UNIVERSIDADE FEDERAL DO AMAPÁ
PRÓ-REITORIA DE ENSINO E GRADUAÇÃO
CURSO DE LICENCIATURA EM MATEMÁTICA

CÁLCULO DE MOMENTOS E APLICAÇÕES

MACAPÁ-AP
2012

MILENA NASCIMENTO LIMA

CÁLCULO DE MOMENTOS E APLICAÇÕES

Trabalho de Conclusão de Curso apresentado ao corpo docente do Curso de Licenciatura em Matemática - UNIFAP, como requisito parcial para a obtenção da Graduação em Licenciatura em Matemática.

Área de Concentração: Estatística Aplicada
Orientador: ***Dr. José Walter Cárdenas Sotil.***

MACAPÁ-AP
2012

CÁLCULO DE MOMENTOS E APLICAÇÕES

por
LIMA, Milena Nascimento

Trabalho de Conclusão de Curso apresentado como pré-requisito para obtenção do título de Licenciatura Plena em Matemática da Universidade Federal do Amapá, submetida à aprovação da banca examinadora composta pelos seguintes membros:

Banca Examinadora

Orientador: Prof. *Dr.* José Walter Cárdenas Sotil.
Universidade Federal do Amapá - UNIFAP

Membro: Prof. *Dr.* Guzman Eulálio Isla Chamilco.
Universidade Federal do Amapá - UNIFAP

Membro: Prof. *Esp.* João Socorro Pinheiro Ferreira.
Universidade Federal do Amapá - UNIFAP

Avaliado em: 10 de Dezembro de 2012.

MACAPÁ-AP
2012

Agradecimentos

Sou grata ao doutor dos doutores, Deus, pela vida, saúde e companheirismo, pois essas três concessões serviram de intermédio para que eu tivesse a honra de ingressar na UNIFAP. E é por estas mesmas razões que neste momento me farto de uma alegria incomensurável, que é a de poder apresentar este TCC.

"Falarei da magnificência gloriosa da tua majestade e das tuas obras maravilhosas. Porque tu és grande e fazes maravilhas; só tu és Deus". Bíblia Sagrada:Salmos [145:5] e [86:10].

⋆ Ao professor Dr. José Walter Cárdenas, pelas orientações e compreensões;
⋆ À minha mãe, Anita por ter me apoiado nesses anos de muita luta;
⋆ Ao meu pai, pela torcida ao meu sucesso e por sempre se dispor em meu favor;
⋆ À minha Tia, Luzia, por ter acreditado em mim e por isso nunca mediu esforços para me ajudar;
⋆ Aos meus avós, Adair e Apolônia que sempre me incentivaram a trilhar por este caminho: obrigada pelas orações;
⋆ Aos meus irmãos, Renê e Cássia;
⋆Às minhas amigas, Sueide e Amanda: obrigada pela companhia. Jamais esquecerei vocês;
⋆ José Ivaney, obrigada pela assistência. Você é inesquecível!
⋆ Ítalo Bruno, valeu pelas aulas de Análise Real. Essas marcaram!
⋆ Graciana Fernandez e Cristiane Fernandez.
⋆ Jeison de Lima Freitas
⋆ Adelson Nazaré Bahia
⋆ Amarildo Maciel Sousa
⋆ Joelma Soeiro de Figueiredo
⋆ Renato Júnior

"Um pouco de ciência nos afasta
de Deus. Muito, nos aproxima."
(PASTEUR, Louis)

Resumo

Após uma momentânea revisão sobre surgimento da teoria das probabilidades, bem como sua evolução até alcançar os dias atuais, optamos por fazer um estudo detalhado sobre os conceitos básicos desta doutrina. Nesse primeiro momento, houve uma preocupação em representar uma parte considerável dos conceitos através de diagramas, gráficos e exemplos. São definidas as variáveis aleatórias (discretas e contínuas), ferramentas de tamanha significância, pois através do conhecimento destas é que foi possível a expansão deste estudo até chegarmos ao segundo período, que foi o de estudar os modelos probabilísticos. Nessa etapa, os modelos probabilísticos mereceram capítulo exclusivo, onde definimos os modelos discretos e contínuos, ambos com devidos exemplos e provas de que constituem verdadeiras funções de distribuição probabilidades ou funções densidade de probabilidades, dependendo da característica do modelo. E para alcançar o ápice da temática deste trabalho, definimos os momentos de ordem 1, de ordem 2, e de ordem n, bem como fizemos as demonstrações de suas propriedades. Avante, fizemos menção da função geratriz de momentos, uma ferramenta poderosa no que diz respeito aos estudos das distribuições de probabilidade, mas que não é eficaz no cálculo dos momentos de todas as distribuições, já que nem sempre existirá. É justamente neste impasse que tratamos de outra função que também gera momentos, e que ao contrário da função geratriz de momentos sempre existirá: esta é a função característica. Finalmente calculamos os momentos de alguns dos principais modelos probabilísticos a partir das definições de momentos e também com o uso dessas duas funções.

Palavras-chave: Variáveis Aleatórias, Esperança, Variância, Covariância, Coeficiente de Correlação, Função Geratriz de Momentos, Função Característica.

Abstract

After a momentary revision on appearance of the theory of the probabilities, as well as your evolution until reaching the current days, we opted to do a detailed study on the basic concepts of this doctrine. On that first moment, there was a concern in representing a considerable part of the concepts through diagrams, graphs and examples. It is defined the random variables (discreet and continuous), tools of great meaning, because through the knowledge of these it is that was possible the expansion of this study to we arrive to the second period, that was it of studying the models of probability. In that stage, the models of probability deserved exclusive chapter, where we defined the discreet and continuous models, both with having owed examples and proofs that they constitute true functions of distribution probabilities or functions density of probabilities, depending on the characteristic of the model. And to reach the apex of the thematic of this work, we defined the moments of order 1, of order 2, and of order n, as well as we made the demonstrations of your properties. Forward, we made mention of the Moment Generating Function, a powerful tool in what concerns the studies of the distributions of probability, but that is not effective in the calculation of the moments of all the distributions, since not always it will exist. It is exactly in this impasse that we treated of other function that also generates moments, and that unlike the fgm, it will always exist: this is the characteristic function. Finally we calculated the moments of some of the principal probability models starting from the definitions of moments and also with the use of those two functions.

keyword: Random Variables, Expectation, Variance, Covariance, Coefficient of Correlation, Moment Generating Function, Characteristic Function.

SUMÁRIO

LISTA DE TABELAS

LISTA DE FIGURAS

Introdução

Muitas características essenciais das variáveis aleatórias podem ser quantificadas através do valor esperado, de potências de uma determinada variável aleatória, no qual podemos analisar o comportamento da mesma e ainda tomá-lo como parâmetro para vários modelos probabilísticos. As noções de valor esperado, ou média, foram desenvolvidas com a teoria das probabilidades, a partir do instante em que o homem passou a se interessar pelos estudos de fenômenos que envolviam possibilidades (ou incertezas), em especial, ao dos jogos. Por tanto, antes de tratarmos efetivamente do tema deste trabalho, é importante fazermos uma revisão dos aspectos de maior relevância na história da teoria das probabilidades bem como alguma de suas aplicações para a modernidade.

Na pré-história e na Antiguidade.

Embora a história considere o início do Cálculo das Probabilidades somente a partir de 1564, é conveniente destacar que a percepção dos fenômenos indeterminísticos já estimulava a razão humana bem antes desse período. Já havia uma distinção entre o "acidental" e o proposital, pois já existiam jogos, apostas e previsões sobre o futuro. Isso é percebido mediante a exposição de diversos objetos, inclusive ossos de tornozelos bovinos e de dados, encontrados em escavações de cemitérios no oriente, o que comprova suas utilizações no período pré-histórico.

Na busca pela compreensão e explicação dos fenômenos incontroláveis, que intrigavam o homem, recorria-se à justificação por intermédio das forças ocultas, e mágicas.

A palavra probabilidade apareceu no dicionário francês Hachett, em 1361, como a ciência que tem o objetivo determinar a probabilidade de um evento.

Na filosofia grega, Aristóteles atribuía às questões de sorte ou azar, a causa acidental, sendo que o azar é tratado como um processo mais amplo, que atinge a natureza e os seres humanos, e que não há como controlar. Porém, na filosofia aristotélica, essa concepção

só pode ser aplicada no mundo terreno, mas nunca no celeste, considerando-o perfeito e ordenado [43].
O termo "provável" aparecia cerca de 1285 vezes, em notação utilizada em jurisprudência do século XV, com o sentido daquilo que se pode provar. A probabilidade seria o caráter daquilo que é provável e o cálculo das probabilidades, a ciência cujo objetivo é o de determinar a probabilidade de ocorrência de um evento.

Na Idade Média.

Até então, os destaques sobre a probabilidade estavam voltados somente para a apresentação da enumeração das possibilidades de um determinado evento vir a acontecer num jogo. Não havia assim, uma preocupação probabilística mais explicita. Diversas obras literárias medievais, inclusive "A Divina Comédia de Dante", retrataram as possibilidades de ocorrerem somas 2,3,5,...12, na jogada de dois dados. Somente a partir de 1500 começou-se a desencadear o desenvolvimento de cálculos, por matemáticos italianos. Luca Bartolomeo de Pacioli (1445-1517), publicou em Veneza a famosa obra "Summa de Arithmetica, Geometria proportioni et propornalità" onde estudou o seguinte problema: "Dois jogadores disputavam um prêmio que seria dado a quem primeiro fizesse 6 pontos no jogo da balla. Quando o primeiro jogador tinha 5 pontos e o segundo tinha 3 pontos, foi preciso interromper o jogo. Como dividir o prêmio?" A solução foi dada ao fazer a divisão proporcional à probabilidade de vitória de cada jogador. Desta forma, estava surgindo a noção intuitiva de valor esperado.
Girolamo Cardano (1501-1576), em sua obra, "Liber de Ludo Aleae" , publicada em 1663, após sua morte, mostra as análises do conceito probabilístico bem como a exposição de argumentos para afirmações feitas por ele, como a de que em um determinado lançamento a probabilidade de ocorrência de uma face de um dado honesto é a mesma para as demais.Apesar de ter introduzido técnicas combinatórias para o cálculo das quantidades possíveis e favoráveis de eventos aleatórios, limitou-se a estudar somente problemas de natureza concreta, não produzindo ainda o formalismo matemático.

Na Idade Moderna

Galileu Galilei (1564-1642) também deixou sua contribuição ao resolver o problema que ficou conhecido como "Problema do Grão-duque de Toscana". Desejava-se saber por que num lançamento de três dados, cujo resultado consistia na soma dos pontos das faces superiores, registrava-se com mais frequência o número 10 do que o número 9, já que a decomposição de cada um desses dois números é dada de seis maneiras (ver Tabela 1).

Resultado 9	Resultado 10
1+2+6	1+3+6
1+3+5	1+4+5
1+4+4	2+2+6
2+2+5	2+3+5
2+3+4	2+4+4
3+3+3	3+3+4

Tabela 1 *Decomposição do número 9 e 10*

Galileu mostrou que isso acontecia pelo fato das somas não serem equiprováveis. As somas de (1,2,6), (1,3,5), (2,3,4), (1,3,6), (1,4,5) e (2,3,5) aparecem em seis formas distintas (ver Tabela 2:

1+2+6	1+3+5	2+3+4	1+3+6	1+4+5	2+3+5
1+6+2	1+5+3	2+4+3	1+6+3	1+5+4	2+5+3
2+1+6	3+1+5	3+2+4	3+1+6	4+1+5	3+2+5
2+6+1	3+5+1	3+4+2	3+6+1	4+5+1	3+5+2
6+1+2	5+1+3	4+2+3	6+1+3	5+1+4	5+2+3
6+2+1	5+3+1	4+3+2	6+3+1	5+4+1	5+3+2

Tabela 2 *Valores equiprováveis*

A soma de (3,3,3,) aparece de uma única forma. Como o jogo consiste no lançamento de três dados, o espaço amostral é composto de 216 pontos, que é o produto (6x6x6) dos valores possíveis de cada dado. Assim,

$$\begin{aligned}
P(soma = 9) &= P(1,2,6) + P(1,3,5) + P(2,2,5) + P(1,4,4) + P(2,3,4) + P(3,3,3) \\
&= \frac{6}{216} + \frac{6}{216} + \frac{3}{216} + \frac{3}{216} + \frac{6}{216} + \frac{1}{216} \\
&= \frac{25}{216} \\
P(soma = 10) &= P(1,3,6) + P(1,4,5) + P(2,2,6) + P(2,3,5) + P(2,4,4) + P(3,3,4) \\
&= \frac{6}{216} + \frac{6}{216} + \frac{3}{216} + \frac{6}{216} + \frac{3}{216} + \frac{3}{216} \\
&= \frac{27}{216}
\end{aligned}$$

Pierre Fermat (1601-1665) e Blaise Pascal (1623-1662) formaram os pilares da doutrina probabilística através de correspondências, em que continham situações que demandavam soluções para jogos com apostas. É conveniente que isso possa ser visto como fator decisivo para o tratamento da probabilidade como ciência, pois outros matemáticos fundamentaram os demais conceitos tomando por base os princípios estabelecidos por estes, mais especificamente por Pascal [05].

A probabilidade era um tema desconhecido por Fermat até então, mas isso lhe impulsionou a buscar por fundamentos matemáticos que explicassem e descrevessem com eficiência e precisão, as leis do acaso. Unido a Pascal, ambos determinaram suas regras constitucionais. Em uma das correspondências, datada de 24 de agosto de 1654, endereçada a Pascal, continha o seguinte problema: "Se num jogo, com dois jogadores A e B. No momento em que o jogador A precisa de 2 pontos para ganhar enquanto que B precisa 3 pontos, o jogo será certamente decidido em quatro jogadas. Quem tem maior chance de ganhar?". Para tentar solucionar esse problema, Fermat escreveu todas as combinações possíveis entre as letras a, que representa uma jogada em favor do jogador A e b que representa uma jogada em favor do jogador B [36] (ver Tabela 3):

E verificou que, em um total de 16, têm-se 11 casos favoráveis para A versus 5 favoráveis

aaaa	baaa
aaab	baab
aaba	baba
aabb	babb
abaa	bbaa
abab	bbab
abba	bbba
abbb	bbbb

Tabela 3 *Combinações de a b e c*

para B, uma vez que a ocorrência de 2 ou mais $a^{'}$s é favorável para o jogador A e a ocorrência de 3 ou mais $b^{'}$s para o jogador B.

A solução respondida por Pascal foi a seguinte: "suponhamos que cada um dos jogadores aposte a mesma quantia, 32 pistolas (moeda da época), aquele que tirar primeiramente três vezes, seguidas ou não, o número que aposta no dado, de 1 a 6, ganhará, num total de quatro partidas".

Embora Fermat tenha se interessado pela probabilidade, ela não foi o centro de sua atenção como matemático, pois as formulações feitas por ele estavam sempre baseadas nos formalismos de Pascal. Este, por sua vez, alcançou prestígio pelas suas descobertas e possibilitou o desenvolvimento das pesquisas por outros matemáticos até alcançar um patamar que até então era inimaginável.

Christian Huygens (1629-1695), físico, astrônomo e geômetra holandês, foi quem deu o primeiro tratamento científico a este assunto com a publicação, em 1657, da obra De Ratiociinis in LudoAleae, onde estabeleceu o conceito de esperança matemática, indispensável para o desenvolvimento da temática deste trabalho, expôs diversas situações sobre o assunto, fundamentou propriedades e fez demonstrações para suas afirmações.

Antoine Arnauld(1612-1694) e Pierre Nicole (1625-1695), ambos franceses, publicaram a obra "La Logique ou l′art de penser", que influenciou fortemente a literatura francesa. O tema central dessa obra é a lógica, incluindo a probabilidade com elevado grau de significância e de credibilidade entre os vários contextos que formam sua base teórica [37].

Gottfried Wilhelm Leibnitz(1646-1716) deu um "passo largo" no que tange a esse assunto. Na tentativa de ampliar seus estudos sobre lógica, percebeu que todos os acontecimentos, inclusive o pensamento, se reduzem a uma combinação, de códigos, letras, números, cores, etc.[38]. Em 1666, publicou a obra "De Arte Combinatória" com os seguintes princípios:

a) Criação de uma nova língua, com notação universal e artificial

b) Fazer o inventário das ideias simples e simbolizá-las de modo a obter um "alfabeto dos pensamentos" também simples e expresso em caracteres elementares.

c) Produção de ideias compostas com a combinação desses caracteres, e estabelecer técnicas automatizaçãode raciocínio, de modo a substituir o pensamento e a intuição por um cálculo de signos.

Jakob Bernoulli (1654-1705), produziu a obra "A Arte da Conjectura" (publicada em 1713 após a morte), que foi a primeira dedicada inteiramente a teoria das probabilidades. Nesse trabalho ele expôs os resultados de alguns problemas, para os quais Pascal e Fermat não publicaram soluções. Fundamentou a probabilidade como um grau mensurável de certeza, necessidade e oportunidade, confrontou a expectativa moral e a esperança matemática.Além disso, estabeleceu a Lei dos Grandes Números, fator indispensável para o conhecimento aprofundado dos fenômenos aleatórios.

Pierre Rémond de Montmort(1678-1719), juntamente com Nicola Bernoulli (1695-1726), após uma intensa troca de correspondências, produziram famosa obra “Essay d′analyse sur Le jeux de hazard” onde dão o tratamento algébrico para combinação e jogos complexos. Montmort foi o primeiro, depois de Pascal, a mencionar o triângulo de Pascal, mas com o nome de tabuleiro de Pascal e ainda conceituou desarranjo e função geratriz.
Nicola Bernoulli registrou em sua tese, publicada em 1711, o conceito da Distribuição Uniforme, sendo o primeiro a fazer este registro. Além disso, articulou um paradoxo que despertou o interesse de seus sucessores.

Abraham de Moivre (1667-1754), descobriu a taxa de convergência da lei dos grandes números, resultando na primeira versão do teorema central do limite, na obra “The Doctrineof Chances” [13]. Apresentou mais de 50 problemáticas distribuídas entre jogos de dados e as probabilidades de se tirar bolas de cores diferentes de uma mesma urna.

Georges Louis Leclerc, Conde de Buffon (1707-1788), foi um naturalista francês, que demonstrou enorme interesse pelas ciências naturais e pela matemática. Sua obra mais importante foi uma enciclopédia de 36 volumes, intitulada Historie Naturelle, onde expôs seus conhecimentos sobre o mundo natural, resultando em grande interesse por parte da comunidade européia daquela época. Mas o que nos interessa nessa produção é o volume de número 4, onde incluiu o trabalho Essai d′ArithmetiqueMorale, onde fez adaptações da matemática com a realidade humana. A essência, inovadora, desse trabalho está na intervenção do Cálculo diferencial e Integral na Teoria das Probabilidades. E, mediante ao estudo que fez sobre o problema que ficou conhecido como “Agulhas de Buffon”, introduziu a probabilidade geométrica, que futuramente seria convertida para Geometria Integral e Estocástica [39].
No problema das agulhas, desejava-se saber a probabilidade de uma agulha de comprimento l cruzar uma de duas retas paralelas, com distância d, de um plano. Esse problema foi resolvido na seguinte forma:

a) Se $l < d$:

Supondo-se que X seja a variável aleatória (Ver capítulo 2) que representa a distância d, do centro da agulha até à reta mais próxima e θ a v.a. que indica o ângulo formado pelo centro da agulha às retas, de forma que $0 \leq X \leq 1/2$ e $0 \leq \theta \leq \pi/2$. Desta forma, uma das retas só será interceptada, pela agulha, se $x \leq l/2sen\theta$. E no caso de $x = 0$, pode-se afirmar que isso só será possível se o ponto correspondente ao centro da agulha for comum a um ponto de uma dessas retas.

Assim, a probabilidade p da agulha interceptar uma das retas é dada pela razão entre área do retângulo R e o gráfico da função $x = l/2sen\theta$.

$$\begin{aligned} A_R &= \pi d/2 \\ A_f &= \int_0^\pi l/2sen\theta d\theta = l/2[-cos\theta]_0^\pi = l \end{aligned}$$

Logo,

$$p = \frac{l}{\frac{\pi d}{2}} = \frac{2l}{\pi d}$$

b) Se $l = d$:

Basta considerar o caso acima, só que fazendo a substituição de d por l. Assim:

$$p = 2/\pi$$

Daniel Bernoulli (1700-1782), publicou 1738 na academia de São Petersburgo, na França,uma obra sobre o paradoxo que havia sido articulado por Daniel Bernoulli. Esse ficou conhecido como Paradoxo de São Petresburgo, que tratava da seguinte questão[40]: No lançamento de uma moeda, considere dois jogadores A e B. Considere ainda, as seguintes situações:

i) Se sair cara no primeiro lançamento, o jogador B pagará duas moedas ao A.

ii) Se cara sair no segundo lançamento, o jogador B pagará quatro moedas ao A.

iii) Se cara aparece só no terceiro lance, o jogador B paga oito moedas ao A.

Considerando que o n-ésimo lançamento exista, o jogador B pagará 2^{n-1} moedas ao jogador A. Mas o impasse dessa questão é o seguinte: Quantas moedas o jogador A deverá pagar ao jogador B, simplesmente pela honra de jogar? Se ele tomar por base o princípio da esperança matemática, ele estará disposto a pagar o próprio valor desta. Como o valor da esperança matemática é dado por:

$$\sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^n 2^{n-1} = 1/2 \cdot 1 + 1/4 \cdot 2 + 1/8 \cdot 4 + \cdots 1/2 + 1/2 + 1/2 + \cdots = +\infty$$

Então o jogador estaria disposto a pagar qualquer quantia para entrar nessa jogada. Até mesmo se este fosse rico, sua fortuna poderia ser a postada, o que foge da realidade, pois não é comum encontrar pessoas dispostas a fazer esse tipo de aposta.A solução que Bernoulli encontrou para esse paradoxo se converteu no fundamento da teoria da utilidade esperada. Para ele, a determinação do valor de um item não pode ser baseada em seu preço, mas sim na utilidade que ele fornece. O preço de um item depende somente do próprio item e, é igual para todo mundo; a utilidade, contudo, depende das circunstâncias particulares do indivíduo que faz a estimativa [41].

Na Contemporaneidade

No século XX, tempo de grandes transformações na ciência e na tecnologia, o desenvolvimento da teoria da probabilidade disparou num ritmo acelerado. Os empenhos realizados para avaliar o retorno num sistema de várias jogadas produziram formalismo matemático, onde as idéias em questão foram estabelecidas em axiomas, definições, teoremas, lemas e proposições, envolvendo os casos discretos e contínuos. Nesse processo de axiomatização, muitos matemáticos se destacaram, entre eles: Félix Édouard Justin Émile Borel (1871-1956), autor da obra "Leçons sur la Théorie dês fonctions". Seus trabalhos são os originais sobre os conjuntos e os cálculos das probabilidades. Além do mais é autor da primeira teoria científica da medida de um conjunto de pontos.

Maurice René Fréchet (1878-1973) formalizou, em seus trabalhos, os estudos sobre esperança matemática de variáveis aleatórias. Em uma de suas de suas principais publicações está incluído o tema "Modernes théoretiques Recherches sur la théorie des probabilités".

Andrei kolmogorov (1903-1987), é tratado como um dos mais importantes cientistas do século XX, pois participou das maiores descobertas científicas desse tempo, principalmente no que se refere ao formalismo da teoria das probabilidades. É autor da obra Grundbegriffe der Wahrscheinlichkeitsrechnung (Fundamentos da Teoria das Probabilidades), onde lançou toda a base axiomática da probabilidade tal como conhecemos hoje.

Atualmente, desenvolvimento da teoria das probabilidades, assim com tantos outros ramos da matemática, tem sido estimulado pelo avanço descontrolado da tecnologia, onde é notável a variedade de suas aplicações. Ciências como a medicina física e engenharias, têm sido verdadeiras aliadas de seus recursos. Porém, suas principais aplicações dizem respeito aos estudos sobre equidade de jogos e respectivos prêmios, e está destinada também à

Estatística Indutiva, na acepção de amostra, extensão dos resultados à população e na previsão de acontecimentos futuros [44].

Capítulo 2

DEFINIÇÕES BÁSICAS

2.1 Modelo Matemático para um Experimento

É fundamental que saibamos qual tipo de fenômeno estamos preocupados em observar, para então fazermos a exposição do modelo matemático para o mesmo. Este, por sua vez, tem a finalidade de simplificar as coisas, desprezando certos detalhes. Na maioria das vezes é bastante fácil afirmar com certeza se um modelo matemático especificado é ou não adequado, antes que alguns dados de observação sejam obtidos. Mas o interessante mesmo, é que o sucesso do modelo depende do fato de que os detalhes desprezados sejam ou não, verdadeiramente, sem importância na explicação do fenômeno estudado [20].

Se um experimento é repetido nas mesmas condições, não sendo possível conhecer a priori o resultado, o experimento é dito aleatório. Um exemplo deste tipo de experimento é o lançamento de uma moeda, em que conhecemos que o resultado é cara ou coroa, entretanto não temos certeza a priori se o lançamento será cara ou coroa.

2.2 Espaços de Probabilidade

Considere um experimento aleatório, o qual estamos interessados em modelar. A etapa inicial é determinar os possíveis resultados do experimento e, em seguida, escolher um conjunto Ω, denominado *Espaço Amostral*, cujos pontos estão associados a esses resultados. Feito isso, suponha que exista uma coleção não-vazia β de subconjuntos de Ω, denominada coleção de eventos. Considere ainda, uma função escalar P definida em β, ver a Figura 2.1 para uma representação da probabilidade.

Definição 2.2.1. *Ao terno* (Ω,β,P) *designamos o espaço de probabilidades , onde a função* P *obedece as seguintes relações:*

i) $0 \leq P(B) \leq 1, \qquad \forall B \in \beta$

ii) $P(\Omega) = 1$

iii) $P\left(\sum_{i=1}^{n} B_i\right) = \sum_{i=1}^{n} P(B_i)$ para todas as seqüências contáveis e disjuntas de $B_i \in \beta$.

iv) $P\left(\bigcup_{i=1}^{n} B_i\right) = \bigcup_{i=1}^{n} P(B_i)$ se cada par dos B_i forem mutuamente exclusivos.

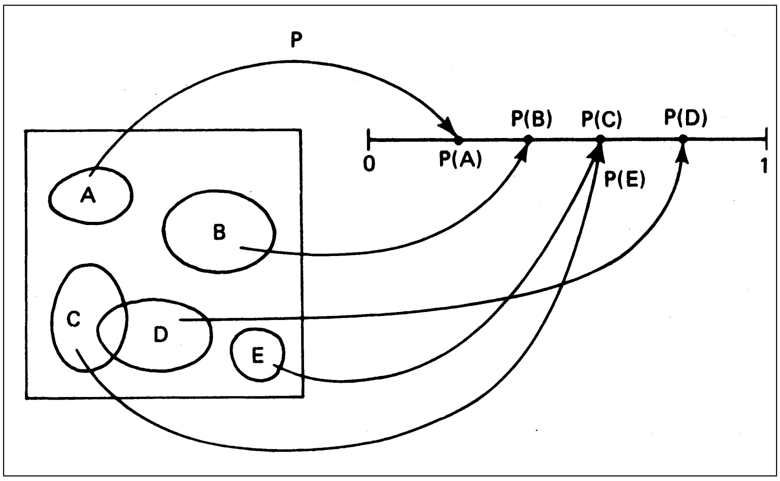

Figura 2.1 *Representação sistemática da probabilidade.*
Fonte: Chung Kai Lai, 2002

2.3 Probabilidade Condicionada

Para muitas situações, nosso interesse estará voltado para a ocorrência não-simultânea de dois ou mais eventos, de uma forma tal que o fato de que um deles tenha ocorrido, ou não, venha a comprometer probabilidade da ocorrência do outro. Na Figura 2.2 se ilustra o conceito de probabilidade condicional, se o evento B ocorre, a ocorrência do evento A fica restrita ao evento $A \cap B$.

Definição 2.3.1. *Considere o tripleto* (Ω,β,P), *e um evento* B *qualquer tal que* $P(B) \neq 0$. *A probabilidade de um evento* A *vier a ocorrer, visto que* B *já ocorreu, é denominada probabilidade condicional e é representada na seguinte forma:*

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad \forall \text{ evento } B \text{ tal que } P(B) \neq 0.$$

Analogamente,

$$P(B|A) = \frac{P(A \cap B)}{P(A)}, \quad \forall \text{ evento } A \text{ tal que } P(A) \neq 0.$$

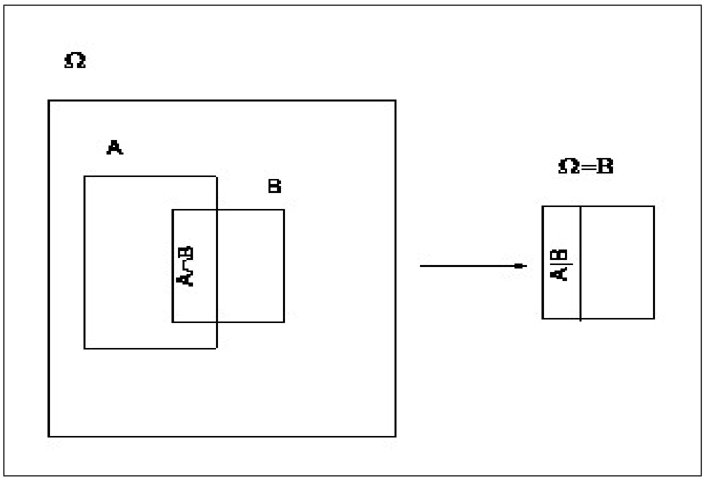

Figura 2.2 *Probabilidade condicionada*
Fonte: cardenas.webnode.com

Exemplo 2.3.1. *[28]: Para o cumprimento de um teste de diagnósticos, foi realizado um grande estudo com 1820 pessoas, com ou sem tuberculose. Estas pessoas foram submetidas ao teste de raios-X, a fim de verificar a capacidade preditiva do exame. Para isso, considere os seguintes eventos:*

$D =$ "A doença é presente"

$D^c =$ "A doença é ausente"

$T^+ =$ "Resultado positivo (teste detecta a doença)"

$T^- =$ "Resultado negativo (teste não detecta a doença)"

$P(T^+|D) =$ "Sensibilidade (taxa de verdadeiro positivo)"

$P(T^+|D^c) =$ "Probabilidade de um resultado falso positivo"

$P(T^-|D) =$ "Probabilidade de um resultado falso negativo"

$P(T^-|D^c) =$ "Especificidade (verdadeiro negativo)"

Considere ainda que, para se obter estimativas do valor preditivo positivo (VPP), temos que o valor preditivo negativo (VPN) é dado na seguinte forma: $VPN = P(D^c|T^-)$, o que implica numa estimativa da probabilidade global da doença (prevalência da doença P(D) na população geral. E mais, que encontrar VPP significa achar a probabilidade de que um indivíduo tenha a doença por intermédio de um teste positivo $P(D|T^+)$. A situação é apresentada nas Tabelas 2.1, 2.2 e 2.3.

