**Teoria dos Conjuntos.**

Um conjunto é uma coleção de elementos distinguíveis, i.e., cada elemento só aparece uma vez no conjunto. É preciso ficar bem claro que elementos pertencem ou não pertencem ao conjunto[[1]](#footnote-1). Geralmente isso é feito através de propriedades partilhadas por todos os elementos do conjunto. Exemplo:

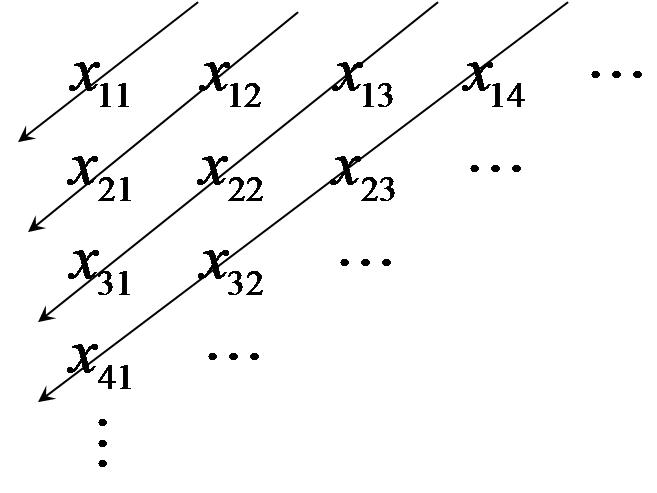
Seja A o conjunto de todos os elementos que possuem a propriedade P, então a sentença

 x possui a propriedade P.

 x é inteiro positivo ou zero.

O conjunto vazio  não possui qualquer elemento. Um conjunto finito tem um número finito de elementos, e um conjunto infinito possui um número infinito de elementos. Dois conjuntos A e B possuem a mesma potênica se for possível estabelecer uma relação biunívoca entre seus elementos, ou seja, a cada elemento de A pode-se associar um, e só um, elemento de B. O conjunto pode ser enumerável [countable] ou não enumerável. Se for enumerável o conjunto tem uma associação biunívoca com o conjnto dos núemros naturais. Um conjunto infinito pode ser enumerável, como o dos números naturais. Todo conjunto finito é enumerável pois podemos ordenar seus elementos e associá-los a 1, 2, 3, etc.

Os elementos de um conjunto de conjuntos enumeráveis formam um conjunto enumerável. Pense em uma matriz



Podemos enumerá-los pela seqüência triangular, e dentro da diagonal pelo primeiro índice, como mostra a tabela xx abaixo:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **x11** | **x12** | **x21** | **x13** | **x22** | **x31** | **x14** | **x23** | **x32** | **x41** |
| **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |

Assim fizemos a correspondência com . Conseqüência dessa fato é que o conjunto dos números racionais é enumerável, pois  contém dois índices, o n e o m, logo pode ser enumerado usando a mesma regra acima.

Entretanto, o conjunto de todos os números em um intervalo não é enumerável. Basta trabalhar com o intervalo . A pergunta é: o conjunto de todos os números no intervalo  é enumerável? Vamos provar que não por absurdo.

Suponha que seja. Então temos  e podemos ordená-los em ordem crescente:

 uma vez que dois números não podem ser iguais. Neste caso  é tal que  e  não pertence ao conjunto dado. Logo o conjunto não incluiu todos os números entre 0 e 1. Note então que existem  números racionais entre 0 e 1 e que também existem  números irracionais entre 0 e 1. Só que o conjunto dos irracionais não é enumerável, e dos racionais é enumerável, ou seja, .

**Álgebra dos conjuntos.**

São duas as operações principais entre conjuntos: a UNIÃO e a INTERSEÇÃO.

**Operação UNIÃO:**

Seja A o conjunto dos elementos com a propriedade PA e B o conjunto dos elementos com a propriedade PB. Se então  ou . Ou seja, x ou tem a propriedade PA  ou tem a propriedade PB. Note que a operação lógica da união é OU. Vamos usar a notação 0 para falso e 1 para verdadeiro. A tabela da verdade para essa operação é dada por:

|  |  |  |
| --- | --- | --- |
| **PA** | **PB** |  |
| **1** | **1** | **1** |
| **1** | **0** | **1** |
| **0** | **1** | **1** |
| **0** | **0** | **0** |

Ou seja se x possui PA e PB então ; se x possui PA mas não PB então ; se x não possui PA mas possui PB então  e, finalmente, se x nem possui PA nem possui PB então . Em linguagem de conjuntos estamos afirmando que:



Na nossa álgebra de lógica em que só existem 0 e 1, falsa ou verdadeira, então , ,  e . Por isso é comum associar o sinal de + à operação lógica OU.

Ou seja  quando A e B são conjuntos.

Propriedades da operação união[[2]](#footnote-2).

Associativa: 

Comutativa: 

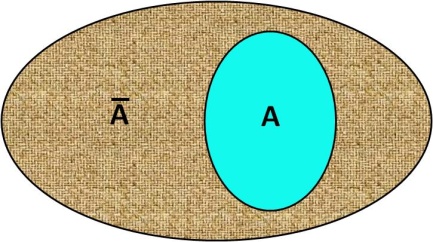
Elemento unitário: 

Observação: apesar da operação união possuir o elemento unitário  ela não admite inversa pois  se  ou .

**CONJUNTO UNIVERSO**

O conjunto universo é definido como o conjunto contendo todos os elementos possíveis, com todas as propriedades existentes em dado contexto e denominado por S. Note que  sempre e que .

**CONJUNTO COMPLEMENTAR .**

****

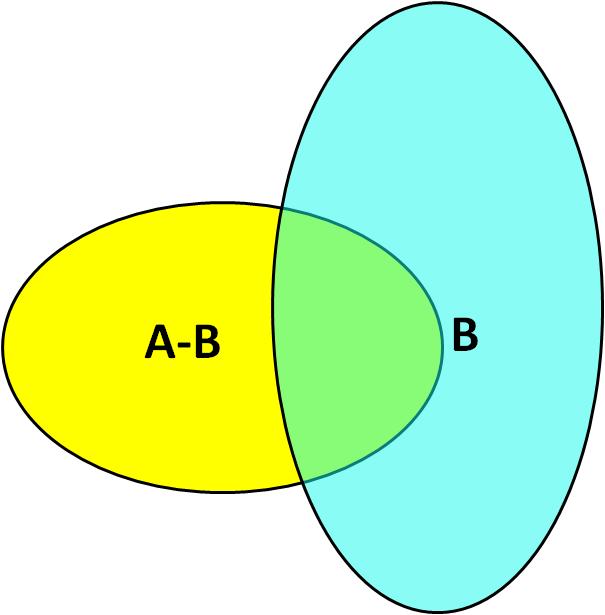
Se  então  e .

Propriedades:

; ; ; se  então  e se  então .

**Operação DIFERENÇA **

Se  então  e  OU  e . Se  então  e .



**Operação INTERSEÇÃO:**

Dizemos que  se  E . A operação lógica nesse caso é E (AND). Ou seja, agora temos que:



A tabela da verdade para essa operação é dada por:

|  |  |  |
| --- | --- | --- |
| **PA** | **PB** |  |
| **1** | **1** | **1** |
| **1** | **0** | **0** |
| **0** | **1** | **0** |
| **0** | **0** | **0** |

Como , ,  e , usamos também a notação de multiplicação na forma .

Propriedades da operação união.

Associativa: 

Comutativa: 

Distributiva frente à união: 

Se  então ; ; ;  e.

Conjuntos disjuntos: se  dizemos que A e B são disjuntos, ou mutuamente exclusivos. Se pertence a A não pertence a B e se pertence a B não pertence a A.

PARTIÇÃO. Uma partição de um conjunto A é uma coleção de subconjuntos Ai tais que: , ou seja, , entretanto .

Algumas partições clássicas:

1.  e 
2.  e 
3.  e 
4.  e 

**Leis de De Morgan:**

São leis super importantes na teoria dos conjuntos e muito úteis para demonstração de teoremas. Podem ser apresentadas em duas formas equivalentes:

Forma 1: 

Forma 2: 

A estratégia para demonstrá-la e usar o fato de que se  e  então .

Forma 1: Se 

Com isso demonstramos que se  então  o que significa que . Entretanto, como todas as setas são bidirecionais também concluímos que se  então  logo , significando que .

Forma 2: Se . Com isso mostramos que se  então  logo que significa que . Com a bidirecionalidade das setas concluímos que se  então  logo , e .

Parecem duas leis mas na realidade é uma só. Dado uma a outra será verdadeira e vice-versa. Passando de uma forma à outra:

Na forma 2 fazer  e  logo  agora tirar o complementar de ambos os lados  a forma 1. Na forma 1 fazer  e  logo  tirar o complementar de tudo novamente .

**Probabilidade.**

Vocabulário:

Experimento. Na estatística designa uma atividade para a qual não se pode especificar antecipadamente o resultado final. Jogar um dado, por exemplo, é um experimento. Jogar um dado duas vezes seguidas é um experimento. Se é possível especificar o resultado antecipadamente se diz que estamos no campo determinístico. Experimento nas ciências exatas possui outra conotação – é uma experiência determinística utilizada para comprovar ou falsificar uma teoria ou modelo.

TRIAL (ensaio, tentativa). Cada performance isolada de um experimento é um trial.

Resultado (outcome). É o resultado do experimento. Exemplo, joguei o dado e obtive 5 – 5 é o resultado. Cada trial dá origem a um resultado. Jogar dois dados, por exemplo, pode dar o resultado (2,3).

Espaço amostral S ou . O conjunto de todos os resultados do experimento é o espaço amostral. Esse conjunto pode conter mais resultados do que os possíveis, mas não pode deixar de conter todos os possíveis. No caso de um dado . Agora suponha o conjunto da quantidade de gordura no leite, x. Sabemos que  embora se saiba que  é praticamente impossível. Logo  também é um espaço amostral. Todo resultado, portanto, é um elemento do espaço amostral. O espaço amostral pode ser finito, infinito, enumeráel ou não enumerável.

Evento. Evento é um sub-conjunto do espaço amostral. São coleções de resultados de um experimento.

**FUNÇÃO**. Uma função é uma regra de associação entre elementos de um conjunto chamado domínio com elementos de outro conjunto chamado contra-domínio. Para ser função a regra deve ser clara, sem dar origem a impasses, deve se saber exatamente a que elemento associar e o que fazer com todos os elementos do domínio. Não pode portanto, associar um elemento do domínio a mais de um elemento do contra-domínio pois haveria dúvida sobre qual regra seguir. Além disso, todos os elementos do domínio devem poder ser associados para evitar não saber o que fazer com um elemento que não se pode associar.

Estamos acostumados à funções de um conjunto de números em outro conjunto de números, mas podemos perfeitamente associar um conjunto a uma número, ou conjuntos a conjuntos. Um exemplo de uma função de conjunto que associa elementos de um conjunto a um número é o indicador do conjunto:



Probabilidade é uma função de conjunto, que deve associar um número real à todo evento A do espaço amostral.

**Definições de probabilidade.**

Subjetiva: uma pessoa julga qual a probabilidade de ocorrência dos eventos.

Freqüência relativa. Executa um experimento N vezes e conta quantas vezes o evento A ocorreu e assim associa à probabilidade . A dificuldade dessa definição é que seria preciso repetir o experimento inifinitas vezes. Também só seria útil se for possível provar que  estabiliza para certo valor à medida que N cresce, ou seja, que  converge.

Clássica. Seja um espaço amostral finito com N resultados igualmente PROVÁVEIS e A um evento com NA elementos, então . A maior dificuldade com essa definição [é que ela usou o conceito de probabilidade para definir probabilidade [igualmente prováveis]. Ou seja, é uma definição circular. Também, da forma como foi definida, seria impossível analisar o comportamento de um dado desonesto. Finalmente restringe o estudo a espaços amostrais finitos.

Dadas todas as dificuldades apontadas acima finalmente chegou-se a conclusão que a probabilidade deveria ser definida através de axiomas.

