Universidade Estadual da Paraíba
Centro de Ciências e Tecnologia
Departamento de Estatística

Juliana Sales de Lima

Relação entre as distribuições t de Student, Normal e Qui-quadrado

Campina Grande
Junho de 2014

Juliana Sales De Lima

Relação entre as distribuições t de Student, Normal e Qui-quadrado

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de bacharel em Estatística.

Orientador:

Gustavo Henrique Esteves

Campina Grande

Junho de 2014

L732r Lima, Juliana Sales de.

Relação entre as distribuições t de Student, normal e qui-quadrado [manuscrito] / Juliana Sales de Lima. - 2014.

44 p. : il. color.

Digitado.

Monografia (Graduação em Estatística) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2014.

"Orientação: Prof. Dr. Gustavo Henrique Esteves, Departamento de Estatística".

1. Teoria da probabilidade. 2. Distribuição t-Student. 3. Simulação. I. Título.

21. ed. CDD 519.53

Juliana Sales De Lima

Relação das distribuições t de Student, Normal e Qui-quadrado

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de bacharel em Estatística.

Aprovado em: 30 / 06 / 2014

Banca Examinadora:

Gustavo H. Esteves

Gustavo Henrique Esteves
Universidade Estadual da Paraíba - DE/CCT
Orientador

Tiago Almeida de Oliveira

Tiago Almeida de Oliveira
Universidade Estadual da Paraíba - DE/CCT
Examinador

DMEsteves

Divanilda Maia Esteves
Universidade Estadual da Paraíba - DE/CCT
Examinadora

Dedicatória

A Deus por tudo que me proporciona na vida.

A vocês que sempre me fizeram acreditar na realização dos meus sonhos e trabalharam muito para que eu pudesse realizá-los, meus pais, Júlia e Roberto.

A meu esposo Edriano que tem aturado minha falta de tempo.

A ao meu "filho"Joaquim pela alegria e diversão.

E a todos os meu professores que me acompanharam até agora.

Agradecimentos

Agradecer primeiramente a Deus, por me iluminar e abençoar minha trajetória.

Ao meu pai Roberto, e minha mãe Júlia, pelo apoio e por tudo que sempre fizeram por mim, pela simplicidade, exemplo, amizade-e carinho, fundamentais na construção do meu caráter. Além da ajuda incansável da minha mãe para ajudar nos cuidados com meu filho, você é demais!

Ao meu amado filho Joaquim, que chegou para alegrar as nossas vidas. Hoje a minha vitória também é dele.

Ao meu esposo, Edriano, pacientemente sempre me dando conselhos, força, coragem e incentivo.

Ao orientador Gustavo Esteves, pelo apoio e conhecimento transmitido.

A todos que de alguma forma ajudaram, agradeço por acreditarem no meu potencial, nas minhas ideias, nos meus devaneios, principalmente quando nem eu mais acreditava.

E por último, e não menos importante, obrigada meus amigos de curso em especial, Evelyne, Moniclaudia e Socorro. Sem vocês nada disso seria possível.

Resumo

O presente trabalho faz uma breve revisão de probabilidade, nos mostrando que o cálculo das probabilidades pertence ao campo da Matemática, entretanto a maioria dos fenômenos de que trata a Estatística são de natureza aleatória ou probabilística. Conhecer muitas das distribuições de probabilidades é de extrema importância, pois as mesmas contribuem para toda teoria estatística. Neste sentido, este trabalho apresenta um importante resultado teórico de probabilidade mostrando a relação entre as distribuições normal, qui-quadrado e t-Student, usando o $software$ estatístico **R**, foi possível verificar a existência de relação entre estas distribuições, através de simulação com diferentes algoritmos para geração de números aleatórios, com amostras de tamanho 1000 para cada distribuição e em cada simulação, graus de liberdade diferentes 1,15 e 30, fazendo assim o quociente de uma Normal padrão por raiz quadrada de uma Qui-quadrado com ν graus de liberdade, e através de histogramas e o gráfico $qq - plot$ observar se as distribuições dadas apresentam um comportameto semelhante à distribuição t de Student.

Palavras-chave: Teoria de Probabilidade, Distribuição t-Student, Simulação.

Abstract

This work contains a brief review about probability theory, showing that the calculation of probabilities belongs to the field of mathematics, however most of the phenomena studied in Statistics are of random or probabilistic nature. The knowledge of probability distributions is of utmost importance because they contribute to the whole statistical theory. In this sense, this work presents an important probabilistic theoretical result showing the relationship between the normal, chi-square and Student's t distributions, and using the R statistical software, it was possible to show this relationship through simulation studies with different algorithms for generating random numbers, with sample sizes of 1,000 for each distribution, and in each simulation, degrees of freedom 1, 15 and 30, thus making the ratio from a standard normal by a chi-squared with ν degrees of freedom, and through histograms and qq-plot graphs to visually show a similar behavior of the Student's t distribution.

Keywords: Probability Theory, Student's t, Simulation.

Sumário

Lista de Figuras

1 Introdução

De modo geral ao estudarmos qualquer experimento devemos procurar um modelo matemático que nos ajude a descrever de forma satisfatória o experimento apresentado. Assim necessitamos materializar uma forma matemática para os fenômenos de observação, tais modelos são de dois tipos: Determinísticos e não Determinísticos.

- Determinísticos são relativos aos experimentos que apresentam um resultado com um padrão matemático.
- Não Determinísticos (probabilístico ou estatístico) se apresentam como resultados irregulares quando analisados individualmente, ou seja, existirão desvios suficientes que podem alterar um dado comportamento de um fenômeno qualquer, mesmo que saibamos todas as possíveis respostas do experimento.

Ao realizarmos um experimento aleatório, o conjunto de todos os resultados possíveis do experimento será chamado de espaço amostral, que é denotado pela letra grega Ω, que poderá ser finito ou infinito.

O resultado obtido quando se realiza um experimento aleatório pode ser formado por um número ou um grupo de números, um atributo ou grupo de atributos, ou, ainda, por uma combinação de aspectos quantitativos e/ou qualitativos. Convém observar que tecnicamente todo subconjunto de um espaço amostral é um evento apenas quando ele for finito ou, infinito enumerável, se o espaço amostral é infinito não numerável é possível construir subconjuntos que não são eventos.

Probabilidade é baseada em informação e conhecimento. Um dos objetivos de probabilidade é saber como atualizar o valor da probabilidade quando esta base de informação ou conhecimento é alterada. Em particular, quando alteramos a probabilidade de um dado evento A quando sabe-se que um determinado evento B ocorreu é chamado de probabilidade condicional.

Embora determinados experimentos possuam espaços amostrais que não são representados por números, muitas vezes estamos interessados em alguma função do resultado, e não do resultado em si. Logo denomina-se variável aleatória uma função que associe a cada elemento A pertencente a Ω um número real $X(A)$.

Se X é uma variável aleatória e se a quantidade de valores possíveis para X for finita, ou infinita enumerável, então X, é uma variável aleatória discreta, e se X puder assumir todo e qualquer valor em algum intervalo $a \leq X \leq b$, onde a e b podem ser respectivamente $-\infty$ e $+\infty$, então X é uma variável aleatória contínua.

Na área de ciências exatas muitas vezes os profissionais fazem uso dos modelos probabilísticos para mostrar situações reais, ou para descrever um experimento aleatório. Porém certas questões não podem ser resolvidas analiticamente, entretanto, mesmo tendo um modelo probabilístico, teremos que recorrer a estudos de simulações para obter aproximações de quantidade de interesse.

Os estudos apresentados até o momento incidiram sobre a finalidade de mostrar a importância de relacionarmos algumas distribuições contínuas aplicadas no $software$ estatístico **R** visando com isso facilitar o entendimento dessas distribuições e observar o comportamento gráfico, nesse sentido, o cálculo da relação de distribuições contínuas é de fundamental interesse para os estudos estatísticos. Faremos estudos das principais distribuições, como a distribuição normal que é uma das mais importantes para análises de fenômenos reais, e também de grande importância para a Inferência Estatística e a Amostragem. Trataremos também da distribuição de qui-quadrado que nos diz em que medida é que os valores observados se desviam do valor esperado, caso as duas variáveis não estivessem associadas, e a t-Student que é uma distribuição de probabilidade teórica.

Este trabalho tem por objetivo, utilizar uma sequência de passos para mostrar o resultado teórico proposto, escolhido aleatoriamente da distribuição Normal, Qui-quadrado e t de Student através de simulação. Neste estudo, uma simulação pode ser entendida como uma particular realização de cada um desses modelos. Assim, os valores simulados podem ser considerados uma amostra aleatória de cada uma das distribuições. Portanto, com as simulações será possível mostrar gráfica e visualmente a relação entre as distribuições Normal, Qui-quadrado e t de Student.

2 Fundamentação Teórica

2.1 Experimentos não Determinísticos

Pode-se classificar um experimento não determinístico como aquele em que, quando realizado sob condições idênticas, não é possível prever, a priori, o resultado particular que irá ocorrer e sim, o conjunto dos possíveis resultados. Ou seja, é um processo de coleta de dados relativos a um fenômeno que acusa variabilidade em seus possíveis resultados, mas não se pode saber a priori qual deles ocorrerá (CUZZOL, 2014).

O experimento não determinístico (probabilístico ou estatístico) se apresenta como resultados irregulares quando analisados individualmente, ou seja, existirão desvios suficientes que podem alterar um dado comportamento de um fenômeno qualquer, mesmo que saibamos todas as possíveis respostas do experimento. Tais experimentos podem ser repetidos diversas vezes sob as mesmas condições iniciais, e mesmo assim não é possível determinar previamente o resultado.

