Síntese dos resultados desse capítulo:

Funções distribuição e densidade de probabilidade:

1. $F(x) = P\left[\{x_v \le x\}\right]$ com F não decrescente, contínua à direita, i.e., $\lim\limits_{x \to x_o^+} F(x) = F(x_o)$, $F(-\infty) = 0$ e $F(+\infty) = 1$.

2. $f(x) = \dfrac{dF(x)}{dx}$ com $f(x) \ge 0$ e $\int\limits_{-\infty}^{+\infty} f(x)\,dx = 1$

3. $F(x) = \int\limits_{-\infty}^{x} f(u)\,du$

4. Mudança de variável $y = g(x)$ contínua: $f(y) = \sum\limits_k \dfrac{f[x_k]}{\left|g'[x_k]\right|} = \sum\limits_k \dfrac{f\left[g^{-1}(y)\right]}{\left|g'\left[g^{-1}(y)\right]\right|}$ onde os x_k são todas as raízes da equação $g(x_k) = y$ ou $x_k = g^{-1}(y)$

Delta de Dirac:

5. $\int\limits_{-\infty}^{+\infty} \delta(x - x_o)\,dx = 1$

6. $\int\limits_{-\infty}^{+\infty} f(x)\delta(x - x_o)\,dx = f(x_o)$

7. $\delta(x - x_o) = \dfrac{1}{2\pi} \int\limits_{-\infty}^{+\infty} e^{i(x - x_o)t}\,dt$

8. $\delta(x - x_o) = \dfrac{d}{dx} H(x - x_o)$ com $H(x - x_o) = \begin{cases} 0 & x < x_o \\ 1 & x \ge x_o \end{cases}$

Derivando distribuições descontínuas:

9. Se $F(x) = F_1(x) + \left[F_2(x) - F_1(x)\right] H(x - x_o) = \begin{cases} F_1(x) & x < x_o \\ F_2(x) & x \ge x_o \end{cases}$ então:

$$f(x) = f_1(x) + \left[f_2(x) - f_1(x)\right] H(x - x_o) + \left[\Delta f(x_o)\right] \delta(x - x_o)$$

com $\Delta f(x_o) = \lim\limits_{x \to x_o^+} f(x) - \lim\limits_{x \to x_o^-} f(x)$

Operação Esperança:

10. $E\left[g(x)\right]=\int_{-\infty}^{+\infty} g(x)f(x)dx$

Momentos de ordem n:

11. $M_n = E\left[x^n\right]=\int_{-\infty}^{+\infty} x^n f(x)dx$

11.1. $M_0 = 1$

11.2. $M_1 = E\left[x\right]=\mu$ é a esperança de x.

Momentos centrados de ordem n:

12. $m_n = E\left[(x-\mu)^n\right]=\int_{-\infty}^{+\infty} (x-\mu)^n f(x)dx$

12.1. $\sigma^2 = m_2 = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)dx$ é a variância de x.

Relações entre os momentos centrados e não centrados:

13. $M_n = \sum_{k=0}^{n}\binom{n}{k}\mu^{n-k}m_k$ ou $m_n = \sum_{k=0}^{n}\binom{n}{k}(-\mu)^{n-k} M_k$

Função geradora dos momentos:

14. $M(t)=E\left[e^{xt}\right]=\int_{-\infty}^{+\infty} e^{xt} f(x)dx$

14.1. $M(0)=1$

14.2. $M_n = \left.\frac{d^n M(t)}{dt^n}\right|_{t=0}$

14.3. $m_n = \left.\frac{d^n\left[e^{-\mu t}M(t)\right]}{dt^n}\right|_{t=0}$

Função característica:

15. $\varphi(t)=E\left[e^{ixt}\right]=FT\left[f(x)\right]=\int_{-\infty}^{+\infty}e^{ixt}f(x)\,dx$ com $\varphi(t)$ absolutamente contínua

15.1. $\varphi(0)=1$

15.2. $\left|\varphi(t)\right|\leq 1$

15.3. $M_n=\frac{1}{i^n}\left.\frac{d^n\varphi(t)}{dt^n}\right|_{t=0}$

15.4. $m_n=\frac{1}{i^n}\left.\frac{d^n\left[e^{-i\mu t}\varphi(t)\right]}{dt^n}\right|_{t=0}$

16. $f(x)=FT^{-1}\left[\varphi(t)\right]=\frac{1}{2\pi}\int_{-\infty}^{+\infty}e^{-ixt}\varphi(t)\,dx$

Função geradora dos cumulantes:

17. $C(t)=\ln\varphi(t)$

18. $C(t)=\sum_{k=0}^{\infty}\frac{i^k c_k}{k!}t^k$ onde c_k é o cumulante de ordem k .

19. $c_0=0$ e $c_k=(-i)^k\left.\frac{d^k}{dt^k}\ln\left[\varphi(t)\right]\right|_{t=0}$.

20. Relações entre os momentos centrados, não centrados e cumulantes até ordem 4:

Momentos não centrados	**Momentos centrados**	**Cumulantes**
$M_o=1$	$m_o=1$	$c_o=0$
$M_1=\mu$	$m_1=0$	$c_1=\mu$
$M_2=\mu^2+\sigma^2$	$m_2=M_2-\mu^2$	$c_2=\sigma^2$
$M_3=\mu^3+3\mu\sigma^2+m_3$	$m_3=M_3-3\mu M_2+2\mu^3$	$c_3=m_3$
$M_4=\mu^4+6\mu^2\sigma^2+4\mu m_3+m_4$	$m_4=M_4-4\mu M_3+6\mu^2 M_2-3\mu^4$	$c_4=m_4-3\sigma^4$

21. Cumulantes normalizados: $\alpha_k = \frac{c_k}{\sigma^k}$

21.1. Skewness $\alpha_3 = \frac{c_3}{\sigma^3}$. Se $\alpha_3 < 0$ assimétrica à esquerda [skewed-to-the-left] e se $\alpha_3 > 0$ assimétrica à direita [skewed-to-the-right].

21.2. Curtose $k = \frac{c_4}{\sigma^4} = \frac{m_4}{\sigma^4} - 3$. Se $k > 0$ temos uma distribuição leptocúrtica, se $k > 0$ temos uma distribuição platicúrtica e se $k = 0$ temos uma distribuição mesocúrtica.

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 $f:\Omega\to\mathbb{R}$ que permite associar um número a cada evento. Assim poderemos trocar $P(A)$ por $P(x)$ onde x é uma variável aleatória definida pela função de conjunto $x(A):\Omega\to\mathbb{R}$ é uma função de conjunto com imagem R_x. 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 x_v a função e x o seu valor. Para ser uma variável aleatória a função de conjunto precisa satisfazer poucas condições.

1. O conjunto $\{x_v \le x\}$ é um evento para $\forall\, x\in\mathbb{R}$
2. $P(\{x_v=\infty\})=P(\{x_v=-\infty\})=0$

Note que a um conjunto evento do espaço amostral estamos associando uma probabilidade e um valor da variável aleatória x_v. 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 $x_v=x$ qual o conjunto A a ele associado e qual $P(A)$. Como garantir que o mapeamento inverso $x_v^{-1}(x):\mathbb{R}\to\Omega$ tenha a estrutura definida para a probabilidade?

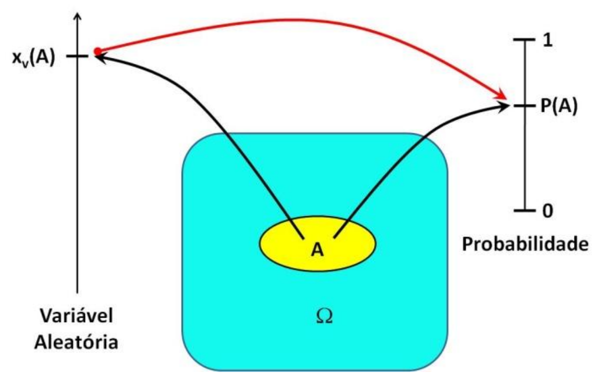


Figura xxx. Mapeamento do conjunto dos eventos para a probabilidade e para a variável aleatória. Seta vermelha: mapeamento direto da variável aleatória para a probabilidade.

Suponha que o conjunto de pontos $B\subset R_x$ seja $x_v^{-1}(B)\equiv\{A/\,A=x_v^{-1}(x)\quad\forall x\in B\}$. Esse conjunto tem que ser um evento. Assim, dado o espaço de probabilidade $[\Omega,\Im,p]$ a função $x_v(A)=x$ é uma função variável aleatória relativa ao campo de Borel $\Im$ se, e somente se, ela é uma função com domínio Ω e imagem R_x tais que $\{A/\,x_v(A)\le x,\, A\in\Omega\}\in\Im$ para todo $x\in R_x$.

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: $\Omega=\{(H,H);(H,T);(T,H);(T,T)\}$. O conjunto de todos os possíveis subconjuntos terá $2^4=16$ elementos. Vejamos se

$\Im_1=\{\phi;\Omega;\{(T,H)\};\{(H,H);(H,T);(T,T)\}\}$ é uma álgebra. Nesse caso $\phi\in\Im_1$, $\overline{\phi}=\Omega\in\Im_1$, $\{\overline{(T,H)}\}=\{(H,H);(H,T);(T,T)\}\in\Im_1$ e $\{\overline{(H,H);(H,T);(T,T)}\}=\{(T,H)\}\in\Im_1$. Assim todos A_i e $\overline{A}_i$ pertencem a $\Im_1$. Vejamos as uniões: $\phi+\Omega=\Omega\in\Im_1$; $\phi+A_i=A_i\in\Im_1$; $\Omega+A_i=\Omega\in\Im_1$ e $A_i+A_i=A_i\in\Im_1$. Sobrou apenas a união $\{(H,H);(H,T);(T,T)\}+\{(T,H)\}=\Omega\in\Im_1$. Então $\Im_1$ é uma álgebra.

Vejamos se $\Im_2=\{\phi;\Omega;\{(H,H);(T,T)\}\}$ é uma álgebra. Basta notar que $\{\overline{(H,H);(T,T)}\}=\{(H,T);(T,H)\}\notin\Im_2$ para decidir que $\Im_2$ não é uma álgebra.

Vejamos

$$\Im_3=\begin{Bmatrix}\phi;\Omega;\{(H,H)\};\{(T,T)\};\{(H,H);(T,T)\};\{(H,T);(T,H)\};\{(H,T);(T,H);(H,H)\};\\ \{(H,T);(T,H);(T,T)\}\end{Bmatrix}$$

Chamando $A_1=\{(H,H)\}$; $A_2=\{(T,T)\}$; $A_3=\{(H,H);(T,T)\}$;

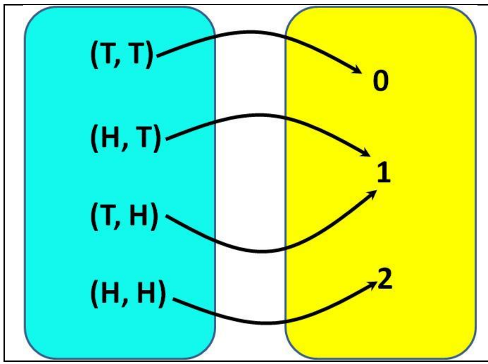
$A_4=\{(H,T);(T,H)\}$ $A_5=\{(H,T);(T,H);(H,H)\}$ e $A_6=\{(H,T);(T,H);(T,T)\}$

notamos que $\overline{A}_1=A_6$, $\overline{A}_2=A_5$, $\overline{A}_3=A_4$, $A_3=A_1+A_2$, $A_5=A_1+A_4$ e $A_6=A_2+A_4$, logo $\Im_3$ é uma álgebra.

Vamos agora definir a variável aleatória $x_v:A\to\mathbb{R}$ como:

$$x_v(A)=\begin{cases}0 & se\ A=\{(T,T)\}\\ 1 & se\ A\in\{(H,T);(T,H)\}\\ 2 & se\ A=\{(H,H)\}\end{cases}$$

Ou seja $x_v(A)$ é o número de vezes em que H aparece. A pergunta é $x_v(A)$ é uma v.a. frente a $\Im_3$? Vejamos quem é $x_v^{-1}[\forall x\le x_o]$.



$$x_v^{-1}\left[(-\infty,x)\right]=\begin{cases}\phi \quad se\; x<0\\ \{(T,T)\} \quad se\; 0\le x<1\\ \{(T,T);(H,T);(T,H)\} \quad se\; 1\le x<2\\ \Omega=\{(T,T);(H,T);(T,H),(H,H)\} \quad se\; x\ge 2\end{cases}$$

Como $\phi \in \Im_3$; $\{(T,T)\} \in \Im_3$ $\{(T,T);(H,T);(T,H)\} \in \Im_3$ e $\Omega \in \Im_3$ então x_v é uma v.a. frente a $\Im_3$.

Exemplo que não é v.a.

$x_v(A)=\begin{cases}0 \quad se\; A\in\{(T,T);(T,H)\}\\ 1 \quad se\; A\in\{(H,T);(H,H)\}\end{cases}$ ou seja, $x=1$ se obtém-se H na primeira jogada e zero caso contrário. Note que agora

$$x_v^{-1}\left[(-\infty,x)\right]=\begin{cases}\phi \quad se\; x<0\\ \{(T,T);(T,H)\} \quad se\; 0\le x<1\\ \Omega \quad se\; x\ge 1\end{cases}$$

Mas nesse caso o conjunto $\{(T,T);(T,H)\} \notin \Im_3$ e x não é uma v.a. frente a $\Im_3$. Será que é possível encontrar um $\Im_4$ frente ao qual esse x será uma v.a.? Basta partir dos conjuntos básicos:

$x_v^{-1}(0)=\{(T,T);(T,H)\}$ e $x_v^{-1}(1)=\{(H,T);(H,H)\}$ e gerar uma σ-álgebra através das uniões, interseções e complementos: $x_v^{-1}(0)x_v^{-1}(1)=\phi$, $x_v^{-1}(0)+x_v^{-1}(1)=\Omega$, $\overline{x_v^{-1}(0)}=x_v^{-1}(1)$ e $\overline{x_v^{-1}(1)}=x_v^{-1}(0)$ então $\Im_4=\{\phi;\Omega;\{(T,T);(T,H)\};\{(H,T);(H,H)\}\}$ é o menor σ-field para o qual x é uma v.a.