**Definição Axiomática.**

São apenas três os axiomas para uma função de conjuntos , com , que pode representar uma probabilidade:

1. 
2. , é chamado de evento certo.
3. Se  então 

Tudo o que pode ser demonstrado através dos axiomas é teorema e não deve ser colocado na mesma categoria de axioma. Com esses 3 axiomas podemos mostrar vários teoremas:

1. .

Prova:  mas , logo  e .

1. 

Prova:  e  logo pelo axioma 3 mas , logo  . Por outro lado  pelo axioma 2. Então  e .

Corolário: Se ,  e  então  e , pois  e . Repetindo o argumento temos também que . Como  e  e  pelo axioma 1 então  e .

1. 

Prova:  logo . Fazendo a união com A temos  entretanto  logo  e . Note que  e  representam duas partições pois  e . Aplicando axioma 3 nas duas partições temos:  e . Extraindo  da primeira partição e substituindo na segunda temos:



Esse teorema implica em que a probabilidade é sub-aditiva, ou seja, a união dos conjuntos leva a uma probabilidade menor do que a da soma das probabilidades.

1. Se então . Prova  logo  . Mas como  então , então  e como  então.

**Eventos independentes:**

Os eventos  e  são independentes se .

Daí podemos mostrar como teoremas que se  e  são independentes então ( e ), ( e ) e ( e ) também são independentes entre si. Isso significa que os eventos complementares também são independentes.

Prova: Usando  e , percebemos que , logo . Como  e  são independentes, então , provando que  e  são independentes. Chamando  de  e vice-versa temos que  e  são independentes. Se  e  são independentes então mudando  para  temos que  e  são independentes.

**Probabilidade Condicional.**

Pergunta: qual a probabilidade do evento  sabendo que o evento  ocorreu? Denotamos essa probabilidade por , [leia-se: p de A dado B]. Se  ocorreu então  e podemos restringir o espaço amostral para . Agora basta mostrar que  obedece aos axiomas da probabilidade.

1. 
2.  pois  e 
3. Se  então:



Teorema da propabilidade total:

Seja  uma partição de  e  um evento arbitrário. Então:



Prova:  e , logo  é uma partição de . Nesse caso:



Agora basta substituir  para provar o teorema.



**Teorema de Bayes.**

 logo:



Thomas Bayes [1701 – 1761] estabeleceu o teorema de Bayes em uma obra póstuma Bayes ***“An Essay towards solving a Problem in the Doctrine of Chances”*** [1763] editada pelo seu amigo Richard Price, da tabela Price.



A inferência de Bayes está sendo hoje, cada vez mais, considerada mais robusta do que a inferência frequentista de Fisher.

**Exemplos da utilização da inferência Bayesiana:**

**[extraídos do livro “The Signal and the Noise” de Nathan Silver.**

**Exemplo 1:** Após uma viagem Ella encontra uma calcinha estranha em sua gaveta. Deseja saber a probabilidade de estar sendo traída pelo seu parceiro dado que a calcinha foi encontrada. O evento  foi: encontrar calcinha estranha na sua gaveta. O evento  é estar sendo traída sem qualquer outra informação e o evento complementar  é não estar sendo traída. Nesse caso o evento  só tem duas opções: traição  ou não traição , e . A probabilidade  é chamada de probabilidade à priori, ou prior. Sem qualquer informação extra pode-se apelar para estatísticas da sociedade como um todo na qual se sabe que 4% dos parceiros traem durante um ano. Assim, a probabilidade de estar sendo traída vale  e de não estar sendo traída vale . Agora precisamos estimar, probabilidade da calcinha estranha aparecer dado que está sendo traída, e , a probabilidade da calcinha estranha aparecer mesmo sem existir traição. Vamos supor que exista a traição. Mas nesse caso se espera que o parceiro fosse mais cuidadoso em evitar que a calcinha fosse abandonada na gaveta – digamos então que seja de 50% a chance dele não ter percebido a calcinha estranha no momento da traição. Assim. Agora pode-se imaginar a probabilidade da calcinha estranha aparecer sem que exista traição – tipo ele comprou a calcinha e esqueceu de contar, comprou a calcinha para ele mesmo, uma amiga em comum confiável dormiu em casa, a empregada misturou tudo, etc. Digamos que essa probabilidade seja de . Agora pode-se calcular a probabilidade de estar sendo traída dado que a calcinha estranha apareceu como:





**Exemplo 2:** Uma mulher na faixa dos 40 anos fez uma mamografia que deu positiva para câncer de mama. Qual a probabilidade de ela ter câncer de mama dada a mamografia positiva. Evento é ter câncer de mama e  é não ter câncer de mama na faixa dos 40 anos. O evento  é a mamografia ser positiva quando há câncer de mama, e o evento  é a mamografia ser positiva quando não há câncer de mama [são chamados falso-positivos]. Sem qualquer outra informação sabe-se que a probabilidade de uma mulher na faixa dos 40 desenvolver câncer de mama é de apenas , logo . Também se sabe que a probabilidade da mamografia dar positivo na presença de câncer de mama é de , ou seja, 25% dos cânceres de mama deixam de ser detectados pela mamografia. Por outro lado a freqüência de falsos positivos é de . Assim a probabilidade dela ter câncer de mama dado que a mamografia deu positiva é de:



Ainda bastante baixa. Por isso só se recomenda exame periódico de mamografia depois dos 50, quando a prior aumenta. Nathan Silver demonstra esse resultado da seguinte forma. Suponha 1000 mulheres fazendo mamogramas. Dessas 1000 apenas 14 estarão com câncer de mama enquanto 986 não estarão com câncer de mama. Das 986 sem câncer de mama aparecerão 99 com mamografia positiva. Das 14 com câncer de mama aparecerão 11 com mamograma positivo e 3 cânceres não serão detectados. Ou seja, do total de 110 mamogramas positivos apenas 11 são realmente câncer de mama.

**Exemplo 3:** No ataque terrorista de 2001 qual a probabilidade da primeira colisão com a world trade center ser um ataque terrorista. Nesse caso evento  é haver um ataque terrorista e  é não haver um ataque terrorista. Vamos chutar que a chance de haver um ataque terrorista contra um arranha céu de Manhattan seja de 1 em 20.000, contando as tentativas que já ocorreram no passado. Trata-se de um chute, pois só houve um atentado terrorista de sucesso contra um arranha céu de Manhattan em toda a história da cidade, justamente o de 11 de setembro de 2001. Isso significa que a estimativa dessa probabilidade é muito pouco precisa. Mas deixa assim mesmo. Então  e . Agora, estimando o caso pior aceitamos que a probabilidade de acertar o WTC dado que é realmente um ataque terrorista seja de 100%, então . Agora a probabilidade de uma colisão não intencional de um avião com um arranha céu em Manhattan é bem precisa pois em 25000 dias de aviação sobre Manhattan antes de 2001 só existiram 2 acidentes, um no Empire State Building em 1945 e outro em Wall Street em 1946, ou seja,  . Nesse caso a probabilidade da primeira colisão ser de uma ataque terrorista é:



Agora vamos atualizar a prior para  e recalcular a probabilidade do segundo ataque ser terrorista:



**Inferência Bayesiana.**

Fisher não gostou das estimativas da prior e das probabilidades  e  que lhe pareceram muito subjetivas. No fundo o método de Fisher é estimar os parâmetros da distribuição através das observações e com eles calcular as probabilidades.

A robustez da aparente fraqueza da metodologia de Bayes, entretanto, é o fato de que uma atualização dos dados usando as probabilidades posteriores como prior do processo seguinte converge para a probabilidade real. Mas para isso é necessário provar que a probabilidade converge para a probabilidade real à medida que se atualizam as informações da inferência Bayesiana.

http://en.wikipedia.org/wiki/Bayesian\_inference

**Variável aleatória.**

Trabalhar com funções de conjuntos é bem mais complicado do que trabalhar com funções numéricas. Por isso pode ser interessante criar uma associação entre os conjuntos A do espaço amostral e os númeors. Ou seja, vamos criar uma nova função de conjuntos  que permite associar um número a cada evento. Assim poderemos trocar  por  onde  é uma variável aleatória definida pela função de conjunto  é uma função de conjunto com imagem . Ou seja, a variável aleatória não é uma variável mas uma função. Para distinguir a função de conjuntos do valor que ela pode assumir vamos designar por  a função e  o seu valor. Para ser uma variável aleatória a função de conjunto precisa satisfazer poucas condições.

1. O conjunto é um evento para 
2. 

Note que a um conjunto evento do espaço amostral estamos associando uma probabilidade e um valor da variável aleatória . Queremos eliminar a necessidade de passar pelo estágio intermediário dos conjuntos para chegar diretamente na probabilidade. Nossa questão então é como andar na direção inversa. Ou seja, dado que  qual o conjunto A a ele associado e qual . Como garantir que o mapeamento inverso  tenha a estrutura definida para a probabilidade?

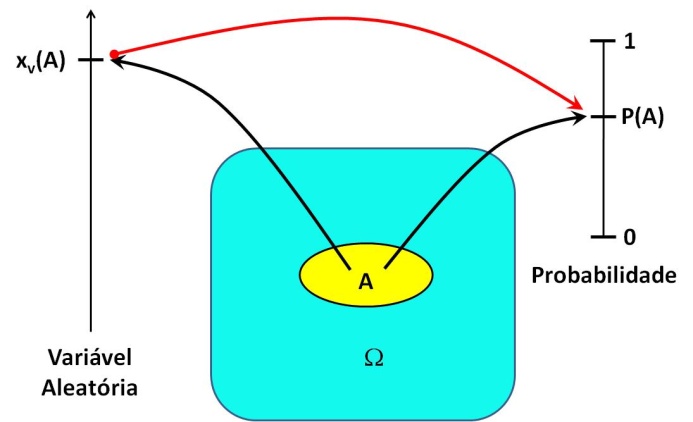


Figura xxx. Mapeamento do conjunto dos eventos para a probabilidade e para a variável aleatória. Seta vermelha: mapeamento direto da variável aleatória para a probabilidade.

Suponha que o conjunto de pontos  seja . Esse conjunto tem que ser um evento. Assim, dado o espaço de probabilidade a função  é uma função variável aleatória relativa ao campo de Borel  se, e somente se, ela é uma função com domínio  e imagem  tais que  para todo . Vamos dar um exemplo para evitar que o tópico fique muito abstrato. Jogar uma moeda duas vezes seguidas. Qual o conjunto de possibilidades, ou o espaço amostral? Como em português as possibilidades são CARA e COROA, ambas começando com C, vamos chamar as possibilidades pelas iniciais H e T dos nomes em inglês, Head ou Tail.

Nesse caso o espaço amostral é dado pelo conjunto: . O conjunto de todos os possíveis subconjuntos terá  elementos. Vamos agora definir a variável aleatória  como:



Ou seja  é o número de vezes em que H aparece. A pergunta é  é uma v.a. frente a  dado por:? Vejamos quem é .

|  |  |
| --- | --- |
|  |  |

Como ;  e  então  é uma v.a. frente a 

Um aspecto importante a notar aqui é que se então os eventos .

**Função Distribuição de Probabilidade**

Também chamada Função Distribuição Acumulada [Cumulative Density Function] [CDF]. Para evitar confusão com a Função Densidade de Probabilidade vamos denotar a Função Distribuição de Probabilidade por CDF. Sabendo que o conjunto  é um evento, podemos calcular  para qualquer valor de x. Assim a CDF é definida por:



Exemplo: jogar uma moeda desonesta, com probabilidade  de e  de , , uma vez. Vamos criar a v.a.  e . Essa distribuição é conhecida como distribuição de Bernoulli. Nesse caso



Como ,  e  então  .

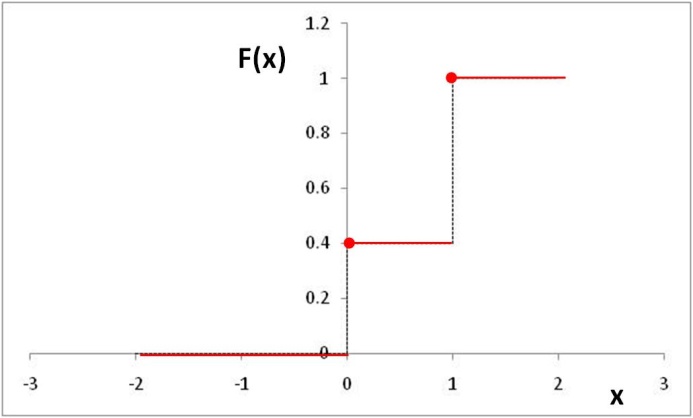


Figura xxx. FDA ou CDF da Distribuição de Bernoulli.