Pode-se considerar como experimentos não determinísticos os fenômenos produzidos pelo homem, tais como:

- lançamento de uma moeda;
- lançamento de um dado;
- determinação da vida útil de um componente eletrônico;
- previsão do tempo

Dessa forma, os modelos não determinísticos, em geral, estão associados a fenômenos probabilísticos.

As características dos experimentos não determinísticos são: podem ser repetidos indefinidamente sob as mesmas condições e mesmo assim não se pode adiantar um resultado particular, mas pode-se descrever todos os resultados possíveis; se repetidos muitas vezes apresentarão uma regularidade em termos de frequência de resultados (CUZZOL, 2014).

Em síntese, pode-se dizer que experimentos não determinísticos, os resultados não serão previsíveis, mesmo que haja um grande número de repetições do mesmo fenômeno. Assim sendo, no experimento não determinístico não é possível explicitar ou definir um resultado particular, mas sim somos capazes de descrevermos o conjunto de todos os possíveis resultados.

2.2 Espaço Amostral

Espaço amostral é o conjunto de todos os possíveis resultados de um experimento aleatório, isto é, o conjunto de todos os resultados elementares, mutuamente exclusivo e coletivamente exaustivo, relativo aos resultados de um experimento. A letra que representa o espaço amostral, é Ω.

Exemplos:

- jogar uma moeda e observar a face obtida

  $\Omega = \{cara, coroa\}$

- selecionar uma peça para inspeção e observar se ela tem ou não algum defeito

  $\Omega = \{defeituosa, \textit{não defeituosa}\}$

- lançar um dado e anotar o número da face que ficou virada para cima

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

Quando o espaço amostral consiste em um número finito ou infinito enumerável de resultados do experimento, é chamado espaço amostral discreto; se consistir em todos os números reais de determinado intervalo, é um espaço amostral contínuo. Dessa forma, espaço amostral é o conjunto estabelecido por todos os possíveis resultados de um experimento.

Os pesquisadores trabalham com amostras. Independentemente se a população é infinita ou finita muito grande, mas o objetivo é "obter informações sobre o todo, baseando-se no resultado de uma amostra"(BOLFARINE; BUSSAB, 2005).

2.3 Eventos

Em teoria de probabilidades, um evento é um conjunto de resultados ao qual é associado um valor de probabilidade. Habitualmente, quando o espaço amostral é finito, qualquer subconjunto seu é um evento (i. e., todos os elementos do conjunto de partes do espaço amostral são definidos como eventos) (WIKIPEDIA, 2014).

Assim, um evento é um conjunto, classe ou grupo de possíveis resultados incertos. Eventos podem ser de dois tipos. Um evento composto pode ser decomposto em 2 ou mais sub eventos não vazios, enquanto um evento elementar não pode.

Os eventos usualmente são simbolizados por letras maiúsculas do início do alfabeto, como, A, B, C, etc.

Portanto, compreende-se que evento é um conjunto de resultados do experimento, em termos de conjuntos. Para um espaço amostral em particular, Ω e $\emptyset$ (conjunto vazio) são eventos. Ω é dito o evento certo e $\emptyset$ o evento impossível. Em síntese, evento é a representação de um subconjunto do espaço amostral.

2.4 Probabilidade

A palavra probabilidade deriva do Latim probare (provar, testar), figurativamente conhecida por sorte, azar, certeza, incerteza e risco, dependendo do contexto (GRIARSTEAD; LAURIE, 2011). Probabilidade é uma coleção ampla de conceitos que trata dos estudos de experimentos aleatórios ou não determinísticos. Probabilidade pode significar também um número num intervalo de 0 a 1, o que fornece um significado ao avaliar a ocorrência de um resultado num experimento. A probabilidade não é uma ciência exata, que não tem o intuito de obter um resultado concreto para o sucesso do evento, e sim, apresentar todos os resultados possíveis que possam ocorrer. Conforme Cassiano e Alves (2011), a prática informal da probabilidade vem dos tempos primórdios em que ocorriam jogos com ossos. Com o decorrer dos tempos foram surgindo problemas casuais e com eles a necessidade de busca de modelos não casuais para o esclarecimento e/ou justificativa dos eventos.

O estudo da probabilidade começou na Itália, quando o matemático e médico Giloramo Cordano $(1501 - 1576)$ relacionou noções elementares de probabilidade com jogos de azar. Assim, vem da necessidade de em certas situações, prever-se a possibilidade de ocorrência de determinados fatos. A probabilidade consiste num ramo da Matemática que estuda as possibilidades de um fenômeno ocorrer. Possui aplicações em algumas áreas do conhecimento humano, como Genética, Finanças, Marketing, Economia. Dito isto, compreende-se que a probabilidade deve ser vista como um conjunto de ideias e procedimentos que permitem aplicar a Matemática em questões do mundo real, quantificar e interpretar conjuntos de dados ou informações que não podem ser quantificados direta ou exatamente (BRASIL, 2002).

Segundo Costa e Hurtado (2005), a probabilidade pode se constituir em um poderoso instrumento social, na medida em que pode permitir ao estudante uma melhor compreensão das estatísticas oficiais, tornando-o capacitado a exercer mais conscienciosamente sua cidadania. O cálculo de uma probabilidade para ser aplicado, deve ser iniciado pelo conhecimento de todos os dados que envolvem o fenômeno a se testar ou provar, sua finalidade e, a partir daí, selecionar dentre os vários modelos, os adequados para o fenômeno distinto. Segundo Meyer (1983), todas as vezes que emprega-se a Matemática com o objetivo de estudar alguns fenômenos de observação:

> Deveremos essencialmente começar por construir um modelo matemático (determinístico ou probabilístico) para esses fenômenos. Inevitavelmente, o modelo deve simplificar as coisas e certos pormenores devem ser desprezados. O bom resultado do modelo depende de que os pormenores desprezados sejam ou não realmente sem importância na elucidação do fenômeno estudado. A resolução do problema matemático pode estar correta e, não obstante, estar em grande discordância com os dados observados, simplesmente porque as hipóteses básicas feitas não sejam confirmadas. Geralmente é bastante difícil afirmar com certeza se um modelo matemático especificado é ou não adequado, antes que alguns dados de observação sejam obtidos. A fim de verificar a validade de um modelo, deveremos deduzir um certo número de consequências de nosso modelo e, a seguir, comparar esses resultados previstos com observações.

Nesta perspectiva, o estudo da probabilidade é essencial para termos entendimento na leitura dos fenômenos apresentados que ocorrem no dia-a-dia. Na escola básica, ao se estudar probabilidades o que ocorre com frequência é buscar a compreensão desses eventos casuais ou não e, principalmente, conseguir decifrar os gráficos estatísticos vivenciados no dia-a-dia.

Os Parâmetros Curriculares Nacionais, Brasil (2002), estabelecem como principal finalidade para o estudo de probabilidade que o educando compreenda que grande parte dos acontecimentos do cotidiano é de natureza aleatória e que é possível identificar prováveis resultados desses acontecimentos. Nesta perspectiva, as noções de acaso e incerteza, que se manifestam intuitivamente, podem ser exploradas na escola, em situações nas quais o aluno realiza experimentos e observa eventos. Dessa forma, o termo probabilidade é aplicado de modo muito amplo no nosso dia-a-dia para propor um certo nível de incerteza sobre o que ocorreu no passado, ocorrerá no futuro ou o que está ocorrendo no presente. Portanto, probabilidades são utilizadas em situações em que dois ou mais resultados diferentes podem ocorrer, mas não é possível saber antecipadamente qual deles realmente acontecerá.

Quando o espaço amostral e finito com elementos equiprováveis a probabilidade é determinada com base na proporção de vezes que ocorre um resultado favorável em um certo número de observações ou experimentos (LEITE, 2007). Diante do exposto, conclui-se que a probabilidade pode ser pensada como a teoria matemática utilizada para estudar a incerteza oriunda de fenômenos que envolvem o acaso.

A probabilidade de um acontecimento, associada a certa experiência aleatória, é a frequência relativa esperada desse acontecimento, ou seja, o quociente entre o número de vezes que o acontecimento se realiza ao fim de n repetições. Logo esta definição de probabilidade é muitas vezes usada em experiências de interesse científico em que as probabilidades são calculadas à posteriori a partir das frequências relativas, do acontecimento em estudo, num número de provas consideráveis.

2.4.1 Lei de Laplace

Nas situações em que os vários resultados elementares possíveis são equiprováveis (têm todos a mesma probabilidade) podemos calcular a probabilidade de um acontecimento desse espaço amostral através da Lei de Laplace (SOARES, 2010).

A probabilidade de um acontecimento, associada a uma certa experiência aleatória, é dada pelo quociente entre o número de casos favoráveis ao acontecimento e o número de caso possíveis. Representado da seguinte forma, se A é um evento associado a uma certa experiência aleatória, cujo espaço amostral é Ω, tendo-se $A \subseteq \Omega$, então a sua probabilidade de ocorrência pode ser denotada por $p(A)$ e calculada por,

$$p(A) = \frac{A}{n},$$

onde A é o números de casos favoráveis a A e n é o números de casos possíveis.

Segundo Meyer (1983), para que o número que convencionarmos tenha significado, qualquer experimentação subsequente deverá produzir uma frequência relativa que seja “próxima”do valor convencionado, particularmente se o número de repetições, no qual a frequência relativa calculada se tenha baseado, for muito grande.