Um aspecto importante a notar aqui é que se $x_1<x_2$ então os eventos $x_v^{-1}\left[(-\infty,x_1)\right]\subset x_v^{-1}\left[(-\infty,x_2)\right]$.

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 $A=\{x_v\le x\}$ é um evento, podemos calcular $P(A)$ para qualquer valor de x. Assim a a CDF é definida por:

$$F(x) = P(\{x_v \le x\}) \quad x \in \mathbb{R}$$

Exemplo: jogar uma moeda desonesta, com probabilidade p de H e q de T, $p+q=1$, uma vez. Vamos criar a v.a. $x_v(H)=1$ e $x_v(T)=0$. Essa distribuição é conhecida como distribuição de Bernoulli. Nesse caso

$$x_v^{-1}\left[(-\infty, x)\right] = \begin{cases} \phi & se\ x<0 \\ \{T\} & se\ 0 \le x < 1 \\ \Omega = \{\{T\},\{H\}\} & se\ x \ge 1 \end{cases}$$

Como $P(\phi)=0$, $P(\{T\})=q$ e $P(\Omega)=1$ então $F(x) = \begin{cases} 0 & se\ x<0 \\ q & se\ 0 \le x < 1 \\ 1 & se\ x \ge 1 \end{cases}$.

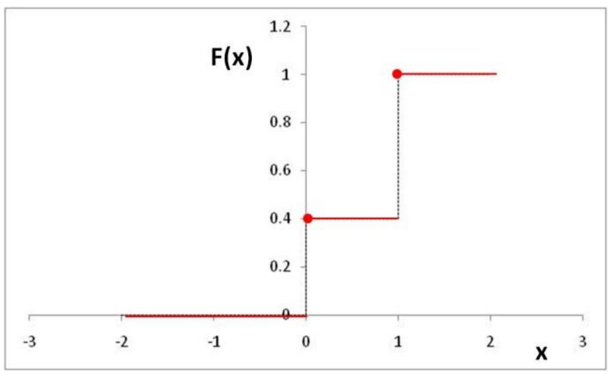


Figura xxx. FDA ou CDF da Distribuição de Bernoulli.

Propriedades da CDF

1. $F(+\infty)=1$ e $F(-\infty)=0$ pois $F(+\infty)=P(\{x_v \le \infty\})=P(\Omega)=1$ e $F(-\infty)=P(\{x_v \le -\infty\})=P(\phi)=0$.
2. Se $x_2 > x_1$ então $F(x_2) \ge F(x_1)$, ou seja, F é sempre crescente. Prova: $\{x_v \le x_1\} \subset \{x_v \le x_2\}$ logo $P(\{x_v \le x_1\}) < P(\{x_v \le x_2\})$ e $F(x_1) \le F(x_2)$.
3. Se $F(x_o)=0$ então $F(x)=0 \quad \forall\ x < x_o$. Como $F(x) \ge 0$ e $F(x \le x_o) \le F(x_o)=0$ então $F(x)=0$.
4. $P(\{x_v > x\}) = 1-F(x)$. Os eventos $E_1=\{x_v \le x\}$ e $E_2=\{x_v > x\}$ são mutuamente exclusivos e $E_1+E_2=\Omega$, logo $P(\Omega)=P(E_1+E_2)=P(E_1)+P(E_2)=1$ e $P(\{x_v > x\}) = 1-P(\{x_v \le x\}) = 1-F(x)$.

5. $F(x)$ é contínua pela direita, ou seja, $F(x^{+}) = F(x)$. Prova: $P(\{x_v \le x+\varepsilon\}) = F(x+\varepsilon)$ e $\underset{\varepsilon \to 0}{Lim}\{x_v \le x+\varepsilon\} = \{x_v \le x\}$ logo $F(x^{+}) = F(x)$.
6. $P(\{x_1 < x_v \le x_2\}) = F(x_2) - F(x_1)$. Os eventos $\{x_v \le x_1\}$ e $\{x_1 < x_v \le x_2\}$ são mutuamente exclusivos e, além disso, $\{x_v \le x_1\} + \{x_1 < x_v \le x_2\} = \{x_v \le x_2\}$ logo $P(\{x_v \le x_1\}) + P(\{x_1 < x_v \le x_2\}) = P(\{x_v \le x_2\})$, ou seja, $F(x_1) + P(\{x_1 < x_v \le x_2\}) = F(x_2)$ logo $P(\{x_1 < x_v \le x_2\}) = F(x_2) - F(x_1)$.
7. $P(\{x_v = x\}) = F(x) - F(x^{-})$. $P(\{x-\varepsilon < x_v \le x\}) = F(x) - F(x-\varepsilon)$ fazendo $\varepsilon \to 0$ temos que $P(\{x_v = x\}) = F(x) - F(x^{-})$. Se F é contínua, então $P(\{x_v = x\}) = 0$ mas senão se F for descontínua, nesse caso, $P(\{x_v = x\}) = \Delta F$ será a descontinuidade no ponto x.

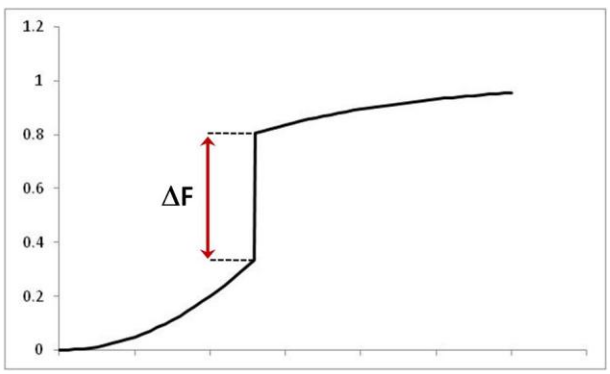
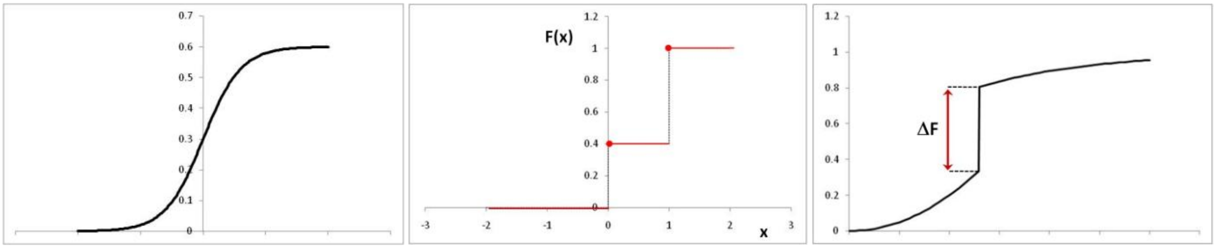


Figura xxx. Descontinuidade na Função Distribuição de Probabilidade fornece $P(\{x_v = x\})$.

8. $P(\{x_1 \le x_v \le x_2\}) = F(x_2) - F(x_1^{-})$. Aqui, novamente, usamos o fato de que $\{x_1 \le x_v \le x_2\} = \{x_1 < x_v \le x_2\} + \{x_v = x_1\}$ e que $\{x_1 < x_v \le x_2\}\{x_v = x_1\} = \phi$, logo $P(\{x_1 \le x_v \le x_2\}) = P(\{x_1 < x_v \le x_2\}) + P(\{x_v = x_1\})$ que pode ser reescrito como $[F(x_2) - F(x_1)] + [F(x_1) - F(x_1^{-})] = F(x_2) - F(x_1^{-})$.

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

a. Contínua
b. Discreta
c. 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 da função distribuição de probabilidade: $f(x)=\frac{dF(x)}{dx}$. 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: $F(x)=\int_{-\infty}^{x} f(x)dx$. 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. $f(x)\geq 0$ pois $\lim_{h\to 0}\frac{F(x+h)-F(x)}{h}\geq 0$ uma vez que $F(x+h)>F(x)\quad se\quad h>0$ e $F(x+h)<F(x)\quad se\quad h<0$, pela propriedade (2) da CDF. Poderíamos simplesmente ter afirmado que se F é sempre crescente então $f(x)=\frac{dF}{dx}\geq 0$.
2. $\int_{-\infty}^{+\infty} f(x)dx=F(\infty)=1$
3. $P(\{x_1<x_v\leq x_2\})=F(x_2)-F(x_1)=\int_{x_1}^{x_2} f(x)dx$
4. $P(\{x<x_v\leq x+dx\})=f(x)dx$, ou seja, $f(x)dx$ é a probabilidade de encontrar a v.a. x_v entre x e $x+dx$.

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:

$$\delta_{ij} = \begin{cases} 1 & se \quad i = j \\ 0 & se \quad i \neq j \end{cases}$$

A matriz identidade pode ser escrita em termos do delta de Kronecker como $I_{ij} = \delta_{ij}$. Em particular ela tem a propriedade de que:

$$\sum_{j=n_1}^{n_2} a_j \delta_{ij} = \begin{cases} a_i & se \quad i \in [n_1, n_2] \\ 0 & se \quad i \notin [n_1, n_2] \end{cases}$$

Pois o único termo não nulo do produto $a_j \delta_{ij}$ será o termo com $j = i$.

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

$$\int_a^b f(x)\delta(x - x_o)\,dx = f(x_o)\mathbf{1}_{[a,b]}(x_o) = f(x_o)\begin{cases} 1 & se \quad x_o \in [a,b] \\ 0 & se \quad x_o \notin [a,b] \end{cases}$$

Nesse caso queremos $\delta(x - x_o) = 0$ se $x \neq x_o$, entretanto a área sobre a delta tem que ser 1 pois:

$$\int_{x_o-\varepsilon}^{x_o+\varepsilon} f(x)\delta(x - x_o)\,dx = f(x_o)\int_{x_o-\varepsilon}^{x_o+\varepsilon} \delta(x - x_o)\,dx = f(x_o)$$

Em outras palavras, estamos em busca de uma função que é nula para todo $x \neq x_o$ mas que tenha área unitária, ou seja $\int_{-\infty}^{+\infty} \delta(x - x_o)\,dx = 1$. Note que a exigência de que $\int_{-\infty}^{+\infty} \delta(x - x_o)\,dx = 1$ 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 $\delta_n(x - x_o)$ de largura limitada, ou seja, $\delta_n(x - x_o) \to 0$ quando $|x - x_o| > m$, mas cuja área seja 1 e independente de n. Além disso, é preciso que $n \to \infty$ então $m \to 0$, ou seja, a largura da delta vai a zero. Fazendo o n tender a infinito então teremos a função Delta de Dirac como

$\delta\left(x-x_o\right)=\lim_{n\to\infty}\delta_n\left(x-x_o\right)$. Qualquer função $\delta_n\left(x-x_o\right)$ com as propriedades acima pode ser usada para construir a função Delta de Dirac.

Um exemplo é a função retângulo: $\delta_n\left(x-x_o\right)=\begin{cases}0 & se \quad \left|x-x_o\right|>\dfrac{\Delta}{n} \\ \dfrac{n}{\Delta} & se \quad \left|x-x_o\right|\le\dfrac{\Delta}{n}\end{cases}$.

A área sobre a curva **vale** $\int_{-\infty}^{+\infty}\delta_n\left(x-x_o\right)dx=\int_{x_o-\frac{\Delta}{2n}}^{x_o-\frac{\Delta}{2n}}\frac{n}{\Delta}dx=\frac{n}{\Delta}\frac{\Delta}{n}=1$. Se n 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 $\int_{-\infty}^{+\infty}f\left(x\right)dx=1$. 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 x_o.

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:

$$H\left(x-x_o\right)=\begin{cases}1 & se \quad x\ge x_o \\ 0 & se \quad x<x_o\end{cases}$$

Essa função é descontínua em $x=x_o$ e, portanto, não diferenciável. Agora considere a função logística dada por $H_n\left(x-x_o\right)=\dfrac{1}{1+e^{-n\left(x-x_o\right)}}$. Note que se $x<x_o$ então $x-x_o<0$ e $-n\left(x-x_o\right)>0$ logo para $x\to-\infty$ $H_n\left(x-x_o\right)\sim\dfrac{1}{e^{-n\left(x-x_o\right)}}\to 0$. Já para $x\to\infty$ então $e^{-n\left(x-x_o\right)}\to 0$ e $H_n\left(x-x_o\right)\to 1$. Assim temos que $H_n\left(-\infty\right)=0$ e $H_n\left(+\infty\right)=1$. Para $x=x_o$ $H_n\left(0\right)=\dfrac{1}{2}$. A função H_n é diferenciável $\dfrac{dH_n}{dx}=\dfrac{ne^{-n\left(x-x_o\right)}}{\left[1+e^{-n\left(x-x_o\right)}\right]^2}$ e essa derivada tem obviamente uma área sobre a curva igual a 1, pois $\int_{-\infty}^{+\infty}\dfrac{dH_n}{dx}dx=H_n\left(+\infty\right)-H_n\left(-\infty\right)=1-0=1$. 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 H_n e $\frac{dH_n}{dx}$ para $n = 1, 2, 4, 10\ e\ 20$.

