipo ordinários quando $a = 0$ ou centrados na média quando $a = \overline{x}$. No primeiro caso, podem ser denotados por m'_r; no segundo, por m_r[01]. Em ambos os casos, tem singular importância os momentos m'_1, m_2, m_3 e m_4, pois através desses podemos obter os coeficientes de assimetria (a_3) e de curtose (a_4), bem como, as principais medidas de tendência central e dispersão. Os coeficientes citados podem ser obtidos como segue:

$$a_3 = \frac{m_3}{m_2\sqrt{m_2}} \;\; e \;\; a_4 = \frac{m_4}{m_2^2}.$$

3 Probabilidade e variável aleatória

Neste capitulo, contemplaremos conceitos que, de maneira sucinta, deram amplitude de atuação a Estatística, bem como, contribuíram para que a ciência, em geral, pudesse evoluir no estudo de fenômenos não determinísticos. E que, hoje, sabemos serem tão diversos e importantes para compreensão e preservação da vida em seus vários aspectos.

3.1 Sobre probabilidade

Nesta seção, discorreremos sobre o conceito e algumas propriedades relacionadas ao cálculo de probabilidade. Assim o faremos, porque o desenvolvimento da Estatística Inferencial somente se fez possível com o advento da teoria sobre probabilidade. Como tantos outros ramos da matemática, a origem e o desenvolvimento deste envolve participação de diversos matemáticos e relatos passíveis de críticas ou que requerem maior aprofundamento histórico. Entretanto, como tal linha de pesquisa não é um dos focos deste trabalho, nos deteremos a breves informações a esse respeito.

Apesar de sua origem ser considerada do século XVII [01], com ações vinculadas a jogos de azar, e atribuídas aos matemáticos francês Pierre de Fermat e Blaise Pascal; Na obra Divina Comédia de Dante Alighieri, que viveu entre os seculos XIII e XIV, existe menção ao uso de probabilidade em jogos de dados [05]. Tal fato reforça a crença de uma origem bem mais atinga atrelada aos jogos, porém com participação de outros estudiosos atraídos pela natural importância que se atribuía aos jogos na cultura da época.

Contudo, sabe-se que foi somente no período que se estende do século XIX ao século

XX que a citada teoria se consolidou sob uma organização mais formal e axiomática. Desse processo vários outros Matemáticos de renome participaram, dentre eles, Laplace e Gauss.

3.1.1 Conceito de Probabilidade

A compreensão do que seja probabilidade, em qualquer nível de formalidade, passa pelo entendimento de conceitos iniciais como **experimento aleatório**, **espaço amostral** e **eventos**. Neste sentido, entende-se por experimento aleatório todo aquele que mesmo realizado sob as mesmas condições não nos permite prever seu resultado, embora conheçamos suas possibilidades. Por espaço amostral, o conjunto formado por todas as possibilidades de resultados para um experimento aleatório. E por evento, qualquer subconjunto de um espaço amostral. O qual pode ser simples, se é um conjunto unitário ou impossível, se é o conjunto vazio. E se dois eventos não podem ocorrer simultaneamente, esses são ditos mutuamente excludentes.

Mediante todas as considerações, que certamente podem ser enriquecidas com maiores detalhes históricos, aqui não contemplados, a compreensão do que seja probabilidade passou por níveis diferentes de formalidade até chegarmos a concepção atual.

De um modo geral, a pergunta que norteou a busca pelo conceito de probabilidade, foi também a que levou ao surgimento de toda uma teoria, e que ainda hoje pode ser vista como a razão da sua existência que é : **Como mensurar a possibilidade de ocorrência de um evento para um experimento aleatório?**

Na busca por tal resposta, e considerando que a Matemática seguiu e segue evoluindo em seus diversos momentos, quatro definições podem ser destacadas nesse processo evolutivo: **Definição Clássica**; **Definição Frequencial**; **Definição Subjetiva** e **Definição Axiomática**.

Em geral, a evolução de um conceito na Matemática não implica, necessariamente, na revogação de todas as premissas que formava sua concepção já existente. E assim ocorreu com a definição de probabilidade, como buscaremos evidenciar.

Definição 12. *(Definição Clássica) Seja E um experimento aleatório cujo o espaço amostral S é formado por $n(S)$ eventos simples equiprováveis. Dado A um evento desse espaço com $n(A)$ eventos simples, chama-se probabilidade de A o número $P(A)$ dado por*

$$P(A) = \frac{n(A)}{n(S)}.$$

Tal definição também costuma ser designada por probabilidade de Laplace, probabilidade a priori ou razão entre casos favoráveis e casos possíveis. Conforme consta em [05], foi a primeira formalização do conceito. Nota-se que em tal concepção a ideia de probabilidade está restrita a espaços enumeráveis finitos, onde cada evento simples tenha a mesma chance de ocorrer o que traduz a expressão equiprováveis.

Decorre também da sua interpretação, a expressão da probabilidade de um evento como soma das probabilidades dos eventos simples que o forma. Bem como, as propriedades: i) $0 \leq P(A) \leq 1$; ii) $P(S) = 1$ iii) $P(\varnothing) = 0$. Consequências imediatas por se tratar de uma razão onde o termo antecedente pode ter no mínimo o valor 0, e no máximo o valor do termo consequente.

Definição 13. *(Definição Frequencial) Seja E um experimento aleatório, e A um evento deste experimento cuja probabilidade de ocorrência $P(A)$ deseja-se aferir. Se n_A é o número de vezes em que o evento A ocorreu em n repetições do experimento E, tem-se*

$$P(A) = \lim_{n \to \infty} \frac{n_A}{n}.$$

Como o próprio nome sugere, nesta definição, a probabilidade é estimada pela frequência com a qual o evento desejado ocorre em relação a um número de repetições do experimento. Razão pela qual costuma também ser designada de probabilidade a posteriori. Ao contrário da clássica, esta definição não carrega os pressupostos de espaço finito e equiprovável, porém trás a imprecisão do quão grande necessita ser o número de repetições do experimento para se fazer a estimativa.

Semelhante a definição clássica, possibilita também, de modo geral, fazer a seguinte caracterização como propriedades da probabilidade : i) $0 \leq P(A) \leq 1$; ii) $P(S) = 1$ iii) $P(\varnothing) = 0$. Pois, dado um evento qualquer A, este pode não ocorrer ou pode ocorrer em todas as vezes que o experimento seja realizado, de onde se chega a caracterização posta.

Definição 14. *(Definição subjetiva) Consiste na expressão do grau de crença pessoal, próprio de um especialista, sobre a ocorrência de um evento em um experimento aleatório.*

Pode-se perceber que tal concepção de probabilidade encontra seu espaço nas lacunas que as outras definições não preenchem. A exemplo, como podemos mensurar a chance de duas pessoas casarem e a união durar mais que 10 anos? Nesta situação, não visualizamos as ideias de equiprobabilidade, nem da repetição de experimento.

Vemos, assim, que essa definição evidencia o aspecto de incerteza que é inerente ao processo de mensurar uma probabilidade. A qual se apresenta como a manifestação da incerteza que cada pessoa, mediante suas possibilidades ou condições, é capaz de atribuir a um experimento.

A busca por uma definição mais formal para probabilidade, sobretudo no que tange a espaços amostrais infinitos não enumeráveis, teve como uma das mais significativas contribuições a formulação proposta, no século XX, pelo matemático Russo Andrei Nikolaevich Kolmogorov [01]. Em cuja essência, reside no tratamento deste conceito como medida expressa por uma função que satisfaz algumas condições postas como axiomas.

Conhecida como definição axiomática, a conceituação proposta por Kolmogorov influenciou toda a teoria moderna sobre probabilidade. Sendo estruturada em torno de conceitos não tão difundidos em cursos de graduação, como o de σ-álgebra e outras ideias oriundas da Teoria da Medida, a presença de tal definição na maioria dos livros didáticos não contempla toda a sua formalidade.

Nesse sentido, embora algumas obras até empreguem o título de definição axiomática, nota-se, na verdade, uma busca por uma expressão mais intuitiva do que foi proposto por Kolmogorov. Podendo, em alguns casos, até mesmo serem interpretadas como um caso particular daquilo que de fato compreende a definição. De todo modo, é possível perceber que, em tese, essas "simplificações"tem como finalidade se ajustar ao nível de aprofundamento buscado em cada contexto de estudo.

No que segue, apresentaremos, brevemente, o conceito de σ-álgebra e a definição axiomática proposta por Kolmogorov como base em [06].

Definição 15. *Uma classe de subconjuntos de S, representada por $\mathcal{F}$, recebe o nome de σ-álgebra se satisfaz as seguintes propriedades:*

I) $S \in \mathcal{F}$;

II) *Se* $A \in \mathcal{F}$ *,então* $A^c \in \mathcal{F}$;

III) *Se* $A_i \in \mathcal{F}, i \geq 1$, *então,* $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$;

Essencialmente, essa definição diz que uma σ-álgebra de subconjuntos de S é uma classe de subconjuntos fechada em relação as operações de união, interseção e complementar.

Definição 16. *(Definição Axiomática) Uma função $\mathcal{P}$, definida na σ-álgebra $\mathcal{F}$ de subconjuntos de um espaço amostral S e com valores em $[0,1]$, é uma probabilidade se satisfaz os Axiomas de Kolmogorov:*

I) $P(S) = 1$;

II) *Para todo subconjunto* $A \in \mathcal{F}, P(A) \geq 0$;

III) *Para toda sequência* $A_1, A_2 \cdots \in \mathcal{F}$ *mutuamente excludentes*

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

Observamos confome [06] que a triade $(S, \mathcal{F}, \mathcal{P})$ recebe o nome de espaço de probabilidade. Além disso, que tal definição não estabelece como deve ser o cálculo de uma probabilidade. Que tal medida (função) deve ser estabelecida de acordo com as características do experimento aleatório. Observando ainda, que a mesma não refuta o teor das definições frequencial e clássica que podem ser vistas como casos particulares.

Apesar de apresentarmos a definição mais rigorosa de Kolmogorov, não a seguiremos nas demostrações e demais resultados que estão por vir nesse trabalho. Optamos por uma abordagem menos formal e mais intuitiva a exemplo do que propõe Morgado e outros autores [05] e que externamos a seguir.

Definição 17. *Seja S um espaço amostral (conjunto). Uma função P definida para todos os subconjuntos de S (chamados eventos) é uma probabilidade se satisfaz os seguintes axiomas:*

1. $0 \leq P(A) \leq 1$;

2. $P(S) = 1$;

3. *Se* $A \subset S$ *e* $B \subset S$, *com* $A \cap B = \varnothing$, *então* $P(A \cup B) = P(A) + P(B)$.

3.1.2 Teoremas sobre o cálculo de probabilidade

Em geral, todos os teoremas que tratam sobre o cálculo de probabilidade são consequências dos três axiomas contidos na definição 17. No que segue, procuramos evidenciar isso admitindo como conhecidos alguns conceitos e propriedades da linguagem de conjuntos para que possamos realizar algumas demonstrações a luz de concepções de autores como MORGADO e outros presentes em [05].

Observação: Denotaremos o evento impossível(aquele que não pode ocorrer) por $\varnothing$. Já para dizermos que os eventos A e B são mutuamente excludentes(não ocorrem ao mesmo tempo) indicaremos que $A \cap B = \varnothing$. E para denotar o complementar de um evento A, usaremos A^c.

Teorema 1. *Se $\varnothing$ é um evento impossível, então $P(\varnothing) = 0$.*

Demonstração. $S = S \cup \varnothing \Rightarrow P(S) = P(S \cup \varnothing)$.Pelos axiomas 2 e 3, temos que $P(S) = 1$ e $P(S \cup \varnothing) = P(S) + P(\varnothing)$. Assim, $P(S) = P(S \cup \varnothing) \Rightarrow P(\varnothing) = P(S) - P(S) = 0$ □

Teorema 2. *Se A^c é o evento complementar de A, então $P(A^c) = 1 - P(A)$.*

Demonstração. Pelo axioma 3, $A \cup A^c = S \Rightarrow P(A) + P(A^c) = P(S)$. Pelo axioma 2, $P(S) = 1. Logo, P(A) + P(A^c) = P(S) \Rightarrow P(A^c) = 1 - P(A)$. □

Teorema 3. *(Probabilidade da Soma) Sendo A e B dois eventos quaisquer, então $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.*

Demonstração. Dados A e B eventos quaisquer, podemos afirmar que :

$A = (A - B) \cup (A \cap B) \qquad e \qquad (A \cup B) = (A - B) \cup B$.

E pelo axioma 3, decorre que

$A = (A - B) \cup (A \cap B) \Rightarrow P(A) = P(A - B) + P(A \cap B)(1)$. E

$(A \cup B) = (A - B) \cup B \Rightarrow P(A \cup B) = P(A - B) + P(B)(2)$.

Fazendo $(2) - (1)$ membro a membro, obtemos $P(A \cup B) - P(A) = P(B) - P(A \cap B)$. □

Observação: A probabilidade da união de n eventos ocorre de modo análogo a extensão do principio da inclusão exclusão para n conjuntos[2]

$P(A_1 \cup A_2 \cup A_3 \cup \cdots A_n) = P(A_1) + P(A_2) + P(A_3) + \cdots + P(A_n) - P(A_1 \cap A_2) \cdots - P(A_1 \cap A_n) + \cdots + P(A_1 \cap A_2 \cap A_n) \cdots$

Definição 18. *Sejam A e B eventos tal que $P(A) > 0$. Denota-se por $P(B|A)$ o número $\dfrac{P(A \cap B)}{P(A)}$ chamado probabilidade condicional de B na certeza de A. Em que $P(B|A) = \dfrac{P(A \cap B)}{P(A)}$ implica que $P(A \cap B) = P(A)P(B|A)$.*

O entendimento do que seja probabilidade condicional relaciona-se, sobretudo, a situações nas quais precisamos realizar o produto de probabilidades. Em geral, quando o experimento em estudo envolve estágios ditos condicionados, em que a ocorrência de um evento pode alterar a probabilidade de outro ocorrer.

Teorema 4. *(Produto de Probabilidades)Se $A_1, A_2, A_3, \cdots A_n$ são eventos tal que $P(A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_n) > 0$, então $P(A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_{n-1})$.*

Conforme orientação presente em [05], faremos a demostração deste teorema por indução finita sobre n.

Demonstração. Para n =2 temos, pela definição de probabilidade condicional, que $P(A_1 \cap A_2) = P(A_1)P(A_2|A_1)$. Logo, é verdade. Suponha, por hipótese de indução, que seja verdade para algum $n > 2$. Isto é, que $P(A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_{n-1})$. Desejamos mostrar que vale para $n + 1$.

[2]Segundo o qual, a cardinalidade da união de n conjuntos é igual a soma das cardinalidades uma a uma, menos as cardinalidades das interseções duas a duas, mais as cardinalidades das interseções três a três, etc. Subtraindo interseções pares e somando interseções impares.

Pela propriedade associativa da interseção de conjuntos, temos que $A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_n \cap A_{n+1} = (A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_n) \cap A_{n+1}$.E assim, $P(A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_n \cap A_{n+1}) = P[(A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_n) \cap A_{n+1}]$. Usando o caso base, $n = 2$, temos que $P[(A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_n) \cap A_{n+1}] = P(A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_n)P(A_{n+1}|(A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_n)$. Usando a hipótese de indução , obtemos que $P[(A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_n) \cap A_{n+1}] = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_{n-1})P(A_{n+1}|(A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_n)$. Logo, vale para $n + 1$. E por indução finita, concluímos que vale $\forall n \in \mathbb{N}$. $\square$

Relaciona-se a este último teorema o conceito de eventos independentes. De um modo geral, dois ou mais eventos recebem essa designação quando a ocorrência de um não altera as chances de ocorrência dos demais.

Definição 19. *Dois eventos A e B tais que $P(A \cap B) = P(A)P(B)$ são ditos independentes.*

Observamos que para mais de dois eventos, a exemplo três A, B, C, a verificação de que sejam mutuamente independentes não se traduz apenas em saber se $P(A \cap B \cap C) = P(A)P(B)P(C)$. Faz-se necessário verificar que dois a dois valem igualdades análogas. Logo, teríamos $2^3 - 3 - 1 = 4$ condições a serem verificados nesse exemplo. De modo geral, para n eventos, o número de condições é dado por $2^n - n - 1$ [07]. O quê inclui verificações 2 a 2, 3 a 3 etc.

Para ilustrar alguns dos conceitos abordados até o momento, exploraremos a seguinte situação:

Exemplo 1. *Sabe-se que 95% das pessoas são portadoras do vírus varicela-zoster causador da catapora e também da doença infeciosa herpes-zoster, popularmente conhecida por cobreiro[08]. Suponha que um teste para detectar a presença deste vírus tenha sido* *desenvolvido com eficiência de 85%, quando a pessoa é portadora ; e de 80% quando a pessoa não é portadora do vírus. Qual o percentual de pessoas que testam positivo, e de fato tem o vírus?*

Nesta situação, observamos que o problema consiste em determinar uma probabilidade condicional: A probabilidade de uma pessoa ter o vírus na certeza do teste ter dado positivo. Definamos os seguintes eventos V: {a pessoa tem o vírus }; W:{a pessoa não tem o vírus }; Tp:{teste positivo}; Tn:{teste negativo}. Do enunciado, temos as seguintes probabilidades condicionais: P(Tp|V)= 85%; P(Tn|V)=15% ; P(Tn|W)= 80% e P(Tp|W)=20% .

Assim, em termos das notações definidas, desejamos calcular P(V|Tp). Usando a probabilidade condicional, temos que: P(V|Tp) $=\dfrac{P(V\cap Tp)}{P(Tp)}$.

Pelo teorema do produto de probabilidades, $P(V\cap Tp) = P(V)P(Tp|V)$. Como $P(V) = 0,95$ e $P(Tp|V) = 0,85$; segue que $P(V\cap Tp) = 0,95 \times 0,85 \approx 0,81$.

Para obtermos a $P(Tp)$, temos dois eventos disjuntos a considerar: $P(Tp\cap V)$ e $P(Tp\cap W)$. Logo, pela adição de probabilidade com eventos disjuntos, segue $P(Tp) = P(Tp\cap V) + P(Tp\cap W)$.

Usando mais uma vez o teorema do produto, vemos que $P(Tp\cap W) = P(W).P(Tp|W)$. Dados que $P(W) = 5\%$ e $P(Tp|W) = 20\%$) obtemos que $P(Tp\cap W) = 0,05\times 0,2 = 0,01$.

Assim , $P(Tp) = 0,81 + 0,01 = 0,82$. E $P(V|Tp) = \dfrac{0,81}{0,82} \approx 0,98$ ou 98%.

Observação: Situações com a explorada podem ser resolvidas com mais agilidade fazendo uso de um esquema gráfico chamado de *diagrama em árvore* [05].

Definição 20. *Sejam $A_1, A_2, A_3, \cdots, A_n$ uma coleção de eventos de um espaço amostral S, tais que $A_i\cap A_j = \varnothing, \forall i \neq j$, e que $A_1\cup A_2\cup A_3\cup\cdots\cup A_n = S$. Nestas condições, a coleção de eventos forma uma partição de S.*

O teorema seguinte nos mostra como calcular a probabilidade de um evento correr em função das probabilidades condicionais de outros, desde que algumas condições sejam atendidas.

Teorema 5. *(Teorema da probabilidade Total)Se* $A_1, A_2, A_3, \cdots, A_n$ *é uma partição de um espaço amostral* S*, tal que* $P(A_i) > 0, \forall i = 1, 2 \cdots n$*, então sendo* B *um outro evento de* S*, tem-se que*

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

Demonstração. Dado que $B \subset S$, então B possui interseção com vários eventos A_{is}. Como esses eventos formam uma partição de S, tais interseções são disjuntas. Logo, podemos escrever B da forma

$$B = (A_1 \cap B) \cup (A_2 \cap B) \cup (A_3 \cap B) \cup \cdots \cup (A_n \cap B).$$

Pelo teorema da soma de probabilidades, temos que $P(B) = P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B) + \cdots + P(A_n \cap B)$. Usando o teorema do produto de probabilidades ou definição de probabilidade condicional, obtemos que $P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \cdots + P(A_n)P(B|A_n)$. □

O próximo teorema se relaciona ao anterior, de forma "inversa"as ideias apresentadas pelo primeiro. Nele, usamos um dado evento do espaço para calcularmos a probabilidade de uma unidade da partição.

Teorema 6. *(Teorema de Bayes)Se* $A_1, A_2, A_3, \cdots, A_n$ *é uma partição de um espaço amostral* S*,* $P(A_i) > 0 \, \forall i = 1, 2 \cdots n$*, e* B *é um outro evento de* S*, tal que* $P(B) > 0$*, então tem-se* $\forall A_i$ *que*

$$P(A_i|B) = \frac{P(A_i)P(B|A_1)}{\sum_i^n P(A_i)P(B|A_i)}.$$

Demonstração. Dado que $P(B) > 0$, temos pela definição de probabilidade condicional que $P(A_i|B) = \dfrac{P(A_i \cap B)}{P(B)}$. Como $P(A_i) > 0$, temos que $P(A_i \cap B) = P(A_i)P(B|A_i)$. Assim, $P(A_i|B) = \dfrac{P(A_i)P(B|A_i)}{P(B)}$. Pelo teorema da probabilidade total, temos que $P(B) = \sum_{i=1}^{n} P(A_i)P(B|A_i)$. Logo, $P(A_i|B) = \dfrac{P(A_i)P(B|A_1)}{\sum_{i=1}^{n} P(A_i)P(B|A_i)}$. $\square$

Finalizaremos essa seção com apresentação e exploração de mais uma situação ilustrativa.

Exemplo 2. *A avaliação final do estágio probatório de professores numa rede estadual de ensino é realizada por uma comissão da qual participa um aluno como representante dos discentes na unidade de ensino. Numa dada escola onde 27,8%, 36,6% e 35,6% são, respectivamente, alunos do 3º, 2º e 1º ano, a escolha do aluno para compor tal comissão ocorrerá de forma aleatória. Sabendo que, respectivamente, 31,1%, 28,7% e 28,9% são percentuais de alunos indisciplinados dessas turmas, qual a probabilidade de um aluno indisciplinado fazer parte da comissão avaliadora? Uma vez que se confirme a escolha de um aluno indisciplinado, qual a probabilidade de que seja do 1ºAno?*

Inicialmente, observamos que as séries formam uma participação do conjunto dos alunos que é o espaço amostral desse experimento. Além disso, que o conjunto dos alunos indisciplinados é um evento com interseção com as séries. Logo, temos uma situação na qual podemos usar o Teorema da probabilidade total.

Para tanto, definamos os seguintes eventos: A_1:{o aluno escolhido é do primeiro ano}, A_2:{o aluno escolhido é do segundo ano}; A_3:{o aluno escolhido é do terceiro ano}; B:{o aluno escolhido é indisciplinado}.

A primeira pergunta nos pede $P(B)$. Pelo teorema citado,

$$P(B) = P(B \cap A_1) + P(B \cap A_2) + P(B \cap A_3).$$

Expressando cada parcela por meio da probabilidade condicional, temos

$$P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + P(A_3)P(B|A_3)$$

Como $P(A_1) = 0,356; P(B|A_1) = 0,289;$ $P(A_2) = 0,366; P(B|A_2) = 0,287$ e $P(A_3) = 0,278; P(B|A_3) = 0,331$, segue que

$$P(B) = 0,356 \times 0,289 + 0,366 \times 0,287 + 0,278 \times 0,331 \approx 0,294.$$

Na segunda pergunta, desejamos a probabilidade condicional $P(A_1|B)$. Como o evento A_1 é uma unidade da partição do espaço amostral, e o evento B tem probabilidade positiva, o cálculo dessa probabilidade condicional é uma aplicação do Teorema de Bayes. Logo,

$$P(A_1|B) = \frac{P(A_1 \cap B)}{P(B)} = \frac{P(A_1)P(B|A_1)}{P(B)} = \frac{0,356 \times 0,289}{0,294} \approx 0,35.$$

3.2 Variáveis aleatórias

Nesta parte discorremos sobre conceitos que, em linguagem simples, podemos dizer que fazem a ponte entre probabilidade e estatística. O primeiro deles é o de *variável aleatória*. No que segue, denotaremos variáveis aleatórias por letras maiúsculas(X, Y, Z, etc.) e os valores por elas assumidos por letras minúsculas correspondentes.