Considera-se uma experiência aleatória e seja Ω o espaço amostral dessa experiência. Seja $p(A)$ a probabilidade do acontecimento de A, que satisfaça os seguintes axiomas de Kolmogorov que são um conjunto de propriedades que definem que tipo de funções matemáticas podem ser adotadas para descrever um modelo probabílistico. Segundo Rego (2012), os primeiros quatros axiomas podem ser motivados pelas propriedades de frequência relativa.

- Inicial é quando o experimento aleatório é descrito pelo espaço de probabilidade, que consiste do espaço amostral Ω, de uma coleçao A de eventos de Ω e de uma função de valores reais $X : A \to \mathbb{R}$
- Não-negatividade. $\forall A \in \Omega, P(A) \geq 0$.
- Normalização Unitária. $P(\Omega) = 1$.
- Aditividade finita: Seja $A_i, i = 1, 2, ..., n$, uma coleção finita de eventos disjuntos par a par. Então , $P(\cup_{i=1}^{n} Ai) = \sum_{i=1}^{n} P(Ai)$.

  Um último axioma foi proposto por Kolmogorov para garantir um certo grau de continuidade da medida de probabilidade.
- σ -aditividade. Se A_i é uma coleção enumerável de eventos disjuntos dois a dois então:$P(\cup_{i=1}^{\infty} A_i) = \displaystyle\sum_{i=1}^{\infty} P(A_i)$.

Podemos ainda concluir outras propriedades desses axiomas:

$$P(A^c) = 1 - P(A),$$

sendo A^c o evento complementar de A;

Se A,B dois acontecimentos de Ω tais que $A \subseteq B$ entao:

$$P(A) \leq p(B) = P(B - A) = P(B) - P(A),$$

se A, B são eventos quaisquer de Ω, então:

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

2.4.2 Probabilidade condicional

Seja Ω o espaço de probabilidade para um determinado experimento aleatório, suponhamos que existe a princípio alguma informação a respeito do resultado do experimento aleatório.

De acordo com Bussab e Morettin (2002) para dois eventos quaisquer A e B, sendo $P(B) > 0$, definimos a probabilidade condicional de A dado B, $P(A|B)$, como sendo:

$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$, portanto é chamada a probabilidade condicional de A dado B.

Isso significa que a probabilidade de A ocorrer, dado que B ocorreu, é igual à probabilidade de ocorrência simultânea de A e B dividida pela probabilidade de ocorrência de B.

2.4.3 Independência

Segundo Correa (2003), um evento A é considerado independente de outro evento B se a probabilidade de A é igual à probabilidade condicional de A dado B.

Frequentemente, um experimento aleatório consiste em realizar um sequência de ensaios, por exemplo se o experimento aleatório é lançar uma moeda repetidamente, cada lançamento pode ser considerado como um ensaio. Neste caso dizer que os ensaios são independentes significa dizer que as seguintes condições são válidas: Se A_i é um evento cuja ocorrência é completamente determinada pelo resultado do i-ésimo ensaio, então A_1, A_2,... são independentes.

Observe-se que se A e B são independentes, tem-se:

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

Usando o exemplo do lançamento do dado e uma moeda, ambos não-viciados, para exemplificar com mais precisão o conceito de independência estatística, temos:

ϵ: Lançamento simultâneo de um dado e uma moeda, ambos não-viciados.

$$\Omega = \{(CA, 1), (CA, 2), (CA, 3), (CA, 4), (CA, 5), (CA, 6,)$$
$$(CO, 1), (CO, 2), (CO, 3), (CO, 4), (CO, 5), (CO, 6)\}$$

Sejam os eventos:

$A = \{(6, CA), (6, CO)\}$: resultado 6

$B = \{(2, CA), (2, CO), (4, CA), (4, CO), (6, CA), (6, CO)\}$: resultado par

$C = \{(CO, 1), (CO, 2), (CO, 3), (CO, 4), (CO, 5), (CO, 6)\}$: resultado coroa

Como o espaço amostral é finito e com elementos equiprováveis, tem-se:

$$P(A) = \frac{n(A)}{n(\Omega)} = \frac{2}{12} = \frac{1}{6}$$

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{1}{3}$$

$\Rightarrow A$ e B *são dependentes.*

$$P(A \mid C) = \frac{P(A \cap C)}{P(C)} = \frac{1}{6}$$

$\Rightarrow A$ e C *são independentes.*

Logo podemos dizer que os eventos A, B e C, são independentes se e somente se:

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

$$P(A \cap C) = P(A)P(C,),$$

$$P(B \cap C) = P(B)P(C),$$

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

De acordo com Bussab e Morettin (2002) se apenas as três primeiras relações estiverem satisfeitas, dizemos que os eventos A, B e C são mutuamente independentes. É possível que três eventos sejam mutuamente independentes, mas não sejam completamente independentes. A definição pode ser estendida facilmente para um número infito qualquer de eventos.

2.5 Variáveis aleatórias

Uma variável aleatória é uma função que atribui um valor a cada resultado individual de uma experiência aleatória. Entende-se como uma variável quantitativa, onde o resultado (*valor*) depende de fatores aleatórios.

Logo a variável aleatória é uma função que associa elementos do espaço amostral a valores numéricos. Mais precisamente, é uma função (mensurável) $X : \Omega \to \mathbb{R}$ que associa um número real a cada resultado de um experimento aleatório.

Isso equivale a descrever os resultados de um experimento aleatório por meio de número ao invés de palavras, o que é uma grande vantagem pois possibilita melhor tratamento matemático, inclusive através de parâmetros.

2.5.1 Tipos de variáveis aleatórias

Na grande maioria dos problemas práticos, encontramos dois tipos de variáveis aleatórias, as discretas e as contínuas.

Discreta

Para Portnoi (2005), uma variável aleátoria é discreta se todos os seus valores podem ser listados, e estes valores pertencem a um conjunto finito ou infinito, enumerável.

Seja X uma variável aleatória discreta, a função de probabilidade de X é uma função f_X que associa a cada valor possível x de X a sua probabilidade:

$$f_X(x) = P(X = x)$$

Estas funções tem as seguintes propriedades:

- $0 \leq f_X(x) \leq 1, \forall x \in \mathbb{R}$
- $\sum_x f_X(x) = P(\Omega) = 1$

Ao lançar um dado honesto, sempre nos dará um valor inteiro, não exite possibilidade que ele caia de lado ou nos dando um valor decimal, portanto é a melhor maneira de exemplificar o conceito de variável aleatória discreta finita e associarmos a um exemplo real.

Já para variável aleatória discreta infinita, tenhamos o número de carros que chegam em determinado pedágio, partindo do pressuposto que sabemos que virão infinitos carros, logo não existirão fração no número de carros, o resultado não é completamente conhecido, mais sempre descritível.

Valor médio esperado de uma Variável Aleatória discreta

A esperança de uma variável aleatória nada mais é que o valor médio esperado da variável. Por este motivo a esperança é usualmente denominada de valor esperado. Isto é se X é uma variável aleatória discreta, então:

$$E(X) = \sum_{x} x.P(x) = \mu_1$$

Variância de uma variável aleatória discreta

A variância de uma variável aleatória discreta, é a medida que dá o grau de dispersão (ou de concentração) de probabilidade em torno da média. O fato de conhecermos a média de uma distribuição de probabilidade já nos ajuda bastante, porém, precisamos de uma medida que nos dê o grau de dispersão de probabilidade em torno dessa média. Define-se a variância de X, σ_X^2 ou $Var(X)$ então:

$$Var(X) = E[X^2] - [E(X)]^2 = \mu_2 - \mu_1^2$$

Alguns dos modelos mais usuais para variaveis aleatórias discretas seguem a seguir, porém, existem ainda muitos outros modelos, entretanto, estes são os mais conhecidos:

- Distribuição de Bernoulli
- Distribuição binomial
- Distribuição binomial negativa
- Distribuição geométrica
- Distribuição hipergeométrica
- Distribuição de Poisson
- Distribuição uniforme

Contínua

Uma variável aleatória, cujos valores são expressos em uma escala contínua, é chamada de variável aleatória contínua (CORREA, 2003).

Ao estudarmos as variáveis aleatórias, foi observado que os valores assumiam apenas um número finito (ou infinito enumerável). Portanto uma variável aleatória é contínua se os valores não podem ser listados, mais podem assumir um número infinito de valores em um intervalo finito ou infinito. Sejam:

X: Duração de uma conversa telefônica;

Y: O tempo de vida de uma lâmpada.

Nestes exemplos, os possíveis valores das variáveis aleatórias em estudo, não podem ser contados (ou enumerados), ou muito dificilmente são caracterizadas por um número finito e enumerável de valores. Nestes casos é razoável considerar que as variáveis aleatórias associadas, X e Y, tomam um número indeterminado de valores, por exemplo, nos intervalos $[x_1, x_2]$ e $[y_1, y_2]$, respectivamente.

Desta maneira, para estes tipos de variáveis, que podem tomar todos os valores num determinado intervalo de números reais, vamos ter de considerar distribuições de probabilidades diferentes das dadas nas variáveis aleatórias discretas. Mais concretamente, iremos descrever a probabilidade de uma variável aleatória X assumir um valor no intervalo $[a, b]$ através de uma função f oportunamente escolhida, como sendo a área limitada pelo eixo $[a, b]$ e pelo gráfico da função f, dado na Figura 1.