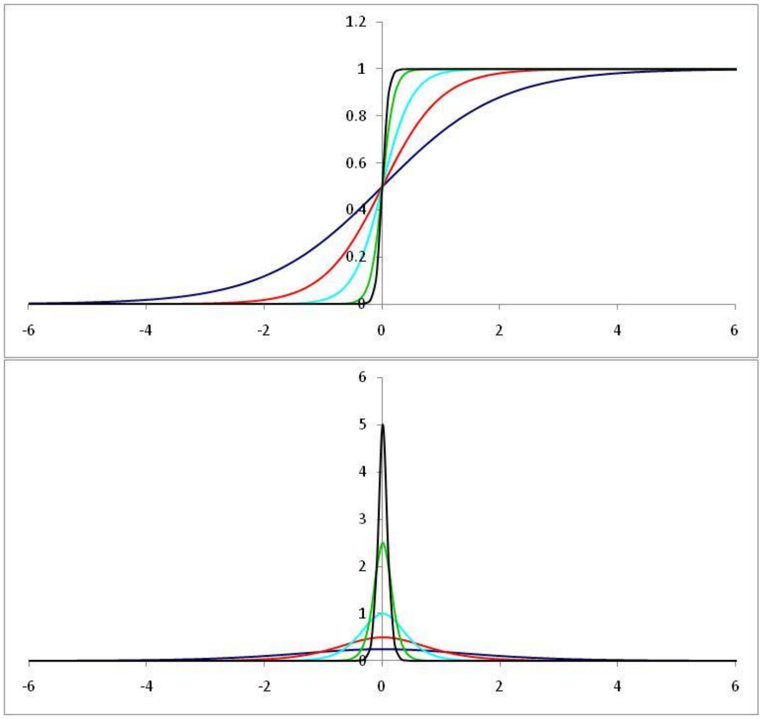


Figura xxx. Função Logística H_n e e sua derivada $\frac{dH_n}{dx}$ para $n = 1, 2, 4, 10\ e\ 20$.

Daí se percebe, então, que para $n \to \infty$, teremos: $\frac{d}{dx} H(x - x_o) = \delta(x - x_o)$.

Derivando funções descontínuas[1]. Uma função descontínua da forma $f(x) = \begin{cases} f_1(x) & se \quad x < x_o \\ f_2(x) & se \quad x \geq x_o \end{cases}$ pode ser escrita como $f(x) = f_1(x) + \left[f_2(x) - f_1(x) \right] H(x - x_o)$. Agora podemos derivar esse função pela regra do produto como:

$$f'(x) = f_1'(x) + \left[f_2'(x) - f_1'(x) \right] H(x - x_o) + \left[f_2(x_o) - f_1(x_o) \right] \delta(x - x_o)$$

Ou seja $f'(x) = f_1'(x) + \left[f_2'(x) - f_1'(x) \right] H(x - x_o) + \Delta f\ \delta(x - x_o)$ onde Δf é a descontinuidade em x_o.

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

[1] 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.

1. $\int_a^b f(x)\delta(x-x_o)\,dx = f(x_o)\mathbf{1}_{[a,b]}(x_o)$

2. $\frac{d}{dx}H(x-x_o) = \delta(x-x_o)$

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: $f(x) = \frac{dF(x)}{dx}$ e $F(x) = \int_{-\infty}^{x} f(x)\,dx$ são válidas em geral. Um bom exemplo é o caso da distribuição de Bernoulli onde $F(x) = qH(x) + pH(x-1)$ e $f(x) = q\delta(x) + p\delta(x-1)$. 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:

$$f(x) = \frac{1}{6}\left[\delta(x-1)+\delta(x-2)+\delta(x-3)+\delta(x-4)+\delta(x-5)+\delta(x-6)\right]$$

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

$$F(x) = \frac{1}{6}\left[\int_{-\infty}^{x}\delta(x-1)\,dx+\int_{-\infty}^{x}\delta(x-2)\,dx+\int_{-\infty}^{x}\delta(x-3)\,dx+\int_{-\infty}^{x}\delta(x-4)\,dx+\int_{-\infty}^{x}\delta(x-5)\,dx+\int_{-\infty}^{x}\delta(x-6)\,dx\right]$$

que leva a $F(x) = \frac{1}{6}\left[H(x-1)+H(x-2)+H(x-3)+H(x-4)+H(x-5)+H(x-6)\right]$ pois $\int_{-\infty}^{x}\delta(x-x')\,dx = H(x-x')$.

Distribuições condicionadas:

Se $P(B) \neq 0$ então $P(A\,|\,B) = \frac{P(AB)}{P(B)}$ então:

$$F(x\,|\,B) = P\left[\{x_v \leq x\,|\,B\}\right] = \frac{P\left[\{x_v \leq x\}B\right]}{P[B]}$$

onde $\{x_v \le x\}B$ é o conjunto dos eventos $A / A \in \{x_v \le x\}$ e $A \in B$, ou seja, é a interseção entre os conjuntos $\{x_v \le x\}$ e B. Agora a fdp condicional é definida por: $f(x \mid B) = \frac{d}{dx} F(x \mid B)$.

Casos especiais:

1. Se $B \equiv \{x_v \le a\}$ então $F(x \mid x_v \le a) = \frac{P[\{x_v \le x\}\{x_v \le a\}]}{P[x_v \le a]}$.

Se $x \ge a$ então $\{x_v \le x\}\{x_v \le a\} = \{x_v \le a\}$ logo:

$$F(x \ge a \mid x_v \le a) = \frac{P[\{x_v \le a\}]}{P[\{x_v \le a\}]} = 1.$$

Agora se $x < a$ então $\{x_v \le x\}\{x_v \le a\} = \{x_v \le x\}$ e:

$$F(x < a \mid x_v \le a) = \frac{P[\{x_v \le x\}]}{P[\{x_v \le a\}]} = \frac{F(x)}{F(a)}.$$

Então podemos expressar, para qualquer valor de x

$$F(x \mid x_v \le a) = \begin{cases} 1 & se\ x \ge a \\ \dfrac{F(x)}{F(a)} & se\ x < a \end{cases}$$

ou $F(x \mid x_v \le a) = \frac{1}{F(a)}\{F(x) + [F(a) - F(x)]H(x-a)\}$

Portanto:

$$f(x \mid x_v \le a) = \frac{1}{F(a)} \frac{d}{dx} F(x \mid x_v \le a)$$

$$f\left(x \mid x_v \le a\right) = \frac{1}{F(a)}\left\{\frac{d}{dx}F(x) - \left[\frac{d}{dx}F(x)\right]H(x-a) + \left[F(a)-F(x)\right]\frac{d}{dx}H(x-a)\right\}$$

$$f\left(x \mid x_v \le a\right) = \frac{1}{F(a)}\left\{f(x) - f(x)H(x-a) + \left[F(a)-F(x)\right]\delta(x-a)\right\}$$

$$f\left(x \mid x_v \le a\right) = \frac{f(x)}{F(a)}\left[1-H(x-a)\right] = \frac{f(x)}{F(a)}H(a-x)$$

Com resultado final:

$$f\left(x \mid x_v \le a\right) = \frac{f(x)}{F(a)}H(a-x) = \frac{f(x)}{\int\limits_{-\infty}^{a} f(x)\,dx}H(a-x)$$

2. Se $B \equiv \{b < x_v \le a\}$ então $F\left(x \mid b < x_v \le a\right) = \dfrac{P\left[\{x_v \le x\}\{b < x_v \le a\}\right]}{P\left[b < x_v \le a\right]}$.

Se $x \ge a$ então $\{x_v \le x\}\{b < x_v \le a\} = \{b < x_v \le a\}$ n, logo:

$$F\left(x \ge a \mid b < x_v \le a\right) = \frac{P\left[\{b < x_v \le a\}\right]}{P\left[\{b < x_v \le a\}\right]} = 1.$$

Se $b \le x < a$ então $\{x_v \le x\}\{b < x_v \le a\} = \{b < x_v \le x\}$ portanto:

$$F\left(b \le x < a \mid b < x_v \le a\right) = \frac{F(x)-F(b)}{F(a)-F(b)}.$$

Se $x < b$ então $\{x_v \le x\}\{b < x_v \le a\} = \phi$ portanto:

$$F\left(x < b \mid b < x_v \le a\right) = 0\,.$$

Então podemos expressar, para qualquer valor de x

$$F\left(x \mid x_v \le a\right) = \begin{cases} 1 & se\ x \ge a \\ \dfrac{F(x) - F(b)}{F(a) - F(b)} & se\ b \le x < a \\ 0 & se\ x < b \end{cases}$$

Em termos da função de Heaviside $H(x)$ temos que nnnnnnnnn $F\left(x \mid b < x_v \le a\right) = \dfrac{F(x) - F(b)}{F(a) - F(b)}\left[H(x-b) - H(x-a)\right] + H(x-a)$.

Daqui se extrai que:

$$f\left(x \mid b < x_v \le a\right) = \frac{d}{dx} F\left(x \mid b < x_v \le a\right)$$

$$f\left(x \mid b < x_v \le a\right) = \frac{\frac{d}{dx}F(x) - \frac{d}{dx}F(b)}{F(a) - F(b)}\left[H(x-b) - H(x-a)\right] + \frac{d}{dx}H(x-a) +$$
$$+ \frac{F(x) - F(b)}{F(a) - F(b)}\left[\frac{d}{dx}H(x-b) - \frac{d}{dx}H(x-a)\right]$$

$$f\left(x \mid b < x_v \le a\right) = \frac{f(x)}{F(a) - F(b)}\left[H(x-b) - H(x-a)\right] + \delta(x-a) + \frac{F(x) - F(b)}{F(a) - F(b)}\left[\delta(x-b) - \delta(x-a)\right] =$$
$$= \frac{f(x)}{F(a) - F(b)}\left[H(x-b) - H(x-a)\right] + \delta(x-a) + \frac{F(b) - F(b)}{F(a) - F(b)}\delta(x-b) - \frac{F(a) - F(b)}{F(a) - F(b)}\delta(x-a) =$$
$$= \frac{f(x)}{F(a) - F(b)}\left[H(x-b) - H(x-a)\right] + \delta(x-a) - \delta(x-a)$$

Logo:

$$f\left(x \mid b < x_v \le a\right) = \frac{f(x)}{F(a) - F(b)} H(x-b) H(a-x) = \frac{f(x)}{\int_a^b f(x)\,dx} H(x-b) H(a-x)$$

Note que a operação realizada foi a truncagem da distribuição no intervalo entre $b < x \le a$ e que foi necessário dividir tudo por $\int_a^b f(x)dx$ para garantir que $\int_{-\infty}^{+\infty} f(x \mid b < x_v \le a)dx = 1.$

Teoremas da Probabilidade Total e de Bayes:

1. $P[x_v \le x] = P[x_v \le x \mid A_1]P(A_1) + \cdots + P[x_v \le x \mid A_n]P(A_n)$ então:

$$F(x) = F(x \mid A_1)P(A_1) + F(x \mid A_2)P(A_2) + \cdots + F(x \mid A_n)P(A_n)$$

$$f(x) = f(x \mid A_1)P(A_1) + f(x \mid A_2)P(A_2) + \cdots + f(x \mid A_n)P(A_n)$$

2. $P(A \mid B) = \dfrac{P(B \mid A)P(A)}{P(B)}$ então:

$$P(A \mid x_v \le x) = \frac{P(x_v \le x \mid A)}{P(x_v \le x)}P(A) = \frac{F(x \mid A)}{F(x)}P(A)$$

3. $P(A \mid x_1 < x_v \le x_2) = \dfrac{P(x_1 < x_v \le x_2 \mid A)}{P(x_1 < x_v \le x_2)}P(A) = \dfrac{F(x_2 \mid A) - F(x_1 \mid A)}{F(x_2) - F(x_1)}P(A)$

4. $P(A \mid x_v = x) = \lim_{h \to 0} P(A \mid x < x_v \le x + h) = \lim_{h \to 0} \dfrac{F(x+h \mid A) - F(x \mid A)}{F(x+h) - F(x)}P(A)$

$$P(A \mid x_v = x) = \lim_{h \to 0} \frac{\dfrac{F(x+h \mid A) - F(x \mid A)}{h}}{\dfrac{F(x+h) - F(x)}{h}}P(A) = \frac{f(x \mid A)}{f(x)}P(A)$$

5. $P(A \mid x_v = x)f(x) = f(x \mid A)P(A)$

$$\int_{-\infty}^{+\infty} P(A \mid x_v = x)f(x)dx = P(A)\int_{-\infty}^{+\infty} f(x \mid A)dx = P(A)$$

$$P(A)=\int_{-\infty}^{+\infty} P(A \mid x_v = x)\, f(x)\, dx$$

6. $$f(x \mid A)=\frac{P(A \mid x_v = x)}{P(A)} f(x)=\frac{P(A \mid x_v = x)\, f(x)}{\int_{-\infty}^{+\infty} P(A \mid x_v = x)\, f(x)\, dx}$$

Função de uma Variável Aleatória

Uma nova v.a. $y=g(x)$ pode ser criada a partir de uma v.a. x desde que os seguintes requisitos sejam satisfeitos:

1. O conjunto $\{g(x)\leq y\}$ é um evento.

1. Os eventos $\{g(x)=\pm\infty\}$ devem ter probabilidade nula, ou seja, $\{g(x)=\pm\infty\}=\phi$.

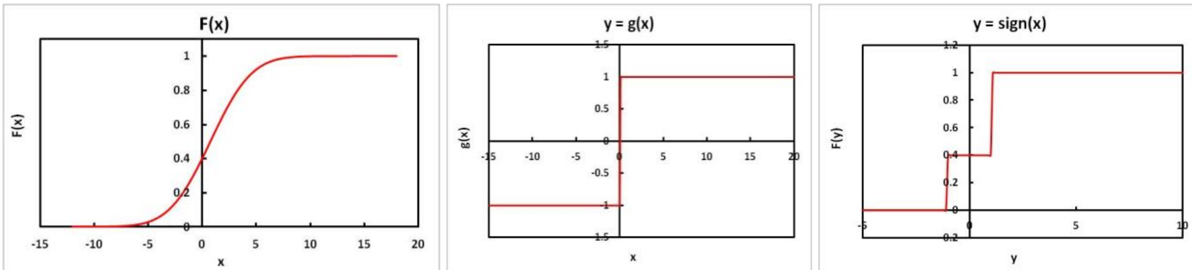
2. Imagem de x está contida no domínio de g.