Resultado de Raios-X	Tuberculose		Total
	SIM	NÃO	
POSITIVO	22	51	73
NEGATIVO	8	1739	1757
TOTAL	40	1790	1830

Tabela 2.1 *Freqüências do teste de raios X*

Resultado do Teste	Doença		Total
	SIM	NÃO	
POSITIVO	$n(D \cap T^+)$	$n(D^c \cap T^+)$	$n(T^+)$
NEGATIVO	$n(D \cap T^-)$	$n(D^c \cap T^-)$	$n(T^-)$
TOTAL	$n(D)$	$n(D^c)$	n

Tabela 2.2 *Representação das freqüências para o teste de raios X.*

Resultado do Teste	Doença		Total
	SIM	NÃO	
POSITIVO	$P(D \cap T^+)$	$P(D^c \cap T^+)$	$P(T^+)$
NEGATIVO	$P(D \cap T^-)$	$P(D^c \cap T^-)$	$P(T^-)$
TOTAL	$P(D)$	$P(D^c)$	1

Tabela 2.3 *Representação da probabilidade para o teste de raios X*

Considere ainda que a prevalência da doença na população geral seja $P(D) = 0,000093$. A partir da tabela, podemos derivar estimativas aproximativas para a sensibilidade e especificidade do raios-x como um teste de diagnósticos. Assim:

$$P(T^+|D) \approx \frac{22}{40} \quad = \quad 0,55$$

Observe que, desde que D seja dado Ω é formado apenas pelos 30 casos. Analogamente,

$$\begin{aligned} P(T^-|D^c) \approx \frac{1739}{1790} &= 0,9715 \\ P(T^-|D) \approx \frac{8}{40} &= 0,2 \\ P(T^+|D^c) \approx \frac{51}{1790} &= 0,0285 \end{aligned}$$

Nota: Não devemos usar como estimativa de prevalência a relação dos $\frac{40}{1820} = 0,021978$; assim como a de 1820 indivíduos não podem ser representáveis como população em geral. Na realidade, a prevalência de tuberculose na amostra é de $2,1\%$, ou cerca de $2197,8$ em 100000. Isto é, mais de 70 vezes maior do que a prevalência na população geral.
Agora, vamos calcular $P(D|T^+)$:

$$P(D|T^+) = \frac{(P(D \cap T^+))}{(P(T^+))}$$

Uma vez que não conhecemos $P(T^+)$, devemos decompor T^+ em relação a D e D^c como mostra a Figura 2.3.

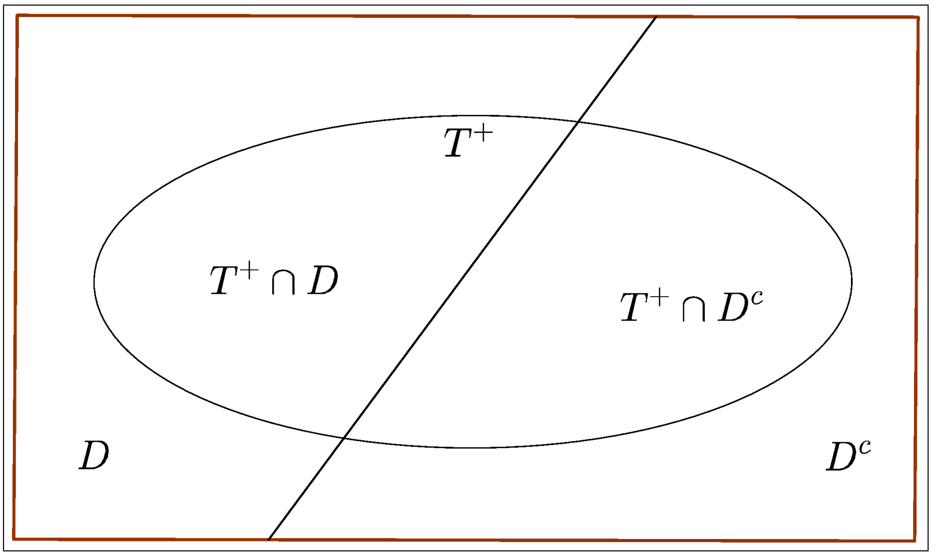

Figura 2.3 *Decomposição de* T^+
Fonte: cardenas.webnode.com

Isto é, de $T^+ = (D \cap T^+) \cup (T^+ \cap D^c)$, temos que

$$\begin{aligned} P(T^+) &= P[(D \cap T^+) \cup (T^+ \cap D^c)]) \\ &= P(D \cap T^+) \cup P(T^+ \cap D^c) \\ &= (T^+|D)P(D) + (T^+|D^c)P(D^c). \end{aligned}$$

Substituindo, resulta

$$
\begin{aligned}
P(D|T^+) &= \frac{P(T^+|D)P(D)}{P(T^+|D)P(D) + P(T^+|D^c)P(D^c)} \\
&= \frac{sensibilidade \times \text{prevalência}}{sensibilidade \times \text{prevalência} + falsopositivo \times (1 - prevalencia)} \\
&= \frac{0,55 \times 0,000093}{0,55 \times 0,000093 + 0,0285 \times 0,999907} \\
&= 0,028497358149.
\end{aligned}
$$

No Exemplo 2.3.1 o espaço amostral foi particionado em pessoas doentes D e pessoas sadias D^c, isto permitiu expressar o tratamento T^+ como a soma de eventos disjuntos $T^+ \cap D$ e $T^+ \cap D^c$. A seguir foi determinada a probabilidade $P(T^+)$, como a soma das probabilidade destes dois eventos, este processo é conhecido como Lei da Probabilidade Total, enquanto o calculo da probabilidade condicional $P(D|T^+)$ é conhecida como Regra de Bayes. A seguir generalizamos este processo.

Definição 2.3.2. *Uma coleção de eventos $E_1, E_2, \ldots\ldots E_k$ é dita uma Partição do Espaço Amostral Ω se verificam*

$$
\begin{aligned}
E_i \cap E_j &= \phi, \quad \forall i \neq j \\
\bigcup_{i=1}^{k} E_i &= \Omega.
\end{aligned}
$$

Teorema 2.3.1. *(Lei Da probabilidade Total). Se $E_1, E_2, \ldots\ldots, E_n$ é uma partição de Ω e A um evento, então $P(A)$ é calculada por*

$$P(A) = \sum_{i=1}^{n} P(E_i)P(A|Ei)$$

Demonstração. Como $\Omega = \bigcup_{i=1}^{n} E_i$, então $A = A \cap \Omega = A \cap \left(\bigcup_{i=1}^{n} E_i\right) = \bigcup_{i=1}^{n}(A \cap E_i)$. Considerando que $(A \cap E_j) \cap (A \cap E_k) = A \cap (E_j \cap E_k) = A \cap \emptyset = \emptyset$ se $j \neq k$, resulta $P(A) = \sum_{i=1}^{n} P(A \cap E_i)$ (ver Figura 4.4). Da probabilidade condicional temos $P(A \cap E_i) = P(A|E_i)P(E_i)$, e portanto $P(A) = \sum_{i=1}^{n} P(A|E_i)P(E_i)$.

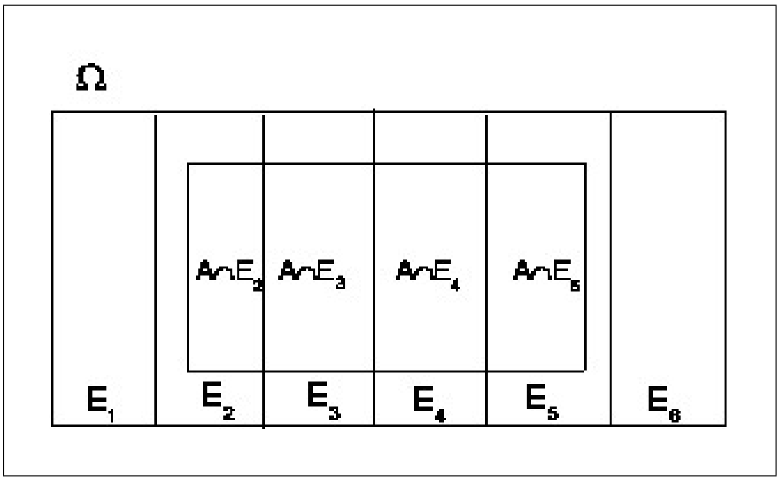

Figura 2.4 *Partição de Ω para k = 6*
Fonte: cardenas.webnode.com

Teorema 2.3.2. *(Bayes): Suponha que os eventos $E_1, E_2, \cdots, E_n$ formam uma partição do espaço amostral Ω e que suas probabilidades sejam conhecidas. Suponha ainda que, para algum evento A se conheçam as probabilidades $P(A|E_i)$ $\forall$ $i = 1, 2, \cdots, n$. Então para qualquer j, temos:*

$$P(E_j|A) = \frac{P(A|E_j)P(E_j)}{\sum_{i=1}^{n} P(E_i)P(A|E_i)}, \quad i = 1, 2, ..., n.$$

Demonstração. Da representação de probabilidade condicional, temos:

$$P(E_j|A) = \frac{P(E_j \cap A)}{P(A)} = \frac{P(A|E_j)P(E_j)}{P(A)},$$

do Teorema da Probabilidade Total, temos:

$$P(E_j|A) = \frac{P(A|E_j)P(E_j)}{\sum_{i=1}^{n} P(E_i)P(A|E_i)}, \quad j = 1, 2, \cdots, n.$$

Exemplo 2.3.2. *[2]: Suponha que em um levantamento de dados, uma determinada população foi classificada de acordo com as características apresentadas na Tabela 2.4. Imagine ainda que, de levantamentos estatísticos anteriores os riscos de transmissão do HIV entre as populações P_1, P_2, P_3, P_4 sejam de 2,3%, 9,3%, 12%, e 17,1%, respectivamente. Se A é o evento HIV^+, então $P(A|P_1) = 0,023$, $P(A|P_2) = 0,12$ e $P(A|P_3) = 0,17$. Daí, para saber o risco de um HIV^+ ser proveniente do grupo de hete-*

P_1	HETEROSSEXUAIS	60%
P_2	HOMOSSEXUAIS	21%
P_3	HEMOFÍLICOS	6%
P_4	USUÁRIOS DE DROGAS	13%

Tabela 2.4 *Levantamento de dados de uma determinada população.*

rossexuais, usamos o teorema de Bayes,

$$\begin{aligned}
P(P_1|A) &= \frac{(P(A|P_i)P(P_i)}{\sum_{i=1}^{4} P(P_i)P(A|P_i)} \\
&= \frac{P(A|P_1)P(P1)}{P(A|P_1)P(P_1) + P(A|P_2)P(P_2) + P(A|P_3)P(P_3) + P(A|P_4)P(P_4)} \\
&= \frac{0,023 \times 0,6}{0,023 \times 0,6 + 0,093 \times 0,21 + 0,12 \times 0,06 + 0,171 \times 0,13} \\
&= \frac{0,0138}{0,0138 + 0,01953 + 0,0072 + 0,02223} \\
&= \frac{0,0138}{0,06276} = 0,2198853 = 21,99\%.
\end{aligned}$$

2.4 Independência

Intuitivamente, a independência entre dois ou mais eventos está no fato de que a ocorrência de qualquer um deles, não intervém na ocorrência dos demais. Desta forma, se dois eventos A e B forem independentes, temos que $P(A|B) = P(A)$ e $P(B|A) = P(B)$.

Definição 2.4.1. *Considere A e B eventos contidos em Ω. Dizemos que estes são independentes se $P(A \cap B) = P(A)P(B)$.*

Exemplo 2.4.1. *Suponhamos que a probabilidade de cura, após t anos de tratamento, de uma determinada doença seja de 62% para os homens, enquanto que para as mulheres essa porcentagem seja de 75%, visto que foi levado em consideração a freqüência e o grau de ingestão de bebida alcoólica, bem como a prática do tabagismo de ambos.*

Considere os eventos:

H: "Homem curado após t anos de tratamento".

M: "Mulher curada após t anos de tratamento".

Com as informações dadas e considerando que os eventos H e M são independentes, podemos calcular as probabilidades das seguintes situações:

a) Ambos obtêm sucesso após o período t de tratamento

$$P(H \cap M) = P(H \mid M)P(M) = P(H)P(M) = 0,62 \times 0,75 = 0,465$$

b) Somente um deles obtenha sucesso ao final do tratamento

$P(H \cap M^c) = P(H \mid M^c)P(M) = P(H)P(M^c) = 0,62 \times 0,25 = 0,155$
$P(H^c \cap M) = P(H^c \mid M)P(M) = P(H^c)P(M) = 0,38 \times 0,75 = 0,285.$
Como $(H \cap M^c) \cap (H^c \cap M) = (H \cap H^c) \cap (M \cap M^c) = \phi \cap \phi = \phi$, temos

$$P[(H \cap M^c) \cup (H^c \cap M)]) = P(H \cap M^c) + P(H^c \cap M) = 0,15 + 0,285 = 0,435$$

c) Nenhum deles tenha sucesso no tratamento

$$P(H^c \cap M^c) = P(H^c \mid M^c)P(M^c) = P(H^c)P(M^c) = 0,38 \times 0,25 = 0,095$$

d) Pelo menos um deles alcance a cura

$$P(H \cup M) = P(H) + P(M) - P(H \cap M) = 0,62 + 0,75 - 0,465 = 0,905$$

2.5 Variáveis Aleatórias

Em muitos experimentos estamos interessados em associar um número real ao espaço amostral, mesmo quando este não for constituído de resultados numéricos (ver Figura 2.5). É neste contexto que se enquadram as variáveis aleatórias. Desta forma, podemos entendê-las da seguinte maneira:

Definição 2.5.1. *(Variável Aleatória): Realizamos um experimento x_i qualquer e obtemos espaço amostral Ω ; Definimos a variável aleatória (v.a.) X como a função X que será encarregada de associar a cada elemento $w \in \Omega$ a um número real X(w).*

$$X : \Omega \to R$$

Definição 2.5.2. *Seja $X : \Omega \to R$ uma variável aleatória:*
a) X é discreta se seu contradomínio for finito ou infinito enumerável.
b) X é contínua se seu contradomínio for um intervalo ou uma coleção de intervalos.

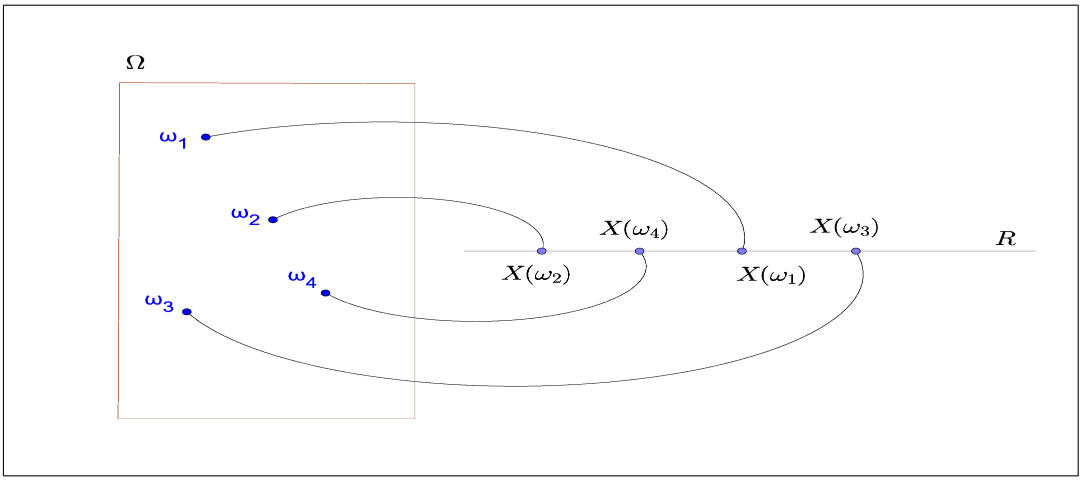

Figura 2.5 *Variável aleatória*
Fonte: cardenas.webnode.com

Definição 2.5.3. *(Função de Probabilidade). Se X é uma variável aleatória discreta, dizemos que $P : X(\Omega) \to R$ é uma função de probabilidades se associa a cada $x \in X(\Omega)$ sua probabilidade $P(x)$.*

Exemplo 2.5.1. *Sejam*

E={lançamento de duas moedas}

$\Omega = \{(c, c), (c, k), (k, c), (k, k)\}$

X: Número de caras obtidas nos dois lançamentos.

Nestas condições a função de probabilidades (ver Figura 2.6 é descrita na Tabela 2.5.

x	0	1	2
P(x)	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Tabela 2.5 *Função de Probabilidades*

Definição 2.5.4. *X uma variável aleatória continua, dizemos que $f : X(\Omega) \to R$ é uma Função Densidade de Probabilidade se verifica as seguintes condições:*

1. $f(x) \geq 0, \quad \forall x \in X(\Omega)$
2. $\int_{X(\Omega)} f(x)dx = 1$

Além mais, definimos para qualquer $a < b$ em $X(\Omega)$ a probabilidade:

$$P(a < X < b) = \int_a^b f(x)dx.$$

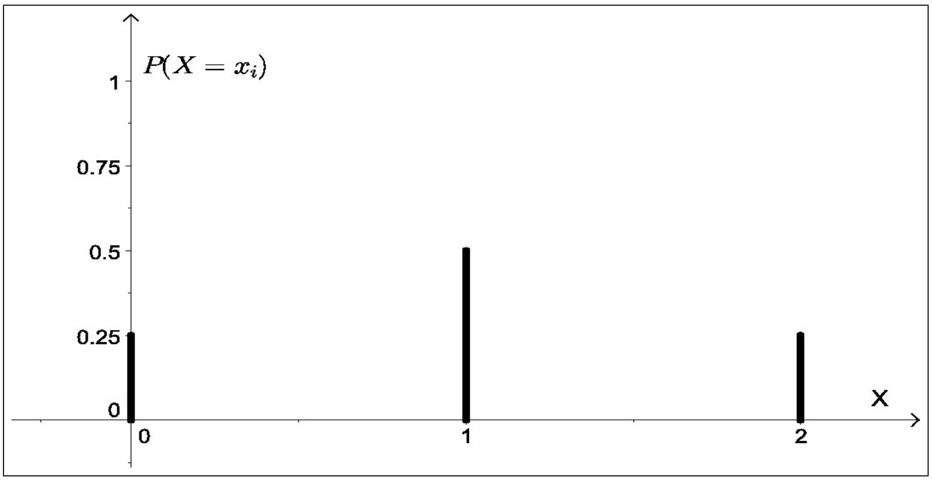

Figura 2.6 *Função de distribuição discreta* ($y = P[X = x_i]$)
Fonte: cardenas.webnode.com

Na Figura 2.7 se ilustra o conceito da função densidade de probabilidade, sendo a área sombreada a probabilidade da variável aleatória assumir valores entre a e b.

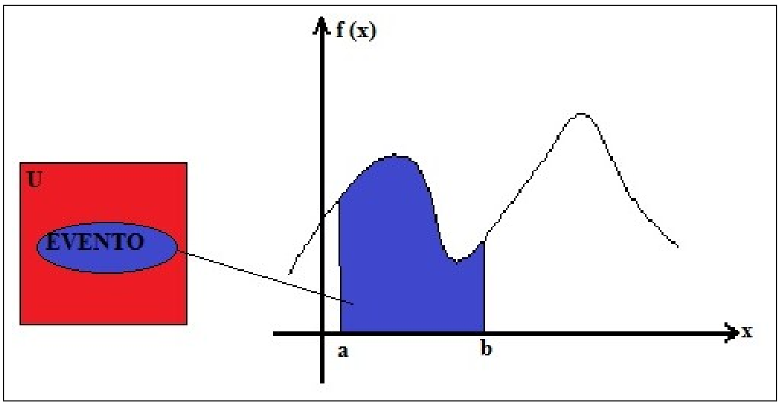

Figura 2.7 *fdp. A área sombreada corresponde a* $P(a < X < b)$.

Exemplo 2.5.2. *O tempo, em minutos, para o efeito de um determinado analgésico é uma variável aleatória contínua* T*, com a seguinte função densidade:*

$$f(t) = \begin{cases} (t-4)/40, & se\ 8 \leq t < 10 \\ 3/20, & se\ 10 \leq t \leq 15 \\ 0, no\ complementar & \end{cases}$$

f é uma fdp pois $f(x) > 0$ e

$$\int_{\infty}^{8} 0dt + \int_{8}^{10} \frac{1}{40}(t-4)dt + \int_{10}^{15} \frac{3}{20}dt + \int_{15}^{\infty} 0dt$$

$$= \frac{1}{40}\left[\frac{t^2}{2} - 4t\right]\Bigg|_8^{10} + \frac{3}{20}t\Bigg|_{10}^{15} = 1.$$

Definição 2.5.5. *(Função de Distribuição Acumulada): Seja X uma variável aleatória. Dizemos que $F : \mathcal{R} \to \mathcal{R}$ é uma função de distribuição acumulada ou função repartição, se*

$$F(x) = P[X \leq x] = \begin{cases} \sum_{x_k \leq x} P[X = x_k], & \textit{se } X \textit{ é discreta} \\ \\ \int_{-\infty}^{x} f(t)dt, & \textit{se } X \textit{ é contínua} \end{cases}$$

No caso discreto temos a seguinte representação:

$$F(X) = \begin{cases} 0, & \text{se } x < x_1 \\ \\ p_1, & \text{se } x < x_2 \\ \\ p_1 + p_2, & \text{se } x < x_3 \\ \\ \vdots \\ \\ p_1 + p_2 + \ldots + p_{n-1}, & \text{se } x < x_n \\ \\ 1, & \text{se } x \geq x_n \end{cases}$$

As Figuras 2.8 e 2.9 ilustram o comportamento da função repartição discreta e contínua respectivamente.

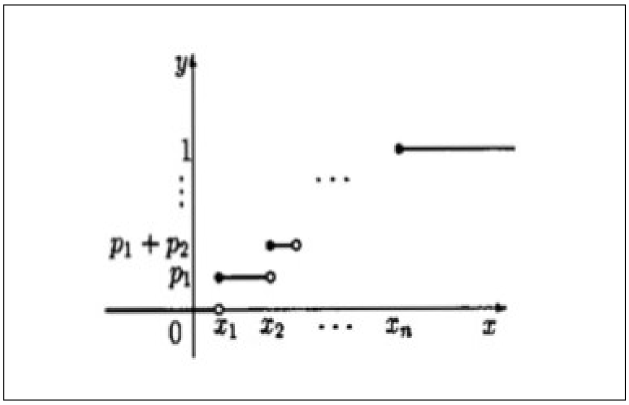

Figura 2.8 *Função repartição discreta*

Propriedade 2.5.1.

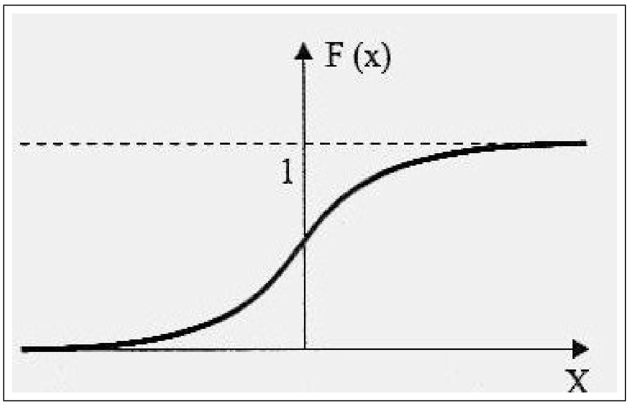

Figura 2.9 *Função repartição contínua*

a) $0 \leq F(x) \leq 1, \forall x \in \mathcal{R}$

b) Em qualquer dos casos, F é monótona não-decrescente.

c) $\lim_{x\to\infty} F(x) = 0$ e $\lim_{x\to+\infty} F(x) = 1$

d) Se X é uma v.a. discreta de forma que $x_1 < x_2 < \ldots$, então

$$P[X = x_i] = F(x_i) - F(x_i - 1).$$

e) Se X é uma v.a. contínua, então

$$P(X = x) = 0, \quad \forall x \in \mathcal{R}.$$

f) Se X é uma variável aleatória $a < b$, então:

i) $P[X \leq a] = F(a)$

ii) $P[X \geq a] = 1 - P[X < a]$

iii) $P[a < x \leq b] = F(b) - F(a)$

iv) $P[a \leq x \leq b] = F(b) - F(a) + P[X = a]$

v) $P[a < x < b] = F(b) - F(a) - P[X = b]$

Exemplo 2.5.3. *Em um laboratório de pesquisas clínicas, foram feitos estudos sobre os níveis de ácido úrico, em* $(mg/100ml)$*, encontrados nos exames bioquímicos de sangue em uma população de* 1000 *pacientes. Os resultados são os seguintes na Tabela 2.6*

Ácido Úrico (mg%)	4,0	4,5	5,0	5,2	5,5	6,0	6,5	7,0	9,0
Frequência	201	30	42	354	29	19	199	59	67

Tabela 2.6 *Níveis de ácido úrico em exames bioquímicos*

Utilizando a idéia de atribuir probabilidade por intermédio da freqüência de ocorrência, a função de probabilidade da variável aleatória discreta nível de ácido úrico é apresentada na Tabela 2.7.

x	4,0	4,5	5,0	5,2	5,5	6,0	6,5	7,0	9,0
$P[X = x_i]$	0,201	0,03	0,042	0,354	0,029	0,019	0,199	0,059	0,067

Tabela 2.7 *Probabilidade do nível de ácido úrico*

Supondo que uma pessoa seja sorteada ao acaso, deseja-se calcular a probabilidade de ter até o nível 5 de ácido úrico. Em outras palavras, deseja-se a função repartição no ponto 5, ou seja a probabilidade acumulada de ocorrência de valores menores ou iguais a 5. Desta forma:

$$F(5) = P[X \leq 5] = P[X = 4] + P[X = 4,5] + P[X = 5] = 0,273$$

E os valores completos da função de distribuição são os seguintes:

$$F(X) = \begin{cases} 0, & \text{se } x < 4,0 \\ 0,201, & \text{se } x < 4,5 \\ 0,231, & \text{se } x < 5,0 \\ 0,273, & \text{se } x < 5,2 \\ 0,627, & \text{se } x < 5,5 \\ 0,656, & \text{se } x < 6,0 \\ 0,675, & \text{se } x < 6,5 \\ 0,874, & \text{se } x < 7,0 \\ 0,933, & \text{se } x < 9,0 \\ 1, & \text{se } x \geq 9,0 \end{cases}$$

Exemplo 2.5.4. *Considere que o efeito, em relação ao tempo, de um determinado anestésico seja dado de acordo com a fdp*

$$f(t) = \begin{cases} \dfrac{2t+1}{12}, & se\ 0 \leq t < 3 \\ \\ 0, & no\ complementar \end{cases}$$

de uma variável aleatória X. Então a função repartição é dada na seguinte forma:

$$f(t) = \begin{cases} 0, & \text{se } t < 0 \\ \\ \displaystyle\int_0^t \frac{2\tau+1}{12} d\tau = \frac{t^2+t}{12}, & \text{se } 0 \leq t < 3 \\ \\ 1, & \text{se } t \geq 3 \end{cases}$$

2.6 Funções de variáveis aleatórias

Há experimentos que apresentam uma relação entre duas ou mais variáveis aleatórias. Isto ocorre por intermédio da associação de seus respectivos eventos entre si e o espaço amostral Ω.

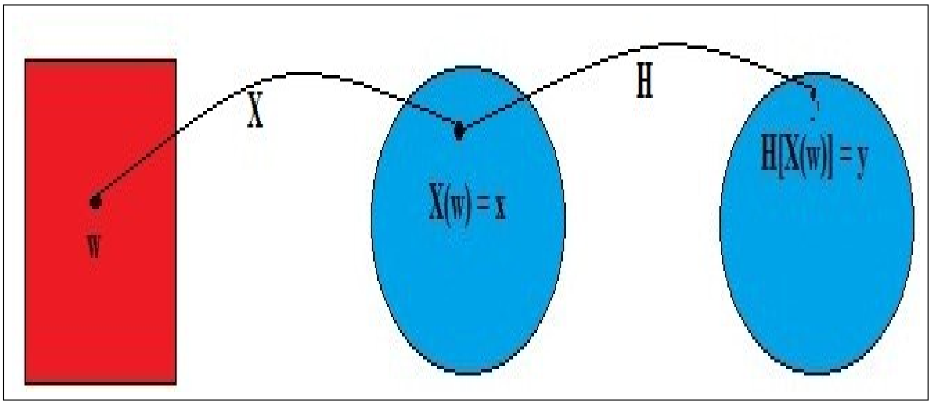

Figura 2.10 *Esboço sistemático de funções de variáveis aleatórias*

Analisando a figura (2.10), podemos definir um evento B no conjunto de valores $X(\Omega)$ a partir do conhecimento de um evento C do conjunto de valores $H(X(\Omega))$, e vice-versa.

$$B = \{x \in X(\Omega) : \quad H(x) \in C\} = \{\omega \in \Omega : \quad H[X(w)] \in C\}$$

Ou seja, B é o conjunto de todos os valores de X tais que $H(x) \in C$. E se B e C forem relacionados deste modo, os denominaremos de eventos equivalentes.

Assim que B e C sejam conhecidas, pode-se estabelecer suas respectivas probabilidades através do manuseamento adequado da definição de eventos equivalentes.

$$\begin{cases} P(B) = & \{\omega \in \Omega : \quad X(w) \in B\} \\ \\ P(C) = & \{x \in X(\Omega) : \quad H(x) \in C\} \end{cases}$$

Quando tratamos de uma v.a. X discreta, e obtemos $Y = H(X)$, decorre da definição de eventos equivalentes que $P[X = x_i] = P[Y = y_i]$, se $y_i = H(x_i))$.

