**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 .

Prova: , então  logo . Troca o nome das v.a.s e o teorema continua válido. Para a segunda parte basta notar que  logo .

1. ****

Para mostrar isso basta usar a seguinte partição: , logo , ou seja:



De modo análogo é claro que 

1. ****

Vamos tomar a partição:



Então:





1. ****

Fazendo , ,  e  em (3) e usando o fato de que:



Então



Note que  sempre.

Densidade de probabilidade conjunta [joint density probability]

Definimos a fdp conjunta agora como: .

O reverso é dado por:

Prova do reverso:  e  logo:



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 distrib uições 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: . O apêndice xxx traz a demonstração dessa regra. Podemos aplicá-la ao caso da transformação de coordenadas retangulares para polares, em que  e . Portanto:

.

Isso significa que .

**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: 



**Matriz de variância-covariância:**

Definindo a matriz : , como a matriz de variância-covariância percebe-se que os termos da diagonal são as variâncias de cada v.a. específica.

**Propriedades da matriz de variância-covariância:**

Nota: com os dados reais só podemos calcular a matriz , onde  foi o valor da késima observação da v.a.  e , em lugar da matriz . Ou seja estamos substituindo a operação esperança . Nesse caso  é um estimador de  que só seria matematicamente idêntico se a média fosse tomada com infinitos pontos, impossível na prática. Entretanto, esse fato não muda as propriedades da matriz de variância-covariância, quer sejam definidas como  ou .

1. É simétrica: 
2. É uma matriz definida positiva [Ver apêndice de matrizes]:

Uma matriz é definida positiva se para qualquer vetor  não nulo o produto: , é positivo.

Prova: considere . Obviamente que  e que . Então, o fato de que , implica que 

**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 .



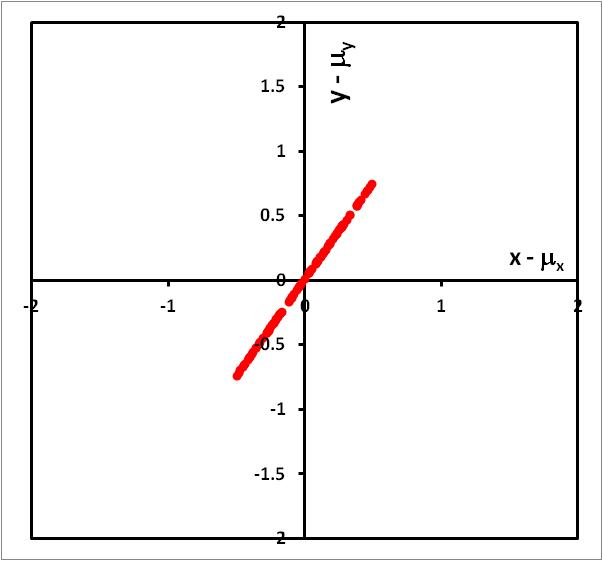
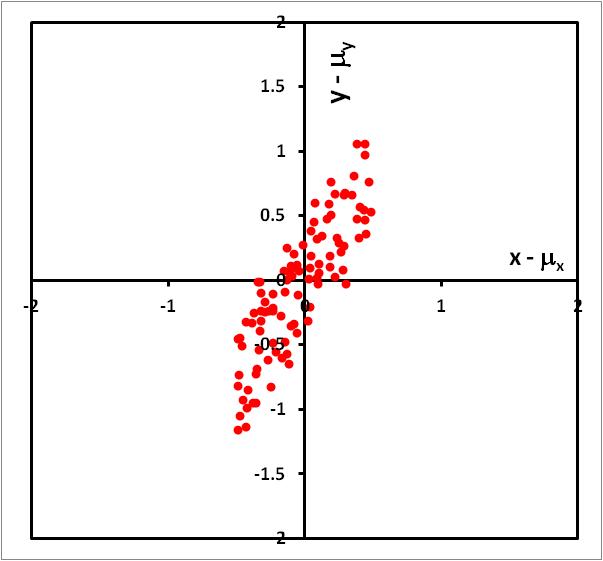
**Teorema 3.** Se  e são independentes, então .



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 segundoe quarto quadrantes, .