Note a necessidade desses requisitos. Se $\{g(x)\leq y\}$ não é um evento não existe probabilidade associada ao mesmo. O segundo requisito garante que $f_y(\pm\infty)=0$, exigido para uma FDP. O terceiro é um pouco mais sutil. Precisamos ter certeza de que ao varrer y todos os valores possíveis de x, ou seja, a imagem da função de conjuntos $x=x_v(A)$, estarão incluídos. Não podem faltar valores de x nem pode haver superposição de intervalos de x. A ausência de superposição é garantida pelo fato de que $g(x)$ é uma função, ou seja, o mesmo valor de x só pode ser associado a apenas um valor de $y=g(x)$. Podemos calcular $F_y(y)$ da seguinte forma:

1. Encontrar todos os intervalos de x para os quais $g(x)\leq y$

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

Note que $g(x)$ pode ser inclusive descontínua, constante, divergir, que mesmo assim poderemos encontrar a nova distribuição de probabilidade. Vejamos alguns exemplos dos casos mais patológicos. Para simplificar considere que x segue uma distribuição contínua bem comportada, como a normal da figura 17(a), por exemplo. Agora vamos fazer $y=sign(x)$ mostrada na figura 17(b). Note que nesse caso o conjunto $g(x)\leq y<-1$ é vazio, logo tem probabilidade nula; o conjunto $g(x)\leq y$, para qualquer $0 \le y < 1$, corresponde ao conjunto $x \le 0$, ou seja, com probabilidade $p = 0,4$ pelo gráfico da $F(x)$. Note que, por outro lado, $g(x) \le y$ para qualquer $1 \le y$ representa todo o espaço amostral, $\forall x \in \mathbb{R}$, logo é associado à probabilidade 1. A função distribuição de probabilidade de y 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)

Suponha agora o caso em que $y(x) = \begin{cases} -1 & se\ x \le -1 \\ x & se\ -1 < x \le +1 \\ +1 & se\ x > 1 \end{cases}$ mostrada na figura 18. O conjunto $g(x) \le y$ para $y < -1$ é vazio, logo tem probabilidade nula. O conjunto $g(x) \le y$ para $-1 \le y < +1$ é $x \le y$ e com probabilidade $F_X(y)$ e o conjunto $g(x) \le y$ para $+1 \le y$ corresponde a todo o espaço amostral $x \in \mathbb{R}$ com probabilidade 1.

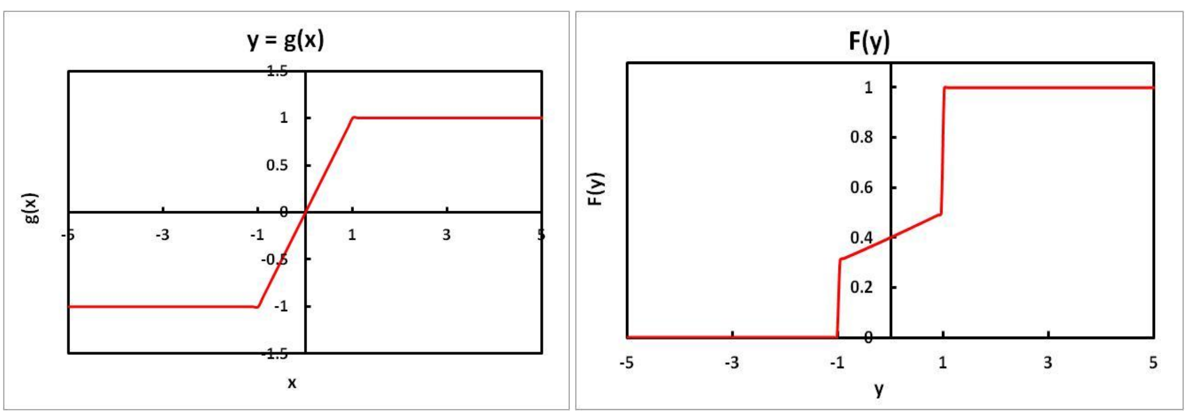


Figura 18 - Gráficos da $g(x)$ e da $F(y)$. Note que a $F(y)$ é nula para $y < -1$, salta para $F_x(-1)$ em $x = -1$, acompanha a curva $F_x(x)$ no intervalo entre -1 e +1 e salta para o valor de 1 a seguir.

Vamos tomar o caso em que y é descontínua, por exemplo, $y(x)=\begin{cases} x & se\ x \le 0 \\ x+1 & se\ x>0 \end{cases}$, como mostra a figura 19. Nesse caso o conjunto $g(x)\le y$ para $y \le 0$ será $x \le y$ e $F_y(y)=F_x(x)$. Já o conjunto $g(x)\le y$ para $0<y\le 1$ será $x\le 0$ e $F_y(y)=F_x(0)$. Isso significa que na descontinuidade de y sua FDA permanece constante, como se percebe no gráfico da $F(y)$ da figura 19. Finalmente o conjunto $g(x)\le y$ para $y>1$ será $x\le y-1$ e $F_y(y)=F_x(y-1)$, que é a própria $F(x)$ transladada de 1 para a direita.

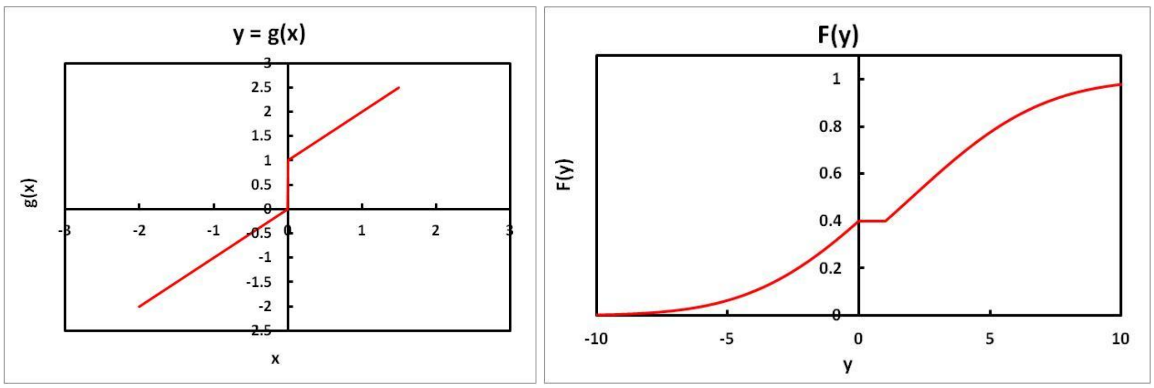


Figura 19 - *g(x)* e *F(y)*

No caso especial em que $g(x)$ é diferenciável a função densidade de probabilidade da nova variável é dada por $fdp(y)=\sum_k \frac{fdp\left[g^{-1}(y)\right]}{\left|g'\left[g^{-1}(y)\right]\right|}$ onde $g(x_i)=y$.

Prova: $fdp(y)=\lim_{\delta y\to 0}\frac{CDF(y+\delta y)-CDF(y)}{\delta y}=\lim_{\delta y\to 0}\frac{P(y<y_v\le y+\delta y)}{\delta y}$, ou seja, $P(y<y_v\le y+\delta y)=fdp(y)dy$. A pergunta é, então, qual o conjunto de pontos de x que leva ao conjunto para $y<y_v\le y+\delta y$? É aquele em que $y<g(x)\le y+\delta y$, como mostra a figura 20 onde existem 3 raízes $g(x)=y$. Note que o dx_2 da figura é negativo por que $\frac{dg}{dx}$ é negativa nessa região, enquanto dx_1 e dx_3 são positivos.

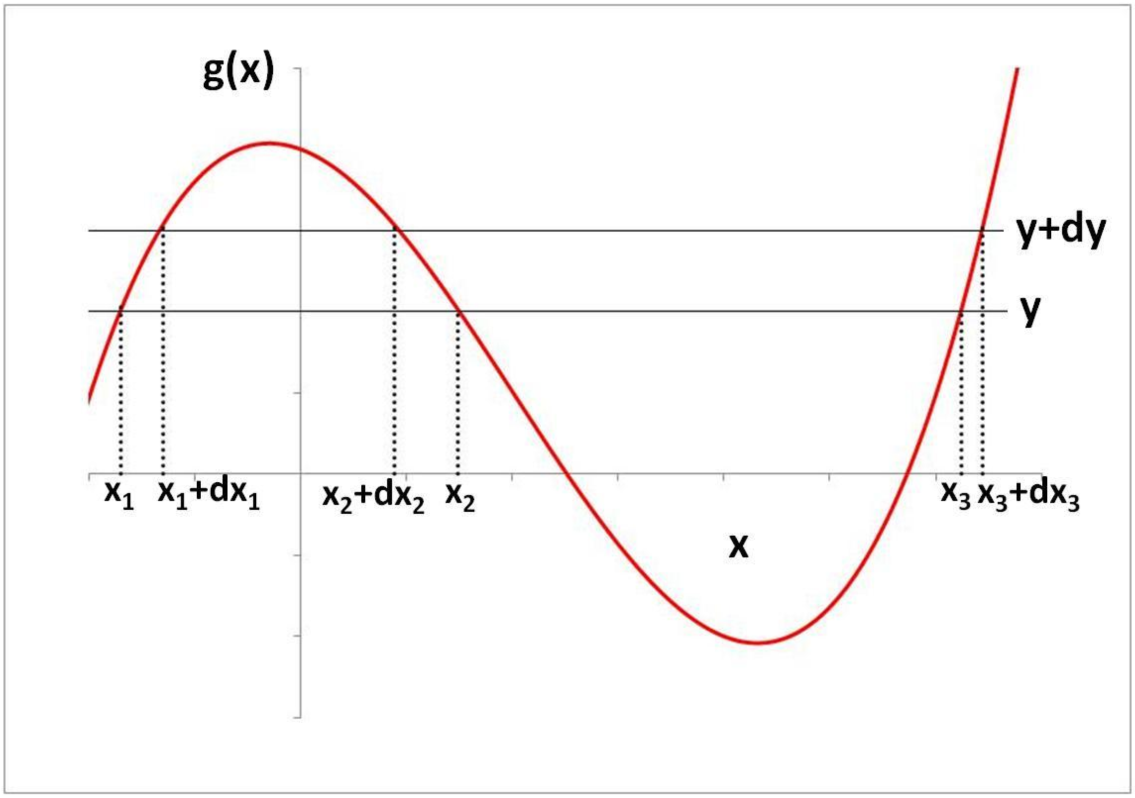


Figura 20 - Três regiões em que $y < g(x) \le y + \delta y$, definindo x_1, x_2 e x_3. Nas regiões 1 e 3 $g'(x)$ é positiva, logo dx_1 e dx_3 também são positivos. Já na região 2 $g'(x)$ é negativa então dx_2 é negativo.

Note que $dy = g'(x)dx$, então $dx = \dfrac{1}{g'(x)}dy$ será negativo onde $g'(x) < 0$. Vamos separar as raízes com $g'(x) > 0$ e denotá-las pelo índice i, das raízes com $g'(x) < 0$ denotadas pelo índice j. Nesse caso:

$$P(y < y_v \le y + \delta y) = \sum_i P(x_i < x_v \le x_i + dx_i) + \sum_j P(x_j + dx_j < x_v \le x_j)$$

com $dx_i > 0$ e $dx_j < 0$. Usando as propriedades da fdp temos que:

$$P(y < y_v \le y + \delta y) = \sum_i fdp(x_i)dx_i - \sum_j fdp(x_j)dx_j = \sum_i fdp(x_i)\frac{dy}{g'(x_i)} - \sum_j fdp(x_j)\frac{dy}{g'(x_j)}$$

$$P(y < y_v \le y + \delta y) = \sum_i fdp(x_i)\frac{dy}{g'(x_i)} + \sum_j fdp(x_j)\frac{dy}{\left[-g'(x_j)\right]} = \sum_k fdp(x_k)\frac{dy}{\left|g'(x_k)\right|}$$

Onde a somatória é feita em k tal que $g(x_k)=y$, independente do sinal de $g'(x)$. O módulo dá conta dos casos em que $g'(x)$ é positiva ou negativa. Com isso temos, no final:

$$fdp(y)=\sum_k \frac{fdp\left[g^{-1}(y)\right]}{\left|g'\left[g^{-1}(y)\right]\right|}$$

Exemplo 1: Vamos transformar a variável x da distribuição normal cuja fdp é dada por $fdp(x)=\frac{1}{\sqrt{2\pi\sigma}}e^{-\frac{x^2}{2\sigma^2}}$ para $y=x^2$, $\operatorname{Im} y=[0,+\infty)$. Note que a função inversa de $x^2=y$ admite duas raízes: $x=+\sqrt{y}$ e $x=-\sqrt{y}$. Não há raízes para $y<0$, logo $fdp(y)=0 \quad se \quad y<0$. Além disso, $g'(x)=\frac{dy}{dx}=2x$ então $g'(x_1)=2\sqrt{y}$ e $g'(x_2)=-2\sqrt{y}$. Nesse caso, então:

$$fdp(y)=\frac{1}{\sqrt{2\pi\sigma}}\left[\frac{e^{-\frac{y}{2\sigma^2}}}{2\sqrt{y}}+\frac{e^{-\frac{y}{2\sigma^2}}}{2\sqrt{y}}\right]H(y)=\frac{1}{\sqrt{2\pi\sigma}}y^{-1/2}e^{-\frac{y}{2\sigma^2}}H(y).$$

Essa é a distribuição $\chi^2\left(\frac{y}{\sigma^2},\nu=1\right)$. A distribuição χ^2 faz parte das distribuições Gama que será apresentada no capítulo de aplicações e distribuições.