Exemplo 2.6.1. *Considere uma v.a. X, representada na Taabela 2.8.*

x	-$\sqrt{2}$	0	$\sqrt{2}$
$P[X = x_i]$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$

Tabela 2.8 *Função de probabilidades da variável X*

Considere ainda outra v.a. Y tal que $Y = X + 2$. Na Tabela 2.9 se apresentam as probabilidade da variável $Y = X + 2$.

x	2-$\sqrt{2}$	2	$2 + \sqrt{2}$
$P[X = x_i]$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$

Tabela 2.9 *Função de probabilidades da variável $Y = X + 2$*

Veja que para cada valor que X assume, há um correspondente no contradomínio R_y, de forma que as probabilidades são iguais. No entanto, há muitos casos em que a função H não apresenta o mesmo comportamento deste exemplo e por isso, ocorrerá que um ou mais valores de X produzirão os mesmos valores em Y. Em situações como esta, devemos aplicar o seguinte procedimento:

$$P[Y = y_i] = P[X = x_{i_1}] + P[X = x_{i_2}] + \ldots + P[X = x_{i_n}]$$

Exemplo 2.6.2. *No exemplo anterior, considere $Y = X^2$. Na Tabela 2.10 se representa a transformação da variável X para a variável Y, e na Tabela 2.11 sua fun;'ao de probabilidades.*

$$P[Y = 2] = P[X = -\sqrt{2}] + P[X = \sqrt{2}] = 2/3$$

x	$-\sqrt{2}$	0	$\sqrt{2}$
$y = x^2$	2	0	2
$P[Y = y_j]$	?	$\frac{1}{3}$	?

Tabela 2.10 *Determinação das probabilidades de* $Y = X^2$

y	0	2
P[Y = 2]	$\frac{1}{3}$	$\frac{2}{3}$

Tabela 2.11 *Função de probabilidades de* $Y = X^2$

Antes de tratarmos exclusivamente dos casos contínuos, é imprescindível destacar que poderão acontecer situações em que X é uma v.a.d., enquanto que Y é contínua. Assim, para o cálculo das probabilidades de X, é mister que tenhamos o conhecimento da fdp de Y. Desta forma, se $X = x_i$ for equivalente a um evento B de R_y, então

$$P[X = x_i] = \int_B f(x)d_x$$

Agora, daremos o tratamento adequado para as v.a.c. Ou seja, para os casos em que X for uma vac com fdp f e H for apenas uma função contínua. Desta forma, temos que $Y = H(X)$ será uma vac, e para esta, temos de encontrar sua fdp g.

A fim de alcançarmos este objetivo, segue-se o seguinte procedimento:

i) Obter a função repartição de Y, denotada por G, na qual $G(y) = P[Y \leq y]$, o que permite encontrar um evento A qualquer em R_x, que seja equivalente ao evento $Y \leq y$.

ii) Derivar G(y) em relação a y para descobrir g(y).

iii) Definir valores em R_y, para os quais $g(y) > 0$.

Exemplo 2.6.3. *Suponha que X esteja distribuída uniformemente no intervalo* $[0, 1]$*. Nosso objetivo é determinar a função densidade de* $Y = -\lambda^{-1}\log(1 - X)$ *para* $\lambda > 0$*. Para isso, considere G a função de distribuição de Y. Observe ainda que Y é uma variável aleatória positiva e, por essa razão* $G(Y) = 0$ *para* $y \leq 0$*. Para* $y > 0$*, temos:*

$$\begin{aligned} G(y) &= P[Y \leq y] = P[\lambda^{-1}log(1-X) \leq y] \\ &= P[log(1-X) \geq -\lambda y] \\ &= P[1-X \geq e^{-\lambda y}] \\ &= P[X \leq 1-e^{-\lambda y}] = \int_0^{1-e^{-\lambda y}} \frac{1}{1-0}dz \\ &= 1-e-\lambda y \end{aligned}$$

Daí,

$$\begin{aligned} g(y) &= \frac{d}{dy}(1-e^{-\lambda y}) \text{ para } y > 0 \\ g(y) &= 0 \text{ para } y < 0 \end{aligned}$$

Por tanto, a densidade de Y é dada na seguinte forma:

$$g(y) = \begin{cases} \lambda e^{-\lambda y}, & \text{se } y > 0 \\ \\ 0, & \text{se } y \leq 0 \end{cases}$$

A densidade obtida é chamada de densidade da distribuição exponencial de parâmetro λ, a qual receberá o tratamento adequado no Capítulo 4.

2.7 Variáveis aleatórias bidimensionais

Em determinadas circunstâncias, estaremos interessados em analisar dois ou mais resultados, simultaneamente. De tais situações, podemos articular que o tratamento dado aos casos em que o números de variáveis aleatória estudadas em um mesmo experimento é maior ou igual a dois consiste na extensão dos pontos do plano euclidiano [1]. Porém, nos restringiremos somente aos casos que tratam da análise simultânea de duas situações.

Definição 2.7.1. *(variáveis aleatórias bidimensionais): Se em um determinado experimento tivermos duas funções, X e Y, por exemplo, tais que desempenham o papel de associar um número real para cada elemento de Ω, diremos que o vetor (X,Y) define uma variável aleatória bidimensional (vab).*

Definição 2.7.2. *(Variáveis Aleatórias Discretas Bidimensionais): Designamos o vetor aleatório (X,Y) de variável aleatória discreta bidimensional (vadb) , se o número de valores possíveis de (X,Y) forem finitos ou infinitos numeráveis, e ainda, se houver uma função* *de probabilidade P[$X = x_i$, $Y = y_j$], que seja a correspondência de cada realização x,y $\in$ (X, Y) tal que obedeça as seguintes condições:*

i) P[X = x_i, Y = y_j] $\geq$0

ii) 0$\leq$ P[X = x_i,Y = y_j] $\leq$1

iii) $\displaystyle\sum_{j=1}^{\infty}\sum_{i=1}^{\infty}$ P[X = x_i,Y = y_j]=1

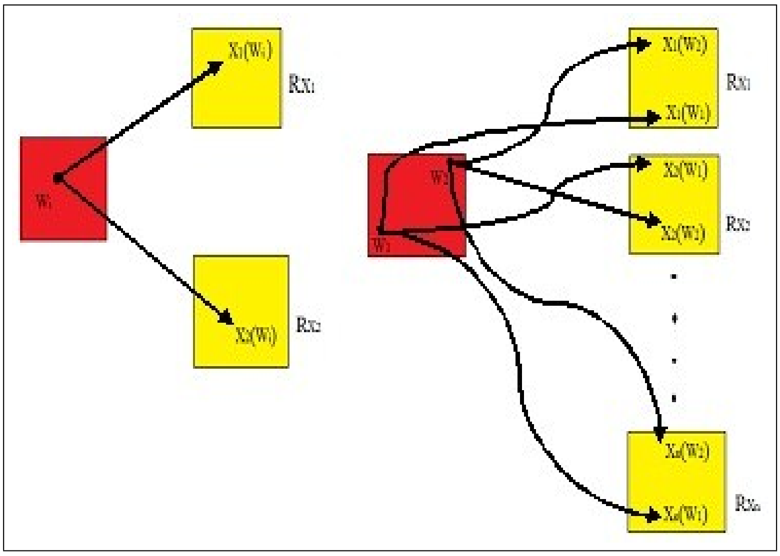

Figura 2.11 *Variáveis aleatórias bidimensionais e multidimensionais, respectivamente*

Definição 2.7.3. *(Variáveis Aleatórias Contínuas Bidimensionais): Chamamos de variável aleatória contínua bidimensional (vacb) o vetor (X, Y) que toma todos os valores em alguma região $\Re$ do plano euclidiano, de forma que exista uma função f, denominada função densidade de probabilidade conjunta tal que satisfaça as condições a seguir:*

i) f(x,y)$\geq$ 0

ii) $\displaystyle\int_{\infty}^{\infty}\int_{-\infty}^{\infty}$ f(x,y)dxdy=1

iii) $\forall$ a, b, c, d $\in$ $\Re$ temos $P[a \leq X \leq b,\ c \leq X \leq d]$= P[$a < X \leq b,\ c < X \leq d$]= P[$a \leq X < b,\ c \leq X < d$]= P[$a < X < b, c < X < d$]=$\int_c^d \left[\int_a^b f(x,y)dx\right] dy$

Definição 2.7.4. *Se (X, Y) é uma vab, podemos estabelecer a função repartição F que, é definida na seguinte forma:*

$$F(x,y) = \begin{cases} \displaystyle\sum_{y_j \leq y} \sum_{x_i \leq x} P[X = x_i, Y = y_j], & \text{se (X,Y) é discreta} \\ \displaystyle\int_{-\infty}^{y} \int_{-\infty}^{x} f(u,v)dudv, & \text{se (X,Y) é contínua} \end{cases}$$

2.7.1 Distribuição Marginal

Quando associamos duas variáveis aleatórias unidimensionais à outra que seja bidimensional, podemos optar por calcular a distribuição de probabilidade de X ou a de Y. Ocorrências desse tipo nos dão a idéia de marginal. Desta forma, se tivermos o objetivo de calcular a distribuição de X, podemos fazer isso desconsiderando Y, e vice-versa. Denotamos a distribuição marginal por $P[X = x_i]$ no caso discreto e $g(x)$ no caso contínuo.

Para o caso discreto, fazemos o uso das seguintes relações:

$$\begin{cases} P[X = x_i] = \displaystyle\sum_{j=1}^{n} P[X = x_i, Y = y_j], & \text{i=1,2,..., m} \\ P[Y = y_j] = \displaystyle\sum_{i=1}^{n} P[X = x_i, Y = y_j], & \text{j=1,2,..., n} \end{cases}$$

Para o caso contínuo, devemos partir do conhecimento da fdp conjunta e estabelecer as seguintes relações:

$$\begin{cases} g(x) = \displaystyle\int_{-\infty}^{\infty} f(x,y)dy \\ h(y) = \displaystyle\int_{-\infty}^{\infty} f(x,y)dx \end{cases}$$

Exemplo 2.7.1.1. *Em uma pesquisa com duas maternidades A e B foram fornecidos dados sobre o número de recém-nascidos portadores de algum tipo de anomalia congênita. Considere que em qualquer das maternidades, esses números representam uma variável aleatória e que (X, Y) seja uma variável aleatória bidimensional que forneça o número de portadores de anomalias, nascidos nas maternidades A e B, respectivamente. Observe a Tabela 2.12:*

Y \ X	0	1	2	3
0	0	0,01	0,01	0,01
1	0,01	0,02	0,03	0,02
2	0,03	0,04	0,05	0,04
3	0,05	0,05	0,05	0,06
4	0,07	0,06	0,05	0,06
5	0,09	0,08	0,06	0,05

Tabela 2.12 *Distribuição da probabilidade conjunta de (X, Y)*

Desta forma, $P[X = 3, Y = 4] = 0,06$ e assim sucessivamente. Para o cálculo das marginais, considere a extensão das linhas e colunas, como na Tabela 2.13.

	X	0	1	2	3	$\sum_{i=1}^{m} P[y_j]$
Y						
0		0	0,01	0,01	0,01	0,03
1		0,01	0,02	0,03	0,02	0,08
2		0,03	0,04	0,05	0,04	0,16
3		0,05	0,05	0,05	0,06	0,21
4		0,07	0,06	0,05	0,06	0,24
5		0,09	0,08	0,06	0,05	0,28
$\sum_{j=1}^{n} P[x_i]$		0,25	0,26	0,25	0,24	1

Tabela 2.13 *Cálculo das marginais de X e de Y*

Observe que cálculo das marginais é feito através da soma das linhas e colunas. Sendo assim, $P[X = 2] = 0,25$ e $P[Y = 5] = 0,28$, e de forma similar para as demais.

Exemplo 2.7.1.2. *Seja a função*

$$f(x,y) = \begin{cases} x^3y + \dfrac{y}{4}, & \textit{se } 0 \leq x \leq 1 \textit{ e } 0 \leq y \leq 2 \\ 0, & \textit{no complementar} \end{cases}$$

definida sobre o vetor aleatório (X,Y). Observe que f é não negativa e

$$\begin{aligned}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y,)dxdy &= \int_0^2\left[\int_0^1 (x^3y+\frac{y}{4})dx\right]dy = \int_0^2\left[\frac{x^4}{4}y+\frac{xy}{4}\right]_0^1 dy\\ &= \int_0^2\left(\frac{y}{4}+\frac{y}{4}\right)dy = \int_0^2\frac{2y}{4}dy = \frac{1}{2}\int_0^2 ydy = \frac{y^2}{4} = 1.\end{aligned}$$

Para calcular as marginais, é só aplicar a definição

$$\begin{aligned}g(x) &= \int_{-\infty}^{\infty} f(x,y)dy = \int_0^2\left(x^3y+\frac{y}{4}\right)dy = \int_0^2 x^3ydy+\frac{1}{4}\int_0^2 ydy\\ &= \left.\frac{(x^3y^2)}{2}\right|_0^2 + \left.\frac{1}{4}\frac{y^2}{2}\right|_0^2 = 2x^3+\frac{1}{2}\end{aligned}$$

Analogamente,

$$\begin{aligned}h(y) &= \int_{-\infty}^{\infty} f(x,y)dx = \int_0^1 (x^3y+\frac{y}{4})dx = \int_0^1 x^3ydx+\int_0^1\frac{y}{4}dx\\ &= \left.\frac{x^4}{4}y\right|_0^1 + \left.\frac{xy}{4}\right|_0^1 = \frac{y}{2}\end{aligned}$$

2.7.2 Distribuição Condicional

Agora que tomamos conhecimento das distribuições marginais, podemos estender os conhecimentos adquiridos nos estudos sobre probabilidade condicionada e estabelecer relações de (X,Y) para os casos discretos e contínuos.
No caso discreto, definimos:

$$\begin{cases} P[X|Y=y_j] = \dfrac{P[X=x_i,Y=y_j]}{P[Y=y_j]}, & P[Y=y_j]\neq 0 \text{ e } j \text{ fixo e } i=1,2,\ldots m\\ \\ P[Y|X=x_i] = \dfrac{P[X=x_i,Y=y_j]}{P[X=x_i]}, & P[X=x_i]\neq 0 \text{ e } i \text{ fixo e } j=1,2,\ldots n\end{cases}$$

Exemplo 2.7.2.1. *Considerando o exemplo(2.7.1.1), da definição de probabilidade condicionada vamos determinar $P[X=0|Y=5]$,*

$$P[X=0|Y=5] = \frac{(P[X=0,Y=5])}{(P[Y=5])} = \frac{0,09}{0,28} = 0,32$$

Analogamente,

$$P[Y=5|X=0] = \frac{(P[X=0,Y=5])}{(P[X=0])} = \frac{0,09}{0,25} = 0,36$$

Para o caso contínuo, considere g e h as fdp marginais de X e Y, respectivamente. Daí, definimos:

$$\begin{cases} g(x|y) = \dfrac{(f(x,y))}{(h(y))}, \quad h(y) > 0 \\ \\ h(y|x) = \dfrac{(f(x,y))}{(g(x))}, \quad g(x) > 0 \end{cases}$$

Exemplo 2.7.2.2. *Considerando o exemplo 2.7.1.2, temos que:*

$$g(x|y) \quad = \quad \frac{f(x,y)}{h(y)} = \frac{x^3y + \frac{y}{4}}{\frac{y}{2}} = \frac{8x^3y+2}{4y}$$

enquanto que,

$$h(y|x) \quad = \quad \frac{f(x,y)}{g(x)} = \frac{x^3y + \frac{y^3}{4}}{2x^3 + \frac{1}{2}} = \frac{4x^3+y}{8x^3+2}$$

Definição 2.7.2.1. *(Variáveis Aleatórias Discretas Independentes): Considere (X, Y) uma vad. Diz-se que X e Y são independentes se* $P[X = x_i, Y = y_j]$*=*$P[X = x_i]P[Y = y_j]$*, para todo par (*x_i*,* y_j*),* $i = 1, 2, \ldots, m$ *e* $j = 1, 2, \ldots, n$*.*

Em outras palavras, X e Y são independentes se, e somente se, $P[X|Y = y_j] = P[X = x_i]$ $\forall$ i e j. Analogamente, $P[Y|X = x_i] = P[Y = y_j]$.

Exemplo 2.7.2.3. *Em uma determinada localidade, foi feito um levantamento sobre a incidência do vírus* HIV^+*em relação ao número de parceiros sexuais. As probabilidades de cada caso são apresentadas na tabela a seguir, em que* X *representa a variável "soro positivo"e* Y *é a variável "parceiros sexuais", ver Tabela 2.14.*
Analisando estas informações, conclui-se que as variáveis X *e* Y *não são independentes, uma vez que, por exemplo:*

$$P[X = 0, Y = 0] \quad = \quad 0,08 \neq P[X = 0]P[Y = 0] = 0,04$$

Y \ X	0	1	2	3 ou +	$\sum_{j=1}^{n} P[x_i]$
0	0	0,08	0,08	0,04	0,02
1	0,12	0,24	0,16	0,28	0,8
$\sum_{i=1}^{m} P[y_j]$	0,2	0,32	0,2	0,28	1

Tabela 2.14 *Distribuição conjunta de (X, Y)*

Definição 2.7.2.2. *(Variáveis Aleatórias Contínuas Independentes): Considere* (X, Y) *uma vac. Diz-se que* X *e* Y *são independentes se, e somente se,* $f(x, y) = g(x)h(y)$*. Ou seja, se, e somente se,* $g(x|y) = g(x)$*. Equivalentemente,* $h(y|x) = h(y)$*.*

Exemplo 2.7.2.4. *[11]: Sejam* X *e* Y *a duração de vida de dois dispositivos eletrônicos. Suponha-se que sua fdp conjunta seja dada por:*

$$f(x, y) = \begin{cases} \dfrac{4}{3}(x + xy), & \text{se x}\in(0,1) \text{ e y}\in(0,1) \\ \\ 0, & \text{no complementar} \end{cases}$$

Integrando em relação a y, obtemos $g(x) = 2x$.
Integrando em relação a x, obtemos $h(y) = \dfrac{2}{3}(1 + y)$.
Daí, conclui-se que X e Y são independentes, pois:

$$f(x, y) = \frac{4}{3}(x + xy) = 2x \cdot \frac{2}{3}(1 + y) = g(x)h(y).$$

2.7.3 Funções de variáveis aleatórias

Para os casos discretos, considere $Z = H_1(X, Y)$, uma função das variáveis aleatórias X e Y. Logo, $Z = Z(\omega) = H_1(X(\omega), Y(\omega))$ é uma função que associa um número real $Z(w)$ a todo resultado $w \in \Omega$, seu cálculo segue os seguintes passos

i) Executamos o experimento ε e obtemos o resultado w.

ii) Calculamos os números $X(w)$ e $Y(w)$.

iii) Calculamos o número $Z = H_1[X(w), Y(w)]$.

Exemplo 2.7.3.1. *Considere que o vetor bidimensional (X, Y) tenha a distribuição dada no exemplo (2.7.1.1). Seja Z = Máx (X, Y) = maior número de casos registrados pelas duas maternidades.*

Note que os valores possíveis de Z, são: 4 e 5. Para calcular $P[Z = 4]$, raciocinaremos que $Z = 0$, se e somente se, algum dos casos acontecer: $Y = 4$, $X = 0$ ou $Y = 4$, $X = 1$ ou $Y = 4$, $X = 2$ ou $Y = 4$, $X = 3$ e para $P[Z = 5]$ temos que: $Y = 5$,$X = 0$ ou $Y = 5$, $X = 1$ ou $Y = 5$, $X = 2$ ou $Y = 5$, $X = 3$. Logo,

$$\begin{aligned} P[Z=4] &= 0,07 + 0,06 + 0,05 + 0,06 = 0,24 \\ P[Z=5] &= 0,09 + 0,08 + 0,06 + 0,05 = 0,28 \end{aligned}$$

logo, a distribuicao de probabilidades da v.a. Z é apresentada na Tabela 2.15.

Z	4	5
P[Z = z]	0,24	0,28

Tabela 2.15 *Variável* $Z = Máx(X, Y)$

Se (X, Y) for um vetor aleatório bidimensional contínuo, e se $Z = H_1(X, Y)$ for uma função contínua de (X, Y), então Z será uma *vac* unidimensional, cuja fdp é dada por:

$$g(z) = \int_{-\infty}^{\infty} h(z, u)du.$$

Para definir a fdp de Z, introduzimos uma segunda variável aleatória $U = H_2(X, Y)$ e por conseguinte, ficou perceptível a necessidade de obtermos a fdp conjunta de Z e U, que ficou descrita como $h(z, u)$ [20]. Em que $h(z, u) = f[G_1(z, w),\ G_2(z, w)]|J(z, w)|$ e $J(z, w)$ é o Jacobiano da transformação $(x, y) \longrightarrow (z, w)$. Sendo ainda, $z = H_1(x, y)$ e $w = H_2(x, y)$ equações univocamente resolvidas para x e y, em termos de z e w ou seja $x = G_1(z, w)$ e $y = G_2(z, w)$.

$$J(z, w) = det \begin{vmatrix} \dfrac{\partial x}{\partial z} & \dfrac{\partial x}{\partial w} \\ & \\ \dfrac{\partial y}{\partial z} & \dfrac{\partial y}{\partial w} \end{vmatrix}$$

Exemplo 2.7.3.2. *[12]: Sejam* X *e* Y *variáveis aleatórias contínuas, cuja densidade de probabilidade conjunta é* $f(x, y)$. *Sejam,*

$$\begin{cases} R = \sqrt{x^2 + y^2}, & \text{para r}\geq 0 \\ \Phi = \arctan \dfrac{y}{x}, & \text{para } 0 < \Phi < 2\pi \end{cases}$$

Para encontrar a densidade de probabilidade conjunta de (R, Φ), façamos $X = H_1(R, \Phi)$ e $Y = H_2(R, \Phi)$. Assim, transformação que induze (x, y) em (R, Φ) no domínio referido é biunívoca e sua inversa é $x = H_1(r, \Phi) = rcos\Phi$ e $Y = H_2(r, \Phi) = rsen\Phi$. Daí, o jacobiano de transformação fica expresso na seguinte maneira:

$$J(r, \Phi) = det \begin{vmatrix} \dfrac{\partial x}{\partial r} & \dfrac{\partial x}{\partial \Phi} \\ \dfrac{\partial y}{\partial r} & \dfrac{\partial y}{\partial \Phi} \end{vmatrix} = det \begin{vmatrix} \cos\Phi & -r\sin\Phi \\ \sin\Phi & r\cos\Phi \end{vmatrix} = r$$

Consequentemente,

$$f(r, \Phi) \quad = \quad rf(r\cos\Phi, r\sin\Phi).$$

Capítulo 3

MOMENTOS

Agora que temos todas as ferramentas necessárias, iremos nos dedicar ao estudo das características adicionais (ou momentos), inclusos nas examinações de fenômenos indeterminísticos. As informações buscadas a apartir dessas características, também são conhecidas como parâmetros que uma vez associados a uma variável aleatória, população ou amostra, exercem a função de informar os diferentes comportamentos que possuem. No geral, quanto maior o número de momentos, maior é a informação que temos sobre a distribuição de probabilidade. Sendo assim, possuem grande significância no que tange ao estudo global dos fenômenos aleatórios.

3.1 Momento de Ordem 1

É um dos mais importantes parâmetros das variáveis aleatórias, onde sua característica é a de ser o centro de distribuição de probabilidade dos valores da v.a. X e, por essa razão, costuma-se chama-lo de medida de tendência central ou esperança matemática, ou ainda, média [21].

Definição 3.1.1. *Considere X uma variável aleatória com função de distribuição de probabilidades $P(X)$ no caso discreto ou com função densidade de probabilidades $f(x)$ no caso contínuo. A Esperança Matemática de $\varphi(X)$, denotada por $E(\varphi(X))$ é definida na seguinte forma:*

$$E(\varphi(X)) = \begin{cases} \displaystyle\sum_{x_i \in R_X} \varphi(x_i) P\left[X = x_i\right], & \textit{se } X \textit{ é discreta} \\ \displaystyle\int_{-\infty}^{\infty} \varphi(x) f(x) dx, & \textit{se } X \textit{ é contínua} \end{cases}$$

Em particular se $\varphi(X) = X$, a esperança matemática de X, denotada por $E(X)$ ou μ_X, é

$$E(X) = \begin{cases} \displaystyle\sum_{x_i \in R_x} x_i P\left[X = x_i\right], & \text{se } X \text{ é discreta} \\ \displaystyle\int_{-\infty}^{\infty} x f(x) dx, & \text{se } X \text{ é contínua} \end{cases}$$

Proposição 3.1.1. *Se X é uma variável aleatória, então a esperança matemática verifica as seguintes propriedades:*

a) $E[aX + b] = aE(X) + b$, onde a e b são constantes reais

b) $E(c) = c$, onde c é uma constante real

c) $E(cX) = cE(X)$, onde c é uma constante real

d) $E(X - \mu_X) = 0$

Demonstração

a) Se X é uma variável aleatória discreta e $\varphi(X) = aX + b$, temos

$$\begin{aligned} E(\varphi(X)) = E(aX + b) &= \sum_{i=1}^{n} \varphi(x_i) P\left[X = x_i\right] \\ &= \sum_{i=1}^{n} (ax_i + b) P\left[X = x_i\right] \\ &= a \sum_{i=1}^{n} x_i P\left[X = x_i\right] + b \sum_{i=1}^{n} P\left[X = x_i\right] \\ &= aE(X) + b.1 = aE(x) + b \end{aligned}$$

Se X é uma variável aleatória contínua e $\varphi(X) = aX + b$, temos

$$\begin{aligned} E(\varphi(X)) = E(aX + b) &= \int_{-\infty}^{\infty} \varphi(ax + b) f(x) dx \\ &= \int_{-\infty}^{\infty} (ax + b) f(x) dx \\ &= a \int_{-\infty}^{\infty} x f(x) dx + b \int_{-\infty}^{\infty} f(x) dx \\ &= aE(X) + b \cdot 1 = aE(X) + b. \end{aligned}$$

b) Fazendo $a = 0$ e $b = c$ em a) temos:

$$E(c) = E(0 \cdot X + c) = 0 \cdot E(X) + c = 0 + c = c.$$

c) Fazendo $a = c$ e $b = 0$ em a) temos:

$$E(c) = E(c \cdot X + 0) = c \cdot E(X) + 0 = cE(X) + 0 = cE(X).$$

d) Fazendo $a = 1$ e $b = \mu_X$ em a) temos:

$$\begin{aligned} E(X - \mu_X) &= E(1 \cdot X + (-\mu_X)) = 1 \cdot E(X) - \mu_X \\ &= E(X) - \mu_X = \mu_X - \mu_X = 0. \quad \blacksquare \end{aligned}$$

No caso de variáveis aleatórias conjuntas X e Y, definimos a Esperança Matemática da função $h(X, Y)$:

Definição 3.1.2. *Sejam X, Y variáveis aleatórias com distribuição conjunta de probabilidades $P(X, Y)$ no caso discreto ou função densidade conjunta f(x,y) no caso contínuo. A Esperança Matemática da função $h(X, Y)$ é definida como:*

$$E[h(X,Y)] = \begin{cases} \displaystyle\sum_{x_i \in R_X} \sum_{y_j \in R_Y} h(x_i, y_j) P[X = x_i, Y = y_j], & \textit{se } X, Y \textit{ são discretas} \\ \displaystyle\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x,y) f(x,y) dx dy, & \textit{se } X \textit{ e } Y \textit{ são contínuas.} \end{cases}$$

Proposição 3.1.2. *Se X é Y são variáveis aleatórias, então a esperança matemática verifica as seguintes propriedades:*

a) $E(H_1(X, Y) + H_2(X, Y)) = E(H_1(X, Y)) + E(H_2(X, Y))$.

b) $E(aX + bY) = aE(X) + bE(Y)$, *onde a e b são constantes reais.*

c) $E(X \pm Y) = E(X) \pm E(Y)$.

d) Se X e Y são duas variáveis independentes, então

$$E(XY) = E(X) \cdot E(Y)$$

Demonstração

a) Se X e Y são variáveis aleatórias discretas e $h(X,Y) = H_1(X.Y) + H_2(X,Y)$, temos

$$\begin{aligned} E(H_1(X,Y) + H_2(X,Y)) &= \sum_{i=1}^{m}\sum_{j=1}^{n} H_1(x_i, y_j)\ P[X = x_i, Y = y_j] \\ &\quad + \sum_{i=1}^{m}\sum_{j=1}^{n} H_2(x_i, y_j)\ P[X = x_i, Y = y_j] \\ &= E(H_1(X,Y)) + E(H_2(X,Y)) \end{aligned}$$

Se X e Y são variáveis aleatórias contínuas e $h(X,Y) = H_1(X.Y) + H_2(X,Y)$, temos

$$\begin{aligned} E(H_1(X,Y) + H_2(X,Y)) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [H_1(x,y)]\, f(x,y) dx dy \\ &\quad + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [H_2(x,y)]\, f(x,y) dx dy \\ &= E(H_1(X,Y)) + E(H_2(X,Y)). \end{aligned}$$

b) Se X e Y são variáveis aleatórias discretas e $h(X,Y) = aX + bY$, temos

$$\begin{aligned} E(aX + bY) &= \sum_{i=1}^{m}\sum_{j=1}^{n}(ax_i + by_j) P\left[X = x_i, Y = y_j\right] \\ &= \sum_{i=1}^{m}\sum_{j=1}^{n} ax_i\ P[X = x_i, Y = y_j] \\ &\quad + \sum_{i=1}^{m}\sum_{j=1}^{n} by_j\ P[X = x_i, Y = y_j] \\ &= a\sum_{i=1}^{m} x_i \sum_{j=1}^{n}\ P[X = x_i, Y = y_j] \\ &\quad + b\sum_{j=1}^{n} y_j\ \sum_{i=1}^{m} P[X = x_i, Y = y_j] \\ &= a\sum_{i=1}^{m} x_i\ P[X = x_i] + b\sum_{j=1}^{n} y_j\ P[Y = y_j] \\ &= aE(X) + bE(Y) \end{aligned}$$

Se X e Y são variáveis aleatórias contínuas e $h(X,Y) = aX + bY$, temos

$$\begin{aligned}
E(aX+bY) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(ax+by)f(x,y)xdy \\
&= a\int_{-\infty}^{\infty} x\left(\int_{-\infty}^{\infty} f(x,y)dy\right)dx \\
&\quad +b\int_{-\infty}^{\infty} y\left(\int_{-\infty}^{\infty} f(x,y)dx\right)dy \\
&= a\int_{-\infty}^{\infty} xg(x)dy + b\int_{-\infty}^{\infty} yh(y)dy \\
&= aE(X) + bE(Y).
\end{aligned}$$

c) Fazendo $a = 1$ e $b = \pm 1$ em b) resulta

$$E(X \pm Y) = E(1 \cdot X + (\pm)Y) = 1 \cdot E(X) + (\pm)E(Y) = E(X) \pm E(Y)$$

d) Se X e Y são variáveis aleatórias discretas independentes, então

$$P\left[X = x_i, Y = y_j\right] = P\left[X = x_i\right] P\left[Y = y_j\right].$$

Considerando $h(X,Y) = XY$ resulta

$$\begin{aligned}
E(XY) &= \sum_{i=1}^{m}\sum_{j=1}^{n} x_i y_j P\left[X = x_i, Y = y_j\right] \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n} x_i y_j P\left[X = x_i\right] P\left[Y = y_j\right] \\
&= \sum_{i=1}^{m} x_i P\left[X = x_i\right] \sum_{j=1}^{n} y_j P\left[Y = y_j\right] \\
&= E(X)E(Y)
\end{aligned}$$

Se X e Y são variáveis aleatórias contínuas independentes, então $f(x,y) = g(x)h(y)$. Considerando $h(X,Y) = XY$ resulta

$$\begin{aligned}
E(XY) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyg(x)h(y)dxdy \\
&= \int_{-\infty}^{\infty} xg(x)dx \int_{-\infty}^{\infty} yh(y)dy \\
&= E(X)E(Y). \quad \blacksquare
\end{aligned}$$

Vamos agora definir a probabilidade condicional de uma variável aleatória em relação a um valor de outra variável aleatória, o qual permite definir a probabilidade condicional entre duas variáveis aleatórias.

