Análise Multivariada.

Vamos agora criar uma função vetorial de conjunto $\vec{x}_v(A):\Omega \to \mathbb{R}^n$ que possui as componentes $x_1, x_2, \cdots, x_n$ em que cada x_j é uma v.a. Neste caso $\{x_{vi} \le x_i\}$ e $\{x_{vj} \le x_j\}$ são dois eventos, assim como $\{x_{vi} \le x_i\} + \{x_{vj} \le x_j\} = \{x_{vi} \le x_i, x_{vj} \le x_j\}$. Para facilitar a compreensão e as demonstrações vamos trabalhar apenas com o caso bivariado, ou seja, duas v.a.s, e depois generalizar para n. Facilita, nesse estágio, chamar uma v.a. de x e a outra de y.

Distribuição conjunta [Joint Distribution]

$$F(x,y) = P\{x_v \le x, y_v \le y\}$$

Propriedades:

1. $F(-\infty, y) = F(x, -\infty) = 0$ e $F(+\infty, +\infty) = 1$.

Prova: $\{x_v \le -\infty, y_v \le y\} \subset \{x_v = -\infty\} \quad \to \quad P\{x_v \le -\infty, y_v \le y\} \le P\{x_v = -\infty\} = 0$, então $0 \ge P\{x_v \le -\infty, y_v \le y\} \le 0$ logo $F(-\infty, y) = 0$. Troca o nome das v.a.s e o teorema continua válido. Para a segunda parte basta notar que $\{x_v \le +\infty, y_v \le +\infty\} = \Omega$ logo $P\{x_v \le +\infty, y_v \le +\infty\} = P\{\Omega\} = 1$.

2. $P\{x_1 < x_v \le x_2, y_v \le y\} = F(x_2, y) - F(x_1, y)$

Para mostrar isso basta usar a seguinte partição: $\{x_v \le x_2, y_v \le y\} \equiv \{x_v \le x_1, y_v \le y\} + \{x_1 < x_v \le x_2, y_v \le y\}$, logo $P\{x_v \le x_2, y_v \le y\} = P\{x_v \le x_1, y_v \le y\} + P\{x_1 < x_v \le x_2, y_v \le y\}$, ou seja:

$$F(x_2, y) = F(x_1, y) + P\{x_1 < x_v \le x_2, y_v \le y\}$$

De modo análogo é claro que $P\{x_v \le x, y_1 < y_v \le y_2\} = F(x, y_2) - F(x, y_1)$

3. $P\{x_1 < x_v \le x_2, y_1 < y_v \le y_2\} = F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1)$

Vamos tomar a partição:

$$\left\{x_1 < x_v \le x_2, y_v \le y_2\right\} \equiv \left\{x_1 < x_v \le x_2, y_v \le y_1\right\} + \left\{x_1 < x_v \le x_2, y_1 < y_v \le y_2\right\}$$

Então:

$$P\left\{x_1 < x_v \le x_2, y_v \le y_2\right\} = P\left\{x_1 < x_v \le x_2, y_v \le y_1\right\} + P\left\{x_1 < x_v \le x_2, y_1 < y_v \le y_2\right\}$$

$$F\left(x_2, y_2\right) - F\left(x_1, y_2\right) = F\left(x_2, y_1\right) - F\left(x_1, y_1\right) + P\left\{x_1 < x_v \le x_2, y_1 < y_v \le y_2\right\}$$

4. $P\left\{x < x_v \le x+\delta x, y < y_v \le y+\delta y\right\} = \dfrac{\partial^2 F}{\partial x \partial y}\delta x \delta y$

Fazendo $x_1 = x$, $x_2 = x+\delta x$, $y_1 = y$ e $y_2 = y+\delta y$ em (3) e usando o fato de que:

$$F\left(x+\delta x, y+\delta y\right) - F\left(x, y+\delta y\right) = \frac{F\left(x+\delta x, y+\delta y\right) - F\left(x, y+\delta y\right)}{\delta x}\delta x = \frac{\partial F\left(x, y+\delta y\right)}{\partial x}\delta x$$

$$F\left(x+\delta x, y\right) - F\left(x, y\right) = \frac{F\left(x+\delta x, y\right) - F\left(x, y\right)}{\delta x}\delta x = \frac{\partial F\left(x, y\right)}{\partial x}\delta x$$

Então

$$\left[F\left(x+\delta x, y+\delta y\right) - F\left(x, y+\delta y\right)\right] - \left[F\left(x+\delta x, y\right) - F\left(x, y\right)\right] = \frac{\partial F\left(x, y+\delta y\right)}{\partial x}\delta x - \frac{\partial F\left(x, y\right)}{\partial x}\delta x$$

$$P\left\{x < x_v \le x+\delta x, y < y_v \le y+\delta y\right\} = \left[\frac{\dfrac{\partial F\left(x, y+\delta y\right)}{\partial x} - \dfrac{\partial F\left(x, y\right)}{\partial x}}{\delta y}\right]\delta y \delta x = \frac{\partial^2 F\left(x, y\right)}{\partial x \partial y}\delta x \delta y$$

Note que $\dfrac{\partial^2 F\left(x, y\right)}{\partial x \partial y} \ge 0$ sempre.

Densidade de probabilidade conjunta [joint density probability]

Definimos a fdp conjunta agora como: $f\left(x,y\right)=\dfrac{\partial^2 F\left(x,y\right)}{\partial x\partial y}$.

O reverso é dado por: $F\left(x,y\right)=\int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y} f\left(u,v\right)dudv$

Prova do reverso: $\dfrac{\partial}{\partial y}\int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y} f\left(u,v\right)dudv=\int\limits_{-\infty}^{x} du\dfrac{\partial}{\partial y}\int\limits_{-\infty}^{y} f\left(u,v\right)dv=\int\limits_{-\infty}^{x} f\left(u,y\right)du$ e $\dfrac{\partial}{\partial x}\dfrac{\partial}{\partial y}\int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y} f\left(u,v\right)dudv=\dfrac{\partial}{\partial x}\int\limits_{-\infty}^{x} f\left(u,y\right)du=f\left(x,y\right)$ logo:

$$f\left(x,y\right)=\frac{\partial^2}{\partial x\partial y}\int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y} f\left(u,v\right)dudv$$

Se queremos a probabilidade de encontrar $\left(x,y\right)\in A$ então devemos fazer a seguinte integral múltipla:

$$P\left\{\left(x,y\right)\in A\right\}=\iint\limits_{\left(x,y\right)\in A} f\left(x,y\right)dxdy$$

Também exigimos aqui que:

$$F\left(\infty,\infty\right)=\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} f\left(x,y\right)dxdy=1$$

Para ser uma densidade de probabilidade multivariada, então, $f\left(x,y\right)\geq 0$ e $\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} f\left(x,y\right)dxdy=1$.

Distribuição e Densidades Marginais:

Suponha que queremos a estatística de apenas uma das variáveis sem interessar o valor da outra. Notamos que $\left\{x_v\leq x\right\}\equiv\left\{x_v\leq x,y_v\leq\infty\right\}$ assim como $\left\{y_v\leq y\right\}\equiv\left\{x_v\leq\infty,y_v\leq y\right\}$. Então $F_x\left(x\right)=F\left(x,\infty\right)$ e $F_y\left(y\right)=F\left(\infty,y\right)$ são as distribuições marginais de x e de y. Note então que:

$$F_x(x)=\int_{-\infty}^{x}\int_{-\infty}^{+\infty} f(u,y)\,du\,dy=\int_{-\infty}^{x}\left[\int_{-\infty}^{+\infty} f(u,y)\,dy\right]du$$

e

$$F_y(y)=\int_{-\infty}^{x}\int_{-\infty}^{+\infty} f(x,v)\,dx\,dv=\int_{-\infty}^{y}\left[\int_{-\infty}^{+\infty} f(x,v)\,dx\right]dv$$

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:

$$f_x(x)=\frac{d}{dx}F_x(x)=\frac{d}{dx}\int_{-\infty}^{x}\left[\int_{-\infty}^{+\infty} f(u,y)\,dy\right]du=\int_{-\infty}^{+\infty} f(x,y)\,dy$$

e

$$f_y(y)=\frac{d}{dy}F_y(y)=\frac{d}{dy}\int_{-\infty}^{y}\left[\int_{-\infty}^{+\infty} f(x,v)\,dx\right]dv=\int_{-\infty}^{+\infty} f(x,y)\,dx$$

Fica claro então que:

$$\int_{-\infty}^{+\infty} f_x(x)\,dx=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\,dx\,dy=1 \text{ e } \int_{-\infty}^{+\infty} f_y(y)\,dy=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\,dx\,dy=1.$$

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:

$$\delta(\vec{x}-\vec{x}_o)=\delta(x_1-x_{1o})\delta(x_2-x_{2o})\cdots\delta(x_n-x_{no}).$$

Funções escalares $f(\vec{x}):\mathbb{R}^n\rightarrow\mathbb{R}$ multivariadas.

Vamos criar a v.a. z_v à partir das v.a.s x e y através da função escalar $z_v = g(x, y)$ que associa um vetor em $\mathbb{R}^2$ a um número real em $\mathbb{R}$. Nesse caso o evento $\{z_v \leq z\} \equiv \{g(x, y) \leq z\}$ e a distribuição de probabilidade de z será dada por:

$$F_z(z) = \iint_{g(x,y)\leq z} f(x, y)\,dxdy$$

enquanto a fdp da v.a. z dada por:

$$f_z(z) = \iint_{z<g(x,y)\leq z+dz} f(x, y)\,dxdy$$

A integral pode complicar devido à restrição $g(x, y) \leq z$ ou $z < g(x, y) \leq z + dz$. Em vários casos pode ser vantajoso trocar as variáveis de integração para $u(x, y)$ e $w(x, y)$ através da regra do Jacobiano:

$$\iint_V f(x, y)\,dxdy = \iint_V f(u, w)\,J(u, w)\,dudw$$

Onde o Jacobiano é dado pela matriz: $J = \det\begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} \\ \frac{\partial x}{\partial w} & \frac{\partial y}{\partial w} \end{pmatrix}$. O apêndice xxx traz a demonstração dessa regra. Podemos aplicá-la ao caso da transformação de coordenadas retangulares para polares, em que $x = r\cos\theta$ e $y = r\sin\theta$. Portanto:

$$J = \det\begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial y}{\partial r} \\ \frac{\partial x}{\partial \theta} & \frac{\partial y}{\partial \theta} \end{pmatrix} = \det\begin{pmatrix} \cos\theta & \sin\theta \\ -r\sin\theta & r\cos\theta \end{pmatrix} = r\cos^2\theta + r\sin^2\theta = r\,.$$

Isso significa que $dxdy = rdrd\theta$.