Exemplo 2: Distribuição log-Normal. Vamos transformar a variável x da distribuição normal para $y=e^x$, $x=\ln y$. Nesse caso $g(x)=e^x$, $g'(x)=e^x=y$ e a função inversa $g^{-1}(y)=\ln y$. Como a função $y=e^x$ é injetora então só há uma raiz $x=\ln y$ na qual $g'(x)=y$. Não há raízes para $y<0$. O resultado da transformação é:

$$fdp(y)=\frac{1}{\sqrt{2\pi\sigma y}}e^{-\frac{(\ln y)^2}{2\sigma^2}}H(y)$$

Operação ESPERANÇA.

Esperança de uma v.a.: $E[x_v]=\int_{-\infty}^{+\infty} x\, f(x)\,dx$

Os físicos também gostam da notação $\langle x \rangle = E[x]$, as vezes também se usa $\overline{x}$ embora seja necessário tomar cuidado porque $\overline{x} = \frac{1}{n}\sum_i x_i$ é a média obtida em uma amostragem e não a esperança da população completa. Nem sempre $\overline{x} = \langle x \rangle$ embora se espere que sejam próximos, ou seja, $\overline{x}$ é uma boa inferência de $\langle x \rangle = E[x]$. Note que o caso discreto sai automaticamente da utilização das funções delta de Dirac, pois: $E[x_v] = \int_{-\infty}^{+\infty} x \sum_i p_i \delta(x - x_i)\, dx = \sum_i p_i \int_{-\infty}^{+\infty} x\, \delta(x - x_i)\, dx = \sum_i p_i x_i$ pelas propriedades da delta.

Esperança condicional: $E[x_v \mid A] = \int_{-\infty}^{+\infty} x\, f(x \mid A)\, dx$

Esperança de uma $g(x)$ onde x é uma v.a.

Vamos criar a v.a. $y = g(x)$, então $E[y] = \int_{-\infty}^{+\infty} y\, f(y)\, dy$ e mostrar que $E[g(x)] = \int_{-\infty}^{+\infty} g(x)\, f(x)\, dx$. Novamente $f(y)\, dy = \sum_i f(x_i)\, dx_i$ e devemos notar que $x_i \neq x_j$ 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 $y\, f(y)\, dy = \sum_i g(x_i)\, f(x_i)\, dx_i$ e $E[y] = \int_{-\infty}^{+\infty} y\, f(y)\, dy = \int_{-\infty}^{+\infty} g(x)\, f(x)\, dx$. Daí extraímos que:

$$E[g(x)] = \int_{-\infty}^{+\infty} g(x)\, f(x)\, dx$$

Casos particulares:

1. $g(x) = ax$ logo $E[ax] = \int_{-\infty}^{+\infty} axf(x)\, dx = a\int_{-\infty}^{+\infty} xf(x)\, dx = aE[x]$. Constantes entram e saem da operação esperança.

2. $g(x)=q(x)+h(x)$ então $E\left[g(x)\right]=\int_{-\infty}^{+\infty}\left[q(x)+h(x)\right]f(x)dx$,

$E\left[q(x)+h(x)\right]=\int_{-\infty}^{+\infty}q(x)\ f(x)\,dx+\int_{-\infty}^{+\infty}h(x)\ f(x)\,dx$ finalmente chegamos a que a esperança da soma é a soma das esperanças: $E\left[q(x)+h(x)\right]=E\left[q(x)\right]+E\left[h(x)\right]$.

Momentos de ordem n:

O momento de ordem n, se existir, é definido por: $M_n=E[x^n]=\int_{-\infty}^{+\infty}x^n f(x)dx$. A condição para a existência do momento é que a integral acima exista. Se para valores muito grandes de $|x|$, i.e, comportamento assimptótico de f(x) para $x\to\pm\infty$, a f(x) cai com uma lei de potência do tipo $f(x)\propto\frac{1}{x^m}$, então só existirão momento até ordem $n=m-2$. Note que se $n=m-1$ então $M_n\propto\int_{x_o}^{+\infty}\frac{x^n}{x^m}dx=\int_{x_o}^{+\infty}\frac{1}{x}dx\propto\ln(x)$ que diverge.

Algumas propriedades dos momentos são:

1. $M_o=1$ e $M_1=E\left[x\right]=\mu$, pois $M_o=\int_{-\infty}^{+\infty}f(x)dx=1$ e $M_1=E[x]=\int_{-\infty}^{+\infty}x\ f(x)dx$.

Momentos Centrados de ordem n.

O momento centrado de ordem n é definido por: $m_n=E[(x-\mu)^n]=\int_{-\infty}^{+\infty}(x-\mu)^n\ f(x)dx$. Os momentos centrados possuem as propriedades:

1. $m_o=1$, novamente $m_o=\int_{-\infty}^{+\infty}f(x)dx=1$
2. $m_1=0$, pois $m_1=\int_{-\infty}^{+\infty}(x-\mu)\,f(x)dx=\int_{-\infty}^{+\infty}x\ f(x)dx-\mu\int_{-\infty}^{+\infty}f(x)dx=\mu-\mu=0$.

3. A variância é definida pelo $m_2 = \sigma^2$, pois $m_2 = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)dx$.

3.1. Desigualdade de Tchebycheff: A desigualdade de Tchebycheff $P[|x-\mu| \geq \varepsilon] \leq \frac{\sigma^2}{\varepsilon^2}$ demonstra a importância da variância na teoria da probabilidade. Pode-se mostrar essa desigualdade da seguinte forma: $P[|x-\mu| \geq \varepsilon] = \int_{-\infty}^{-\mu-\varepsilon} f(x)dx + \int_{\mu+\varepsilon}^{+\infty} f(x)dx = \int_{|x-\mu|\geq\varepsilon} f(x)dx$. Por outro lado a variância é dada por: $\sigma^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)dx \geq \int_{|x-\mu|\geq\varepsilon} (x-\mu)^2 f(x)dx \geq \varepsilon^2 \int_{|x-\mu|\geq\varepsilon} f(x)dx$, o que leva à desigualdade: $P[|x-\mu| \geq \varepsilon] \leq \frac{\sigma^2}{\varepsilon^2}$.

3.2. **Generalização da desigualdade de Tchebycheff:** Se $f(x) = 0$ para $x < 0$ então $P[x \geq \alpha] \leq \frac{\mu}{\alpha} \quad \forall \alpha > 0$.

Prova: $E[x] = \int_0^\infty x f(x)dx \geq \int_\alpha^\infty x f(x)dx \geq \alpha \int_\alpha^\infty f(x)dx = \alpha P[x > \alpha]$, logo $P[x \geq \alpha] \leq \frac{\mu}{\alpha} \quad \forall \alpha > 0$. A v.a. $|x-a|^n$ só tem valores positivos, logo obedece as condições acima, então: $P\left[|x-a|^n \geq \varepsilon^n\right] \leq \frac{E\left[|x-a|^n\right]}{\varepsilon^n} \quad \forall \varepsilon > 0$. Mas $|x-a|^n \geq \varepsilon^n \quad \Rightarrow \quad |x-a| \geq \varepsilon$, então:

$$P\left[|x-a| \geq \varepsilon\right] \leq \frac{E\left[|x-a|^n\right]}{\varepsilon^n} \quad \forall \varepsilon > 0$$

O caso particular $n = 2$ e $a = \mu$ cai na desigualdade de Tchebycheff.

4. Se $f(x)$ é simétrica, ou seja, $f(\mu + x) = f(\mu - x)$, então todos os momentos centrados ímpares serão nulos.

4.1. Integração em intervalo simétrico de funções pares, $f(-x)=f(x)$, e ímpares, $f(-x)=-f(x)$. Queremos $\int_{-a}^{+a} f(x)dx=\int_{-a}^{0} f(x)dx+\int_{0}^{+a} f(x)dx$. Na primeira integral mudar a variável para $x\to -x$ então $\int_{-a}^{+a} f(x)dx=-\int_{a}^{0} f(-x)dx+\int_{0}^{+a} f(x)dx=\int_{0}^{a} f(-x)dx+\int_{0}^{+a} f(x)dx=\int_{0}^{+a}\left[f(x)+f(-x)\right]dx$.

Se f é par então $\left[f(x)+f(-x)\right]=2f(x)$ e $\int_{-a}^{+a} f(x)dx=2\int_{0}^{+a} f(x)dx$. Já se f é ímpar então $\left[f(x)+f(-x)\right]=0$ e $\int_{-a}^{+a} f(x)dx=0$.

No nosso caso $m_n=\int_{-\infty}^{+\infty}(x-\mu)^n f(x)dx=\int_{-\infty}^{\mu}(x-\mu)^n f(x)dx+\int_{\mu}^{+\infty}(x-\mu)^n f(x)dx$. Mudando, então a variável de integração para $z=x-\mu$, $x=z+\mu$, teremos portanto: $m_n=\int_{-\infty}^{0} z^n f(z+\mu)dz+\int_{\mu}^{+\infty} z^n f(z+\mu)dz=\int_{0}^{\infty}(-z)^n f(\mu-z)dz+\int_{0}^{+\infty} z^n f(\mu+z)dz$, logo se $f(\mu+x)=f(\mu-x)$ e n é ímpar então $m_n=\int_{0}^{+\infty} z^n\left[f(\mu+z)-f(\mu-z)\right]dz=0$.

Relação entre os Momentos Centrados e não centrados.

Podemos usar binômio de Newton $(a+b)^n=\sum_{k=0}^{n}\frac{n!}{k!(n-k)!}a^{n-k}b^k$ para encontrar a relação entre os momentos centrados e não centrados.

$$m_n=E[(x-\mu)^n]=\int_{-\infty}^{+\infty}(x-\mu)^n f(x)dx=\sum_{k=0}^{n}\frac{n!}{k!(n-k)!}(-1)^{n-k}\mu^{n-k}\int_{-\infty}^{+\infty}x^k f(x)dx \text{ logo}$$

$$m_n=\sum_{k=0}^{n}\binom{n}{k}(-1)^{n-k}\mu^{n-k}M_k$$

Casos particulares:

1. $m_o = M_o = 1$
2. $m_1 = \binom{1}{0}(-1)^{1-0}\mu^{1-0}M_o + \binom{1}{1}(-1)^{1-1}\mu^{1-1}M_1 = -\mu M_o + M_1 = -\mu + \mu = 0$
3. $m_2 = \binom{2}{0}(-1)^2\mu^2 M_o + \binom{2}{1}(-1)^1\mu^1 M_1 + \binom{2}{2}(-1)^0\mu^0 M_2 = \mu^2 - 2\mu^2 + M_2 = M_2 - \mu^2$
4. $m_3 = \binom{3}{0}(-1)^3\mu^3 M_o + \binom{3}{1}(-1)^2\mu^2 M_1 + \binom{3}{2}(-1)^1\mu^1 M_2 + \binom{3}{3}M_3 =$
   $= -\mu^3 + 3\mu^3 - 3\mu M_2 + M_3 = M_3 - 3\mu M_2 + 2\mu^3$
5. $m_4 = \mu^4 M_o - 4\mu^3 M_1 + 6\mu^2 M_2 - 4\mu M_3 + M_4 = M_4 - 4\mu M_3 + 6\mu^2 M_2 - 3\mu^4$

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

$$M_n = E[x^n] = E[(x-\mu+\mu)^n] = \int_{-\infty}^{+\infty}\left[(x-\mu)+\mu\right]^n f(x)\,dx = \sum_{k=0}^{n}\frac{n!}{k!(n-k)!}\mu^{n-k}\int_{-\infty}^{+\infty}(x-\mu)^k f(x)\,dx$$

logo: $M_n = \sum_{k=0}^{n}\binom{n}{k}\mu^{n-k}m_k$

1. $M_o = m_o = 1$
2. $M_1 = \mu^1 m_o + \mu^0 m_1 = \mu$ pois $m_o = 1$ e $m_1 = 0$.
3. $M_2 = \mu^2 m_o + 2\mu^1 m_1 + m_2 = \mu^2 + \sigma^2$
4. $M_3 = \mu^3 m_o + 3\mu^2 m_1 + 3\mu^1 m_2 + m_3 = \mu^3 + 3\mu\sigma^2 + m_3$
5. $M_4 = \mu^4 + 6\mu^2\sigma^2 + 4\mu m_3 + m_4$

Função Geradora de Momentos (FGM)

Considere a seguinte função da variável t: $M(t)=E(e^{tx})=\int_{-\infty}^{+\infty} e^{tx} f(x)\,dx$. Podemos usar a expansão em série de Taylor: $e^{tx}=\sum_{n=0}^{\infty}\frac{t^n x^n}{n!}$, para obter: $M(t)=\sum_{n=0}^{\infty}\frac{t^n}{n!}\int_{-\infty}^{+\infty} x^n f(x)\,dx=\sum_{n=0}^{\infty}\frac{t^n}{n!}M_n$. Se comparamos com a série de Taylor da própria $M(t)$ $M(t)=\sum_{n=0}^{\infty}\frac{M^{(n)}(0)}{n!}t^n$ vemos que $M_n=\frac{d^n}{dt^n}\left[M(t)\right]\Big|_{t=0}$. Por isso a função é chamada de geradora dos momentos. Para gerar os momentos centrados devemos multiplicar a função geradora dos momento por $e^{-\mu t}$, uma vez que $e^{-\mu t}M(t)=\int_{-\infty}^{+\infty} e^{t(x-\mu)} f(x)\,dx$, logo $e^{-\mu t}M(t)=\sum_{n=0}^{\infty}\frac{t^n}{n!}\int_{-\infty}^{+\infty}(x-\mu)^n f(x)\,dx=\sum_{n=0}^{\infty}\frac{m_n}{n!}t^n$. Daí se percebe que $m_n=\frac{d^n}{dt^n}\left[e^{-\mu t}M(t)\right]\Big|_{t=0}$.