**Propriedades da CDF**

1.  e .
2. Se  então .
3. Se  então .
4. .
5.  é contínua pela direita, ou seja, .
6. .
7. .  fazendo  temos que . Se F é contínua, então  mas se F for descontínua, nesse caso,  será a descontinuidade no ponto x.

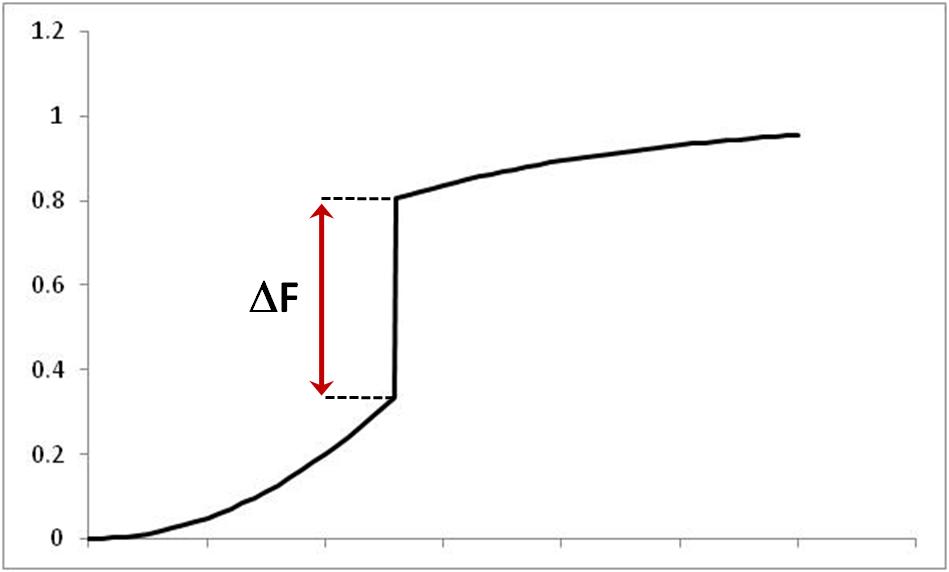
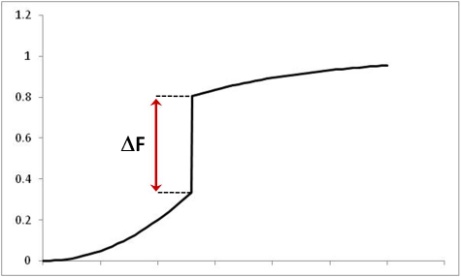
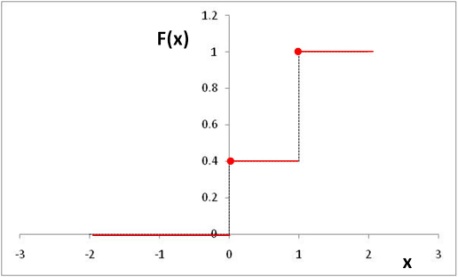
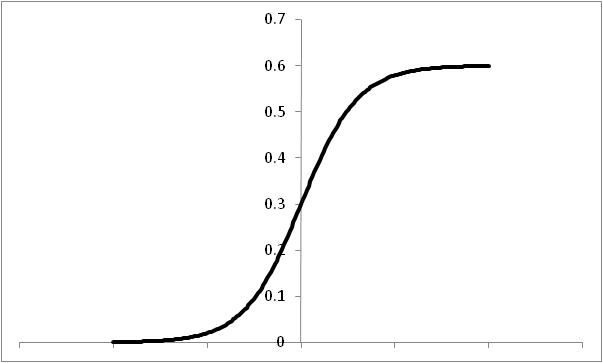


Figura xxx. Descontinuidade na Função Distribuição de Probabilidade fornece .

1. .

Um aspecto importante aqui é o fato de que a distribuição pode ser

1. Contínua
2. Discreta
3. Mista



(a) (b) (c)

Figura xxx. (a) Distribuição contínua; (b) Discreta e (c) Mista

**Função Densidade de Probabilidade [fdp]**

Essa função é definida como a derivada da função distribuição de probabilidade: . Note que dimensão da fdp é probabilidade por unidade de x, e não probabilidade. Assim também podemos definir a CDF em termos da fdp como: . O único problema aqui é que a F(x) pode ter pontos de descontinuidade nos quais a função não é diferenciável. Antes de lidar com as descontinuidades, através das funções Delta de Dirac, vamos extrair as propriedades da fdp supondo que F é diferenciável.

**Propriedades da Função Densidade de Probabilidade [fdp]**

1.  pois  uma vez que  e  , pela propriedade (2) da CDF. Poderíamos simplesmente ter afirmado que se F é sempre crescente então .
2. 
3. 
4. , ou seja, é a probabilidade de encontrar a v.a.  entre  e .

**Função Delta de Dirac ou Função Impulso**

Para lidar com as derivadas das descontinuidade necessitaremos das funções Delta de Dirac. Kronecker definiu a delta de Kronecker, muito útil no cálculo matricial e tensorial, dada por:



A matriz identidade pode ser escrita em termos do delta de Kronecker como . Em particular ela tem a propriedade de que:



Pois o único termo não nulo do produto  será o termo com .

Agora queremos uma função que opere da mesma forma para as integrais, ou seja, para :



Nesse caso queremos  se , entretanto a área sobre a delta tem que ser 1 pois:



Em outras palavras, estamos em busca de uma função que é nula para todo  mas que tenha área unitária, ou seja . Note que a exigência de que  implica em que a dimensão da delta é de 1/x. Se a largura da função delta de Dirac vai a zero então a altura deve ir para o infinito para garantir a área sobre a curva.

**Como construir a Delta de Dirac**

Partindo de uma função  de largura limitada, ou seja,  quando , mas cuja área seja 1 e independente de n. Além disso, é preciso que  então , ou seja, a largura da delta vai a zero. Fazendo o n tender a infinito então teremos a função Delta de Dirac como . Qualquer função  com as propriedades acima pode ser usada para construir a função Delta de Dirac.

Um exemplo é a função retângulo: .

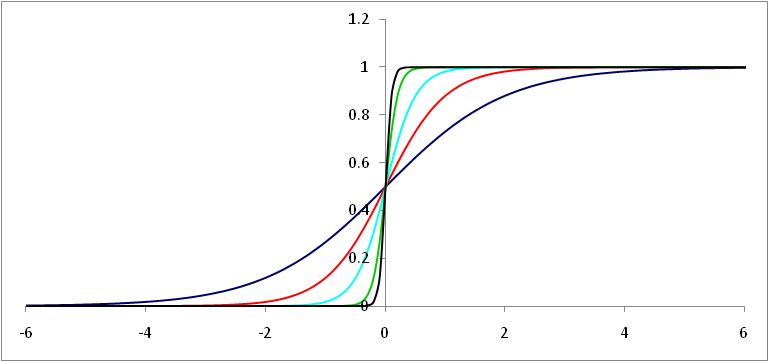
A área sobre a curva **vale **. Se  vai a infinito a largura vai a zero, a altura a infinito e a área se mantém constante. Vale notar que as funções densidade de probabilidade são excelentes candidatas à função Delta de Dirac, pela propriedade . Assim poderíamos usar distribuições Normais, de Cauchy ou qualquer outra com a propriedade da largura ir diminuindo e tendendo a zero quando determinado parâmetro vai a inifinito ou zero. Uma boa representação gráfica para a função delta de Dirac é a de uma seta vertical na posição xo.

**Delta de Dirac como a derivada da função Degrau.**

O importante nesse ponto é o uso da delta de Dirac para obter a derivada de funções descontínuas. Vamos considerar a função de Heaviside, ou função degrau, definida como:



Essa função é descontínua em e, portanto, não diferenciável. Agora considere a função logística dada por . Note que se  então  e  logo para  . Já para  então  e . Assim temos que  e . Para  . A função Hn é diferenciável  e essa derivada tem obviamente uma área sobre a curva igual a 1, pois . Aumentando o n se percebe que a função logística se parece mais e mais com a função degrau e que a largura de sua derivada vai diminuindo. Figura xx mostra  e para .



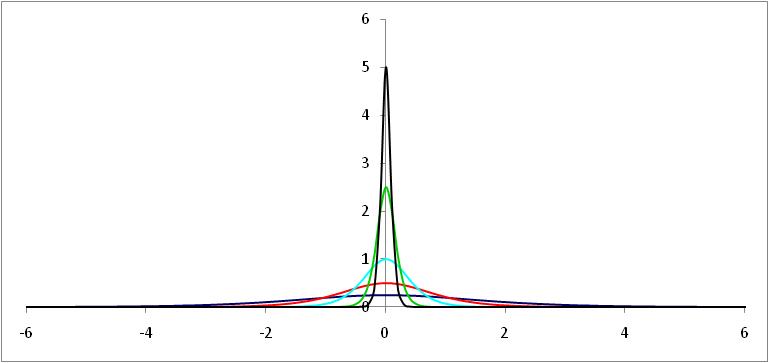


Figura xxx. Função Logística  e e sua derivada para .

Daí se percebe, então, que para , teremos: .

**Derivando funções descontínuas[[3]](#footnote-3).** Uma função descontínua da forma  pode ser escrita como . Agora podemos derivar esse função pela regra do produto como:



Ou seja  onde é a descontinuidade em xo.

Apêndice xxx mostra algumas propriedades da função Delta de Dirac. As duas que mais utilizaremos são:

1. 
2. 

**Função densidade de probabilidade de funções descontínuas.**

Agora a função Delta de Dirac dá conta de todas as descontinuidades da distribuição de probabilidade e não é mais necessário distinguir os casos discretos, mistos e contínuos e as definições:  e  são válidas em geral. Um bom exemplo é o caso da distribuição de Bernoulli onde  e . Notem que a fdp ficou com a dimensão de probabilidade por unidade de x após a multiplicação pelas deltas. Outro exemplo interessante é o de um dado honesto com probabilidade 1/6 para cada face e a v.a. sendo o número da face. Nesse caso a fdp será dada por:



Note que a CDF sai automaticamente da fdp através da integração:

 que leva a  pois .

## Função de uma Variável Aleatória

Uma nova v.a.  pode ser criada a partir de uma v.a.  desde que os seguintes requisitos sejam satisfeitos:

O conjunto  é um evento.

Os eventos  devem ter probabilidade nula, ou seja, .

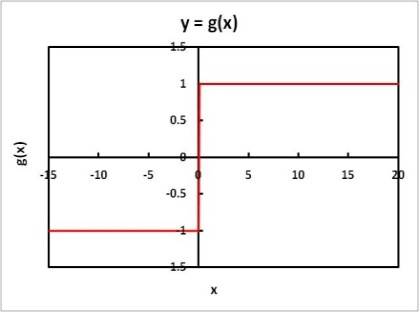
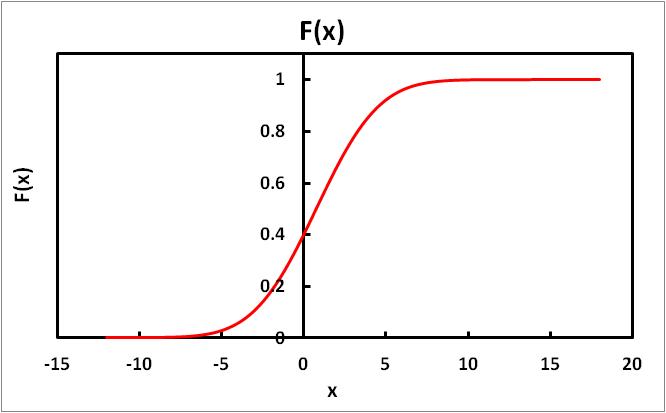
Imagem de  está contida no domínio de .