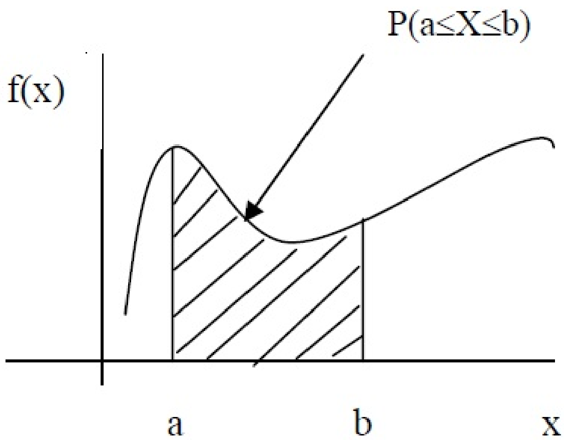

Figura 1: Gráfico da função f

De acordo com Wikipedia (2013), tecnicamente, as variáveis aleatórias contínuas $v.a.c$ são variáveis para qual o conjunto A é um conjunto infinito não enumerável, ou seja, é uma variável que assume valores dentro de intervalos de números reais.

Em muitos fenômenos que analisamos na prática, as variáveis aleatórias envolvidas têm natureza contínua podendo assumir uma quantidade não enumerável de valores. Logo uma variável aleatória X diz-se (absolutamente) contínua se existir um função $f_X(x)$ não negativa ($f_X(x) \geq 0, \forall x \in \mathbb{R}$) e integrável. Uma função $f_X \geq 0$ é densidade de alguma variável aleatória se somente se;

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

Além das funções de densidade f_X, também é possível se definir uma nova função F, dada por:

$$F(x) = \int_{-\infty}^{x} f_X(t)dt,$$

que é a função de distribuição acumulada. Logo, a distribuição de uma variável aleatória contínua X pode ser determinada tanto pela função de distribuição acumulada Fx ou pela sua função de densidade fx.

Portanto, para qualquer variável contínua X, sua função de distribuição acumulada é contínua, o que implica que a probabilidade de ela assumir qualquer número real a é igual a 0. Então:

$$P(X = a) = F_X(a) - F_X(a^-) = \int_{-\infty}^{a} fx(t)dt - \int_{-\infty}^{a-} fx(t)dt = 0.$$

Para melhor compreender este tipo de variável aleatória considere a probabilidade de escolher um número real entre 0 e 1 considerando que todos os valores têm a mesma chance de serem escolhidos, neste caso só faz sentido falar na probabilidade do número escolhido pertencer a um determinado subintervalo. Em geral, se a e b são números reais tais que $a < b$, tem-se no caso que X é uma variável aleatória contínua;

$$P(a < X \leq b) = \int_{a}^{b} f(t)dt.$$

Esperança de uma variável aleatória contínua

Se X é uma variável aleatória contínua com função de densidade de probabilidade f, então sua esperança é dada pela formúla:

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx,$$

desde que a integral esteja bem definida.

Portanto $E(X)$ está definida desde que a integral acima convirja para algum valor finito. Em caso contrário, dizemos que $E(X)$ não existe (ou que X não tem valor esperado.) (LEBENSZTAYN; COLETTI, 2008)

O valor médio de uma variável aleatória, possui algumas propriedades. De maneira análoga tanto *v.a.d.* quanto para *v.a.c.* Segundo Soares (2010) O valor esperado de uma variável aleatória não é necessariamente um valor dessa variável assim como o centro de massa de um corpo pode não pertencer ao próprio corpo.

- $E(c) = c, \forall c \in \mathbb{R}$
- $E[aX + b] = aE[X] + b, \forall (a, b) \in \mathbb{R}^2$ (operador linear)

Variância de uma variável aleatória contínua

A variância de uma variável aleatória contínua X com função densidade de probabilidade$f$ é dada pela fórmula:

$$Var(x) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x)dx.$$

A variância de uma variável aleatórias tem algumas propriedades a serem seguidas, sendo elas:

- $Var[X] \geq 0$
- $Var[c] = 0, \forall c \in \mathbb{R}$
- $Var[X] = 0 \Leftrightarrow P(X = E[X]) = 1$
- $Var[aX + b] = a^2 Var[X], \forall (a, b) \in \mathbb{R}^2$
- $Var[X] = E[X^2] - [E(X)]^2 \Leftrightarrow M_2 = \mu_2 - \mu_1^2$

Vamos apresentar alguns dos modelos mais usuais para variáveis aleatórias contínuas, mas existem ainda muitos outros modelos, entretanto, estes são os mais conhecidos.

- Uniforme contínua
- Normal
- Exponencial
- Gama

- Qui-quadrado
- Beta
- Weibull
- t de Student
- F de Fisher-Snedcor

Este trabalho foi desenvolvido para demonstrar a relação de alguns modelos probabilísticos das variáveis aleatórias contínuas, logo adentraremos nas distribuições de probabilidade sendo elas Normal, Qui-quadrado e t de Student.

Método do Jacobiano

Segundo Barros (2008), se uma variável aleatória contínua definida num intervalo (a, b), com densidade $f(x)$ e função de distribuição $F(x)$, e se $Y = h(X)$ onde $h(.)$ é uma função contínua e injetora ou seja, cada x é levado num y diferente, então a densidade de $y, g(y)$, pode ser encontrada da seguinte maneira:

$$g(y) = f(x)\left|\frac{dx}{dy}\right|$$

O módulo $|dx/dy|$ aparece na fórmula anterior para garantir que $g(y)$ seja sempre ≥ 0, pois dx/dy pode ser negativo.

Onde, x na expressão anterior está escrito em função de y, ou seja, a variável "velha" está em função da variável "nova". E também podemos observar que, $x = h^{-1}(y)$ é expresso em termos da "nova" variável y, e que $h(.)$ é uma função crescente, isto é, $x_1 \leq x_2$ implica em $h(x_1) \leq h(x_2)$, onde serão usadas as variáveis aleatórias multidimensionais.

Escrevemos a transformada inversa como $X = h^{-1}(Y)$ e definimos os jacobianos: (ROLLA, 2013)

$$J_h(y) = det\left(\frac{\partial x}{\partial y}\right) = det\begin{bmatrix} \frac{\partial_{x1}}{\partial_{y1}} & \cdots & \frac{\partial_{x1}}{\partial y_d} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_d}{\partial_{y1}} & \cdots & \frac{\partial x_d}{\partial y_d} \end{bmatrix}$$

e

$$J_g(x) = det\left(\frac{\partial y}{\partial x}\right) = det\begin{bmatrix} \frac{\partial_{y1}}{\partial_{x1}} & \cdots & \frac{\partial_{y1}}{\partial x_d} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_d}{\partial_{x1}} & \cdots & \frac{\partial y_d}{\partial x_d} \end{bmatrix}$$

O jacobiano satisfaz a seguinte identidade:

$$J_h(y) = \frac{1}{J_g(x)}$$

Se X é um vetor aleatório absolutamente contínuo onde $Y = g(X)$, então a densidade f_Y pode ser obtidoa a partir da densidade f_X pela relação:

$$f_{Y(y)} = |J_h(y)|.f(x)(h(y)) = \frac{1}{|J_g(x)|}f_X(h(y))$$

Este resultado pode ser generalizado para uma função de várias variáveis aleatórias.

2.5.2 A distribuição Normal

Segundo Meyer (1983), a distribuição normal (ou gaussiana) é uma das mais importante distribuições de variáveis aleatórias contínua.

De um modo geral, a maior parte dos fenômenos probabilísticos são de natureza contínua, e mesmo alguns de natureza discreta, tendem a seguir uma lei de distribuição designada por função de distribuição Normal, ou Gaussiana.

A função de distribuição Normal estabelece que os valores mais frequentes (isto é, os valores a que correspondem as maiores probabilidades) se encontram em torno da média (estimativa do valor médio ou esperança matemática) da variável aleatória; quanto mais afastados os valores estão da média onde este afastamento é quantificado em termos de variância ou segundo momento em relação a média, quer acima quer abaixo desta, menos frequentes são. Esta interpretação imediata da distribuição Normal é coerente com o que se passa com a maior parte dos fenômenos que ocorrem na natureza.

Dizemos que X tem uma distribuição *Normal (ou Gaussiana)* com parâmetros μ e $\sigma^2 > 0$, que são números reais, se a função densidade de X for igual a

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma^2}exp\left(-\frac{1}{2}\left[\frac{x-\mu}{\sigma^2}\right]^2\right), \tag{2.1}$$

Os parâmetros μ e σ^2 devem satisfazer às condições $-\infty < \mu < \infty, \sigma^2 > 0$. Já que teremos muitas ocasiões de nos referir à distribuição acima, empregaremos a seguinte notação: $X \sim N(\mu, \sigma^2)$ se, e somente se, sua distribuição de probabilidade for dada pela Eq (2.1)

Historicamente, esta distribuição foi chamada de Normal porque ela era amplamente aplicada em fenômenos biológicos e sociais que era sempre tida como a distribuição antecipada ou normal. Aplicações da distribuição normal incluem variabilidade em parâmetros de componentes manufaturados e de organismos biológicos por exemplo, altura, peso, inteligência. Pode parecer estranho, modelar quantidades que só assumem valores positivos por uma distribuição normal onde valores negativos aparecem. Nestes casos o que ocorre é que os parâmetros μ e σ^2 devem ser escolhidos de modo que a probabilidade da variável assumir um valor negativo seja aproximadamente nula de modo que a representação seja válida.