Função Característica

A grande dificuldade da função geradora dos momentos é a convergência da integral $M(t)=E(e^{tx})=\int_{-\infty}^{+\infty} e^{tx} f(x)\,dx$ por conta do e^{tx} . Se usarmos e^{itx} entretanto, não teremos mais tantos problemas de convergência uma vez que $\left|e^{itx}\right|=1$ para qualquer x e t . Assim a função característica é definida por: $\varphi(t)=E(e^{itx})=\int_{-\infty}^{+\infty} e^{itx} f(x)\,dx$. Se a função geradora dos momentos existe então $\varphi(t)=M(it)$. Note que $\varphi(0)=1$. Além disso, podemos mostrar que $|\varphi(t)|\leq 1$, pois $|\varphi(t)|=\left|\int_{-\infty}^{+\infty} e^{itx} f(x)\,dx\right|\leq\int_{-\infty}^{+\infty}\left|e^{itx} f(x)\right|dx=\int_{-\infty}^{+\infty}\left|e^{itx}\right| f(x)dx=\int_{-\infty}^{+\infty} f(x)dx$. A relação com os momentos só precisa ser ligeiramente modificada uma vez que $e^{itx}=\sum_{n=0}^{\infty}\frac{i^n t^n x^n}{n!}$, levando a: $\varphi(t)=\sum_{n=0}^{\infty}\frac{i^n t^n}{n!}\int_{-\infty}^{+\infty} x^n f(x)\,dx=\sum_{n=0}^{\infty}\frac{i^n M_n}{n!}t^n$. Se comparamos com a série de Taylor $\varphi(t)=\sum_{n=0}^{\infty}\frac{\varphi^{(n)}(0)}{n!}t^n$

vemos que $M_n = (-i)^n \frac{d^n}{dt^n}\left[\varphi(t)\right]\Big|_{t=0}$. Novamente, os momentos centrados podem ser obtidos multiplicando a função característica por $e^{-i\mu t}$, obtendo $m_n = (-i)^n \frac{d^n}{dt^n}\left[e^{-i\mu t}\varphi(t)\right]\Big|_{t=0}$.

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:

$$\varphi(t) = FT\left[f(x)\right] = \int_{-\infty}^{+\infty} e^{ixt} f(x)\,dx$$

A associação entre $\varphi(t)$ e $f(x)$ é biunívoca de modo que ela admite transformada inversa dada por:

$$f(x) = FT^{-1}\left[\varphi(t)\right] = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-ixt}\varphi(t)\,dt$$

Podemos verificar a transformada inversa facilmente, substituindo a $\varphi(t)$ abaixo e usando o fato de que $\delta(x'-x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{i(x'-x)t}dt$, demonstrado no apêndice xxx:

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-ixt}\int_{-\infty}^{+\infty} e^{ix't} f(x')\,dx'dt = \int_{-\infty}^{+\infty} f(x')\,dx'\left[\frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{i(x'-x)t}dt\right] = \int_{-\infty}^{+\infty} f(x')\,\delta(x'-x)\,dx' = f(x)$$

Isso nos permite reconhecer a $f(x)$ dada a $\varphi(t)$ 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 $\varphi(s)=\int_{t_1}^{t_2} K(s,t)\, f(t)\, dt$, onde $K(s,t)$ 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 $\varphi(s)=\int_{-\infty}^{\infty} e^{ist}\, f(t)\, dt$, em que o Kernel é dado por $K(s,t)=e^{ist}$, e a de Laplace $\varphi(s)=\int_{0}^{\infty} e^{-st}\, f(t)\, dt$, em que o Kernel é dado por $K(s,t)=e^{-st}$. 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 transformade 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.

Cumulantes:

A função geradora dos cumulantes é dada por $C(t)=\ln\left[E\left(e^{ixt}\right)\right]=\ln\left[\int_{-\infty}^{+\infty} e^{ixt} f(x)dx\right]=\sum_{k=0}^{\infty}\frac{i^k c_k}{k!}t^k$.

Note que os cumulantes se acumulam, por isso o nome cumulante. Se x e y são independentes então $\varphi_{x+y}(t)=\varphi_x(t)\varphi_y(t)$ e:

$$C_{x+y}(t)=\ln\left[\varphi_x(t)\varphi_y(t)\right]=\ln\left[\varphi_x(t)\right]+\ln\left[\varphi_x(t)\varphi_y(t)\right]=C_x(t)+C_y(t)$$

A operação se torna semelhante então a transformação do log-retorno que transforma operações de multiplicação em operações de adição.

Comparando com a série de Taylor vemos que $c_k = (-i)^k \frac{d^k}{dt^k} \ln\left[\varphi(t)\right]\Big|_{t=0}$. É diferente do $M_k = (-i)^k \frac{d^k}{dt^k}\left[\varphi(t)\right]\Big|_{t=0} = (-i)^k \varphi^{(k)}(0)$, por causa do logaritmo. Podemos extrair a relação entre os cumulantes e os momentos derivando o logaritmo pela regra da cadeia e lembrando que $\varphi(0) = 1$:

1. $\frac{d^0}{dt^0} \ln\varphi = \ln\varphi$; $c_o = 0$

2. $\frac{d}{dt} \ln\varphi = \frac{\varphi'}{\varphi} = \varphi^{-1}\varphi'$; $c_1 = M_1 = \mu$

3. $\frac{d^2}{dt^2} \ln\varphi = \varphi^{-1}\varphi'' - \varphi^{-2}\varphi'\varphi' = \varphi^{-1}\varphi'' - \varphi^{-2}\varphi'^2$; $c_2 = M_2 - \mu^2 = \sigma^2$

4. $\frac{d^3}{dt^3} \ln\varphi = \varphi^{-1}\varphi''' - \varphi^{-2}\varphi'\varphi'' - 2\varphi^{-2}\varphi'\varphi'' - 2\varphi^{-3}\varphi'^3$ logo $\frac{d^3}{dt^3} \ln\varphi = \varphi^{-1}\varphi''' - 3\varphi^{-2}\varphi'\varphi'' + 2\varphi^{-3}\varphi'^3$

   então $c_3 = M_3 - 3\mu M_2 + 2\mu^3$ que pode ser colocado em termos dos momentos centrados como $c_3 = \mu^3 + 3\mu\sigma^2 + m_3 - 3\mu\left(\mu^2 + \sigma^2\right) + 2\mu^3 = m_3$. Logo $\frac{c_3}{\sigma^3} = \frac{m_3}{\sigma^3}$.

5. $\frac{d^4}{dt^4} \ln\varphi = \varphi^{-1}\varphi^{(4)} - 4\varphi^{-2}\varphi'\varphi''' - 3\varphi^{-2}\varphi''^2 + 12\varphi^{-3}\varphi'^2\varphi'' - 6\varphi^{-4}\varphi'^4$, ou seja:

   $c_4 = M_4 - 4M_1M_3 - 3M_2^2 + 12M_1^2M_2 - 6M_1^4$. Colocando em termos dos momentos centrados

   $c_4 = m_4 + 4\mu m_3 - 4\mu m_3 + 6\mu^2\sigma^2 - 12\mu^2\sigma^2 - 6\sigma^2\mu^2 + 12\mu^2\sigma^2 + \mu^4 - 4\mu^4 - 3\mu^4 + 12\mu^4 - 6\mu^4 - 3\sigma^4$

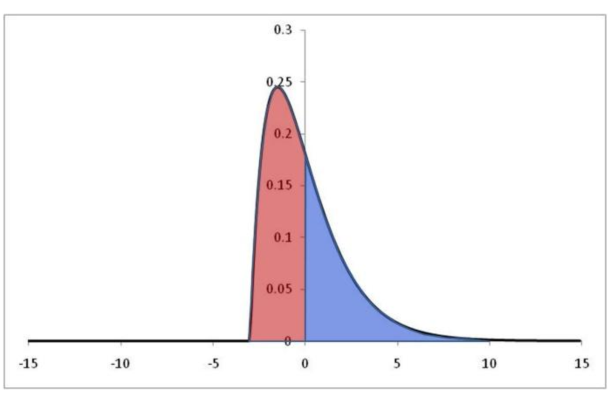
   finalmente $c_4 = m_4 - 3\sigma^4$. $k = \frac{c_4}{\sigma^4} = \frac{m_4}{\sigma^4} - 3$.

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

Momentos não centrados	Momentos centrados	Cumulantes
$M_o = 1$	$m_o = 1$	$c_o = 0$
$M_1 = \mu$	$m_1 = 0$	$c_1 = \mu$
$M_2 = \mu^2 + \sigma^2$	$m_2 = M_2 - \mu^2$	$c_2 = \sigma^2$
$M_3 = \mu^3 + 3\mu\sigma^2 + m_3$	$m_3 = M_3 - 3\mu M_2 + 2\mu^3$	$c_3 = m_3$
$M_4 = \mu^4 + 6\mu^2\sigma^2 + 4\mu m_3 + m_4$	$m_4 = M_4 - 4\mu M_3 + 6\mu^2 M_2 - 3\mu^4$	$c_4 = m_4 - 3\sigma^4$

Skewness a Kurtosis:

Se dividimos $\frac{c_k}{\sigma^k}$ temos uma grandeza adimensional. Até $k=2$ já caracterizamos uma distribuição pela esperança e a variância. Para $k=3$ definimos a skewness, que mede o grau de assimetria da distribuição, através de $\alpha_3 = \frac{m_3}{\sigma^3}$ dada pelo cumulante normalizado. Se $\alpha_3 > 0$ dizemos que a distribuição está skewed to the right e se $\alpha_3 < 0$ está skewed to the left. Note que $m_3 = \int_{-\infty}^{\mu} (x-\mu)^3 f(x)dx + \int_{\mu}^{+\infty} (x-\mu)^3 f(x)dx$ e que a integral da esquerda é sempre negativa enquanto a da direita é sempre positiva. O lado que possuir cauda mais longa [em uma distribuição unimodal] define o sinal da skewness. Figura xxx mostra exemplo dos dois casos.



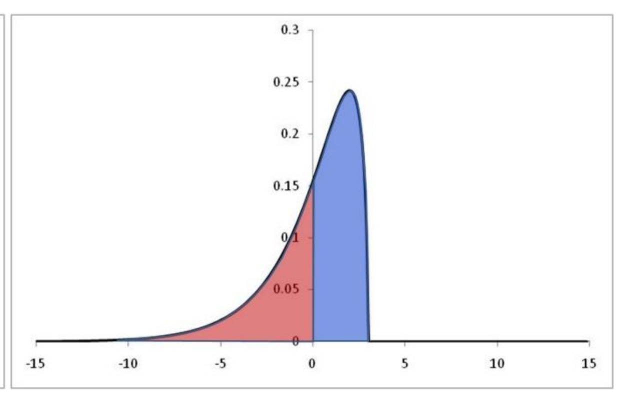


Figura xxx. Esquerda: distribuição skewed to the right. Direita: distribuição skewed to the left.

Outra grandeza que caracteriza a distribuição é a CURTOSE [kurtosis], dada pelo quarto cumulante normalizado, $k = \frac{m_4}{\sigma^4} - 3$. Para comparar duas distribuições é preciso que elas tenham o mesmo σ e m_4 diferentes. Vamos examinar duas distribuições simétricas com curtoses diferentes. A área sobre a curva tem que ser 1, logo iguais. Além disso, queremos o mesmo σ e que f_1 esteja mais concentrada em μ do que f_2. Para manter a área e σ iguais é preciso, então, que f_1 se espalhe mais do que f_2 para x longe de μ. Com isso ela ganha mais m_4, pois $(x-\mu)^4 f(x)$ é maior para x mais longe de μ, e a curtose de f_1 será maior do que a curtose de f_2. Se $k > 0$ a distribuição é chamada de LEPTOCÚRTICA, se $k = 0$ de MESOCÚRTICA e se $k < 0$ de PLATICÚRTICA. Lepto, do grego, significa fino, delgado; plati significa achatado e meso significa médio, no meio. A distribuição platicúrtica por excelência é a distribuição uniforme, em que $k = -6/5$ sempre.

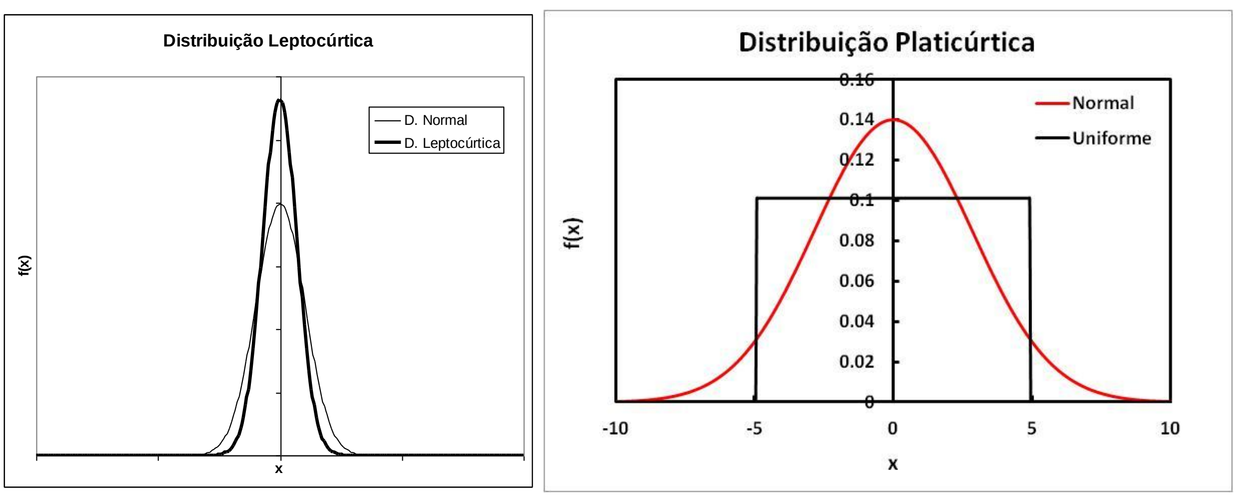


Figura 5.2. Uma distribuição leptocúrtica k < 0 (a) e platicúrtica k > 0 (b).