Operação esperança multivariada:

Agora a operação esperança de qualquer função escalar das v.a.s x e y $z = g(x, y)$ é dada por:

$$E\left[g(x,y)\right]=\int_{-\infty}^{\infty} z\, f_z(z)dz=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\, f(x,y)dxdy$$

Dessa definição podemos extrair as seguintes propriedades da esperança:

1. $E[k]=k$

Se $g(x,y)=k$ é uma constante então: $E[k]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} k\, f(x,y)dxdy=k\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy=k$

2. $E\left[\alpha g(x,y)+\beta h(x,y)\right]=\alpha E\left[g(x,y)\right]+\beta E\left[h(x,y)\right]$

$$E\left[\alpha g(x,y)+\beta h(x,y)\right]=\alpha\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\, f(x,y)dxdy+\beta\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y)\, f(x,y)dxdy$$

2.1. $E[x+y]=E[x]+E[y]$

Momentos conjuntos:

No caso multivariado definimos os momentos por:

$$M_{kp}=E\left[x^k y^p\right]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^k y^p\, f(x,y)dxdy$$

A generalização para n v.a.s é:

$$M_{k_1k_2\cdots k_n}=E\left[x_1^{k_1}x_2^{k_2}\cdots x_n^{k_n}\right]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} x_1^{k_1}x_2^{k_2}\cdots x_n^{k_n}\, f(x_1,x_2\cdots,x_n)dx_1dx_2\cdots dx_n$$

Notamos imediatamente que: $M_{00}=m_{00}=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy=1$. Alguns desses momentos possuem nomes específicos:

$$M_{10}=\mu_x=E[x]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\, f(x,y)dxdy=\int_{-\infty}^{\infty} x\, f_x(x)dx$$

$$M_{01} = \mu_y = E[y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\, f(x,y)\,dxdy = \int_{-\infty}^{\infty} y\, f_y(y)\,dy$$

Com eles podemos definir os momentos centrados por:

$$m_{kp} = E\left[(x-\mu_x)^k (y-\mu_y)^p\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_x)^k (y-\mu_y)^p\, f(x,y)\,dxdy$$

Novamente percebe-se que: $m_{00} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dxdy = 1$ e que:

$$m_{10} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_x)\, f(x,y)\,dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\, f(x,y)\,dxdy - \mu_x \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dxdy$$

$m_{10} = \mu_x - \mu_x 1 = 0$, da mesma forma que $m_{01} = 0$.

Os momentos centrados com nomes específicos são as variâncias:

$$V(x) = \sigma_x^2 = m_{20} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_x)^2\, f(x,y)\,dxdy$$

$$V(y) = \sigma_y^2 = m_{02} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (y-\mu_y)^2\, f(x,y)\,dxdy$$

e a covariância:

$$\text{cov}(x,y) = m_{11} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_x)(y-\mu_y)\, f(x,y)\,dxdy$$

Nota-se então que:

$$V(x) = \text{cov}(x,x) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_x)^2\, f(x,y)\,dxdy$$

$$V(y) = \text{cov}(y,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (y-\mu_y)^2\, f(x,y)\,dxdy.$$

A covariância tem as seguintes propriedades:

1. $\operatorname{cov}(x_1,x_2)=\operatorname{cov}(x_2,x_1)$ pois $(x_1-\mu_1)(x_2-\mu_2)=(x_2-\mu_2)(x_1-\mu_1)$

2. $\operatorname{cov}(x_1+x_2,x_3)=\operatorname{cov}(x_1,x_3)+\operatorname{cov}(x_2,x_3)$, pois:

$$(x_1+x_2-\mu_1-\mu_2)(x_3-\mu_3)=(x_1-\mu_1)(x_3-\mu_3)+(x_2-\mu_2)(x_3-\mu_3)$$

3. $\operatorname{cov}(x,y)=E[xy]-E[x]E[y]$

$$\operatorname{cov}(x,y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_x)(y-\mu_y)f(x,y)dxdy=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(xy-\mu_x y-\mu_y x+\mu_y\mu_x)f(x,y)dxdy$$

$$\operatorname{cov}(x,y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xy\,f(x,y)dxdy-\mu_x\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}y\,f(x,y)dxdy+$$

$$-\mu_y\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x\,f(x,y)dxdy+\mu_y\mu_x\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)dxdy$$

$$\operatorname{cov}(x,y)=E[xy]-\mu_x E[y]-\mu_y E[x]+\mu_y\mu_x$$

$$\operatorname{cov}(x,y)=E[xy]-\mu_x\mu_y-\mu_y\mu_x+\mu_y\mu_x=E[xy]-\mu_x\mu_y=E[xy]-E[x]E[y]$$

4. $\operatorname{cov}(\alpha x,\beta y)=\alpha\beta\operatorname{cov}(x,y)$

$$\operatorname{cov}(\alpha x,\beta y)=E[\alpha\beta xy]-E[\alpha x]E[\beta y]=\alpha\beta E[xy]-\alpha\beta E[x]E[y]=\alpha\beta(E[xy]-E[x]E[y])$$

5. $\operatorname{cov}(x,k)=0$ onde k é uma constante.

$$\operatorname{cov}(x,k)=E[kx]-E[x]E[k]=kE[x]-kE[x]=0$$

Essas propriedades dão origem as seguintes propriedades da variância:

1. $V(x)=E\left[x^2\right]-(E[x])^2$ pois $V(x)=\operatorname{cov}(x,x)=E\left[x^2\right]-E[x]E[x]$

2. $V[kx]=k^2V[x]$ pois $V[kx]=\operatorname{cov}(kx,kx)=k^2\operatorname{cov}(x,x)$

3. $V[\alpha+\beta x]=\beta^2V[x]$

$$V\left[\alpha+\beta x\right]=\operatorname{cov}\left(\alpha+\beta x,\alpha+\beta x\right)=\operatorname{cov}\left(\alpha,\alpha+\beta x\right)+\operatorname{cov}\left(\beta x,\alpha\right)+\operatorname{cov}\left(\beta x,\beta x\right)=\beta^2\operatorname{cov}\left(x,x\right)$$

4. $V\left[\alpha x+\beta y\right]=\alpha^2 V\left[x\right]+\beta^2 V\left[y\right]+2\alpha\beta\operatorname{cov}\left(x,y\right)$

Corolário: $V\left[x\pm y\right]=V\left[x\right]+V\left[y\right]\pm 2\operatorname{cov}\left(x,y\right)$

$$V\left[\alpha x+\beta y\right]=\operatorname{cov}\left(\alpha x+\beta y,\alpha x+\beta y\right)=\operatorname{cov}\left(\alpha x,\alpha x\right)+\operatorname{cov}\left(\alpha x,\beta y\right)+$$
$$+\operatorname{cov}\left(\beta y,\alpha x\right)+\operatorname{cov}\left(\beta y,\beta y\right)==\alpha^2\operatorname{cov}\left(x,x\right)+\alpha\beta\operatorname{cov}\left(x,y\right)+\alpha\beta\operatorname{cov}\left(y,x\right)+\beta^2\operatorname{cov}\left(y,y\right)=$$
$$=\alpha^2\operatorname{cov}\left(x,x\right)+2\alpha\beta\operatorname{cov}\left(x,y\right)+\beta^2\operatorname{cov}\left(y,y\right)$$

Matriz de variância-covariância:

Definindo a matriz $n\times n$: $V_{ij}=\operatorname{cov}\left(x_i,x_j\right)$, 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 $V_{ij}=\frac{1}{n}\sum_k\left(x_{ik}-\bar{x}_i\right)\left(x_{jk}-\bar{x}_j\right)$, onde x_{ik} foi o valor da késima observação da v.a. x_i e $\bar{x}_i=\frac{1}{n}\sum_k x_{ik}$, em lugar da matriz $V_{ij}=E\left[\left(x_i-\mu_i\right)\left(x_j-\mu_j\right)\right]$. Ou seja estamos substituindo a operação esperança $E\left[Q\right]\to\frac{1}{n}\sum_k Q_k$.

Nesse caso $\bar{Q}=\frac{1}{n}\sum_k Q_k$ é um estimador de $E\left[Q\right]$ 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 $V_{ij}=\frac{1}{n}\sum_k\left(x_{ik}-\bar{x}_i\right)\left(x_{jk}-\bar{x}_j\right)$ ou $V_{ij}=E\left[\left(x_i-\mu_i\right)\left(x_j-\mu_j\right)\right]$.

1. É simétrica: $V_{ij} = V_{ji}$

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

Uma matriz é definida positiva se para qualquer vetor $\vec{h}$ não nulo o produto:

$$\vec{\tilde{h}} M \vec{h} = (h_1, h_2, \cdots, h_n) M \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_n \end{pmatrix} = \sum_i \sum_j M_{ij} h_i h_j > 0$$, é positivo.

Prova: considere $z = \sum_i h_i (x_i - \mu_i) \neq 0$. Obviamente que $z = \sum_j h_j (x_j - \mu_j)$ e que $z^2 = \sum_i \sum_j (x_i - \mu_i)(x_j - \mu_j) h_i h_j > 0$. Então, o fato de que $E[z^2] > 0$, implica que

$$\sum_i \sum_j E[(x_i - \mu_i)(x_j - \mu_j)] h_i h_j = \sum_i \sum_j V_{ij} h_i h_j > 0$$

Variáveis aleatórias independentes:

Se os eventos $\{x_v \in A\}$ e $\{y_v \in B\}$ são independentes então $P[\{\{x_v \in A\}\{y_v \in B\}\}] = P[\{x_v \in A\}] P[\{y_v \in B\}]$. Neste caso então:

$$F(x, y) = F_x(x) F_y(y) \text{ e } f(x, y) = f_x(x) f_y(y)$$

Experimentos independentes:

Suponha que o espaço dos eventos da v.a. x seja Ω_x e o espaço da v.a. y seja Ω_y, e que ao realizar um experimento conjunto, cujos eventos pertencem ao espaço amostral $\Omega = \Omega_x \times \Omega_y$, o resultado de um não interfere no outro. Matematicamente estamos afirmando que:

$$x(\omega_1, \omega_2) = x(\omega_1) \text{ e } y(\omega_1, \omega_2) = y(\omega_2)$$

Então as v.a.s x e y são independentes.