Definição 3.1.3. *Sejam as variáveis aleatórias X e Y com distribuição de probabilidade conjunta $P(X,Y)$ no caso discreto ou função densidade conjunta $f(X,Y)$ no caso contínuo. Definimos*

a) Se X e Y são variáveis aleatórias discretas

$$\begin{aligned} E(X|Y=y_j) &= \sum_{i=1}^{m} x_i P\left[X=x_i|Y=y_j\right] = \sum_{i=1}^{m} x_i \frac{P\left[X=x_i, Y=y_j\right]}{P\left[Y=y_j\right]} \\ E(X|Y) &= \sum_{j=1}^{n} E(X|Y=y_j) \end{aligned}$$

analogamente,

$$\begin{aligned} E(Y|X=x_i) &= \sum_{j=1}^{n} y_j P\left[Y=y_j|X=x_i\right] = \sum_{j=1}^{n} y_j \frac{P\left[X=x_i, Y=y_j\right]}{P\left[X=x_i\right]} \\ E(Y|X) &= \sum_{i=1}^{m} E(Y|X=x_i) \end{aligned}$$

b) Se X e Y são variáveis aleatórias contínuas

$$\begin{aligned} E(X|Y=y) &= \int_{-\infty}^{\infty} x g(x|y) dx = \int_{-\infty}^{\infty} x \frac{f(x,y)}{h(y)} dx, \qquad h(y)>0 \\ E(X|Y) &= \int_{-\infty}^{\infty} E(X|Y=y) dy \end{aligned}$$

analogamente

$$\begin{aligned} E(Y|X=x) &= \int_{-\infty}^{\infty} y h(y|x) dy = \int_{-\infty}^{\infty} y \frac{f(x,y)}{g(x)} dy, \qquad h(y)>0 \\ E(Y|X) &= \int_{-\infty}^{\infty} E(Y|X=x) dx. \end{aligned}$$

Proposição 3.1.3. *A esperança condicional do vetor aleatório (X,Y) verifica*

a) $E[E(X|Y)] = E(X)$

b) *Se X e Y são variáveis aleatórias independentes, então $E(X|Y) = E(X)$.*

Demonstração.

a) Para o caso discreto:

$$\begin{aligned} E(X|Y) &= \sum_{i=1}^{m} x_i P\left[X = x_i | Y = y_j\right] \\ &= \sum_{i=1}^{m} x_i \frac{P[X = x_i, Y = y_j]}{P[Y = y_j]}, P[Y = y_j] > 0, \end{aligned}$$

logo,

$$\begin{aligned} E[E(X|Y)] &= \sum_{j=1}^{n} E(X|Y) P[Y = y_j] \\ &= \sum_{j=1}^{n} \left(\sum_{i=1}^{m} x_i \frac{P[X = x_i, Y = y_j]}{P[Y = y_j]} \right) P[Y = y_j] \\ &= \sum_{j=1}^{n} \sum_{i=1}^{m} x_i P[X = x_i, Y = y_j] \\ &= \sum_{i=1}^{m} x_i \left(\sum_{j=1}^{n} P[X = x_i, Y = y_j] \right) \\ &= \sum_{i=1}^{m} x_i P[X = x_i] = E(X) \end{aligned}$$

Para o caso contínuo:

Considere f a fdp conjunta do vetor aleatório (X, Y) e h a fdp da marginal de Y. Daí, por definição, temos que:

$$E(X|Y) = \int_{-\infty}^{\infty} x g(x|y) dx = \int_{-\infty}^{\infty} x \frac{f(x, y)}{h(y)} dx, \quad h(y) > 0$$

Por isso,

$$\begin{aligned} E[E(X|Y)] &= \int_{-\infty}^{\infty} E(X|Y) h(y) dy = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} x \frac{f(x, y)}{h(y)} dx \right) h(y) dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x, y) dx dy = \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f(x, y) dy \right) dx \\ &= \int_{-\infty}^{\infty} x g(x) dx = E(X) \end{aligned}$$

b) No caso discreto:

$$\begin{aligned}
E(X|Y) &= \sum_{i=1}^{m} x_i P[X = x_i|Y = y_j] \\
&= \sum_{i=1}^{m} x_i \frac{P[X = x_i, Y = y_j]}{P[Y = y_j]}, \quad P[Y = y_j] > 0 \\
&= \sum_{i=1}^{m} x_i \frac{P[X = x_i][Y = y_j]}{P[Y = y_j]} \\
&= E(X).
\end{aligned}$$

No caso contínuo

$$E(X|Y) = \int_{-\infty}^{\infty} xg(x|y)dx = \int_{-\infty}^{\infty} x\frac{f(x,y)}{h(y)}dx$$

Da hipótese, X e Y são independentes, e portanto $f(x,y) = g(x)h(y)$. Logo,

$$E(X|Y) = \int_{-\infty}^{\infty} x\frac{g(x)h(y)}{h(y)}dx = \int_{-\infty}^{\infty} xg(x)dx = E(X).\blacksquare$$

3.2 Momento de Ordem 2

A partir do conhecimento da média, e do cálculo do momento de ordem 2, podemos estabelecer outro parâmetro, que tem a função de fornecer o nível de concentração de probabilidade em torno do valor esperado.

Definição 3.2.1. *Se X é uma variável aleatória com função de probabilidade ou fdp então o seu momento de ordem 2 é definido pela seguinte relação:*

$$E(X^2) = \begin{cases} \sum_{x_i \in R_x} x_i^2 P[X = x_i], & \textit{se } X \textit{ é discreta} \\ \int_{-\infty}^{\infty} x^2 f(x)dx, & \textit{se } X \textit{ é contínua.} \end{cases}$$

Definição 3.2.2. *Seja X uma variável aleatória com função de probabilidade P ou fdp f e média $E(X) = \mu_x$. Define-se variância de X como sendo o valor esperado da v.a. $(X - \mu x)^2$.*

$$Var(X) = E[(X - \mu_X)^2] = \begin{cases} \sum_{i=1}^{n} (X - \mu_x)^2 P[X = x_i], & \textit{se } x \textit{ é discreta} \\ \int_{-\infty}^{\infty} (X - \mu_x)^2 f(x)dx, & \textit{se } x \textit{ é contínua.} \end{cases}$$

Verifica-se a seguinte propriedade.

Proposição 3.2.1.

a) $Var(X) = E[X^2] - (E[X])^2$

b) $Var(X + c) = Var(X)$, *se* c *é uma constante real*

c) $Var(cX) = c^2Var(X)$, *sendo* c *uma constante real*

d) $Var(c) = 0$, *sendo* c *uma constante real*

e) $Var(aX + b) = a^2Var(X)$.

Demonstração

a)

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

b) Considere a variável aleatória $X + c$

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

c) A variância da variável aleatória cX é,

$$\begin{aligned} Var(cX) &= E[(cX) - E(cX))^2] \\ &= E[(cX) - cE(X))^2] \\ &= E[c^2 (X - E(X))^2] \\ &= c^2E[(X - E(X))^2] \\ &= c^2Var(X) \end{aligned}$$

d) $Var(c) = E[c^2] - (E[c])^2 = c^2 - c^2 = 0.$

e) De b) e c) temos:

$$\begin{aligned} Var(aX + b) &= Var(aX) \\ &= a^2 Var(X). \blacksquare \end{aligned}$$

No caso de um vetor aleatório vamos definir o conceito de covariância e correlação.

Definição 3.2.3. *O parâmetro que exerce a função de determinar o grau de dependência entre duas variáveis aleatórias X e Y, definidas no mesmo espaço de probabilidade, é chamado de covariância, denotado por $Cov(X,Y)$ ou σ_{XY}, se sua expressão for dada na seguinte forma:*

$$Cov(X,Y) = \sigma_{XY} = E[(X - E(X))(Y - E(Y))].$$

Se verifica a seguinte propriedade:

Proposição 3.2.2.

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

b) $Var(X \pm Y) = Var(X) + Var(Y) \pm 2Cov(X,Y)$

c) *Se a e b são reais, então $Var(aX \pm bY) = a^2 var(X) + b^2 var(Y) \pm 2abCov(X,Y)$.*

Demonstração

a)

$$\begin{aligned} Cov(X,Y) &= E[(X - E(X))(Y - E(Y))] \\ &= E[XY - XE[Y] - YE[X] + E[X]E[Y]] \\ &= E[XY] - E[X]E[Y] - E[X]E[Y] + E[X]E[Y] \\ &= E[XY] - E[X]E[Y] \end{aligned}$$

b)

$$\begin{aligned}
Var(X \pm Y) &= E[(X \pm Y)^2] - [E(X \pm Y)]^2 \\
&= E[X^2 \pm 2XY + Y^2] - [(E(X))^2 \\
&\quad \pm 2E(X)E(Y) + (E(Y))^2] \\
&= E[X^2] - (E[X])^2 + E[Y]^2 - (E[Y])^2 \\
&\quad \pm 2(E[XY] - E[X]E[Y]) \\
&= Var(X) + Var(Y) \pm 2Cov(X,Y)
\end{aligned}$$

c) De vi) e vii) temos,

$$\begin{aligned}
var(aX \pm bY) &= Var[aX] + Var[bY] \pm 2Cov(aX, bY) \\
&= a^2 Var[X] + b^2 Var[Y] \pm 2(E(aXbY) - E[aX]E[bY]) \\
&= a^2 Var[X] + b^2 Var[Y] \pm 2abCov(X,Y)
\end{aligned}$$

Definição 3.2.4. *(Desvio Padrão). Seja X uma variável aleatória, discreta ou contínua. Seu desvio padrão é definido como a raiz da sua variância:*

$$\sigma_x = \sqrt{Var(X)}$$

Definição 3.2.5. *(Coeficiente de Correlação). Considere o vetor bidimensional (X,Y). O parâmetro que mede o grau de associação entre X e Y é chamado de coeficiente de correlação e é definido na seguinte forma:*

$$\rho = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} \quad \textbf{\textit{ou}} \quad \rho = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y}.$$

Quando X e Y são independentes, ou seja prossuem a propriedade $E(XY) = E(X)(Y)$, dizemos que elas são descorrelacionadas. Logo, sua covariância é nula. Consequentemente, o coeficiente de correlação também o é.

Podemos assim reafirmar que, a covariância mede, de alguma maneira, o grau de dependência entre X e Y. Ela tem o inconveniente de depender das unidades de medida. Sejam a e b duas constantes reais. Se em vez de X e Y tivermos as variáveis aX e aY. Decorre, imediatamente da definição que $Cov(aX, aY) = abCov(X,Y)$. É, justamente, para evitar esse tipo de inconveniência que fazemos o uso do coeficiente de correlação [23].

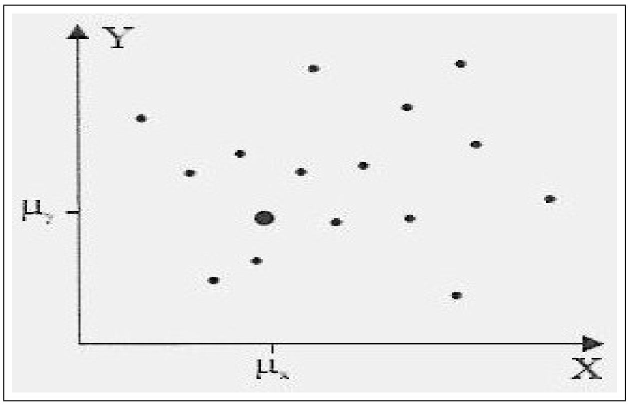

Figura 3.1 *Comportamento grafico de uma vab quando $\rho = 0$*

Proposição 3.2.3. *O coeficiente de correlação das variáveis X e Y verifica:*

a) $|\rho(X,Y)| \leq 1$

b) *Se $\rho(X,Y) = 1$, a ligação entre X e Y é linear: $Y = \alpha X + \beta$, onde $\alpha > 0$.*

c) *Se $\rho(X,Y) = -1$, a ligação entre X e Y é linear: $Y = \alpha X + \beta$, onde $\alpha < 0$.*

Demonstração.

a) Sendo a uma constante real, considere a seguinte expressão:

$$E\left[(a(X - E[X]) + (Y - E[Y]))^2\right] = \tag{3.1}$$

$$= a^2 E[(X - E(X))^2] + 2aE[(X - E(X)(Y - E(Y)] \tag{3.2}$$

$$+E[(Y - E(Y))^2] \tag{3.3}$$

Por ser a esperança matemática de um quadrado, essa expressão é sempre maior ou igual a zero. Portanto,

$$a^2\sigma^2(X) + 2aCov(X,Y) + \sigma^2(Y) \geq 0 \tag{3.4}$$

Como o trinômio é não negativo, seu discriminante fica:

$$\Delta = 4Cov^2(X,Y) - 4\sigma^2(X)\sigma^2(Y) \leq 0 \tag{3.5}$$

Dividindo por $4\sigma^2(X)\sigma^2(Y)$, resulta:

$$\rho^2(X,Y) \leq 1$$

Consequentemente,

$$|\rho(X,Y)| \leq 1.$$

b) Se $\rho\ (X,Y) = 1$, decorre da definição do coeficiente de correlação, que:

$$Cov(X,Y) = \sigma(X)\sigma(Y)$$

Substituindo em (3.1), vemos que $\Delta = 0$. Logo, o trinômio tem raiz dupla que é igual a:

$$a = \frac{-\sigma(Y)}{\sigma(X)}$$

Substituindo a em (3.1), resulta:

$$E\left[\left(\left[\frac{-\sigma(Y)}{\sigma(X)}(X - E(X)\right] + (Y - E(Y)\right)^2\right] = 0$$

Porém, como uma $v.a.$ não negativa tem esperança nula se, e somente se, ela for identicamente nula, temos:

$$\frac{-\sigma(Y)}{\sigma(X)}(X - E[X]) + (Y - E[Y]) = 0.$$

Reescrevendo, segue que:

$$Y - E(Y) = \frac{\sigma(Y)}{\sigma(X)}(X - E(X))$$

que é a equação da reta do tipo $Y = \alpha X + \beta$, onde $\alpha > 0$.

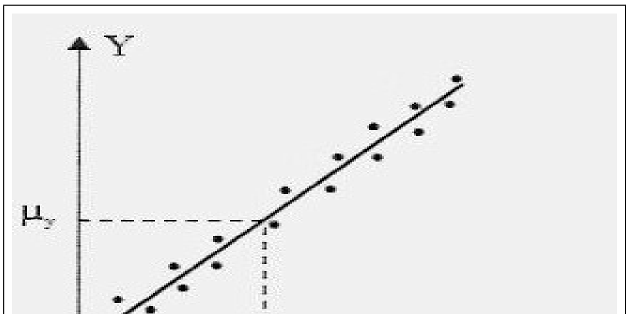

Figura 3.2 *Comportamento grafico de uma vab quando* $\rho = +1$

c) Se $\rho(X,Y) = -1$, temos de modo similar a b) a equação

$$Y - E(Y) = -\frac{\sigma(Y)}{\sigma(X)}(X - E(X))$$

que é a equação da reta do tipo $Y = \alpha X + \beta$, onde $\alpha < 0$. ■

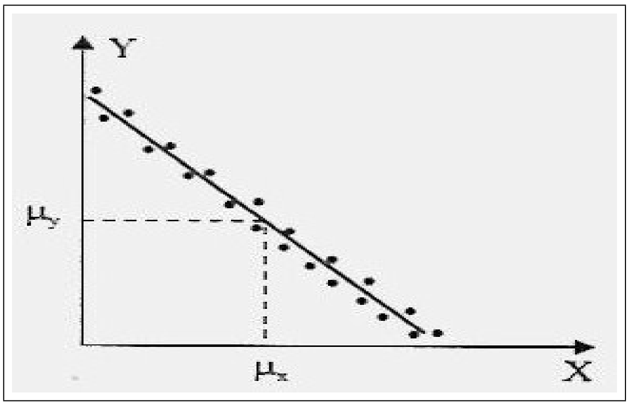

Figura 3.3 *Comportamento gráfico de uma vab quando* $\rho = -1$

Quanto mais próximo ρ for de $+1$ ou de -1, maior é o grau de dependência entre as $v.a.$ e maior se torna a confiabilidade de escrever uma variável em função de outra, por meio do processo dos mínimos quadrados por exemplo. [21].

O conceito de Momento pode-se estender para maior ordem, segundo a seguinte definição:

Definição 3.2.6. *O momento n de uma variável aleatória, para n = 0,1,2,..., em relação à média é definido por*

$$\mu_n = E[(X-\mu)^n].$$

Para $n = 0, 1, 2$, verifica-se

$$\begin{cases} n = 0 & \Rightarrow \mu_0 = 1 \\ n = 1 & \Rightarrow \mu_1 = 0 \\ n = 2 & \Rightarrow \mu_2 = \sigma^2. \end{cases}$$

O momento de ordem n se escreve por extenso como,

$$\mu_n = E[(X-\mu)^n] = \begin{cases} \sum_{i=1}^{k}[(X-\mu)^n]\, P\,[X = x_i]\,, & \text{para variável discreta} \\ \\ \int_{-\infty}^{\infty}[(X-\mu)^n]\, f(x)dx, & \text{para variável contínua} \end{cases}$$

3.3 Função Geratriz de Momentos

Considere o desenvolvimento da função e^x da série de Maclaurim:

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots + \frac{x^n}{n!} \ldots$$

Como a série converge para todos os valores de x, temos:

$$e^{tx} = 1 + tx + \frac{(tx)^2}{2!} + \frac{(tx)^3}{3!} + \ldots + \frac{(tx)^n}{n!} + \ldots$$

Definimos a função $M_X(t)$ como,

$$M_X(t) = E(e^{tX}) = E\left[1 + tX + \frac{(tX)^2}{2!} + \frac{(tX)^3}{3!} + \ldots + \frac{(tX)^n}{n!} + \ldots\right]$$

Como t é uma constante, podemos escrever:

$$M_X(t) = E[e^{tX}] = 1 + tE[X] + \frac{t^2}{2!}E[X^2] + \frac{t^3}{3!}E[X^3] + \ldots + \frac{t^n}{n!}E[X^n] + \ldots$$

O motivo da denominação função geratriz de momentos está no fato de ser possível determinar os momentos a partir de uma estrutura que consiste num processo de derivação, da expressão acima, em relação a t, e após isso, a avaliação para $t = 0$.

$$\frac{dM_X(t)}{dt} = E(X) + tE(X^2) + t^2\frac{E(X^3)}{2!} + \ldots + t^{n-1}\frac{E(X^n)}{(n-1)!} + \ldots$$

Fazendo-se $t = 0$, resulta:

$$\frac{dM_X(0)}{dt} = E(X)$$

Prosseguindo neste raciocínio, obteremos

$$\frac{d^2M_X(t)}{dt} = E(X^2) + tE(X^3) + \ldots + t^{n-2}\frac{E(X^n)}{(n-2)!} + \ldots$$

Para $t = 0$, temos:

$$\frac{d^2M_X(0)}{dt} = E(X^2)$$

Generalizando, por admitir que n exista, temos:

$$\frac{d^nM_X(0)}{dt} = E(X^n)$$

Definição 3.3.1. *Considere X uma variável aleatória com função de distribuição de probabilidade, ou com função densidade. Para tais situações a função geratriz de momentos é definida na forma:*

$$M_X(t) = E(e^{tx}) = \begin{cases} \sum_{i=1}^{k} e^{tx}P[X = x_i], & \text{se X é discreta} \\ \int_{-\infty}^{\infty} e^{tx}f(x)dx, & \text{se X é contínua} \end{cases}$$

Teorema 3.3.1. *Se Y é uma variável aleatória tal que $Y = \alpha X + \beta$, sendo que X tem fgm. Então a fgm de Y é dada por:*

$$M_Y(t) = e^{\beta t} M_X(\alpha t).$$

Demonstração.

$$\begin{aligned} M_Y(t) &= E(e^{Yt}) = E(e^{(\alpha X+\beta)t}) \\ &= e^{\beta t} E(e^{\alpha t X}) = e^{\beta t} M_X(\alpha t). \blacksquare \end{aligned}$$

Verificam-se os seguintes Teoremas:

Teorema 3.3.2. *Considere Z uma variável aleatória tal que $Z = X + Y$, sendo que X e Y são variáveis aleatórias independentes. Então $M_Z(t) = M_X(t) M_Y(t)$.*

Demonstração.

$$\begin{aligned} M_Z(t) &= E(e^{tZ}) = E(e^{t(X+Y)}) \\ &= E(e^{(tX)}))E(e^{tY}) = M_X(t) M_Y(t). \blacksquare \end{aligned}$$

Teorema 3.3.3. *Sejam X e Y são duas v.a. com fgm $M_X(t)$ e $M_Y(t)$, respectivamente. Se $M_X(t) = M_Y(t)$ para todos os valores de t, então elas têm a mesma distribuição de probabilidade.*

Demonstração.
Considere R_X e R_Y os contradomínios de X e Y respectivamente e $P_X(x)$, $P_Y(y)$ suas funções de probabilidades. Considere ainda o conjunto $B = R_X \cup R_Y$ e $b_1, b_2, \cdots, B_n$ elementos de B. Temos,

A fgm de X pode-se escrever como

$$\begin{aligned} M_X(t) &= E[e^{tx}] = \sum_{x \in R_X} e^{tx} P_X(X = x) = \sum_{i=1}^{k} e^{tb_i} P_X(b_i) \\ P_X(b_i) &= 0, \qquad \text{se } b_i \notin R_X \end{aligned}$$

Analogamente, a fgm de Y pode-se escrever como

$$\begin{aligned} M_Y(t) &= E[e^{ty}] = \sum_{x \in R_Y} e^{ty} P_Y(Y = y) = \sum_{i=1}^{k} e^{tb_i} P_Y(b_i) \\ P_Y(b_i) &= 0, \qquad \text{se } \; b_i \notin R_Y \end{aligned}$$

Se X e Y têm a mesma fgm, logo para qualquer t pertenecente a uma vizinhança fechada de zero $V(0)$, temos

$$\sum_{i=1}^{k} e^{tb_i} P_X(b_i) = \sum_{i=1}^{k} e^{tb_i} P_Y(b_i).$$

Reajustando os termos resulta

$$\sum_{i=1}^{k} e^{tb_i} [P_X(b_i) - P_Y(b_i)] = 0 \Leftrightarrow P_X(b_i) - P_Y(b_i) = 0, \quad t \in V(0).$$

Dai conclui-se que as funções de probabilidade de X e Y são iguais e portanto elas têm a mesma distribuição de probabilidades. ■

Capítulo 4

MODELOS PROBABILÍSTICOS

Como salientamos na introdução, estamos trabalhando com fenômenos representantes de incertezas. É lógico que entre esses acontecimentos há diferenças em formas de ocorrências, daí a necessidade de estudarmos cada modelo probabilístico bem como seus momentos, por intermédio da própria definição e pela função geratriz de momentos (fgm), de forma detalhada. Sendo assim, nos casos discretos, trataremos com todos os valores possíveis que podem ser assumidos por uma variável aleatória (v.a.) bem como suas respectivas probabilidades. Agora, quando formos tratar dos casos contínuos, devemos lembrar que estamos apresentando o espaço amostral "idealizado", no qual todos os números reais possíveis (em algum intervalo especificado ou um conjunto de intervalos) podem ser observados como resultados prováveis.

4.1 Distribuição de Bernoulli

Considere uma experiência, cujo resultado implique em dois valores que chamamos, por conveniência, de sucesso e fracasso.

Definição 4.1.1. *Considere X uma v.a.d. Diz-se que X segue o modelo de Bernoulli quando atribuímos os valores 0 e 1 ao seu conjunto de valores possíveis, em que 0 poderá representar a ocorrência de fracasso, enquanto que 1, a ocorrência de sucesso e p representando a probabilidade de sucesso.*

$$P\left[X = x_i\right] = p^x(1-p)^{1-x}; x = 0, 1e0 \leq p \leq 1$$

Veja que a expressão acima é uma função de distribuição de probabilidade, pois:

$$\sum_{i=0}^{1} p^x(1-p)^{1-x} = p^0(1-p)^1 + p^1(1-p)^0 = (1-p) + p = 1$$

Denotamos $X \sim Ber(p)$, se a variável aleatória X segue uma distribuição de Bernoulli com probabilidade p de sucesso.

Exemplo 4.1.1. *Ao fazer um estudo com os jovens concluintes do ensino médio, sobre a preferência do curso de matemática para o vestibular da UNIFAP, podemos fazer a exposição dos resultados conforme cada jovem tenha preferência ou não pelo curso de matemática.*

Neste caso,

$X = 0$, *se o jovem não tem preferência pelo curso de matemática.*

$X = 1$, *se o jovem tem preferência pelo curso de matemática.*

4.1.1 Esperança Matemática e Variância

Seja $X \sim Ber(p)$,

a) pela definição de esperança matemática e variância, temos:

$$\begin{aligned} E(X) &= \sum_{i=0}^{1} xip^x(1-p)^{1-x} = 0p^0(1-p)^1 + 1p^1(1-p)^0 = p \\ E(X^2) &= \sum_{i=0}^{1} xi^2p^2(1-p)^{1-x} = 0^2p^0(1-p)^1 + 1^2p^1(1-p)^0 = p \end{aligned}$$

Aplicando estes resultados na expressão da variância, vem:

$$Var(X) = p - p^2 = p(1-p)$$

b) pela função geratriz de momentos, temos:

$$\begin{aligned} M_X(t) &= \sum_{i=0}^{1} e^{tx}p^x(1-p)^{1-x} = \sum_{i=0}^{1}(pe^t)^x(1-p)^{1-x} = \left[pe^t + (1-p)\right] \\ M'_X(t) &= pe^t \\ M''_X(t) &= M''_X(t) = pe^t \end{aligned}$$

Assim,

$$E(X) = M'{}_X(0) = p$$

E como as derivas são iguais então, $M''_X(0) = p = E(X^2)$. Daí,

$$\begin{aligned} Var(X) &= E(X^2) - [E(X)]^2 \\ &= p - p^2 = p(1-p) \end{aligned}$$

Temos provado pela definição e pelo uso de momentos o seguinte teorema:

Teorema 4.1.1. *Se $X \sim Ber(p)$ com probabilidade de sucesso p, então:*

$$\begin{aligned} E(X) &= p \\ Var[X] &= p(1-p) \end{aligned}$$

4.2 Distribuição Binomial

Considere a seguinte situação: temos um experimento aleatório, do qual pode resultar um sucesso A com probabilidade p, ou um fracasso A^c, com probabilidade $q = 1 - p$. Daí, um número n de provas é realizado tal que, observando-se que cada vez se obtêm A ou A^c, resulta no espaço amostral Ω constituído pelos 2^n elementos seguintes:

$$AAA\ldots AA, AAA\ldots AA^c, AAA\ldots A^cA, \ldots, A^cA^cA^c\ldots A^c,$$

em que cada elemento goza de n letras, obtidas entre A e A^c, de todas as formas possíveis, e levando-se em consideração a ordem. Assim, o número k de sucessos A, em cada prova, é o que dá origem à variável aleatória X, cujos valore possíveis são 0, 1, 2, ..., n com as respectivas probabilidades, p_0, p_1,..., p_n [23].

Definição 4.2.1. *Considere n repetições independentes, e do mesmo tipo, de um experimento ε. Diz-se que uma variável X, expressa em termos binários, possui a distribuição binomial, com parâmetros n e p se sua função de distribuição é expressa na seguinte forma:*

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

em que as n repetições de ε são denominadas como provas de Bernoulli ou n ensaios de Bernoulli.

Denotamos $X \sim B(n, p)$, se a variável aleatória X segue uma distribuição binomial com probabilidade de sucesso p e n provas de Bernoulli.

Observe que a soma dos termos $P[X = k]$ é igual à unidade:

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^{n} d\frac{n!}{(n-k)!k!} p^k (1-p)^{n-k} = [p + (1-p)]^n = 1^n = 1$$

A expressão $\binom{n}{k} = d\frac{n!}{(n-k)!k!}$ corresponde às combinações de n em k, e é também conhecida como coeficiente binomial de n em k.