Apêndices

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

1. $\int_{-\infty}^{+\infty} f(x)\delta'(x-x_o)dx = -f'(x_o)$ com $\delta'(x) = \frac{d}{dx}\delta(x)$

Para mostyrar essa propriedade basta fazer a integral por partes: $\frac{d}{dx}(uw) = w\frac{du}{dx} + u\frac{dw}{dx}$ logo: $udw = d(uw) - wdu$ e $\int_a^b udw = \int_a^b d(uw) - \int_a^b wdu = uw\Big|_a^b - \int_a^b wdu$. No nosso caso $u = f(x)$ e $dw = \frac{d}{dx}\delta(x-x_o)dx$, portanto $w = \delta(x-x_o)$ e $du = f'(x)dx$. Então:

$$\int_{-\infty}^{+\infty} f(x)\delta'(x-x_o)dx = f(x)\delta(x-x_o)\Big|_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} f'(x)\delta(x-x_o)dx = -f'(x_o) \text{ CQD}$$

2. $x\frac{d}{dx}\delta(x) = -\delta(x)$

$\frac{d}{dx}\left[x\delta(x)\right] = x\frac{d}{dx}\delta(x) + \delta(x)\frac{d}{dx}x = x\frac{d}{dx}\delta(x) + \delta(x)$ por outro lado:

$\int_a^b \frac{d}{dx}\left[x\delta(x)\right]dx = x\delta(x)\Big|_a^b = 0$ que nos leva a $\frac{d}{dx}\left[x\delta(x)\right] = 0$ ou seja $x\frac{d}{dx}\delta(x) + \delta(x) = 0$ e $x\frac{d}{dx}\delta(x) = -\delta(x)$ CQD.

3. $\delta\left[f(x)\right] = \sum_j \frac{\delta(x-x_j)}{\left|\frac{df}{dx}\right|_{x_j}}$ onde x_j são as raízes da f, ou seja, $f(x_j) = 0$.

Sabemos que $\int_{-f_o}^{+f_o} \delta(f)\,df = 1$ em torno de uma raiz de f. Suponha que $f'(x_j) > 0$, que é o caso mostrado na figura xx (a). Nesse caso $x_2 > x_1$ onde $f(x_2) = +f_o$ e $f(x_1) = -f_o$. Então:

$$\int_{-f_o}^{+f_o} \delta(f)\,df = \int_{x_1}^{x_2} \delta(f)\frac{df}{dx}dx = 1$$ logo $\delta(f) = \dfrac{\delta(x - x_j)}{\dfrac{df}{dx}}$.

Por outro lado se $f'(x_j) < 0$ nós temos o caso da figura xxx (b) com $x_2 < x_1$. Nesse caso:

$$\int_{-f_o}^{+f_o} \delta(f)\,df = \int_{x_1}^{x_2} \delta(f)\frac{df}{dx}dx = -\int_{x_2}^{x_1} \delta(f)\frac{df}{dx}dx = \int_{x_2}^{x_1} \delta(f)\left[-\frac{df}{dx}\right]dx = 1$$

logo $\delta(f) = \dfrac{\delta(x - x_j)}{-\dfrac{df}{dx}}$.

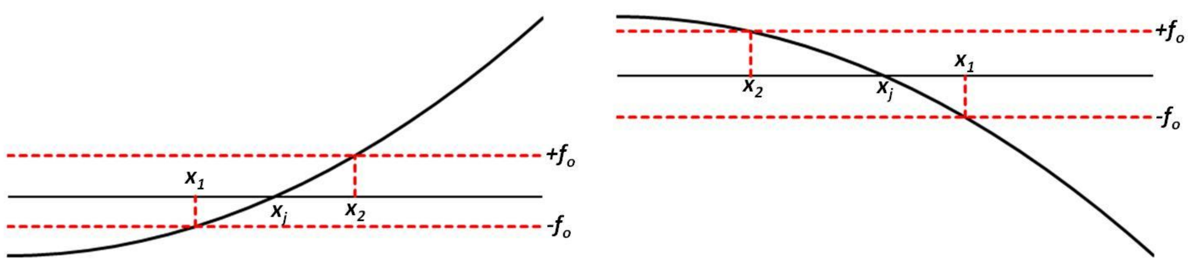


Figura xxx. (a) caso em que $f(x_j) = 0$ e $f'(x_j) > 0$. (b) caso em que $f(x_j) = 0$ e $f'(x_j) < 0$.

Podemos juntar os dois casos afirmando então que $\delta(f) = \dfrac{\delta(x - x_j)}{\left|\dfrac{df}{dx}\right|}$. Finalmente, devemos somar em todas as raízes, obtendo:

$$\delta\left[f(x)\right]=\sum_j \frac{\delta\left(x-x_j\right)}{\left|\frac{df}{dx}\right|_{x_j}}$$

Casos particulares:

a. $\delta\left[a\left(x-x_o\right)\right]=\dfrac{\delta\left(x-x_o\right)}{|a|}$

b. $\delta\left[-1\left(x-x_o\right)\right]=\dfrac{\delta\left(x-x_o\right)}{|-1|} \quad \rightarrow \quad \delta\left(x_o-x\right)=\delta\left(x-x_o\right)$

2. 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:

$$\int_{-\infty}^{+\infty} e^{-x^2} dx = \sqrt{\pi}$$

Para demonstrar esse resultado definimos $I = \int_{-\infty}^{+\infty} e^{-x^2} dx$. Mas como a variável de integração é muda, então $I = \int_{-\infty}^{+\infty} e^{-y^2} dy$ logo $I^2 = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-\left(x^2+y^2\right)} dx\, dy$. Agora podemos mudar de sistema de coordenadas cartesianas para polares no qual $x = r\cos\theta$, $y = r\sin\theta$, $x^2 + y^2 = r^2$ e o elemento de área vale $dx\, dy \to r\, dr\, d\theta$. Para fechar todo o plano x-y, r varia de zero a infinito e θ de 0 a 2π. Neste caso $I^2 = \int_0^{+\infty}\int_0^{2\pi} e^{-r^2} rdr\, d\theta$. A integral em θ é imediata e ficamos com $I^2 = 2\pi \int_0^{+\infty} e^{-r^2} rdr = \pi \int_0^{+\infty} e^{-r^2} 2rdr$.

Agora mudamos a variável para $u = r^2$ logo $du = 2\, r\, dr$ e ficamos com $I^2 = \pi \int_0^{+\infty} e^{-u} du$. Mas $\int_0^{+\infty} e^{-u} du = -e^{-u}\Big|_0^{+\infty} = e^0 = 1$, então $I^2 = \pi$ e $I = \sqrt{\pi}$, ou seja, $\int_{-\infty}^{+\infty} e^{-x^2} dx = \sqrt{\pi}$.

Agora podemos mostrar que $\delta_n\left(x - x_o\right) = \frac{n}{\sqrt{\pi}} e^{-n\left(x-x_o\right)^2}$. 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:

$$\int_{-\infty}^{+\infty} \delta_n\left(x - x_o\right) dx = \frac{n}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-n\left(x-x_o\right)^2} dx = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-n\left(x-x_o\right)^2} d\left[n\left(x - x_o\right)\right] = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-u^2} du = 1$$

Dessa forma $\delta\left(x - x_o\right) = \lim_{n\to\infty}\left[\frac{n}{\sqrt{\pi}} e^{-n\left(x-x_o\right)^2}\right]$.

A distribuição Normal com parâmetros μ e σ segue a forma $f_{Normal}\left(x\right) = Ae^{-\frac{\left(x-\mu\right)^2}{2\sigma^2}}$. Deve-se colocá-la dessa forma porque x, μ e σ possuindo a mesma dimensão tornam a fração $\frac{\left(x-\mu\right)^2}{\sigma^2}$ adimensional. Para ser uma fdp é necessário que $A\int_{-\infty}^{+\infty} e^{-\frac{\left(x-\mu\right)^2}{2\sigma^2}} dx = 1$. Podemos achar o valor de A que torna essa igualdade verdadeira fazendo a mudança de variável $z=\frac{(x-\mu)}{\sqrt{2}\,\sigma}$, $dx=\sqrt{2}\,\sigma dz$, logo $A\sqrt{2}\sigma\int_{-\infty}^{+\infty}e^{-z^2}dz=A\sqrt{2\pi}\sigma=1$, logo $A=\frac{1}{\sqrt{2\pi}\sigma}$ e a fdp da distribuição Normal é dada por:

$$f_{Normal}(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

3. Função Delta Especial

Uma função delta especial pode ser obtida da seqüência $\delta_n\left(x-x_o\right)=A\frac{\sin\left[n\left(x-x_o\right)\right]}{\left(x-x_o\right)}$ cuja curva é mostrada na figura xxx.

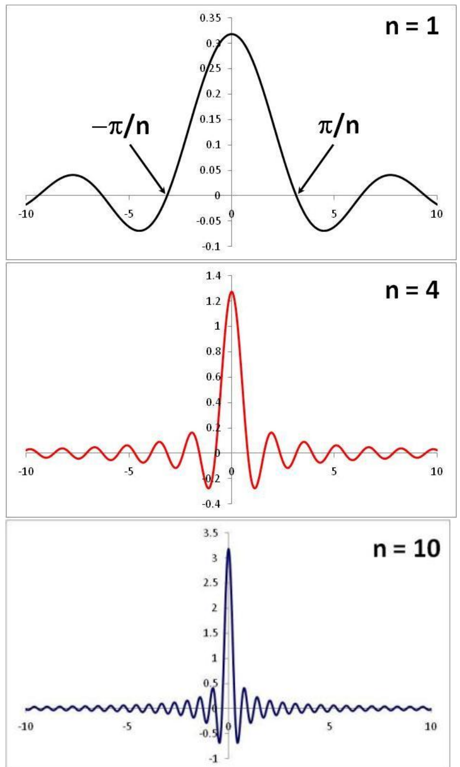


Figura xxx. Gráfico da funçãov $\delta_n\left(x-x_o\right)=\frac{\sin\left[n\left(x-x_o\right)\right]}{\pi\left(x-x_o\right)}$ para $x_o=0$ e $n=1,\ 4\ e\ 10$. Note que a altura sobe com n e a largura diminui. A distância entre as duas primeiras raízes vale $2\pi/n$.

Vamos ver a área sobre essa curva:

$$I=\int_{-\infty}^{+\infty}\delta_n\left(x-x_o\right)dx=A\int_{-\infty}^{+\infty}\frac{\sin\left[n\left(x-x_o\right)\right]}{\left(x-x_o\right)}dx=A\int_{-\infty}^{+\infty}\frac{\sin\left[n\left(x-x_o\right)\right]}{n\left(x-x_o\right)}d\left(nx\right)=A\int_{-\infty}^{+\infty}\frac{\sin u}{u}du$$

Precisamos mostrar que integral $\int_{-\infty}^{+\infty}\frac{\sin u}{u}du$ converge. Note que $\frac{\sin u}{u}$ é uma função par e portanto $\int_{-\infty}^{+\infty}\frac{\sin u}{u}du=2\int_{0}^{+\infty}\frac{\sin u}{u}du$. Por causa do u do denominador de $\frac{\sin u}{u}$ 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 $A_o + A_1 + A_2 + A_3 + \cdots > 0$, mas, por outro lado, que $A_1 + A_2 + A_3 + \cdots < 0$. Então sabemos que $2\left(A_o - |A_1|\right) \le \int_{-\infty}^{+\infty} \frac{\sin u}{u} du \le 2A_o$, logo a integral converge. Sabendo que $\lim_{u\to 0} \frac{\sin u}{u} = 1$ podemos afirmar usando o retângulo de altura 1 e largura π que $0 < \int_{-\infty}^{+\infty} \frac{\sin u}{u} du < 2\pi$.

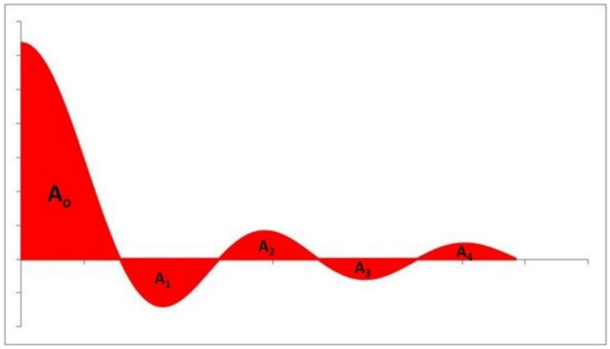


Figura xxx. área sobre a curva $\frac{\sin u}{u}$. Quebrando as áreas entre duas raízes se nota que as áreas pares são positivas e as ímpares negativas e que as mesmas vão diminuindo com a distância $|A_{n+1}| < |A_n|$.