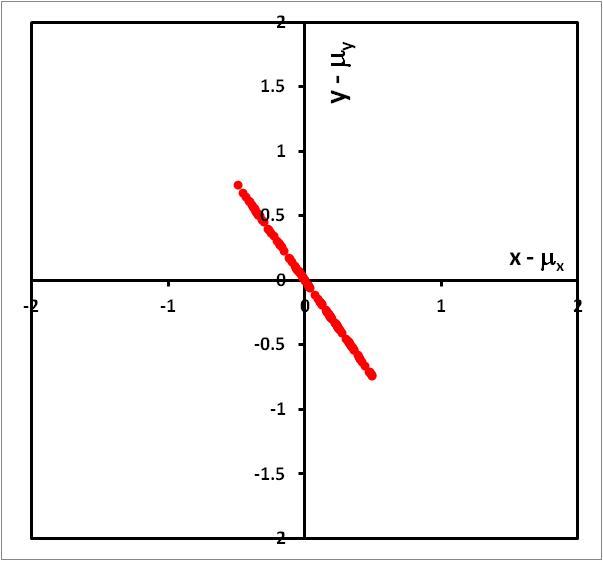
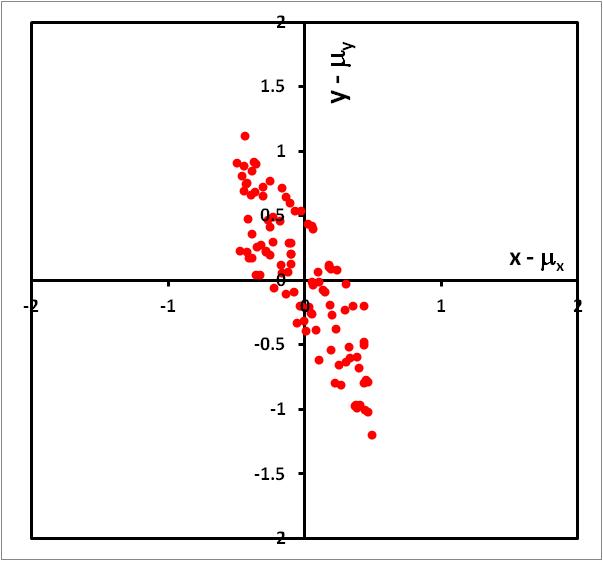
A figura xxx (a) mostra uma núvem de pontos com uma concentração maior de pontos no primeiro e terceiro quadrantes, terá  positiva, ou seja, com uma covariância positiva. Percebe-se dessa núvem que a v.a.  tende a crescer quando a v.a.  cresce, e a decrescer quando  decresce. O espalhamento da núvem 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 núvem 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 , perfeitament 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 núvem se espalha igualmente pelos quatro quadrantes levando a  como mostra a figura xxx.

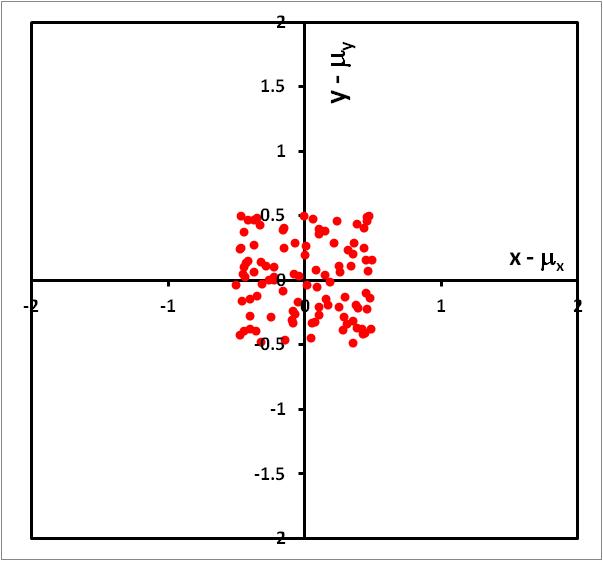


Figura xxx. Caso de duas variáveis descorrelacionadas.

**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, .

**Espaços métricos e distância de correlação:**

Um espaço é métrico se existe uma função distância  para  satisfazeno aos axiomas:

1. Desigualdade triangular: 
2. Se  então 
3. 

Com esses axiomas podemos demonstrar o teorema:

1. 