Exemplo 4.2.1. *No exemplo anterior, podemos estar interessados no número de vezes em que a preferência dada foi para o curso de matemática. Para isso, considere, por exemplo, que um grupo de* 3 *jovens seja escolhido e que a probabilidade de preferência seja de* 60%. *Veja a Figura* (4.1)

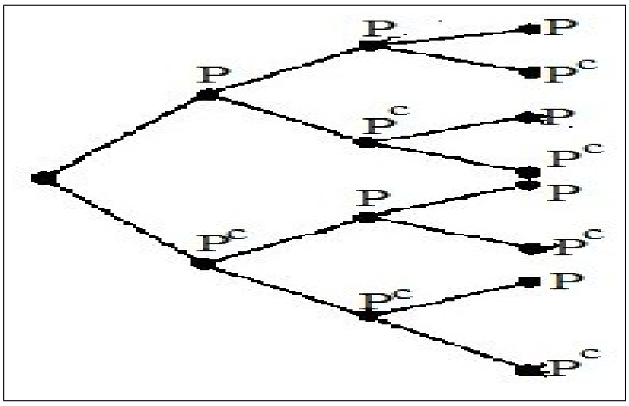

Figura 4.1 *Árvore de probabilidades das preferências do curso para* 3 *jovens*

PPP	0,6 × 0,6 × 0,6	$0,6^3$
PPP^C	0,6 × 0,6 × 0,4	$0,6^2$ x 0,4
PP^CP^C	0,6 × 0,4 × 0,4	$0,4^2$ x 0,6
P^CPP	0,4 × 0,6 × 0,6	$0,6^2$ x 0,4
P^CPP^C	0,6 × 0,6 × 0,4	$0,4^2$ x 0,6
P^CP^CP	0,4 × 0,4 × 0,6	$0,4^2$ x 0,6
$P^CP^CP^C$	0,4 × 0,4 × 0,4	$0,4^3$

Tabelas 4.1:Probabilidades de preferência da matemática para 3 pessoas

Onde p representa o evento ”preferência pelo curso de matemática”. Desta forma, Xfica representada assim:

X	0	1	2	3
P[X = xi]	$0,4^3$	3 × 0,6× $0,4^2$	3× $0,6^2$× 0,4	$0,6^3$

Tabelas 4.2:Variável aleatória ”Preferência pelo curso de matemática”

$$P[X = k] = \binom{3}{k} \times 0,6^k \times 0,4^{3-k}, k = 1, 2, 3$$

4.2.1 Esperança Matemática e Variância

Seja $X \sim B(n, p)$,

a) pela definição de esperança matemática e variância, temos:

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

pois, a componente para $k = 0$ se anula. Considerando a identidade $\binom{n}{k} = n\binom{n-1}{k-1}$ temos.

Considere $j = k - 1$ e a identidade. Daí,

$$E(X) = \sum_{k=1}^{n} n\binom{n-1}{k-1}p^{k-1}p(1-p)^{n-1-(k-1)}$$

e fazendo a substituição $j = k - 1$ resulta,

$$E[x] = np\sum_{j=0}^{n-1} n\binom{n-1}{j}p^j(1-p)^{n-1-j} = np\,[p + (1-p)]^{n-1} = np.1^{n-1} = np,$$

do binômio de Newton tem-se a expressão para a esperança matemática

$$E[x] = np\,[p + (1-p)]^{n-1} = np.1^{n-1} = np.$$

A seguir avaliamos pela definição o termo $E[X^2]$:

$$\begin{aligned}
E(X^2) &= \sum_{k=0}^{n} k^2\binom{n}{k}p^k(1-p)^{n-k} = \sum_{k=0}^{n}[k(k-1)+k]\binom{n}{k}p^k(1-p)^{n-k} \\
&= \sum_{k=0}^{n} k(k-1)\binom{n}{k}p^k(1-p)^{n-k} + \sum_{k=0}^{n} k\binom{n}{k}p^k(1-p)^{n-k} \\
&= \sum_{k=2}^{n} k(k-1)\frac{n(n-1)(n-2)!}{(n-k)!k(k-1)(k-2)!}p^k(1-p)^{n-k} + E(X) \\
&= n(n-1)p^2\sum_{k=2}^{n}\binom{n-2}{k-2}p^{k-2}(1-p)^{(n-2)-k} + np
\end{aligned}$$

Considerando-se $j = k-2$ vem.

$$\begin{aligned}
&= n(n-1)p^2\sum_{j=0}^{n-2}\binom{n-2}{j}p^j(1-p)^{(n-2)-j} + np \\
&= n(n-1)p^2[p+(1-p)]^{n-2} + np = n^2p^2 + np(1-p)
\end{aligned}$$

Aplicando esses resultados em

$$\begin{aligned}
Var(X) &= E(X^2) - [E(X)]^2, \text{segue-se que} \\
Var(X) &= n^2p^2 + n(n-1)p^2 - n^2p^2 = np(1-p)
\end{aligned}$$

b) Avaliamos a esperança e a variância pelo uso de momentos,

$$\begin{aligned}
M_X(t) &= \sum_{k=0}^{n} e^{tk}\binom{n}{k}p^k(1-p)^{n-k} \\
&= \sum_{k=0}^{n}\binom{n}{k}(pe^t)^k(1-p)^{n-k} \\
&= [pe^t + (1-p)]^n
\end{aligned}$$

Agora, considere $u = pe^t + (1-p)$. Por Leibniz temos que

$$\frac{dM_X}{dt} = \frac{dM_X}{du}\cdot\frac{du}{dt} = nu^{n-1}\cdot pe^t = n[pe^t + (1-p)]^{n-1}pe^t$$

Assim,

$$E(X) = M'_X(0) = n[p+(1-p)]^{n-1}p = n1^{n-1}p = np.$$

E o momento de ordem 2 advém de $M'_X(t)$, calculado na forma a seguir:

$$\frac{d^2M_X}{dt} = \{n[pe^t + (1-p)]^{n-1}\}', pe^t + n[pe^t + (1-p)]^{n-1}(pe^t)'$$

Considerando novamente $u = pe^t + (1-p)$ na primeira parcela, e fazendo o mesmo procedimento de derivação já utilizado, segue:

$$\frac{d^2M_X}{dt} = n(n-1)[pe^t + (1-p)]^{n-2}(pe^t)^2 + n[pe^t + (1-p)]^{n-1}(pe^t)$$

Pondo $t = 0$ obtemos:

$$\frac{d^2M_X(0)}{dt^2} = n(n-1)p^2 + np$$

Substituindo os valores encontrados em $Var(X) = E(X^2) - [E(X)]^2$ teremos:

$$\begin{aligned} Var(X) &= n(n-1)p^2 + np - n^2p^2 \\ &= np[(n-1)p + (1-np)] \\ &= np[np - p + 1 - np] \\ &= np(1-p). \end{aligned}$$

Observa-se que o cálculo da esperança matemática e variância foi avaliada diretamente pelo método dos momentos, sem precisar de manipulação das somatórias e mudança de índices como foi o caso na avaliação direta pela definição. Estes resultados são resumidos no seguinte teorema:

Teorema 4.2.1. *Se $X \sim B(n,p)$ com probabilidade de sucesso p, e n provas de Bernoulli, então:*

$$\begin{aligned} E[X] &= np \\ Var[X] &= np(1-p) \end{aligned}$$

4.3 Distribuição Geométrica

Imagine uma ou várias situações que consistem na persistência para se alcançar um objetivo, de forma que sejam levados em consideração os fracassos antecedentes ao sucesso. É justamente que para situações como essa, que o modelo a seguir é aplicável e eficiente.

Definição 4.3.1. *Seja X uma variável aleatória que consiste de uma sequência de ensaios de Bernoulli. Dizemos que X possui distribuição geométrica quando sua distribuição de probabilidade $P[X = k]$ mede a probabilidade para obter o primeiro sucesso no k-ésimo ensaio de Bernoulli.*

Em outras palavras, a variável X é definida como o número de tentativas necessárias até que se obtenha a primeira ocorrência de um evento Desta maneira, $X = k$ se, e somente se, as primeiras $k-1$ repetições apontarem o fracasso. Logo, se p é a probabilidade de sucesso de um ensaio de Bernoulli, a distribuição de probabilidade geométrica é representada na seguinte forma:

$$\begin{aligned} \sum_{k=1}^{\infty} P[X = k] &= \sum_{k=1}^{\infty} p(1-p)^{k-1} = p\sum_{k=1}^{\infty}(1-p)^{k-1} \\ p\left[\frac{1}{1-(1-p)}\right] &= p\left[\frac{1}{p}\right] = 1 \end{aligned}$$

Exemplo 4.3.1. *Observe a figura a seguir:*

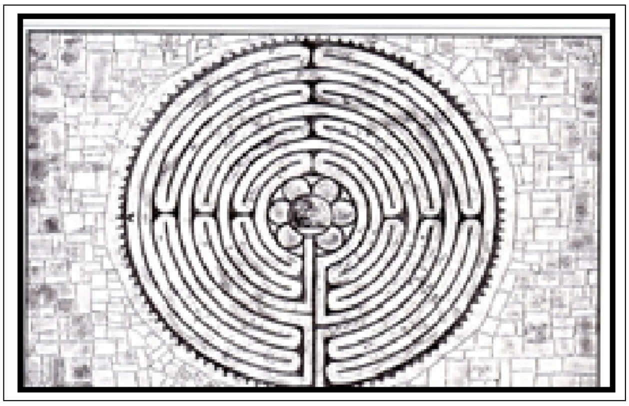

Figura 4.2 *Jogo de tabuleiro Disponível em:< http://www.labirintosnosotao.com/*

Trata-se de um jogo que consiste na movimentação do tabuleiro até uma bola de gude alcance o centro num determinado período de tempo t.

Considerando que a probabilidade p, do primeiro acerto, seja igual a $0,2$ e admitindo a independência de um jogo para outro, é perceptível que a ocorrência do primeiro sucesso após onze tentativas fracassadas é obtida na seguinte forma:

$$\begin{aligned} P[X = 11] &= (1-p)^{11-1}p \\ &= (1-0,2)^{10}0,2 \\ &= 0,8^{10} \times 0,2 \approx 0,02 \end{aligned}$$

4.3.1 Esperança Matemática e Momentos

Seja $X \sim G(p)$,

a) pela definição de esperança matemática e variância, temos:

$$\begin{aligned} E(X) = \sum_{k=1}^{\infty} kpq^{k-1} &= p\sum_{k=1}^{\infty} \frac{d}{dq}(q^k) = p\frac{d}{dq}\left[\sum_{k=1}^{\infty}(q^k)\right] \\ p\frac{d}{dq}\left[\frac{1}{(1-q}\right] = p\frac{[-1(-1)]}{[1-q]^2} &= \frac{p}{[1-q]^2} = \frac{p}{p^2} = \frac{1}{p}. \end{aligned}$$

O termo $E(X^2)$ é calculado pela expressão:

$$\begin{aligned} E(X^2) &= \sum_{k=1}^{\infty} k^2pq^{k-1} = p\sum_{k=1}^{\infty}\frac{d}{dq}(kq^k) \\ = p\frac{d}{dq}\left[\sum_{k=1}^{\infty}(kq^k)\right] &= p\frac{d}{dq}\left[\frac{q}{(1-q)}\sum_{k=1}^{\infty}kq^{k-1}p\right] \\ = p\frac{d}{dq}\left[\frac{q}{(1-q)}E(X)\right] &= p\frac{d}{dq}\left[\frac{q}{(1-q)}\cdot\frac{1}{(1-q)}\right] = p\frac{d}{dq}\left[\frac{q}{(1-q)^2}\right] \\ = p\frac{((1-q)^2+2q(1-q))}{(1-q)^4} &= p\frac{((1-q)\left[(1-q)+2q\right])}{(1-q)^4} \\ = p\frac{((1+q))}{(1-q)^3} &= \frac{(1+q)}{p^2} \end{aligned}$$

Logo,

b) Avaliamos a esperança e a variância pelo uso de momentos,

$$\begin{aligned} M_X(t) &= \sum_{k=1}^{\infty} e^{tk}pq^{k-1} \\ &= \sum_{k=1}^{\infty}\frac{e^{tk}pq^k}{q} = \sum_{k=1}^{\infty}\frac{p(qe^t)^k}{q} \\ &= \frac{p}{q}\sum_{k=1}^{\infty}(qe^t)^k = \frac{p}{q}qe^t[1+qe^t+(qe^t)^2+(qe^t)^3+\ldots] \\ &= \frac{p}{q}\left[\frac{qe^t}{1-qe^t}\right] = \frac{pe^t}{1-qe^t} \end{aligned}$$

Assim,

$$
\begin{aligned}
M'_X(t) &= \frac{(pe^t)'(1-qe^t)-(pe^t)(1-qe^t)'}{(1-qe^t)^2} = \frac{pe^t(1-qe^t)-(pe^t)(-qe^t)}{(1-qe^t)^2} \\
&= \frac{pe^t[1-qe^t+qe^t]}{(1-qe^t)^2} = \frac{pe^t}{(1-qe^t)^2} \\
M''_X(t) &= \frac{(pe^t)'(1-qe^t)^2-(pe^t)[(1-qe^t)^2]'}{[(1-qe^t)^2]^2} \\
&= \frac{pe^t(1-qe^t)^2-2pe^t(1-qe^t)(-qe^t)}{(1-qe^t)^4} \\
&= \frac{pe^t[1-qe^t][(1-qe^t)+2qe^t]}{(1-qe^t)^4} = \frac{pe^t[(1-qe^t)+2qe^t]}{(1-qe^t)^3} = \frac{pe^t(1+qe^t)}{(1-qe^t)^3}
\end{aligned}
$$

Consequentemente,

$$
\begin{aligned}
M'_X(0) &= E(X) = \frac{p}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p} \\
M''_X(0) &= E(X^2) = \frac{p(1+q)}{(1-q)^3} = \frac{p(1+q)}{p^3} = \frac{(1+q)}{p^2}
\end{aligned}
$$

E,

$$
Var(X) = \frac{(1+q)}{p^2} - \left[\frac{1}{p}\right]^2 = \frac{q}{p^2}
$$

Observa-se novamente que o cálculo da esperança matemática e variância foi avaliada diretamente pelo método dos momentos, sem precisar de manipulação das somatórias e mudança de índices como foi o caso na avaliação direta pela definição. Estes resultados são resumidos no seguinte teorema:

Teorema 4.3.1. *Se $X \sim G(p)$ com probabilidade de sucesso p, então:*

$$
\begin{aligned}
E[X] &= \frac{1}{p} \\
Var[X] &= \frac{q}{p^2}
\end{aligned}
$$

4.4 Distribuição de Pascal

Seja o experimento aleatório composto por repetições de ensaios de Bernoulli independentes, todos com probabilidade de "sucesso" p, resultantes da seleção aleatória e com reposição de elementos até obter r sucessos. O espaço amostral é dado por Ω $\{a_1$, a_2, $\ldots$, a_k: $a_k = S$ e $(r-1)$ dos ai são S, $i < k, k \geq r\}$ [3].

Definição 4.4.1. *Dizemos que X tem distribuição de Pascal quando o experimento é repetido independentemente até que o evento um evento A ocorra pela r-ésima vez. Assim, X: número de tentativas necessárias para que A ocorra r vezes [21]. Se $X = k$, o evento ocorre pela r-ésima vez na repetição de número k. Sendo assim, a expressão fica definida na seguinte maneira:*

$$P[X=k] = \binom{k-1}{r-1} p^r (1-p)^{k-r}$$

Em que

- A ocorre $(r-1)$ vezes nas $(k-1)$ repetições anteriores.
- $P(A) = P$ (sucesso)
- $P(A^C) = 1-p$ (fracasso)

Notação: $X \sim Pa(p)$

É evidente que para os casos em que n = 1, X tem distribuição geométrica; o que nos faz compreender que a distribuição de Pascal é uma generalização desta.

Exemplo 4.4.1. *[21] Suponha que a probabilidade de um sinal de trânsito estar aberto, em uma esquina, seja de 20*

$$\begin{aligned} P[X=10] &= \binom{10-1}{4-1} 0,2^4 (1-0,2)^6 \\ \binom{9}{3} 0,2^4 (0,8)^6 &= 0,035232 \end{aligned}$$

Exemplo 4.4.2. $X \sim Pa(k,p)$

i)

$$\begin{aligned} E(X^n) &= \sum_{k=r}^{\infty} k^n \binom{k-1}{r-1} p^r (1-p)^{k-r} \\ &= \frac{r}{p} \sum_{k=r}^{\infty} k^{n-1} \binom{k}{r} p^{r+1} (1-p)^{k-r} = \\ &= \frac{r}{p} \sum_{m=r+1}^{\infty} (m-1)^{n-1} \binom{m-1}{r} p^{r+1} (1-p)^{m-(r+1)} \\ &= \frac{r}{p} E[(X-1)^{n-1}] \end{aligned}$$

Em que $X \sim Pa(r+1, p)$.

Utilizamos as identidades $k\binom{k-1}{r-1} = r\binom{k}{r}$ *e colocamos* $m = k+1$. *Assim , para* $n = 1$ *no resultado acima, decorre que:*

$$E(X) = \frac{r}{p}$$

E, para $n = 2$*:*

$$E(X^2) = \frac{r}{p}E(X-1) = \frac{r}{p}\left(\frac{r+1}{p} - 1\right)$$

Logo,

$$Var(X) = \frac{r}{p}\left(\frac{r+1}{p} - 1\right) - \left(\frac{r}{p}\right)^2 = \frac{r(1-p)}{p^2}$$

4.5 Distribuição Hipergeométrica

O modelo que estudaremos agora envolve uma população, finita, de elementos de forma que as observações sejam feitas por intermédio de uma seqüência, da qual o objetivo seja o número de sucessos.

Definição 4.5.1. *Diz-se que* X *tem distribuição hipergeométrica quando a tratamos como um modelo para amostragem sem reposição de uma população com um número finito de elementos, de forma que cada elemento possa ser um, de dois tipos.*

Se a população possui N elementos, sendo M de um tipo e $N - M$ de outro, então podemos representá-la por um problema, de maneira que a expressão fique definida na seguinte forma:

$$P[X = k] = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$$

em que n representa o número de elementos retirados.

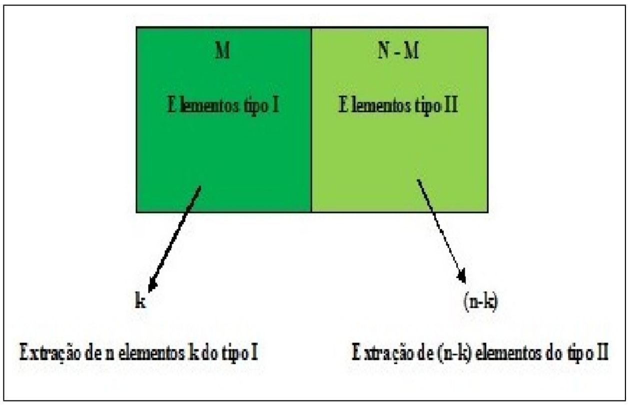

Figura 4.3 *Representação sistemática da distribuição hipergeométrica*

Notação: $H \sim (p, M, N - M, n)$

Nota: Quando o tamanho da população é muito maior do que a amostra (ou seja, N é muito maior que n) a distribuição hipergeométrica é razoavelmente bem aproximada pela distribuição binomial com parâmetros n (número de tentativas) e$p = K/N$ (probabilidade de sucesso numa tentativa única) [33].

Exemplo 4.5.1. *Para um determinado concurso de música inscreveram-se* 100 *pessoas, das quais 15 possuem experiência na área musical, enquanto que as demais não. Dez pessoas são escolhidas, por sorteio, para se apresentarem na abertura do evento. Qual é a probabilidade de que cinco, das dez escolhidas para a apresentação, sejam experientes ?*

$$P[X = 5] \quad = \quad \frac{\binom{15}{5}\binom{85}{5}}{\binom{100}{10}} = \frac{\frac{15!85!}{10!5!80!5!}}{\frac{100!}{90!10!}}$$

Exemplo 4.5.2. $X \sim H(n, N, k)$

i)

$$
\begin{aligned}
E(X) &= \sum_{k=0}^{n} k \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} = \\
= \sum_{k=0}^{n} k \frac{\frac{M!}{(M-k)!k!}\binom{N-M}{n-k}}{\frac{N!}{(N-n)!n!}} &= \sum_{k=0}^{n} \frac{\frac{k \cdot M(M-1)!}{(M-k)!k(k-1)!}\binom{N-M}{n-k}}{\frac{N(N-1)!}{(N-n)!n(n-1)!}} = \\
&= n\frac{M}{N}\sum_{k=0}^{n} \frac{\binom{M-1}{k-1}\binom{N-M}{n-k}}{\binom{N-1}{n-1}}
\end{aligned}
$$

Fazendo-se $j = k - 1$ *vem*

$$
n\frac{M}{N}\sum_{j=0}^{n-1} = \frac{\binom{M-1}{j}\binom{N-1-M+1}{n-k}}{\binom{N-1}{n-1}}
$$

Admitindo-se que a soma acima é a soma das probabilidades de uma variável aleatória, com distribuição hipergeométrica, de parâmetros $M-1$, $N-1$ *e* $n-1$, *segue-se que:*

$$
\begin{aligned}
E(X) &= n\frac{M}{N} \cdot 1 = n\frac{M}{N} \\
E(X^2) &= \sum_{k=0}^{n} k^2 \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} = n\frac{M}{N}\sum_{k=1}^{n} k \frac{\binom{M-1}{k-1}\binom{N-M}{n-k}}{\binom{N-1}{n-1}} \\
&= n\frac{M}{N}\sum_{j=0}^{n-1}(j+1)\frac{\binom{M-1}{j}\binom{N-M}{n-1-j}}{\binom{N-1}{n-1}} \\
&= n\frac{M}{N}E(X+1) = n\frac{M}{N}\left[\left(\frac{(n-1)(M-1)}{(N-1)} + 1\right)\right]
\end{aligned}
$$

Consequentemente,

$$Var(X) = n\frac{M}{N}\left[\left(\frac{(n-1)(M-1)}{(N-1)}+1\right)-n\frac{M}{N}\right]$$

4.6 Distribuição de Poisson

A expressão da distribuição binomial é difícil de calcular diretamente para valores grandes de n e k. Sendo assim, há uma necessidade de substituir a expressão binomial por outra que seja melhor manejo e mais eficiente para situações como esta.

Agora, imagine a seguinte situação: Existe possibilidade para valores pequenos de p, de forma que o produto np seja relativamente pequeno para valores grandes de n. Desta forma, temos o problema de buscar o limite da função de probabilidade binomial para o caso em que p tende a zero ao passo em que n tende ao infinito; e ainda, de maneira que o produto np se mantenha igual a uma constante positiva λ. Ou seja, que fixemos a seguinte relação:

$$\lambda = np$$

Agora, podemos fazer uso destas considerações substituindo-as na expressão da distribuição binomial. Assim,

$$\begin{aligned}
\binom{n}{k}(1-p)^{n-k} &= \binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} \\
&= \frac{n(n-1)(n-2)\dots(n-k+1)}{k!\left(1-\frac{\lambda}{n}\right)^k} \\
&= \lambda^k\frac{n(n-1)(n-2)\dots(n-k+1)}{k!\left(1-\frac{\lambda}{n}\right)^k}\left(\frac{1}{n^k}\right)\left(1-\frac{\lambda}{n}\right)^n \\
&= \frac{\lambda^k}{k!}\frac{1\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\dots\left(1-\frac{(k-1)}{n}\right)}{\left(1-\frac{\lambda}{n}\right)^k}\left(1-\frac{\lambda}{n}\right)^n
\end{aligned}$$

Note que se mantivermos k fixo, resulta que para $n \longrightarrow \infty$, o numerador do segundo termo tende a 1, visto que se trata do produto de fatores, de uma forma tal que cada um deles tende a 1. Uma vez que k está fixo, decorre que o denominador deste mesmo termo também tende a 1, e o último fator tende a $e^{-\lambda}$. Sendo assim, já podemos definir a distribuição de Poisson.

Definição 4.6.1. *Diz-se que uma v.a. X tem distribuição de Poisson quando atribuímos a ela o número de ocorrências de um determinado evento, num intervalo com uma taxa fixa $\lambda > 0$, de maneira que expressão seja dada na seguinte forma:*

$$\begin{aligned} P[X=k] &= \lim_{n\longrightarrow\infty} \frac{\lambda^k}{k!}\left(1-\frac{1}{n}\right)^n = \frac{\lambda^k}{k!}\lim_{n\longrightarrow\infty}\left(1-\frac{1}{n}\right)^n \\ P[X=k] &= \frac{\lambda^k e^{-\lambda}}{k!}, k=0,1,2,\ldots \end{aligned}$$

Onde:

$K =$valor da v.a. número de ocorrências em um intervalo.

$\lambda =$taxa de ocorrência do evento.

$e = 2,71828\ldots$(constante real)

Note que $P[X=k]$ é uma legítima função de distribuição de probabilidade, pois:

$$\sum_{k=0}^{\infty}\frac{\lambda^k e^{-\lambda}}{k!} = e^(-\lambda)\left(1+\lambda+\frac{\lambda^2}{2!}+\frac{\lambda^3}{3!}+\ldots\right) = e^{-\lambda}e^{\lambda} = 1$$

Notação: $X \sim P(\lambda)$

Exemplo 4.6.1. *Se um posto de gasolina recebe em média seis carros a cada minuto, então a probabilidade de não receber carros durante esse mesmo intervalo de tempo é calculada na seguinte maneira:*

$$\begin{aligned} P[X=0] &= \frac{6^0 e^{-6}}{0!} \\ &= e^{-6}\frac{1}{e^6} \\ &= 0,002479 \end{aligned}$$

Exemplo 4.6.2. $X \sim P(\lambda)$.

i)

$$\begin{aligned} E(X) &= \sum_{k=0}^{\infty} k\frac{\lambda^k e^{-\lambda}}{k!} = \sum_{k=1}^{\infty} k\frac{\lambda^k e^{-\lambda}}{(k-1)!} \\ &= \lambda\sum_{k=1}^{\infty}\frac{\lambda^{k-1}e^{-\lambda}}{(k-1)!} \end{aligned}$$

Considerando-se que a soma acima é soma de uma variável aleatória com distribuição de Poisson com $k-1$ ocorrências segue que:

$$E(X) = \lambda$$

Analogamente,

$$\begin{aligned}
E(X^2) &= \sum_{k=0}^{\infty} k^2 \frac{\lambda^k e^{-\lambda}}{k!} = \sum_{k=0}^{\infty} [k(k-1)+k] \frac{\lambda^k e^{-\lambda}}{k!} \\
&= \sum_{k=0}^{\infty} k(k-1) \frac{\lambda^k e^{-\lambda}}{k!} + \sum_{k=0}^{\infty} k \frac{\lambda^k e^{-\lambda}}{k!} \\
&= \lambda^2 \sum_{k=2}^{\infty} k(k-1) \frac{\lambda^{k-2} e^{-\lambda}}{k(k-1)(k-2)!} E(X) \\
&= \lambda^2 \sum_{k=2}^{\infty} k(k-1) \frac{\lambda^{k-2} e^{-\lambda}}{(k-2)!} E(X) + \lambda = \lambda^2 + \lambda
\end{aligned}$$

Daí, segue que

$$Var(X) = E(X) = \lambda$$

ii)

$$M_X(t) = \sum_{k=0}^{\infty} e^{tk} \frac{\lambda^k e^{-\lambda}}{k!} = \sum_{k=0}^{\infty} \frac{(\lambda e^t)^k e^{-\lambda}}{k!} = e^{\lambda(e^t - 1)}$$

Para o cálculo do primeiro momento, vamos fazer a diferenciação:

Na equação acima, considere $u = e^t - 1$ o que resulta na seguinte situação em $M_X(t) = e^{\lambda u}$. Daí,

$$\begin{cases}
\frac{dM_X(t)}{du} = \frac{dM_X(t)}{dv} \frac{dv}{du} \\
\frac{du}{dt} = e^t \\
v = \lambda u
\end{cases}$$

$$\begin{aligned}
\frac{dM_X(t)}{dt} &= \frac{dM_X(t)}{du} \frac{du}{dt} \\
&= \lambda e^v e^t = \lambda e^t e^{\lambda u}
\end{aligned}$$

Assim,

$$= \lambda e^t e^{\lambda[e^t-1]} \Rightarrow \frac{dM_X(0)}{dt} = \lambda = E(X)$$

Analogamente, podemos encontrar o segundo momento. Confira!

$$\begin{aligned}
\frac{d^2M_X(t)}{dt} &= (\lambda e^t)' e^{\lambda[e^t-1]} + (\lambda e^t) M'_X(t) \\
&= \lambda e^t e^{\lambda[e^t-1]} + (\lambda e^t)\lambda e^t e^{\lambda[e^t-1]} \\
&= \lambda e^t e^{\lambda[e^t-1]} + (\lambda e^t)^2 e^{\lambda[e^t-1]} \Rightarrow \frac{d^2M_X(0)}{dt} = \lambda + \lambda^2 = E(X^2)
\end{aligned}$$

Inserindo estes resultados na expressão da variância, segue que:

$$Var(X) = \lambda + \lambda^2 - \lambda^2 = \lambda$$

4.7 Distribuição Multinomial ou Polinomial

Definição 4.7.1. *Considere um experimento ε, seu espaço amostral Ω, e uma partição em k eventos mutuamente exclusivos C_1, C_2, $\cdots$, C_k, com probabilidades p_1, p_2, $\cdots$, p_k, respectivamente. Considere ainda n ensaios deste experimento, de forma que os p_i, $i = 1, 2, \cdots, k$, permanecem inalterados durante as repetições, com $\sum_{i=1}^{k} p_i = 1$.*
Se X_1, X_2, $\cdots$, X_k forem os números de ocorrências de C_1, C_2,...,C_k, respectivamente, com $\sum_{i=1}^{k} X_i = n$, então podemos definir a distribuição multinomial ou polinomial na seguinte forma:

$$P[X_1 = n_1, X_2 = n_2, \ldots, X_k = n_k] = \frac{n!}{n_1!n_2!\ldots n_k!} p_1^{n_1} p_2^{n_2} \ldots p_k^{n_k}$$

onde, $\sum_{i=1}^{k} n_i = n$.

Coeficiente multinomial: Se temos n objetos, não necessariamente distintos. Suponha n_1 de um tipo, n_2 de outro e assim sucessivamente até n_k objetos do tipo k, onde $n_1 + n_2 + \ldots + n_k = n$. Assim todos os n objetos podem ordenar-se por detrás do outro de tantas formas distintas, como indica o que chamamos de coeficiente multinomial:

$$\binom{n}{n_1 n_2 \ldots n_{k-1} n_k} = \frac{n!}{n_1! n_2! \ldots n_k!}$$

O motivo da existência desse coeficiente é muito simples, pois se considerarmos que os n objetos são todos distintos, então claramente as distintas formas em que podemos escrevê-los é $n!$. [22]

Observação 4.7.1. *quando $k = 2$, a distribuição se reduz à distribuição binomial, pois:*

$$P[X_1 = n_1, X_2 = n_2] = \frac{n!}{n_1! n_2!} p_1^{n_1} p_2^{n2}, \qquad com, \begin{cases} n_2 = n - n_1 \\ \\ p_2 = 1 - p_1 \end{cases}$$

Notação: $X \sim Po\ (n, p)$ denota a distribuição polinomial com parâmetros n e p.