Note a necessidade desses requisitos. Se  não é um evento não existe probabilidade associada ao mesmo. O segundo requisito garante que , exigido para uma FDP. O terceiro é um pouco mais sutil. Precisamos ter certeza de que ao varrer  todos os valores possíveis de , ou seja, a imagem da função de conjuntos , estarão incluídos. Não podem faltar valores de  nem pode haver superposição de intervalos de . A ausência de superposição é garantida pelo fato de que  é uma função, ou seja, o mesmo valor de  só pode ser associado a apenas um valor de . Podemos calcular  da seguinte forma:

Encontrar todos os intervalos de  para os quais 

Calcular a probabilidade de cada um dos intervalos e somá-los

Note que  pode ser inclusive descontínua, constante, divergir, que mesmo assim poderemos encontrar a nova distribuição de probabilidade. Vejamos um exemplo de caso patológico. Para simplificar considere que segue uma distribuição contínua bem comportada, como a normal da figura 17(a). Agora vamos fazer  mostrada na figura 17(b). Note que nesse caso o conjunto  é vazio, logo tem probabilidade nula; o conjunto , para qualquer , corresponde ao conjunto , ou seja, com probabilidade  pelo gráfico da . Note que, por outro lado,  para qualquer  representa todo o espaço amostral, , logo é associado à probabilidade 1. A função distribuição de probabilidade de nesse caso é dada pelo gráfico da figura 17 (c).

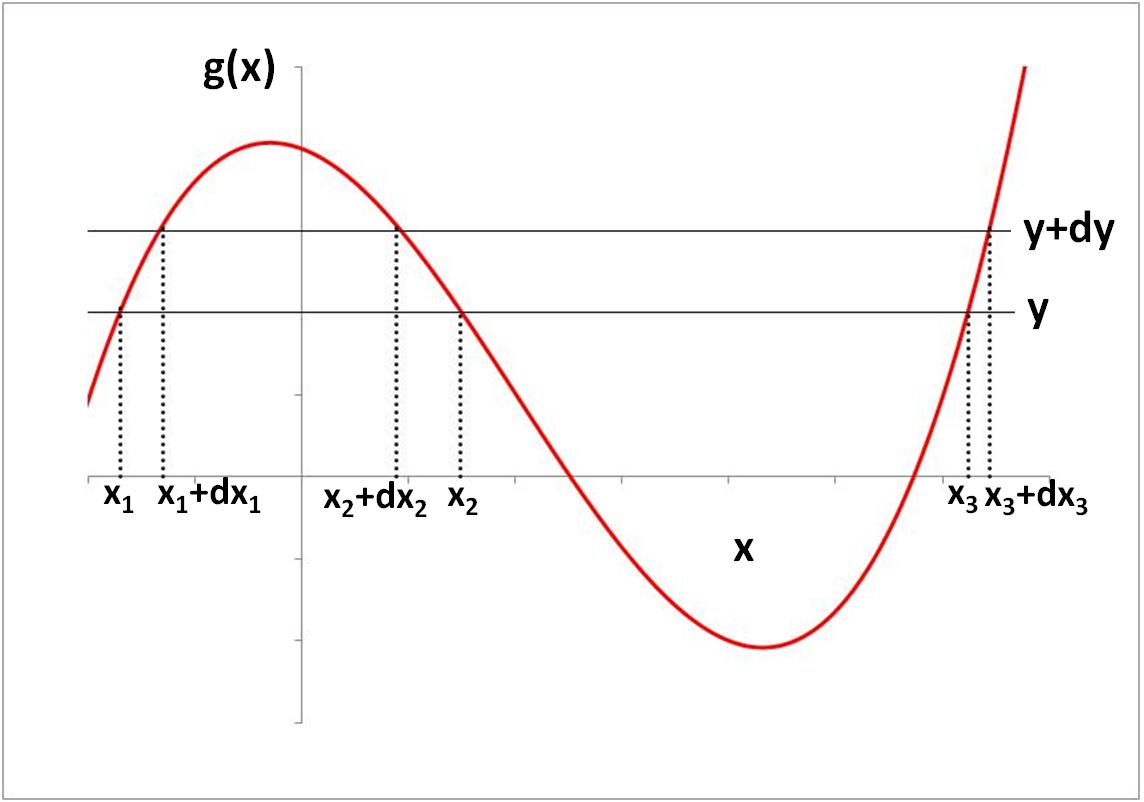


(a) (b) (c)

**Figura 17 -** F (x) de uma Normal (a), a CDF da função sign(x) (b) e a CDF da F(y) (c)

No caso especial em que  é diferenciável a função densidade de probabilidade da nova variável é dada por onde .

Prova: , ou seja, . A pergunta é, então, qual o conjunto de pontos de que leva ao conjunto para ? É aquele em que , como mostra a figura 20 onde existem 3 raízes . Note que o  da figura é negativo por que  é negativa nessa região, enquanto  e  são positivos.



**Figura 20 -** Três regiões em que , definindo ,  e . Nas regiões 1 e 3  é positiva, logo  e  também são positivos. Já na região 2  é negativa então  é negativo.

Note que , então  será negativo onde . Vamos separar as raízes com  e denotá-las pelo índice , das raízes com  denotadas pelo índice . Nesse caso:



com  e . Usando as propriedades da fdp temos que:





Onde a somatória é feita em  tal que , independente do sinal de . O módulo dá conta dos casos em que  é positiva ou negativa. Com isso temos, no final:



**Exemplo 1:** Vamos transformar a variável  da distribuição normal cuja fdp é dada por  para , . Note que a função inversa de admite duas raízes:  e . Não há raízes para , logo . Além disso,  então  e . Nesse caso, então:

.

Essa é a distribuição . A distribuição  faz parte das distribuições Gama que será apresentada no capítulo de aplicações e distribuições.

**Exemplo 2:** Distribuição log-Normal. Vamos transformar a variável  da distribuição normal para ,  . Nesse caso , e a função inversa . Como a função  é injetora então só há uma raiz  na qual . Não há raízes para . O resultado da transformação é:



**Operação ESPERANÇA.**

Esperança de uma v.a.: 

Os físicos também gostam da notação , as vezes também se usa  embora seja necessário tomar cuidado porque é a média obtida em uma amostragem e não a esperança da população completa. Nem sempre  embora se espere que sejam próximos, ou seja,  é uma boa inferência de . Note que o caso discreto sai automaticamente da utilização das funções delta de Dirac, pois:  pelas propriedades da delta.

**Esperança de uma  onde x é uma v.a.**

Vamos criar a v.a. , então  e mostrar que . Novamente  e devemos notar que  e que não há superposição dos intervalos correspondentes a diferentes raízes. Assim quando y varre o eixo vertical, os intervalos de x vão preenchendo completamente o eixo horizontal. Nesse caso  e . Daí extraímos que:



Casos particulares:

1.  logo . Constantes entram e saem da operação esperança.
2.  então ,  finalmente chegamos a que a esperança da soma é a soma das esperanças:.

**Momentos de ordem n:**

O momento de ordem n, se existir, é definido por: . A condição para a existência do momento é que a integral acima exista. Se para valores muito grandes de , i.e, comportamento assimptótico de f(x) para , a f(x) cai com uma lei de potência do tipo , então só existirão momento até ordem . Note que se  então  que diverge.

Algumas propriedades dos momentos são:

1.  e , pois  e .

**Momentos Centrados de ordem n.**

O momento centrado de ordem  é definido por: . Os momentos centrados possuem as propriedades:

1. , novamente 
2. , pois .
3. A variância é definida pelo , pois .
4. Se  é simétrica, ou seja, , então todos os momentos centrados ímpares serão nulos.
   1. Integração em intervalo simétrico de funções pares, , e ímpares, . Queremos . Na primeira integral mudar a variável para  então . Se f é par então  e . Já se f é ímpar então  e .

No nosso caso . Mudando, então a variável de integração para , , teremos portanto: , logo se  e é ímpar então .

**Relação entre os Momentos Centrados e não centrados.**

Podemos usar binômio de Newton  para encontrar a relação entre os momentos centrados e não centrados.

 logo



Casos particulares:

1. 
2. 
3. 
4. 
5. 

A volta, obter os momentos não centrados em termos dos centrados, pode ser feita da seguinte forma:

 logo: 

1. 
2.  pois  e .
3. 
4. 
5. 

**Função Geradora de Momentos (FGM)**

Considere a seguinte função da variável t: . Podemos usar a expansão em série de Taylor: , para obter: . Se comparamos com a série de Taylor da própria   vemos que . Por isso a função é chamada de geradora dos momentos. Para gerar os momentos centrados devemos multiplicar a função geradora dos momento por , uma vez que , logo . Daí se percebe que .

**Função Característica**

A grande dificuldade da função geradora dos momentos é a convergência da integral  por conta do . Se usarmos  entretanto, não teremos mais tantos problemas de convergência uma vez que  para qualquer  e . Assim a função característica é definida por: . Se a função geradora dos momentos existe então . Note que . Além disso, podemos mostrar que , pois . A relação com os momentos só precisa ser ligeiramente modificada uma vez que , levando a: . Se comparamos com a série de Taylor  vemos que . Novamente, os momentos centrados podem ser obtidos multiplicando a função característica por , obtendo .

Além da função característica ser mais poderosa do que a função geradora dos momentos a operação para sua obtenção é conhecida desde o século XIX e chama-se Transformada de Fourier:



A associação entre  e  é biunívoca de modo que ela admite transformada inversa dada por:



Podemos verificar a transformada inversa facilmente, substituindo a  abaixo e usando o fato de que , demonstrado no apêndice xxx:



Isso nos permite reconhecer a  dada a  e vice-versa. Transformadas de Fourier, e transformadas em geral, são uma ferramenta das ciências exatas e da matemática há longo tempo e existem milhares de tabelas associando as funções e suas transformadas assim como um listagem extensa de suas muitas propriedades. Uma propriedade muito importante na teoria da probabilidade é o teorema da convolução. No apêndice apresentamos uma lista introdutória das transformadas de Fourier e mostramos como calcular essas transformadas numericamente usando o Excel.

**Transformadas**

Transformadas integrais são relações entre duas funções através da equação integral , onde  é chamado de Kernel da transformada. Note que após a integração em t a função resultante só depende de s. Entre as mais conhecidas temos a transformada de Fourier , em que o Kernel é dado por , e a de Laplace , em que o Kernel é dado por . Note que as funções geradoras dos momentos são uma transformada de Laplace de dois lados. Mas essas não são as únicas, existem transformadas de Cauchy, de Hadamard, de Hankel, etc. São aplicadas em muitas áreas desde processamento de sinais e imagens (tomografia utiliza as transformadas de Hadamard), solução de equações diferenciais até a estatística avançada. Pode-se usar a transformada de Fourier de um sinal acústico de um tiro captado por um microfone para distinguir que tipo de arma foi utilizada e a distância do disparo ao microfone. Com três desses microfones saberíamos onde o disparo foi feito, com que arma e em que momento.

**Análise Multivariada.**

Vamos agora criar uma função vetorial de conjunto  que possui as componentes  em que cada  é uma v.a. Neste caso  e  são dois eventos, assim como . Para facilitar a compreensão e as demonstrações vamos trabalhar apenas com o caso bivariado, ou seja, duas v.a.s, e depois generalizar para . Facilita, nesse estágio, chamar uma v.a. de  e a outra de .

**Distribuição conjunta [Joint Distribution]**



Propriedades:

1.  e .
2. ****
3. ****
4. ****

**Densidade de probabilidade conjunta [joint density probability]**

Definimos a fdp conjunta agora como: .

O reverso é dado por:

Se queremos a probabilidade de encontrar  então devemos fazer a seguinte integral múltipla:



Também exigimos aqui que:



Para ser uma densidade de probabilidade multivariada, então,  e .

**Distribuição e Densidades Marginais:**

Suponha que queremos a estatística de apenas uma das variáveis sem interessar o valor da outra. Notamos que  assim como . Então  e  são as distribuições marginais de  e de . Note então que:



e



Ou seja integra-se em todas as possibilidades das outras variáveis para se obter a distribuição de uma variável independente dos valores das outras.