Vamos examinar o aspecto do gráfico de f. Ele apresenta a bem conhecida forma de sino, mostrada na Figura 2. Visto que f depende de x somente através da expressão $(x-\mu)^2$, torna-se evidente que o gráfico de f será *simétrico* em relação a μ. Por exemplo, se $x = \mu+2$, $(x-\mu)^2 = (\mu+2-\mu)^2 = 4$, enquanto para $x = \mu-2$, $(x-\mu)^2 = (\mu-2-\mu)^2 = 4$, também. (MEYER, 1983)

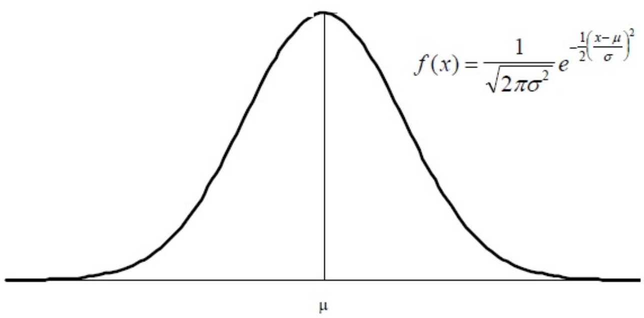

Figura 2: Gráfico da função de densidade de uma variável aleatória normal

A distribuição normal (ou gaussiana), têm como características:

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

Quando temos em mãos uma variável aleatória com distribuição normal, nosso principal interesse é obter a probabilidade dessa variável aleatória assumir um valor em um determinado intervalo.

Pela propriedade de funções de densidade, para calcular $P(a < X < b)$, deveríamos solucionar a seguinte integral:

Suponha-se que X tenha distribuição $N(0,1)$. Nesse caso,

$$P(a < X < b) = \int_a^b f_x(x)dx = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2}dx.$$

Para calcular esta integral toda vez, não seria fácil.

A fim de ultrapassar este inconveiente, Gauss (um dos estatísticos que inicialmente estudou essa função de distribuição) desenvolveu uma metodologia que conduz a um padrão estabelecido, ou redução a um caso único, de qualquer que seja a função de distribuição normal, caracterizada por μ, σ^2. Este padrão a ser estabelecido transforma qualquer função de distribuição normal $N(\mu, \sigma^2)$ numa única função de distribuição normal caracterizada por média $\mu = 0$ e variância $\sigma^2 = 1$, logo $N(0,1)$, modifica a função a ser integrada, tornando inviável a tarefa de se fazer a avaliação numérica de cada uma dessas integrais. Fazemos uso então do seguinte fato: se $X \sim N(\mu, \sigma^2)$, então

$$Z = \frac{X - \mu}{\sigma} \sim N(0,1),$$

logo Z tem distribuição normal padrão ou normal reduzida, basta reduzi-la à $N(0,1)$ e, então, consultar as probabilidades desejadas nessas tabelas.

2.5.3 A distribuição Qui-quadrado

Segundo Meyer (1983). Um caso especial importante do modelo gama é obtido fazendo-se $\alpha = \frac{\nu}{2}$ e $\beta = 2$, com $\nu > 0$ inteiro.

A distribuição qui-quadrado pode ser interpretada de duas formas, como um caso particular da distribuição gama, ou como sendo a soma de normais padronizada ao quadrado, logo:

$$X_i \sim N(0,1),$$

então,

$$\sum_{j=1}^{r} X_j^2 = \chi^2_{(r)}.$$

De acordo com Bussab e Morettin (2002), o quadrado de uma variável aleatória com distribuição normal padrão é uma variável aleátoria com distribuição qui-quadrado com $1g.l.$ Ou seja, uma variável aleatória $\chi^2_{(\nu)}$ pode ser vista como soma de variáveis aleatórias independentes com distribuição normal padrão ao quadrado.

Dizemos que uma variável aleatória X contínua de distribuição do qui-quadrado com $\nu \in \mathbb{N}$ graus de liberdade $X \sim \chi^2_{(\nu)}$ se a sua função de densidade de probabilidade $(f.d.p.)$ for da forma:

$$f(x) = \frac{e^{-x/2}x^{\nu/2-1}}{2^{\nu/2}\Gamma(\nu/2)}, \nu > 0, x > 0$$

Seja X um variável aleatória contínua com distribuição qui-quadrado (χ^2) com n graus de liberdade. Graficamente, a distribição χ^2 pode ser representada na Figura 3.

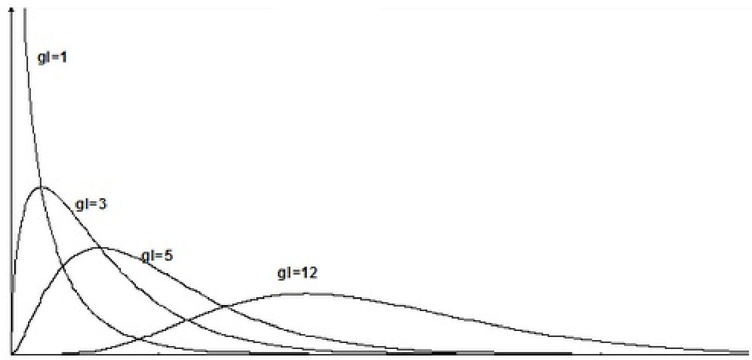

Figura 3: Gráfico da distribuição qui-quadrado $\chi^2(\nu)$

A distribuição qui-quadrado é uma estrutura probabilística adequada para variáveis qualitativas com duas ou mais categorias. A figura acima apresenta a densidade de modelo χ^2, com a região crítica (RC) do teste, isto é, RC $= \{X > \chi^2_c\}$.

A distribuição χ^2, tem alguma características a ser observadas:

- Curva positiva e não simétrica;
- Se $X \sim \chi^2_{(\nu)}$ então:

  $E(X) = \nu$

  $Var(X) = 2\nu$
- A soma de funções de distribuição qui-quadrado tem distribuição do qui-quadrado com graus de liberdade igual à soma dos graus de liberdade dessas distribuições, isto é:

$$X_i \sim \chi^2_{(\nu i)} \Rightarrow \sum X_i \sim \chi^2_{(\sum \nu i)}$$

- O seu aspecto gráfico depende do parâmetro n.

2.6 Distribuição t de Student

A distribuiçao t de Student é uma das distribuições mais utilizadas na estatística, com aplicações que vão desde a modelagem estatística até testes de hipóteses.

Uma variável aleatória contínua X tem distribuição t de Student com ν graus de liberdade se sua função densidade de probabilidade é dada por:

$$f(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}\left(1+\frac{x^2}{\nu}\right)^{-(\nu+1)/2}, \quad -\infty < x < \infty,$$

e utilizamos a notação $X \sim t_{(\nu)}$.

A distribuição t de Student foi descoberta por William S. Gosset em 1908. Esta distribuição surgiu na seguinte situação: Suponhamos uma amostra $X_1, ..., X_n$ $i.i.d$ de uma distribuição normal com média μ e desvio padrão σ. Seja,

$$\overline{X} = \frac{\sum_{i=1}^{n} Xi}{n} \quad \text{e} \quad s^2 = \frac{\sum_{i=1}^{n}(Xi - \overline{X})^2}{n-1},$$

a média amostral e o desvio padrão amostral, então pode-se provar que;

$$t = \frac{\overline{X} - \mu}{\sqrt{\frac{s^2}{n-1}}},$$

tem distribuição t, onde s^2 é a variância amostral com $n-1$ graus de liberdade. Este fato é decorrente do teorema a seguir.

Teorema: Obtenção da t de Student apartir do quociente da normal padrão por raíz quadrada de uma qui-quadrado com ν graus de liberdade, considerando Y e Z duas variáveis aleatórias independentes tal que $Y \sim N(0,1)$ e $Z \sim \chi^2_\nu$. Definindo X como sendo uma variável aleatória de tal forma que:

$$X = \frac{Y}{\sqrt{Z/\nu}},$$

a variável aleatória X tem distribuição t de Student com ν graus de liberdade.

Demonstração: De acordo Action (2014) a função densidade de probabilidade conjunta de F e Z é dada por:

$$f_{F,Z}(f,z) = \frac{1}{\sqrt{2\pi}}\frac{1}{\Gamma(\nu/2)}\left(\frac{1}{2}\right)^{\nu/2} u^{(\nu/2)^{-1}}e^{-z/2}e^{f^2/2}\mathbb{I}_{(0,\infty)}(z)$$

considerando a transformação

$$X = \frac{Z}{\sqrt{Z/\nu}} \text{ e } Y = Z$$

o jacobiano é $\sqrt{y/\nu}$ e então

$$f_{X,Y}(x,y) = \frac{y}{\nu}\frac{1}{\sqrt{2\pi}}\frac{1}{\Gamma(\nu/2)}\left(\frac{1}{2}\right)^{\nu/2} y^{(\nu/2)-1}e^{-y/2}e^{-x^2y/2\nu}\mathbb{I}_{(0,\infty)(y)}$$

e

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy = \frac{1}{\sqrt{2\nu\pi}}\frac{1}{\Gamma(\nu/2)}\left(\frac{1}{2}\right)^{\nu/2}\int_0^{\infty} y^{\nu/2-1+1/2}e^{-1/2(1+x^2/\nu)y}dy$$

Ao fazermos a mudança de variável

$$w = y\frac{1}{2}\left(1+\frac{x^2}{\nu}\right),$$

obtemos;

$$f_X(x) = \frac{\Gamma[(\nu+1)/2]}{\Gamma(\nu/2)}\frac{1}{\sqrt{\nu\pi}}\frac{1}{(1+x^2/\nu)^{(\nu+1)/2}}$$

que é a função densidade de probabilidade de uma distribuição t com ν graus de liberdade.

A distribuição t de Student é essencialmente uma distribuição normal com forma aproximada de um sino para todas as amostras de tamanho n onde veremos na Figura 4. É simétrica, campaniforme e semelhante a curva normal, porém com caldas mais largas, ou seja, uma simulação da t de Student pode gerar valores mais extremos que uma simulação da normal.