Exemplo 4.7.1. *Sabe-se que o sangue humano é classificado em quatro tipos. Numa certa população, as distribuições de probabilidades destes tipos foram dadas de acordo com a tabela a seguir:*

A	*B*	*AB*	*O*
0,4	*0,45*	*0,10*	*0,05*

Tabelas 4.3: Distribuições de probabilidade para tipagem sanguínea

Qual a probabilidade de que em dez indivíduos, escolhidos ao acaso, existam: dois do tipo A, seis do tipo B, e um de cada dos outros tipos?

Considere o evento
S_i= sair o número i, $i = 1, 2, 3, 4$ e
$X_1 = 2$, X_2= 6, X_3 =1, $X_4 = 1$.
Observe ainda que, $p_1 = 0,4,\ p_2 = 0,45,\ p_3 = 0,10,\ p_4 = 0,05$ e $\sum_{i=1}^{4} X_i = 10$. Portanto, da definição da distribuição polinomial, temos que:

$$P[X_1 = 2, X_2 = 6, X_3 = 1, X_4 = 1] \quad = \quad \frac{10!}{2!6!1!1!} (0,4)^2 (0,45)^6 (0,10)^1 (0,05)^1$$

4.8 Distribuição Uniforme

Definição 4.8.1. *Diz-se que uma v.a.c. X tem distribuição uniforme em um intervalo* $[a, b]$*, se sua fdp é expressa na forma:*

$$f(x) = \begin{cases} \dfrac{1}{b-a}, \text{se para a} \leq \text{x} \leq \text{b} \\ \\ 0, \text{no complementar} \end{cases}$$

A única restrição para os valores de a e b está no fato de a ser menor que b.

Notação: $X \sim U[a, b]$

Assim, o modelo nos faz entender que os valores possíveis da v.a., no intervalo dado, possuem a mesma probabilidade de ocorrência.

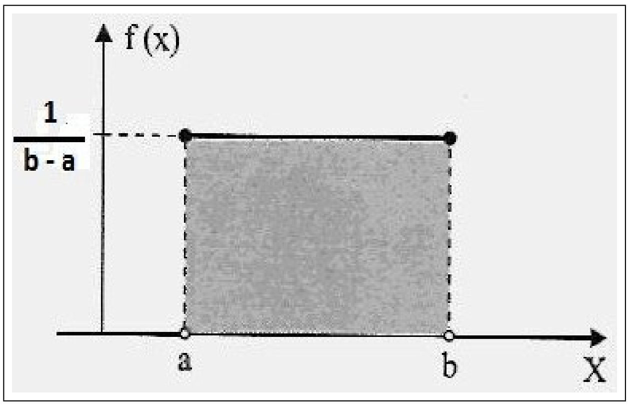

Figura 4.4 *Esboço sistemático de funções de variáveis aleatórias*

Propriedade 4.8.1.

i) f é uma fdp

Demonstração 4.8.1.

$$\begin{aligned} I &= \int_{-\infty}^{\infty} f(x)dx = \int_a^b \frac{1}{b-a}dx \\ &= \frac{1}{b-a}\int_a^b dx \\ &= \frac{1}{b-a}x \Big|_a^b \\ &= \frac{1}{b-a}.(b-a) = 1 \end{aligned}$$

ii)

$$f(x) = \begin{cases} 0, \text{se } x < a \\ \\ \dfrac{(x-a)}{(b-a)}, \text{se } a \leq x < b \\ \\ 1, \text{ se } x \geq b \end{cases}$$

Demonstração 4.8.2. *Da definição, temos que*

$$\begin{aligned} F(x) &= \frac{1}{(b-a)} \int_{-\infty}^{x} dx = \frac{1}{(b-a)} \int_{a}^{x} dx \\ = \frac{1}{(b-a)}\, x|_a^x &= \frac{(x-a)}{b-a} \end{aligned}$$

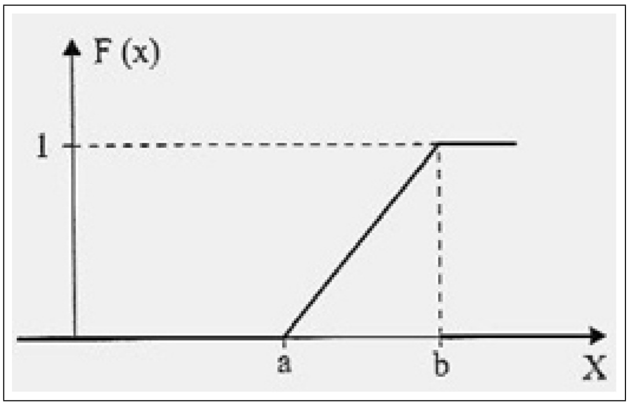

Figura 4.5 *Gráfico da função repartição uniforme*

Exemplo 4.8.1 (INTERNET)**.** *: Grande parte das linguagens de programação, pacotes estatísticos ou de planilhas de cálculo, possui capacidade de gerar números pseudo-aleatórios, visto que é possível repetir a mesma sequência a partir de um mesmo valor inteiro. O gerador desses números os produz partindo de um mesmo valor inteiro. Desta forma, se tivermos uma função que gera números entre 0 e 5, podemos perfeitamente calcular a probabilidade de um número gerado estar entre qualquer intervalo. Para* $(1 \leq x \leq 2,5)$, *temos:*

$$\begin{aligned} P(1 \leq x \leq 2,5) &= \frac{1}{5} \int_{1}^{2,5} dx \\ \frac{1}{5}\, x|_1^{2,5} &= \frac{1}{5}(2,5-1) \\ &= 0,3 \end{aligned}$$

Exemplo 4.8.2. $X \sim U(a,b)$

i)

$$
\begin{aligned}
E(X) &= \int_a^b x\frac{1}{b-a}dx = \frac{1}{b-a}\int_a^b xdx = \frac{1}{b-a}\left[\frac{x^2}{2}\right]\Bigg|_a^b \\
&= \frac{b^2-a^2}{2(b-a)} = \frac{(b+a)}{2} \\
E(X^2) &= \frac{1}{(b-a)}\int_a^b x^2dx = \frac{1}{(b-a)}\left[\frac{x^3}{3}\right]\Bigg|_a^b = \frac{a^2+ab+b^2}{3}
\end{aligned}
$$

Desta forma,

$$
Var(X) = \frac{a^2+ab+b^2}{3} - \left[\frac{(b+a)}{2}\right]^2 = \frac{(b-a)^2}{12}
$$

ii)

$$
M_X(t) = \frac{1}{(b-a)}\int_a^b e^{tx}dx = \frac{1}{(b-a)}\left(\frac{1}{t}e^{tx}\right)\Bigg|_a^b = \frac{t}{t(b-a)}(e^{tb}-e^{ta})
$$

4.9 Distribuição Exponencial

Trata-se de uma das mais importantes, no que tange a descrição de uma imensa classe de fenômenos. Isso, pelo fato de ser exponencialmente distribuída. Com isso sua utilidade é eficiente nos estudos de tempos a desintegração de partículas radioativas, tempo de vida, tempo até à falha de um determinado equipamento (teoria da confiabilidade), etc.

Definição 4.9.1. *Uma variável aleatória X, tem distribuição exponencial, com parâmetro $\lambda > 0$ se sua fdp é dada por:*

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

Propriedade 4.9.1.

i)

$$
\int_{-\infty}^{\infty} \lambda e^{\lambda x} = 1
$$

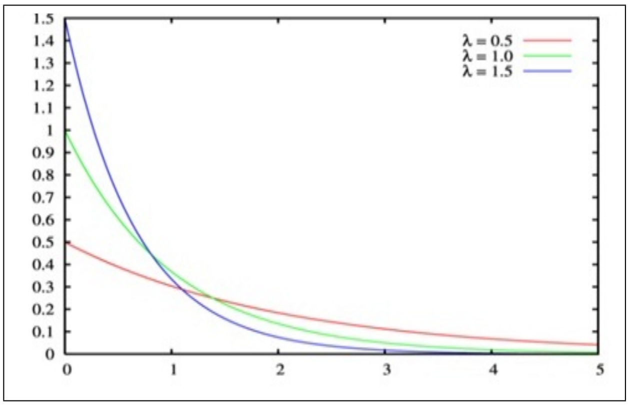

Figura 4.6 *Representação gráfica da distribuição exponencial. Disponível em:* <*http://upload.wikimedia.org/wikipedia/commons/b/b1/Exponential_distribution_pdf.png*>

Demonstração 4.9.1.

Considere que λ é uma constante. Daí, temos:

$$\int_0^{\infty} \lambda e^{\lambda x} dx = (e^{-\lambda x})|_0^{\infty} = -\frac{1}{e^{\lambda x}}|_0^{\infty} = \left[-\frac{1}{e^{\infty}} - \left(-\frac{1}{e^0}\right)\right] = 1$$

ii)

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

Demonstração 4.9.2.

$$F(x) = P[X \leq x] = \int_0^x \lambda e^{-\lambda x} dx = 1 - e^{-\lambda})$$

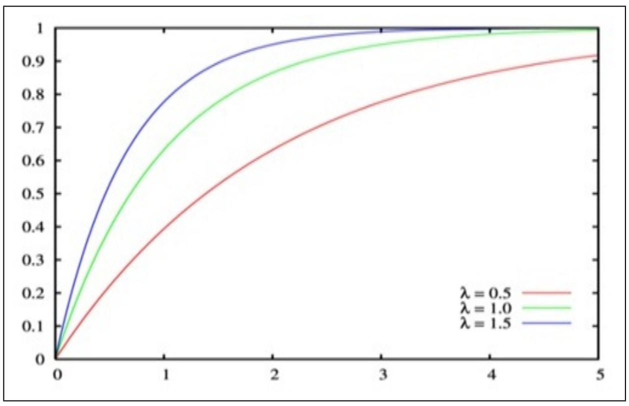

Figura 4.7 *Representação gráfica da distribuição disponivel em: exponencial*<*http://upload.wikimedia.org/wikipedia/commons/7/77/Exponential_distribution _cdf.png*>

iii) Considere para quaisquer $a, b > 0$, então $P[X > a + b|X > a] = P[X > b]$

Demonstração 4.9.3.

$$\begin{aligned} P[X > a + b|X > a] &= \frac{P[X > a + b]}{P[X > a]} \\ &= \frac{e^{-\lambda(a+b)}}{e^{-\lambda a}} = \frac{e^{-\lambda a}.e^{-\lambda b}}{e^{-\lambda a}} = e^{-\lambda b} = P[X > b] \end{aligned}$$

Exemplo 4.9.1. *Sabe-se que o tempo de vida de uma determinada peça mecânica é em média* $\lambda = 4200^1$*horas. Considere que essa situação pode ser representada por uma variável aleatória* X*, com distribuição exponencial, de parâmetro. Qual é a probabilidade de que* X*ultrapasse essa média?*

$$\begin{aligned} P[X > 4200] &= \frac{-1}{e^{4200}}(4200) \\ &= e^{-1} < \frac{1}{2} \end{aligned}$$

Exemplo 4.9.2. $X \sim \in (\lambda)$

i)

$$E(X) = \int_0^\infty x\lambda e^{\lambda - x}dx$$

Vamos integrar por partes!

Considere:

$$\begin{cases} u = \lambda x \\ \\ dv = e^{-\lambda x} \end{cases} \Longrightarrow \begin{cases} du = \lambda dx \\ \\ v = \int e^{-\lambda x}dx = \dfrac{-1}{\lambda}e^{-\lambda x} \end{cases}$$

$$\begin{aligned} E(X) &= -\lambda x\frac{1}{\lambda}e^{-\lambda x}\Big|_0^\infty + \int_0^\infty \frac{1}{\lambda}e^{-\lambda x}\lambda dx \\ \frac{-x}{e^{\lambda x}}\Big|_0^\infty + \int_0^\infty e^{-\lambda x}dx &= 0 - \frac{e^{-\lambda x}}{\lambda}\Big|_0^\infty = \frac{1}{\lambda} \end{aligned}$$

Semelhantemente,

$$E(X^2) \;=\; \int_0^\infty x^2\lambda e^{-\lambda x}dx$$

$$\begin{cases} u = \lambda x^2 \\ \\ dv = e^{-\lambda x} \end{cases} \Longrightarrow \begin{cases} du = \lambda 2x dx \\ \\ v = \int e^{-\lambda x}dx = \dfrac{-1}{\lambda}e^{-\lambda x} \end{cases}$$

Sendo assim,

$$\begin{aligned} E(X^2) &= \left.\frac{\lambda x^2 e^{-\lambda x}}{\lambda}\right|_0^\infty + \int_0^\infty \frac{1}{\lambda}e^{-\lambda x}\lambda 2x dx \\ &= \left.\frac{x^2}{e^{\lambda x}}\right|_0^\infty + \frac{2}{\lambda}\int_0^\infty e^{-\lambda x}\lambda x dx \\ &= 0 + \frac{2}{\lambda}E(X) = \frac{2}{\lambda}\cdot\frac{1}{\lambda} = \frac{2}{\lambda^2} \\ Var(X) &= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2} \end{aligned}$$

ii)

$$\begin{aligned} M_X(t) &= \int_0^\infty e^{tx}\lambda e^{-\lambda x}dx = \int_0^\infty \lambda e^{-x(\lambda - t)}dx = \frac{\lambda}{(\lambda - t)} \\ M'_X(t) &= \frac{\lambda}{(\lambda - t)^2} \Longrightarrow M'_X(0) = \frac{1}{\lambda} = E(X) \\ M''_X(t) &= \frac{2\lambda}{(\lambda - t)^3} \Longrightarrow M''_X(0) = \frac{2}{\lambda^2} = E(X^2) \end{aligned}$$

Assim,

$$Var(X) \;=\; \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

4.10 Distribuição Normal

Definição 4.10.1. *Uma v.a. X, que assuma todos os valores reais* $-\infty < x < \infty$ *e que seja equipada dos parâmetros* μ *e* σ^2*, tem distribuição normal se sua fdp é da forma:*

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

Notação: $X \sim N(\mu, \sigma^2)$ denota a distribuição normal com parâmetros μ e σ.

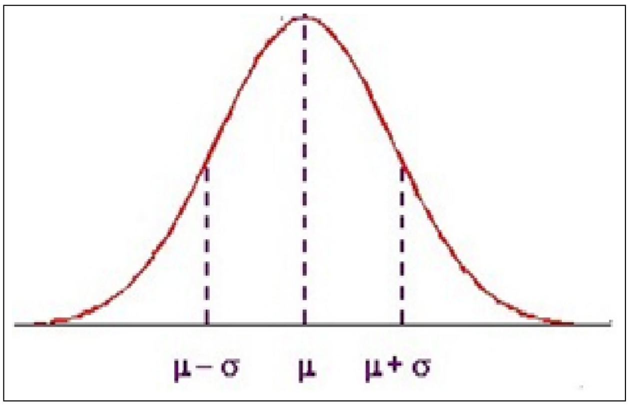

Figura 4.8 *Comportamento gráfico da densidade normal*

Características:

a) f(x) é simétrica em relação a μ, visto que depende somente da expressão $(x - \mu)^2$.

b) o valor máximo de f(x) ocorre para $x = \mu$, onde média, mediana e moda da distribuição coincidem.

c) $f(x) \longrightarrow 0$ quando $x \longrightarrow \pm\infty$

d) é assintótica em relação ao eixo das abscissas.

Propriedade 4.10.1. *f é uma fdp, isto é, $f(x) \leq 0$ e $\int_{-\infty}^{\infty} f(x)dx = 1$.*

Prova

Façamos $z = \dfrac{(x-\mu)}{\sigma}$ em $\int_{-\infty}^{\infty} f(x)dx$. Assim, obtemos a integral $I = \dfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-z^2}{2}} dz$.
O mecanismo aplicado para calcular I é considerar I^2 no lugar de I. Desta forma:

$$
\begin{aligned}
I^2 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{\frac{-z^2}{2}} dz \int_{-\infty}^{\infty} e^{\frac{-t^2}{2}} dt \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{\frac{-(t^2+z^2)}{2}} dtdz
\end{aligned}
$$

Para o cálculo da integral dupla, considere $t = r\cos(\theta)$ e $z = r\sin(\theta)$.
Daí, $-\infty < t < \infty$, $-\infty < z < \infty$, $0 \leq \theta \leq 2\pi$, $0 \leq r \leq \infty$ e $dtdz = rdrd\theta$. Logo

$$\begin{aligned} I^2 &= \frac{1}{2\pi}\int_0^{2\pi}\int_0^{\infty} re^{\frac{-r^2}{2}}drd\theta \\ &= \frac{1}{2\pi}\int_0^{2\pi} -e^{\frac{-r^2}{2}}\Big|_0^{\infty} d\theta \\ &= \frac{1}{2\pi}\int_0^{2\pi} d\theta = \frac{1}{2\pi}\theta\Big|_0^{2\pi} = \frac{2\pi}{2\pi} = 1 \end{aligned}$$

Propriedade 4.10.2.

$$F(x) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{\frac{-t^2}{2}} dt$$

Prova

Considere novamente $z = \dfrac{((x-\mu))}{\sigma}$, daí

$$\begin{aligned} F(z) = P[Z \le z] &= P\left[\frac{(x-\mu)}{\sigma} \le z\right] P\left[X \le az + \mu\right] \\ &= \int_{-\infty}^{az+\mu} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \end{aligned}$$

Fazendo-se outra vez o processo de substituição, só que agora $t = \dfrac{(x-\mu)}{\sigma}$ daí segue que:

$$F(x) = \int_{-\infty}^{z} \frac{1}{\sigma\sqrt{2\pi}} \sigma e^{-\frac{t^2}{2}} dt = \int_{-\infty}^{z} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

4.10.1 Tabulação da Normal

Definição 4.10.2. *Suponha-se que X tenha a distribuição $N \sim (\mu, \sigma^2)$ tal que $\mu = 0$ e $\sigma^2 = 1$. Dizemos então que X tem distribuição normal reduzida. Isto é, a fdp de X é:*

$$\Phi(X) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$$

Desta forma, $P[a \le x \le b] = \frac{1}{\sqrt{2\pi}}\int_a^b e^{\frac{-x^2}{2}} dx$. Porém, esta integral não pode ser calculada pelas ferramentas conhecidas, por nós, até aqui. Isto é, não podemos aplicar o teorema fundamental do cálculo, visto que não existe uma função cuja derivada seja igual a $e^{\frac{-x^2}{2}}$.

Para isso, existem métodos de integração numérica que permitem o cálculo de integrais desse tipo pelo fato de $P[Z \leq z]$ ter sido tabelada.
A fdp da normal reduzida é determinada a partir de $Z = \dfrac{(x-\mu)}{\sigma}$, donde segue que:

$$\begin{aligned}\Phi(z) = P[Z \leq z] &= P\left[\frac{(x-\mu)}{\sigma} \leq z\right] = P[X \leq \sigma z + \mu] \\ &= \int_{-\infty}^{\sigma z+\mu} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{\frac{-x^2}{2}} dx\end{aligned}$$

em que usamos $x = \dfrac{(x-\mu)}{2}$.

Nota: É evidente que $\Phi(-x) = 1 - \Phi(x)$

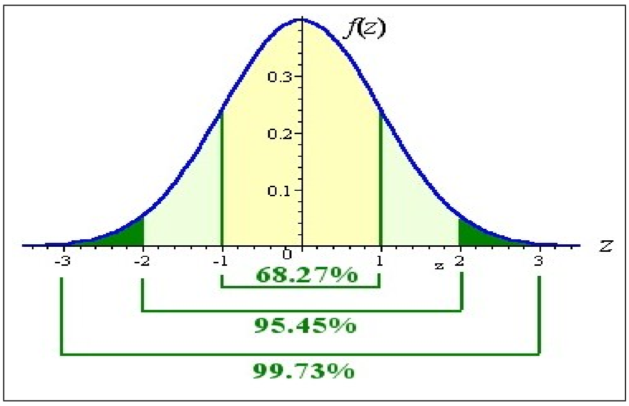

Figura 4.9 *Densidade da Normal padronizada*

Deste modo, considere $N \sim (\mu, \sigma^2)$ em que desejamos obter a probabilidade $P[a < x < b]$.
Para isso, temos:

$$\begin{aligned}P[a < x < b] = &= P\left[\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right] \\ &= \Phi\left[\frac{b-\mu}{\sigma}\right] - \Phi\left[\frac{a-\mu}{\sigma}\right]\end{aligned}$$

cujo valor pode ser obtido com o uso da tabela, isto é, desde que as constantes sejam conhecidas.

Exemplo 4.10.1.1. *Um determinado laboratório supôs que o tempo de coagulação sanguínea em seres humano é variável aleatória que admite distribuição normal de média 8 minutos e desvio padrão* 2 *minutos. Se este mesmo laboratório admitiu um tempo de coagulação de*

SEGUNDA PARTE DECIMAL

PARTE INTEIRA E 1ª PARTE DECIMAL	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1770	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621

Figura 4.10 $P[Z \leq 0,32] = 0,1255$

pelo menos 6 *minutos, então a probabilidade de que haja exceção dessa admissão é dada por:*

$$\begin{aligned} P[X < 6] &= P\left[z < \frac{6-8}{2}\right] = P[Z < -1] \\ &= 0,5 - P[0 < Z < 1] = 0,5 - 0,34134 = 0,1587 \end{aligned}$$

Exemplo 4.10.1.2. $X \sim N(\mu, \sigma)$

i)

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

Fazendo-se novamente a transformação,

$$z = \frac{(x-\mu)}{\sigma}$$

Segue-se que

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

Na primeira parcela, a integral é nula, visto que o integrando comporta a seguinte propriedade: $f(z) = -f(-z)$, o que se refere a uma função ímpar. Já a segunda, representa a área total (desconsiderando μ) sobre a fdp normal. Consequentemente,

$$E(X) = 0 + \mu.1 = \mu$$

Usando os mesmos mecanismos acima, podemos encontrar o segundo momento,

$$\begin{aligned}
E(X^2) &= \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx \\
&= \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} (\sigma z + \mu)^2 e^{\frac{-z^2}{2}} \sigma dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sigma^2 z^2 e^{\frac{-z^2}{2}} dz + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} 2\sigma z \mu e^{\frac{-z^2}{2}} dz \\
&\quad + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mu^2 e^{\frac{-z^2}{2}} dz \\
&= \frac{1}{\sqrt{2\pi}} \sigma^2 \int_{-\infty}^{\infty} z^2 e^{\frac{-z^2}{2}} + \frac{1}{\sqrt{2\pi}} 2\sigma\mu \int_{-\infty}^{\infty} z e^{\frac{-z^2}{2}} dz \\
&\quad + \frac{1}{\sqrt{2\pi}} \mu^2 \int_{-\infty}^{\infty} e^{\frac{-z^2}{2}} dz
\end{aligned}$$

Para a segunda e terceira parcelas, o argumento é o mesmo utilizado acima. E para a primeira, vamos fazer o processo de integração por partes:

Considere:

$$\begin{cases} u = z \\ \\ dv = z e^{\frac{-z^2}{2}} \end{cases} \Longrightarrow \begin{cases} du = dz \\ \\ v = \int z e^{\frac{-z^2}{2}} = -e^{\left(\frac{-z^2}{2}\right)} \end{cases}$$

Daí, podemos resolvê-la na seguinte maneira:

$$\begin{aligned}
\frac{1}{\sqrt{2\pi}} \sigma^2 I &= \frac{1}{\sqrt{2\pi}} \sigma^2 \left[uv|_0^{\infty} - \int_{-\infty}^{\infty} v du \right] \\
&= \frac{1}{\sqrt{2\pi}} \sigma^2 \left[z \left(-e^{\left(\frac{-z^2}{2}\right)} \right) \Big|_0^{\infty} + \int_{-\infty}^{\infty} e^{\left(\frac{-z^2}{2}\right)} dz \right] \\
&= \frac{z e^{\left(\frac{-z^2}{2}\right)}}{\sqrt{2\pi}} dz|_0^{\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\left(\frac{-z^2}{2}\right)} dz \\
&= 0 + 1 = 1
\end{aligned}$$

Ou seja,

$$E(X^2) = \sigma^2 + \mu^2$$

Sendo assim,

$$Var(X) = \sigma^2 + \mu^2 - \mu^2 = \sigma^2$$

ii)

$$\begin{aligned}
M_X(t) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} e^{tx} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\left(\frac{x^2}{2}\right)+tx} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\left(\frac{x^2}{2}\right)+tx-\frac{t^2}{2}\frac{t^2}{2}} \\
&= e^{t^2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{(x-t)^2}{2}} dx = e^{\frac{t^2}{2}}
\end{aligned}$$

Fazendo $x = \sigma y + \mu$, vem:

$$M_X(t) = e^{t\mu} M(\alpha t) = e^{t\mu} e^{\frac{t^2\alpha^2}{2}} = e^{t\mu \frac{t^2\alpha^2}{2}}$$

Seja $u = t\mu + \dfrac{t^2\alpha^2}{2}$

$$\begin{aligned}
\frac{dM_X}{dt} &= \frac{dM_X}{du} \times \frac{du}{dt} = e^u(\mu + t\alpha^2) e^{t\mu + \frac{t^2\alpha^2}{2}} (\mu + t\alpha^2) \\
\frac{d^2M_X}{dt} &= \left(e^{t\mu+\frac{t^2\alpha^2}{2}}(\mu + t\alpha^2)\right) \mu + t\alpha^2 e^{t\mu+\frac{t^2\alpha^2}{2}} = \mu^2 + \alpha^2
\end{aligned}$$

Pondo $t = 0$ nas expressões acima, segue que:

$$E(X) = \mu, E(X^2) = \mu^2 + \alpha^2$$

Consequentemente,

$$Var(X) = \alpha^2$$

4.11 Distribuição Gama

Definição 4.11.1. *Se X é uma v.a. contínua, que tome todos os valores não negativos, e ainda, os parâmetros r (de forma) e λ (de escala), tais que $r > 0$, $\lambda > 0$. Diremos que X tem distribuição gama se sua fdp é dada por:*

$$f(x) = \begin{cases} \dfrac{\lambda}{\Gamma(r)} (\lambda x)^{r-1} e^{-\lambda x}, & se\ x > 0 \\ \\ 0, & no\ complementar \end{cases}$$

em que $\Gamma(r)$ é a função gama, definida na seguinte forma:

$$\Gamma(r) = \int_0^{\infty} x^{r-1} e^{-x} dx, se\ r > 0$$

Propriedade 4.11.1. *A função gama é uma extensão da função fatorial, aos reais, por intermédio de uma reparametrização.*

Prova

Vamos utilizar a integração por partes. Considere:

$$\begin{cases} u = x^{r-1} \\ \\ dv = e^{-x} \end{cases} \Longrightarrow \begin{cases} du = (r-1)x^{(r-2)}dx \\ \\ v = \int e^{-x} = -e^{-x} \end{cases}$$

Assim,

$$\begin{aligned} I &= uv|_0^{\infty} - \int_0^{\infty} vdu \\ &= -x^{r-1}e^{-x}\big|_0^{\infty} + \int_0^{\infty} e^{-x}(r-1)x^{(r-2)} \\ &= \left.\frac{x^{r-1}}{x^2}\right|_0^{\infty} + (r-1)\int_0^{\infty} e^{-x}x^{(r-2)}dx \\ &= (r-1)\Gamma(r-1) \end{aligned}$$

ou seja, a integral imprópria existe (converge para $r > 0$). ■

Nota: Observe que a função gama obedece a uma relação de ocorrência: Considere $r = n$, daí,

$$\begin{aligned} \Gamma(n) &= (n-1)\Gamma(n-1) \\ &= (n-1)(n-2)\Gamma(n-2) = \Gamma = (n-1)(n-2)\ldots(1) \end{aligned}$$

Mas,

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

Logo, $\Gamma(n) = (n-1)!$ quando n for um inteiro positivo. Por tanto, a função gama é uma extensão da função fatorial aos complexos, por meio de uma reparametrização. Veja também que a f é uma autêntica função densidade de probabilidades:

$$\int_0^{\infty} \frac{\lambda}{\Gamma(r)}(\lambda x)^{r-1}e^{-\lambda x}dx = \frac{\lambda^r}{\Gamma(r)}\int_0^{\infty} x^{r-1}e^{-\lambda x}dx$$

Considere $y = \lambda x$, daí

$$= \frac{\lambda^r}{\Gamma(r)}\int_0^{\infty} \frac{y^{r-1}e^{-y}}{\lambda^r}dy = \frac{1}{\Gamma(r)}\int_0^{\infty} y^{r-1}e^{-y}dy$$

observe que o integrando da segunda igualdade corresponde a $\Gamma(r)$, daí é imediato verificar que o desenvolvimento da expressão acima equivale a 1 , pois:

$$\frac{1}{\Gamma(r)}\Gamma(r) = 1$$

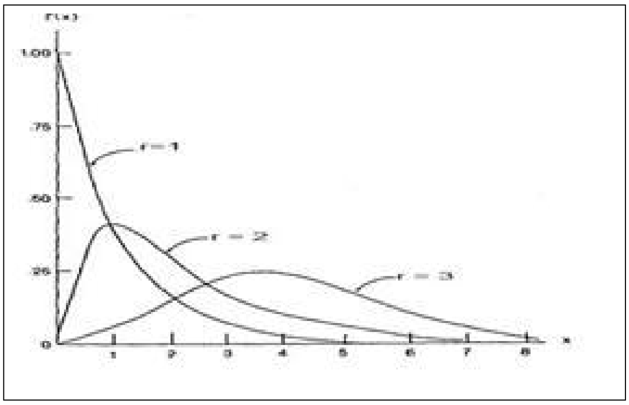

Figura 4.11 *Representação da distribuição Gama*

Propriedade 4.11.2. *A fda da distribuição Gama, fica definida na seguinte forma:*

$$F(x) = P[X \leq x] = \int_0^x \frac{\lambda}{\Gamma(r)}(\lambda x)^{r-1}e^{-\lambda x}dx = 1 - \sum_{k=0}^{r-1}\frac{e^{-\lambda x}(\lambda x)^k}{k!}$$

Prova.
Considere $r \in Z_+^*$, $a > 0$ e $I = \int_a^\infty \dfrac{e^{-y}y^r}{r!}dy$, daí, podemos fazer a seguinte relação:

$$Ir! = \int_a^\infty e^{-y}y^r dy$$

Para fazer a integração por partes, considere:

$$\begin{aligned} u &= y^r \Rightarrow = du = ry^{r-1}dy \\ dv &= e^{-y} \Rightarrow v = \int e^{-y} = -e^{-y} \\ Ir! = &= uv - \int_a^\infty vdu \\ &= -y^r e^{-y}\Big|_a^\infty e^{-y}ry^{r-1}dy \\ &= a^r e^{-a} + r\int_a^\infty e^{-y}y^{r-1}dy \end{aligned}$$

Veja que a integral da expressão é obtida da mesma forma que a integral modelo, só que com $r = (r-1)$. Daí, se continuarmos a integrar por partes teremos:

$$Ir! = a^r e^{-a} + ra^{r-1}e^{-a} + r(r-1)a^{r-2} + r\int_a^\infty e^{-y}(r-1)y^{r-2}dy$$

e isso sucede até que obtenhamos:

$$Ir! = e^{-a}[a^r + ra^{r-1} + r(r-1)a^{r-2} + \ldots + r!]$$

Logo,

$$I = e^{-a}\left[1 + a + \frac{a^2}{2!} + \ldots + \frac{a^r}{r!}\right] = \sum_{k=0}^{r} P[Y=k] = \frac{\lambda^k e^{-\lambda}}{k!}$$

em que Y tem distribuição de Poisson, com parâmetro λ.
Agora, considere r $\in Z_{+}^{*}$, tal que $\Gamma(r) = (r-1)!$. Substituindo esta função na expressão da fdp, vem:

$$f(x) = \frac{\lambda}{(r-1)!}(\lambda x)^{r-1}e^{-\lambda x}, \text{se } r > 0.$$

Consequentemente, a função repartição se torna assim:

$$F(x) = 1 - P[X > x] = 1 - \int_x^\infty \frac{\lambda}{(r-1)!}(\lambda t)^{r-1}e^{-\lambda t}dt$$

Considere $u = \lambda t$, daí

$$F(x) = 1 - \int_{\lambda x}^\infty \frac{u^{r-1}e^{-u}}{(r-1)!}du$$

que é da mesma forma que I, só que com $a = \lambda x$. Por tanto,

$$F(x) = 1 - \sum_{k=0}^{r-1} \frac{e^{-\lambda x}(\lambda x)^k}{k!}$$

desta forma, a função repartição da distribuição gama pode ser expressa em termos da função repartição da distribuição de Poisson.