Definição 21. *Dado um espaço amostral S qualquer de um experimento aleatório, chama-se de variável aleatória uma função $X : S \to S_x \subset \mathbb{R}$. A qual associa cada elemento $s \in S$ a um $X(s) = x \in S_x$, formando assim um espaço amostral numérico e real S_x.*

Em essência, busca-se dizer que X deve ser uma função definida de S para $\mathbb{R}$ tal que para seus valores seja possível calcular probabilidade. Nesse sentido, ainda, ressaltamos que, embora intuitiva, a definição posta não contém toda formalidade e precisam que o conceito comporta. E que uma formulação mais completa pode ser encontrada em [06] fazendo uso da ideia de espaço de probabilidade.

Dada uma variável aleatória X, se S_x for um conjunto finito ou infinito enumerável, a variável é dita **Discreta**. Se não enumerável, quando S_x é um intervalo real, a variável é dita **Continua** [01]. Decorrem de tais conceitos, na sequência de explanação que segue, os de **Função de probabilidade** e **Função densidade de probabilidade**.

Em síntese, tais termos designam funções que associam os valores de uma variável aleatória X a suas probabilidades de ocorrência. Em se tratando de variável discreta, teremos a função de probabilidade. Caso seja contínua, a função densidade de probabilidade.

Na sequência, definiremos tais funções conforme referência [01], bem como, apresentaremos ilustrações.

Definição 22. *Seja X uma variável aleatória discreta e S_x o conjunto dos seus valores. A função de probabilidade denotada por $P(X = x)$, $P(x)$ ou simplesmente fp, será a função que associa a cada valor de X a sua probabilidade de ocorrência, desde que satisfaça as duas condições seguintes:*

i) $P(x) \geq 0, \forall x \in S_x$;

ii) $\sum_{S_x} P(x) = 1$;

Sobre tal definição e suas subsequentes derivações neste trabalho, observamos que S_x pode ser um conjunto infinito enumerável. E, assim, que para sermos mais formais deveríamos, em tese, expressar o item $ii)$ como soma infinita. Algo que poderá ser feito no decorrer desta dissertação sempre que julgarmos ser o mais adequado. A escolha por tal notação reflete uma visão de maior clareza que este autor atribui a mesma.

Exemplo 3. *Considere o experimento E que consiste em retirar de uma vez duas bolas de urna, na qual consta três bolas pretas e duas brancas de mesma massa e tamanho. Tomando por variável aleatória X o número de bolas pretas retiradas, determine a função de probabilidade [01].*

Denotando os eventos simples bolas pretas por P_1, P_2, P_3; e os eventos simples bolas brancas por B_1e B_2, representaremos o espaço amostral de E por

$$S = \{P_1B_1, P_1B_2, P_2B_1, P_2B_2, P_3B_1, P_3B_2, P_1P_2, P_1P_3, P_2P_3, B_1B_2\}.$$

Aplicando a variável X sobre S, obtemos o espaço amostral numérico e enumerável $S_x = \{0, 1, 2\}$. Por isso X recebe a classificação de variável discreta.

Buscando uma representação para a função de probabilidade, podemos escolher uma dentre as formas tabular, gráfica ou analítica. Nesta ilustração, nos deteremos a forma analítica. Para tanto, calcularemos de início, a probabilidade de ocorrência de $X = 1$. Dessa forma, teremos uma melhor compreensão do cálculo e podemos fazer a sua generalização para um valor x qualquer de X, obtendo assim a forma analítica[01].

Seguindo conforme exposto, calcularemos

$$P(X = 1) = P(P_1B_1) + P(P_1B_2) + P(P_2B_1) + P(P_2B_2) + P(P_3B_1) + P(P_3B_2).$$

Tal expressão se justifica pelo fato de S ser composto por eventos que são dois a dois mutuamente excludentes. Além disso, dada a sua construção, temos que S é um espaço equiprovável. Logo,

$$P(X = 1) = \frac{1}{10} + \frac{1}{10} + \frac{1}{10} + \frac{1}{10} + \frac{1}{10} + \frac{1}{10} = \frac{6}{10}.$$

Considerando a observação de S como espaço enumerável, finito e equiprovável, podemos calcular essa mesma probabilidade pela forma clássica. Usando combinação simples para fazer a contagem dos casos possíveis e dos favoráveis, obtemos que:

$$P(X=1)=\frac{C_3^1C_2^1}{C_5^2}=\frac{6}{10}.$$

Generalizando esse último raciocínio, obtemos a representação analítica da função de probabilidade. Isto é, uma vez escolhidas x bolas pretas dentre as 3 possíveis, então serão escolhidas $2-x$ bolas brancas, dentre as duas possíveis. Portanto, $P(X=x)$ ou simplesmente $P(x)$ é dada por :

$$P(x)=\frac{C_3^xC_2^{2-x}}{C_5^2},\quad \text{com}\quad x\in S_x=\{0,1,2\}.$$

Em geral, quando dispomos de informações que nos permitem conhecer a associação dos possíveis valores de uma variável aleatória a suas respectivas probabilidades de ocorrência, seja por recurso analítico, gráfico ou tabular, dizemos que conhecemos a sua **Distribuição de Probabilidade**.

Contudo, para sabermos caracterizar, reconhecer e diferenciar distribuições de probabilidades, precisamos nos apropriar de outros conceitos que estão relacionados. Neste sentido, seguiremos com outras definições ainda com base em [01].

Definição 23. *Se X é uma variável aleatória discreta, cujo o espaço amostral é S_x, denota-se por $F(x)$, $P(X\leq x)$ ou fda a função de probabilidade Acumulada ou função de distribuição, definida por*

$$F(x)=P(X\leq x)=\sum_{t\leq x}P(x=t).$$

De modo conciso, essa expressão diz que $F(x)$ representa a probabilidade de ocorrência de um valor menor do que ou igual a x. A qual se expressa como sendo uma soma de probabilidades.

No que diz respeito a variáveis continuas, falaremos de funções com papéis análogos. Porém, dada a natureza distinta das variáveis, as operações com as mesmas não ocorrem de forma tão direta. Fazendo-se necessário a conexão com outros conceitos matemáticos.

Definição 24. *Seja X uma variável aleatória contínua e S_x o seu espaço amostral. Uma função f associada à variável X é denominada função densidade de probabilidade (fdp) se satisfizer as duas condições a seguir:*

i) $f(x) \geq 0, \forall x \in S_x$;

ii) $\int_{S_x} f(x)dx = 1$.

Observamos que S_x, no contexto contínuo, denota um intervalo real, que, em tese, pode se ser de $-\infty$ a $+\infty$. E, assim, que a definição posta, bem com outras que dela descendem, sob uma abordagem mais formal, comportam integrais impróprias. Tal como se observa no Guidorizzi [09] ao falar de aplicações do cálculo. Onde se expressa que $f(x)$ precisa ser definida $\forall x \in \mathbb{R}$, ser integrável em todo intervalo fechado $[a, b]$, com $a < b$, e atender as condições $i)$ e $ii)$ da definição 24.

Por fim, que embora este autor tenha feito opção por tal notação por conferir a mesma mais simplicidade, em momentos que sejam necessários maior formalidade, ocorrerá as devidas alterações e considerações sem ônus a compreensão do leitor.

Essencialmente, essa definição diz que para uma função ter a classificação dada ela não pode ser negativa para $x \in S_x$. E que a área abaixo do gráfico da função no intervalo que representa S_x deve ser igual a 1.

Observamos também, que no intuito de verificarmos se uma dada função cumpre com as condições apresentadas na definição 24, precisamos nos valermos de procedimentos algébricos/analíticos e gráficos. Pois, ao contrário das variáveis discretas, não fará sentido a construção de uma tabela para representar a distribuição de probabilidade.

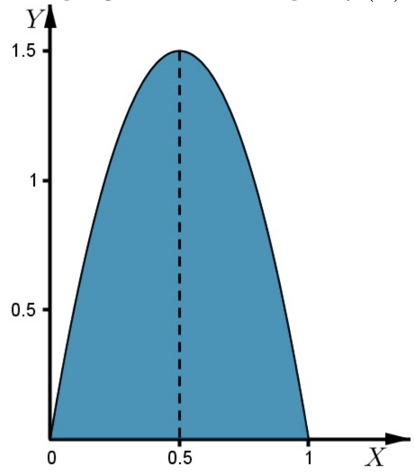
A fim de tornar mais claras as últimas afirmações, analisaremos como ilustração a seguinte função.

Exemplo 4. *Seja a função $f(x) = -6x^2+6x$ definida no intervalo $S_x = [0, 1]$. Verifique se a mesma é uma função densidade de probabilidade. Exemplo presente em [01].*

Ressaltamos que nosso intuito é verificar se as duas condições da definição 24 são atendidas pela função $f(x) = -6x^2 + 6x$. Neste sentido, observando que a mesma é uma função quadrática. Logo, que seu gráfico é uma parábola. Para esboçá-lo, basta observarmos o valor que a função assume nos extremos do intervalo em questão $(f(0) = f(1) = 0)$, as coordenadas do vértice, $V = \left(\frac{1}{2}, \frac{3}{2}\right)$, e vermos que a concavidade é para baixo, pois o coeficiente a é negativo($a = -6$).

Convém observar que, de modo geral, fazemos uso das ferramentas do cálculo para obtenção de informações necessárias ao esboço de gráficos. E que não foi essa a opção aqui adotada devido a simplicidade da função ilustrada.

Figura 3.1: Esboço gráfico da função $f(x) = -6x^2 + 6x$



Fonte:Ilustração do autor

A partir desse esboço, constatamos que a primeira condição da definição 24 é atendida, pois para $0 \leq x \leq 1$, temos $f(x) \geq 0$. Para concluir a análise, resta verificar se área destacada sob o gráfico no mesmo intervalo é igual a 1. Para tanto, calculamos a integral da função no citado intervalo.

$$\begin{aligned}
\int_0^1 (-6x^2 + 6x)dx &= -6\int_0^1 x^2dx + 6\int_0^1 xdx = -6\left[\frac{x^3}{3}\right]_0^1 + 6\left[\frac{x^2}{2}\right]_0^1 \\
\int_0^1 (-6x^2 + 6x)dx &= -6\left[\frac{1}{3}\right] + 6\left[\frac{1}{2}\right] = 1.
\end{aligned}$$

Segue do resultado, a verificação da segunda condição posta na definição 24.

A expressão da função de distribuição ou de probabilidade acumulada para variáveis continuas, traz a integração como expressão natural da soma para valores contínuos, conforme mostra a definição.

Definição 25. *Se X é uma variável aleatória continua, cujo o espaço amostral é S_x, denota-se por $F(x)$, $P(X \leq x)$ ou fda a função de probabilidade acumulada ou função* *distribuição, definida por*

$$F(x) = P(X \leq x) = \int_a^x f(t)dt \qquad com \qquad S_x = [a, b].$$

Como a função distribuição $F(x)$ é dada por uma integral da função densidade $f(x)$, a qual é integrável em qualquer intervalo fechado por definição, resulta que $F(x)$ é contínua, e que $F'(x) = f(x), \forall$x em que f for contínua[09]. Decorre que :

$$P(a \leq x \leq b) = \int_a^b f(x)dx = F(b) - F(a).$$

Uma outra forma de chegarmos a tal conclusão é trabalhando diretamente com a função distribuição.

Demonstração. Como $F(b) = P(X \leq b) = \int_{-\infty}^{b} f(x)dx$, e $F(a) = P(X \leq a) = \int_{-\infty}^{a} f(x)dx$. Sendo $a < b$, temos que $\int_{-\infty}^{b} f(x)dx = \int_{-\infty}^{a} f(x)dx + \int_a^b f(x)dx$. Assim, $F(b) - F(a) = \int_{-\infty}^{b} f(x)dx - \int_{-\infty}^{a} f(x)dx = \int_{-\infty}^{a} f(x)dx + \int_a^b f(x)dx - \int_{-\infty}^{a} f(x)dx = \int_a^b f(x)dx$. $\square$

Observarmos, ainda, que dá interpretação de integral definida como área, resulta que $P(a \leq x \leq b) = P(a < x \leq b) = P(a \leq x < b) = P(a < x < b)$. Pois, $\int_d^d f(x)dx = 0$, qualquer que seja f e d.

Partindo da compreensão de que os espaços amostrais de variáveis aleatórias são sempre conjuntos numéricos, ressaltamos que as distribuições de probabilidades também são caracterizadas por medidas descritivas[01]. No que segue, buscamos contemplar as principais informações a esse respeito.

Medidas Descritivas para Variáveis aleatórias

Essencialmente, essas medidas são números que caracterizam distribuições de probabilidades teóricas. Assumindo, em alguns modelos probabilísticos, a função de parâmetros.

Definição 26. *Dada uma variável aleatória discreta X, denota-se por E(X) ou* μ *o valor esperado ou média dos valores assumidos por X ponderada pelas respectivas probabilidades de ocorrência. E se expressa por* $E(X) = \sum_{S_x} x_i P(x_i)$*.*

Definição 27. *Sendo* X *uma variável discreta, denota-se por* $V(X)$ *ou* σ^2 *a média dos quadrados dos desvios em relação ao valor esperado, ponderada pelas probabilidades de ocorrência. E se expressa por* $V(X) = \sum_{S_x}(x_i - E(X))^2 P(x_i)$*.*

Reiteramos o sentido da variância já comentado nesse trabalho: Sua importância é aferir o grau médio de dispersão dos valores da variável em relação a média (valor esperado).

Observamos também que, de modo geral, a expressão que defini a variância pode ser reescrita da seguinte forma: $V(X) = E(X^2) - (E(X))^2$.

Demonstração. Seja X uma variável aleatória discreta. Temos por definição que

$$E(X) = \sum x_1 P(x_i) \quad \text{e que} \quad V(X) = E(X - E(X))^2$$

, decorrendo que $V(X) = \sum (x_i - E(X))^2 P(x_i)$. Segue que

$$V(X) = \sum ({x_i}^2 - 2x_i E(X) + (E(X))^2) P(x_i).$$

Pelas propriedades dos somatórios, e observando que $E(X)$ é um número, temos que

$$V(X) = \sum {x_i}^2 P(x_i) - 2E(X) \sum x_i P(x_i) + (E(X))^2 \sum P(x_i).$$

Como $\sum P(x_i) = 1$, temos que $V(X) = E(X^2) - 2(E(X))^2 + (E(X))^2$. Logo,

$$V(X) = E(X^2) - (E(X))^2.$$

. $\square$

Definição 28. *Dada X uma variável aleatória continua, denota-se por $E(X)$ ou μ o valor esperado de X que se expressa por meio da seguinte integral*

$$E(X) = \int_{S_x} x f(x) dx.$$

Nota-se que essa definição, bem como a seguinte e outras do contexto, trás em sua essência a extensão do raciocínio presente nas definições feitas no contexto das variáveis discretas. Sendo que para as variáveis contínuas $f(x)dx$ figura como aproximação da probabilidade de X assumir valor no intervalo infinitesimal de extremos x e $x+dx$ [09].

Definição 29. *Denota-se por $V(X)$ ou σ^2 a variância de uma variável aleatória contínua que se obtém por meio da seguinte integral.*

$$V(X) = \int_{S_x} [x - E(X)\ ]^2 f(x) dx.$$

A expressão posta na definição pode ser apresentada sob a forma

$$V(X) = \int_{S_x} x^2 f(x) dx - [E(X)]^2.$$

Demonstração. Como $[x - E(X)]^2 = x^2 - 2xE(X) + [E(X)]^2$, temos pela definição que $V(X) = \int_{s_x} [x^2 - 2xE(X) + [E(X)]^2] f(x) dx$. Como a integral da soma é igual a soma das integrais das parcelas, segue que $V(X) = \int_{S_x} x^2 f(x) dx - 2E(X) \int_{S_x} x f(x) dx +$

$[E(X)]^2 \int_{S_x} f(x)dx$. Das definições de função densidade e valor esperado, temos que $\int_{S_x} f(x)dx = 1$ e que $\int_{S_x} xf(x)dx = E(X)$. Fazendo as substituições, obtemos o resultado desejado. □

Propriedades do Valor Esperado e da Variância

Enfatizamos que o valor esperado e a variância possuem propriedades análogas as que possui a média aritmética e variância de dados numéricos de uma distribuição de frequência. Dentre as quais, destacamos as seguintes, onde X e Y são variáveis e c é uma constante.

I) $E(X \pm Y) = E(X) \pm E(Y)$;

II) $E(XY) = E(X)E(Y)$, desde que Xe Y sejam variáveis independentes;

III) $E(cX) = cE(X)$ e $E(X + c) = E(X) + c$;

IV) $V(X \pm Y) = V(X) + V(Y)$, se X e Y forem variáveis independentes;

V) $V(X + c) = V(X)$ e $V(cX) = c^2V(X)$.

Obs.: Essas propriedades são extensíveis a mais de duas variáveis. As demonstrações omitidas aqui podem ser vistas em [07].

Funções de Variáveis Aleatórias

Em diversas situações, faz-se necessário envolver mais de uma variável no estudo de um experimento aleatório. Neste contexto, tem-se como premissa geral, conforme Meyer [07], que se X é uma variável aleatória, então $Y = h(X)$ também é variável aleatória. De onde se justifica precisarmos também conhecer e caracterizar a distribuição dessa variável Y através da obtenção do valor esperado, variância etc.

Dentro deste contexto, destacamos que: Em principio, podemos caracterizar a distribuição de probabilidade de Y a partir do conhecimento da distribuição de X; Os processos podem se diferenciar a depender se ambas são discretas, se X é contínua e Y é discreta ou se ambas são contínuas. Contemplaremos mais detalhes a esse respeito à medida que se tornarem necessários a realização deste trabalho.

Momentos de Variáveis Aleatórias

Como já conceituado para distribuições de frequências, momentos no contexto das variáveis aleatórias também podem ser compreendidos como valores ou medidas descritivas de caráter geral. A partir destas, torna-se possível estudar e caracterizar diversos aspectos de uma distribuição de probabilidade como dispersão, assimetria e outras informações que se relacionam ao seu formato.

Na literatura é consensual o entendimento de que tão melhor seremos capazes de conhecer a distribuição de uma variável, quanto mais soubermos sobre seus momentos [01]. Neste sentido, é de bastante utilidade conhecer o momento genérico de uma variável ou o r-ésimo momento como denotaremos nas definições a seguir.

Definição 30. *Seja X uma variável aleatória. Chama-se de momento centrado na origem de ordem r ou r-ésimo momento de X o valor denotado e obtido por* $\mu'_r = E(X^r)$.

De início, observamos que o termo centrado na origem faz distinção a uma classe de momentos ditos centrados na média, cuja indicação e cálculo se dá por $\mu_r = E(X - E(X))^r$. Dessa classe faz parte a variância, cuja expressão pode ser vista como uma função de momentos centrados na origem, bem como, os demais momentos centrados na média. Esse fato já foi evidenciado, quando partindo da definição, mostramos que $V(X) = E(X^2) - [E(X)]^2$. E assim, que $V(X) = \mu'_2 - {\mu'_1}^2$. Algo que pode ser feito para qualquer outro momento centrado na média.

Depreende-se da definição, que sendo X uma variável discreta, temos $\mu'_r = \sum_{s_x} x_i^r P(x_i)$. Caso seja contínua, $\mu'_r = \int_{S_x} x^r f(x)dx$. Além disso, em ambos os casos os momentos são, essencialmente, valores esperados de potências de uma variável aleatória. E no caso dos momentos centrados na média, valores esperados das potências dos desvios.

Ainda sobre a expressão momentos centrados na origem, antecipamos que a mesma faz referência ao emprego da *Função Geratriz de Momentos* (fgm) de uma variável aleatória. Conceito que discutiremos a seguir.

Função Geratriz de Momentos

No estudo de uma variável aleatória, objetiva-se, de modo geral, obter o máximo de informações a respeito de sua distribuição de probabilidade. E, na maioria das vezes, o ponto de partida para o processo investigativo são as funções definidas até aqui(fp, fdp e fda). Pois, além de serem as principais representações, assumem um papel de identidade para as distribuições.

Contudo, outras funções, potencialmente importantes, podem ser definidas para uma variável aleatória. E que, a depender da variável e da situação a ser contemplada, podem se revelar até mais significativas para o estudo de uma distribuição. É com essa concepção que definiremos a função geratriz de momentos, bem como, citaremos algumas de suas propriedades.

Definição 31. *Seja X uma variável aleatória. Designa-se por Função Geratriz de Momentos(fgm) dessa variável a que se expressa por $M_X(t) = E(e^{tX})$, com $t \in \mathbb{R}$.*

Essa definição contém, em sua essência, uma condição de existência para a fgm. Pois, sendo tal função dada pelo valor esperado de e^{tX}, o mesmo pode não existir no caso discreto em que a série não convirja, e no caso contínuo, em que a integral imprópria também seja divergente para algum valor de t [07].

Afim de tornar mas claro essas ideias, esse é um dos momentos nos quais é mais adequado expressar $E(X) = \sum_{i=1}^{\infty} x_i P(x_i)$, no caso discreto ; e $E(X) = \int_{-\infty}^{+\infty} x f(x) dx$, no caso contínuo. Pois, como já mencionamos, em tese, X pode assumir infinitos valores enumeráveis, caso seja discreta; ou qualquer valor real, se for contínua.

Segue do exposto, as expressões que definem a fgm quando X é discreta ou contínua, respectivamente: $M_X(t) = \sum_{i=1}^{\infty} e^{tx_i}P(x_i)$ e $M_X(t) = \int_{-\infty}^{+\infty} e^{tX}f(x)dx$.

Dentre as propriedades da fgm, citaremos as seguintes:

i) Se X é uma variável aleatória, então a derivada r-ésima da $M_X(t)$ aplicada em $t = 0$ gera $E(X^r)$;

ii) Se X e Y são variáveis aleatórias tal que $M_X(t) = M_Y(t)$, então essas variáveis tem a mesma distribuição de probabilidade;

iii) Se X e Y são variáveis independentes, então $Z = X+Y$ tem $M_Z(t) = M_X(t)M_Y(t)$;

No decorrer desta dissertação, buscaremos ao menos justificar a propriedade $i)$, a exemplo do que faz Meyer [07].

Tendo em vista a possibilidade de mencionarmos o conceito de vetor aleatório no decorrer deste trabalho, julgamos coerente ao menos darmos uma definição intuitiva a esse respeito.

Em algumas situações, a ocorrência de um resultado pode depender de duas ou mais variáveis aleatórias relacionadas ao mesmo espaço amostral. Nestes casos, em geral, deve-se levar em consideração todas as variáveis envolvidas para estimar a probabilidade de ocorrência deste resultado. Também dita probabilidade conjunta das variáveis.

Definição 32. *Sejam $X_1, X_2, X_3, \cdots, X_n$ variáveis aleatórias definidas num mesmo espaço amostral. Chama-se vetor aleatório n dimensional ou n variado o vetor $X = (X_1, X_2, X_3, \cdots X_n)$.*

Por hora, finalizamos essa seção ressaltando que é vasta a teoria sobre variável aleatória e que a exposição feita até aqui não teve como intuito contemplar todo esse universo. Mas, mencionar partes que sejam mais pertinentes a este trabalho.

4 Modelos Probabilísticos

Neste capítulo, falaremos, de modo geral, sobre modelos de distribuição de probabilidade. Sem nos aprofundamos nesta temática, almejamos, por meio da revisão de alguns modelos, nos apropriarmos de ideias que serão estendidas no decorrer desta dissertação ou serviram como base para o melhor entendimento de outras.

Modelos de Probabilidade

Essencialmente, a Estatística trata do estudo de comportamentos coletivos e aleatórios buscando compreender como determinadas características (as variáveis), se apresentam numa dada população. Com este propósito, estudar tal população tornar-se equivalente a definir adequadamente uma variável aleatória e saber como os seus valores se comportam [01].

Para tanto, busca-se, em geral, dispor de relação ou lei matemática, que sob dadas suposições, permita mensurar as chances da variável assumir valores, bem como, caracterizá-la em termos de medidas descritivas. Por **modelos probabilísticos**, designam-se tais expressões matemáticas. Conhecer e saber caracterizar a associação dos valores a suas probabilidades de ocorrências, consiste em conhecer a **distribuição de probabilidade da variável**.

Dada a diversidade de variáveis aleatórias que possam permear diversos estudos, a tarefa de estudá-las partindo sempre do início para obter as suas distribuições de probabilidades é algo bastante complexo e moroso. Por isso, a partir do estudo de algumas variáveis aleatórias chegou-se ao que se chama de **Modelos de probabilidades**. Que são modelos genéricos que sob dadas suposições descrevem o comportamento de variáveis reais.