Fazer no axioma 1:  usando os axiomas (2) e (3) , logo  .

Então a função distância deve ser um número real e positivo. Se essa função existe então ela é a métrica do espaço e podemos medir distâncias entre os elementos do conjunto . Nesse caso dizemos que o espaço é métrico.

Distância Euclidiana:

A distância Euclidiana entre os vetores  e  é definida como  que já apresenta naturalmente as propriedades  e . Notamos que  onde o produto escalar entre dois vetores é definido da forma anterior como .

Falta mostrar o axioma 1:







Agora usamos o fato de que  para perceber que , ou seja, . Nesse caso:



Logo: .

Essa não é a única distância possível. Existem outras distâncias como a distância Manhattan . É chamada de distância Manhattan, ou distância do motorista de Taxi, taxicab distance, porque em uma cidade quadriculada o motorista nunca pode tomar o caminho da hipotenusa, como mostra a figura xxx abaixo extraída da wikipedia:

[http://upload.wikimedia.org/wikipedia/commons/0/08/Manhattan\_distance.svg].

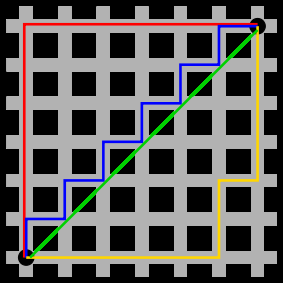


Figura xxx. Manhattan distance. Note que as distâncias vermelha, azul e amarela são iguais.

Em lugar de somar os quadrados das diferenças o motorista tem que somar os módulos das diferenças nas duas dimensões.

Distância p-ádica:

Kurt Hensel em 1897 introduziu a noção dos números p-ádicos. Seja  um inteiro e  um número primo, então:  com .

Exemplo:  e . Nesse caso: , ,  e  já não interessa mais. Assim começamos da potência mais alta e vamos descendo: ,  e , logo: .

Vamos tomar agora , então  e . Nesse caso , , , , logo .

Qualquer inteiro positivo pode ser escrito como:

 e um número primo.

O que acontece se permitimos  negativos? Teríamos também ,  etc e assim poderíamos escrever qualquer número racional como:  com , logo  pode ser negativo. Note dessa definição então que é divisível por , então é a maior potência divisora de . A distância p-ádica é definida como: .

Exemplos:

1. .  com , ,  e . Neste caso .
2. O número ZERO é divisível por qualquer número, portanto .

Operações adição e multiplicação de números p-ádicos:

Suponha agora dois números  e , com  e  então . Fazendo ,  e  então podemos escrever . Nesse caso percebemos que .

Suponha agora dois números  e , com  e  então , onde . O número que ficou entre colchetes  pode ser, ou não, divisível por , ou seja . Neste caso . Assim  e . Dessa forma  como a distância, mas a distância p-ádica apresenta uma restrição maior que é:  . Note que isso significa que , ou seja,  em lugar dos costumeiro . Em 1944 Mark Krasner criou o termo espaço ultramétrico para a distância satisfazendo aos axiomas:

1. Se  então 
2. 
3. 

Um espaço que satisfaz à esses axiomas é chamado de ultramétrico. Note que se trocou a desigualdade triangular pela desigualdade do máximo.

Com a distância Euclidiana podemos definir uma distância de correlação.

Suponha as v.a.s  e , com , , , , em que  e  são os estimadores de  e de  e  e  os estimadroes de  e . A covariância será dada por  e o coeficiente de correlação . Algebricamente vemos que:



Agora vamos definir aa v.a.s padronizadas  e . Notamos que  e, da mesma forma, , que é equivalente ao fato de que . Por outro lado , ou seja, . Da mesma forma . Podemos então pensar em dois vetores unitários.

Agora vamos definir uma distância Euclidiana entre esses dois vetores unitários como:





Mas agora notamos que  logo  e:



Então vemos que a grandeza  se comporta como uma distância. Chamamos essa distância de distância de correlação. Como  a distância de correlação varia entre . Quanto maior a correlação menor a distância.

Vale notar um ponto importante aqui. Para ser uma distância exigimos que se  então . Mas  significa que . Duas v.a.s relacionadas da forma  com  apresentam correlação  embora . Entretanto, as duas variáveis  e  são iguais, pois:

1. , ou seja,  .
2. , ou seja, .
3.  logo .