Nesse caso as densidades marginais serão dadas por:



e



Fica claro então que:

 e .

**Caso discreto:**

De forma análoga à distribuições univariadas os casos de distribuições discretas pode ser implementado com a função delta de Dirac generalizada para mais de uma dimensão definida como:

.

**Funções escalares  multivariadas.**

Vamos criar a v.a.  à partir das v.a.s  e  através da função escalar  que associa um vetor em  a um número real em . Nesse caso o evento  e a distribuição de probabilidade de  será dada por:



enquanto a fdp da v.a.  dada por:



A integral pode complicar devido à restrição  ou . Em vários casos pode ser vantajoso trocar as variáveis de integração para  e  através da regra do Jacobiano:



Onde o Jacobiano é dado pela matriz: .

**Operação esperança multivariada:**

Agora a operação esperança de qualquer função escalar das v.a.s  e  é dada por:



Dessa definição podemos extrair as seguintes propriedades da esperança:

1. 

Se  é uma constante então:

1. 



* 1. 

**Momentos conjuntos:**

No caso multivariado definimos os momentos por:



A generalização para  v.a.s é:



Notamos imediatamente que: . Alguns desses momentos possuem nomes específicos:





Com eles podemos definir os momentos centrados por:



Novamente percebe-se que:  e que:



, da mesma forma que .

Os momentos centrados com nomes específicos são as variâncias:





e a covariância:



Nota-se então que:



.

A covariância tem as seguintes propriedades:

1.  pois 
2. , pois:



1. 







1. 



1.  onde  é uma constante.



Essas propriedades dão origem as seguintes propriedades da variância:

1.  pois 
2.  pois 
3. 



1. 

Corolário: 



**Variáveis aleatórias independentes:**

Se os eventos  e  são independentes então . Neste caso então:

 e 

**Experimentos independentes:**

Suponha que o espaço dos eventos da v.a.  seja  e o espaço da v.a.  seja , e que ao realizar um experimento conjunto, cujos eventos pertencem ao espaço amostral , o resultado de um não interfere no outro. Matematicamente estamos afirmando que:

e 

Então as v.a.s  e  são independentes. Exemplo de v.a.s independentes: lançar dois dados de cores diferentes simultaneamente e definir  como o resultado de uma cor e  como o resultado da outra cor. O resultado de um dado não interfere no resultado do outro dado. Exemplo de v.a.s não independentes: pintar metade das faces de um dado de uma cor e a outra metade de outra cor. Nesse a cor e a numeração do dado estão associadas e o resultado numérico interfere no resultado da cor. Por exemplo se o resultado para  foi 1, o resultado para  jamais poderá ser 1.

**Teorema 1:** Se  e são independentes, então  e  também são independentes.

Prova: se  e  são independentes, quaisquer dois sub-conjuntos de  e  serão independentes. Assim  e  é a condição para poder calcular as funções  e . Portanto se  e são independentes, então  e  são independentes.

**Teorema 2.** Se  e são independentes, então .



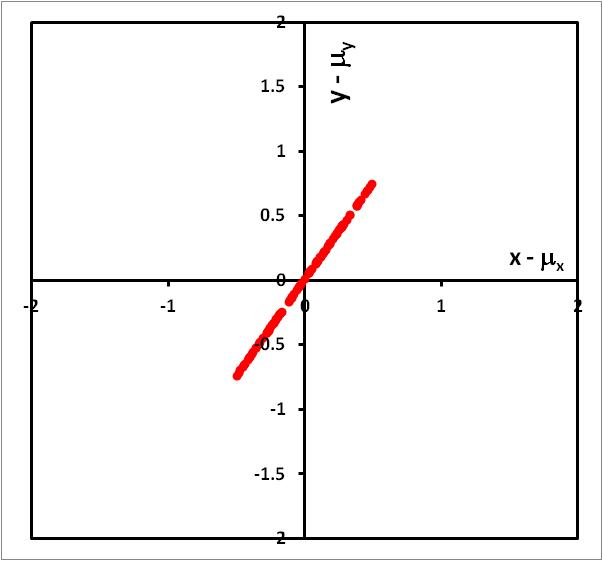
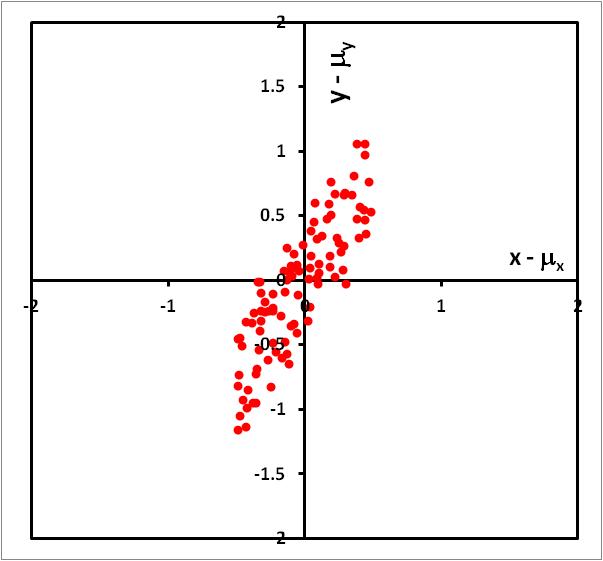
**Teorema 3.** Se  e são independentes, então .



A covariância, portanto, nos fornece alguma informação sobre a independência entre v.a.s. Se  então  e  são independentes. O que ocorre se  ou ?

Note que os produtos  e  em um gráfico  ou  serão positivos no primeiro e terceiro quadrantes, , e negativos no segundo e quarto quadrantes, .

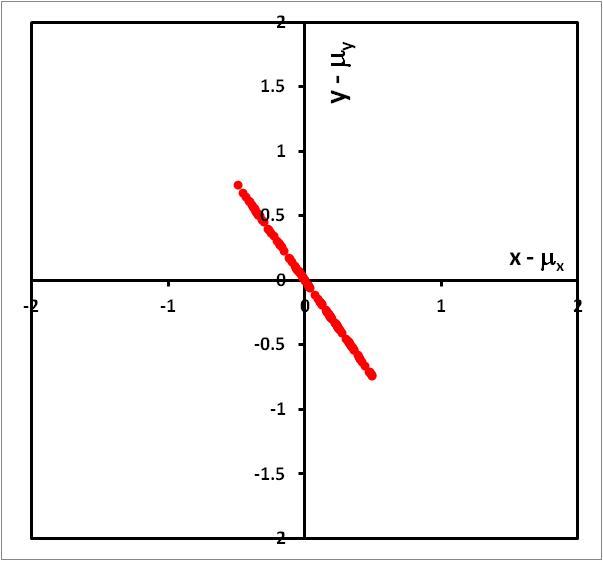
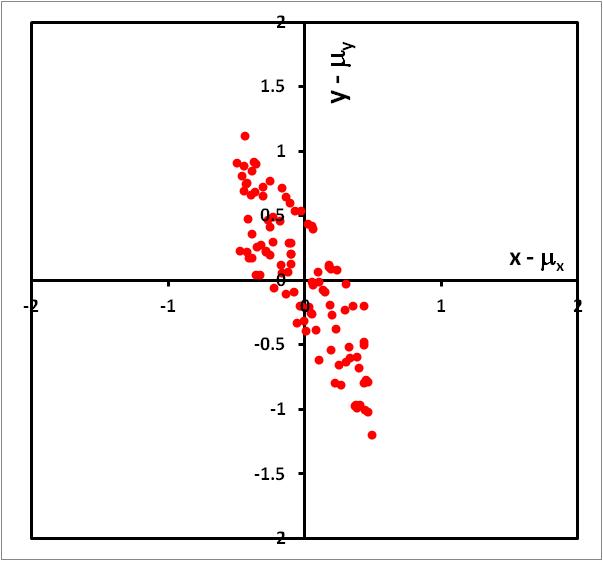
A figura xxx (a) mostra uma nuvem de pontos com uma concentração maior de pontos no primeiro e terceiro quadrantes, terá  positiva, ou seja, com uma covariância positiva. Percebe-se dessa nuvem que a v.a.  tende a crescer quando a v.a.  cresce, e a decrescer quando  decresce. O espalhamento da nuvem informa que essa tendência não é perfeita é que existe algum grau de independência estatística da v.a.  em relação à v.a. . Nesse caso afirmamos que as v.a.s  e  são positivamente correlacionadas. O gráfico da figura xxx(b) mostra o caso em que , totalmente dependente, ou totalmente correlacionas, e se percebe a reta perfeita em que nenhum dos pontos se desvia da reta.



(a) (b)

Figura xxx. (a) caso de duas variáveis positivamente, mas não perfeitamente, correlacionadas. (b) Caso de duas variáveis positivamente e perfeitamente correlacionadas.

Já a figura xxx (a) mostra uma nuvem de pontos com uma concentração maior de pontos nos segundo e quarto quadrantes, com  negativa, ou seja, com uma covariância negativa. Percebe-se dessa núvem que a v.a.  tende a decrescer quando a v.a.  cresce, e a crescer quando  decresce. Nessa situação afirmamos que as v.a.s  e  são negativamente correlacionadas. O gráfico da figura xxx(b) mostra o caso em que , perfeitamente anti-correlacionada, em que nenhum dos pontos se desvia da reta negativamente inclinada.



(a) (b)

Figura xxx. (a) caso de duas variáveis positivamente, mas não perfeitamente, correlacionadas. (b) Caso de duas variáveis positivamente e perfeitamente correlacionadas.

Se as v.a.s são independentes então a nuvem se espalha igualmente pelos quatro quadrantes levando a  como mostra a figura xxx.

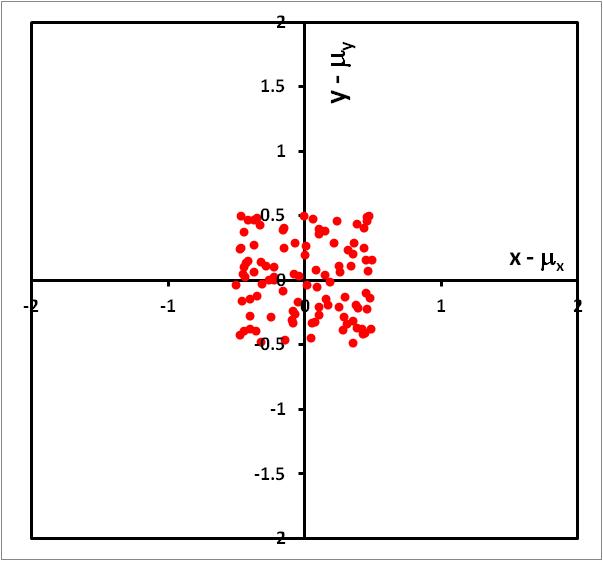


Figura xxx. Caso de duas variáveis descorrelacionadas.

**Coeficiente de Correlação:**

A medida da covariância como uma medida da independência entre duas v.a.s, entretanto, apresenta alguns problemas. Primeiro trata-se de uma medida com dimensão, . Se  e têm dimensão de distância, ou massa, por exemplo, a covariância terá dimensão de área, ou massa ao quadrado. Precisamos de uma grandeza adimensional relacionada à covariância para ser utilizada como um grau de independência entre v.a.s. Então vamos construir o coeficiente de correlação adimensional definido por:



Com essa definição ganhamos mais do que simplesmente a obtenção de uma grandeza adimensional porque podemos mostrar que se trata de um número que varia entre +1 e -1, com zero significando independência estatística, +1 correlação positiva perfeita e -1 correlação negativa, ou anti-correlação, perfeita.

**Teorema do coeficiente de correlação: .**

Prova usando a desigualdade de Schwartz:

 pois se trata da esperança de uma quantidade positiva. Desenvolvendo o quadrado temos:



Logo

que pode ser escrito em termos das variâncias e covariâncias como:



Isso nos leva à desigualda da equação quadrática em  dada por:

 com  e 

A desigualdade  com  só pode ser satisfeita se  não admite raízes reais ou apenas uma raiz que toca o eixo . Essa condição implica que . Agora fazendo ,  e  percebe-se que  ou seja,  que implica em .