Para o que segue, observamos que é usual a notação $X \sim Nome(p, s, t \cdots)$ para dizer que a avariável X segue distribuição de probabilidade definida pelo modelo $Nome$, nos parâmetros $(p, s, t \cdots)$ que pode ser um dois ou mais.

4.1 Modelos Discretos

Dentre os vários modelos que existem para esse tipo de variável, destacam-se entre os mais conhecidos, os seguintes: **Distribuição Uniforme Discreta; Distribuição de Bernoulli; Distribuição Binomial; Distribuição Hipergeométrica e Distribuição de Poisson**.

No que segue, descreveremos brevemente alguns desses modelos. Reservando-se o modelo de Poisson para o último capítulo.

Distribuição Uniforme Discreta

Definição 33. *Dada uma variável aleatória discreta X, com valores* $x_1, x_2, x_3, ..., x_n$*, afirma-se que essa variável possui distribuição amostral uniforme discreta, que pode ser indicada por* $X \sim U(n)$*, se, somente se,* $P(X = x_i) = \frac{1}{n}$*, para* $i = 1, 2, 3, \cdots n$*.*

Essencialmente, esse modelo traduz a concepção da definição clássica de probabilidade com os pressupostos de espaço finito e equiprovável. De outro modo, está dito que $P(x_1) = P(x_2) = \cdots = P(x_n) = p$. Como $P(x_1) + P(x_2) + \cdots + P(x_n) = 1$, pelos axiomas 3) e 2) da definição 17,$1 = np \Rightarrow p = \frac{1}{n}$.

Tomando X como sendo a pontuação obtida no lançamento de um dado honesto, temos uma ilustração clássica [10] de uma varável com tal distribuição, onde $P(X = x_i) = \frac{1}{6}, \forall x_i \in S_x = \{1, 2, 3, 4, 5, 6\}$.

Segundo Meyer [07], na maioria das vezes em que se define uma variável aleatória, não se deve dar ênfase a natureza funcional desta, mas sim aos seus valores. Contrariando essa premissa, observamos que na ilustração X é uma função identidade. Algo que ocorre em varias situações quando o espaço amostral do experimento aleatório em questão já tem a característica numérica desejada.

Neste modelo, obtemos E(X), V(X) e F(x) como segue :

i) $E(X) = \sum_{(x_1 \in S_x)} P(x_i)x_1 = \frac{1}{n} \sum_{(x_1 \in S_x)} x_i;$

ii) $V(X) = E(X^2) - (E(X))^2 = \frac{1}{n} \sum_{(x_1 \in S_x)} .x_i^2 - \left(\frac{1}{n} \sum_{(x_1 \in S_x)} x_i \right)^2.$

Nota-se que $E(X)$ coincide com a média aritmética simples dos valores assumidos por X. Bem como, $V(X)$ é a variância estatística normal destes valores. Como de fato, mostra os dados na ilustração:

- $E(X) = \frac{1}{6} + \frac{2}{6} + \frac{3}{6} + \frac{4}{6} + \frac{5}{6} + \frac{6}{6} = \frac{21}{6} = 3,5;$
- $V(X) = \frac{1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2}{6} = \frac{91}{6} - 3,5^2 = 15,2 - 12,3 = 2,9.$

Para obtermos a função de distribuição acumulada (fda)deste modelo, procedemos conforme a definição, segundo a qual

$$F(x) = P(X \leq x) = \sum_{t \leq x} P(X = t).$$

Essencialmente, podemos entender essa função como aplicação do terceiro axioma da definição17. Segundo o qual a probabilidade da união de eventos disjuntos é dada pela soma das probabilidades destes eventos.

Na situação ilustrada, observe que $F(3) = P(X \leq 3) = P(\{X = 1\} \cup \{X = 2\} \cup \{X = 3\}) = P\{X = 1\} + P\{X = 2\} + P\{X = 3\} = \frac{3}{6}$. Assim procedendo, obteríamos todos os valores de F(x) que poderiam ser dispostos numa tabela ou num gráfico que mostraria uma função descontinua $\forall x \notin S_x$.

iii) De modo analítico, para esse modelo, temos $F(x) = P(X \leq x) = \sum_{t \leq x} P(X = t) = \frac{k}{n}$, onde k é o número de valores t menores que ou iguais a x.

Distribuição de Bernoulli

Esse modelo carrega o nome de seu idealizador, o matemático suíço Jacob Bernoulli, que segundo a fonte [01], o desenvolveu durante o século XVII. De modo geral, tal modelo se ajusta bem a experimentos que possam ser vistos sob a perspectiva de dois possíveis resultados que costumam ser designados por *sucesso*, e *fracasso*. Esses termos identificam os resultados desejado e preterido, respectivamente, no experimento.

Dado um experimento com tal característica, para fazer a estruturação do modelo definimos uma variável X que assume o valor 1, em caso de sucesso, e 0, se ocorrer fracasso em uma realização do experimento, também dito ensaio de Bernoulli.

Definição 34. *Dada uma variável aleatória discreta X que assume apenas os valores 1 ou 0, com probabilidade p e 1-p, respectivamente, dizemos que a mesma possui distribuição de Bernoulli, e denotamos por $X \sim Ber(p)$.*

Aplicando-se as definições para valor esperado E(X), e para variância V(X), obtemos :

i) $E(X) = \sum_{(x_1 \in S_x)} P(x_i)x_1 = p \cdot 1 + (1-p) \cdot 0 = p$;

ii) $V(X) = E(X^2) - (E(X))^2 = (p \cdot 1^2 + (1-p) \cdot 0^2) - P^2 = p - p^2 = p(1-p)$.

Muitos experimentos podem ser estudados a partir deste modelo. Até mesmo o lançamento de um dado honesto em que, de modo clássico, se visualiza sob a perspectiva do modelo uniforme discreto. Pois, ao se atribuir o valor sucesso a um resultado de interesse, o evento complementar recebe o valor fracasso e tem probabilidade determinada mediante a conferida ao valor sucesso.

Contudo, a maior importância deste modelo reside em sua compreensão ser pré-requisito para o entendimento de outros. Neste sentido, fazemos menção a função geratriz de momentos da variável de Bernoulli.

Proposição 1. *Se $X \sim Ber(p)$, então $M_X(t) = (p-1) + e^t p$.*

Demonstração. $M_X(t) = E(e^{tX}) \Rightarrow M_X(t) = \sum_{x=0}^{1} e^{tx}P(x) \Rightarrow M_X(t) = e^{t\cdot 0}(1-p) + e^{t\cdot 1}p \Rightarrow M_X(t) = (1-p) + e^t p.$ □

Distribuição Binomial

Alguns experimentos aleatórios podem ser vistos como repetições de n ensaios de Bernoulli. Nesses casos, os possíveis resultados são sequências de eventos independentes do tipo sucesso ou fracasso. Para experimentos que se enquadram em tal descrição, defini-se a variável X como sendo o número de sucessos em n repetições independentes. E toma-se para sua distribuição o modelo binomial.

Definição 35. *Dado um experimento aleatório formado por uma sequência de n ensaios de Bernoulli independentes, onde p é a probabilidade de sucesso em um ensaio; dizemos que possui distribuição binominal, e denotaremos por $X \sim Bin(n, p)$, a variável X definida como número de sucessos nas n repetições independentes.*

Como situação ilustrativa, suponha que numa escola 80% dos alunos tenham sido vacinados contra a covid 19. Escolhendo-se um aluno ao caso, temos duas possibilidades quanto a vacina: Está vacinado ou não está. Definindo o resultado vacinado como sucesso, e o contrário como fracasso, temos um experimento de Bernoulli. Caso escolhamos dois, três ou mais alunos ao acaso, temos uma repetição de ensaios de Bernoulli, e assim, um experimento que pode ser modelado segundo a Distribuição Binomial.

Neste modelo, o cálculo da probabilidade $P(X = x)$ é dado em função do número n de repetições, e da probabilidade de sucesso em um ensaio p. Por isso, costuma-se dizer que essas informações (n,p) são os parâmetros do modelo. De modo geral, considerando p a probabilidade de sucesso em um ensaio, a probabilidade de X = x sucessos, em n repetições é dada por:

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

Tal expressão é a representação analítica da função de probabilidade(fp) para a variável binomial X. Mas, como se chega a essa relação? Que conceitos matemáticos se relacionam na sua definição? E quais pressupostos estão envolvidos ?

Essencialmente, podemos compreendê-la como aplicação do Teorema do Produto de Probabilidade e do Princípio Fundamental da Contagem. Considerando um experimento formado por n ensaios de Bernoulli, podemos visualizar cada possibilidade que compõe o espaço amostral como sequências cujos os termos são 0 ou 1. Supondo que desejamos calcular a probabilidade de ocorrer x sucessos, precisamos saber quantas sequências formam esse evento e como calcular a sua probabilidade.

Para melhor compreender como se dará o processo como um todo, podemos olhar para uma sequência em particular que atenda as condições descritas. Por exemplo, a sequência na qual os x primeiros termos são iguais a 1, e os n-x termos restantes são iguais a 0:

$$(1, 1, 1, 1, 1, 1, \cdots, 1, 0, 0, 0, 0. \cdots, 0).$$

Como cada ensaio ocorre de forma independente, e essa é uma condição necessária a caracterização do modelo Binomial, temos pelo Teorema do produto de probabilidades que a sequência exibida, bem como outras que formam o evento, tem probabilidade igual $p^x(1-p)^{n-x}$. Na qual, p é a probabilidade do resultado ser 1; E $1-p$, a probabilidade de ser 0. Mas, quantas sequências formam o evento desejado?

Para chegarmos ao número de sequências, podemos supor, inicialmente, que os n experimentos dão origem a números distintos. Dessa forma para dispô-los em n posições, temos $n!$ modos (permutações) pelo Princípio Fundamental da Contagem. Logo, teríamos $n!$ sequências.

Como na realidade temos um elemento que se repete x vezes, e outro que se repete $n-x$ vezes, na contagem inicial cometemos o erro de contar como diferentes as sequencias em que esses elementos iguais trocam de posição entre si. Mais precisamente, as sequências diferentes estão repetidas em quantidade igual a $x!(n-x)!$.

E para corrigir o erro citado, dividimos os $n!$ por $x!(n-x)!$. Assim obtemos o total de sequências distintas que é o número de combinação simples de n tomados x a x, denotado por C_n^x ou número binomial$\binom{n}{x}$. Logo, a probabilidade do evento é dada por $\binom{n}{x}p^x(1-p)^{n-x}$, uma vez que as sequências comportam-se como eventos disjuntos.

Na ilustração, se tomássemos o número de repetições n=3, teríamos oito possíveis sequências de resultados (Ensaios de Bernoulli) formando o espaço amostral do experimento. Caso desejássemos mensurar as chances de dois alunos estarem vacinados, teríamos três sequências formando o evento desejado. Logo, a probabilidade desejada seria $3 \cdot 0,8^2 \cdot 0,2 = 0,384$.

O valor esperado e a variância de uma variável binomial X se obtém, conforme relações:

i) $E(X) = np$;

ii) $V(X) = np(1-p)$.

A demonstração dessas relações pode ser feita de forma direta, aplicando as definições ou via propriedades.

Demonstração. Inicialmente, observamos que sendo $X \sim Bin(n,p)$, a mesma pode ser expressa como soma de n variáveis do tipo $Y_i \sim Ber(p)$, com $i = 1, 2, 3 \cdots n$. Isto é, $X = Y_1 + Y_2 + Y_3 + \cdots + Y_n$. Segue, pela propriedade do valor esperado da soma, que $E(X) = E(Y_1 + Y_2 + Y_3 + \cdots + Y_n) = E(Y_1) + E(Y_2) + E(Y_3) + \cdots + E(Y_n)$. Como, $E(Yi) = p$, $\forall i$. Resulta que $E(X) = p + p + p + \cdots + n = np$. De modo análogo, temos que $V(X) = V(Y_1 + Y_2 + Y_3 + \cdots + Y_n) = V(Y_1) + V(Y_2) + V(Y_3) + \cdots + V(Y_n)$, por aplicação da propriedade da variância da soma para variáveis independentes. Sabendo que $V(Y_i) = p(1-p), \forall i$, decorre que $V(X) = np(1-p)$. □

Como já exposto, a variância e outras medidas descritivas podem ser obtidas através dos momentos centrados na origem, que por sua vez descendem da função geradora de momentos (fgm) quando essa existe.

Proposição 2. *Se $X \sim Bin(n,p)$, então $M_X(t) = (e^t p + 1 - p)^n$.*

Para chegarmos a essa função, podemos aplicar a definição $M_X(t) = E(e^{tX})$ e o teorema Binomial ou usarmos o fato de ser a variável binomial uma soma de variáveis independentes do tipo Bernoulli. Opção que adotamos.

Demonstração. Se $X \sim Bin(n,p)$, então $X = Y_1 + Y_2 + Y_3 + \cdots + Y_n$, com $Y_i \sim Ber(p), \forall i = 1, 2, 3 \cdots n$. Segue, pela propriedade da soma da fgm para variáveis independentes, que $M_X(t) = M_{Y_1}(t) M_{Y_2}(t) M_{Y_3}(t) \cdots M_{Y_n}(t)$. Como $M_{Y_i}(t) = e^t p + (1 - p), \forall i$. Resulta que $M_X(t) = (e^t p + 1 - p)^n$. □

Aplicação

Conforme Meyer [07], ao se buscar construir um modelo probabilístico para uma situação real, em algum momento, pode ser necessário tomar uma probabilidade proveniente de uma experimentação ou até mesmo oriunda de uma experiência prévia, uma probabilidade subjetiva. E partindo desse ponto, conduz -se um processo dedutivo.

Apoiando-se nessa ideia, trazemos a seguinte situação:

Exemplo 5. *Como parte de sua política de assistência aos funcionários, uma empresa paga a cada pessoa que se torne incapacitada por acidente de trabalho o valor de R$ 150000,00. Estima-se que a probabilidade de ocorrer um acidente seja de 0,02% no ano, e que os acontecimentos são independentes. Sabendo que a empresa possui 10000 funcionários, deseja-se saber, com probabilidade de pelo menos 98%, quanto a mesma precisa* *ter em caixa, por ano, para quitar as possíveis despesas com acidentes.(Adaptada de [11])*

Dadas as suposições, percebemos de imediato que podemos fazer uso do modelo binomial. Haja visto que os acidentes ocorrem de modo independentes e com probabilidade de ocorrência constante $p = 0,0002$.

Definindo a variável aleatória A : número de pessoas que sofrem acidentes em um ano. Pelas suposições, temos que $A \sim Bin(10000, 0.0002)$.

De onde, temos que :

$$P(A = a) = \binom{10000}{a} 0.0002^a 0.9998^{10000-a}.$$

A princípio, pode parecer suficiente termos o número médio de pessoas que sofrem acidentes por ano. E para analisarmos esse caminho calculamos $E(A)$. Do exposto sobre o modelo, sabemos que $E(A) = np \Rightarrow E(A) = 10000 \cdot 0.0002 = 2$. Logo, esperasse que em média duas pessoas sofram acidentes em um ano. É correto dizer que R$300000,00 é o valor procurado?

Na verdade, não. Pois, nem ao menos sabemos qual é a probabilidade de que dois acidentes ocorra em um ano. E ainda que já tivéssemos essa informação, ela deveria atender a condição de ser pelo menos igual a 98%, o quê não é verdade, como será mostrado.

Para potencializarmos a exploração da situação, bem como, estabelecer conexões com a teoria exposta até aqui, definiremos outra variável.

Seja $D = 150000A$ a variável que representará a despesa da empresa no ano com acidentes. Da forma como definida, observamos que os valores por ela assumidos são enumeráveis $\{0, 15000, 300000, \cdots\}$, logo D, é também discreta. Dito isso, observamos que podemos calculara P(D=d) a partir do que sabemos sobre a variável A. Veja que

$$P(D = d) \Leftrightarrow P(150000A = d) \Leftrightarrow P\left(A = \frac{d}{150000}\right).$$

Com isso, observamos que nosso objetivo é encontrar o valor "máximo"d que a variável D pode assumir com probabilidade maior ou igual a 98%. Isto é, desejamos d, tal que $P(D \leq d) \geq 0.98$.

Da observação feita anteriormente, temos que

$P(D \leq d) \geq 0.98 \Rightarrow P\left(A \leq \dfrac{d}{150000}\right) \geq 0.98 \Rightarrow P(A \leq a) \geq 0.98.$

Assim, precisamos encontrar o valor da variável A cuja probabilidade acumulada seja maior ou igual 98%. Fazendo os cálculos :

$$P(A = 0) = \binom{10000}{0} 0.0002^0 0.9998^{10000} = 0.1353;$$

$$P(A = 1) = \binom{10000}{1} 0.0002^1 0.9998^{9999} = 0.2707;$$

$$P(A = 2) = \binom{10000}{2} 0.0002^2 0.9998^{9998} = 0.2707;$$

$$P(A = 3) = \binom{10000}{3} 0.0002^3 0.9998^{9997} = 0.1805;$$

$$P(A = 4) = \binom{10000}{4} 0.0002^4 0.9998^{9996} = 0.0902;$$

$$P(A = 5) = \binom{10000}{5} 0.0002^5 0.9998^{9995} = 0.0361.$$

Somando essas probabilidades obtemos 0.9835. Logo , $P(A \leq 5) \geq 0.98$. E, assim, o valor que a empresa precisa ter em caixa no ano é 5×150000 =R$750000,00.

Modelo Hipergeométrico

Essencialmente, esse modelo se adéqua a um experimento aleatório formados por sequência de ensaios de Bernoulli dependentes.

Definição 36. *Considerando um conjunto com N elementos, dentre os quais, apenas r possui um atributo de interesse. Se um experimento aleatório consiste em retirar n elementos deste conjunto sem reposição (isso caracteriza a dependência), então dizemos que possui distribuição hipergeométrica, e denotamos por $X \sim Hip(N, r, n)$, a variável X que contabiliza o número de elementos com atributo de interesse obtidos dentre os n escolhidos.*

Este modelo, assim como os já contemplados até momento, tem ampla adequação a problemas clássicos envolvendo diversos experimentos aleatórios.

A função de probabilidade do modelo é dada por

$$P(X = x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}.$$

Nesta expressão, fica subentendido que os valores de X satisfazem as condições de existência dos números binomiais. O que pode ser resumido ao expressarmos que $max\{0, n - N + r\} \leq x \leq min\{n, r\}$, conforme consta em [10].

Demonstração. De fato, pela definição de número binomial, temos que $x \leq r$, $n \leq N$ e $n - x \leq N - r$. Assim, $n - x \leq N - r \Rightarrow x \geq n - N + r$. Como $x \geq 0 \Rightarrow x \geq max\{0, n - N + r\}$. Por outro lado, $x \leq r$ e $x \leq n \Rightarrow x \leq min\{r, n\}$. □

A expressão da fp como dada é, essencialmente, a aplicação da definição clássica de probabilidade onde, para se calcular os casos favoráveis, bem como o número de casos possíveis, faz-se o uso do princípio fundamental da contagem e do conceito de combinações simples. Ilustração clássica dessa distribuição foi dada na seção sobre variáveis aleatórias, exemplo 3.

Proposição 3. *Se* $X \sim Hip(N, r, n)$, *então* $E(X) = \dfrac{nr}{N}$.

Demonstração. Por definição, $E(X) = \sum_{x=0}^{n} xP(x)$, Assim,

$$E(X) = \sum_{x=0}^{n} xP(x) \quad \Rightarrow \quad E(X) = \sum_{x=1}^{n} x\frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}.$$

Observando que

$$\binom{r}{x} = \frac{r(r-1)!}{x(x-1)!(r-x)!} = \frac{r(r-1)!}{x(x-1)![(r-1)-(x-1)]!}.$$

Logo que, $\binom{r}{x} = \frac{r}{x}\binom{r-1}{x-1}$, e de modo análogo, que $\binom{N}{n} = \frac{N}{n}\binom{N-1}{n-1}$. Segue que,

$$E(X) = \sum_{x=1}^{n} x\left[\frac{\frac{r}{x}\binom{r-1}{x-1}\binom{N-r}{n-x}}{\frac{N}{n}\binom{N-1}{n-1}}\right] \Rightarrow E(X) = \frac{rn}{N}\sum_{x=1}^{n}\frac{\binom{r-1}{x-1}\binom{N-r}{n-x}}{\binom{N-1}{n-1}}.$$

Fazendo $h = r-1$, $g = x-1$, $k = N-1$ e $z = n-1$, buscaremos mostrar que a soma obtida vale 1, o que nos leva ao resultado desejado para $E(X)$. Veja que para $x = 1$, temos $g = 0$, e para $x = n$, temos $g = z$. Logo,

$$\sum_{x=1}^{n}\frac{\binom{r-1}{x-1}\binom{N-r}{n-x}}{\binom{N-1}{n-1}} \Leftrightarrow \sum_{g=0}^{z}\frac{\binom{h}{g}\binom{(k+1)-(h+1)}{(z+1)-(g+1)}}{\binom{k}{z}} \Leftrightarrow \sum_{g=0}^{z}\frac{\binom{h}{g}\binom{k-h}{z-g}}{\binom{k}{z}} \Leftrightarrow \sum_{g=0}^{h} P(g)$$

onde $G \sim Hip(K, h, z)$. Portanto, $\sum_{g=0}^{h} P(g) = 1$, pois $P(g) = \dfrac{\binom{h}{g}\binom{k-h}{z-g}}{\binom{k}{z}}$ é uma função de probabilidade.

$\square$

Proposição 4. *Se* $X \sim Hip(N, r, n)$*, então* $V(X) = \dfrac{nr(N-r)(N-n)}{N^2(N-1)}$.

Demonstração. Sabemos que $V(X) = E(X^2) - [E(X)]^2$, assim vamos obter $E(X^2)$, uma vez que já temos $E(X)$.

$$E(X^2) = \sum_{x=0}^{n} x^2 P(x) \Rightarrow E(X^2) = \sum_{x=1}^{n} x^2 \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}.$$

Observando que $\binom{r}{x} = \frac{r}{x}\binom{r-1}{x-1}$ e, de modo análogo, que $\binom{N}{n} = \frac{N}{n}\binom{N-1}{n-1}$, segue que

$$E(X^2) = \sum_{x=1}^{n} x^2 \frac{\frac{r}{x}\binom{r-1}{x-1}\binom{N-r}{n-x}}{\frac{N}{n}\binom{N-1}{n-1}}.$$

Disto resulta que $E(X^2) = \dfrac{nr}{N} \sum_{x=1}^{n} x \dfrac{\binom{r-1}{x-1}\binom{N-r}{n-x}}{\binom{N-1}{n-1}}$. Fazendo $t = r-1$, $w = x-1$, $z = N-1$, e $k = n-1$, obtemos $E(X^2) = \dfrac{nr}{N} \sum_{w=0}^{k} (w+1) \dfrac{\binom{t}{w}\binom{z-t}{k-w}}{\binom{z}{k}}$. Assim,

$$E(X^2) = \frac{nr}{N} \sum_{w=0}^{k} w \frac{\binom{t}{w}\binom{z-t}{k-w}}{\binom{z}{k}} + \frac{nr}{N} \sum_{w=0}^{k} \frac{\binom{t}{w}\binom{z-t}{k-w}}{\binom{z}{k}}.$$

Observando que $\dfrac{\binom{t}{w}\binom{z-t}{k-w}}{\binom{z}{k}}$ é a fp de uma variável $W \sim Hip(z, t, k)$, decorre que

$$E(X^2) = \frac{nr}{N} E(W) + \frac{nr}{N} \Rightarrow E(X^2) = \frac{nr}{N}\frac{kt}{z} + \frac{nr}{N} \Rightarrow E(X^2) = \frac{nr}{N}\frac{(n-1)(r-1)}{N-1} + \frac{nr}{N}.$$

Logo, $V(X) = \dfrac{nr}{N}\dfrac{(n-1)(r-1)}{N-1} + \dfrac{nr}{N} - \dfrac{n^2r^2}{N^2} \Rightarrow V(X) = \dfrac{rn}{N^2}\dfrac{(N-r)(N-n)}{(N-1)}$.