Exemplo 4.11.1. $X \sim \mathit{Gama}(\alpha, \beta)$

i)

$$\begin{aligned} E(X) &= \int_0^\infty x\frac{\lambda}{\Gamma(r)}(\lambda x)^{r-1}e^{-\lambda x}dx \\ &= \frac{\lambda^r}{\Gamma(r)}\int_0^\infty x^r e^{-\lambda x}dx \end{aligned}$$

Vamos integrar por partes. Para isso, considere:

$$\begin{cases} u = x^r \\ dv = e^{-\lambda x} \end{cases} \Longrightarrow \begin{cases} du = rx^{r-1} \\ v = \int e^{-\lambda x} = -\frac{1}{\lambda}e^{-\lambda x} \end{cases}$$

Daí,

$$\begin{aligned} E(X) &= \frac{\lambda r}{\Gamma(r)}\left[-x^r\frac{1}{\lambda}e^{-\lambda x}\Big|_0^\infty + \frac{r}{\lambda}\int_0^\infty e^{-\lambda x} r x^{r-1}dx\right] \\ &= \frac{\lambda r}{\Gamma(r)}\left[0 + \frac{r}{\lambda}\int_0^\infty e^{-\lambda x}x^{r-1}dx\right] \end{aligned}$$

Considere $y = \lambda x$

$$\begin{aligned} &= \frac{\lambda^r}{\Gamma(r)}\frac{r}{\lambda}\int_0^\infty \frac{e^{-y}y^{r-1}}{\lambda^r}dy = \frac{r}{\lambda\Gamma(r)}\int_0^\infty e^{-y}y^{r-1}dy \\ &= \frac{r}{\lambda\Gamma(r)}\Gamma(r) = \frac{r}{\lambda}. \end{aligned}$$

Analogamente,

$$\begin{aligned} E(X^2) &= \int_0^\infty x^2\frac{\lambda}{\Gamma(r)}(\lambda x)^{r-1}e^{-\lambda x}dx \\ &= \frac{\lambda^2}{\Gamma(r)}\int_0^\infty x^{r+1}e^{-\lambda x}dx \end{aligned}$$

Considere:

$$\begin{cases} u = x^{r+1} \\ dv = e^{-\lambda x} \end{cases} \Longrightarrow \begin{cases} du = (r+1)x^r \\ v = -\dfrac{e^{-\lambda x}}{\lambda} \end{cases}$$

Assim,

$$\begin{aligned} E(X^2) &= \frac{\lambda^r}{\Gamma(r)}\left[-x^{r+1}\frac{e^{\lambda x}}{\lambda}\Big|_0^\infty + \int_0^\infty \frac{e^{\lambda x}}{\lambda}(r+1)x^r dx\right] \\ &= \frac{\lambda^r}{\Gamma(r)}\left[\frac{(r+1)}{\lambda}\int_0^\infty e^{-\lambda x}x^r dx\right] \end{aligned}$$

Integrando novamente por partes, segue que:

$$\begin{aligned} E(X^2) &= \frac{\lambda^r}{\Gamma(r)}\left[\frac{(r+1)}{\lambda}\left[\frac{1}{\lambda}\int_0^\infty e^{-\lambda x} r x^{r-1} dx\right]\right] \\ &= \frac{\lambda^r}{\Gamma(r)}\left[\frac{(r+1)}{\lambda} . \frac{r}{\lambda}\int_0^\infty e^{-\lambda x} x^{r-1} dx\right] \end{aligned}$$

Fazendo-se $y = \lambda x$

$$\begin{aligned} \frac{\lambda^r}{\Gamma(r)}\left[\frac{r^2+r}{\lambda^2} \cdot \int_0^\infty \frac{e^{-y} y^{r-1}}{\lambda^r} dy\right] &= \frac{\lambda^r}{\gamma(r)}\left[\frac{r^2+r}{\lambda^2}\frac{\gamma(r)}{\lambda^r}\right] \\ &= \frac{r^2+r}{\lambda^2} \end{aligned}$$

Desta forma,

$$Var(X) = \frac{r}{\lambda^2}$$

ii)

$$\begin{aligned} M_X(t) &= \int_0^\infty e^{tx}\frac{\lambda^r}{\Gamma(r)}(\lambda x)^{r-1} e^{-\lambda x} dx \\ &= \frac{\lambda^r}{\Gamma(r)}\int_0^\infty e^{-x(\lambda - t)} x^{r-1} dx \end{aligned}$$

Considere $y = x(\lambda - t)$ daí

$$\begin{aligned} &= \frac{\lambda^r}{\Gamma(r)}\int_0^\infty \frac{e^{-y} y^{r-1}}{(\lambda - t)(\lambda - t)^{r-1}} dy \\ &= \left(\frac{\lambda}{\lambda - t}\right)^r \frac{\lambda^r}{\Gamma(r)}\int_0^\infty e^{-y} y^{r-1} dy \end{aligned}$$

$$M_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^r$$

Ou seja, se r for um inteiro, então a fgm encontrada corresponde a r-ésima potência da função geratriz de momentos da distribuição exponencial, de parâmetroλ.

Agora, vamos derivá-la, para encontrar o primeiro momento, a partir do método de derivação de Leibniz. Considere $u = \left(\frac{\lambda}{\lambda - t}\right)$. Daí

$$\begin{aligned}
\frac{dM_X}{dt}\frac{dM_X}{du}\frac{du}{dt} &= ru^{r-1}\frac{\lambda}{(\lambda-t)^2} = r\left(\frac{\lambda}{\lambda-t}\right)^{r-1}\frac{\lambda}{(\lambda-t)^2}\\
&= r\frac{\left(\frac{\lambda}{\lambda-t}\right)^r}{\left(\frac{\lambda}{\lambda-t}\right)}\frac{\lambda}{(\lambda-t)^2} = \left(\frac{\lambda}{\lambda-t}\right)^r\frac{r}{(\lambda-t)}\\
\frac{(d^2M_X)}{dt} &= \left(\frac{\lambda}{\lambda-t}\right)^r\frac{r^2}{(\lambda-t)^2} + \left(\frac{\lambda}{\lambda-t}\right)^r\frac{r}{(\lambda-t)^2}\\
&= \left(\frac{\lambda}{\lambda-t}\right)^r\left[\frac{r^2}{(\lambda-t)^2} + \frac{r}{(\lambda-t)^2}\right]
\end{aligned}$$

Fazendo-se $t = 0$ nas expressões acima, é imediato constatar que

$$\begin{aligned}
E(X) &= \frac{r}{\lambda}\\
E(X^2) &= \frac{r^2+r}{\lambda^2}
\end{aligned}$$

Consequentemente,

$$Var(X) = \frac{r^2+r}{\lambda^2} - \frac{r^2}{\lambda^2} = \frac{r}{\lambda^2}$$

4.12 Distribuição de Qui-Quadrado

Trata-se de um caso particular, e muito importante, da distribuição gama. Corresponde à distribuição de probabilidade da soma dos quadrados de n variáveis aleatórias independentes, $X^2 = X_1^2 + X_2^2 + \ldots + X_n^2$, que são distribuídas normalmente e padronizadas com $\mu = 0$ e $\sigma^2 = 1$. Sua aplicabilidade é eficaz na realização de testes de hipóteses.

Definição 4.12.1. *Uma v.a. X, com as mesmas características da distribuição gama para $\lambda = \frac{1}{2}$ e $r = \frac{n}{2}$, onde $n \in Z_+^*$, tem distribuição de qui-quadrado com n graus de liberdade se sua fdp é dada na seguinte forma:*

$$f(x) = \begin{cases} \dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\dfrac{n}{2}\right)}e^{-x/2}x^{\frac{n}{2}-1}, & \text{se } x > 0\\ \\ 0, & \text{se } x \leq 0 \end{cases}$$

Vamos integrar essa função:

$$\frac{\left(\frac{1}{2}\right)^n}{\Gamma\left(\frac{n}{2}\right)}\int_0^\infty \frac{x^{\frac{n}{2}-1}e^{\frac{-x}{2}}}{2}dx \quad = \quad \frac{\left(\frac{1}{2}\right)^n}{2\Gamma\left(\frac{n}{2}\right)}\int_0^\infty x^{\frac{n}{2}-1}e^{\frac{-x}{2}}dx$$

Na segunda igualdade, considere $y = \frac{x}{2}$, daí

$$\frac{\left(\frac{1}{2}\right)^n}{2\Gamma\left(\frac{n}{2}\right)}\int_0^\infty \frac{(2y)^{\frac{n}{2}-1}e^{-y}}{\left(\frac{1}{2}\right)^n}dy \quad = \quad 1$$

Note que o numerador do integrando corresponde a $2\Gamma\left(\frac{n}{2}\right)$. Daí é imediato verificar o resultado acima.

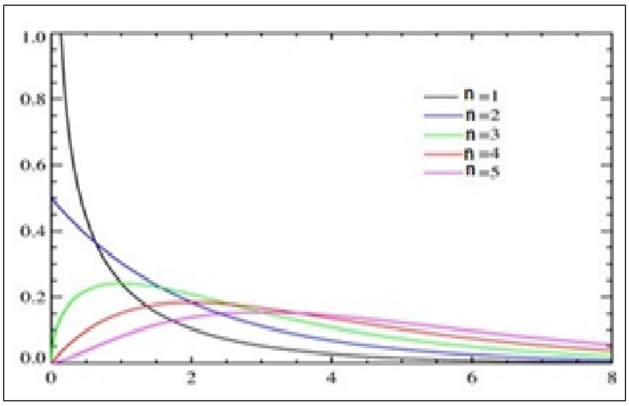

Figura 4.12 *Representação gráfica da distribuição Qui-quadrado Disponív1el em: ¡http://pt.wikipedia.org/wiki/Ficheiro:Chi-square-distributionPDF.png*

4.12.1 Tabulação de Qui-quadrado

Há uma tabela, incumbida de fornecer valores da abscissa para diversas áreas (probabilidades) da cauda direita. Sendo assim, não serão consideradas áreas da cauda esquerda. Deste modo, podemos encontrar, na tábua, valores que satisfaçam as condições exigidas nas aplicações.

Exemplo 4.12.1.1. *Se X^2_{25}, podemos determinar então $P[x_1 \leq X \leq x_2] = 0,95$ na seguinte forma:*

$P[x_1 \leq X \leq x_2] = P[X \leq x_2] - P[X \leq x_1] = 0,975 - 0,25 = P[X \leq 40,6] - P[X \leq 13,1] \Longrightarrow x_1 = 13,1$ *e* $x_2 = 40,6$

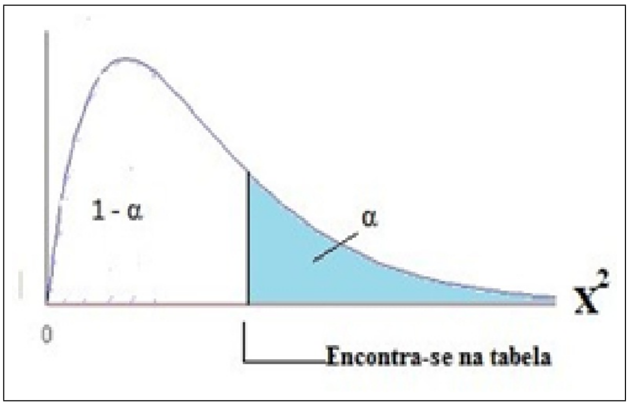

Figura 4.13 $P[X^2 \geq x] = \alpha$

α		97,5%	95%	90%	80%	30%	10%	5%	2,5%	2%	1%	0,2%	0,1%
GRAUS DE LIBERDADE (V)	1	0,001	0,004	0,016	0,064	1,074	2,706	3,841	5,024	5,412	6,635	9,549	10,827
	2	0,051	0,103	0,211	0,446	2,408	4,605	5,991	7,878	7,824	9,210	12,429	13,815
	3	0,216	0,352	0,584	1,005	3,665	6,251	7,815	9,348	9,837	11,345	14,796	16,266
	4	0,48	0,71	1,06	1,65	4,88	7,78	9,49	11,14	11,67	13,28	16,92	18,47
	5	0,83	1,15	1,61	2,34	6,06	9,24	11,07	12,83	13,39	15,09	18,91	20,51
	6	1,24	1,64	2,20	3,07	7,23	10,64	12,59	14,45	15,03	16,81	20,79	22,46
	7	1,69	2,17	2,83	3,82	8,38	12,02	14,07	16,01	16,62	18,48	22,60	24,32
	8	2,18	2,73	3,49	4,59	9,52	13,36	15,51	17,53	18,17	20,09	24,35	26,12
	9	2,70	3,33	4,17	5,38	10,66	14,68	16,92	19,02	19,68	21,67	26,06	27,88
	10	3,25	3,94	4,87	6,18	11,78	15,99	18,31	20,48	21,16	23,21	27,72	29,59
	15	6,26	7,26	8,55	10,31	17,32	22,31	25,00	27,49	28,26	30,58	35,63	37,70
	20	9,59	10,85	12,44	14,58	22,77	28,41	31,41	34,17	35,02	37,57	43,07	45,31
	25	13,12	14,61	16,47	18,94	28,17	34,38	37,65	40,65	41,57	44,31	50,22	52,62
	30	16,79	18,49	20,60	23,36	33,53	40,26	43,77	46,98	47,96	50,89	57,17	59,70

Figura 4.14 $P[X_9^2 \geq 16,92] = 0,05$

4.13 Distribuição T de Student

Em determinadas situações iremos trabalhar com amostras de tamanhos pequenos, e por isso, em geral não teremos o conhecimento do desvio padrão da população. É justamente nesse momento que faremos uso de uma distribuição eficiente para esse tipo de estudo: a distribuição t de Student.

Considere uma população com distribuição normal, da qual extraímos aleatoriamente uma amostra com v graus de liberdade, $x_1, x_2, \ldots, x_v$. Em seguida, calculamos $X^2_{(v)}$. Seja X a v.a. com distribuição de Qui-quadrado, e Y outra v.a. independente de X com distribuição normal $N(0,1)$ [15]. Então os valores da v.a. T são obtidos por:

$$T = \frac{\sqrt{Y}}{\sqrt{\dfrac{X}{v}}}$$

Definição 4.13.1. *Uma variável aleatória T segue o modelo t de Student se sua fdp for* *expressa na seguinte forma:*

$$f(t) \;=\; \frac{\Gamma\left(\dfrac{v+1}{2}\right)}{\Gamma\left(\dfrac{v}{2}\right)}\left(1+\frac{t^2}{v}\right)^{\frac{-(v+1)}{2}}, -\infty < t < \infty$$

O nome T-de Student é devido ao inglês Willian Sealy Gosset (1876-1937), que publicou este modelo com o pseudônimo Student, porque a Cervejaria Guinness não admitia a divulgação de resultados de análises alcançados por seus funcionários, em 1908 [23].

Enfatizando o que foi dito no início, o uso desta distribuição de probabilidades está associado a estudos com pequenas amostras, ou seja, para graus de liberdade menores que trinta. Assim, quanto menor for o valor de v mais achatada a curva se torna. Porém, quando a amostra é grande, o modelo se aproxima da distribuição Normal. E, quando v se aproxima e cem, a curva de Student é quase igual à curva de Gauss [2].

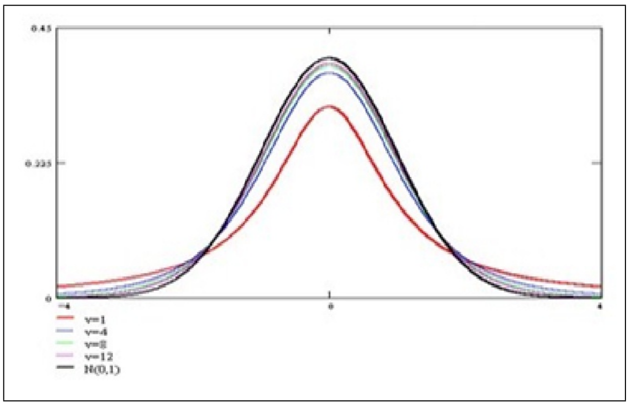

Figura 4.15 *Gráfico da fdp da distribuição T de Student Disponível em: ¡http://pt.wikipedia.org/wiki/Ficheiro:TStudent.png¿*

Características:

A curva $f(t,v)$ simétrica em relação a $t = 0$, campaniforme, e semelhante à curva normal padrão, porém com caudas mais largas, ou seja, uma simulação da t de Student pode gerar valores mais extremos que uma simulação da normal. O único parâmetro v que a define é o grau de liberdade, que caracteriza a sua forma. [INTERNET].

Para determinar a probabilidade $P[T > t_0]$ na figura a seguir, devemos primeiro lembrar que esse valor, dependente de t_0 e v, corresponde à soma das áreas rachuradas. Assim,

$$P[|T| > t_0] \;=\; \int_{-t_0}^{\infty} f(v,t)dt = p$$

4.13.1 Tabulação de Student

Felizmente a distribuição de Student está tabulada, o que nos fornece, de maneira mais fácil, o valor das abscissas para diversas áreas (probabilidades). Trata-se de uma tabela bicaudal.

Desta forma, para acharmos, por exemplo, o valor de t tal que $P[t_{11} \geq t] = 0,02$ devemos seguir os caminhos indicados na figura abaixo:

PROBABILIDADE DE UM t EM VALOR ABSOLUTO SUPERIOR AO TABELADO

GRAUS DE LIBERDADE -V	50%	40%	30%	20%	10%	5%	4%	2%	1%	0,2%
1	1,000	1,376	1,963	3,078	6,314	12,706	15,894	31,821	63,656	318,28
2	0,816	1,061	1,386	1,886	2,920	4,303	4,849	6,965	9,925	22,328
3	0,765	0,978	1,250	1,638	2,353	3,182	3,482	4,541	5,841	10,214
4	0,741	0,941	1,190	1,533	2,132	2,776	2,999	3,747	4,604	7,173
5	0,727	0,920	1,156	1,476	2,015	2,571	2,757	3,365	4,032	5,894
6	0,718	0,906	1,134	1,440	1,943	2,447	2,612	3,143	3,707	5,208
7	0,711	0,896	1,119	1,415	1,895	2,365	2,517	2,998	3,499	4,785
8	0,706	0,889	1,108	1,397	1,860	2,306	2,449	2,896	3,355	4,501
9	0,703	0,883	1,100	1,383	1,833	2,262	2,398	2,821	3,250	4,297
10	0,700	0,879	1,093	1,372	1,812	2,228	2,359	2,764	3,169	4,144
11	0,697	0,876	1,088	1,363	1,796	2,201	2,328	2,718	3,106	4,025

Figura 4.16 $P[t_{15} \geq 2,764] = 0,02$

Isto significa que a probabilidade de que (em valor absoluto) o estatístico t de Student calculado numa amostra de 11 observações, exceda 2,764 é 0,02.

4.14 Distribuição F de Snedecor

Considere duas variáveis aleatórias independentes U e V, ambas com distribuição de Qui-quadrado, respectivamente. Desta forma, uma v.a. W com distribuição F de Snedecor com v_1 graus de liberdade no numerador e v_2 graus de liberdade no denominador é expressa por:

$$w = \frac{v_2 U}{v_1 V}$$

A denominação de F foi indicada por George Snedecor (1881-1974), em tributo a Ronald Fisher (1890-1962), que foi o primeiro a estudar esta distribuição.

Definição 4.14.1. *Diz-se que uma v.a. X tem distribuição F de Snedecor se sua fdp é* *definida na seguinte forma:*

$$f(x, v_1, v_2) \quad = \quad \frac{\Gamma\left(\dfrac{v_1+v_2}{2}\right)}{\Gamma\left(\dfrac{v_1}{e}\right)\Gamma\left(\dfrac{v_2}{2}\right)}\left(\frac{v_1}{v_2}\right)^{\frac{v_1}{2}} \frac{x^{\frac{(v_1-2)}{2}}}{\left(\dfrac{1+v_1}{v_2 x}\right)^{\frac{(v_1+v_2)}{2}}}, \, se \ x > 0$$

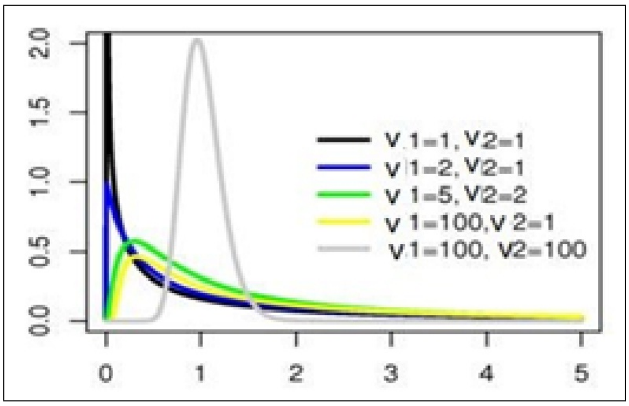

Figura 4.17 *Comportamento gráfico da distribuição de Snedecor*

GRAUS DE LIBERDADE DO NUMERADOR (V1)

GRAUS DE LIBERDADE DO DENOMINADOR (V2)	1	2	3	4	5	6	7	8	9	10
1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5	241,9
2	18,513	19,000	19,164	19,247	19,296	19,329	19,353	19,371	19,385	19,396
3	10,128	9,552	9,277	9,117	9,013	8,941	8,887	8,845	8,812	8,785
4	7,709	6,944	6,591	6,388	6,256	6,163	6,094	6,041	5,999	5,964
5	6,608	5,786	5,409	5,192	5,050	4,950	4,876	4,818	4,772	4,735
6	5,987	5,143	4,757	4,534	4,387	4,284	4,207	4,147	4,099	4,060
7	5,591	4,737	4,347	4,120	3,972	3,866	3,787	3,726	3,677	3,637
8	5,318	4,459	4,066	3,838	3,688	3,581	3,500	3,438	3,388	3,347
9	5,117	4,256	3,863	3,633	3,482	3,374	3,293	3,230	3,179	3,137
10	4,965	4,103	3,708	3,478	3,326	3,217	3,135	3,072	3,020	2,978
11	4,844	3,982	3,587	3,357	3,204	3,095	3,012	2,948	2,896	2,854

Figura 4.18 $P[F_{7,10} \geq 3,135] = 0,05$

Ou seja, num universo de duas amostras n_1 e n_2, respectivamente tais que $n_1 = 8$ e $n_2 = 11$ pode-se estimar que a probabilidade de o estatístico F exceder $3,135$ é $0,05$.

4.15 Distribuição Beta

Definição 4.15.1. *Diz-se que a variável aleatória X, com parâmetros $\alpha\lambda$, onde $\alpha > 0$ e $\lambda > 0$, tem distribuição Beta se sua densidade é dada por:*

$$f(x) = \begin{cases} \dfrac{1}{\beta} x^{\alpha-1}(1-x)^{\lambda-1}, & \textit{para } 0 < x < 1 \\ \\ 0, & \textit{em outro caso} \end{cases}$$

Em que β é a função beta, definida na seguinte maneira:

$$\beta = \int_0^1 x^{\alpha-1}(1-x)^{\lambda-1}dx = \frac{\Gamma(\alpha)\Gamma(\lambda)}{\Gamma(\alpha+\lambda)}$$

Agora vamos integrar a função f da definição:

$$\int_0^1 \frac{1}{\beta} x^{\alpha-1}(1-x)^{\lambda-1}dx = \frac{1}{\beta}\int_0^1 x^{\alpha-1}(1-x)^{\lambda-1}dx = 1$$

Veja que o integrando da segunda igualdade, na expressão acima, corresponde à função beta, daí resulta que f é de fato uma fdp.

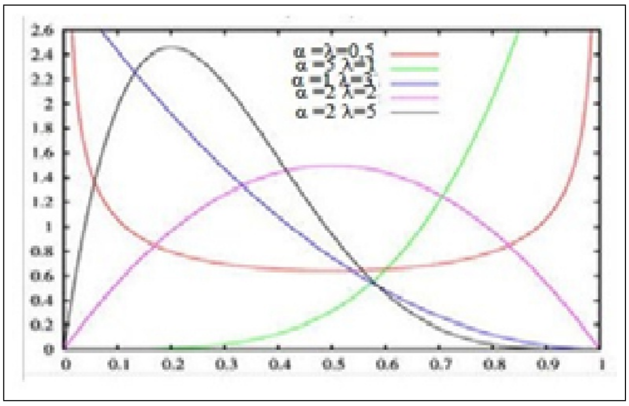

Figura 4.19 *Função densidade de uma v.a. com distribuição Beta*

A função de distribuição acumulada fica expressa na forma a seguir:

$$F(x) = \frac{\beta_y(\alpha,\lambda)}{\beta(\alpha,\lambda)} = I_x(\alpha,\lambda)$$

Demonstração 4.15.1.

$$F(x) = P[X \leq x] = \int_{-\infty}^{x} f(y)dy$$

$$\int_0^1 \frac{1}{\beta} y^{\alpha-1}(1-y)^{\lambda-1}dy = \frac{1}{\beta}\int_0^1 y^{\alpha-1}(1-y)^{\lambda-1}dy = \frac{\beta_y(\alpha,\lambda)}{\beta(\alpha,\lambda)} = I_x(\alpha,\lambda)$$

Um caso especial da distribuição beta, com $a = 1$ e $b = 1$ é a probabilidade uniforme.

Exemplo 4.15.1. $X \sim \beta$

i)

$$\begin{aligned}
E(X) = \frac{1}{\beta}\int_0^1 x x^{\alpha-1}(1-x)^{\lambda-1}dx &= \frac{\Gamma(\alpha+\lambda)}{\Gamma(\alpha)\Gamma(\lambda)}\int_0^1 x^{1+\alpha-1}(1-x)^{\lambda-1}dx \\
= \frac{\Gamma(\alpha+\lambda)}{\Gamma(\alpha)\Gamma(\lambda)}\beta(\alpha+1,\lambda) &= \frac{\Gamma(\alpha+\lambda)}{\Gamma(\alpha)\Gamma(\lambda)}\frac{\Gamma(\alpha+1)T(\lambda)}{T(\alpha+1+\lambda)} \\
&= \frac{T(\alpha+\lambda)}{T(\alpha)}\frac{T(\alpha+1)}{T(\alpha+1+\lambda)} = \frac{\alpha}{\alpha+\lambda}
\end{aligned}$$

$$\begin{aligned}
E(X^2) = \frac{\Gamma(\alpha+\lambda)}{\Gamma(\alpha)\Gamma(\lambda)}\int_0^1 x^2 x^{\alpha-1}(1-x)^(\lambda-1)dx &= \frac{\Gamma(\alpha+\lambda)}{\Gamma(\alpha)\Gamma(\lambda)}\beta(\alpha+2,\lambda) \\
= \frac{\Gamma(\alpha+\lambda)}{\Gamma(\alpha)\Gamma(\lambda)}\frac{T(\alpha+2)\Gamma(\lambda)}{T(\alpha+2+\lambda)} &= \frac{\alpha(\alpha+1)}{(\alpha+\lambda)(\alpha+1+\lambda)}
\end{aligned}$$

Substituindo esses resultados na fórmula da variância, resulta que:

$$Var(X) = \frac{\alpha\lambda}{(\alpha+\lambda)^2(\alpha+1+\lambda)}$$

4.16 Distribuição de Cauchy

Considere a seguinte função:

$$f(x) = \frac{1}{\pi\sigma\left[1+\left(\frac{x-\mu}{\sigma}\right)^2\right]} = \frac{1}{\pi}\left[\frac{1}{\sigma+\sigma\left(\frac{x-\mu}{\sigma}\right)^2}\right] =$$

$$\frac{1}{\pi}\left[\frac{1}{\frac{\sigma^2+(x-\mu)^2}{\sigma}}\right] = \frac{1}{\pi}\left[\frac{\sigma}{\sigma^2+(x-\mu)^2}\right]$$

Em que:

$\mu =$ parâmetro que localiza o ápice da distribuição

$\sigma =$ parâmetro de escala

$\frac{1}{\gamma\pi} =$ amplitude

Um caso exclusivo desta função é obtido quando consideramos$\mu = 0$ e $\sigma = 1$, cuja definição é dada a seguir.