Exemplo de v.a.s independentes: lançar dois dados de cores diferentes simultaneamente e definir x como o resultado de uma cor e y 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 x foi 1, o resultado para y jamais poderá ser 1.

Teorema 1: Se x e y são independentes, então $g(x)$ e $h(y)$ também são independentes.

Prova: se $\{x_v \in A\}$ e $\{y_v \in B\}$ são independentes, quaisquer dois sub-conjuntos de $\{x_v \in A\}$ e $\{y_v \in B\}$ serão independentes. Assim $\{g(x) \leq g\} \subset \{x_v \in A\}$ e $\{h(y) \leq h\} \subset \{y_v \in B\}$ é a condição para poder calcular as funções $g(x)$ e $h(y)$. Portanto se x e y são independentes, então $\forall$ $g(x)$ e $h(y)$ são independentes.

Teorema 2. Se x e y são independentes, então $E[xy] = E[x]E[y]$.

$$E[xy] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\,y\,f(x,y)\,dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\,y\,f_x(x)\,f_y(y)\,dxdy = \left[\int_{-\infty}^{\infty} x f_x(x)\,dx\right]\left[\int_{-\infty}^{\infty} y\,f_y(y)\,dy\right]$$

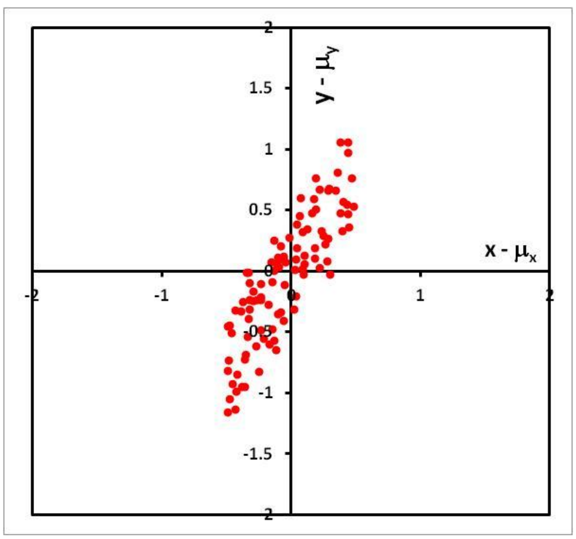
Teorema 3. Se x e y são independentes, então $\operatorname{cov}(x,y) = 0$.

$$\operatorname{cov}(x,y) = E[xy] - E[x]E[y] = E[x]E[y] - E[x]E[y] = 0$$

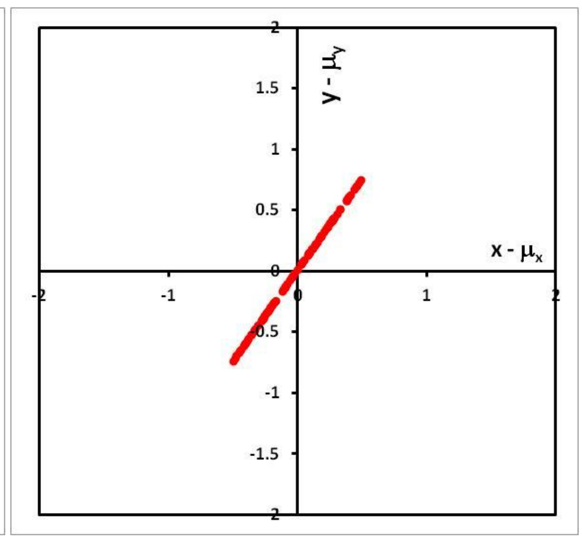
A covariância, portanto, nos fornece alguma informação sobre a independência entre v.a.s. Se $\operatorname{cov}(x,y) = 0$ então x e y são independentes. O que ocorre se $\operatorname{cov}(x,y) > 0$ ou $\operatorname{cov}(x,y) < 0$?

Note que os produtos $(x-\mu_x)(y-\mu_y)$ e $(x-\overline{x})(y-\overline{y})$ em um gráfico $(x-\mu_x)\ \text{vs}\ (y-\mu_y)$ ou $(x-\overline{x})\ \text{vs}\ (y-\overline{y})$ serão positivos no primeiro e terceiro quadrantes, $++\, e\, --$, e negativos no segundoe quarto quadrantes, $-+\, e\, +-$.

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



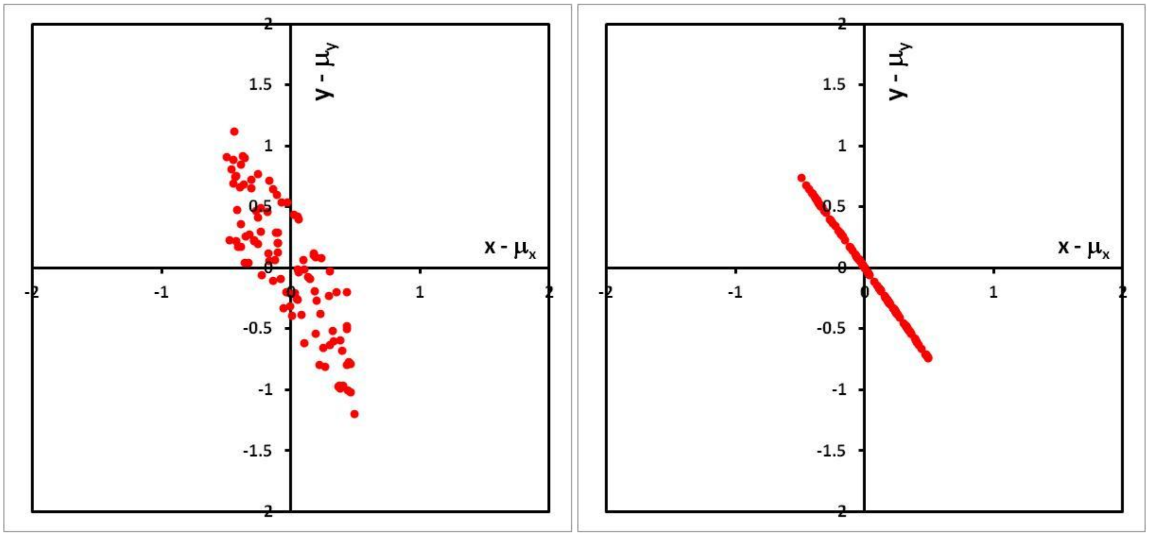
(a)



(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 $\sum_i (x_i-\overline{x})(y_i-\overline{y})$ negativa, ou seja, com uma covariância negativa. Percebe-se dessa núvem que a v.a. y tende a decrescer quando a v.a. x cresce, e a crescer quando x decresce. Nessa situação afirmamos que as v.a.s x e y são negativamente correlacionadas. O gráfico da figura xxx(b) mostra o caso em que $y = -x$, 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 $\sum_i (x_i - \bar{x})(y_i - \bar{y}) = 0$ como mostra a figura xxx.

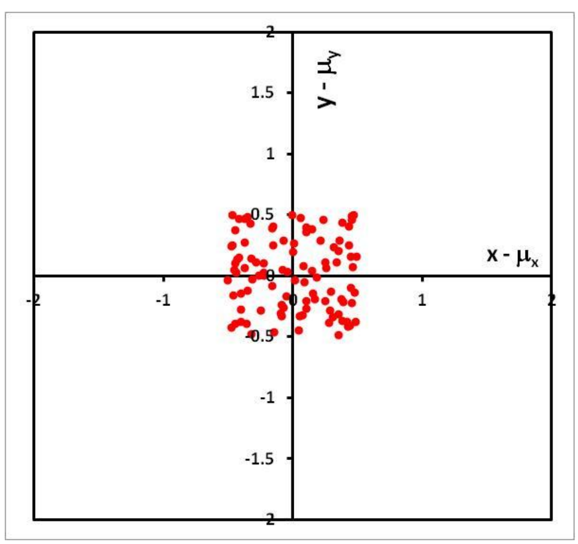


Figura xxx. Caso de duas variáveis descorrelacionadas.

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, $\dim\left[\operatorname{cov}(x,y)\right]=\dim[x]\dim[y]$. Se x e y 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:

$$r_{xy}=r[x,y]=\frac{\operatorname{cov}(x,y)}{\sqrt{\operatorname{cov}(x,x)}\sqrt{\operatorname{cov}(y,y)}}=\frac{\operatorname{cov}(x,y)}{\sqrt{V[x]\,V[y]}}$$

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: $-1\le r_{xy}\le 1$.

Prova usando a desigualdade de Schwartz:

$E\left\{\left[\lambda\left(x-\mu_x\right)-\left(y-\mu_y\right)\right]^2\right\}\ge 0 \quad \forall\lambda\in\mathbb{R}$ pois se trata da esperança de uma quantidade positiva.