□

Exemplo 6. *Em determinado posto de polícia rodoviária estadual, veículos de cargas são sempre inspecionados por amostragem. Numa dada abordagem, um veículo com 50 caixas, cada uma com 20 celulares, foi inspecionada pelos policiais Antônio e Fábio.*

Nesse tipo de abordagem, um dos policiais conversa com o motorista e vê a sua documentação e a do veiculo, enquanto outro faz a inspeção da carga. A escolha por uma ou outra função ocorre de forma aleatória entre eles. Sabe-se que nessas situações, Fábio sempre escolhe aleatoriamente uma caixa, colhe três aparelhos e verifica a situação fiscal destes. Já Antônio, sempre escolhe aleatoriamente três caixas, de cada uma pega um item, e vê a questão fiscal destes. Caso ao menos um item esteja irregular, toda a carga será detida, passará por uma vistoria completa, e as devidas penalidades são aplicadas aos responsáveis. Sabendo que nessa carga 10 caixas contém cada uma 8 aparelhos irregulares, e considerando que todas as caixas são indistinguíveis por aspectos físicos, qual a probabilidade dessa carga ser detida ?

Essencialmente, precisamos determinar a probabilidade de ao menos um celular irregular ser encontrado. Para tanto, vamos definir os seguintes eventos: A: Antônio é a pessoa que inspeciona; F:Fábio é a pessoa que inspeciona ; C: Uma caixa que contém celulares irregulares é escolhida; D: Ao menos um celular é irregular é encontrado; E: Ao menos uma caixa que contém celulares irregulares é selecionada.

Assim, $P(D) = P(F \cap C \cap D) + P(A \cap E \cap D)$. Portanto, temos dois casos gerais a se analisar:

1ºCaso: Probabilidade de Fábio fazer a inspeção, escolher uma caixa que contenha celulares irregulares e encontrar ao menos um celular irregular $P(F \cap C \cap D)$.

Pela teorema do produto de probabilidades, temos que
$P(F \cap C \cap D) = P(F)P(C|F)P(D|F \cap C)$. Como a decisão de quem faz a inspeção é aleatória, temos que $P(F) = P(A) = \frac{1}{2}$. Dadas as considerações da situação, temos também que $P(C|F) = \frac{1}{5}$. Já a $P(D|F \cap C) = P(X = 1) + P(X = 2) + P(X = 3)$, sendo $X \sim Hip(20, 8, 3)$. Segue que,

$$P(D|F \cap C) = \frac{\binom{8}{1}\binom{12}{2}}{\binom{20}{3}} + \frac{\binom{8}{2}\binom{12}{1}}{\binom{20}{3}} + \frac{\binom{8}{3}\binom{12}{0}}{\binom{20}{3}} = 0,807.$$

Logo , $P(F \cap C \cap D) = 0,5 \times 0,2 \times 0,807 = 0,081$.

$2^{\underline{o}}$Caso: Probabilidade de Antônio fazer a inspeção e encontrar ao menos uma caixa com celulares irregulares e escolher ao menos um celular irregular $P(A \cap E \cap D)$.

Também pelo teorema do produto de probabilidades,

$P(A \cap E \cap D) = P(A)P(E|A)P(D|A \cap E)$. Temos que $P(E|A) = P(Y = 1) + P(Y = 2) + P(Y = 3)$, sendo $Y \sim Hip(50, 10, 3)$. Decorre que

$$P(E|A) = \frac{\binom{10}{1}\binom{40}{2}}{\binom{50}{3}} + \frac{\binom{10}{2}\binom{40}{1}}{\binom{50}{3}} + \frac{\binom{10}{3}\binom{40}{0}}{\binom{50}{3}} = 0,496.$$

Para obtermos $P(D|A \cap E)$, precisamos considerar três situações :

Primeira: Apenas uma caixa com celulares irregulares foi escolhida .

Logo , $P(D|A \cap E) = \dfrac{8}{20}$.

Segunda: Duas caixas com celulares irregulares foram escolhidas. $P(D|A \cap E) = P(Z = 1) + P(Z = 2)$, com $Z \sim Bin\left(\dfrac{8}{20}, 2\right)$.

Resulta que $P(D|A \cap E) = 2.\dfrac{8}{20}\dfrac{12}{20} + \left(\dfrac{8}{20}\right)^2 = 0,64$.

Terceira: três caixas com celulares irregulares foram escolhidas. $P(D|A \cap E) = P(W = 1) + P(W = 2) + P(W = 3)$, sendo $W \sim Bin\left(\dfrac{8}{20}, 3\right)$.

Assim,$P(D|A \cap E) = 3.\dfrac{8}{20}\left(\dfrac{12}{20}\right)^2 + 3.\left(\dfrac{8}{20}\right)^2\dfrac{12}{20} + \left(\dfrac{8}{20}\right)^3 = 0,784$.

Segue que ,

$$P(A \cap E \cap D) = 0,5 \times 0,496 \times \frac{8}{20} + 0,5 \times 0,496 \times 0,64 + 0,5 \times 0,496 \times 0,784 \Rightarrow$$

$P(A \cap E \cap D) = 0,452$. Portanto, $P(D) = 0,533$.

4.2 Modelos Contínuos

No estudo de situações reais que envolvem variáveis aleatórias contínuas, geralmente, o processo tem início com pesquisa buscando verificar se já existe estudo sobre a variável em questão. De todo modo, essa tarefa requer uma compreensão de como a variável se comporta em um intervalo.

Análogo ao caso das variáveis aleatórias discretas, existe uma variedade de modelos que se ajusta bem a diversos experimentos reais que envolvem variáveis aleatórias contínuas. Nessa seção, falaremos do modelo uniforme contínuo que é o mais simples deste seguimento. E reservamos momento do capítulo final a apresentação de outros com foco em funções presentes em suas estruturas.

Por hora, objetivamos com o modelo a seguir, ilustrar procedimentos que podem ser usados nos demais modelos contínuos como forma de verificar as condições postas nas definições apresentadas. Bem como, ilustrar a presença e a conexão de conceitos matemáticos na estruturação do modelo, a exemplo do que se faz em [09], ao mostrar aplicações do cálculo.

Modelo Uniforme contínuo

Definição 37. *Dizemos que possui distribuição uniforme continua, e denotamos por* $X \sim U[\alpha, \beta]$, *(com* $\alpha < \beta$*) a variável aleatória* X *que assume valores no intervalo* $[\alpha, \beta]$, *e que para subintervalos deste com mesmo comprimento, possui probabilidades iguais.*

Neste modelo, a função densidade e definida como

$$f(x) = \begin{cases} \dfrac{1}{\beta - \alpha} & se \quad \alpha \leq x \leq \beta \\ 0 & se \quad x < \alpha \ ou \ x > \beta. \end{cases}$$

De fato, f(x) é uma função definida $\forall x \in \mathbb{R}$ e temos f(x)$\geq$ 0, pois $\beta > \alpha$. Logo, satisfaz a condição inicial da definição 24.

Com relação a integrabilidade, temos que $f(x)$ também atende a esse quesito. Pois, é uma função limitada $\left(0 \leq f(x) \leq \dfrac{1}{\beta - \alpha}\right)$ e descontínua em número finito de pontos:

$x = \alpha$ e $x = \beta$. Tal conclusão pode ser obtida tomando os limites laterais de $f(x)$ nos pontos citados e observando que tais limites são distintos. Logo, que $\nexists \lim_{x \to \alpha} f(x)$ e $\nexists \lim_{x \to \beta} f(x)$ e, assim, que a função é descontínua nesses pontos. Algo reforçado pela análise do esboço gráfico, percebendo que o mesmo apresenta "saltos"nesses pontos.

Para verificarmos a segunda condição, calculamos a integral:

$$\int_{-\infty}^{+\infty} f(x)dx = \int_{-\infty}^{\alpha} 0dx + \int_{\alpha}^{\beta} \frac{1}{\beta - \alpha}dx + \int_{\beta}^{+\infty} 0dx = \left[\frac{x}{\beta - \alpha}\right]_{\alpha}^{\beta} = \frac{\beta - \alpha}{\beta - \alpha} = 1.$$

A verificação da segunda condição se faz necessário por conta do axioma 2 da definição17. Segundo a qual a probabilidade de ocorrência do evento certo é 1.

Conforme definição, obtemos a função distribuição $F(x)$ também definida por mais de uma sentença:

$$F(x) = \int_{-\infty}^{x} f(t)dt = \begin{cases} \int_{-\infty}^{x} 0dt & se \quad x < \alpha \\ \int_{-\infty}^{\alpha} 0dt + \int_{\alpha}^{x} \frac{1}{\beta - \alpha}dt & se \quad \alpha \leq x \leq \beta \\ \int_{-\infty}^{\alpha} 0dt + \int_{\alpha}^{\beta} \frac{1}{\beta - \alpha}dt + \int_{\beta}^{x} 0dt & se \quad x > \beta. \end{cases}$$

Observamos que sendo a função densidade definida por mais de uma sentença e integrável em qualquer intervalo fechado, no processo para obtenção da função distribuição, precisamos analisar cada um dos intervalos que definem as sentenças e, quando necessário, desmembrar a integral imprópria em parcelas. Como mostrado:

$$F(x) = \begin{cases} 0 & se \quad x < \alpha \\ \dfrac{x - \alpha}{\beta - \alpha} & se \quad \alpha \leq x \leq \beta \\ 1 & se \quad x > \beta. \end{cases}$$

Para ilustrarmos os aspectos mencionados, bem como, para servir de suporte a outras considerações, vamos explorar a seguinte situação:

Exemplo 7. *Marcos é aluno do curso de Matemática da Urca e todos os dias precisa se deslocar de ônibus de Brejo Santo-Ce, onde reside, para Juazeiro do Norte onde faz a faculdade. Sabe-se que o ônibus que transporta Marcos passa no seu ponto de embarque entre 17 horas e 6 minutos e 17 horas e 12 minutos. Qual a probabilidade de Marcos conseguir pegar esse ônibus se ele chegar no ponto de embarque às 17 horas e 10 minutos ?*

Pelo enunciado, é razoável supormos que o ônibus pode passar em qualquer instante pertencente ao intervalo dado, pois não são fornecidos argumentos contrários. Desse modo, tomando a variável X como o instante em que ônibus passa, dadas as suposições, podemos considerar que sua distribuição é uniforme.

Para sermos mais objetivos, vamos considerar apenas o intervalo entre 6 e 12 minutos, pois de fato, temos a mesma amplitude. Dessa forma, temos uma função densidade na qual $\alpha = 6$ e $\beta = 12$, como segue:

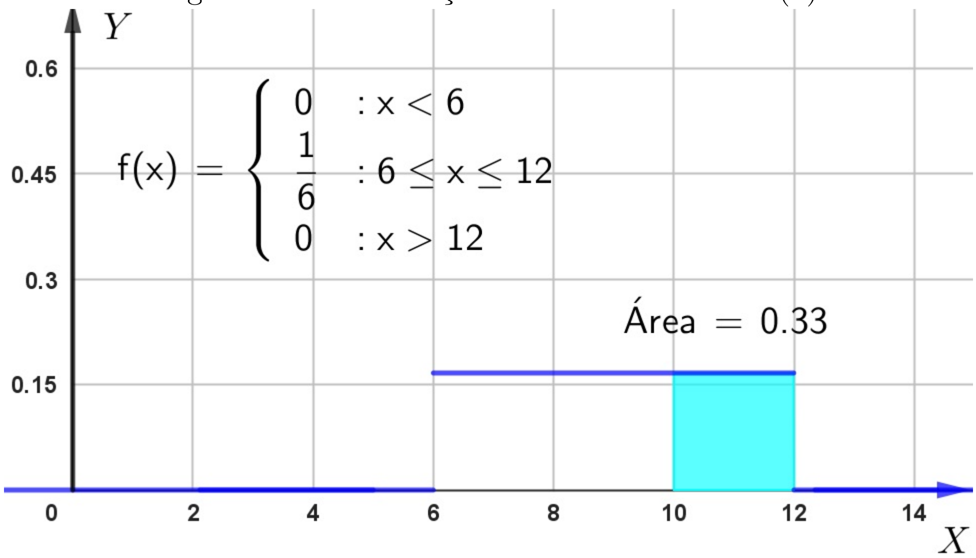
$$f(x) = \begin{cases} \dfrac{1}{6} & se \qquad 6 \leq x \leq 12 \\ 0 & se \quad x < 6,\ ou\ x > 12. \end{cases}$$

Como já exposto, a probabilidade de X assumir valor em qualquer intervalo é dada pela integral dessa função nesse intervalo. Cujo significado geométrico é a área sob o gráfico da função no mesmo intervalo. Por essa razão, nesta distribuição, podemos usar apenas o cálculo de área de retângulos para obtenção das probabilidades. Observe que, de modo geral, tal área delimita-se verticalmente pelo eixo das abscissas e por uma reta horizontal, já que uma das sentenças que define a função densidade é uma constante.

Para respondermos a pergunta proposta, precisamos calcular a probabilidade do ônibus passar entre 10 e 12 minutos. Isto é, a probabilidade de X assumir valor nesse intervalo $(10 \leq X \leq 12)$.

Portanto, precisamos calcular a área de um retângulo de comprimento 2 e altura $\frac{1}{6}$, a qual nos dá $\frac{1}{3}$ ou aproximadamente 33%, como se mostra no esboço a seguir.

Figura 4.1: Distribuição Uniforme Contínua f(x)



Fonte:Ilustração do autor

Observamos que esse mesmo resultado pode ser obtido a partir da função distribuição como diferença entre duas probabilidades acumuladas: $P(X \leq 12)$ e $P(X \leq 10)$, conforme é mostrado. Da $F(x)$, como segue:

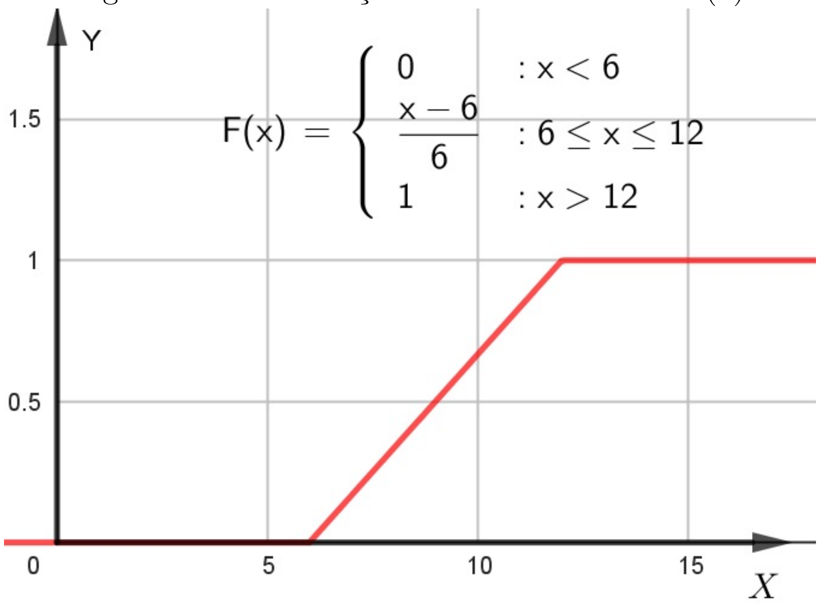
$$F(x) = \begin{cases} 0 & se \quad x < 6 \\ \dfrac{x-6}{6} & se \quad 6 \leq x \leq 12 \\ 1 & se \quad x > 12. \end{cases}$$

Tem-se que $P(X \leq 12) - P(X \leq 10) = F(12) - F(10) = \frac{6}{6} - \frac{4}{6} = \frac{1}{3}$.

Esse resultado é $P(10 \leq X \leq 12)$ que é a mesma para qualquer subintervalo de amplitude 2 nessa distribuição. Nesse sentido, podemos até usar de um raciocínio proporcional para determinarmos a probabilidade procurada. Perceba que o evento desejado é um intervalo cuja amplitude representa $\frac{1}{3}$ da amplitude do intervalo no qual a variável está definida $6 \leq X \leq 12$. Logo, como a área de um retângulo é proporcional a base e a altura, temos que a probabilidade desejada é $\frac{1}{3}$ da $P(6 \leq X \leq 12) = 1$.

Sobre a função distribuição $F(x)$, podemos ver, do ponto de vista analítico e gráfico, que a mesma é contínua, conforme já mencionado de modo geral, a luz da teoria presente em [09].

Figura 4.2: Distribuição Uniforme Contínua F(x)



Fonte:Ilustração do autor

Além disso, evidencia-se no aspecto gráfico presente na figura 4.2 o fato de $F(x)$ não ser derivável em $x = 6$ e $x = 12$. Característica que costuma ser apontada dizendo que o gráfico apresenta "bicos"[3]

[3]Pontos, onde graficamente, temos mais que uma reta tangente.

5 Algumas Funções Aplicadas à Estatística

Em estudos que envolvem modelos determinísticos, existem funções que aparecem em várias aplicações e por isso são notáveis. Dentre essas, citamos as funções dos tipos afim e quadrática. De modo análogo, as funções Exponencial de base natural, a Gama e a Beta estão presentes na estruturação de modelos probabilísticos que se prestam a diversos estudos estatísticos. E por esta, dentre outras razões, ocupam posição importante na literatura. Neste capítulo, faremos uma apresentação dessas funções e buscaremos evidenciar a importância que as mesmas possuem aplicadas à Estatística.

5.1 Sobre a Função Exponencial de Base Natural

Embora estejamos interessados em um caso particular, apresentaremos a definição geral de função exponencial e, de forma concisa, sua caracterização, com base em [12]. Com este propósito, admitiremos conhecidas as propriedades sobre o cálculo de potências. Bem como, a análise necessária para se estender o conceito de exponenciação em $\mathbb{N}$ para $\mathbb{Z}$, $\mathbb{Q}$ e, por fim $\mathbb{R}$. Quando, em termos práticos, já se chega a função em questão. De todo modo, na fonte citada consta tal análise.

Definição 38. *Dado $a \in \mathbb{R}$, positivo e diferente de 1, chama-se de função exponencial de base a, $f : \mathbb{R} \to \mathbb{R}_+$, a função dada por $f(x) = a^x$.*

Pode-se mostrar que essa definição satisfaz uma série de propriedades. Dentre as quais, destacam-se como essenciais as seguintes :$f(x + y) = f(x) \cdot f(y)$; Se $a > 1$, f é crescente; Se $0 < a < 1$, f é decrescente.

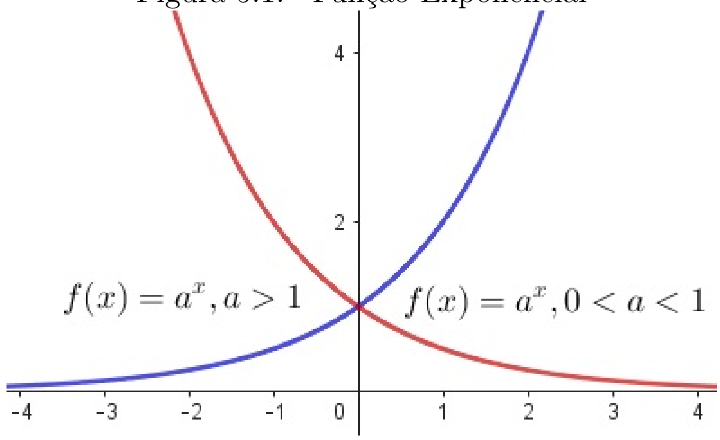
Tais propriedades estão presentes na composição do resultado a seguir conhecido como*Teorema de Caracterização da Função Exponencial.* O qual é importante para o reconhecimento da função exponencial e de situações que justifiquem o seu emprego como modelo matemático.

Teorema 7. *Seja* $f : \mathbb{R} \to \mathbb{R}_+$ *uma função estritamente crescente ou estritamente decrescente. São equivalentes as afirmações :* $1) f(nx) = f(x)^n$*, para todo* $n \in \mathbb{Z}$ *e todo* $x \in \mathbb{R}$; $2) f(x) = a^x$*, para todo* $x \in \mathbb{R}$*, onde* $a = f(1)$; $3) f(x+y) = f(x) \cdot f(y)$ *, quaisquer que sejam* $x, y \in \mathbb{R}$;

A demonstração deste teorema consiste em mostrar que $1 \Rightarrow 2 \Rightarrow 3 \Rightarrow 1$. Embora aqui omitida, pode ser encontrada em [12].

O esboço gráfico da função nos remete a percepção de outras características como continuidade, sobrejetividade, e limites quando $x \to \pm\infty$. Como se evidencia na figura 5.1, $\lim_{x\to\infty} f(x) = \infty$, se $a > 1$; E 0, se $0 < a < 1$. Já $\lim_{x\to-\infty} f(x) = 0$, se $a > 1$; E $+\infty$, se $0 < a < 1$.

Figura 5.1: Função Exponencial



Fonte: Ilustração do autor

Em geral, as aplicações mais conhecidas e atribuídas a função exponencial, envolvem funções da forma $g(x) = ba^x$, com $g : \mathbb{R} \to \mathbb{R}_+$, onde a e b são constantes reais positivas e $a \neq 1$. Tais funções são denominadas funções do tipo exponencial. E tem na sua base o conhecimento da função exponencial.

Função Exponencial de Base Natural

Ainda com base em [12], podemos definir esse caso particular da função exponencial como segue:

Definição 39. *Chama-se de função exponencial de base natural, a função* $f : \mathbb{R} \to \mathbb{R}_+$*, dada por* $f(x) = e^x$*. Em que* $e = \lim_{n\to\infty} \left(1 + \dfrac{1}{n}\right)^n$ *é um número real irracional.*

Uma propriedade que justifica a presença dessa função em várias situações que envolvem cálculo infinitesimal, diz respeito a sua derivada ser igual a própria função.

Demonstração. $f(x) = e^x \Rightarrow f'(x) = \lim_{h\to 0} \dfrac{e^{x+h} - e^x}{h} \Rightarrow f'(x) = e^x \lim_{h\to 0} \dfrac{e^h - 1}{h}$ $\Rightarrow f'(x) = e^x$. Pois, $\lim_{h\to 0} \dfrac{e^h - 1}{h} = 1$. De fato, fazendo $e^h - 1 = y$, temos que $\lim_{h\to 0} \dfrac{e^h - 1}{h} = \lim_{y\to 0} \dfrac{y}{\ln(y+1)} = \dfrac{1}{\lim_{y\to 0} \dfrac{\ln(y+1)}{y}} = \left[\lim_{y\to 0} \ln(y+1)^{\frac{1}{y}}\right]^{-1} =$ $\left[\ln\left(\lim_{y\to 0}(y+1)^{\frac{1}{y}}\right)\right]^{-1} = \left[\ln\left(\lim_{u\to\infty}\left(\frac{1}{u}+1\right)^u\right)\right]^{-1} = 1.$

□

Essencialmente, isso significa que a taxa de variação instantânea dessa função, em qualquer ponto, é igual ao valor que a função assume nesse ponto. Particularmente, essa propriedade é explorada quando ao invés de utilizar $g(x) = ba^x$, utiliza-se $g(x) = be^{dx}$, tomando-se $a = e^d$ no emprego de uma função do tipo exponencial. Pois, dessa forma, além de se conhecer o valor inicial(b), também tem-se destacado o fator de proporcionalidade da função que é d. Tornando-se possível saber quão rápido a função cresce ou decresce.

Outro fato importante, acerca da função em questão, diz respeito a séries de Taylor. Em síntese, se $f(x)$ é uma função infinitamente derivável(f é de classe C^∞), ela possui uma série de potências da forma $\sum_{n=0}^{\infty} C_n(x - x_0)^n$ onde $C_n = \dfrac{f^{(n)}(x_0)}{n!}$, a qual chama-se série de Taylor da função $f(x)$ em torno de x_0.

De um modo geral, essa série é uma extensão do que se conhece por polinômio de Taylor. Um polinômio de ordem k que tão melhor se aproxima de uma dada função na vizinhança de um ponto x_0, quanto maior for o valor de k. E que para sua construção, a função que se deseja aproximar precisa ser derivável até a ordem K desejável.

Considerando que $f^{(n)}(x_0)$ denota a n-ésima derivada de $f(x)$ aplicada em x_0, podemos expressar o polinômio de Taylor de ordem k desta função em torno do ponto x_0 como segue :

$$P_{k,x_0}(x) = \sum_{n=0}^{k} \frac{f^{(n)}(x_0)(x - x_0)^n}{n!}.$$

Vemos que a estrutura de tal polinômio é, essencialmente, a mesma da série. Diverge por ser uma soma finita e ter o propósito de aproximar o valor da função numa vizinhança de x_0. Razão pela qual existe um erro envolvido chamado de resto de Taylor. Para avaliarmos essa diferença num ponto x fixado, a denotamos por $R_{k,x_0}(x)$ e consideramos que a mesma é da forma de um termo do polinômio, expressa por:

$$R_{k,x_0}(x) = \frac{f^{k+1}(c)(x - x_0)^{k+1}}{(k+1)!},$$

onde c é um valor entre x e x_0. Algo expresso em algumas obras sob o título de resto de Taylor na fórmula de Lagrange, conforme consta em [13].