Esse teorema pode ser generalizado e utilizado para definir ortogonalidade entre v.a.s.

**Teorema generalizado para independência entre v.a.s: .**

Basta fazer o mesmo começando com  que nos leva diretamente à  e, consequentemente, a . O fato de que esse é um número entre -1 e +1 significa que sempre existirá um ângulo  para o qual . Se definimos , ou seja root-mean-square, porque utilizamos  como estimador de , podemos afirmar então que:

 ou 

Em que o coseno mede o grau de relação entre as v.a.s  e . Se , mas e  então  e dizemos que  e  são ortogonais entre si, ou seja, .

**Adição de v.a.s independentes:** se  em que  e  são v.a.s independentes com fdp´s  e , então a nova v.a.  terá a fdp dada por .

Prova: 

Então[[4]](#footnote-4): 



**Convolução e Correlação:** A operação entre duas funções  e  definida por  é tão importante que ganhou nome próprio: é chamada de CONVOLUÇÃO e é simbolizada por . Ela tem uma prima denominada por operação CORRELAÇÃO definida de forma um pouco diferente por . Note que a diferença está no argumento da função , o qual na convolução é  e na correlação é .

**Intuição sobre as operações convolução e correlação:**

Note que a operação  é simplesmente transladar a função  no eixo horizontal pela quantidade  para a direita. Já a  translada a função para a esquerda. A figura xxx mostra a função , preta, com a  em azul e a  em vermelho. Note que a curva azul deslocou de 2 para a direita e a vermelha de 2 para a esquerda. Já a operação  significa uma reflexão da função em torno do eixo . A figura xx mostra o gráfico das curvas  e .

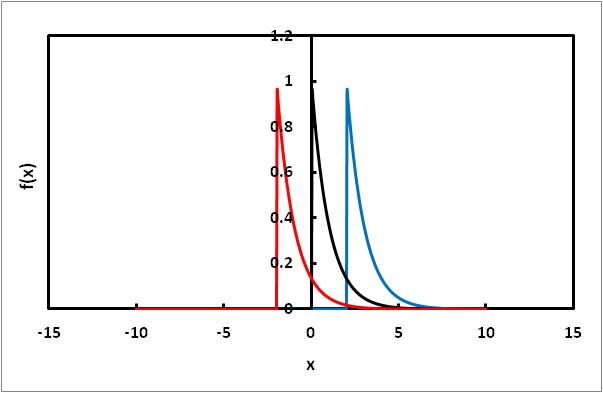


Figura xxx. Gráfico das curvas  em preto,  em azul e  em vermelho.

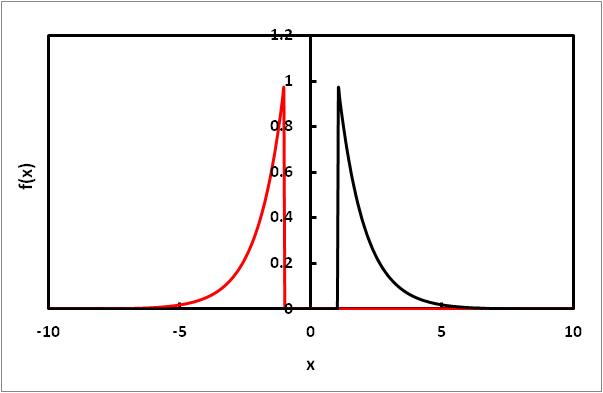


Figura xxx. Gráfico das curvas  em preto e  em vermelho.

Vamos analisar uma auto-convolução e uma auto-correlação da função  com ela mesma. Na auto-correlação a  é a própria função deslocada por . Mas na auto-convolução  a função é deslocada e refletida no eixo .

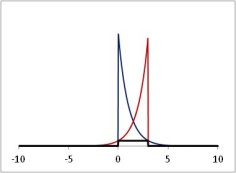
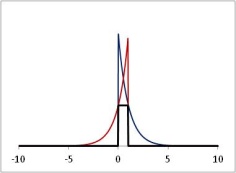
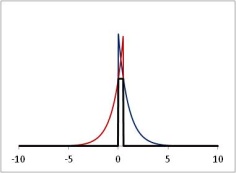
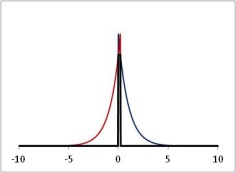
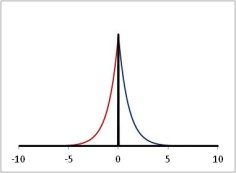
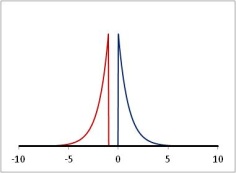


Figura xxx. Multiplicação das curvas  por  para 

A figura xxx mostra a curva da autoconvolução  em função de .

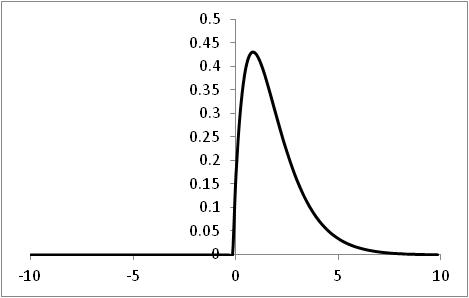


Figura xxx. Autoconvolução de  em função de 

Já a figura xxx mostra a multiplicação de  por  da auto-correlação e a figura xxx o resultado da auto-correlação em função de .

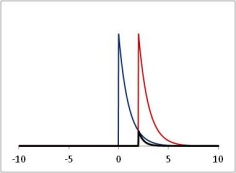
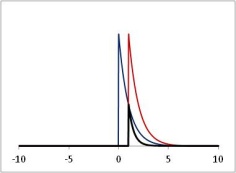
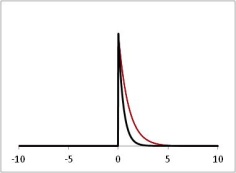
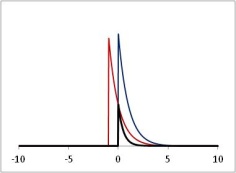
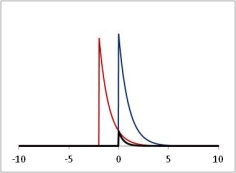


Figura xxx. Multiplicação das curvas  por  para 

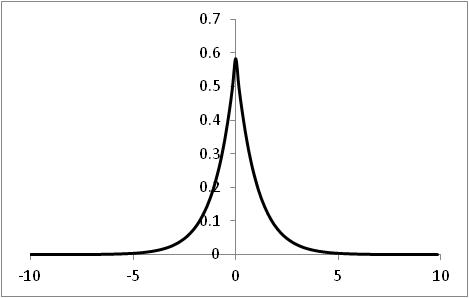


Figura xxx. Auto-correlação de  em função de 

**FGM e Função Característica de v.a.s independentes:**

Se as v.a.s  e  são independentes então . Nesse caso então 

Da mesma forma:



Ou seja a função geradora dos momentos e a função caraterística da v.a.  serão os produtos das respectivas funções de cada uma das v.a.s.

**Teorema da convolução:** Daqui podemos extrair o teorema da convolução afirmando que:

Sejam: ;  e .

Então  é dado por .

**O truque do logaritmo:**

A expansão em série de Taylor-McLaurin da função  pode ser feita notando que, e . As derivadas de ordem superior a um podem ser facilmente calculadas usando: , para obter . Desse resultado mostramos que:



e:

.

O truque do logaritmo é muito útil em casos em que a convergência da série de Taylor é problemática. Suponha o caso da função , com  mas . Melhor dizendo, com  e . Se fizermos a expansão de Taylor-McLaurin para esta função, obteremos:

.

Cuja convergência depende se o produto  é maior ou menor do que 1. Em lugar de fazer a expansão direta da função vamos expandir seu logaritmo na forma: , que não apresenta problemas de convergência para . Agora retorna-se à função inicial para reescreve-la como . O truque do logaritmo levou à definição da função geradora dos cumulantes.

**Cumulantes:**

A função geradora dos cumulantes é dada por . Note que os cumulantes se acumulam, por isso o nome cumulante. Se  e  são independentes então  e:



Comparando com a série de Taylor vemos que . É diferente do , por causa do logaritmo. Podemos extrair a relação entre os cumulantes e os momentos derivando o logaritmo pela regra da cadeia e lembrando que :

1. ; 
2. ; 
3. ; 
4.  logo  então que pode ser colocado em termos dos momentos centrados como . Logo .
5. , ou seja: . Colocando em termos dos momentos centrados  finalmente . .

**Resumo das relações entre os momentos centrados, não centrados e cumulantes até ordem 4:**

|  |  |  |
| --- | --- | --- |
| **Momentos não centrados** | **Momentos centrados** | **Cumulantes** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**Aplicações:**

**Distribuição de Bernoulli:**

Jogar a moeda, só temos duas possibilidades, cara ou coroa. A v.a. será definida como cara = 1 e coroa = 0. Qualquer jogo com apenas duas respostas, sim = 1 e não = 0, segue uma distribuição de Bernoulli. Se a probabilidade de SIM é , a de Não será  e a função densidade de probabilidade é dada por: . A função distribuição de probabilidade acumulada vale:



A funçao geradora dos momentos é dada por: . A função característica .

Momentos:  logo , então . Momentos centrados: .

Agora , ou seja,  ou ainda .

Casos particulares:

1. ;
2. ;
3. ;
4. .

Cumulantes:

, então  logo:

1. . Logo 
2. , logo .
3.  logo, .
4. . Após alguma álgebra temos , logo  ou .

**Distribuição Binomial:**

Vamos jogar a moeda n vezes de forma independente. Nesse caso a v.a. soma são i.i.d., e a função característica vale: . Sabendo a  queremos a  dada por . Expandindo em binômio de Newton temos:



Aqui vale a acumulação dos cumulantes , então , ,  e .

**Distribuição de Poisson:**

Essa distribuição é um caso limite da binomial quando , mas  de tal forma que o produto  é constante. Agora . Nesse ponto usamos o fato de que  para achar . Agora queremos a fdp: 

.

Uma expressão para .



Então  .

Cumulantes: , portanto todos os cumulantes valem , daí , , e , a skewness vale , skewed to the right, e a curtose , sempre leptocúrtica. Se então a skewness e curtose tendem a zero.

**Distribuição Normal:**

Vamos fazer o limite de n tendendo a infinito na distribuição binomial e usar o truque do logaritmo. Nesse caso:

. Mas já sabemos que  e vamos truncar a série na ordem 2. Chamando  temos:



Logo  e essa é a distribuição normal, cuja função característica vale: . Note que, nesse caso, só existem dois cumulantes, pois ,  e , todos os outros são nulos. Os momentos centrados são dados pela função geradora:

.

Percebe-se que não existem momentos ímpares e que os pares valem . Comparando extraímos . Podemos reescrever esse resultado em termos dos fatoriais duplos . Notando que  e, além disso, que:



logo, substituindo  chegamos na expressão mais simples . Então vemos que ; ;  e assim por diante.

Falta a função densidade de probabilidade:  . O truque aqui é completar quadrado no expoente , obtendo . Substituindo de volta na integral temos:



Finalmente obtemos a função densidade de probabilidade da distribuição Normal:



A Normal Padrão tem esperança nula e variância unitária dada por  . A Normal padrão cumulativa é definida como . Note que é sempre possível escrever o resultado de uma normal cumulativa em termos da  por uma mudança de variável. Se queremos a função distribuição de probabilidade cumulativa de uma normal com  e , ou seja:  a mudança de variável  nos leva a . Por isso as tabelas da normal são sempre feitas para a normal padrão usando como argumento o desvio da esperança medido em desvios padrão .

**Distribuição Log-Normal.**

A distribuição Log-Normal é obtida da Normal através da mudança de variável . A regra para mudança de variável é dada por, com a somatória sobre todos os x’s possíveis para as raízes da equação , ou, . Neste caso a função é biunívoca e só existe uma raiz dada por , . Vamos precisar da derivada . Logo:



A log-Normal como uma aproximação da normal:

Vamos reescrever a log-normal como  , , e ver o que acontece para  com .