Desenvolvendo o quadrado temos:

$$\left[\lambda\left(x-\mu_x\right)-\left(y-\mu_y\right)\right]^2=\lambda^2\left(x-\mu_x\right)^2-2\lambda\left(x-\mu_x\right)\left(y-\mu_y\right)+\left(y-\mu_y\right)^2$$

Logo

$$E\left\{\left[\lambda\left(x-\mu_x\right)-\left(y-\mu_y\right)\right]^2\right\}=\lambda^2E\left[\left(x-\mu_x\right)^2\right]-2\lambda E\left[\left(x-\mu_x\right)\left(y-\mu_y\right)\right]+E\left[\left(y-\mu_y\right)^2\right]$$

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

$$E\left\{\left[\lambda\left(x-\mu_x\right)-\left(y-\mu_y\right)\right]^2\right\}=\lambda^2V[x]-2\lambda\operatorname{cov}(x,y)+V[y]$$

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

$$V[x]\lambda^2 - 2\operatorname{cov}(x,y)\lambda + V[y] \ge 0 \text{ com } V[x] > 0 \text{ e } V[y] > 0$$

A desigualdade $a\lambda^2 + b\lambda + c \ge 0$ com $a > 0$ só pode ser satisfeita se $a\lambda^2 + b\lambda + c = 0$ não admite raízes reais ou apenas uma raiz que toca o eixo x. Essa condição implica que $b^2 - 4ac \le 0$. Agora fazendo $a = V[x]$, $b = -2\operatorname{cov}(x,y)$ e $c = V[y]$ percebe-se que $4\operatorname{cov}^2(x,y) - 4V[x]V[y] \le 0$ ou seja, $\frac{\operatorname{cov}^2(x,y)}{V[x]V[y]} \le 1$ que implica em $-1 \le \frac{\operatorname{cov}(x,y)}{\sqrt{V[x]V[y]}} \le 1$.

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

Teorema generalizado para independência entre v.a.s: $-1 \le \frac{E[xy]}{\sqrt{E[x^2]E[y^2]}} \le 1$.

Basta fazer o mesmo começando com $E\left[(\lambda x - y)^2\right] \ge 0$ que nos leva diretamente à $E[x^2]\lambda^2 - 2E[xy]\lambda + E[y^2] \ge 0$ e, consequentemente, a $-1 \le \frac{E[xy]}{\sqrt{E[x^2]E[y^2]}} \le 1$. O fato de que esse é um número entre -1 e +1 significa que sempre existirá um ângulo θ para o qual $E[xy] = \sqrt{E[x^2]}\sqrt{E[y^2]}\cos\theta$. Se definimos $x_{RMS} = \sqrt{E[x^2]}$, ou seja root-mean-square, porque utilizamos $\sqrt{\left(\frac{1}{n}\sum_k x_k^2\right)}$ como estimador de $\sqrt{E[x^2]}$, podemos afirmar então que:

$$E[xy] = x_{RMS}\, y_{RMS} \cos\theta \text{ ou } \cos\theta = \frac{E[xy]}{x_{RMS}\, y_{RMS}}$$

Em que o coseno mede o grau de relação entre as v.a.s x e y. Se $E[xy]=0$, mas $E\left[x^2\right]\neq 0$ e $E\left[y^2\right]\neq 0$ então $\theta=\frac{\pi}{2}=90^o$ e dizemos que x e y são ortogonais entre si, ou seja, $x\perp y$.

Espaços métricos e distância de correlação:

Um espaço é métrico se existe uma função distância $d(x,y):E\times E\to\mathbb{R}$ para $\forall x,y\in E$ satisfazeno aos axiomas:

1. Desigualdade triangular: $d(x,z)\leq d(x,y)+d(y,z)$
2. Se $d(x,y)=0$ então $x=y$
3. $d(x,y)=d(y,x)$

Com esses axiomas podemos demonstrar o teorema:

a. $d(x,y)\geq 0$

Fazer $z=x$ no axioma 1: $d(x,x)\leq d(x,y)+d(y,x)$ usando os axiomas (2) e (3) $2d(x,y)\geq 0$, logo $d(x,y)\geq 0$.

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 E. Nesse caso dizemos que o espaço é métrico.

Distância Euclidiana:

A distância Euclidiana entre os vetores $\vec{x}=(x_1,x_2,\cdots,x_n)$ e $\vec{y}=(y_1,y_2,\cdots,y_n)$ é definida como $d(\vec{x},\vec{y})=\sqrt{(x_1-y_1)^2+(x_2-y_2)^2+\cdots+(x_n-y_n)^2}=\sqrt{\sum_{j=1}^{n}\left(x_j-y_j\right)^2}$ que já apresenta naturalmente as propriedades $d(\vec{x},\vec{y})\geq 0$ e $d(\vec{x},\vec{y})=d(\vec{y},\vec{x})$. Notamos que

$d\left(\vec{x},\vec{y}\right)=\sqrt{\left(\vec{x}-\vec{y}\right)\cdot\left(\vec{x}-\vec{y}\right)}$ onde o produto escalar entre dois vetores é definido da forma anterior como $\vec{a}\cdot\vec{b}=\sum_j a_j b_j$.

Falta mostrar o axioma 1:

$$d\left(x,z\right)\le d\left(x,y\right)+d\left(y,z\right)$$

$$d^2\left(\vec{x},\vec{z}\right)=\sum_j\left(x_j-z_j\right)^2=\sum_j\left(x_j-y_j+y_j-z_j\right)^2=$$

$$=\sum_j\left[\left(x_j-y_j\right)^2+2\left(x_j-y_j\right)\left(y_j-z_j\right)+\left(y_j-z_j\right)^2\right]=$$

$$=\sum_j\left(x_j-y_j\right)^2+\sum_j\left(y_j-z_j\right)^2+2\sum_j\left(x_j-y_j\right)\left(y_j-z_j\right)$$

$$d^2\left(\vec{x},\vec{z}\right)=d^2\left(\vec{x},\vec{y}\right)+d^2\left(\vec{y},\vec{z}\right)+2\left(\vec{x}-\vec{y}\right)\cdot\left(\vec{y}-\vec{z}\right)$$

Agora usamos o fato de que $-1\le\frac{\vec{a}\cdot\vec{b}}{\sqrt{\left(\vec{a}\cdot\vec{a}\right)\left(\vec{b}\cdot\vec{b}\right)}}\le+1$ para perceber que $\left(\vec{x}-\vec{y}\right)\cdot\left(\vec{y}-\vec{z}\right)\le\sqrt{\left(\vec{x}-\vec{y}\right)\cdot\left(\vec{x}-\vec{y}\right)}\sqrt{\left(\vec{y}-\vec{z}\right)\cdot\left(\vec{y}-\vec{z}\right)}$, ou seja, $\left(\vec{x}-\vec{y}\right)\cdot\left(\vec{y}-\vec{z}\right)\le d\left(\vec{x},\vec{y}\right)d\left(\vec{y},\vec{z}\right)$. Nesse caso:

$$d^2\left(\vec{x},\vec{z}\right)\le d^2\left(\vec{x},\vec{y}\right)+2d\left(\vec{x},\vec{y}\right)d\left(\vec{y},\vec{z}\right)+d^2\left(\vec{y},\vec{z}\right)=\left[d\left(\vec{x},\vec{y}\right)+d\left(\vec{y},\vec{z}\right)\right]^2$$

Logo: $d\left(\vec{x},\vec{z}\right)\le d\left(\vec{x},\vec{y}\right)+d\left(\vec{y},\vec{z}\right)$.

Essa não é a única distância possível. Existem outras distâncias como a distância Manhattan $d_{xy}=\sum_j\left|x_j-y_j\right|$. 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].

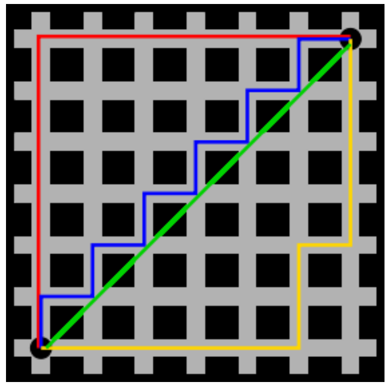


Figura xxx. Manhattan distance. Note que as distâncias vermelha, azul e amarela são iguais.

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 x um inteiro e p um número primo, então: $x = a_o + a_1 p + a_2 p^2 + \cdots + a_k p^k$ com $0 \le a_k \le p - 1$.

Exemplo: $x = 23$ e $p = 2$. Nesse caso: $p^2 = 4$, $p^3 = 8$, $p^4 = 16$ e $p^5 = 32 > 23$ já não interessa mais. Assim começamos da potência mais alta e vamos descendo: $23 - 16 = 7$, $7 - 4 = 3$ e $3 - 2 = 1$, logo: $x = 1 \times 2^0 + 1 \times 2^1 + 1 \times 2^2 + 1 \times 2^4 = 1 + 2 + 4 + 16 = 23$.

Vamos tomar agora $p = 3$, então $p^2 = 9$ e $p^3 = 27 > 23$. Nesse caso $23 - 9 = 14$, $14 - 9 = 5$, $23 - 2 \times 9 = 5$, $5 - 3 = 2$, logo $x = 2 \times 3^0 + 1 \times 3^1 + 2 \times 3^2 = 2 + 3 + 18 = 23$.

Qualquer inteiro positivo pode ser escrito como:

$$x = \sum_{j=0}^{\infty} a_j p^j \quad \rightarrow \quad 0 \le a_j \le p - 1 \text{ e } p \text{ um número primo.}$$

O que acontece se permitimos j negativos? Teríamos também $\frac{1}{p}$, $\frac{1}{p^2}$ etc e assim poderíamos escrever qualquer número racional como: $x = \sum_{j=r}^{\infty} a_j p^j$ com $r \in \mathbb{Z}$, logo r pode ser negativo. Note dessa definição então que x é divisível por p^r, então p^r é a maior potência divisora de x. A distância p-ádica é definida como: $|x|_p = p^{-r}$.

Exemplos:

1. $|20|_2 = ?$. $20 = 2^2 \times 5 = 2^2 \times \left[1 + 2^2\right] = 2^2 + 2^4 = \sum_{j=2}^{\infty} a_j p^j$ com $a_2 = 1$, $a_3 = 0$, $a_4 = 1$ e $a_{j>4} = 0$. Neste caso $|20|_2 = \dfrac{1}{2^2} = \dfrac{1}{4} = 0.25$.

2. O número ZERO é divisível por qualquer número, portanto $|0|_p = \dfrac{1}{p^{\infty}} = 0$.

Operações adição e multiplicação de números p-ádicos:

Suponha agora dois números x e y, com $x = p^{r_1} \sum_{j=0}^{\infty} a_j p^j$ e $y = p^{r_2} \sum_{k=0}^{\infty} b_k p^k$ então $xy = p^{r_1 + r_2} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} a_j b_k p^{j+k}$. Fazendo $j + k = m$, $k = m - j$ e $c_k = a_j b_{m-j}$ então podemos escrever $xy = p^{r_1 + r_2} \sum_{k=0}^{\infty} c_m p^m$. Nesse caso percebemos que $|xy|_p = \dfrac{1}{p^{r_1 + r_2}} = \dfrac{1}{p^{r_1}} \dfrac{1}{p^{r_2}} = |x|_p \, |y|_p$.