A série por sua vez, como soma infinita, pode, em alguns casos, ser idêntica a própria função $f(x)$ no seu intervalo de convergência. Quando existe a equivalência, fala-se que a função é analítica. É o caso da função exponencial e^x, cuja série de Taylor em torno de $x_0 = 0$, fornece a seguinte equivalência $\forall x \in \mathbb{R}$:

$$e^x = \sum_{n=0}^{\infty} \frac{e^0 x^n}{n!} = 1 + \frac{x}{1} + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} \cdots$$

A demonstração desse fato, com base em [13], consiste em mostrar que $|R_{k,x_0}(x)| \to 0$, quando $k \to \infty$.

Demonstração. Como $x_0 = 0$, e $f^{k+1}(c) = e^c, \forall k \in \mathbb{N}$. Fixado $x \in \mathbb{R}$, segue que

$$|R_{k,x_0}(x)| = \left| \frac{f^{k+1}(c)(x-0)^{k+1}}{(k+1)!} \right| = \frac{e^c |x|^{k+1}}{(k+1)!}.$$

Como c está entre $x_0 = 0$ e x fixado , temos que $e^c < 1$, caso $x < 0$; Ou $e^c < e^x$, caso $x > 0$. O que nos permite afirmar que $e^c < e^{|x|}$, observando que $f(x)$ é crescente. Segue que

$$\frac{e^c |x|^{k+1}}{(k+1)!} < \frac{e^{|x|} |x|^{k+1}}{(k+1)!}.$$

Fazendo $k+1 = n$, podemos mostrar que a sequência de termo geral $a_n = \dfrac{|x|^n}{n!}$ converge para zero. De fato, considerando a série $\sum_{n=0}^{\infty} \dfrac{|x|^n}{n!}$, e aplicando o critério da razão[4], temos que

$$\lim_{n\to\infty} \frac{|x|^{n+1}}{(n+1)!} \frac{n!}{|x|^n} = \lim_{n\to\infty} \frac{|x|}{n+1} = 0 < 1 \Rightarrow$$

que a série converge, logo que $a_n \to 0$. Portanto, quando $K \to \infty \Rightarrow |R_{k,X_0}(x)| \to 0$. E como o teste da razão nos levou a zero como limite, a série de Taylor da função $f(x) = e^x$ tem raio de convergência infinito. Logo, converge $\forall x \in \mathbb{R}$. $\square$

A função exponencial de base natural na Estatística

Na teoria de Probabilidade e Estatística, a presença da função exponencial de base natural pode ser vista em contextos gerais, como na definição de função geratriz de momentos(fgm), bem como, na estruturação de modelos probabilísticos como o de Poisson e Exponencial, dentre outros.

[4]Se $(a_n)_{n\geq 1}$ é um sequência de termos não nulos e $\left|\dfrac{a_{n+1}}{a_n}\right| \to L < 1$, então $\sum_{n=1}^{\infty} a_n$ é absolutamente convergente[13].

No que tange a função geratriz de momentos (fgm), evidenciaremos a importância da exponencial de base natural justificando, com base em [07], uma das principais propriedades da (fgm).

Conforme já apresentada na seção sobre variáveis aleatórias, uma das propriedades da (fgm) diz que: Se X é uma variável aleatória, então a derivada r-ésima da sua fgm, dada por $M_X(t) = E(e^{tX})$, aplicada em $t = 0$, gera $E(X^r)$.

Essencialmente, essa propriedade disque que podemos obter momentos de ordem r centrados na origem de uma variável aleatória X, denotados por $E(X^r)$, derivando a função geradora de momentos, indicada por $M_X(t)$, e calculando-a em $t = 0$.

Observando que $t \in \mathbb{R}$, podemos usar série de Taylor para expressar e^{tx}, como segue:

$$e^{tx} = 1 + \frac{tx}{1} + \frac{(tx)^2}{2} + \frac{(tx)^3}{6} + \frac{(tx)^4}{24} \cdots$$

Considerando agora uma variável aleatória X e um t real qualquer, obtemos

$$e^{tX} = 1 + \frac{tX}{1} + \frac{(tX)^2}{2} + \frac{(tX)^3}{6} + \frac{(tX)^4}{24} \cdots$$

Supondo que o valor esperado dessa soma infinita seja igual a soma dos valores esperados de cada parcela, a exemplo do que se faz em [07], segue que

$$M_X(t) = E(e^{tX}) = E\left(1 + \frac{tX}{1} + \frac{(tX)^2}{2} + \frac{(tX)^3}{6} + \frac{(tX)^4}{24} \cdots\right) \Rightarrow$$

$$M_X(t) = 1 + \frac{tE(X)}{1} + \frac{t^2E(X^2)}{2} + \frac{t^3E(X^3)}{6} + \frac{t^4E(X^4)}{24} \cdots$$

Observando que $E(X) \cdots E(X^n)$..., são constantes e que $M_x(t)$ é uma função em t, vamos derivar tal função supondo ser extensível a propriedade da derivada da soma finita para essa soma infinita. Tal suposição se apoia na ideia de que tal soma vem de uma série de potências, que tem como propriedade ser derivável termo a termo no intervalo de convergência.

Disto decorre que

$$M_X'(t) = E(X) + tE(X^2) + \frac{t^2E(X^3)}{2} + \frac{t^3E(X)^4}{6} \cdots$$

$$M_X''(t) = E(X^2) + tE(X^3) + \frac{t^2E(X^4)}{2} \cdots$$

E, assim, constatamos que $M_X'(0) = E(X)$, $M_X''(0) = E(X^2)$. E que essa ideia se estende, conforme descreve a propriedade $M_X^{(r)}(0) = E(X^r)$.

No que segue, contemplamos a presença da função exponencial em modelos probabilísticos.

Modelo de Poisson

Esse modelo carrega o nome de seu idealizador, o matemático francês Siméon Denis Poisson. Em linhas gerais, foi imaginado para descrever, em termos probabilísticos, experimento aleatório formado por uma sequência de infinitos ensaios de Bernoulli independentes. Nesse contexto, em geral, o resultado de interesse tem probabilidade pequena, como o número de lâmpadas defeituosas observadas numa linha de produção [01].

Em termos práticos, definiremos com base em [10].

Definição 40. *Denota-se por $X \sim Poisson(\lambda)$, a variável aleatória discreta X que representa o número de ocorrências de certo evento em um dado intervalo de tempo, comprimento, superfície, volume ou de outras grandezas; E cujas probabilidades são determinadas pela seguinte relação, denominada Modelo de Poisson:*

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

De forma sucinta, podemos dizer com base em [07], que λ representa a taxa de ocorrência do fenômeno estudado por unidade de grandeza do intervalo considerado. Além disso, que fazendo uso das respectivas definições e propriedade da exponencial de base natural, podemos mostrar que $E(X) = \lambda$ e $V(X) = \lambda$.

Demonstração. De fato, $E(X) = \sum_{x=0}^{\infty} xP(x) \Rightarrow E(X) = \sum_{x=0}^{\infty} x\frac{e^{-\lambda}\lambda^x}{x!} \Rightarrow E(X) = e^{-\lambda}\sum_{x=1}^{\infty} x\frac{\lambda^x}{x!} \Rightarrow E(X) = e^{-\lambda}\sum_{x=1}^{\infty} \frac{\lambda^x}{(x-1)!} \Rightarrow E(X) = e^{-\lambda}\lambda\sum_{y=0}^{\infty} \frac{\lambda^y}{y!}$, fazendo $y = x - 1$. Observando que temos a série de Taylor da exponencial, segue que $E(X) = e^{-\lambda}\lambda e^{\lambda} = \lambda$. De modo análogo, mostra-se que $E(X^2) = \lambda^2 + \lambda$. E, assim, que $V(X) = E(X^2) - [E(X)]^2 = \lambda$. $\square$

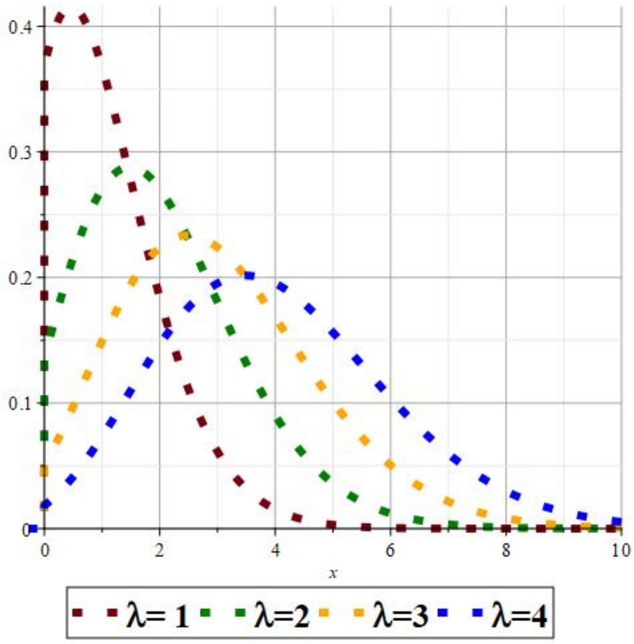
Tendo em vista fazermos menção a relação existente entre este modelo e o que trataremos a seguir, faz sentido mencionar, com base em [07], que o modelo de Poisson resulta de um conjunto de suposições plausíveis e coerentes em diversos contextos. Razão pela qual este modelo é bastante útil no estudo de diversos fenômenos.

Em síntese, tais suposições compreendem admitir, dentre outras considerações, que o número de ocorrências do fenômeno aleatório de interesse, em um dado intervalo, depende do seu comprimento, mas não das suas extremidades; que ocorrências em intervalos não sobrepostos seja independentes; que a probabilidade de ocorrência em intervalos pequenos seja proporcional a λ. E que a probabilidade de duas ocorrências em um intervalo pequeno seja desprezível.

Além disso, que diante de um contexto onde seja coerente fazer tais suposições, dizemos ser adequado empregar o modelo em estudo porque nele ocorre um *Processo de Poisson*.

De um modo geral, temos que a presença da exponencial no modelo tem a finalidade de reparametrizar os valores tendo em vista a soma total das probabilidades ser 1. E fazendo do uso de softwares como o Maple, podemos inferir que a probabilidade neste modelo é máxima quando o número de ocorrências desejado é igual ao parâmetro da distribuição. E que entre distribuições com diferentes parâmetros, o comportamento tende a ser simétrico à médica que o parâmetro aumenta, saindo de uma assimetria positiva, conforme mostra a figura 5.2.

Figura 5.2: Distribuições de Poisson f(x)



Fonte: Ilustração do Autor

Exemplo de aplicação no futebol podemos ver em [14], onde se descreve a utilização deste modelo em um campeonato para estimar os resultados das partidas, apontado chances de vitorias para cada equipe e a possibilidade de empate.

Neste estudo publicado em 2019, procurou-se construir um modelo que possibilitasse as estimativas citadas para o campeonato brasileiro de 1995. Os resultados permitiram concluir acerto de vitórias em 50% das partidas e ocorrência dos placares mais prováveis em 16% destas.

Matematicamente, as probabilidades das partidas foram estimadas pelo modelo seguinte, onde o placar é determinado por um vetor aleatório bivariado. Em que x representa os gols marcados pelo time de casa, e y os gols marcados pelo time visitante. Observando-se ainda que foram desprezadas relações lineares entre as variáveis X,Y; restringindo-se o modelo ao seguinte produto de probabilidades:

$$P(X = x, Y = y) = \frac{e^{-\lambda_x} e^{-\lambda_y} {\lambda_x}^x {\lambda_y}^y}{x!y!}.$$

Sendo a expressão um modelo de probabilidade conjunta de duas variáveis com distribuição de Poisson, o parâmetro λ_x foi definido como sendo a média entre a média dos gols marcados em casa pelo time de casa, e a média dos gols sofridos fora pelo time visitante; já λ_y, representou a média entre a média de gols marcados fora pelo time visitante e a média de gols sofridos em casa pelo time de casa.

A figura 5.3 ilustra estimativas para uma partida ocorrida entre Santos e Bota Fogo na final do campeonato.

Figura 5.3: Tabela de Probabilidades

		Botafogo					
	1x1	**0**	**1**	**2**	**3**	**4**	**5**
Santos	**0**	2,491%	4,120%	3,407%	1,878%	0,777%	0,349%
	1	5,079%	8,399%	6,946%	3,829%	1,583%	0,711%
	2	5,176%	8,561%	7,079%	3,903%	1,614%	0,724%
	3	3,517%	5,817%	4,810%	2,652%	1,096%	0,492%
	4	1,792%	2,964%	2,451%	1,351%	0,559%	0,251%
	5	1,075%	1,778%	1,470%	0,810%	0,335%	0,150%

Fonte: [14]

A pesar de tal partida não ter sido encerrada com o resultado mais provável, que segundo o estudo era de vitória do santos com placar de 2x1 e probabilidade de $8,561\%$; A estimativa mais provável em caso de empate ocorreu, que foi empate de 1x1, com probabilidade de $8,399\%$.

Modelo Exponencial

Definição 41. *Dizemos que uma variável aleatória contínua X possui distribuição exponencial com parâmetro $\lambda > 0$, e denotamos por $X \sim Exp(\lambda)$, quando a sua função densidade de probabilidade for dada por*

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

Aplicando as respectivas definições, podemos mostrar que $E(X) = \dfrac{1}{\lambda}$ e $V(X) = \dfrac{1}{\lambda^2}$.

Demonstração. De fato, temos que

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx \Rightarrow E(X) = \int_{0}^{\infty} x \lambda e^{-\lambda x} dx \Rightarrow E(X) = \frac{1}{\lambda} \int_{0}^{\infty} a e^{-a} da,$$

fazendo a substituição $a = \lambda x$. Segue que

$$E(X) = \frac{1}{\lambda} \lim_{b \to \infty} \int_{0}^{b} a e^{-a} da \Rightarrow E(X) = \frac{1}{\lambda} \lim_{b \to \infty} [-be^{-b}] + \frac{1}{\lambda} \lim_{b \to \infty} [-e^{-b} + 1] \Rightarrow E(X) = \frac{1}{\lambda}.$$

Pois,o primeiro limite é zero e o segundo é 1.(Nessas passagens, usamos integração por partes) De modo análogo, mostra-se que

$$E(X^2) = \frac{2}{\lambda^2}. \text{ De onde se concluir que } V(X) = E(X^2) - [E(X)]^2 \Rightarrow V(X) = \frac{1}{\lambda^2}.$$

□

Observamos que o parâmetro λ é o mesmo da distribuição de Poisson, algo que evidencia a relação entre os dois modelos.

Em termos práticos, podemos entender que em um processo de poisson duas variáveis estão presentes: uma que mapeia o número de ocorrências do fenômeno aleatório de interesse, e outra que que mapeia a "distância"entre as ocorrências. Essa segunda é a variável exponencial. E essa distância, na maioria das aplicações, diz respeito a intervalos de tempo entre as ocorrências. Razão pela qual o inverso do parâmetro λ é o valor esperado da variável, pois retrata a "distância"média entre as ocorrências.

Sob essa concepção, na realidade, podemos chegar a essa forma de apresentação para função densidade partindo de uma abordagem menos formal e usando o que já conhecemos sobre a Distribuição de Poisson.

Para tanto, considere um fenômeno aleatório que ocorre no intervalo $[0, t]$, com t um tempo qualquer em minutos. Supondo que tal fenômeno aconteça com média de λ ocorrências por minuto, faz sentido nos perguntarmos qual a probabilidade da primeira ocorrência se dá após x minutos decorridos ?

Definindo X como sendo o tempo entre as ocorrências, temos assim uma variável contínua. Logo, a pergunta se traduz em saber qual a $P(X > x)$. Que por sua vez é igual a $1 - P(X \leq x)$. Mas, como obter tais probabilidades se não conhecemos a função densidade dessa variável, tão pouco a distribuição acumulada?

Definiremos então a variável Y como sendo o número de ocorrências do fenômeno no intervalo citado. Essa variável por sua vez segue uma distribuição de Poisson. Como se deseja a $P(X > x)$, essa equivale a $P(Y = 0)$ no intervalo [0,x]. Dado que a média é de λ ocorrências por minuto, o parâmetro dessa distribuição de Poisson será de λx ocorrências em x minutos. Segue que $P(X > x) = P(Y = 0) = \dfrac{e^{-\lambda x}(\lambda x)^0}{0!} = e^{-\lambda x}$.

A partir desse ponto, podemos chegar as funções distribuição $F(x)$ e de densidade

$f(x)$. Observe que $e^{-\lambda x} = P(X > x) = 1 - P(X \leq x) \Rightarrow P(X \leq x) = F(x) = 1 - e^{-\lambda x}$. Temos então $F(x) = 1 - e^{-\lambda x}$, para $x \geq 0$, e $F(x) = 0$ para $x < 0$. Da definição de função distribuição $F(x) = \int_0^x f(t)dt$, uma função definida por uma integral, decorre o fato da mesma ser contínua em todo seu domínio. Como também ser derivável em todos os pontos em que $f(x)$ for contínua, sendo $F'(x) = f(x)$ nesses pontos conforme [09]. Destes fatos fatos, resulta que $f(x) = 0$, para $x < 0$; e $f(x) = \lambda e^{-\lambda x}$ para $x \geq 0$.

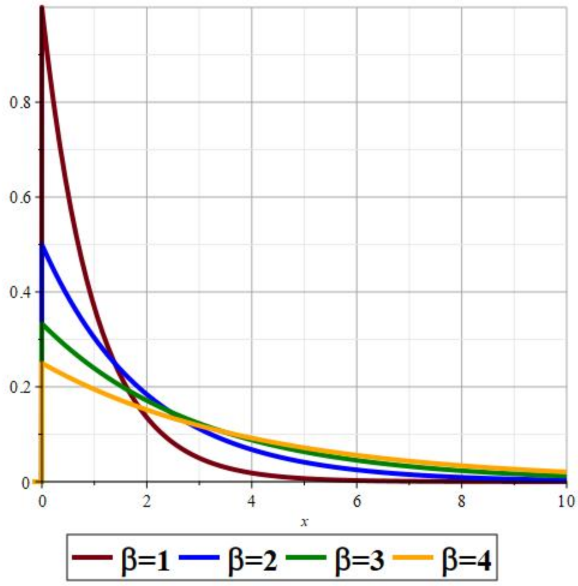
Embora a parametrização apresentada seja mais usual na literatura, a seguinte pode também ser encontrada, a exemplo do que se ver em [09]

$$f(x) = \begin{cases} \dfrac{e^{-x/\beta}}{\beta} & se \quad x \geq 0 \\ 0 & se \quad x < 0. \end{cases}$$

Em que $\beta > 0$ já representa o valor esperado da variável X.

Fazendo uso dessa parametrização, analisamos o comportamento gráfico da distribuição. Conforme mostra a figura 5.4, quanto menor o parâmetro β, mais rapidamente a função densidade decresce. E no intervalo que se estende de 0 até as proximidades de $x = 1$, cada função supera as que possui maior parâmetro, ocorrendo o contrário para $x \geq 4$.

Figura 5.4: Distribuição Exponencial f(x)



Fonte:Ilustração do autor

A essa altura, a interpretação da probabilidade como área sob uma curva definida por uma função densidade já se apresenta como ideia consolidada. Todavia, uma propriedade presente no modelo exponencial, chamada de **Falta de Memória**, contraria a compreensão intuitiva. Algo que nos leva a uma breve análise a esse respeito.

Para tanto, considere a situação ilustrativa a seguir.

Exemplo 8. *Numa dada repartição pública, a duração das ligações recebidas ocorre segundo uma distribuição exponencial com média de 8 minutos. Supondo que apenas uma atendente está disponível, e que essa iniciou o atendimento de uma chamada, imediatamente, antes de uma pessoa chegar, qual a probabilidade dessa pessoa ter que esperar por atendimento mais que* 12 *minutos, se a mesma já esperou* 7*?*

Do enunciado temos que $\beta = 8$ minutos, logo a $F(x)$ dessa distribuição é dada por

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

Observamos também que a situação exige o cálculo de uma probabilidade condicionada: $P[(X > 12)|(X > 7)]$. Decorre do Teorema 4, que

$$P[(X > 12)|(X > 7)] = \frac{P[(X > 12) \cap (X > 7)]}{P(X > 7)}.$$

Como $(X > 12) \cap (X > 7) = (X > 12)$, segue que

$$P[(X > 12)|(X > 7)] = \frac{P(X > 12)}{P(X > 7)}.$$

Usando o conceito de eventos complementares, temos que

$$\frac{P(X > 12)}{P(X > 7)} = \frac{1 - P(X \leq 12)}{1 - P(X \leq 7)}.$$

Pela função distribuição $F(x)$, temos que

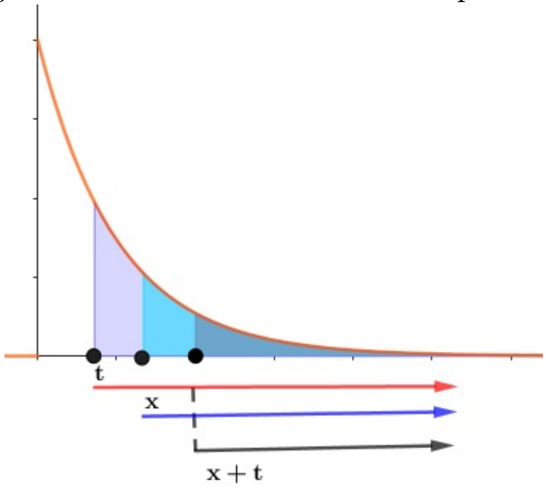
$$\frac{1 - P(X \leq 12)}{1 - P(X \leq 7)} = \frac{e^{-12/8}}{e^{-7/8}} = e^{-5/8}.$$

Mas, como esse resultado pode ser escrito como $1 - (1 - e^{-5/8})$, vemos se tratar de $1 - P(X \leq 5)$. Portanto, é a $P(X > 5)$. Resultados como esse ilustra a propriedade **Falta de Memória** da distribuição exponencial que pode ser expressa como segue:

$$P[(X > x + t)|(X > x)] = P(X > t), \quad \text{com} \quad x, t > 0.$$

Notamos que sob a percepção da probabilidade como área, o resultado da ilustração surpreende e até parece contrariar a nossa intuição, como buscamos representar, simbolicamente, essa visão na figura 5.5.

Figura 5.5: Falta de Memória na Exponencial



Fonte:Ilustração do autor

Nessa ilustração, supomos $x > t$. E para uma melhor compreensão do seu propósito, podemos pensar $x = 7$ e $t = 5$, como ocorrido na ilustração. Assim, temos o significado das probabilidades presentes na discussão pelas áreas ilustradas.

Geometricamente, a intuição que a situação nos trás é que não faria sentido considerar a área em cinza, uma vez que a condição dada já coloca a ocorrência da variável, pelos menos, no intervalo de área em azul claro. Ficando a ideia de que o tempo que já passou não tem influência para o cálculo da probabilidade. Como se o tempo fosse reiniciado. E dessa forma, o cálculo das probabilidades de X não sofrem alterações devido as ocorrências anteriores, conforme [01].

Observamos, com base em [07], que este modelo é o único dentre os contínuos que dispõe dessa propriedade. E além disso, que essa tem o teor de identidade para o modelo. Cuja demonstração, passa por um estágio onde a fdp da variável precisa transformar soma em produto, algo que se encaminha para o modelo em questão dada a caracterização da função exponencial.

Aplicabilidade da Distribuição Exponencial

De acordo com Meyer [07], um dos campos de aplicação deste modelo diz respeito a Teoria de Confiabilidade, a qual podemos compreender como conjunto de conhecimentos a partir dos quais se avalia as circunstâncias em que um sistema ou um dos seus componente pode falhar, por exemplo.

Neste contexto, segundo a mesma referência, o tempo de vida ou de duração de um componente (ou sistema) até ocorrer uma falha é estudado sob a perspectiva de uma variável aleatória contínua. Pois, diversas experiências empíricas comprovam que esse tempo não pode ser previsto por modelos determinísticos.

Dentro dessa concepção, ao tomarmos T a variável em questão, podemos definir a confiabilidade de um componente numa época t $(R(t))$ como sendo a probabilidade deste funcionar adequadamente, dentro das condições para as quais foi projetado, e não falhar no intervalo $[0, t]$. Algo que se expressa matematicamente por

$$R(t) = P(T > t) = \int_t^{\infty} f(x)dx,$$

onde f é uma fdp.