Definição 4.16.1. *Diz-se que uma v.a. X tem distribuição de Cauchy (padrão) se sua fdp for representada na seguinte forma:*

$$f(x) \quad = \quad \frac{1}{\pi(1+x^2}, \textit{para } -\infty < x < \infty$$

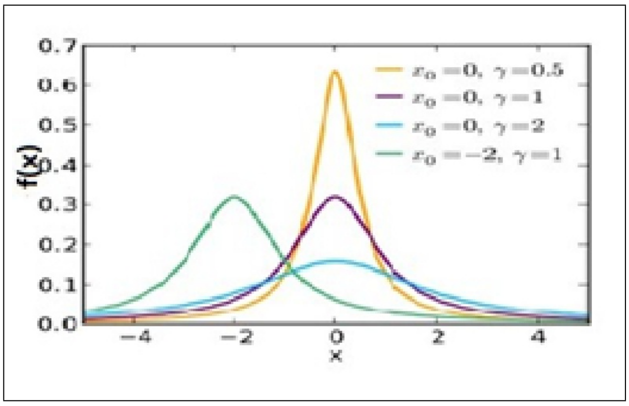

Figura 4.20 *Gráfico da fdp de Cauchy Disponível em: <http://en.wikipedia.org/wiki/Cauchydistribution>*

A função repartição é, então, descrita na seguinte expressão:

$$F(X \leq x) \quad = \quad \frac{1}{2} + \frac{1}{\pi} \arctan x$$

Demonstração 4.16.1.

$$F(X \leq x) \quad = \quad \int_{-\infty}^{x} \frac{1}{\pi(1+y^2)} dy$$

Vamos fazer mudança de variável. Considere $t = \tan\theta$, logo $\theta = \arctan t$ daí

$$\begin{aligned} \frac{1}{1+t^2} dt &= \frac{1}{1+\tan^2\theta} \sec^2\theta d\theta \\ \frac{1}{\dfrac{\cos^2\theta + sen^2\theta}{\cos^2\theta}} \sec^2\theta d\theta &= \frac{1}{\dfrac{1}{\cos^2\theta}} \sec^2\theta d\theta = d\theta \end{aligned}$$

Consequentemente,

$$F(x) = \int_{-\frac{\pi}{2}}^{\arctan x} \frac{1}{\pi} d\theta \quad = \quad \frac{1}{\pi}\, [\theta]|_{-\frac{\pi}{2}}^{\arctan x} = \frac{1}{\pi}\left[\arctan x + \frac{\pi}{2}\right] = \frac{1}{2} + \frac{1}{\pi} \arctan x$$

Exemplo 4.16.1. $X \sim C$

$$E(X) = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{x}{1+x^2}dx = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{2x}{1+x^2}dx$$

E como,

$$\begin{aligned}
\frac{1}{2\pi}\int_{-\infty}^{\infty}\left|\frac{2x}{1+x^2}\right|dx &= \frac{1}{2\pi}\int_{-\infty}^{0}\left|\frac{2x}{1+x^2}\right|dx + \frac{1}{2\pi}\int_{0}^{\infty}\left|\frac{2x}{1+x^2}\right|dx \\
&= \frac{1}{2\pi}\int_{-\infty}^{0}\frac{-2x}{1+x^2}dx + \frac{1}{2\pi}\int_{0}^{\infty}\frac{2x}{1+x^2}dx \\
&= \frac{1}{2\pi}\int_{0}^{\infty}\frac{2x}{1+x^2}dx = \frac{1}{\pi}\log(1+x^2)\Big|_0^{\infty} \longrightarrow \infty
\end{aligned}$$

Resulta que $E(X)$ não existe, portanto os momentos relacionados a $E(X)$ não são determinados.

4.17 Distribuição de Weibull

Devido sua flexibilidade, esta distribuição torna-se importante para estudos sobre tempo de vida de equipamentos eletrônicos e estimativas de falhas. Dependendo do valor de seus parâmetros, ela pode fazer uma imitação de outras distribuições de probabilidade, como por exemplo, as distribuições: normal e exponencial.

Definição 4.17.1. *Uma variável aleatória X, com parâmetros α (de escala) e β (de forma), tem distribuição de Weibull se sua fdp é representada como na forma a seguir.*

$$f(x) = \begin{cases} \frac{\beta}{\alpha^\beta}x^{\beta-1}e^{-\left(\frac{x}{\alpha}\right)^\beta}, & se\ x \geq 0 \\ 0, & se\ x < 0 \end{cases}$$

Note que:

Se $\beta = 1$, a distribuição de Weibull se reduz imediatamente à distribuição exponencial de parâmetro $\lambda = \frac{1}{\alpha}$.

Exemplo 4.17.1. $X \sim W(\alpha, \beta)$

i)

$$E(X) = \int_0^{\infty} x\frac{\beta}{\alpha}\left(\frac{x}{\alpha}\right)^{\beta-1}e^{\left(-\frac{x}{\alpha}\right)^\beta}dx$$

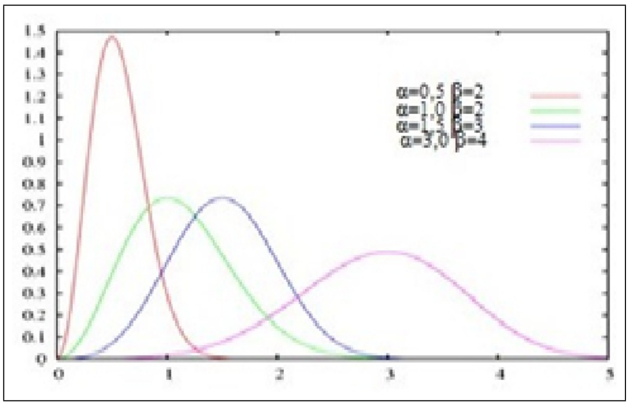

Figura 4.21 *Gráfico da fdp de Weibull Disponível em: http://pt.wikipedia.org/wiki/Ficheiro:Weibulpdf.png*

Considere $z = \left(\frac{x}{\alpha}\right)^{\beta}$. *Logo,* $x = \alpha z^{\frac{1}{\beta}}$ *e* $dx = \frac{\alpha}{\beta} z^{\frac{1-\beta}{\beta}} dz$.

Substituindo esses resultados na expressão acima, vem:

$$\begin{aligned} E(X) &= \int_0^\infty \alpha z^{\frac{1}{\beta}} \frac{\beta}{\alpha} z^{\frac{\beta-1}{\beta}} e^{-z} \frac{\alpha}{\beta} z^{\frac{1-\beta}{\beta}} dz \\ &= \alpha \int_0^\infty z^{\frac{1}{\beta}} e^{-z} dz = \alpha \Gamma\left(\frac{\beta+1}{\beta}\right) \end{aligned}$$

Analogamente,

$$\begin{aligned} E(X^2) &= \int_0^\infty \left(\alpha z^{\frac{1}{\beta}}\right)^2 e^{-z} dz = \alpha^2 \int_0^\infty z^{\frac{2}{\beta}} e^{-z} dz \\ &= \alpha^2 \Gamma\left(\frac{\beta+2}{\beta}\right) \end{aligned}$$

Assim,

$$\begin{aligned} Var(X) &= \alpha^2 \Gamma\left(\frac{\beta+2}{\beta}\right) - \alpha^2 \left(\Gamma\left(\frac{\beta+1}{\beta}\right)\right)^2 \\ &= \alpha^2 \left[\Gamma\left(\frac{\beta+2}{\beta}\right) - \left(\Gamma\left(\frac{\beta+1}{\beta}\right)\right)^2\right] \end{aligned}$$

Capítulo 5

TRANSFORMADA DE FOURIER

É imprescindível que nos recordemos dos logaritmos, que nos foram apresentados simplesmente como um recurso de cálculo. Por exemplo, se nos encontrássemos na responsabilidade de calcular $\frac{x}{y}$ partiríamos da idéia de se obter à priori os valores de $\log x$ e $\log y$ para então aplicar a propriedade $\log x - \log y$, que representa $\log \frac{x}{y}$, e a partir disso seria possível se chegar, finalmente, ao valor de $\frac{x}{y}$. De forma análoga podemos simplificar o cálculo das demais operações aritméticas.

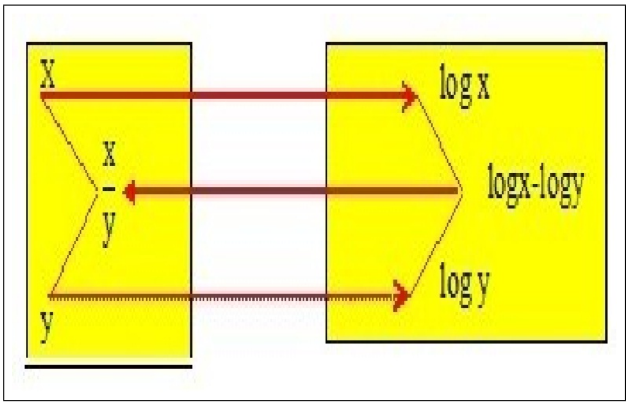

Figura 5.1 *Representação sistemática dos logarítmos*

A correspondência biunívoca existente entre x e $\log x$, é que nos permite, em muitas equações, substituí-las por outras mais simples por intermédio dos números "transformados". Prosseguindo neste raciocínio, lembremo-nos também das equações diferenciais:

O problema da resolução dessas equações está na presença de derivadas na função incógnita[11]. Por exemplo, considere $f(x) = 7y" + 3y' + 2y$. Podemos tentar resolvê-la da seguinte forma: $f(x) = 7D^2y + 3Dy + 2y$, onde $D = \frac{df(X)}{dX}$.

A dificuldade nos é apresentada ao tentarmos resolver esta equação como se fosse algébrica e escrever $y(x) = \frac{f(x)}{7D^2 + 3D + 2}$, o que não faz sentido!

A concepção de transformadas pode ser entendida através da Figura 5.2, onde T representa o processo de transformação de $f(x)$ para $F(E)$. Daí, $F(E) = T[f(x)]$.

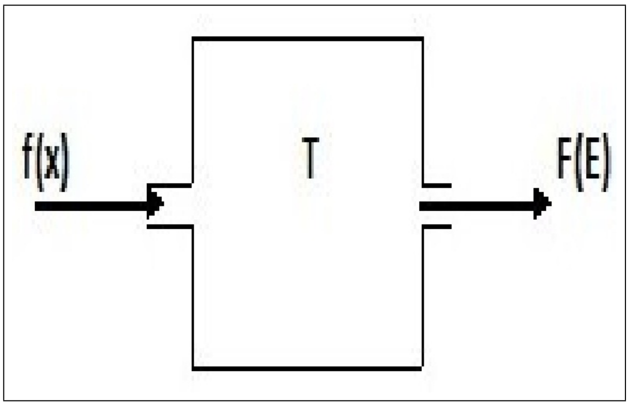

Figura 5.2 *Terminologia de sistema para transformadas*

De forma similar, trataremos de uma função que é de grande importância, uma vez que o emprego desta se dá pelo fato de sempre existir, enquanto que a função geratriz de momentos pode não existir, como é o caso do modelo de Cauchy, visto no capítulo anterior, que em virtude de a integral que define sua fdp, nem sempre ser finita. Entretanto, teremos o trabalho de envolver funções de valores complexos.

Definição 5.0.2. *Diremos que uma variável aleatória X, é complexa se puder ser escrita na seguinte forma:*

$$X = X^a + iX^b$$

Em que, X^a e X^b são variáveis aleatória reais, e $i = \sqrt{-1}$.

Definição 5.0.3. *Seja (Ω, β, P) um espaço de probabilidade, e seja X uma variável aleatória sobre β. A função $\Phi_X : \Re \to C$ dada por, $\Phi_X(t) = E(e^{itX}) = E(\cos tX + i \sin tX)$, se chama Transformada de Fourier ou função característica da variável aleatória X. Daí, temos que:*

$$\Phi_X(t) = E(e^{itX}) = \begin{cases} \sum_{j=1}^{\infty} e^{itxj} P[X = xj], & \text{se } X \text{ é discreta} \\ \int_{-\infty}^{\infty} e^{tx} f(x)dx, & \text{se } X \text{ é contínua} \end{cases}$$

Nota: Dizer que $\Phi x : R \to C$ está bem definida, significa que para cada t, a integral converge para um número,e, assim temos uma função Φ definida em $\Re$.[14]

Proposição 5.0.1. *A função característica verifica:*

a) $\Phi_X(t)$ *é uma função uniformente contínua.*

b) $\Phi_{X(0)} = 1$, *pois a função característica não se anula em uma vizinhança de zero.*

c) $|\Phi_X(t)| \leq 1$

d) $\Phi_{aX+b}(t) = e^{ibt}\Phi_X(at)$

e) $\Phi_X(-t) = \overline{\Phi_X}(t) \forall\ t \in R$

f) *Se as variáveis aleatórias X e Y são independentes, então* $\Phi_{(X+Y)}(t) = \Phi_X(t)\Phi_Y(t)$

g) *Se $E(|X|^n) < \infty$, para um n natural, então existe a derivada de ordem n Φ_X para $t = 0$, e*

$$\frac{d^m}{dt^m}\Phi_X(t) \quad = \quad i^m E(X^m), m \leq n$$

h) $\Phi_X(t)$ *é positiva definida. Isto é, para todo natural n=1, 2, ... e para quaisquer reais t_1, t_2, ..., t_n e complexos z_1, z_2, ..., z_n vale:*

$$\sum_{j=1}^{n}\sum_{k=1}^{n}\Phi_X(t_j - t_k)z_j\overline{z_k} \geq 0$$

Demonstração.

a) Observe que,

$$\begin{aligned}
|\Phi_X(t+h) - \Phi_X(t)| &= |E(e^{i(t+h)X} - e^{itX})| \\
&\leq E|(e^{i(t+h)X} - e^{itX)}| \\
&= E|e^{itX}(e^{ihX} - 1)| \\
&= E(|e^{itX}||(e^{ihX} - 1)|)
\end{aligned}$$

Fazendo-se $|e^{itX}| = 1$, temos:

$$|\Phi_X(t+h) - \Phi_X(t)| \leq E\left(|e^{ihX} - 1|\right)$$

Daí,

$$\begin{aligned}
\lim_{h\to 0}|\Phi_X(t+h)-\Phi_X(t)| &\leq \lim_{h\to 0} E(|e^{ihX}-1|)\\
&= E\left(\lim_{h\to 0}\Big| e^{ihX}-1\Big|\right)=0
\end{aligned}$$

E como o desenvolvimento acima não depende de t, então a continuidade é uniforme.

b)

$$\Phi_X(0)=E(e^{0Xi})=E(1)=1$$

c) Considere $|z|=\sqrt{a^2+b^2}$. Daí,

$$\begin{aligned}
|\Phi_X(t)| &= |E(e^{itX})|=|E[\cos(tX)]+iE[\sin(tX)]|\\
&\leq \sqrt{E^2[\cos(tX)]+E^2[\sin(tX)]}\\
&= \sqrt{E[\cos^2(tX)]+E[\sin^2(tX)]}\\
&= \sqrt{E[\cos^2(tX)]+[\sin^2(tX)]}=1
\end{aligned}$$

d)

$$\begin{aligned}
\Phi_{aX+b}(t) &= E(e^{it(aX+b)})\\
&= (e^{itaX})e^{itb}=e^{itb}E(e^{itaX})\\
&= e^{ibt}\Phi_X(at)
\end{aligned}$$

e) Considere que $\overline{Z}=(a,-b)=a-bi$, daí,

$$\Phi_X(-t)=E(e^{-itX})=(\Phi_X(t))=\overline{\Phi_X(t)}$$

f)

$$\begin{aligned}
\Phi_{X+Y}(t) &= E(e^{it(X+Y)})=E(e^{itX}e^{itY})\\
&= E(e^{itX})E(e^{itY})\\
&= \Phi_X(t)\Phi_Y(t)
\end{aligned}$$

g)

$$\begin{aligned}\frac{d^m}{dt^m}\Phi_X(t) &= \frac{d^m}{dt^m}E(e^{itX}) \\ &= E\left(\frac{d^m}{dt^m}e^{itX}\right) = E(i^m X^m e^{itX})\end{aligned}$$

Fazendo-se $t = 0$, vem

$$\frac{d^m}{dt^m}\Phi_X(0) = E(i^m X^m) = i^m E(X^m)$$

Ou seja, obtemos os momentos, porém levando-se em consideração as potências de i.

h)

$$\begin{aligned}\sum_{j=1}^{n}\sum_{k=1}^{n}\Phi_X(t_j - t_k)z_j\overline{z_k} &= \sum_{j=1}^{n}\sum_{k=1}^{n}E(e^{iX(t_j-t_k)})z_j\overline{z_k} \\ &= \sum_{j=1}^{n}\sum_{k=1}^{n}E(z_j e^{iXt_j}\overline{z_k}e^{-iXt_k}) \\ &= E\left(\sum_{j=1}^{n}\sum_{k=1}^{n}(z_j e^{iXt_j}\overline{z_k e^{iXt_k}})\right) \\ &= E\left[\left(\sum_{j=1}^{n}z_j e^{iXt_j}\right)\overline{\left(\sum_{k=1}^{n}z_k e^{iXt_k}\right)}\right] \\ &= E\left(\left|\sum_{j=1}^{n}z_j e_j^{iXt}\right|^2\right) \geq 0\end{aligned}$$

Por tanto, $\Phi_X(t)$ é, de fato, positiva e definida. ■

5.1 Fórmula de Inversão ou Transformada inversa de Fourier

As fórmulas de inversão da transformada de Fourier permitem obter a função de distribuição de probabilidade bem como a densidade de uma v.a. X a partir do conhecimento de $\Phi_X(t)$, ja que a função característica é finita para todas as variáveis aleatórias X e para todos os números reais t [26].

Definição 5.1.1. *Seja $f : \Re \rightarrow C$ uma função, e seja $F(t)$ sua transformada de Fourier. Se uma v.a. X admite uma densidade $f(x)$, esta pode ser obtida a partir da seguinte função:*

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \Phi_X(t)e^{-tix}dt$$

Caso contrário, temos o seguinte resultado:

$$f(b) - f(a) = \lim_{T\rightarrow\infty} \frac{1}{2\pi}\Phi_X(t)\frac{e^{itb} - e^{ita}}{it}dt$$

Exemplo 5.1.1. *Considere a v.a. $X \sim C(0,1)$. Vamos determinar sua função característica.*

i) Vamos partir da fdp

$$f(x) = \frac{1}{\pi(1+x^2)}$$

De onde se pode encontrar a função característica, que é dada por:

$$\Phi_X(t) = \frac{1}{\pi}\int_{-\infty}^{\infty} e^{itX}\frac{1}{1+x^2}dx$$

Agora, considere a seguinte fdp:

$$f_1(x) = \frac{1}{2}e^{-|x|}$$

Daí,

$$\begin{aligned}
\Phi_1(t) &= \frac{1}{2}\int_{-\infty}^{\infty} e^{itx}e^{-|x|}dx = \frac{1}{2}\int_{-\infty}^{\infty}(\cos tx + i\sin tx)e^{-|x|}dx \\
&= \frac{1}{2}\int_{-\infty}^{0}(\cos tx + i\sin tx)e^{x}dx + \frac{1}{2}\int_{0}^{\infty}(\cos tx + i\sin tx)e^{-x}dx \\
&= -\frac{1}{2}\int_{\infty}^{0}(\cos tx - i\sin tx)e^{-x}dx + \frac{1}{2}\int_{0}^{\infty}(\cos tx + i\sin tx)e^{-x}dx \\
&= \int_{0}^{\infty} e^{-x}\cos tx dx
\end{aligned}$$

Integrando por partes, temos:

$$\begin{aligned}\int_0^\infty e^{-x}\cos tx dx &= -e^{-x}\cos(tx)|_0^\infty - t\int_0^\infty e^{-x}\sin(tx)dx\\ &= 1 - t\int_0^\infty e^{-x}\sin(tx)dx\end{aligned}$$

Analogamente, encontramos:

$$\int_0^\infty e^{-x}\sin(tx)dx = e^{-x}\sin(tx)|_0^\infty + t\int_0^\infty e^{-x}\cos(tx)dx$$

Logo,

$$\begin{aligned}\Phi_1(t) &= 1 - t^2\Phi_1(t)\\ \Phi_1(t) + t^2\Phi_1(t) &= 1\\ \Phi_1(t)(1+t^2) &= 1.\end{aligned}$$

Consequentemente,

$$\Phi_1(t) = \frac{1}{1+t^2}$$

ii) $\Phi_1(t)$ é absolutamente integrável em $(-\infty, +\infty)$, e podemos encontrar a fdp correspondente por intermédio da fórmula de inversão abaixo:

$$f_2(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-itx}}{1+t^2}dt$$

Como a função característica determina de forma única a fdp, encontramos:

$$e^{-|x|} = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{e^{-itx}}{1+t^2}dx$$

Mudando t em $-t$, obtemos:

$$e^{-|x|} = -\frac{1}{\pi}\int_{\infty}^{-\infty}\frac{e^{itx}}{1+t^2}dt = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{e^{itx}}{1+t^2}dt$$

Finalmente, mudando as funções de t e de x, encontramos:

$$e^{-|t|} = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{e^{itx}}{1+x^2}dx$$

e portanto,

$$\Phi_X(t) = e^{-|t|}$$

Exemplo 5.1.2. *Determinar a função densidade da v.a. X, cuja função característica é dada por:*

$$\Phi_X(t) = \begin{cases} 1-|t|, & se\ |t| \leq 1 \\ 0, & se\ |t| > 1 \end{cases}$$

Aplicando a fórmula de inversão de Fourier:

$$\begin{aligned} f(x) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itX}\Phi_X(t)dt \\ &= \frac{1}{2\pi}\int_{-1}^{0}(1+t)e^{-itX}dt + \frac{1}{2\pi}\int_{0}^{1}(1-t)e^{-itX}dt \end{aligned}$$

i)

$$\begin{aligned} \int_{-1}^{0}(1+t)e^{-itX}dt &= \left[\frac{e^{-itX}}{-ix}(1+t)\right]\Bigg|_{-1}^{0} - \frac{1}{-ix}\int_{-1}^{0} e^{-itX}dt \\ &= -\frac{1}{ix} + \frac{1}{ix}\left[\frac{e^{-itX}}{-ix}\right]\Bigg|_{-1}^{0} \\ &= -\frac{1}{ix} - \frac{1}{(ix)^2}(1-e^{iX}) \end{aligned}$$

ii)

$$\begin{aligned} \int_{0}^{1}(1-t)e^{-itX}dt &= \left[\frac{e^{-itX}}{-ix}(1-t)\right]\Bigg|_{0}^{1} + \frac{1}{-ix}\int_{0}^{1} e^{-itX}dt \\ &= \frac{1}{ix} - \frac{1}{ix}\left[\frac{e^{-itX}}{-ix}\right]\Bigg|_{0}^{1} = \frac{1}{ix} + \frac{1}{(ix)^2}(e^{-iX}-1) \end{aligned}$$

Logo,

$$\begin{aligned} f(x) &= \frac{1}{2\pi x^2}\left(2 - e^{iX} - e^{-iX}\right) = \frac{1}{\pi x^2}\left(1 - \frac{e^{iX} + e^{-iX}}{2}\right) \\ &= \frac{1 - \cos(x)}{\pi x^2},\ -\infty < x < +\infty \end{aligned}$$

Considerações finais

É evidente que o estudo das variáveis aleatórias torna-se incompleto, caso não se pesquise também sobre momentos, acarretando suas relações com estas. Nos modelos matemáticos não-determinísticos, ou aleatórios, estudados, os parâmetros exercem função grande valia, já que fornecem informações sobre a distribuição de probabilidade; eles se enquadram na mesma proporção em que a declividade de uma reta fornece valiosos dados sobre a relação linear que representa. Dentre as tantas informações que se pode ter, dependendo da ordem do momento, em geral fizemos analogia à posição, nível de concentração de probabilidade e no caso de um vetor bidimensional (X, Y) estudamos o grau de associação e de dependência entre as $v.a.$ unidimensionais X e Y.

Também apresentamos um tratamento adequado à fgm pela sua importância em determinar momentos ante um processo composto de duas etapas, sendo a primeira baseada no artifício de derivação de M_x em relação a t, e a segunda, no estabelecimento de $t = 0$ para essa derivada. Porém verificamos que há um retrocesso de seu uso quando a integral que a define não existir, como é no caso da distribuição de Cauchy.

Felizmente mencionamos outra função que supre essa falha: a função característica Φ_X, que ao contrário desta, sempre existe para todos os valores de t. Vimos que, após aplicarmos Φ_X, efetuamos o cálculo de interesse e obtemos o resultado no domínio das transformadas. E se for de intersse, podemos voltar ao domínio das funções originais, por intermédio da fórmula de inversão [MAGALHÃES].

Uma extensão deste trabalho está na abordagem das desigualdades de Tchebycheff

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

e de Markov,

$$P(|X|) \leq \frac{E[X]}{\lambda^2}$$

bem como no estudo das variáveis aleatórias em dimensões de maior ordem, cujas formulações dependem de vários argumentos utilizados aqui.

BIBLIOGRAFIA

[1] APOSTOL, Tom M. *Cálculus volume II.*2.ed.New York: Springer,1969.

[2] ARANGO, Héctor G.*Bioestatística teórica e computacional.*3.ed.Rio de Janeiro: Editora Guanabara koogan S.A., 2009.

[3] ALVES, Mariane B.*Cálculo das probabilidades II.*Rio de Janeiro: Instituto de Matemática e Estatística, 2006.

[4] BERTSEKAS, Dimitry P.;TSITSIKLIS, J.N. *Introduction to probability.*(notas).
[5] BOYER, Carl B. *História da matemática.* 2.ed.São Paulo: ABDR, 1996.

[6] CORINA,R.;LEBLANC, R.; RÉMILLARD, B.;LAROC,D.*Théorie des probabilités.*Québec :Presses de L'université du Québec, 2002.

[7] CALLEGARI-JACQUES, Sídia M.*Bioestatística princípios e aplicações.*Porto Alegre: Artmed, 2003.

[8] CANCHO, Vicente Garibay. *Noções de Estatística e Probabilidade.* Ouro Preto: Departamento de ciências exatas e biológicas, 2004.

[9] CORREA, Sônia M.B.B. *Probabilidade e estatística.* 2ª ed. Belo Horizonte: PUC Minas Virtual, 2003.

[10] CHUNG, Kai Lai.*Elementary probability theory with stochastic processes and an introduction to mathematical finance.* New York: Academic Press, 2002.

[11] CUADRAS, Charles M.*Problemas de probabilidad e estadística.*vol1.Barcelona: PPU, 1990.

[12] DANTAS, C.A.B.*Probabilidade: um curso introdutório.*3.ed.São Paulo:Edusp, 2008.

[13] DELMAS, Jean-François. *Introductionau calcul des probabilités et à la statistique.* %Paris. LesPresses de L′ensta, 2010.

[14] FIGUEIREDO, Djairo.G.*Análise de Fourier e equações diferenciais parciais.*Rio de janeiro: IMPA-Projeto Euclides, 1977.

[15] GUIDORIZZI, H. L.*Um curso de cálculo*.vol2. Rio de Janeiro: LTC, 2011.

[16] HOEL,P.G.;PORT, S.C.*Introdução à teoria da probabilidade*.Rio de Janeiro: Interciência, 1978.

[17] LIPSCHUTZ, Seymour.*Probabilidade*.3.ed.São Paulo: McGraw-Hill, 1972.

[18] MAGALHÃES, *Noções de probabilidade e estatística*.5.ed.São Paulo:Edusp, 2008.

[19] MAGALHÃES, *Probabilidade e variáveis aleatórias*. 2.ed.São Paulo: Edusp, 2006.

[20] MEYER, Paul L. *Probabilidade aplicações à estatística*. 2.ed.Rio de Janeiro: LTC, 2009

[21] MORETTIN, Luiz G. *Estatística básica*. 7.ed.São Paulo : Pearson Education do Brasil,1999.

[22]RINCÓN, Louis. *Una introduccion a la probabilidade y estadística*. México: Departamento de matemáticas UNAM, 2006.

[23] SANTALÓ, Luis A.*Probabilidad e inferência estadística*.2.ed.Buenos Aires: Eva V. Chesneau, 1975.

[24] SPIEGEL, Murray R.*Probabilidade e Estatística*.São Paulo: McGraw-Hill,1978.

[25] TRIOLA, Mario F.*Elementary Statistics using the graphing calculator*. New York: Pearson, 2005.

[26] VEYSSEYRE, Renée.*Statistique et probabilités pour l'ingenieur*.2.ed.Paris : Dunod, 2006.

[27] VIEIRA, Sônia.*Introdução à bioestatística*.3.ed.Rio de Janeiro: Campus,1980.

[28] YIANNOUTSOS, Constantin.*Principles of Biostatistics*. EUA: Harvard School of Public Health-class notes.

[29] http://www.galileu.esalq.usp.br/mostra_topico.php?cod=285

[30] http://pt.wikipedia.org/wiki/Transformada_de_Fourier

[31] http://pt.wikipedia.org/wiki/Distribui%C3%A7%C3%A3o_uniforme

[32] http://www.mspc.eng.br/matm/prob_est230.shtmldistr_unif

[33] http://www.leg.ufpr.br/ silvia/CE701/node36.html

[34] http://risk.nuvvo.com/lesson/5861-distribuicao-de-probabilidade

[35]http://www.portalaction.com.br/content/63-distribui%C3%A7%C3%A3o-qui-quadrado

[36]http://pt.wikipedia.org/wiki/Pierre_de_Fermat

[37]http://pensee-classique.ens-lyon.fr/spip.php?article635

[38]http://afilosofia.no.sapo.pt/Hist.htm

[39]http://www.moraissilva.com/pdf%20curiosidades/77%20curiosidade%20agulha%20de%20Buffon.pdf

[40]http://fr.wikipedia.org/wiki/Paradoxe_de_Saint-P%C3%A9tersbourg

[41] http://www.lume.ufrgs.br/bitstream/handle/10183/1961/000362539.pdf?sequence=1

[42] http://educacao.uol.com.br/biografias/felix-borel.jhtm

[43] http://www.ppge.ufpr.br/teses/M07_rotunno.pdf

[44] http://www.learn-math.info/portugal/historyDetail.htm?id=Frechet

[45] http://pt.wikipedia.org/wiki/Andrei_Kolmogorov

[46] http://www.brasilescola.com/matematica/historia-probabilidade.htm