Suponha agora dois números x e y, com $x = p^{r_1} \sum_{j=0}^{\infty} a_j p^j$ e $y = p^{r_2} \sum_{k=0}^{\infty} b_k p^k$ então $x + y = p^{r_{\min}} \left[p^{r_1 - r_{\min}} \sum_{j=0}^{\infty} a_j p^j + p^{r_2 - r_{\min}} \sum_{j=0}^{\infty} b_j p^j \right]$, onde $r_{\min} = \min\{r_1, r_2\}$. O número que ficou entre colchetes $[\ ]$ pode ser, ou não, divisível por p, ou seja $[\ ] = p^r \sum c_k p^k$. Neste caso $x + y = p^{r_{\min} + r} \sum c_k p^k$. Assim $|x + y|_p = \dfrac{1}{p^{r + r_{\min}}} \le \dfrac{1}{p^{r_{\min}}} \le \dfrac{1}{p^{r_1}} + \dfrac{1}{p^{r_2}}$ e $\dfrac{1}{p^{r_{\min}}} > \dfrac{1}{p^{r_{\max}}}$. Dessa forma $|x + y|_p \le |x|_p + |y|_p$ como a distância, mas a distância p-ádica apresenta uma restrição maior que é: $\left|x+y\right|_p \le \max\left[\left|x\right|_p, \left|y\right|_p\right]$. Note que isso significa que $\left|x+x\right|_p \le \max\left[\left|x\right|_p, \left|x\right|_p\right] \le \left|x\right|_p$, ou seja, $\left|2x\right|_p \le \left|x\right|_p$ em lugar dos costumeiro $\left|2x\right| \ge \left|x\right|$. Em 1944 Mark Krasner criou o termo espaço ultramétrico para a distância satisfazendo aos axiomas:

1. Se $\delta(x,y)=0$ então $x=y$
2. $\delta(x,y)=\delta(y,x)$
3. $\delta(x,z) \le \max\left[\delta(x,y), \delta(y,z)\right]$

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 x e y, com $\bar{x}=\frac{1}{N}\sum_j x_j$, $s_x^2=\frac{1}{N}\sum_j\left(x_j-\bar{x}\right)^2$, $\bar{y}=\frac{1}{N}\sum_j y_j$, $s_y^2=\frac{1}{N}\sum_j\left(y_j-\bar{y}\right)^2$, em que $\bar{x}$ e s_x^2 são os estimadores de $\mu_x=E[x]$ e de $\sigma_x^2=V(x)$ e $\bar{y}$ e s_y^2 os estimadroes de $\mu_y=E[y]$ e $\sigma_y^2=V(y)$. A covariância será dada por $\operatorname{cov}(x,y)=\frac{1}{N}\sum_j\left(x_j-\bar{x}\right)\left(y_j-\bar{y}\right)$ e o coeficiente de correlação $r_{xy}=\frac{\operatorname{cov}(x,y)}{s_x s_y}$. Algebricamente vemos que:

$$s_x^2=\frac{1}{N}\sum_j\left(x_j^2-2\bar{x}\,x_j+\bar{x}^2\right)=\frac{1}{N}\sum_j x_j^2-2\bar{x}\frac{1}{N}\sum_j x_j+\bar{x}^2\frac{1}{N}\sum_j 1=\frac{1}{N}\sum_j x_j^2-2\bar{x}\,\bar{x}+\bar{x}^2\frac{1}{N}N$$
$$s_x^2=\frac{1}{N}\sum_j x_j^2-\bar{x}^2=\overline{x^2}-\bar{x}^2$$

Agora vamos definir aa v.a.s padronizadas $z_j = \frac{x_j - \overline{x}}{\sqrt{N}\, s_x}$ e $s_j = \frac{y_j - \overline{y}}{\sqrt{N}\, s_y}$. Notamos que $\sum_j z_j = \frac{1}{\sqrt{N}\, s_x} \sum_j \left(x_j - \overline{x} \right) = \frac{1}{\sqrt{N} s_x} \left[\sum_j x_j - \overline{x} \sum_j 1 \right] = \frac{1}{\sqrt{N} s_x} \left[N\,\overline{x} - N\,\overline{x} \right] = 0$ e, da mesma forma, $\sum_j s_j = 0$, que é equivalente ao fato de que $E\left[x - \mu \right] = 0$. Por outro lado $\sum_j z_j^2 = \frac{1}{N\, s_x^2} \sum_j \left(x_j - \overline{x} \right)^2 = \frac{1}{s_x^2} \left[\frac{1}{N} \sum_j \left(x_j - \overline{x} \right)^2 \right] = \frac{1}{s_x^2} s_x^2$, ou seja, $\sum_j z_j^2 = 1$. Da mesma forma $\sum_j s_j^2 = 1$. Podemos então pensar em dois vetores unitários.

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

$$d_{xy}^2 = \sum_j \left(z_j - s_j \right)^2$$

$$d_{xy}^2 = \sum_j \left(z_j - s_j \right)^2 = \sum_j z_j^2 + \sum_j s_j^2 - 2 \sum_j z_j s_j = 2 \left(1 - \sum_j z_j s_j \right)$$

Mas agora notamos que $\sum_j z_j s_j = \frac{1}{s_x s_y} \frac{1}{N} \sum_j \left(x_j - \overline{x} \right) \left(y_j - \overline{y} \right) = \frac{\operatorname{cov}(x, y)}{s_x s_y}$ logo $\sum_j z_j s_j = r_{xy}$ e:

$$d_{xy}^2 = 2\left(1 - r_{xy} \right)$$

Então vemos que a grandeza $d_{xy} = \sqrt{2\left(1 - r_{xy} \right)}$ se comporta como uma distância. Chamamos essa distância de distância de correlação. Como $-1 \le r_{xy} \le +1$ a distância de correlação varia entre $0 \le d_{xy} \le 2$. Quanto maior a correlação menor a distância.

Vale notar um ponto importante aqui. Para ser uma distância exigimos que se $d\left(x, y \right) = 0$ então $x = y$. Mas $d\left(x, y \right) = 0$ significa que $r_{xy} = +1$. Duas v.a.s relacionadas da forma $x_j = a\, y_j + b$ com $a > 0$

apresentam correlação $r_{xy} = +1$ embora $x \neq y$. Entretanto, as duas variáveis $z_j = \dfrac{x_j - \overline{x}}{\sqrt{N} s_x}$ e $s_j = \dfrac{y_j - \overline{y}}{\sqrt{N} s_y}$ são iguais, pois:

1. $\overline{x} = \dfrac{1}{N}\sum_j x_j = \dfrac{1}{N}\sum_j a\,y_j + \dfrac{1}{N}\sum_j b = a\dfrac{1}{N}\sum_j y_j + b$, ou seja, $\overline{x} = a\overline{y} + b$.
2. $s_x^2 = \dfrac{1}{N}\sum_j \left(x_j - \overline{x}\right)^2 = \dfrac{1}{N}\sum_j \left(ay_j + b - a\overline{y} - b\right)^2 = a^2\dfrac{1}{N}\sum_j \left(y_j - \overline{y}\right)^2$, ou seja, $s_x^2 = a^2 s_y^2$.
3. $z_j = \dfrac{x_j - \overline{x}}{\sqrt{N} s_x} = \dfrac{ay_j + b - a\overline{y} - b}{\sqrt{N} a s_y} = \dfrac{a}{a}\dfrac{y_j - \overline{y}}{\sqrt{N} s_y}$ logo $z_j = s_j$.

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 $z = x + y$ em que x e y são v.a.s independentes com fdp´s $f_x(x)$ e $f_y(y)$, então a nova v.a. z terá a fdp dada por $f_z(z) = \int_{-\infty}^{+\infty} f_x(x)\, f_y(z-x)\,dx$.

Prova: $F_Z(z) = \iint\limits_{x+y\le z} f(x,y)\,dxdy = \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{z-x} f(x,y)\,dxdy = \int\limits_{-\infty}^{+\infty} f_X(x)\,dx \int\limits_{-\infty}^{z-x} f_Y(y)\,dx_2$

Então[1]: $f_Z(z) = \dfrac{d}{dz}F_Z(z) = \int\limits_{-\infty}^{+\infty} f_X(x)\,dx \dfrac{d}{dz}\int\limits_{-\infty}^{z-x} f_Y(y)\,dx_2$

$$f_Z(z) = \int\limits_{-\infty}^{+\infty} f_X(x)\, f_Y(z-x)\,dx$$

Convolução e Correlação: A operação entre duas funções $f(x)$ e $g(x)$ definida por $c(z) = \int\limits_{-\infty}^{+\infty} f(x)g(z-x)\,dx$ é tão importante que ganhou nome próprio: é chamada de CONVOLUÇÃO e é simbolizada por $c(z) = f * g$. Ela tem uma prima denominada por operação CORRELAÇÃO definida de forma um pouco diferente por $C(z) = f \otimes g = \int\limits_{-\infty}^{+\infty} f(x)g(x-z)\,dx$. Note que a diferença está no argumento da função $g(x)$, o qual na convolução é $z-x$ e na correlação é $x-z$.

Intuição sobre as operações convolução e correlação:

Note que a operação $f(x-a)$ é simplesmente transladar a função $f(x)$ no eixo horizontal pela quantidade a para a direita. Já a $f(x+a)$ translada a função para a esquerda. A figura xxx mostra a função $f(x) = e^{-x}H(x)$, preta, com a $f(x-2)$ em azul e a $f(x+2)$ em vermelho. Note que a curva azul deslocou de 2 para a direita e a vermelha de 2 para a esquerda. Já a operação $f(-x)$ significa uma reflexão

[1] Estamos usando a seguinte regra para derivar integrais: $\dfrac{d}{dz}\int\limits_{r(z)}^{s(z)} f(u)\,du = \dfrac{d}{dz}\left[F[s(z)] - F[r(z)]\right]$ onde $F'(u) = f(u)$ portanto $\dfrac{d}{dz}\int\limits_{r(z)}^{s(z)} f(u)\,du = \dfrac{d}{ds}F[s(z)]\dfrac{ds}{dz} - \dfrac{dF}{dr}[r(z)]\dfrac{dr}{dz}$ ou seja $\dfrac{d}{dz}\int\limits_{r(z)}^{s(z)} f(u)\,du = f[s(z)]\dfrac{ds}{dz} - f[r(z)]\dfrac{dr}{dz}$.

da função em torno do eixo y. A figura xx mostra o gráfico das curvas $f\left(x\right)=e^{-x+1}H\left(x-1\right)$ e $f\left(-x\right)=e^{x-1}H\left(-x-1\right)$.