Neste caso  e o termo no expoente é aproximado por . No denominador simplesmente fazemos  e vemos que  é uma normal da variável . Se fizermos teremos duas curvas muito semelhantes no caso em que . Note que se  o anulando a função. Grandes diferenças, portanto, entre a normal e a log-normal ocorrerão quando a probabilidade de valores de x negativos na normal forem grandes. A figura xx mostra esse comportamento:

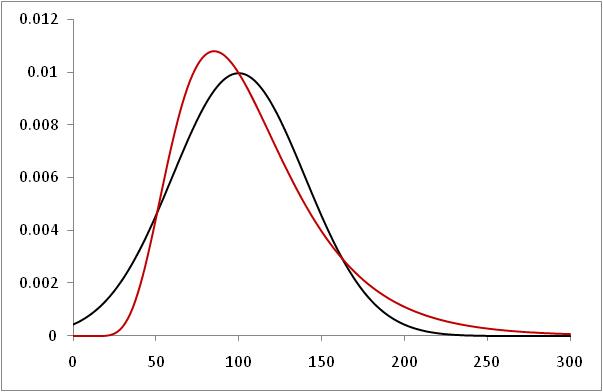
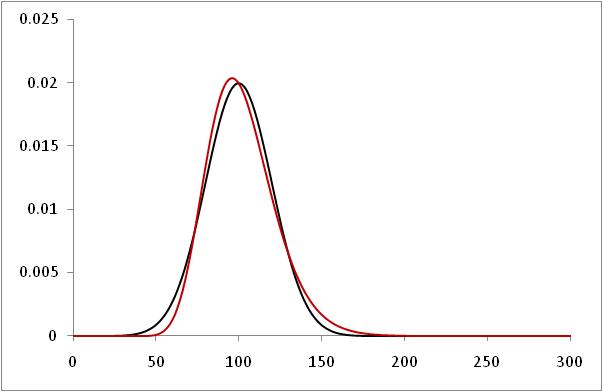
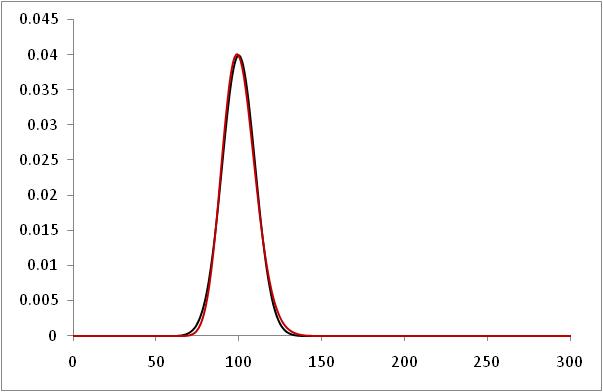


Figura xxx. Normal com . (a)  e ; (b)  e ;(c)  e .

Tanto a função geradora dos momentos quanto a função característica apresentam problemas de convergência, mas podemos calcular os momentos da Log-Normal  mudando a variável de integração para , , , quando  e quando . Nesse caso: . O truque aqui é completar quadrado no expoente:  continuando . Daí vemos que . A integral entre colchetes vale 1 e temos todos os momentos de ordem n dados por . Em particular temos ; ; ;  e . Podemos calcular os momentos centrados usando binômio de Newton  . Já sabemos que  e . Para a variância m2 temos:  logo ,  e . Para o momento centrado de ordem 3 temos  de onde extraímos que e . Fatorando o termo entre colchetes ainda chegamos a .

**Teorema Central do Limite e o truque do logaritmo:**

Agora vamos tomar uma variável aleatória  dada pela adição de  v.a. independentes no limite . Sabemos que . Note que se as v.a. fossem, além de independentes, idênticas [i.i.d.], teríamos , um caso semelhante ao utilizado no truque do logarítmo. Também sabemos que  e que . Um número menor do que 1 elevado à uma potência muito alta tende a zero. Mas não em  porque , o que significa que a função  se torna concentrada em torno de , caindo a zero para fora desse intervalo. Com isso podemos fazer uma expansão em série de Taylor-McLaurin da função característica, mas usando o truque do logarítmo, . Mas essa é a expansão dos cumulantes . Se as v.a. não são idênticas a expansão em Taylor agora será dada por .

**Teorema Central do Limite:**

Truncando a expansão até segunda ordem em  temos . Mas  e , logo  que é a função característica de uma normal com  e . Se as variáveis são independentes e idênticas [i.i.d.] então  e  e a distribuição tende para uma normal com  e . Então notamos que a variável  tem esperança  e desvio padrão , ambos crescendo com .

Vamos usar agora  em lugar de  . Note que nesse caso ,  e , portanto a nova fdp será  e a nova função característica será:



Nesse caso a função característica da distribuição da média será dada por:



que é a função característica da Normal , com  e . Note que se as v.a. são iid então  e . Agora e esperança fica parada e o desvio padrão vai diminuindo com o aumento de . Para manter os dois parados fazemos a última mudança de variável ,  e , logo  e a nova função característica será dada por:



Ou seja  que é a função característica da distribuição Normal padrão .

A melhor forma, portanto, de especificar o Teorema Central do Limite é afirmando que:

Se  e  são independentes, e  existem e são finitos então a v.a.:



Note que não foi necessário que as v.a. fossem idênticas ou que sigam uma distribuição normal mas apenas que a média seja feita em um número muito grande de v.a.s.

Se as v.a.s são iid então:



**Cuidados com o Teorema Central do Limite.**

Primeiro cuidado é em relação as condições de validade do teorema: momentos de ordem 1 e 2 finitos. Se o momento de ordem 2 for infinito então o teorema pode falhar. Entretanto, mesmo no caso em que a variância é finita, garantindo a validade do teorema, cuidados extras são necessários para o comportamento das caudas. Note que o TCL depende da validade da expansão dos cumulantes, que truncamos na ordem 2. Isso significa que a região central, próxima do pico, vai coincidir com a Normal, mas essa aproximação vai se tornando pior nas caudas, bem longe do pico. No limite de  o teorema é 100% válido, mas dado um número grande  mas finito, a região de validade é um função de  que vai com  se a skewness é diferente de zero ou  se apenas a curtose existe.

Quem se interessa pelas caudas? A probabilidade nas caudas é obviamente pequena, mas para muitas situações é essa probabilidade que interessa. No caso em que a probabilidade é muito pequena mas os efeitos do evento são devastadores o estudo das caudas é fundamental.

O gráfico da figura xxx mostra os cuidados necessários com TCL. A curva azul é uma distribuição generalizada de Student para  e a vermelha a distribuição normal padrão. A curva preta mostra a distribuição para . Nota-se que na região central a curva preta se superpõe com a curva da distribuição normal até determinada distância quando se afasta se torna uma reta paralela à da curva azul.

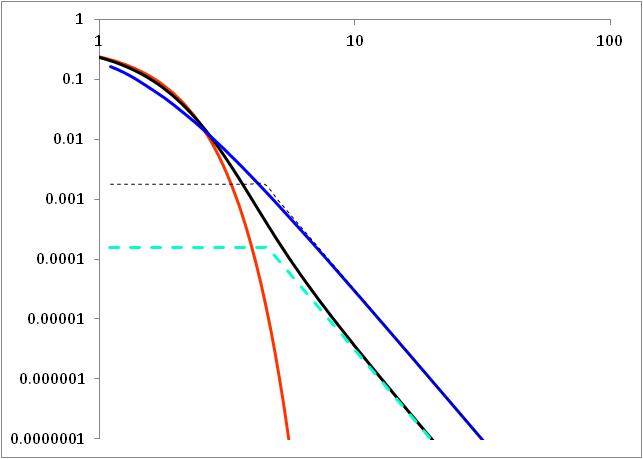


Figura 3. Curvas da distribuição de Student generalizada para , após a adição de 10 v.a.s, da distribuição normal padrão e das leis de potência das caudas em gráfico log-log.

#### Processos estocásticos

#### Processos estocásticos aditivos

Sabemos que para  a binomial converge para a normal9: . Como n representa o número de períodos vamos expressá-lo simplesmente como o tempo . Suponha a situação aditiva em que  com probabilidade , ou  com probabilidade , sendo . Seja  o número de passos Up e  o número de passos Down. Depois de  períodos o preço será . Mudando a variável para  temos que , onde . Além disso, , logo, o preço da ação após  períodos segue o Movimento Browniano, ou processo estocástico de Wiener, dado pela normal . Esse processo é conhecido como Movimento Browniano ou processo estocástico de Wiener.

#### Processos estocásticos multiplicativos

Suponha agora o processo multiplicativo  com probabilidade  ou  com probabilidade . Depois de  passos o preço será . Tirando o logaritmo de ambos os lados  e . Nesse caso  e o preço da ação segue um movimento Browniano Geométrico dado pela log-normal: .

Note as diferenças entre os dois processos, o Browniano e o Browniano Geométrico:

 e 

O capítulo 2 apresentou as propriedades das distribuições Normal e log-Normal. No primeiro caso  e . No segundo caso  para a qual vale a regra .

**Apêndices**

1. **Propriedades da Função Delta de Dirac:**
2.  com 

Para mostyrar essa propriedade basta fazer a integral por partes:  logo:  e . No nosso caso  e , portanto  e . Então:

 CQD

1. 

 por outro lado:

 que nos leva a  ou seja  e  CQD.

1.  onde  são as raízes da , ou seja, .

Sabemos que  em torno de uma raiz de . Suponha que , que é o caso mostrado na figura xx (a). Nesse caso  onde  e . Então:  logo .

Por outro lado se  nós temos o caso da figura xxx (b) com . Nesse caso:



logo .

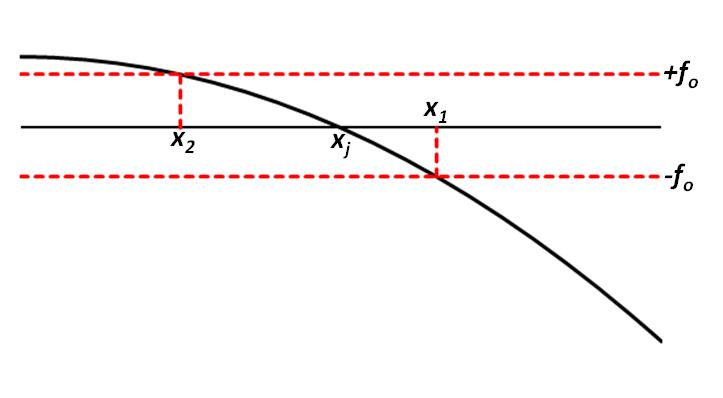
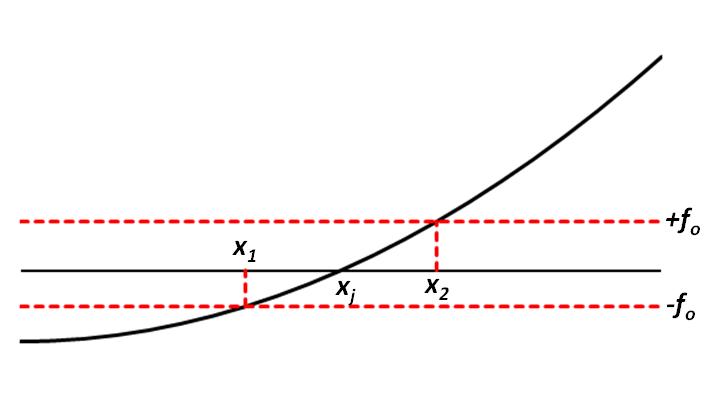


Figura xxx. (a) caso em que  e . (b) caso em que  e .

Podemos juntar os dois casos afirmando então que . Finalmente, devemos somar em todas as raízes, obtendo:



Casos particulares:

1. 
2. 
3. **Função Delta Gaussiana, ou da Distribuição Normal.**

Vale a pena mostrar como uma Gaussiana se transforma na função Delta de Dirac porque precisaremos do resultado da seguinte integral:



Para demonstrar esse resultado definimos . Mas como a variável de integração é muda, então  logo . Agora podemos mudar de sistema de coordenadas cartesianas para polares no qual , ,  e o elemento de área vale . Para fechar todo o plano x-y, r varia de zero a infinito e  de 0 a 2. Neste caso . A integral em  é imediata e ficamos com . Agora mudamos a variável para  logo e ficamos com . Mas , então  e  , ou seja, .