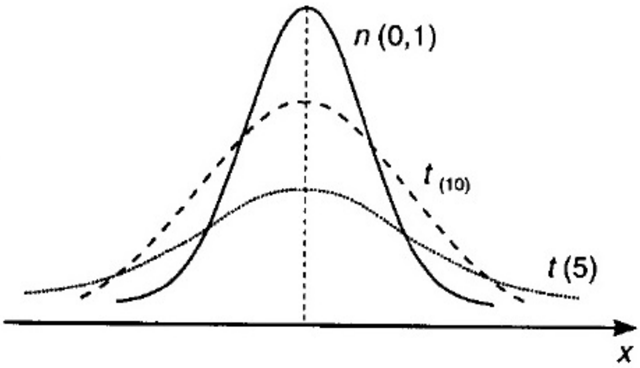

Figura 4: Gráfico da função de distribuição t de Student.

A distribuição t de Student tem como propriedades:

$$E(X) = 0, \text{ para } \nu > 1$$

$$Var(X) = \frac{\nu}{\nu-2}, \text{ para } \nu > 2$$

Podemos observar ainda as suas principais características, sendo elas:

- Cada número de grau de liberdade da origem a uma distribuição t diferente.
- A função densidade tem a mesma forma de sino da distribuição normal, mais reflete uma maior variabilidade (com curvas mais alargadas) que é de se esperar em amostras pequenas.
- A distribuição t de Student se aproxima da normal, quando aumentamos o número de graus de liberdade.
- A curva é simétrica em torno do zero.

Ao contrário da distribuição normal, não existe uma relação entre as diferentes distribuições t, assim seria necessária uma tabela para cada valor de ν.

Utilizamos a distribuição t de Student quando o desvio padrão σ não é conhecido, mas a amostra é pequena, isto é, n é pequeno, pouco se sabe sobre a distribuição da estatística Z, no entanto, se adicionarmos a condição de a amostra ter sido retirada de uma distribuição Normal, então definimos a estatítica t. Onde foi visto que a distribuição t de Student trata-se de um modelo derivado do quociente da distribuição Normal sobre a raiz de uma qui-quadrado com ν graus de liberdade.

Para ilustrar graficamente a relação demonstrada acima entre as distribuições normal, qui-quadrado e t de Student, usamos simulação com diferentes graus de liberdade (1, 15 e 30) para a qui-quadrado, utilizando o $software$ **R**, esclarecendo assim como se comportão as distribuições.

3 Aplicação

Com o propósito de mostrar a relação entre as distribuições Normal, Qui-quadrado e t de Student, executamos algumas simulações computacionais. As simulações foram realizadas com o uso do *software* estatístico **R** *version* 3.0.2, (2013-09-25) *interface* **RStudio** *version* 0.98.507. Usando o *software* estatístico, pode-se gerar valores aleatórios das distribuições Normal padrão e Qui-quadrado, visando mostrar a construção da distribuição t de Student demonstrada no teorema do final da seção anterior, alterando os graus de liberdade.

Os resultados apresentados seguem a formulação proposta nos objetivos do trabalho. Em relação a distribuição t de Student, a título meramente informativo, temos que, se duas variáveis aleatórias independentes tal que $Y \sim N(0,1)$ e $Z \sim \chi^2_\nu$, onde Y é uma variável normal padrão e Z e uma distribuição qui-quadrado com ν graus de liberdade, então a variável aleatória tem distribuição t de Student com ν graus de liberdade.

De início será gerada uma mostra aleatória com 1000 (mil) observações, para ambas distribuições, e será mudado apenas os graus de liberdade da distribuição Qui-quadrado. Logo faremos a divisão de uma variável normal padrão, pela raiz quadrada de uma distribuição qui-quadrado com ν graus de liberdade, onde mostrará a relação de distribuição entre a Normal, Qui-quadrado e t de Student.

Na primeira simulação foi gerada uma variável aleatória $Y \sim N(0,1)$, onde foi feito o histograma e traçada a curva da distribuição, como podemos ver à esquerda da Figura 5.

```
y<-rnorm(1000)
hist(y,prob=T)
curve(dnorm(x),-4,4,col=2,add=T)
```

Foi gerada uma variável aleatória $Z \sim \chi^2_{(1)}$, onde foi feito o histograma e traçada a curva da distribuição, como podemos ver ao centro da Figura 5.

```
gl=1
```

```
z<-rchisq(1000,gl)
hist(z,prob=TRUE)
curve(dchisq(x, df=gl),0,25,col=2,add=T)
```

Portanto veremos agora como se comporta o quociente da variável aleatória $Y \sim N(0,1)$ por raiz quadrada da $Z \sim \chi^2_1$, logo ficará claro a diferença de comportamento das variáveis que seguem as distribuções Normal e a t de Sudent, a direita na Figura 5.

```
x<-y/sqrt(z/gl)
hist(x,prob=TRUE)
curve(dt(x, df=gl),-100,100,col=2,add=T)
```

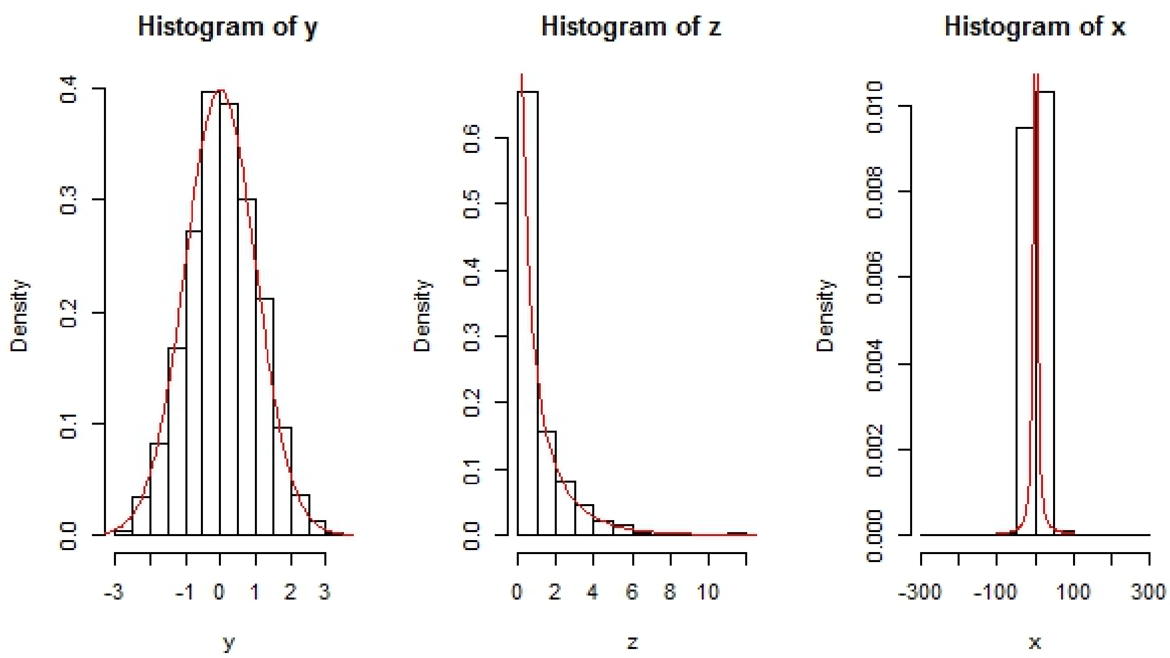

Figura 5: Gráficos da relação de distribuição de probabilidade com 1 grau de liberdade.

Temos outra maneira para determinar se dois conjuntos de dados pertencem a mesma distribuição de probabilidade, o gráfico *qq-plot*, em tais gráficos os pontos são formados pelos quantis amostrais e se no resultado os pontos alinham-se numa reta de inclinação 1, as distribuições das duas amostras podem ser consideradas as mesmas.

O gráfico *qq-plot*, é usado para se comparar os quantis de duas distribuições, que podem ser amostrais ou teóricas.

Como veremos a seguir, de forma visual para 1 grau de liberdade, se todos os pontos plotados estiverem próximos à reta, a variável X se aproxima de uma distribuição Normal, como apresentado a seguir. Porém é visível que quando comparamos os quantis do conjunto de dados gerados com a distribuição Normal padrão com 1 grau de liberdade, o *qq-plot* nos indica uma particularidade que em torno do 0 seguem sim um distribuição Normal, mas quando partimos para os pontos fora da reta, nos indica um distaciamento da distribuição Normal, como podemos observar à esquerda da Figura 6.