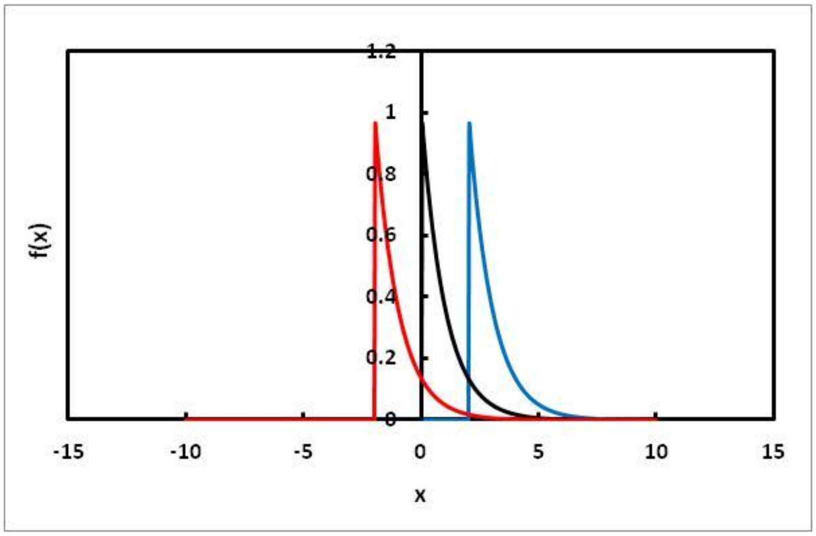


Figura xxx. Gráfico das curvas $f\left(x\right)=e^{-x}H\left(x\right)$ em preto, $f\left(x-2\right)=e^{-x+2}H\left(x-2\right)$ em azul e $f\left(x+2\right)=e^{-x-2}H\left(x+2\right)$ em vermelho.

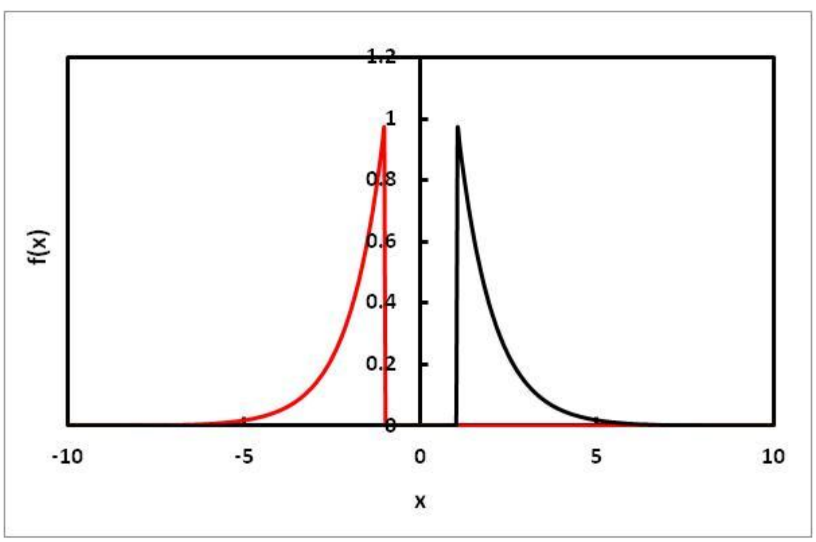


Figura xxx. Gráfico das curvas $f\left(x\right)=e^{-x+1}H\left(x-1\right)$ em preto e $f\left(-x\right)=e^{x+1}H\left(-x-1\right)$ em vermelho.

Vamos analisar uma auto-convolução e uma auto-correlação da função $f(x)$ com ela mesma. Na auto-correlação a $f(x-z)$ é a própria função deslocada por z. Mas na auto-convolução $z-x=-(x-z)$ a função é deslocada e refletida no eixo y.

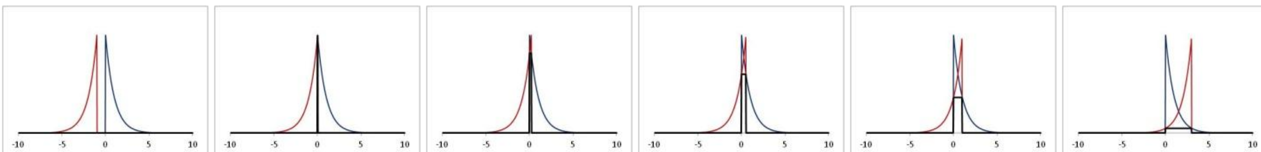


Figura xxx. Multiplicação das curvas $f(x)=e^{-x}H(x)$ por $f(z-x)$ para $z=-1;0;0,5;1;2$ e 4

A figura xxx mostra a curva da autoconvolução $c(z)=\int_{-\infty}^{+\infty} f(x)g(z-x)dx$ em função de z.

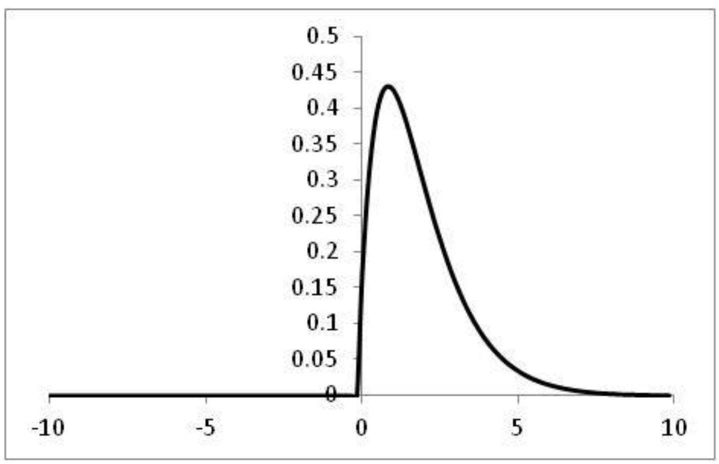


Figura xxx. Autoconvolução de $f(x)=e^{-x}H(x)$ em função de z

Já a figura xxx mostra a multiplicação de $f(x)$ por $f(x-z)$ da auto-correlação e a figura xxx o resultado da auto-correlação em função de z.

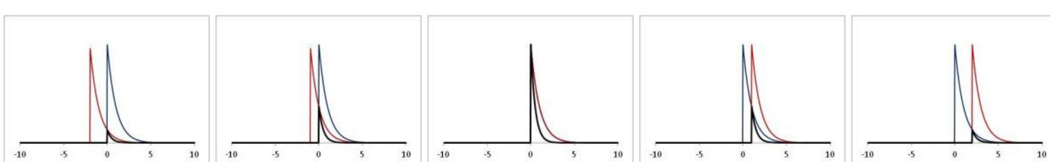


Figura xxx. Multiplicação das curvas $f\left(x\right)=e^{-x}H\left(x\right)$ por $f\left(x-z\right)$ para $z=-2;-1;0;1$ e 2

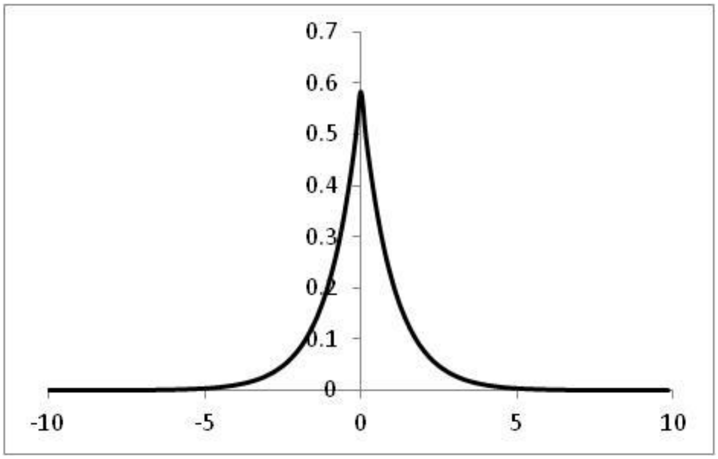


Figura xxx. Auto-correlação de $f\left(x\right)=e^{-x}H\left(x\right)$ em função de z

Propriedades da convolução:

1. $f*g=g*f$

Prova: $g*f=\int_{-\infty}^{+\infty} f\left(z-x\right)g\left(x\right)dx$. Fazendo a mudança de variável $z-x=u$, $g*f=-\int_{+\infty}^{-\infty} f\left(u\right)g\left(z-u\right)du=\int_{-\infty}^{+\infty} f\left(u\right)g\left(z-u\right)du=f*g$. Note que $f\otimes g\neq g\otimes f$, pois $\int_{-\infty}^{+\infty} f\left(x\right)g\left(x-z\right)dx=\int_{-\infty}^{+\infty} f\left(x+z\right)g\left(x\right)dx\neq\int_{-\infty}^{+\infty} f\left(x-z\right)g\left(x\right)dx$.

2. A propriedade distributiva frente à adição $f*\left(g+h\right)=f*g+f*h$ é trivial.

3. Propriedade distributiva frente à convolução: $f*\left(g*h\right)=\left(f*g\right)*h$.

Prova: $f*\left(g*h\right)=\int_{-\infty}^{+\infty} f\left(x\right)g*h\left(x-z\right)dx$ e $g*h\left(z-x\right)=\int_{-\infty}^{+\infty} g\left(u\right)h\left(z-x-u\right)du$, então $f*\left(g*h\right)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f\left(x\right)g\left(u\right)h\left(z-x-u\right)dxdu$. Por outro lado:

$$(f*g)*h=\int_{-\infty}^{+\infty} f*g(x)h(z-x)dx=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(u)g(x-u)h(z-x)dxdu\ .$$

Fazendo $w=x-u$, $x=w+u$ e $dw=dx$ temos que:

$$(f*g)*h=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(u)g(w)h(z-w-u)dwdu\ .$$

Chamando $w=u$ e $u=x$ temos $(f*g)*h=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x)g(u)h(z-x-u)dxdu=f*(g*h)$.