Ainda de acordo com [07], ao buscarmos discutir que fdp pode ser usada na expressão exposta, nos deparamos com uma situação clássica que é saber que modelo usar para representar um fenômeno observado. E que tal escolha deve se adequar ao que sabemos sobre as falhas. Mas, que a priori, pode ser qualquer uma fdp numa perspectiva, estritamente, matemática.

Por fim, o mesmo autor considera que uma das situações mais importantes do contexto ocorre quando a variável T tem distribuição Exponencial. Algo que resulta da suposição que a taxa de falhas instantânea seja constante. Tal taxa é definida como sendo $Z(t) = \dfrac{f(t)}{R(t)}$. De fato, se $f(t) = \lambda e^{-\lambda t}(t > 0)$ for a fdp, então λ será a constante em questão. Pois, obtemos $R(t) = e^{-\lambda t}$. Maiores informações a esse respeito podem ser obtidas em [07].

A título de ilustração, considere a seguinte situação construída com base em [07]:

Exemplo 9. *Um circuito elétrico possui dois componentes dos tipos A e B ligados em paralelo. Considerando que tais componentes funcionam de forma independente, e que sob condições adequadas de temperatura, tensão e corrente, operam com taxas de falhas constantes e iguais a* $0,00001$*e* $0,0002$*, respectivamente, determine a confiabilidade do circuito para um período de 8640 horas.*

Dado que os componentes operam com taxas de falhas constantes, temos que a duração dos mesmos até a falha, que denotaremos por T_A e T_B, seguem distribuição exponencial com parâmetros iguais as respectivas taxas que representaremos por $\lambda_A = 0,00001$ e $\lambda_B = 0,0002$.

Considerando também que os componentes estão ligados em paralelo, o circuito falha quando ambos os componentes falham. Logo, podemos expressar a confiabilidade do sistema $R(t) = P(T > t)$ em função das confiabilidades dos componentes.

De fato,

$$R(t) = P(T > t) \Rightarrow R(t) = 1 - P(T \leq t) \Rightarrow R(t) = 1 - P[(T_A \leq t) \cap (T_B \leq t)] \Rightarrow$$

pela independência, dos eventos em questão que

$$R(t) = 1 - [P(T_A \leq t)P(T_B \leq t)] \Rightarrow R(t) = 1 - \{[1 - P(T_A > t)][1 - P(T_B > t)]\} \Rightarrow$$

$$R(t) = 1 - \{[1 - R_A(t)][1 - R_B(t)]\} \Rightarrow R(t) = R_A(t) + R_B(t) - R_A(t)R_B(t).$$

Logo ,

$$R(t) = e^{-\lambda_A t} + e^{-\lambda_B t} - e^{-(\lambda_A+\lambda_B)t}.$$

Avaliando para $t = 8640$,obtemos

$$R(8640) = \frac{1}{e^{0,0864}} + \frac{1}{e^{1,728}} - \frac{1}{e^{1,8144}} \approx 0,9319.$$

Outro campo no qual encontramos aplicação da distribuição exponencial, bem como, da distribuição de Poisson, diz respeito a Teoria das Filas. Por esse título designa-se um conjunto de técnicas de Estatística e Probabilidade que possibilita o estudo e a caracterização das filas.

Sob a concepção dessa teoria, as filas são vistas como processos de natureza aleatória originados ao se buscar um atendimento que exija uma espera. E por meio de técnicas oriundas da estatística e da probabilidade, descreve-se as filas através de modelos que são aplicados em sistemas de comunicação, na engenharia de tráfego e na estruturação de atendimentos bancários, dentre outros setores da economia, conforme é mencionado em [15].

Em linhas gerais, podemos dizer que tal ramo de estudo busca otimizar os tempos envolvidos na prestação de serviços que demandam filas. Para tanto, concebe a ideia de que uma fila de espera caracteriza-se pelo regime de chegadas de pessoas, pelo regime de serviço e pela organização da fila [15]. Tais aspectos compreendem saber, dentre outras informações, se existe uma procura máxima ou não pelo serviço, que tipo de distribuição descreve as taxas de chegadas e de atendimentos, bem como, se a prestação do serviço é realizada em um único ponto de atendimento e de que forma (com prioridades ou não, etc.)

Dentro dessa concepção, tal teoria estabelece modelos que descrevem um sistema de fila por meio de indicadores. Dentre eles, média de cliente no sistema (compreende fila e atendimento), média de clientes na fila, tempo médio de permanência de um cliente no sistema, tempo médio de espera na fila e probabilidade de haver determinado número de clientes no sistema.

Entre os modelos existentes, considera-se como o mais simples o que se denota de forma simplificada pela simbologia M/M/1. A qual se traduz na concepção de que os tempos entre chegadas, bem como os tempos entre atendimentos, seguem distribuições exponenciais independentes. E além disso, que a população de clientes não é limitada, e que os atendimentos ocorrem em um único ponto por ordem de chegada.

Essencialmente, supõe-se em tal modelo a existência de um processo de Poisson caracterizando as chegadas e as saídas dos clientes. E, a partir desta concepção, pode-se deduzir, conforme se mostra em [15], que a probabilidade de haver n clientes no sistema $P(n)$ é dada por $P(n) = \rho^n(1-\rho)$, em que $\rho = \dfrac{\lambda}{\mu}$, sendo λ e μ, as taxas de chegadas e de atendimentos, respectivamente.

Observa-se ainda que tais relações fazem sentido se $\rho < 1$. Situação descrita como sistema em estado estacionário. Do contrário, teria-se uma situação em que as taxas ao menos seriam iguais gerando uma tendência de fila crescente, conforme consta na mesma referência.

Decorre das relações postas que a probabilidade do sistema está desocupado ($n = 0$) é $P(0) = 1 - \rho$. Logo, que a de está ocupado é $P(N \geq 1) = \rho$. Em que N denota a variável número de pessoas no sistema.

Ainda de acordo com [15], o tempo de espera no modelo em questão segue distribuição exponencial com parâmetro $(\mu - \lambda)$. Onde denotando esse tempo por W, tem-se como fdp: $f(w) = (\mu - \lambda)e^{-(\mu-\lambda)w}$. Consequentemente, o tempo médio de espera no sistema é dado por $E(W) = \dfrac{1}{\mu - \lambda}$, não incluindo a possibilidade do sistema ter sido encontrado desocupado. Situação em que a média passa a ser descrita por $E(W) = \dfrac{\lambda}{\mu(\mu - \lambda)}$.

Demonstração. Note que podemos chegar a esse resultado, subtraindo de $\dfrac{1}{\mu - \lambda}$ o tempo médio de atendimento $\dfrac{1}{\mu}$, que é único tempo necessário ao se encontrar o sistema desocupado. □

A esse respeito considere a seguinte ilustração :

Exemplo 10. *Numa loja de auto atendimento apenas um caixa encontra-se disponível em certo dia. Sabe-se que, em média, chegam 6 pessoas a cada 4 minutos, enquanto o atendimento médio é de 7 pessoas a cada 4minutos. Supondo que os tempos entre chegadas e os tempos de atendimentos seguem distribuições exponenciais com as respectivas taxas informadas, determine: A) O tempo médio de espera nesse sistema; B) O número médio de pessoas no sistema; C) A probabilidade de haver fila no sistema.*

Dadas as suposições, podemos tomar o modelo M/M/1 para caracterizar o sistema em questão. Logo, conforme exposto, o tempo de permanência no sistema (W) segue distribuição exponencial com parâmetro $\mu - \lambda$. Assim, $E(W) = \dfrac{1}{\mu - \lambda} = 4$min representa o tempo médio de espera no sistema. Como a taxa média de chegada é $\lambda = 1,5$ pessoas /min. Segue que $\dfrac{\lambda}{\mu - \lambda} = 6$ representa o número médio de pessoas no sistema.

Por fim, a probabilidade de haver fila é $P(N \geq 2)$, onde N representa o número de pessoas no sistema. Visto que $P(n) = \rho^n(1 - \rho)$, decorre que $P(N \geq 2) = 1 - [P(0) + P(1)]$. Logo, $P(N \geq 2) = 1 - \left[\dfrac{1}{7} + \dfrac{6}{49}\right] = \dfrac{36}{49} \approx 73,47\%$.

Ainda podemos mencionar que diversas publicações relatam o emprego da distribuição exponencial em estudos meteorológicos, como precipitação pluviométrica. Exemplo pode ser encontrado em[16], onde se conclui que este modelo foi o mais adequado para descrever as chuvas na região centro sul do Ceará nos períodos considerados secos.

5.2 Sobre a Função Gama

Compreendida como extensão do conceito de fatorial para além dos inteiros não negativos, sua definição mais usual remonta a 1730. É atribuída a Euler e faz uso de integral imprópria. Entretanto, a depender da finalidade ou da complexidade do estudo, o emprego de limite ou de produtório pode ser feito como opções de definição creditadas a Gaus e a Weierstrass, respectivamente[17].

Neste trabalho, faremos uso da definição via integral imprópria por entendermos ser a mais adequada aos objetivos construídos.

Definição 42. *Chama-se função gama,* $\Gamma : \mathbb{R}_+ \to \mathbb{R}$*, a que se expressa por*

$$\Gamma(x) = \int_0^\infty e^{-t}t^{x-1}dt.$$

Nesta definição, especificamos como domínio o conjunto de valores de x para os quais a integral sempre converge. Contudo, conforme se argumenta em [18], tal domínio pode ser estendido a $\mathbb{R} \setminus \mathbb{Z}_-$.

A esse respeito, observamos que uma análise mais completa desse domínio passa pela percepção de que

$$\int_0^\infty e^{-t}t^{x-1}dt = \int_0^1 e^{-t}t^{x-1}dt + \int_1^\infty e^{-t}t^{x-1}dt.$$

Onde a primeira integral da soma é também imprópria porque o integrando é não limitado em $t = 0$. E assim, verificar a convergência da integral que defini $\Gamma(x)$, consiste em verificar a convergência de ambas as integrais desta soma.

Nesse sentido, nos limitaremos a seguinte argumentação:

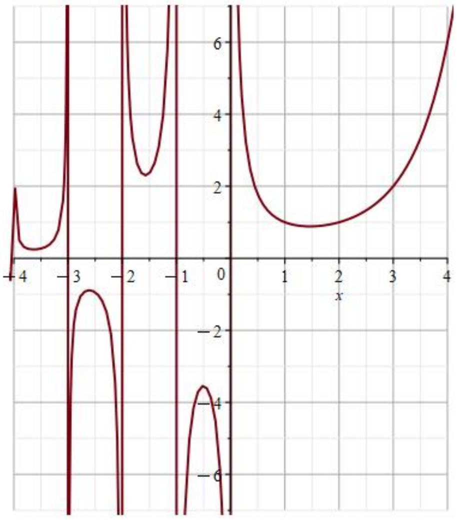
Demonstração. Note que para $x > 0$, $\int_0^1 e^{-t}t^{x-1}dt$ converge. Pois, pelo critério da comparação, $f(t) = t^{x-1}e^{-t} < h(t) = t^{x-1}$. E $\lim_{a \to 0^+} \int_a^1 t^{x-1} dx$ converge para $\frac{1}{x}$. Além disso, como é sabido o crescimento exponencial supera o polinomial, de onde afirmamos que $\int_1^\infty e^{-t}t^{x-1}dt$ também converge. De outro modo, note que $e^t > \frac{t^n}{n!}, \forall n$. Resultando que $e^{-t} < \frac{n!}{t^n}$. Escolhendo n, tal que $n > x$, temos que $t^{x-1}e^{-t} < t^{x-1}\frac{n!}{t^n} = \frac{n!}{t^{n-x+1}}$. Como $\int_1^\infty \frac{n!}{t^{n-x+1}}dt$ converge, pelo critério da comparação chegamos a mesma conclusão. Por outro lado, observe que para $x < 0 \Rightarrow x - 1 < -1 \Rightarrow t^{x-1} > t^{-1} \Rightarrow e^{-t}t^{x-1} > e^{-t}t^{-1}$, com $0 < t < 1$. Decorre, também pelo critério da comparação, que $\int_0^1 e^{-t}t^{x-1}dt$ diverge. Pois, usando integração por partes verifica-se que $\int_0^1 e^{-t}t^{-1}dt$ diverge. Já para $x = 0$, recaímos nesta última análise, chegando-se ao mesmo resultado.

$\square$

Além disso, podemos encontrar na literatura definições que estende o domínio da função gama aos números complexos com parte real positiva, a exemplo do que fez um estudo sobre cálculo fracionário [17]. Areá do cálculo para a qual está função é apontada como importante.

A figura a seguir traz o esboço do gráfico da função e nos permite ter a percepção de onde a mesma apresenta divergência. Informação representada pelas assíntotas verticais. Algo que corrobora com o já comentado sobre seu domínio.

Figura 5.6: Gráfico de $\Gamma(x)$



Fonte:Ilustração do autor

A partir de algumas proposições, buscaremos apresentar alguns dos resultados importantes advindos desta função, bem como, evidenciar a propriedade que lhe confere o título de extensão do conceito de fatorial, conforme tratado em [09].

Proposição 5. *Se* $x \in \mathbb{R}_+$, *então* $\Gamma(x+1) = x\Gamma(x)$.

Essencialmente, essa proposição traz a propriedade da recorrência da função gama. A qual se restringida aos naturais, permite obter a definição convencional de fatorial. Razão pela qual os demais valores obtidos pela função são referenciados como fatoriais.

Demonstração. Tomando $x \in \mathbb{R}; x > 0$, pela definição, temos que

$$\Gamma(x+1) = \int_0^\infty e^{-t}t^x dt = \lim_{a\to\infty}\int_0^a e^{-t}t^x dt.$$

Integrando por partes, temos que

$$\int_0^a e^{-t}t^x dt = -a^x e^{-a} + x\int_0^a e^{-t}t^{x-1}dt.$$

Logo ,

$$\lim_{a\to\infty}\int_0^a e^{-t}t^x dt = \lim_{a\to\infty}\left[-a^x e^{-a} + x\int_0^a e^{-t}t^{x-1}dt\right] =$$

$$\lim_{a\to\infty}\left[-a^x e^{-a}\right] + x\int_0^\infty e^{-t}t^{x-1}dt = \lim_{a\to\infty}\left[-a^x e^{-a}\right] + x\Gamma(x).$$

Resta mostrarmos que $\lim_{a\to\infty}\left[-a^x e^{-a}\right] = 0$.

Se $x \in \mathbb{N}$, por uma sequência finita de aplicações da regra de L'Hôpital, obtemos o resultado desejado. Do contrário, pela propriedade Arquimediana dos reais, sempre existirá $m \in \mathbb{N}$, tal que $m > x$. Logo, $m > x \Rightarrow \dfrac{a^m}{e^a} > \dfrac{a^x}{e^a} > 0$. Como $\lim_{a\to\infty}\dfrac{a^m}{e^a} = 0$, por argumento já usado; pelo Teorema do Confronto, conclui-se que $\lim_{a\to\infty}\left[-a^x e^{-a}\right] = 0$.

□

Proposição 6. $\Gamma(1) = 1$.

Demonstração. Pela definição,

$$\Gamma(1) = \int_0^\infty e^{-t}dt \Rightarrow \Gamma(1) = \lim_{a\to\infty}\left[\frac{-1}{e^t}\right]_0^a = \lim_{a\to\infty}\left[\frac{-1}{e^a}+1\right] = 1.$$

□

Proposição 7. $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

Demonstração.

$$\Gamma(x) = \int_0^\infty e^{-t}t^{x-1}dt \Rightarrow \Gamma\left(\frac{1}{2}\right) = \int_0^\infty e^{-t}t^{-\frac{1}{2}}dt.$$

Fazendo $u^2 = t$, obtemos que $dt = 2udu$, de onde segue que

$$\Gamma\left(\frac{1}{2}\right) = 2\int_0^\infty e^{-u^2}du.$$

Neste ponto, usaremos de um artifício que nos permita fazer uso de integrais duplas. Observando que $f(u) = e^{-u^2}$ é uma função par, temos que $\int_{-\infty}^{\infty} e^{-u^2} du = 2\int_0^{\infty} e^{-u^2} du$. Assim, tomando por R o resultado desejado,

$$R = \int_{-\infty}^{\infty} e^{-u^2} du \Rightarrow R^2 = \int_{-\infty}^{\infty} e^{-u^2} du \int_{-\infty}^{\infty} e^{-u^2} du = \int_{-\infty}^{\infty} e^{-x^2} dx \int_{-\infty}^{\infty} e^{-y^2} dy.$$

Como cada função depende apenas de uma das variáveis, temos que

$$R^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(x^2+y^2)} dx dy.$$

Usando coordenadas polares: $x = r\cos\theta$ e $y = r\sin\theta$.

Como $-\infty < x < \infty$ e $-\infty < y < \infty$, temos que $0 \leq r < \infty$ e $0 \leq \theta \leq 2\pi$.

Decorre que

$$R^2 = \int_0^{2\pi} \int_0^{\infty} e^{-r^2} r dr d\theta \Rightarrow R^2 = \int_0^{2\pi} \lim_{a\to\infty} \left[\frac{-e^{-r^2}}{2}\right]_0^a d\theta \Rightarrow R^2 = \int_0^{2\pi} \frac{1}{2} d\theta.$$

De onde concluímos que

$$R = \sqrt{\pi}.$$

□

De posse das proposições 5 e 6, e fazendo uso da convenção de $0! = 1$, podemos mostrar que $\Gamma(n) = (n-1)!$, $\forall n \in \mathbb{N}$.

Demonstração. Por indução infinita.

i) Para $n = 1$, temos que $\Gamma(1) = 0!$,que é verdade pela proposição 6.

ii)Suponha que para algum $n > 1$, $\Gamma(n) = (n-1)!$. Desejamos mostrar que vale para $n + 1$.

$\Gamma(n+1) = n\Gamma(n)$ pela proposição 5. E pela hipótese de indução, temos que $\Gamma(n) = (n-1)!$. Logo, $\Gamma(n+1) = n(n-1)! = n!$. Por indução finita, concluí-se que $\Gamma(n) = (n-1)!$, $\forall n \in N$. $\square$

Fazendo uso das proposições 5 e 7, podemos evidenciar a extensão do domínio da função Gama para valores negativos e não inteiros, partindo da análise do caso em que $-1 < x < 0$ e prosseguindo para os demais intervalos abertos de comprimento 1, conforme é exposto em [19]. A título de ilustração, veja que

$$\Gamma\left(\frac{1}{2}\right) = -\frac{1}{2}\Gamma\left(-\frac{1}{2}\right) \Rightarrow \Gamma\left(-\frac{1}{2}\right) = -2\sqrt{\pi}.$$

Proposição 8. *Se $r > 0$, então* $\int_0^\infty e^{-rt}t^{x-1}dt = \dfrac{\Gamma(x)}{r^x}$.

Demonstração. Fazendo $t = \dfrac{u}{r}$, temos que $dt = \dfrac{1}{r}du$. Segue que

$$\int_0^\infty e^{-rt}t^{x-1}dt = \int_0^\infty e^{-u}\frac{u^{x-1}}{r^{x-1}}\frac{1}{r}du = \frac{1}{r^x}\int_0^\infty e^{-u}u^{x-1}du = \frac{\Gamma(x)}{r^x}.$$

$\square$

Até aqui fizemos uma breve apresentação da função Gama buscando priorizar aspectos que sejam de maior relevância aos objetivos deste trabalho. Neste sentido, buscaremos, a partir desse momento, situar sua presença na estruturação e aplicação de modelos probabilísticos, amplamente utilizado pela Estatística.

Modelo Gama

Antes da conceituação que vem a seguir, consideramos pertinente um esclarecimento sobre duas terminologias. Por distribuição Gama Generalizada(GG), entende-se o modelo probabilístico de três parâmetros proposto em 1962 por Stacy. O qual possui como

fdp a seguinte relação, conforme consta em [20].

$$f(t) = \frac{\tau}{\alpha\Gamma(k)}\left(\frac{t}{\alpha}\right)^{\tau k-1}\exp\left[-\left(\frac{t}{\alpha}\right)^{\tau}\right], \text{com}\, t > 0, \alpha > 0, \tau > 0\, \text{e}\, k > 0.$$

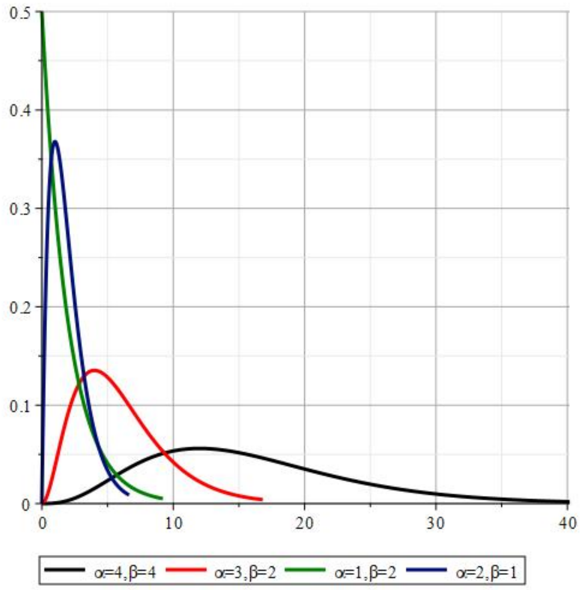
Deste derivam modelos como Exponencial, Weibul, Quiquadrado, Rayleigh, Log-Normal e o Gama como apresentaremos a seguir com dois parâmetros. Razão pela qual o mesmo costuma ser designado por distribuição Gama incompleta. E a expressão Distribuição Gama é usada para fazer referência a uma família de distribuições que derivam da GG.

Definição 43. *Denotamos por $X \sim Gama(\alpha, \beta)$ a variável aleatória contínua X que assume valores positivos e dizemos que segue distribuição gama, com parâmetros $\alpha > 0$ e $\beta > 0$, se sua função densidade de probabilidade é dada por*

$$f(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}.$$

Os parâmetros α e β são responsáveis pelo formato e escala da curva, respectivamente. A figura seguinte trás uma ilustração da variação dos parâmetros e evidencia o caráter assimétrico positiva da distribuição.

Figura 5.7: fdp da distribuição Gama



Fonte:Ilustração do autor

Convém mencionar que, a exemplo da distribuição exponencial, a fdp da distribuição Gama pode ser apresentada de outra forma a depender de como é feita a parametrização. A forma seguinte é encontrada com muita frequência na literatura:

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}.$$

A presença da função gama nesse modelo acarreta em vários ganhos operacionais e permite, de forma quase imediata, a percepção de relações deste com outros modelos de distribuição como será ilustrado.

Proposição 9. *Se $X \sim Gama(\alpha, \beta)$, então $E(X) = \dfrac{\alpha}{\beta}$.*

Demonstração. $E(X) = \int_0^{\infty} xf(x)dx \Rightarrow E(X) = \dfrac{\beta^{\alpha}}{\Gamma(\alpha)} \int_0^{\infty} x^{(\alpha+1)-1} e^{-\beta x} dx$. Usando as proposições 8 e 5, segue que:

$$E(X) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \frac{\Gamma(\alpha+1)}{\beta^{\alpha+1}} \Rightarrow E(X) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \frac{\alpha}{\beta} \frac{\Gamma(\alpha)}{\beta^{\alpha}} \Rightarrow E(X) = \frac{\alpha}{\beta}.$$

$\square$

Proposição 10. *Se $X \sim Gama(\alpha, \beta)$, então $V(X) = \frac{\alpha}{\beta^2}$.*

Demonstração. Como $V(X) = E(X^2) - [E(X)]^2$, precismos obter $E(X^2)$.

$$E(X^2) = \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} x^2 x^{\alpha-1} e^{-\beta x} dx \Rightarrow E(X^2) = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{(\alpha+2)-1} e^{-\beta x} dx \Rightarrow$$

(pela proposição 8)

$$E(X^2) = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+2)}{\beta^{\alpha+2}} \Rightarrow E(X^2) = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{(\alpha+1)\Gamma(\alpha+1)}{\beta^{\alpha+2}} \Rightarrow$$

$$E(X^2) = \frac{(\alpha+1)\alpha\Gamma(\alpha)}{\Gamma(\alpha)\beta^2} \Rightarrow E(X^2) = \frac{\alpha^2+\alpha}{\beta^2}.$$

Logo , $V(X) = \frac{\alpha^2+\alpha}{\beta^2} - \frac{\alpha^2}{\beta^2} = \frac{\alpha}{\beta^2}$.