Logo a direita da Figura 6, visivelmente claro que ao compararmos os quantis do conjunto de dados gerados com a da distribuição t de Student, fica bem definido que o conjunto de dados dos quantis teóricos segue sim um distribuição conhecida neste caso a distribuição t de Student.

```
quantis <- qnorm(ppoints(length(y)))
qqplot(quantis, x)
abline(0,1,col=2)
quantis <- qt(ppoints(length(x)),gl)
qqplot(quantis, x)
abline(0,1,col=2)
```

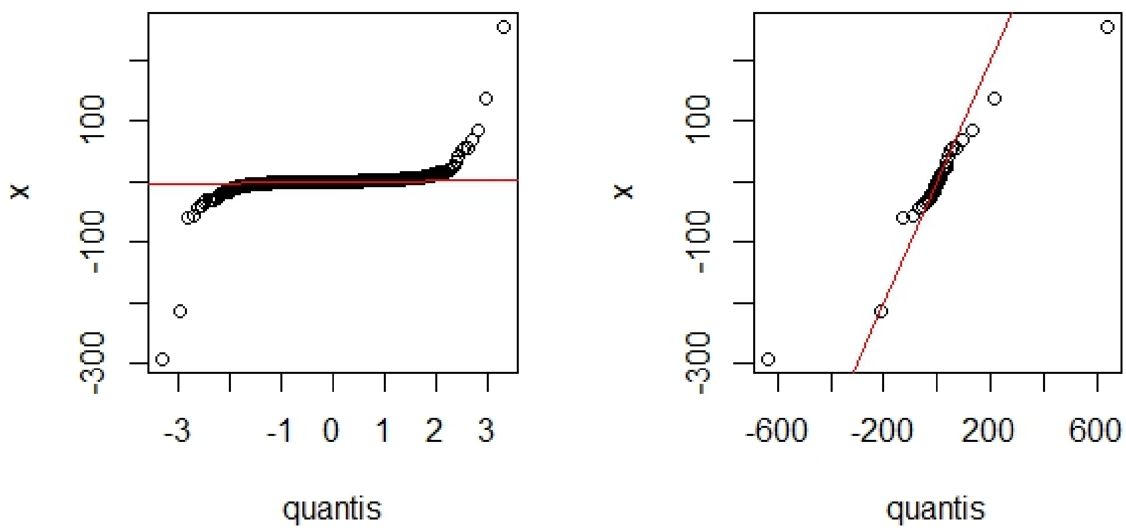

Figura 6: Gráfico *qq-plot* da distribuição Normal com a distribuição t de Student com 1 grau de liberdade.

Na segunda simulação temos uma amostra com 1000 observações para ambas as distribuições onde $Y \sim N(0,1)$ à esquerda, $Z \sim \chi^2_{15}$ ao meio e o quociente de $Y \sim N(0,1)$ por raíz quadrada de $Z \sim \chi^2_{(15)}$ à direita, conforme pode ser obervado na Figura 7.

```
gl=15
y<-rnorm(1000)
z<-rchisq(1000,gl)
x<-y/sqrt(z/gl)
hist(y,prob=TRUE)
curve(dnorm(x),-4,4,col=2,add=T)
hist(z,prob=TRUE)
curve(dchisq(x, df=gl),0,40,col=2,add=T)
hist(x,prob=TRUE)
curve(dt(x, df=gl),-5,5,col=2,add=T)
```

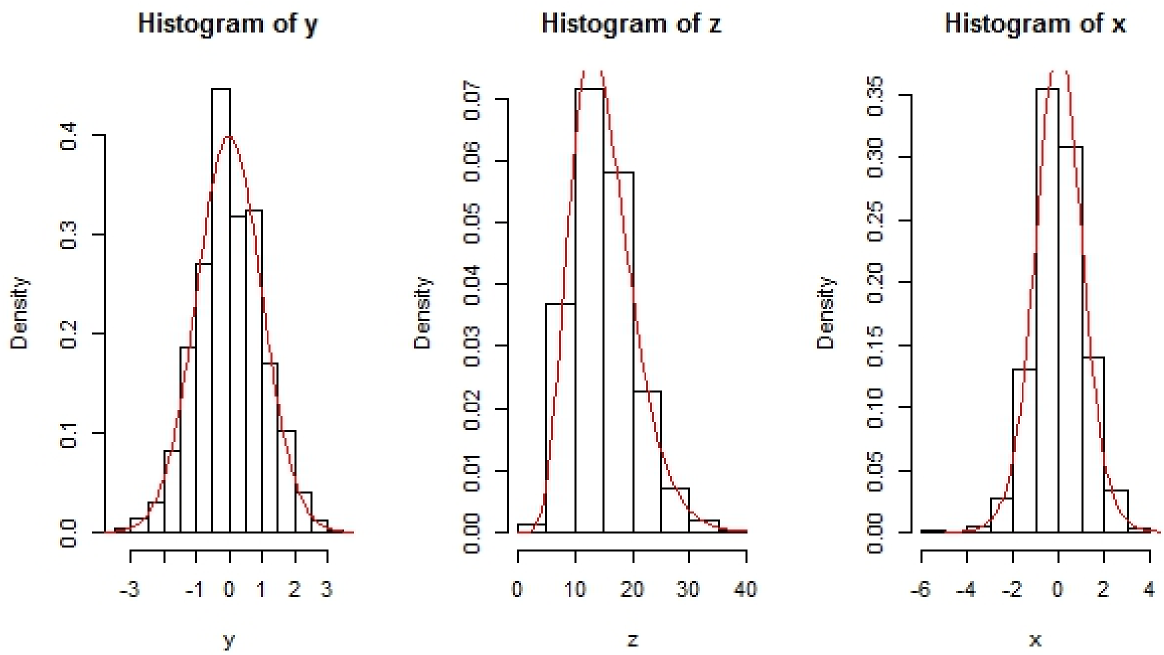

Figura 7: Gráficos da relação de distribuição de probabilidade com 15 graus de liberdade.

Na Figura 8, é possível compararmos os quantis do conjunto de dados referentes a distribuição Normal e os quantis do conjunto de dados da distribuição t de Student ambos com 15 graus de liberdade e ficou bem claro que os quantis seguem sim distribuições teóricas conhecidas, à esquerda segue uma distribuição Normal padrão e à direita uma t de Student.

```
quantis <- qnorm(ppoints(length(y)))
qqplot(quantis, x))
quantis <- qt(ppoints(length(x)),gl)
qqplot(quantis, x)
```

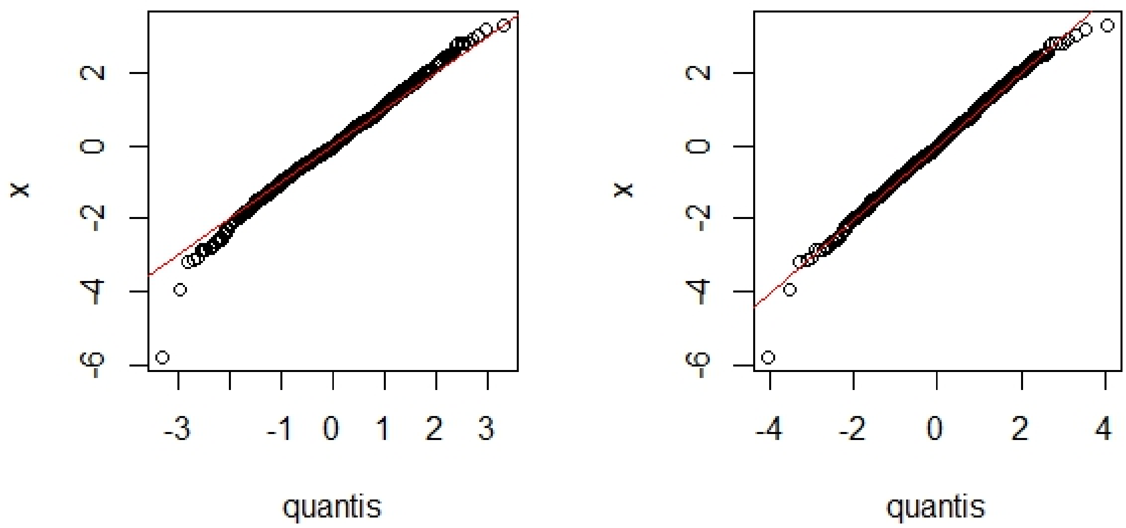

Figura 8: Gráficos *qq-plot* da distribuição Normal com a distribuição t de Student com 15 graus de liberdade.

Na terceira simulação mantemos as amostras de 1000 observações para ambas as distribuições logo, $Y \sim N(0,1)$ á esquerda, $Z \sim \chi^2_{30}$ ao meio e o quociente de $Y \sim N(0,1)$ por raiz quadrada de $Z \sim \chi^2_{30}$ á direita, conforme a Figura 9.

```
gl=30
y<-rnorm(1000)
z<-rchisq(1000,gl)
x<-y/sqrt(z/gl)
hist(y,prob=TRUE)
curve(dnorm(x),-4,4,col=2,add=T)
hist(z,prob=TRUE)
curve(dchisq(x, df=gl),10,60,col=2,add=T)
hist(x,prob=TRUE)
```

```
curve(dt(x, df=gl),-4,4,col=2,add=T)
```

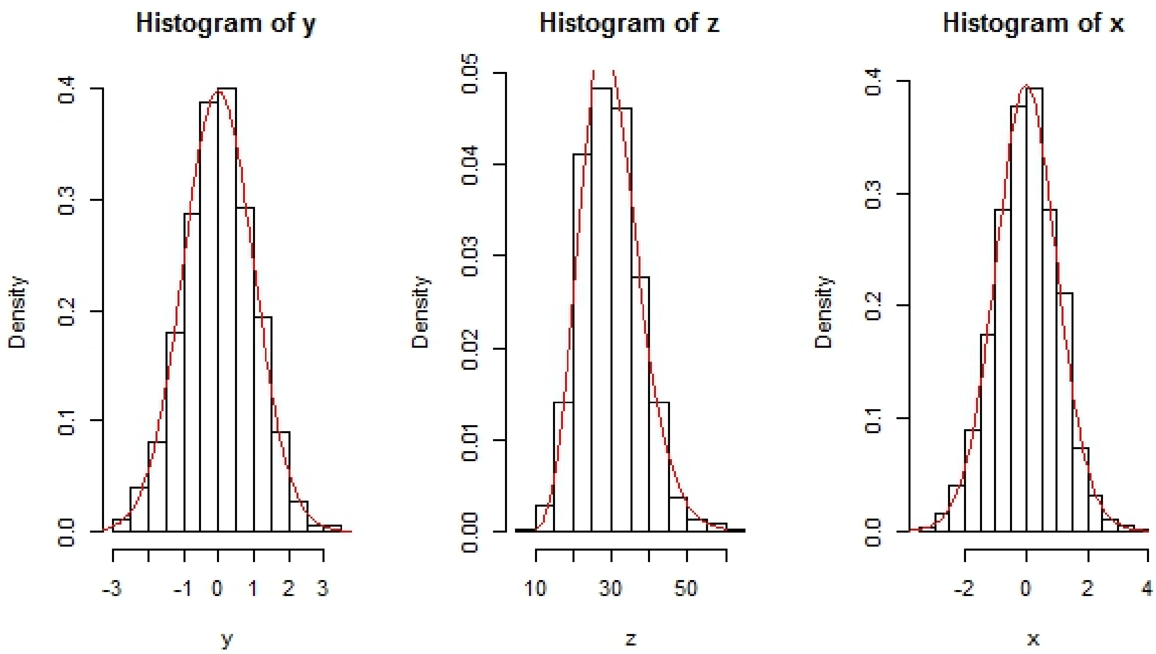

Figura 9: Gráficos da relação de distribuição de probabilidade com 30 graus de liberdade.