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

Se as v.a.s x e y são independentes então $E[xy]=E[x]E[y]$. Nesse caso então $M_{x+y}(t)=E\left[e^{(x+y)t}\right]=E\left[e^{xt}e^{yt}\right]=E\left[e^{xt}\right]E\left[e^{yt}\right]=M_x(t)\times M_y(t)$

Da mesma forma:

$$\varphi_{x+y}(t)=E\left[e^{i(x+y)t}\right]=E\left[e^{ixt}e^{iyt}\right]=E\left[e^{ixt}\right]E\left[e^{iyt}\right]=\varphi_x(t)\times\varphi_y(t)$$

Ou seja a função geradora dos momentos e a função caraterística da v.a. $z=x+y$ 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: $\varphi_x(t)=\int_{-\infty}^{+\infty} e^{ixt} f_x(x)dx$; $\varphi_y(t)=\int_{-\infty}^{+\infty} e^{iyt} f_y(y)dy$ e $f_z(z)=\int_{-\infty}^{+\infty} f_x(x)\, f_y(z-x)dx$.

Então $\varphi_z(t)=\int_{-\infty}^{+\infty} e^{izt} f_z(z)dz$ é dado por $\varphi_z(t)=\varphi_x(t)\ \varphi_y(t)$.

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

FGM e Função Característica conjuntas:

$$M\left(t_1,t_2\right)=E\left[e^{xt_1+yt_2}\right]=\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty}e^{xt_1+yt_2}\,f\left(x,y\right)dx\,dy$$

$$\varphi\left(t_1,t_2\right)=E\left[e^{ixt_1+iyt_2}\right]=\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty}e^{ixt_1+iyt_2}\,f\left(x,y\right)dx\,dy$$

Relações com os momentos: sabemos que $e^{xt_1+yt_2}=\sum\limits_{n=0}^{\infty}\frac{\left(xt_1+yt_2\right)^n}{n!}$. Por outro lado $\left(xt_1+yt_2\right)^n=\sum\limits_{n=0}^{\infty}\frac{n!}{k!\left(n-k\right)!}x^k t_1^k y^{n-k}t_2^{n-k}$.

Então $E\left[e^{xt_1+yt_2}\right]=\sum\limits_{n=0}^{\infty}\sum\limits_{n=0}^{\infty}\frac{1}{n!}\binom{n}{k}E\left[x^k y^{n-k}\right]t_1^k t_2^{n-k}$

Portanto $M\left(t_1,t_2\right)=\sum\limits_{n=0}^{\infty}\sum\limits_{n=0}^{\infty}\frac{1}{n!}\binom{n}{k}M_{k,n-k}t_1^k t_2^{n-k}$

Claro então que $\varphi\left(t_1,t_2\right)=\sum\limits_{n=0}^{\infty}\sum\limits_{n=0}^{\infty}\frac{i^n}{n!}\binom{n}{k}M_{k,n-k}t_1^k t_2^{n-k}$

A série de Taylor multivariada é dada por:

$$f\left(x,y\right)=\sum_{n=0}^{\infty}\frac{1}{n!}\left[\left(x\partial_x+y\partial_y\right)^n\ f\left(x,y\right)\right]_{\left(x,y\right)=\left(0,0\right)}=$$

$$=\sum_{n=0}^{\infty}\sum_{k=0}^{n}\frac{1}{n!}\frac{n!}{k!\left(n-k\right)!}x^k y^{n-k}\left[\frac{\partial^n}{\partial x^k\partial y^{n-k}}\ f\left(x,y\right)\right]_{\left(x,y\right)=\left(0,0\right)}=$$

$$=\sum_{n=0}^{\infty}\frac{1}{n!}\left[\sum_{k=0}^{n}\frac{n!}{k!\left(n-k\right)!}\left(x^k\partial_x^k+y^{n-k}\partial_y^{n-k}\right)f\left(x,y\right)\right]_{\left(x,y\right)=\left(0,0\right)}$$

$$f(x,y)=\sum_{n=0}^{\infty}\sum_{k=0}^{n}\frac{1}{n!}\binom{n}{k}x^k y^{n-k}\left[\frac{\partial^n f(x,y)}{\partial x^k \partial y^{n-k}}\right]_{(x,y)=(0,0)}$$

Nesse caso $M(t_1,t_2)=\sum_{n=0}^{\infty}\sum_{k=0}^{n}\frac{1}{n!}\binom{n}{k}\left[\left.\frac{\partial^n M(t_1,t_2)}{\partial t_1^k \partial t_2^{n-k}}\right|_{(t_1,t_2)=(0,0)}\right]t_1^k t_2^{n-k}$

Comparando com a expansão dos momentos vemos que $M_{k,n-k}=\left.\frac{\partial^n M(t_1,t_2)}{\partial t_1^k \partial t_2^{n-k}}\right|_{(t_1,t_2)=(0,0)}$ e que

$$M_{k,n-k}=\frac{1}{i^n}\left.\frac{\partial^n \varphi(t_1,t_2)}{\partial t_1^k \partial t_2^{n-k}}\right|_{(t_1,t_2)=(0,0)}.$$

Produto de v.a.s independentes: se $z=xy$ em que x e y são v.a.s independentes com fdp´s $f_x(x)$ e $f_y(y)$, então a nova v.a. z terá a fdp dada por $f(z)=\int_{-\infty}^{\infty}\frac{1}{|x|}f_x(x)\,f_y\left(\frac{z}{x}\right)dx$.

Prova: Note que $xy\le z$ pode ser escrito como $y\ge\frac{z}{x}\quad se\quad x<0$ ou $y\le\frac{z}{x}\quad se\quad x>0$. Então

$$F(z)=P[xy\le z]=P\left[y\ge\frac{z}{x}\,|\,x<0\right]+P\left[y\le\frac{z}{x}\,|\,x>0\right]$$

$$F(z)=\int_{-\infty}^{0}f_x(x)dx\int_{\frac{z}{x}}^{+\infty}f_y(y)dy+\int_{0}^{\infty}f_x(x)dx\int_{-\infty}^{\frac{z}{x}}f_y(y)dy$$

Então:

$$f(z)=\frac{dF}{dz}=\int_{-\infty}^{0}f_x(x)dx\frac{d}{dz}\int_{\frac{z}{x}}^{+\infty}f_y(y)dy+\int_{0}^{\infty}f_x(x)dx\frac{d}{dz}\int_{-\infty}^{\frac{z}{x}}f_y(y)dy$$

$$f(z)=\frac{dF}{dz}=\int_{-\infty}^{0} f_x(x)\,dx\left(-\frac{1}{x}\right) f_y\left(\frac{z}{x}\right)dx+\int_{0}^{\infty} f_x(x)\left(\frac{1}{x}\right) f_y\left(\frac{z}{x}\right)dx$$

$$f(z)=\int_{-\infty}^{\infty} \frac{1}{|x|} f_x(x)\, f_y\left(\frac{z}{x}\right)dx$$

Quociente: se $z = x + y$ em que x e y são v.a.s independentes com fdp´s $f_x(x)$ e $f_y(y)$, então a nova v.a. $z = xy$ terá a fdp dada por $f(z)=\int_{-\infty}^{\infty} \frac{1}{|x|} f_x(x)\, f_y\left(\frac{z}{x}\right)dx$.

Note que a fdp do quociente $z=\frac{x}{y}$ agora muda para $y \geq zx \quad se \quad x<0$ ou $y \leq zx \quad se \quad x>0$.

Então $F(z)=P\left[\frac{x}{y}\leq z\right]=P\left[y\geq zx \mid x<0\right]+P\left[y\leq zx \mid x>0\right]$

$$F(z)=\int_{-\infty}^{0}\int_{zx}^{+\infty} f_x(x)f_y(y)\,dxdy+\int_{0}^{\infty}\int_{-\infty}^{zx} f_x(x)f_y(y)\,dxdy$$

$$f(z)=\frac{dF}{dz}=\int_{-\infty}^{0} -xf_x(x)\, f_y(zx)\,dx+\int_{0}^{\infty} xf_x(x)\, f_y(zx)\,dx$$

$$f(z)=\int_{-\infty}^{\infty} |x|\, f_x(x)\, f_y(zx)\,dx$$

Apêndice XX: Mudança de coordenadas e Jacobiano:

Na álgebra de vetores em 3 dimensões podemos definir os vetores unitários $\vec{u}_1 = (1,0,0)$, $\vec{u}_2 = (0,1,0)$ e $\vec{u}_3 = (0,0,1)$ de forma que qualquer vetor é escrito como $\vec{V} = V_1\vec{u}_1 + V_2\vec{u}_2 + V_3\vec{u}_3$. Note que os vetores unitários gozam da propriedade de que $\left\|\vec{u}\right\| = 1$, onde $\left\|\vec{V}\right\| = \sqrt{V_1^2 + V_2^2 + V_3^2}$ é a norma do vetor $\vec{V}$. Também definimos o produto vetorial através da seguinte operação:

$$\vec{A}\times\vec{B} = \det\begin{pmatrix} \vec{u}_1 & \vec{u}_2 & \vec{u}_3 \\ A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \end{pmatrix}$$

Sem perda de generalidade podemos colocar o vetor $\vec{A}$ no eixo 1 e o vetor $\vec{B}$ no plano 1-2 através de uma rotação dos eixos. Nesse caso: $\vec{A} = (A,0,0) = A\vec{u}_1$, onde $A = \left\|\vec{A}\right\|$. Se o vetor $\vec{B}$ faz um ângulo θ com o vetor $\vec{A}$ então $\vec{B} = (B\cos\theta, B\sin\theta, 0)$ com $B = \left\|\vec{B}\right\|$. Colocando esses dois vetores no produto vetorial temos que :

$$\vec{A}\times\vec{B} = \det\begin{pmatrix} \vec{u}_1 & \vec{u}_2 & \vec{u}_3 \\ A & 0 & 0 \\ B\cos\theta & B\sin\theta & 0 \end{pmatrix} = \vec{u}_3\begin{pmatrix} A & 0 \\ B\cos\theta & B\sin\theta \end{pmatrix} = AB\sin\theta\,\vec{u}_3$$

Nesse caso $\left\|\vec{A}\times\vec{B}\right\| = AB\left|\sin\theta\right|$. Agora note que a área de um paralelepípedo composto pelos vetores $\vec{A}$ e $\vec{B}$ é dada pela base que vale A multiplicada pela altura que vale $B\left|\sin\theta\right|$. Ou seja $\left\|\vec{A}\times\vec{B}\right\|$ é a área do paralelepípedo entre os dois vetores.