$\square$

Nestas ultimas demostrações, ilustramos um pouco do poder de síntese que as propriedades da função Gama confere aos cálculos. Sejam eles de natureza algébrica geral ou numérica específica, essa é uma característica da função que vai além do contexto probabilístico e estatístico. Conforme mostrado em [18].

Observamos ainda, que ao tomarmos $\alpha = 1$, obtemos, de forma imediata, a fdp do modelo exponencial, bem como, o valor esperado e a variância. Caracterizando-o como caso particular da distribuição Gama.

Aplicabilidade da distribuição Gama

Além do já exposto sobre a distribuição exponencial, que é caso particular da distribuição Gama, a literatura registra uso da distribuição Gama incompleta ou generalizada em estudos meteorológicos, análise de confiabilidade e tempos de resposta de equipamentos eletrônicos, duração de estoque, e até análise de sobrevivência de pacientes com câncer.

Nesses contextos, os trabalhos mais recentes que encontramos foram: Análise das sequências de dias chuvosos nas capitais brasileiras(2021)[21]; Discriminação de cenários pluviométricos do estado da Paraíba(2016)[22]; Cálculo de estoque mínimo para itens industriais utilizando as distribuições de probabilidade normal, poisson e gama (2021)[23]; Análise de confiabilidade aplicada na otimização de sistemas de produção animal(2017)[24]; Modelos paramétricos de sobrevivência aplicados a dados de câncer (2020)[25] e Modelos de fragilidade aplicados a análise de fatores contribuintes na morte de pacientes portadores de leucemia(2020)[26].

Nota-se que a distribuição Gama está relacionada a estudos cujos os resultados tem reflexos em diversos aspectos da vida social, econômica e até politica. Uma vez que informações fundamentadas a respeito de condições climáticas de regiões, sobre ações que possam otimizar a produção industrial, ou mesmo sobre fatores que se relacionam com saúde pública, tudo isso constitui conhecimentos que podem embasar políticas de aprimoramento, de prevenção ou de combate a problemas vigentes.

No que segue, exploraremos algumas situações para ilustrar operações com o modelo Gama.

Exemplo 11. *Em determinada cidade, o consumo diário de energia elétrica em milhões de quilowatts-hora pode ser tratado como uma variável aleatória com distribuição Gama com* $\alpha = 3$ *e* $\beta = \dfrac{1}{2}$. *Se a usina desta cidade tem capacidade diária de 12 milhões quilowatts-hora, qual a probabilidade de que essa fonte seja insuficiente em um determinado dia ?*

Definiremos X a variável consumo diário de energia em milhões de quilowartts-hora pela cidade, temos que $X \sim Gama\left(3, \dfrac{1}{2}\right)$. Desejamos obter $P(X > 12)$. Logo, pela probabilidade complementar,

$$P(X > 12) = 1 - P(X \leq 12).$$

$$P(X \leq 12) = \frac{1}{8\Gamma(3)} \int_0^{12} x^2 e^{\frac{-x}{2}} dx = \frac{1}{16} \int_0^{12} x^2 e^{\frac{-x}{2}} dx.$$

Fazendo $u = \dfrac{x}{2} \Rightarrow dx = 2du$. Assim,

$$\frac{1}{16} \int_0^{12} x^2 e^{\frac{-x}{2}} dx = \frac{1}{2} \int_0^{6} u^2 e^{-u} du.$$

Integrando por partes, obtemos que

$$\frac{1}{2} \int_0^{6} u^2 e^{-u} du = -18e^{-6} + \int_0^{6} u e^{-u} du.$$

Usando mais uma vez integração por partes, chegamos a

$$-18e^{-6} + \int_0^{6} u e^{-u} du = -24e^{-6} + \int_0^{6} e^{-u} du = -25e^{-6} + 1.$$

Portanto , $P(X > 12) = 25e^{-6} \approx 0,062$.

Exemplo 12. *Se a vida útil de um componente eletrônico de uma impressora, em anos, tem distribuição Gama com média 8 e variância 16, determine a probabilidade de que esse componente tenha uma vida útil de pelo menos 11 anos.*

Seja X a vida útil desse componente. Temos que $X \sim Gama(\alpha, \beta)$, em que $E(X) = 8$ e $V(X) = 16$. Destas informações, decorre que $\alpha = 4$ e $\beta = \dfrac{1}{2}$. Temos por objetivo calcular $P(X \geq 11) = 1 - P(X < 11)$. Dada a natureza contínua da variável, podemos afirmar que $P(X < 11) = P(X \leq 11)$.

De posse dos dados, obtemos a fdp da distribuição $f(x) = \frac{1}{16\Gamma(4)} x^3 e^{\frac{-x}{2}} dx$. Logo,

$P(X \leq 11) = \frac{1}{96}\int_0^{11} x^3 e^{\frac{-x}{2}} dx.$

Usando integração por partes temos :

$$\frac{1}{96}\int_0^{11} x^3 e^{\frac{-x}{2}} dx = \frac{1}{96}\left[-2662 e^{\frac{-11}{2}} + 6\int_0^{11} x^2 e^{\frac{-x}{2}} dx\right] = \frac{-2662}{96} e^{\frac{-11}{2}} + \frac{1}{16}\int_0^{11} x^2 e^{\frac{-x}{2}} dx.$$

Usando mais uma vez integração por partes, segue que :

$$\frac{1}{96}\int_0^{11} x^3 e^{\frac{-x}{2}} dx = -\frac{2662}{96} e^{\frac{-11}{2}} - \frac{121}{8} e^{\frac{-11}{2}} + \frac{1}{4}\int_0^{11} x e^{\frac{-x}{2}} dx.$$

Novamente, usando integração por partes, obtemos:

$$\frac{1}{96}\int_0^{11} x^3 e^{\frac{-x}{2}} dx = \frac{-2662}{96} e^{\frac{-11}{2}} - \frac{121}{8} e^{\frac{-11}{2}} - \frac{11}{2} e^{\frac{-11}{2}} + \frac{1}{2}\int_0^{11} e^{\frac{-x}{2}} dx.$$

Resultando, que

$$\frac{1}{96}\int_0^{11} x^3 e^{\frac{-x}{2}} dx = -\frac{2662}{96} e^{\frac{-11}{2}} - \frac{121}{8} e^{\frac{-11}{2}} - \frac{11}{2} e^{\frac{-11}{2}} - e^{\frac{-11}{2}} + 1.$$

Logo,

$$P(X > 11) = 1 - \left[-\frac{2662}{96} e^{\frac{-11}{2}} - \frac{121}{8} e^{\frac{-11}{2}} - \frac{11}{2} e^{\frac{-11}{2}} - e^{\frac{-11}{2}} + 1\right] \Rightarrow$$

$$P(X > 11) = \frac{2369}{48} e^{\frac{-11}{2}} \approx 0,202.$$

5.3 Sobre a Função Beta

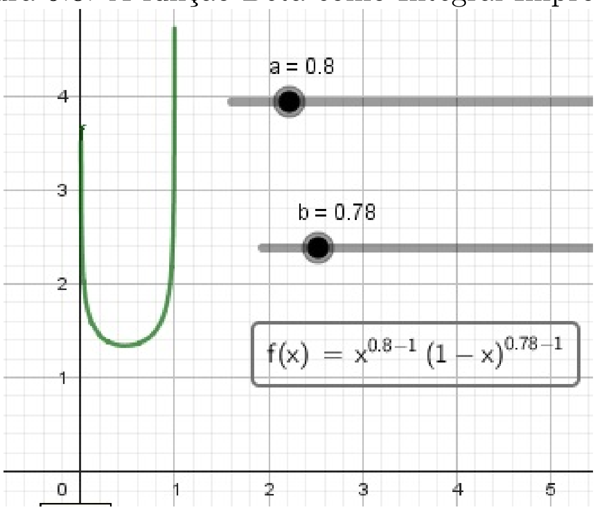
Ao lado da função Gama, a Beta compõe o que a literatura denomina integrais de Euler de segundo e primeiro tipo, respectivamente, ou ainda funções Eulerianas [17]. Semelhante a função Gama, seu estudo está bem relacionado ao cálculo fracionário e sua definição pode ser apresentada de várias formas. Neste trabalho, a definiremos para números reais positivos. Mas, observamos que é usual a extensão desse domínio para números complexos de parte real positiva.

Definição 44. *Chama-se função beta,* $B : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$*, a que se expressa por*

$$B(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt.$$

O domínio especificado é o que usualmente encontramos na literatura ao se definir tal função, a exemplo de [18] e [19]. Entretanto, não encontramos nestas, nem em outras fontes consultadas, menção ao fato de que para $x < 1$ a integral é imprópria, pois não é limitada em $t = 0$. Ocorrendo o mesmo para $y < 1$ em $t = 1$. A imagem da figura 5.8 ilustra essa observação.

Figura 5.8: A função Beta como Integral Imprópria

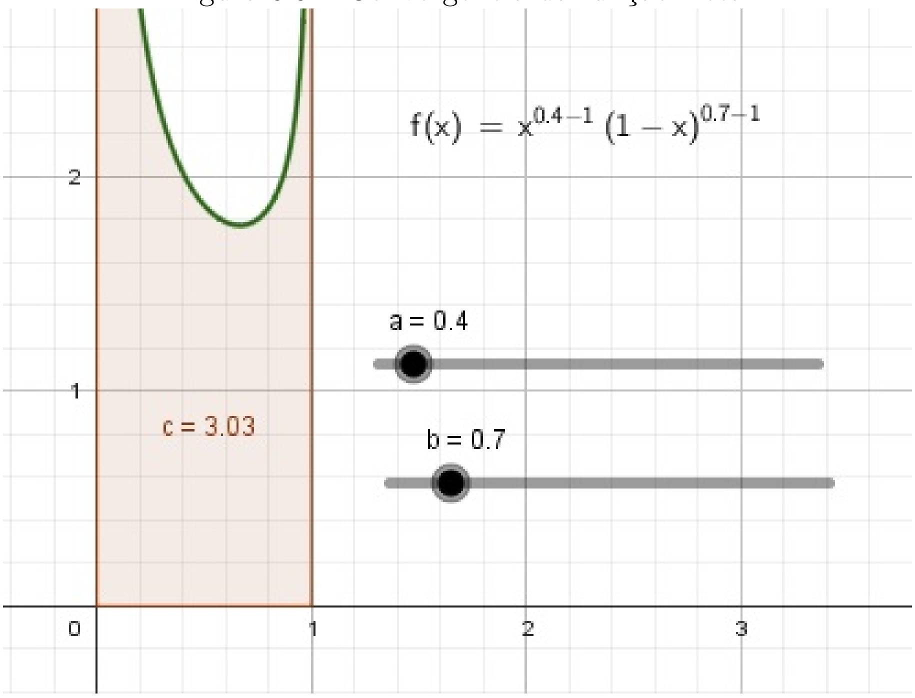


Fonte:Ilustração do autor

Fazendo uso do Geogebra, tomamos para os papeis de x, y controles deslizantes a e b, e observamos o comportamento do integrando no intervalo em questão para esses valores menores que 1. Temos então, evidência gráfica da sua não limitação pelo comportamento assintótico vertical.

Contudo, a integral é convergente para o domínio definido. Fixados valores menores que 1 para os controles deslizantes citados e integrando a função com uso do Geogebra, temos a confirmação de área sob a curva no intervalo $[0, 1]$, conforme mostra a figura 5.9.

Figura 5.9: Convergência da função Beta



Fonte:Ilustração do autor

Para uma análise algébrica, tomamos $0 < r < 1$ e expressamos a integral em questão como sendo

$$A = \int_0^r t^{x-1}(1-t)^{y-1}dt + B = \int_r^1 t^{x-1}(1-t)^{y-1}dt.$$

Analisando A para $0 < x < 1$, concluímos sua convergência.

Demonstração. De fato, tomando $f(t) = t^{x-1}(1-t)^{y-1}$ e $g(t) = \dfrac{1}{t^{1-x}}$, temos que $\lim_{t\to 0} \dfrac{f(t)}{g(t)} = \lim_{t\to 0}(1-t)^{y-1} = 1 > 0$.

Como
$$\int_0^r \frac{1}{t^{1-x}}dt$$
é converge, pois $1-x<1$. Pelo critério do limite do quociente, temos que $A=\int_0^r f(t)dt$ também converge. $\square$

Analisando B para $0<y<1$, também concluímos sua convergência.

Demonstração. Usando o mesmo critério, tomamos agora $f(t) = t^{x-1}(1-t)^{y-1}$ e $g(t) = \dfrac{1}{(1-t)^{1-y}}$, obtemos $\lim_{t\to 1} \dfrac{f(t)}{g(t)} = \lim_{t\to 1} t^{x-1} = 1 > 0$. Decorre que $\int_r^1 f(t)dt$ converge, pois
$$\int_r^1 \frac{1}{(1-t)^{1-y}}dt$$
também converge, visto que $1-y<1$. $\square$

A função Beta pode ser apresentada sob várias formas. E a escolha por uma ou outra, tem por premissa, em geral, atender alguma conveniência da situação. Contemplaremos algumas a seguir.

Proposição 11. *Se $x,y\in\mathbb{R}_+$, então vale que*
$$B(x,y) = 2\int_0^{\frac{\pi}{2}} \sin^{2x-1}\theta\cos^{2y-1}\theta d\theta.$$

Demonstração. Tomando $t=\sin^2\theta$ em $B(x,y)=\int_0^1 t^{x-1}(1-t)^{y-1}dt$, segue que $dt = 2\sin\theta\cos\theta d\theta$. Assim, $B(x,y) = 2\int_0^{\frac{\pi}{2}} (\sin^2\theta)^{x-1}(1-\sin^2\theta)^{y-1}\sin\theta\cos\theta d\theta$. De onde concluímos que $B(x,y) = 2\int_0^{\frac{\pi}{2}} \sin^{2x-1}\theta\cos^{2y-1}\theta d\theta$. $\square$

Proposição 12. *Se $x,y\in\mathbb{R}_+$, então*
$$B(x,y) = \int_0^\infty \frac{a^{x-1}}{(1+a)^{x+y}}da.$$

Demonstração. Fazendo $t = \frac{a}{(1+a)}$ em $B(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt$, temos $dt = \frac{1}{(1+a)^2}da$, quando $t \to 0$ implica em $a \to 0$, e para $t \to 1$ então $a \to \infty$. Decorre que

$$B(x, y) = \int_0^\infty \frac{a^{x-1}}{(a+1)^{x+y-2}} \frac{1}{(1+a)^2} da \Rightarrow B(x, y) = \int_0^\infty \frac{a^{x-1}}{(1+a)^{x+y}} da.$$

□

Proposição 13. *Se $x, y \in \mathbb{R}_+$, então*

$$B(x, y) = \frac{1}{a^{x+y-1}} \int_0^a z^{x-1}(a-z)^{y-1} dz.$$

Demonstração. Em $B(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt$, tomando $t = \frac{z}{a}$, temos que $dt = \frac{1}{a}dz$. Segue que

$$B(x, y) = \int_0^a \frac{z^{x-1}(a-z)^{y-1}}{a^{x+y-2}a} dz.$$

E assim, concluímos que

$$B(x, y) = \frac{1}{a^{x+y-1}} \int_0^a z^{x-1}(a-z)^{y-1} dz.$$

□

Proposição 14. *Se $x, y \in \mathbb{R}_+$, então*

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}.$$

Demonstração. Tomando $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$ e $\Gamma(y) = \int_0^\infty s^{y-1}e^{-s}ds$, temos que $\Gamma(x)\Gamma(y) = \int_0^\infty t^{x-1}e^{-t}dt \int_0^\infty s^{y-1}e^{-s}ds$. Fazendo $t = u^2, s = v^2 \Rightarrow dt = 2udu, ds = 2vdv$. Assim, $\Gamma(x)\Gamma(y) = 2\int_0^\infty u^{2x-1}e^{-u^2}du 2\int_0^\infty v^{2y-1}e^{-v^2}dv$. Como as integrais são em variáveis distintas, segue que:

$$\Gamma(x)\Gamma(y) = 4\int_0^{\infty}\int_0^{\infty} u^{2x-1}v^{2y-1}e^{-(u^2+v^2)}dudv.$$

Fazendo uso de coordenadas polares, tomando $u = r\cos\theta$, $v = r\sin\theta$, obtemos que:

$$\Gamma(x)\Gamma(y) = 4\int_0^{\infty}\int_0^{\frac{\pi}{2}} \cos^{2x-1}\theta\sin^{2y-1}\theta r^{2x+2y-2}re^{-r^2}d\theta dr \Leftrightarrow$$

$$\Gamma(x)\Gamma(y) = 2\int_0^{\frac{\pi}{2}} \cos^{2x-1}\theta\sin^{2y-1}\theta d\theta 2\int_0^{\infty} r^{2x+2y-2}re^{-r^2}dr.$$

Usando a proposição 11, temos que

$$\Gamma(x)\Gamma(y) = B(x,y)\int_0^{\infty}(r^2)^{x+y-1}2re^{-r^2}dr.$$

Fazendo $b = r^2 \Rightarrow db = 2rdr$. E,assim, obtemos que

$$\Gamma(x)\Gamma(y) = B(x,y)\int_0^{\infty} b^{x+y-1}e^{-b}db.$$

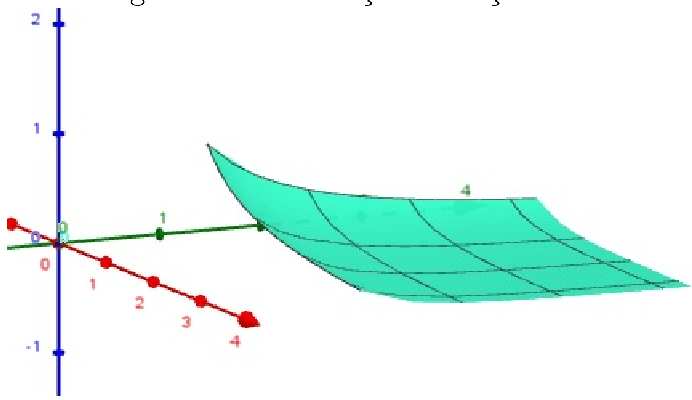
Logo,

$$\Gamma(x)\Gamma(y) = B(x,y)\Gamma(x+y).$$

□

Como mostrado, essa última proposição estabelece a conexão entre as funções Gama e Beta. Por meio desta torna-se mais fácil, por exemplo, inserir a função em softwares como Geogebra para esboçarmos o seu gráfico, como ilustra a figura5.10.

Figura 5.10: Esboço da função Beta



Fonte:Ilustração do autor

A articulação de propriedades de ambas funções permiti contornar dificuldades presentes em processos de integração, que de outro modo, apresentam-se morosos e com mais possibilidade de erros. Em [18], a autora ilustra algumas das possibilidades de uso destas funções.

A exemplo do que ocorre para as funções já contempladas, para esta também existe distribuição de probabilidade que carrega seu nome e presença na estruturação do modelo. Veremos que a última proposição possibilita simplificação das operações na obtenção de resultados importantes, como valor esperado e variância.

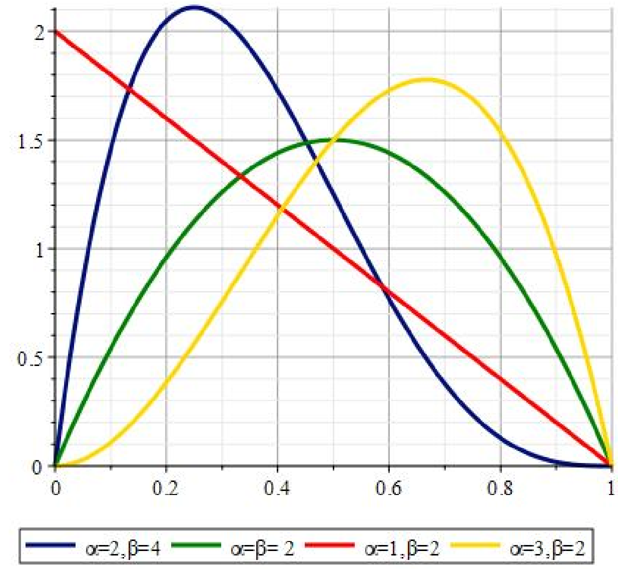
Modelo Beta

Definição 45. *Denotamos por* $X \sim Beta(\alpha, \beta)$ *a variável aleatória contínua* X *que assume valores em* $(0, 1)$ *e dizemos que a mesma possui distribuição beta com parâmetros* $\alpha > 0$ *e* $\beta > 0$*, se a sua função densidade de probabilidade for dada por*

$$f(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1}.$$

Os parâmetros da distribuição são adimensionais e, assim, são responsáveis pela forma da curva da distribuição. A curva pode ter assimetria positiva, negativa ou mesmo ser simétrica a depender dos parâmetros. A figura 5.11 ilustra isso.

Figura 5.11: fdp da distribuição Beta



Fonte:Ilustração do autor

Devido as diversas formas que a distribuição beta pode assumir ao modelar variáveis no intervalo citado, é comum também referi-se a mesma como família de distribuições em tal intervalo. Além disso, conforme consta em [27], a partir de transformações convenientes, variáveis que assumem valores em intervalos finitos (a, b) podem ser descritas pelo modelo beta, uma vez considerando sua fdp sob a forma geral

$$f(y) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{(y-a)^{\alpha-1}(b-y)^{\beta-1}}{(b-a)^{\alpha+\beta-2}},$$

e tomando $x = \dfrac{y-a}{b-a}$, com $a < y < b$, e $\alpha, \beta > 0$. O que nos leva a forma padrão da fdp com $X \sim Beta(\alpha, \beta)$, e $x \in (0, 1)$.

Demonstração. De fato, $x = \dfrac{y-a}{b-a} \Rightarrow y = x(b-a) + a(1) \Rightarrow b - y = (b-a)(1-x)(2)$. Substituindo (1) e (2) na forma geral, obtemos

$$f(y)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{x^{\alpha-1}(b-a)^{\alpha-1}(b-a)^{\beta-1}(1-x)^{\beta-1}}{(b-a)^{\alpha+\beta-2}}=\frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}.\qquad\square$$

Proposição 15. *Se* $X\sim Beta(\alpha,\beta)$*, então* $E(X)=\dfrac{\alpha}{\alpha+\beta}$.

Demonstração.

$$E(X)=\int_0^1 xf(x)dx\Rightarrow E(X)=\frac{1}{B(\alpha,\beta)}\int_0^1 x^{(\alpha+1)-1}(1-x)^{\beta-1}dx$$

$$\Rightarrow E(X)=\frac{1}{B(\alpha,\beta)}B(\alpha+1,\beta)\Rightarrow$$

(pela proposição 14)

$$E(X)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)}\Rightarrow$$

$$E(X)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)}\frac{\Gamma(\alpha+1)}{\Gamma(\alpha+\beta+1)}\Rightarrow E(X)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)}\frac{\alpha\Gamma(\alpha)}{(\alpha+\beta)\Gamma(\alpha+\beta)}\Rightarrow E(X)=\frac{\alpha}{\alpha+\beta}.$$

$\square$

Proposição 16. *Se* $X\sim Beta(\alpha,\beta)$*, então* $V(X)=\dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

Demonstração.

$$E(X^2)=\frac{1}{B(\alpha,\beta)}\int_0^1 x^{(\alpha+2)-1}(1-x)^{\beta-1}dx\Rightarrow E(X^2)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(\alpha+2)\Gamma(\beta)}{\Gamma(\alpha+\beta+2)}\Rightarrow$$

$$E(X^2)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)}\frac{\Gamma(\alpha+2)}{\Gamma(\alpha+\beta+2)}\Rightarrow E(X^2)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)}\frac{(\alpha+1)\alpha\Gamma(\alpha)}{(\alpha+\beta+1)(\alpha+\beta)\Gamma(\alpha+\beta)}$$

$$\Rightarrow E(X^2)=\frac{(\alpha+1)\alpha}{(\alpha+\beta+1)(\alpha+\beta)}.$$

$$V(X)=E(X^2)-[E(X)]^2\Rightarrow V(X)=\frac{(\alpha+1)\alpha}{(\alpha+\beta+1)(\alpha+\beta)}-\frac{\alpha^2}{(\alpha+\beta)^2}\Rightarrow$$

$$V(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

$\square$

Aplicabilidade da Distribuição Beta

De um modo geral, esse modelo é utilizado para descrever o comportamento de variáveis aleatórias contínuas que assumem valores no intervalo (0,1). Porém, através de parametrização adequada, variáveis que assumem valores em outros intervalos finitos tem comportamentos descritos por esse modelo em diferentes ramos da ciência. Neste sentido, encontramos aplicações relacionadas a biodegradação da matéria orgânica natural e a modelagem de trajetórias baseadas em grupos.