Na Figura 10, iremos comparar mais uma vez se os quantis do conjunto de dados referente a distribuição Normal e os quantis do conjunto de dados da distribuição t de Student seguem alguma distribução teórica conhecida, logo fica bem visível que à esquerda seguem sim a distribuição Normal padrão e à direita seguem a distribuição t de Student.

Portanto foi possível observar visualmente que de acordo com que os graus de liberdade aumentam, os dados da amostra aleatória geradas no *software* **R**, seguem uma distribuição de probabilidade conhecida.

```
quantis <- qnorm(ppoints(length(y)))
qqplot(quantis, x)
quantis <- qt(ppoints(length(x)),gl)
qqplot(quantis, x)
```

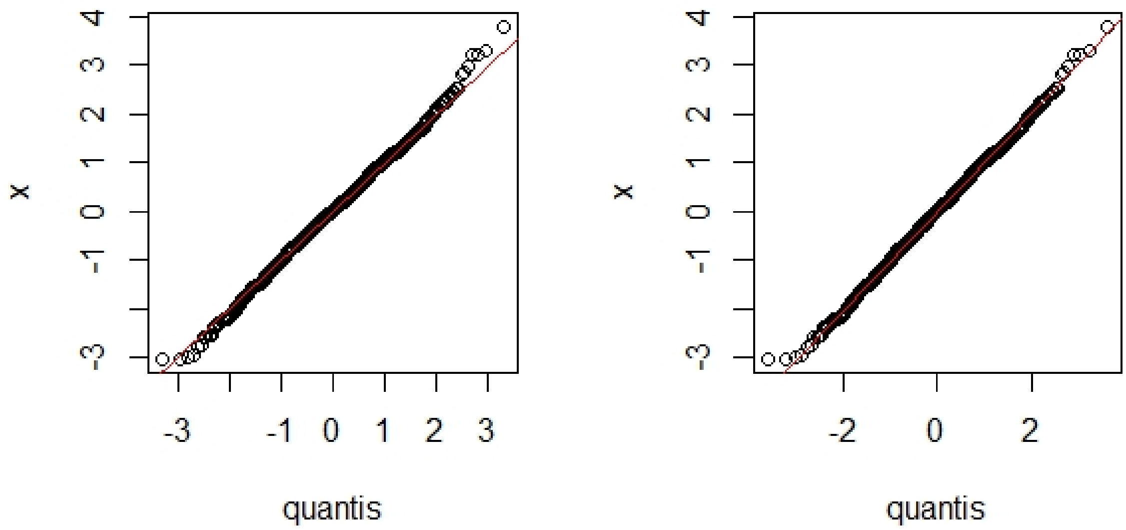

Figura 10: Gráficos *qq-plot* da distribuição Normal com a distribuição t de Student com 30 graus de liberdade.

4 Conclusão

Neste trabalho, tivemos a oportunidade de fazer uma breve revisão de probabilidade onde podemos observar que o problema filosófico da probabilidade coloca-se no centro do conhecimento humano, nomeadamente do conhecimento científico e da natureza. Está fortemente associado aos problemas da indução, do determinismo/indeterminismo, da certeza/incerteza e da ordem/desordem.

O estudo da teoria das probabilidades é um instrumento que nos ajuda a estimar com o máximo de precisão possível o resultado de eventos, dos quais não podemos dizer, antecipadamente, qual será o resultado mais provável. Ele se aplica a quase todos os campos do conhecimento humano. Conhecendo o cálculo de probabilidades o aluno terá a oportunidade de relacioná-lo com dados da experiência cotidiana, dar significado ao aprendizado, fazer a ponte entre a teoria e a prática, a fundamentar a crítica, a argumentar com base em fatos observados a partir de dados.

A distribuição Normal, é uma distribuição de probabilidade contínua das mais importantes e utilizadas, geralmente citada como curva normal ou curva de Gauss. Sua importância em análise matemática resulta do fato de que muitas técnicas estatísticas, como análise de variância, de regressão e alguns testes de hipótese, assumem e exigem a normalidade dos dados. Além disso, a ampla aplicação dessa distribuição vem em parte devido ao teorema do limite central. Este teorema declara que na medida em que o tamanho da amostra aumenta, a distribuição amostral das médias amostrais tende para uma distribuição normal.

Uma outra distribuição de probabilidade bastante importante no contexto da Estatística é a distribuição de Qui-quadrado, que surge no contexto de somas de quadrados de variáveis aleatórias com distribuição Normal.

Já a distribuição t de Student é uma distribuição de probabilidade teórica, simétrica e campaniforme, semelhante à curva Normal padrão, porém com caldas mais largas, ou seja, as realizações da distribuição t de Student podem gerar valores mais extremos (positivos ou negativos) do que a distribuição normal padrão. O único parâmetro ν que a define e caracteriza a sua forma é o número de graus de liberdade, quanto maior for esse parâmetro, mais aproximada da Normal ela será.

Este trabalho tratou sobre relação das distribuições Normal, Qui-quadrado e t de Student, por simulações computacionais, logo foi muito interessante e ao mesmo tempo gratificante demostrar que se estudarmos a teoria de probabilidade com mais precisão descobriremos que há fatos acerca de tal teoria que nos deixará bem curiosos. Porém durante a pesquisa foi muito difícil encontrar material na literatura sobre tais assuntos abordados. Logo a importância deste trabalho é apresentar alguns apontamentos teóricos e metodológicos sobre a realização de simulação para mostrar gráfica e visualmente relações entre distribuições de probabilidades teóricas.

Referências

ACTION, P. *Distribuição t de Student*. 2014. Disponível em: <http://www.portalaction.com.br>. Acesso em: 12/06/2014.

BARROS, M. *Estatística para Metrologia*. [S.l.], 2008.

BOLFARINE, H.; BUSSAB, W. O. *Elementos de amostragem*. 1. ed. São Paulo: Blucher, 2005. 274 p.

BRASIL, S. E. M. T. Parâmetros curriculares nacionais + (pcn+) - ciências da natureza e suas tecnologias. *Ministério da Educação*, Brasília 2002.

BUSSAB, W. O.; MORETTIN, P. A. *Estatística Básica*. 5. ed. São Paulo: Saraiva, 2002. 520 p.

CASSIANO, C. R. P.; ALVES, O. S. Probabilidade e estatística. 2011.

CORREA, S. M. B. *Probabilidade e Estatística*. 2. ed. Belo Horizonte: PUC Minas Virtual: [s.n.], 2003. 116 p.

COSTA, J. F. S.; HURTADO, N. H. A probabilidade no ensino médio. *A importância de jogos como ferramenta didática*, 2005.

CUZZOL, I. *Introdução á Probabilidade*. 2014. Disponível em: <http//www.ebah.com.br/content/ABAAAAp6UAB/intodução-a-probabilidade>. Acesso em: 04/02/2014.

MILLER GRIARSTEAD charles; LAURIE, J. *Introduction to probability*. 2011. Disponível em: <¡snelhttp//pt.widipedia.ort/widi/probabilidade>. Acesso em: 24/11/2011.

LEBENSZTAYN, E.; COLETTI, C. F. *Probabilidade: Teoria e exercícios*. Programa de pós-graduação em estatística departamento de estatística universidade de são paulo. São Paulo: Notas de aula, 2008. 121 p.

LEITE, I. C. C. Distribuições de probabilidade. Salvador 2007.

MEYER, P. L. *Probabilidade: Aplicações à Estatística*. 2. ed. Rio de Janeiro: LTC-Livros Técnicos e Científicos Editora S. A., 1983. 426 p.

PORTNOI, M. *Probabilidades, variáveis aleatórias, distribuição de Probabilidade e geração aleatória*. [S.l.], 2005.

REGO, L. C. *Programa de Pós-Graduação em Engenharia de Produção*. [S.l.], 2012.

ROLLA, L. T. *Introdução à Probabilidade*. [S.l.], 2013.

SOARES, P. *Probabilidade e Estatística*. [S.l.], 2010.

WIKIPEDIA, C. da. *Variável aleatória contínua*. 2013. Disponível em: <http://pt.wikipedia.org>. Acesso em: 17/02/2014.

WIKIPEDIA constribuidores da. *Falácia do apostador*. 2014. Disponível em: <http://pt.wikipedia.org>. Acesso em: 17/05/2014.