Agora vamos tomar o caso em que $\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x(u,w) \\ y(u,w) \end{pmatrix}$. No plano $u-w$ a posição $\begin{pmatrix} u_o \\ w_o \end{pmatrix}$ é transferida para o plano $x-y$ para a posição $\begin{pmatrix} x_o \\ y_o \end{pmatrix} = \begin{pmatrix} x(u_o,w_o) \\ y(u_o,w_o) \end{pmatrix}$. O vetor posição $\vec{r}$ é dado por:

$\vec{r}\left(u,w\right)=x\left(u,w\right)\vec{u}_1+y\left(u,w\right)\vec{u}_2$. Assim

$$\overrightarrow{\delta r}_u=x\left(u_o+\delta u,w_o\right)\vec{u}_1+y\left(u_o+\delta u,w_o\right)\vec{u}_2-x\left(u_o,w_o\right)\vec{u}_1-y\left(u_o,w_o\right)\vec{u}_2=$$
$$=\left[x\left(u_o+\delta u,w_o\right)-x\left(u_o,w_o\right)\right]\vec{u}_1+\left[y\left(u_o+\delta u,w_o\right)-y\left(u_o,w_o\right)\right]\vec{u}_2=$$
$$=\left[\frac{x\left(u_o+\delta u,w_o\right)-x\left(u_o,w_o\right)}{\delta u}\right]\delta u\,\vec{u}_1+\left[\frac{y\left(u_o+\delta u,w_o\right)-y\left(u_o,w_o\right)}{\delta u}\right]\delta u\,\vec{u}_2$$

Ou seja $\overrightarrow{\delta r}_u=\frac{\partial x}{\partial u}\delta u\,\vec{u}_1+\frac{\partial y}{\partial u}\delta u\,\vec{u}_2$.

$$\overrightarrow{\delta r}_w=x\left(u_o,w_o+\delta w\right)\vec{u}_1+y\left(u_o,w_o+\delta w\right)\vec{u}_2-x\left(u_o,w_o\right)\vec{u}_1-y\left(u_o,w_o\right)\vec{u}_2=$$

Já $=\left[x\left(u_o,w_o+\delta w\right)-x\left(u_o,w_o\right)\right]\vec{u}_1+\left[y\left(u_o,w_o+\delta w\right)-y\left(u_o,w_o\right)\right]\vec{u}_2=$

$$=\left[\frac{x\left(u_o,w_o+\delta w\right)-x\left(u_o,w_o\right)}{\delta w}\right]\delta w\vec{u}_1+\left[\frac{y\left(u_o,w_o+\delta w\right)-y\left(u_o,w_o\right)}{\delta w}\right]\delta w\vec{u}_2$$

Ou seja $\overrightarrow{\delta r}_w=\frac{\partial x}{\partial w}\delta w\vec{u}_1+\frac{\partial y}{\partial w}\delta w\vec{u}_2$.

A área entre eles será

$$\left\|\overrightarrow{\delta r}_u\times\overrightarrow{\delta r}_w\right\|=\left\|\det\begin{pmatrix}\vec{u}_1 & \vec{u}_2 & \vec{u}_3\\ \frac{\partial x}{\partial u}\delta u & \frac{\partial y}{\partial u}\delta u & 0\\ \frac{\partial x}{\partial w}\delta w & \frac{\partial y}{\partial w}\delta w & 0\end{pmatrix}\right\|=\left\|\det\begin{pmatrix}\frac{\partial x}{\partial u} & \frac{\partial y}{\partial u}\\ \frac{\partial x}{\partial w} & \frac{\partial y}{\partial w}\end{pmatrix}\delta u\delta w\vec{u}_3\right\|=\left|\det\begin{pmatrix}\frac{\partial x}{\partial u} & \frac{\partial y}{\partial u}\\ \frac{\partial x}{\partial w} & \frac{\partial y}{\partial w}\end{pmatrix}\right|\delta u\delta w$$

Que nos leva ao elemento de área:

$$dxdy=\left|\det\begin{pmatrix}\frac{\partial x}{\partial u} & \frac{\partial y}{\partial u}\\ \frac{\partial x}{\partial w} & \frac{\partial y}{\partial w}\end{pmatrix}\right|dudw$$

Apêndice XX: Teorema da convolução:

Sejam $\varphi_x(t)=\int_{-\infty}^{+\infty} e^{itx} f_x(x)\,dx$ e $\varphi_y(t)=\int_{-\infty}^{+\infty} e^{ity} f_y(y)\,dy$ e suas transformadas inversas $f_x(x)=\frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-itx}\varphi_x(t)\,dt$ e $f_y(y)=\int_{-\infty}^{+\infty} e^{-ity}\varphi_y(t)\,dt$. A convolução é dada por

$$f_x * f_y = \int_{-\infty}^{+\infty} f_x(x-y)\, f_y(y)\,dy = \int_{-\infty}^{+\infty} f_y(y)\left[\frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-it(x-y)}\varphi_x(t)\,dt\right]dy = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\left[e^{ity} f_y(y)\,dy\right]e^{-itx}\varphi_x(t)\,dt$$

logo $f_x * f_y = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-itx}\varphi_y(t)\varphi_x(t)\,dt = FT^{-1}\left[\varphi_x(t)\varphi_y(t)\right]$. Aplicando a transformada de ambos os lados obtemos que $FT\left[f_x * f_y\right]=\varphi_x(t)\varphi_y(t)$, ou seja, $\varphi_z(t)=\varphi_x(t)\varphi_y(t)$, a função característica da variável z é o produto das funções características das variáveis x e y. O teorema também vale na transformada inversa $FT^{-1}\left[\varphi_x * \varphi_y\right]= f_x(x)\, f_y(x)$, ou seja, $\varphi_x * \varphi_y = FT\left[f_x(x)\, f_y(x)\right]$.

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 n v.a.s independentes e queremos a densidade de probabilidade da variável $z = x_1 + x_2 + \cdots + x_n$. Nesse caso temos:

$$f(z)\,dz = \iiint\limits_{z<\sum x_i \le z+dz} \cdots \int f_1(x_1)\, f_2(x_2)\cdots f_n(x_n)\,dx_1 dx_2\cdots dx_n$$

Em vez de colocar a restrição nos limites das integrais, o que nos levaria a um hiper plano de n dimensões, vamos introduzirr uma delta de Dirac na integral que nos garanta a igualdade $\sum x_i = z$, ou seja, $\delta\left[\sum x_i - z\right]dz$, para a desigualdade $z < \sum x_i \le z + dz$. Incluindo a delta de Dirac na integral temos:

$$f(z)\,dz = dz\int\int_{-\infty}^{+\infty}\cdots\int f_1(x_1)\, f_2(x_2)\cdots f_n(x_n)\,\delta\left[\sum x_i - z\right]dx_1 dx_2\cdots dx_n$$

Agora a delta permitiu liberar os limites de integração. Usando $\delta\left[\sum x_i - z\right] = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{i\left(\sum x_i - z\right)t}\,dt$ obtemos

$$f(z)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}e^{-izt}dt\left\{\int_{-\infty}^{+\infty}f_1(x_1)e^{ix_1t}dx_1\right\}\left\{\int_{-\infty}^{+\infty}f_2(x_2)e^{ix_2t}dx_2\right\}\cdots\left\{\int_{-\infty}^{+\infty}f_n(x_n)e^{ix_nt}dx_n\right\}$$

Ou $f(z)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\varphi_1(t)\varphi_2(t)\cdots\varphi_n(t)e^{-izt}dt$

Aplicando a transformada de Fourier de ambos os lados temos que $\varphi_z(t)=\varphi_1(t)\varphi_2(t)\cdots\varphi_n(t)$.

Um Teorema para a Correlação:

Note que $(f\otimes g)dz=\iint_{z<x-y\leq z+dz}f(x)g(y)dxdy=dz\int_{-\infty}^{+\infty}f(x)g(x-z)dx$ gera a correlação $f\otimes g=\int_{-\infty}^{+\infty}f(x)g(x-z)dx$. Por outro lado, podemos usar a delta de Dirac para liberar os limites de integração $(f\otimes g)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f(x)g(y)\delta(x-y-z)dxdy$. Usando novamente a delta da forma $\delta[x-y-z]=\frac{1}{2\pi}\int_{-\infty}^{+\infty}e^{i(x-y-z)t}dt$ chegamos a $f\otimes g=\frac{1}{2\pi}\int_{-\infty}^{+\infty}e^{-izt}dt\int_{-\infty}^{+\infty}f(x)e^{ixt}dx\int_{-\infty}^{+\infty}g(y)e^{-iyt}dy$. Na integral de y trocar de variável para $-y$ e ficamos com $f\otimes g=\frac{1}{2\pi}\int_{-\infty}^{+\infty}e^{-izt}dt\,FT\left[f(x)\right]\times FT\left[g(-y)\right]$ onde $FT\left[g(-y)\right]=\int_{-\infty}^{+\infty}g(-y)e^{iyt}dy$. Assim, $f\otimes g=FT^{-1}\left[FT\left[f(x)\right]\times FT\left[g(-y)\right]\right]$, ou aplicando a transformada de Fourier de ambos os lados $FT\left[f\otimes g\right]=FT\left[f(x)\right]\times FT\left[g(-y)\right]$.