Com relação a biodegradação da matéria orgânica, pudemos ver em [28] a divulgação de um estudo sobre a biodegradação do carbono orgânico dissolvido em água. Neste estudo, os autores partem da premissa de que a matéria orgânica natural é a principal fonte de carbono. E assim, que o estudo da biodegradação do carbono orgânico é utilizado para medir a biodegradação da matéria orgânica. Além disso, que a maior concentração do carbono está na superfície terrestre, seguida das frações restantes presentes em cedimentos e no oceano.

Em tal estudo, fez-se o uso da distribuição beta para expressar a medida da biodegradação do carbono orgânico dissolvido em água através da probabilidade da biodegradação, de modo a ser possível transformar essa probabilidade em um coeficiente de decaimento do carbono. E ao comparar os resultados obtidos com outros cinco modelos de estudo adotados até então, dentre eles um envolvendo o modelo Gama, os autores concluíram que o modelo beta, a exemplo do modelo Gama, descreve bem a cinética da biodegradação. E além disso, que o mesmo se destaca em poder preditivo. Entendido como a capacidade de fazer extrapolações para o futuro.

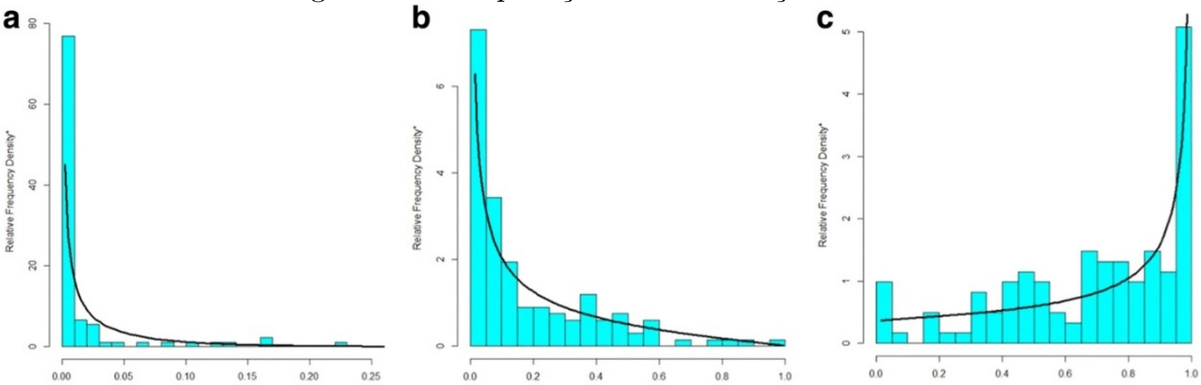
Por tudo isso, os autores defendem o uso do método aplicado para também estudar a a biodegradação da matéria orgânica natural presente na superfície terrestre, e em sedimentos. Pois, julgam estarem de posse de resultados consistentes obtidos por meio da distribuição beta. O estudo ainda é justificado como ação que contribui para o conhecimento da produtividade de ecossistemas, algo que está relacionado com a mineralização da matéria orgânica natural em CO_2 por ação de microrganismos: O que se entende por biodegradação.

Já a modelagem de trajetórias baseadas em grupos, pode ser compreendida como técnica especial empregada para estudar a evolução de comportamentos de indivíduos em situações semelhantes. Por exemplo, em pesquisas biomédicas faz-se uso dessa técnica para estudar a evolução de doenças ou da adesão a determinados tratamentos, conforme consta em [29]. Nesta mesma referência, tivemos conhecimento de estudo que usou a distribuição beta para descrever o nível de atividade neurológica de pacientes em coma, após sofrerem parara cardíaca e serem reanimados.

Pode ser constatado, ainda de acordo com [29], que a distribuição beta, através de parametrização adequada, pode descrever, com melhor adequação que o modelo normal, a taxa de supressão da atividade neurológica de 396 pacientes. Pessoas que após sofrerem paradas cardíacas e terem sido reanimadas, tiveram seus níveis de atividades neurológicas monitorados enquanto estiveram em coma em um determinado hospital.

A figura 5.12 mostra as distribuições da taxa de supressão real de três grupos de pacientes em comparação com suas respectivas distribuições previstas pelo modelo beta na hora 24, após serem reanimados.

Figura 5.12: Aplicação da distribuição beta



Fonte:[29]

Ainda, segundo mesmo estudo, estimou-se que pacientes com o comportamento do grupo a), em que a taxa de supressão decresce rápido e mantém-se baixa na maior parte do tempo, a probabilidade de sobrevivência é de 69,8%; Para o perfil do grupo b) é de 29,6%; Já para o grupo c), no qual a taxa mantém-se alta, a probabilidade é de 2,3%.

Prosseguimos com algumas ilustrações de operações com o modelo contemplado.

Exemplo 13. *Um posto de combustível de uma fazenda de soja tem seu tanque de gasolina completado sempre a cada segunda feira. Se foi observado que o consumo de gasolina na semana segue uma distribuição Beta com $\alpha = 4$ e $\beta = 2$. Qual a probabilidade de que em alguma semana o consumo seja de pelo menos 90%?*

Definindo a variável consumo de gasolina semanal por $X \sim Beta(4,2)$, temos que $f(x) = \dfrac{1}{B(4,2)}x^3(1-x)$. Como desejamos $P(X \geq 0,9)$ e $B(4,2) = \dfrac{1}{20}$, segue que

$$P(X \geq 0,9) = 1 - P(X < 0,9) \Rightarrow P(X \geq 0,9) = 1 - 20\int_0^{0,9} x^3(1-x)dx \Rightarrow$$

$$P(X \geq 0,9) = 1 - 20\left[\frac{x^4}{4} - \frac{x^5}{5}\right]_0^{0,9} = 1 - 20\left(\frac{0,9^4}{4} - \frac{0,9^5}{5}\right) \approx 0,08.$$

Exemplo 14. *Em uma linha de produção em larga escala, a taxa de defeitos de televisores é modelada por uma distribuição Beta com parâmetros $\alpha = 3$ e $\beta = 6$. Qual é a probabilidade de que até 25% dos televisores sejam defeituosos?*

Considerando X a variável taxa de defeitos de televisores, temos que $X \sim Beta(3,6)$ e desejamos obter $P(X \leq 0,25)$. Observando que $B(3,6) = \dfrac{\Gamma(3)\Gamma(6)}{\Gamma(9)} = \dfrac{1}{168}$, segue da forma geral que a fdp desta distribuição é dada por $f(x) = 168x^2(1-x)^5$.
Decorre que

$$P(X \leq 0,25) = 168\int_0^{0.25} x^2(1-x)^5 dx = 168\int_0^{0,25} x^2 - 5x^3 + 10x^4 - 10x^5 + 5x^6 - x^7 dx \Rightarrow$$

$$P(X \leq 0,25) = 168\left[\frac{x^3}{3} - \frac{5x^4}{4} + 2x^5 - \frac{5x^6}{3} + \frac{5x^7}{7} - \frac{x^8}{8}\right]_0^{0.25} \Rightarrow P(X \leq 0,25) \approx 0,3214.$$

5.4 Um olhar para o currículo do Ensino Médio

Nesta seção, fazemos algumas considerações sobre a presença e/ou possibilidade de contemplação de conceitos como variáveis aleatórias e modelos probabilísticos discretos no currículo do ensino médio. Também fazemos menção a presença dos temas probabilidade e estatística nesta etapa de ensino. Buscamos fundamentar as colocações na leitura de documentos que regem a educação escolar no país e outras obras afins. Bem como, na análise de algumas questões extraídas de livros didáticos e do ENEM.

Sob o olhar de muitos alunos, até mesmo de professores, contagem, probabilidade e estatística são, historicamente, tidos como temas difíceis no currículo básico do Brasil. Essa é uma constatação clara, por exemplo, para o PAPMEM-Programa de Aperfeiçoamento de Professores de Matemática do Ensino Médio que vem contemplando tais conteúdos em todas as suas edições, dada a grande solicitação pelos docentes.

Felizmente, essa visão vem sendo transformada graças a programas como o citado e o profmat que tem contribuído para preencher lacunas nas formações de professores. Entretanto, tal problemática ainda é uma realidade no Brasil. Não entramos na discussão sobre as possíveis causas, pois certamente não a esgotaríamos aqui, ainda que fosse esse um dos objetivos deste trabalho. Razão pela qual apenas fazemos menção como forma de melhor contextualizar exposições que vem a seguir.

No que diz respeito a legislação que orienta ou normatiza o currículo no Brasil, partindo da LDB(1996), tivemos PCN, PCN+, DCNEM e mais recentemente a BNCC-Base nacional comum curricular implantada em 2018. E o NEM-novo ensino médio em implantação neste ano de 2023.

De modo geral, depreende-se dessas legislações uma organização curricular flexível que busque promover a formação integral dos educando-os, a valorização de experiências prévias e a articulação dos saberes. Que seja pautado no desenvolvimento de competências e habilidades. E que compreenda ações baseadas em investigação, projetos e atividades interdisciplinares.

No tocante ao tema probabilidade, destacamos a seguinte habilidade presente na BNCC que, em síntese, representa um dos objetivos a ser alcançado no aprendizado dos discentes: "(EM13MAT511)Reconhecer a existência de diferentes tipos de espaços amostrais, discretos ou não, e de eventos, equiprováveis ou não, e investigar implicações no cálculo de probabilidades"[30].

De posse dessa contextualização, externamos algumas concepções a respeito do currículo básico e dos temas em destaque, a partir de breve análise de exercícios presentes em livros didáticos.

Algumas percepções

Em geral, percebemos nos livros didáticos do ensino básico, os quais são ofertados pelo PNLD(Programa Nacional do Livro Didático), uma ênfase a exercícios que versão sobre um mesmo aspecto da teoria, passando a ideia de que os temas se limitam a determinado ponto de vista. Em muitos casos, nota-se que nem ao menos se faz menção a possíveis relações que podem ser estabelecidas com outros temas.

Vários são os exercícios de probabilidade que possibilitam uma introdução, ao menos intuitiva, do conceito de variável aleatória e de modelos probabilísticos discretos como os já contemplados. A título de ilustração, apresentamos alguns exercícios extraídos de dois livros cujas coleções foram aprovadas nos últimos dois PNLD.

Exemplo 15. *Um dado não viciado é lançado duas vezes, sucessivamente. Seja E"a soma dos pontos obtidos é menor ou igual a 9". Determine "E"(Fonte: [31])*

Esse exercício aparece após as definições de espaço amostral, eventos e outras de caráter preliminar.

É possível que algum discente tenha visualizado o espaço amostral como sendo $\{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$. E, assim, entendido que o evento "E"seja $\{2, 3, 4, 5, 6, 7, 8, 9\}$. Na concepção de um professor preso ao material, corre-se o risco de dizer que esta solução esteja errada. Pois, o direcionamento da explicação que precede o exercício deve encaminhar o discente apenas para espaços equiprováveis.

Além da oportunidade de já ir construindo a compreensão de espaços não equiprováveis, vemos nesse exercício também a possibilidade de introduzir, ao menos de forma intuitiva, a noção de variável aleatória discreta. Pois afinal, da forma como descritos, espaço e evento diz respeito a variável soma dos pontos nos lançamentos do dado.

Exemplo 16. *Pretende-se organizar dois eventos no ano; três cidades do Sudeste, duas do Sul e cinco do Nordeste candidataram-se a sede desses eventos. Sabendo que* *uma mesma cidade não pode sediar os dois eventos e que as sedes devem ser sorteadas, ao acaso, entre as cidades candidatas, determine a probabilidade de que: a) os eventos sejam feitos apenas em cidades do Nordeste; (Fonte: [31])*

Este exercício, já com o cálculo de probabilidade, segue após uma seção de exercícios resolvidos onde se espera que os alunos façam uso de combinações simples e da definição clássica de probabilidade. Contudo, por meio de questões assim, podemos introduzir a noção de modelo hipergeométrico. De fato, a solução pelo caminho esperado é, essencialmente, tal aplicação não mencionada. Veja que se $X \sim Hip(10, 5, 2)$ é o número de cidades escolhidas da região nordeste, temos $P(X = 2) = \frac{\binom{5}{2}\binom{5}{0}}{\binom{10}{2}} = \frac{2}{9}$.

Situação análoga, visualizamos no seguinte exemplo, extraído de uma segunda obra aprovada já em consonância com a BNCC.

Exemplo 17. *Segundo a Associação Brasileira de Transplante de Orgãos (ABTO), em 2019, no Brasil, foram registrados 158 transplantes de fígado de doadores vivos, nos quais 128 eram de parentes do transplantado e 30 não eram de parentes. Em um hospital do Brasil, foram feitos dois transplantes de fígado de doadores vivos em 2019. Qual é a probabilidade de os fígados utilizados no procedimento : b)Serem um de parente do transplantado e outro não ?(Fonte: [32])*

No contexto em que tal exercício se apresenta, na referida obra, espera-se que o discente faça uso dos conceitos de probabilidade condicional e produto de probabilidades para resolvê-lo. Entretanto, visualizamos também a oportunidade de se inciar a ideia do modelo hipergeométrico.

Além de fazer essa conexão, pode-se se mencionar que um problema de retiradas sucessivas sem reposição pode ser transformado em outro de retiradas simultâneas. Neste sentido, como o exercício sugere uma ordem: O primeiro de parente e o segundo, não (algo não claro no enunciado, e percebível no gabarito). Tomando

$Y \sim Hip(158, 128, 2)$, em que Y é o número de fígados de parentes dos transplantados, temos que $P(Y = 1) = \dfrac{\binom{128}{1}\binom{30}{1}}{\binom{158}{2}} \approx 0,3096$. Como este resultado representa uma retirada simultânea. A reposta procurada é $0,1548$.

Todavia, essa segunda obra contempla o modelo binomial e o sugere na resolução de questões como a que segue. Algo que vai ao encontro do que defendemos.

Exemplo 18. *A probabilidade de um jogador de basquete converter um lance livre é de 70%. Ao realizar 8 arremessos de lance livre, qual a probabilidade de esse jogador converter apenas 5 ? (Fonte :[32])*

Por fim, ressaltamos a pouca conexão que se faz entre probabilidade e Estatística no Ensino Médio. Quesito no qual a segunda obra aponta uma melhoria, pois ao menos contempla as duas temáticas num mesmo volume e traz questões nesse sentido. Aliás, as obras do último PNLD são temáticas e não organizadas por séries como as anteriores. Isso dá mais possibilidade de organizações curriculares cíclicas nas quais os alunos podem revisitar os temas durante os três anos da etapa em estudo.

É fato que a legislação não prever inferência estatística no ensino básico: Algo que limita a citada conexão. Contudo, negar a possibilidade de fazê-la, ainda que de modo intuitivo, é algo que vai de encontro as orientações documentais e ao fazer pedagógico. Neste sentido, encontramos em [33] ideias que dialogam com esse pensar. E que fazem parte de uma área de pesquisa na educação estatística denominada **inferência informal**. Nesta obra, o autor relata estudo com tal viés feito através de uma sequência didática.

Convém observar que estes temas aparecem relacionados no Exame Nacional do Ensino Médio-ENEM. O qual, em dias atuais, se apresenta como a principal forma de acesso ao ensino superior no Brasil, sobretudo, para aqueles que ainda estão cursando o ensino médio.

Considerando esse fato e a significância desse exame para os discentes, vemos a possibilidade de usar tais questões para abordar os temas sob uma perspectiva diferente das que são usuais. Na exploração das questões, os conceitos podem aparecer e serem ressignificados a partir da conexão que o processo de resolução vai evidenciar.

Tal prática contribui para promover o pensamento investigativo, o exercício intelectual e uma visão integrada da matemática no ensino médio. Algo defendido na BNCC, que além disso transparece que o currículo em ação é flexivo. E que portanto, comporta estratégias diversas desde que visem as aprendizagens essências através do desenvolvimento das competências e habilidades previstas.

A fim de ilustrarmos tais considerações, vamos explorar a seguinte questão do ENEM/2018.

Exemplo 19. *Um rapaz estuda em uma escola que fica longe de sua casa, e, por isso, precisa utilizar o transporte público. Como é muito observador, todos os dias ele anota a hora exata (sem considerar os segundos) em que o ônibus passa pelo ponto de espera. Também notou que nunca consegue chegar ao ponto de ônibus antes de 6h e 15min da manhã. Analisando os dados coletados durante o mês de fevereiro, o qual teve 21 dias letivos, ele concluiu que 6h e 21min foi o que mais se repetiu, e que a mediana do conjunto de dados é 6h e 22min.*

A probabilidade de que, em algum dos dias letivos de fevereiro, esse rapaz tenha apanhado ônibus antes de 6h e 21min é, no máximo:
A) $\frac{4}{21}$ *B)* $\frac{5}{21}$ *C)* $\frac{6}{21}$ *D)* $\frac{7}{21}$ *E)* $\frac{8}{21}$

Nesta questão, vemos uma potencial situação didática a ser explorada. Note que além de evocar a compreensão do que seja moda, mediana e espaço equiprovável, ela ainda trás condições que exige uma análise da relação entre esses conceitos, como a ideia de probabilidade máxima.

Para chegarmos a resposta, é necessário percebermos que o valor mínimo do conjunto, a mediana e a moda delimitam o evento de probabilidade desejada. Entre 6h e 15min e 6h e 22min, temos 9 horários nas condições do problema.

Dentre o conjunto de horários, 6h e 21min deve se repetir pelo menos duas vezes, pois é a moda. Contudo, entre 6h e 15min a 6h e 21min, apenas cinco valores distintos são possíveis nas condições dadas: 6h e 16min, 6h e 17min, 6h e 18min , 6h e 19min e 6h e 20min.

Logo, chegamos a duas possibilidades: A moda é um valor que se repetiu quatro vezes ou um desses valores menores, incluindo o 6h e 15 min, se repetiu duas vezes, e a moda três. A primeira opção nos conduz a $\frac{6}{21}$, já a segunda, a resposta desejada : $\frac{7}{21}$.

Por fim, observamos que excluindo a suposição que desconsidera os segundos, e supondo um horário limite para a passagem do ônibus, torna-se possível fazer menção a ideia de espaço amostral como um intervalo de tempo. E com isso, provocar uma discussão de como se ver esse espaço amostral, quais e como seriam seus eventos. Motivando a formulação de problemas que podem suscitar em noções sobre variáveis aleatórias contínuas.

Referências

[01] **Piana, Clause Fátima de Brum; Machado, Amauri de Almeida; Selau, Lisiane Priscila Roldão.** Estatística Básica,Pelotas, 2009.

[02] **Crespo, Antônio Arnot**. Estatística fácil. Saraiva Educação SA, 2017.

[03] **OBMEP em números.** Disponível em:< http://www.obmep.org.br/em-numeros.htm >. Acesso em: 30 de setembro de 2022.

[04] **Ximenes, Victor.** Os 20 carros mais vendidos do Ceará na 1ª metade de 2022.Diponível em :< https://diariodonordeste.verdesmares.com.br/opiniao/colunistas/victor-ximenes/20-carros-mais-vendidos-do-ceara-2022-1.3252836 >. Acesso em : 28 de outubro de 2022.

[05] **Morgado, Agusto Cesar.et al**.Análise Combinatória e Probabilidade: com resolução de exercícios. Coleção do Professor de Matemática. 10ªEd. Rio de Janeiro: SBM,2016.

[06] **Magalhães, Marcos Nascimento.** Probabilidade e variáveis aleatórias. Edusp, 2006.

[07] **Meyer, Paul L.** Probabilidade: Aplicações à Estatística. 2ª Edição-Tradução de Ruy de CB Lourenço Filho, 1995.

[08] **Lourenço, Tainá**. Vírus do herpes-zoster está presente em 95% das pessoas. Jornal da Usp.Disponível em:<https://jornal.usp.br/atualidades/virus-da-herpes-zoster-esta-presente-em-95-das-pessoas/ >. Acesso em : 21 de fevereiro de 2023.

[09] **Guidorizzi, Hamilton.** Um curso de cálculo. V2, 5ªedição,[reimpressão]. Rio de Janeiro, LTC,2015.

[10] **Morettin, Pedro; Bussab,Wilton.** Estatística Básica.6ª edição,São Paulo, SA, 2017.

[11] **Patriota, Alexandre Galvão.** Quanto a Empresa deve ter em caixa?Disponível em :< https://www.youtube.com/watch?v=hVq7NtCpBFI&t=89s>. Acesso em : 10 de março de 2022.

[12] **Lima, Elon Lages.** Números e funções reais. Sociedade Brasileira de Matemática, 2013.

[13] **Neto, Antonio Caminha Muniz.** Fundamentos de cálculo. Rio de janeiro: SBM, 2015.

[14] **Queiroz, Eduardo Ravaglia Campos et al.** Da (Im) previsibilidade do futebol sob a ótica da distribuição de Poisson.

[15] **Torres, Oswaldo Fadigas.** Elementos da teoria das filas.Revista de Administração de Empresas, v. 6, p. 111-127, 1966.

[16] **Silva, Ítalo Nunes et al.** Distribuição de frequência da chuva para região Centro-Sul do Ceará, Brasil. Revista Ciência Agronômica, v. 44, p. 481-487, 2013.

[17] **Koltun, Ana Paula da Silva et al.** Difusão–Usual e Anômala, Cálculo Fracionário e Aplicações. 2022.

[18] **Melo, Kelma Gomes de.** A função gama como extensão da função fatorial e aplicações. Redenção, 2020.

[19] **Rossato, Rafael Antônio; Ferreira, Vitor Vieira.** Lei dos expoentes envolvendo derivadas e integrais fracionárias segundo Riemann-Liouville.

[20] **Pascoa, Marcelino Alves Rosa de.** Extensões da distribuição gama generalizada: propriedades e aplicações. 2012. Tese de Doutorado. Universidade de São Paulo.

[21] **Borges, Marta Eliane Echeverria; Mazucheli, Josmar.** Análise da duração das sequências de dias chuvosos das capitais Brasileiras via distribuição gama generalizada discreta truncada em zero. Revista Brasileira de Climatologia, v. 29, p. 228-250, 2021.

[22] **Francisco, Paulo Roberto Megna et al.** Discriminação de cenários pluviométricos do estado da Paraíba utilizando distribuição Gama Incompleta e Teste Kolmogorov-Smirnov. Revista Brasileira de Geografia Física, v. 9, n. 01, p. 047-061, 2016.

[23] **Matsushita, Italo Junji et al.** Cálculo de estoque mínimo para itens industriais utilizando as distribuições de probabilidade normal, poisson e gama. 2021.

[24] **Conselvan, Vagner de Almeida.** Análise de confiabilidade aplicada na otimização de sistemas de produção animal. 2017.

[25] **Gonçalves, Daiane de Oliveira et al.** Modelos paramétricos de sobrevivência aplicados a dados de câncer. 2020.

[26] **Isidro, Manoel Joaquim et al.** Modelos de fragilidade aplicados a análise de fatores contribuintes na morte de pacientes portadores de leucemia. Brazilian Journal of Development, v. 6, n. 8, p. 54802-54820, 2020.

[27] **Farnum, Nicholas R.; Stanton, Laverne W.** Some results concerning the estimation of beta distribution parameters in PERT. Journal of the Operational Research Society, p. 287-290, 1987.

[28] **Vahatalo, Anssi V.; Aarnos, Hanna; Mantyniemi, Samu.** Biodegradability continuum and biodegradation kinetics of natural organic matter described by the beta distribution. Biogeochemistry, v. 100, p. 227-240, 2010.

[29] **Elmer, Jonathan; Jones, Bobby L.; Nagin, Daniel S.** Using the Beta distribution in group-based trajectory models. BMC medical research methodology, v. 18, n. 1, p. 1-5, 2018.

[30] **Brasil**. Base Nacional Comum Curricular: Ensino Médio. Brasília: MEC/Secretaria de Educação Básica, 2018.

[31] **Iezzi, Gelson; Dolce, Osvaldo et. al.** Matemática Ciência e Aplicações. 9. ed. São Paulo: Saraiva, v. 2, 2016.

[32] **Chavante, Eduardo; Prestes, Diego.** Quadrante, matemática e suas tecnologias: Estatística, probabilidade e matemática financeira. 1ªed. São Paulo: sm, 2020.

[33] **Rorigues, Vinicius Alves.** Uma experiência de inferência estatística informal na escola básica. 2016. Tese de Doutorado. Universidade de São Paulo.

[34] **Lopes, Celi Espasandin.** O ensino da estatística e da probabilidade na educação básica e a formação dos professores. Cadernos Cedes, v. 28, p. 57-73, 2008.