Agora podemos mostrar que . Percebe-se que a largura da função fica cada vez menor à medida que n cresce e que a área sobre a curva vale sempre 1, pois:



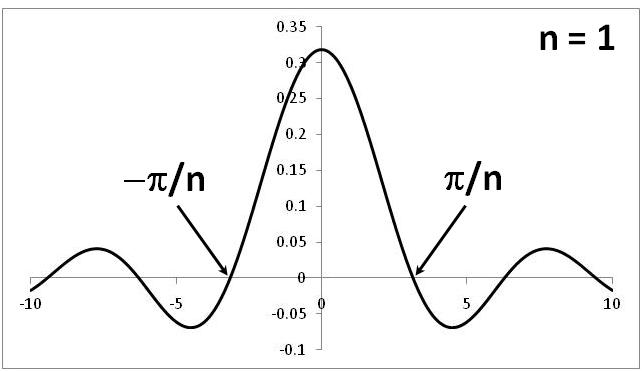
Dessa forma .

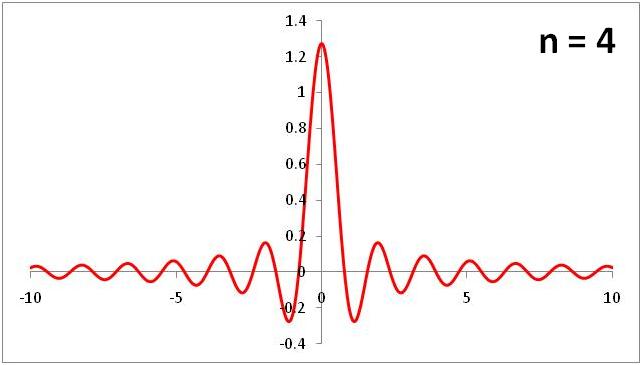
A distribuição Normal com parâmetros  e  segue a forma . Deve-se colocá-la dessa forma porque x,  e  possuindo a mesma dimensão tornam a fração  adimensional. Para ser uma fdp é necessário que . Podemos achar o valor de A que torna essa igualdade verdadeira fazendo a mudança de variável , , logo , logo  e a fdp da distribuição Normal é dada por:



1. **Função Delta Especial**

Uma função delta especial pode ser obtida da seqüência  cuja curva é mostrada na figura xxx.





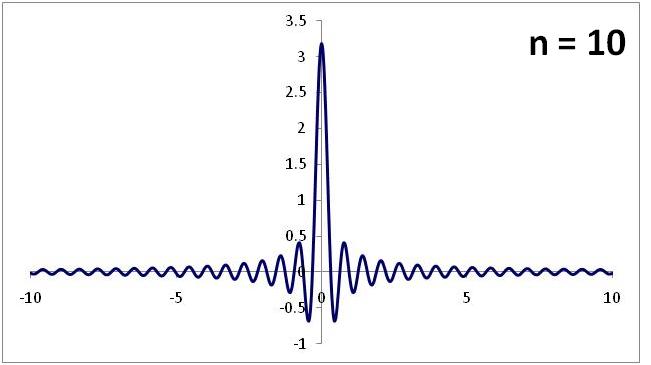


Figura xxx. Gráfico da funçãov para  e . Note que a altura sobe com n e a largura diminui. A distância entre as duas primeiras raízes vale .

Vamos ver a área sobre essa curva:



Precisamos mostrar que integral  converge. Note que é uma função par e portanto . Por causa do do denominador de as áreas entre duas raízes se tornam cada vez menores, como mostra a figura xxx para valores de u positivos apenas. Usando esse fato percebemos que , mas, por outro lado, que  . Então sabemos que , logo a integral converge. Sabendo que  podemos afirmar usando o retângulo de altura 1 e largura  que .

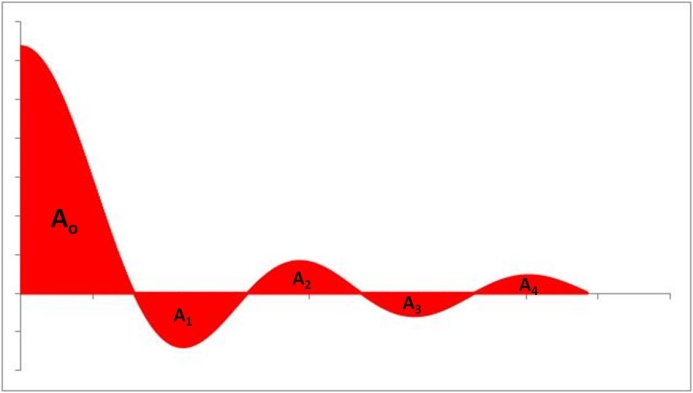


Figura xxx. área sobre a curva . Quebrando as áreas entre duas raízes se nota que as áreas pares são positivas e as ímpares negativas e que as mesmas vão diminuindo com a distância .

No apêndice 5 mostraremos, usando cálculo de resíduos, que . Portanto, para garantir que a área seja unitária com esse resultado precisamos fazer .

Dessa forma a função  se torna a função delta de Dirac no limite :



Por outro lado podemos usar a fórmula de Euler  para calcular . Daqui extraímos a identidade super importante:



Essa identidade será muito útil na transformada inversa de Fourier, no teorema da convolução e no teorema central do limite.

Nota: Seria possível dar a volta no cálculo de resíduos sem especificar que o estamos utilizando, porém trata-se de ferramenta tão poderosa que vale a pena dominá-la, sobretudo para trabalhar com as transformadas de Fourier e as funções características.

1. **Cálculo de variáveis complexas:**

Definimos uma função de variável complexa  em que . A função se chama analítica se for diferenciável. Note entretanto que estamos agora falando de um limite em duas dimensões. O limite só existe se for o mesmo por qualquer caminho.

Condições Cauchy-Riemann para funções analíticas:



Vamos fazer esse limite por dois caminhos:

1. , ou seja,  , e . Nesse caso .
2. , ou seja,  , e . Nesse caso .

A função será diferenciável se , ou seja:  e , que são as condições de Cauchy-Riemann. Note que aqui só provamos que se tratam de condições necessárias para que o limite exista, mas poderiam não ser suficientes. Afirmamos sem provas que também são condições necessárias e suficientes. Então afirmamos que:

 é analítica se  e .

Teorema: se  é analítica então , onde  é qualquer caminho fechado no plano complexo .

Basta fazer a integral no caminho infinitésimal:



mostrado na figura xxx.

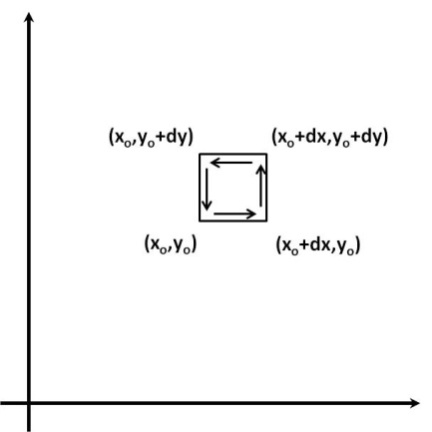


Figura xxx. Circuito infinitésimal para cálculo de 

: 

:



:



:



Portanto:









Agora  e  pela condições de Cauchy-Riemann. Então:



Agora esse resultado pode ser estendido para qualquer caminho  porque podemos quebrar o caminho em sub-caminhos infinitésimais, cancelando os percursos internos e restando apenas o caminho externo, como mostra a figura xxx.

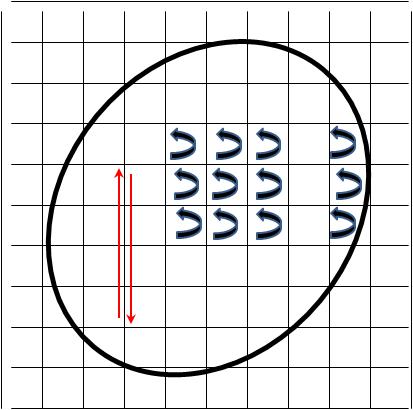


Figura xxx. Note que no interior da região as integrais de caminho se anulam porque enquanto o percurso de um célula está em uma direção o da vizinha está na direção oposta. Esse cancelamento, entretanto, não ocorre na fronteira pois não existe a célula vizinha.

A condição para a validade desse teorema é que a função seja analítica na região envolvida pelo caminho. Entretanto, nos pontos de singularidades a função não é analítica. A figura abaixo mostra como contornar a singularidade escolhendo um caminho apropriado.

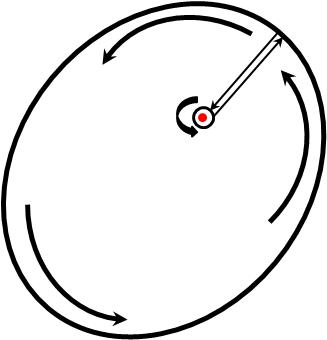


Figura xxx. Isolando uma singularidade do caminho de integração.

Suponha que podemos expandir uma função de variável complexa da forma:



Não é analítica em  por conta dos termos com potências negativas de . Nesse caso dizemos que a função tem um polo de ordem . O coeficiente  é chamado de RESÍDUO. Porque ele é tão importante?

Vamos fazer a integral em torno de . A convenção é que giramos no sentido contrário ao dos ponteiros do relógio, mesmo sentido de crescimento do ângulo  das coordenadas polares. Agora fazemos: , com . Logo  e . Nesse caso:



Independente do valor de , logo o resultado do limite  será o mesmo.

Entretanto note que . Aqui nosso receio é o de que o  do denominador que não cancela com o  do  faria a integral explodir. Entretanto, antes de tomar o limite de  vamos fazer a integral. Usamos a mesma mudança de variável , com . Logo  e . Dessa forma:



Que nos leva a 

pois .

Isso então nos leva ao seguinte resultado:



Para as potências positivas a integral é zero porque a função é analítica. Para as potências negativas só não é nula para o termo com  do resíduo. Então chegamos ao resultado:



Se existirem mais de um ponto de singularidade dentro do caminho de integração o resultado final é:



Esse é o resultado que utilizamos para calcular muitas integrais mesmo no eixo real. Sobra a pergunta: como descobrir o resíduo, ou os resíduos? Suponha uma função com polo de ordem :



Se multiplicamos essa função por  ela se torna analítica, diferenciável, portanto.



Agora derivamos essa função  vezes. Todos os termos com potência  serão nulos, e todos os termos com potência  terão o termo  que vai a zero quando . O único termo que sobra é para . Então:



E o resíduo será: .

Se o polo é de ordem 1, também chamado de polo simples, então  e:



1. **Mostrar que****:**

Com o cálculo de resíduos essa tarefa simplifica. Antes de mais nada fazemos: . Depois fazemos o cálculo da seguinte integral: 

Agora  porque para  então  e o como  para  o termo  anula tudo no limite .

Por outro lado:



O resíduo do polo simples em  vale 

Juntando tudo temos:

 ou seja: 

Portanto: 

Resultado final: 

1. Se a fronteira entre o que pertence e o que não pertence ao conjunto é NEBULOSA, não claramente definida, aceita uma gradação, o con junto é nebuloso, ou FUZZY. Existe toda uma lógica, chamada FUZZY LOGIC, para lidar com esses caso hoje. [↑](#footnote-ref-1)
2. Vamos evitar a letra C para conjuntos por que é a letra usada para estar contido. [↑](#footnote-ref-2)
3. A função delta de Dirac só deve ser usada para descontinuidades finitas, ou seja, para funções com variações finitas. No caso das distribuições as descontinuidades são todas finitas e a representação da derivada da descontinuidade como a funçaõ delta é sempre válida. [↑](#footnote-ref-3)
4. Estamos usando a seguinte regra para derivar integrais:  onde  portanto  ou seja . [↑](#footnote-ref-4)