**Espaços ULTRAMÉTRICOS:**

Partindo dos espaços métricos podemos definir um espaço ultramétrico especialmente adequado para análises de clusters e hierarquias.

**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[[1]](#footnote-1): 



**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 .

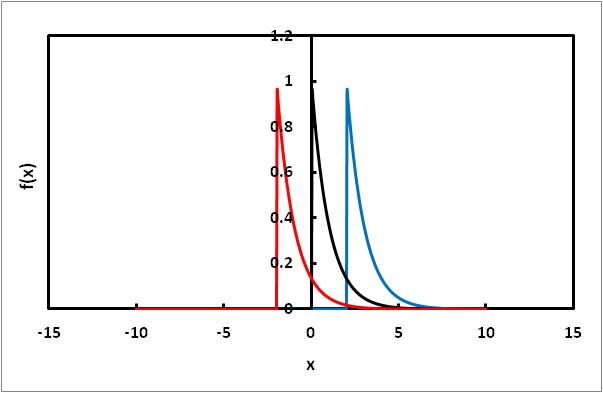


Figura xxx. Gráfico das curvas  em preto,  em azul e  em vermelho.

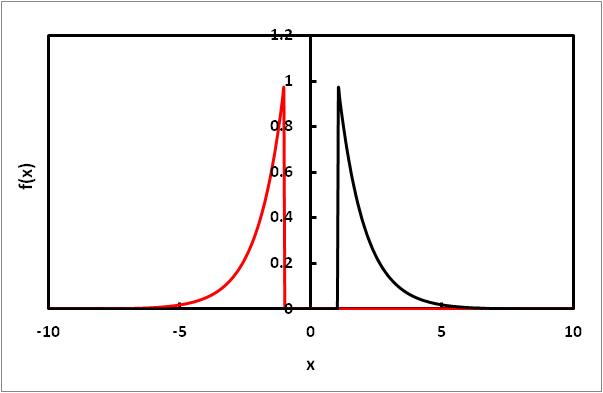


Figura xxx. Gráfico das curvas  em preto e  em vermelho.

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 .

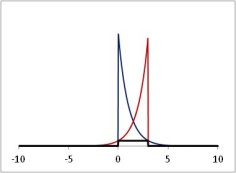
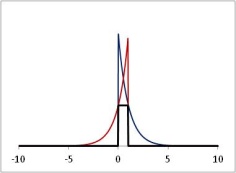
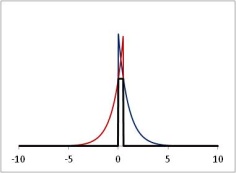
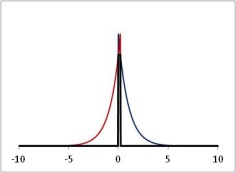
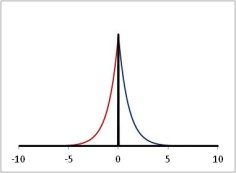
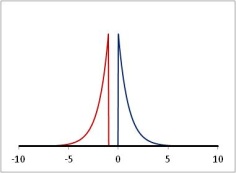


Figura xxx. Multiplicação das curvas  por  para 

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

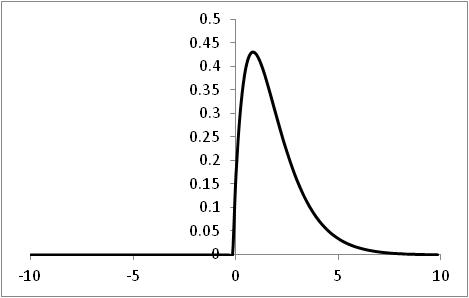


Figura xxx. Autoconvolução de  em função de 

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

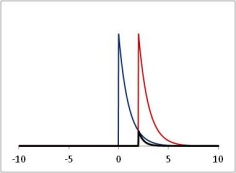
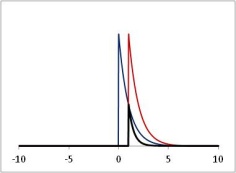
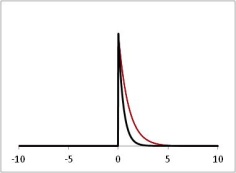
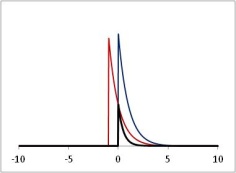
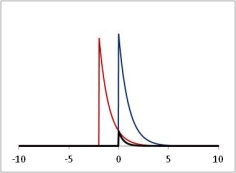


Figura xxx. Multiplicação das curvas  por  para 

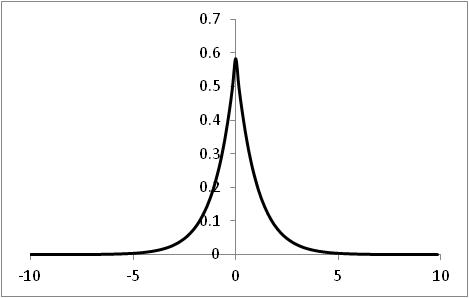


Figura xxx. Auto-correlação de  em função de 

**Propriedades da convolução:**

1. 

Prova: . Fazendo a mudança de variável , . Note que , pois .

1. A propriedade distributiva frente à adição  é trivial.
2. Propriedade distributiva frente à convolução: .

Prova:  e , então . Por outro lado:

.

Fazendo ,  e  temos que:

.

Chamando  e  temos .

**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 teorema da convolução é demonstrado também de outra forma e discutido com mais profundidade no apêndice xxx.

**FGM e Função Característica conjuntas:**





Relações com os momentos: sabemos que . Por outro lado .

Então 

Portanto 

Claro então que 

A série de Taylor multivariada é dada por:





Nesse caso 

Comparando com a expansão dos momentos vemos que  e que .

**Produto 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: Note que  pode ser escrito como  ou . Então 



Então:







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

Note que a fdp do quociente  agora muda para  ou . Então 







**Apêndice XX: Mudança de coordenadas e Jacobiano:**

Na álgebra de vetores em 3 dimensões podemos definir os vetores unitários ,  e  de forma que qualquer vetor é escrito como . Note que os vetores unitários gozam da propriedade de que , onde  é a norma do vetor . Também definimos o produto vetorial através da seguinte operação:



Sem perda de generalidade podemos colocar o vetor  no eixo 1 e o vetor  no plano 1-2 através de uma rotação dos eixos. Nesse caso:  , onde . Se o vetor  faz um ângulo  com o vetor  então  com . Colocando esses dois vetores no produto vetorial temos que :



**Nesse caso** . Agora note que a área de um paralelepípedo composto pelos vetores  e  é dada pela base que vale  multiplicada pela altura que vale . Ou seja  é a área do paralelepípedo entre os dois vetores.

Agora vamos tomar o caso em que . No plano a posição é transferida para o plano  para a posição . O vetor posição  é dado por: . Assim 

Ou seja .

Já 

Ou seja .

A área entre eles será 

Que nos leva ao elemento de área:



**Apêndice XX: Teorema da convolução:**

Sejam e  e suas transformadas inversas  e . A convolução é dada por logo . Aplicando a transformada de ambos os lados obtemos que , ou seja, , a função característica da variável z é o produto das funções características das variáveis x e y. O teorema também vale na transformada inversa , ou seja, .

**Generalização do Teorema da Convolução**

Uma forma muito elegante de demonstrar o teorema da convolução generalizado é através das funções delta de Dirac. Vamos somar  v.a.s independentes e queremos a densidade de probabilidade da variável . Nesse caso temos:



Em vez de colocar a restrição nos limites das integrais, o que nos levaria a um hiper plano de  dimensões, vamos introduzirr uma delta de Dirac na integral que nos garanta a igualdade , ou seja, , para a desigualdade . Incluindo a delta de Dirac na integral temos:



Agora a delta permitiu liberar os limites de integração. Usando  obtemos



Ou 

Aplicando a transformada de Fourier de ambos os lados temos que .

**Um Teorema para a Correlação:**

Note que  gera a correlação . Por outro lado, podemos usar a delta de Dirac para liberar os limites de integração . Usando novamente a delta da forma  chegamos a  . Na integral de y trocar de variável para  e ficamos com onde . Assim, , ou aplicando a transformada de Fourier de ambos os lados .

1. Estamos usando a seguinte regra para derivar integrais:  onde  portanto  ou seja . [↑](#footnote-ref-1)