No apêndice 5 mostraremos, usando cálculo de resíduos, que $\int_{-\infty}^{+\infty} \frac{\sin u}{u} du = \pi$. Portanto, para garantir que a área seja unitária com esse resultado precisamos fazer $A = \frac{1}{\pi}$.

Dessa forma a função $\delta_n\left(x - x_o\right) = \frac{\sin\left[n\left(x - x_o\right)\right]}{\pi\left(x - x_o\right)}$ se torna a função delta de Dirac no limite $n \to \infty$:

$$\delta\left(x - x_o\right) = \lim_{n\to\infty} \delta_n\left(x - x_o\right) = \lim_{n\to\infty} \frac{\sin\left[n\left(x - x_o\right)\right]}{\pi\left(x - x_o\right)}$$

Por outro lado podemos usar a fórmula de Euler $e^{ix} = \cos x + i \sin x$ para calcular $\frac{1}{2\pi}\int_{-n}^{+n} e^{i(x-x_o)t} dt = \frac{1}{2\pi} \frac{e^{i(x-x_o)t}}{i\left(x - x_o\right)}\Bigg|_{-n}^{+n} = \frac{1}{\pi\left(x - x_o\right)} \frac{e^{in(x-x_o)} - e^{-in(x-x_o)}}{2i} = \frac{\sin\left[n\left(x - x_o\right)\right]}{\pi\left(x - x_o\right)}$. Daqui extraímos a identidade super importante:

$$\frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{i(x-x_o)t} dt = \delta\left(x - x_o\right)$$

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.

4. Cálculo de variáveis complexas:

Definimos uma função de variável complexa $f(z)=u(x,y)+iw(x,y)$ em que $z=x+iy$. 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:

$$\frac{df}{dz}=\lim_{(dx,dy)\to(0,0)}\frac{du+idw}{dx+idy}$$

Vamos fazer esse limite por dois caminhos:

1. $dy=0$, ou seja, $y=cte$, e $dx\to 0$. Nesse caso $\frac{df}{dz}=\lim_{dx\to 0}\frac{du+idw}{dx}=\frac{\partial u}{\partial x}+i\frac{\partial w}{\partial x}$.
2. $dx=0$, ou seja, $x=cte$, e $dy\to 0$. Nesse caso $\frac{df}{dz}=\lim_{dx\to 0}\frac{du+idw}{idy}=-i\frac{\partial u}{\partial y}+\frac{\partial w}{\partial y}$

.

A função será diferenciável se $\frac{\partial u}{\partial x}+i\frac{\partial w}{\partial x}=-i\frac{\partial u}{\partial y}+\frac{\partial w}{\partial y}$, ou seja: $\frac{\partial u}{\partial x}=\frac{\partial w}{\partial y}$ e $\frac{\partial w}{\partial x}=-\frac{\partial u}{\partial y}$, 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:

$$f(z)=u(x,y)+iw(x,y) \text{ é analítica se } \frac{\partial u}{\partial x}=\frac{\partial w}{\partial y} \text{ e } \frac{\partial w}{\partial x}=-\frac{\partial u}{\partial y}.$$

Teorema: se $f(z)$ é analítica então $\oint_c f(z)dz=0$, onde c é qualquer caminho fechado no plano complexo xy.

Basta fazer a integral no caminho infinitésimal:

$$(x_o,y_o)\to(x_o+dx,y_o)\to(x_o+dx,y_o+dy)\to(x_o,y_o+dy)\to(x_o,y_o)$$

mostrado na figura xxx.

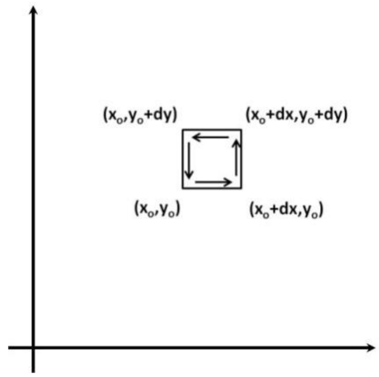


Figura xxx. Circuito infinitésimal para cálculo de $\oint_c f(z)\,dz = 0$

$(x_o, y_o) \rightarrow (x_o + dx, y_o)$:

$$\int_{x_o}^{x_o+dx} f(x, y_o)\,dx = \frac{F(x_o + dx, y_o) - F(x_o, y_o)}{dx} dx = f(x_o, y_o)\,dx$$

$(x_o + dx, y_o) \rightarrow (x_o + dx, y_o + dy)$:

$$\int_{y_o}^{y_o+dy} f(x_o + dx, y)\,i\,dy = i\frac{F(x_o + dx, y_o + dy) - F(x_o, y_o)}{dy} dy = if(x_o + dx, y_o)\,dy$$

$(x_o + dx, y_o + dy) \rightarrow (x_o, y_o + dy)$:

$$\int_{x_o+dx}^{x_o} f(x, y_o + dy)\,dx = \frac{F(x_o, y_o + dy) - F(x_o + dx, y_o + dy)}{dx} dx = -f(x_o, y_o + dy)\,dx$$

$(x_o, y_o + dy) \rightarrow (x_o, y_o)$:

$$\int_{y_o+dy}^{y_o} f(x_o, y)\,i\,dy = i\frac{F(x_o, y_o) - F(x_o, y_o + dy)}{dy} dy = -if(x_o, y_o)\,dy$$

Portanto:

$$\oint_c f(z)\,dz = f(x_o, y_o)\,dx + if(x_o + dx, y_o)\,dy - f(x_o, y_o + dy)\,dx - if(x_o, y_o)\,dy$$

$$\oint_c f(z)dz = i\left[\frac{f(x_o+dx,y_o)-f(x_o,y_o)}{dx}\right]dxdy - \left[\frac{f(x_o,y_o+dy)-f(x_o,y_o)}{dy}\right]dxdy$$

$$\oint_c f(z)dz = i\frac{\partial f}{\partial x}dxdy - \frac{\partial f}{\partial y}dxdy = \left[i\left(\frac{\partial u}{\partial x}+i\frac{\partial w}{\partial x}\right)-\left(\frac{\partial u}{\partial y}+i\frac{\partial w}{\partial y}\right)\right]dxdy$$

$$\oint_c f(z)dz = \left[-\left(\frac{\partial u}{\partial y}+\frac{\partial w}{\partial x}\right)+i\left(\frac{\partial u}{\partial x}-\frac{\partial w}{\partial y}\right)\right]dxdy$$

Agora $\frac{\partial u}{\partial x}-\frac{\partial w}{\partial y}=0$ e $\frac{\partial u}{\partial y}+\frac{\partial w}{\partial x}=0$ pela condições de Cauchy-Riemann. Então:

$$\oint_c f(z)dz = 0$$

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

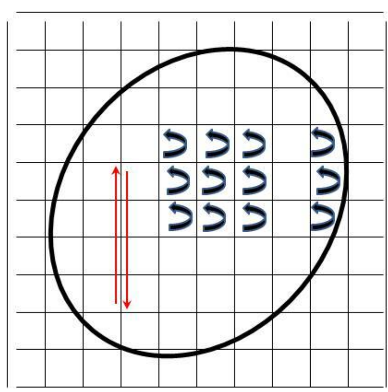


Figura xxx. Note que no interior da região as integrais de caminho se anulam porque enquanto o percurso de um célula está em uma direção o da vizinha está na direção oposta. Esse cancelamento, entretanto, não ocorre na fronteira pois não existe a célula vizinha.

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.

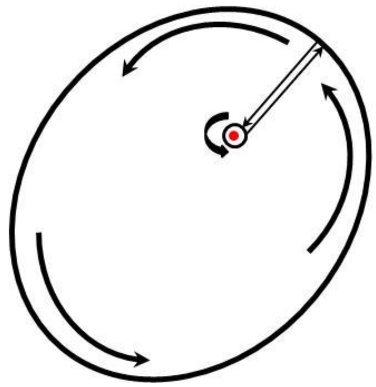


Figura xxx. Isolando uma singularidade do caminho de integração.

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

$$f(z)=\frac{a_{-n}}{(z-z_o)^n}+\frac{a_{-n+1}}{(z-z_o)^{n-1}}+\cdots+\frac{a_{-1}}{(z-z_o)}+a_o+a_1(z-z_o)+a_2(z-z_o)^2+\cdots$$

Não é analítica em $z=z_o$ por conta dos termos com potências negativas de $(z-z_o)$. Nesse caso dizemos que a função tem um polo de ordem n. O coeficiente a_{-1} é chamado de RESÍDUO. Porque ele é tão importante?

Vamos fazer a integral $\oint\frac{1}{(z-z_o)}dz$ em torno de z_o. A convenção é que giramos no sentido contrário ao dos ponteiros do relógio, mesmo sentido de crescimento do ângulo θ das coordenadas polares. Agora fazemos: $z=z_o+\varepsilon e^{i\theta}$, com $\varepsilon=cte$. Logo $z-z_o=\varepsilon e^{i\theta}$ e $dz=i\varepsilon e^{i\theta}d\theta$. Nesse caso:

$$\oint\frac{1}{(z-z_o)}dz=\int_0^{2\pi}\frac{1}{\varepsilon e^{i\theta}}i\varepsilon e^{i\theta}d\theta=i\int_0^{2\pi}d\theta=2\pi i$$

Independente do valor de ε, logo o resultado do limite $\varepsilon\to 0$ será o mesmo.

Entretanto note que $\oint\frac{1}{(z-z_o)^k}dz=0 \quad k\in\mathbb{N}\ e\ k>0$. Aqui nosso receio é o de que o ε do denominador que não cancela com o ε do dz faria a integral explodir. Entretanto, antes de tomar o limite de $\varepsilon\to 0$ vamos fazer a integral. Usamos a mesma mudança de variável $z=z_o+\varepsilon e^{i\theta}$, com $\varepsilon=cte$. Logo $z-z_o=\varepsilon e^{i\theta}$ e $dz=i\varepsilon e^{i\theta}d\theta$. Dessa forma:

$$\oint \frac{1}{\left(z-z_o\right)^n} dz = \int_0^{2\pi} \frac{1}{\varepsilon^n e^{in\theta}} i\varepsilon e^{i\theta} d\theta = \frac{i}{\varepsilon^{n-1}} \int_0^{2\pi} e^{-i(n-1)\theta} d\theta$$

Que nos leva a $\oint \frac{1}{\left(z-z_o\right)^n} dz = \frac{i}{\varepsilon^{n-1}} \left[\int_0^{2\pi} \cos\left[(n-1)\theta\right] d\theta - i \int_0^{2\pi} \sin\left[(n-1)\theta\right] d\theta \right] = 0$

pois $\int_0^{2\pi} \cos\left[(n-1)\theta\right] d\theta = \int_0^{2\pi} \sin\left[(n-1)\theta\right] d\theta = 0 \quad n > 1$.

Isso então nos leva ao seguinte resultado:

$$\oint f(z)dz = a_{-n} \oint \frac{dz}{\left(z-z_o\right)^n} + a_{-n+1} \oint \frac{dz}{\left(z-z_o\right)^{n-1}} + \cdots + a_{-1} \oint \frac{dz}{\left(z-z_o\right)} +$$
$$+ a_o \oint dz + a_1 \oint \left(z-z_o\right) dz + a_2 \oint \left(z-z_o\right)^2 dz + \cdots$$

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 $\frac{1}{\left(z-z_o\right)}$ do resíduo. Então chegamos ao resultado:

$$\oint f(z)dz = 2\pi i\, a_{-1} = 2\pi i \operatorname{Re} s$$

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

$$\oint f(z)dz = 2\pi i \sum_j \operatorname{Res}$$

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 n:

$$f(z) = \frac{a_{-n}}{\left(z-z_o\right)^n} + \frac{a_{-n+1}}{\left(z-z_o\right)^{n-1}} + \cdots + \frac{a_{-1}}{\left(z-z_o\right)} + a_o + a_1\left(z-z_o\right) + a_2\left(z-z_o\right)^2 + \cdots$$

Se multiplicamos essa função por $\left(z-z_o\right)^n$ ela se torna analítica, diferenciável, portanto.

$$\left(z-z_o\right)^n f\left(z\right)=a_{-n}+a_{-n+1}\left(z-z_o\right)+\cdots+a_{-1}\left(z-z_o\right)^{n-1}+$$
$$+a_o\left(z-z_o\right)^n+a_1\left(z-z_o\right)^{n+1}+a_2\left(z-z_o\right)^{n+2}+\cdots$$

Agora derivamos essa função $n-1$ vezes. Todos os termos com potência $k<n-1$ serão nulos, e todos os termos com potência $k>n-1$ terão o termo $\left(z-z_o\right)^{k-n+1}$ que vai a zero quando $z=z_o$. O único termo que sobra é para $k=n-1$. Então:

$$\frac{d^{n-1}}{dz^{n-1}}\left[\left(z-z_o\right)^n f\left(z\right)\right]_{z=z_o}=\left(n-1\right)!a_{-1}$$

E o resíduo será: $\text{Res}=a_{-1}=\dfrac{1}{\left(n-1\right)!}\dfrac{d^{n-1}}{dz^{n-1}}\left[\left(z-z_o\right)^n f\left(z\right)\right]_{z=z_o}$.

Se o polo é de ordem 1, também chamado de polo simples, então $n-1=0$ e:

$$\text{Res}=a_{-1}=\lim_{z\to z_o}\left[\left(z-z_o\right)f\left(z\right)\right]$$

5. Mostrar que $\int_{-\infty}^{+\infty} \frac{\sin u}{u} du = \pi$ **:**

Com o cálculo de resíduos essa tarefa simplifica. Antes de mais nada fazemos: $\int_{-\infty}^{+\infty} \frac{\sin u}{u} du = \mathrm{Im}\left[\int_{-\infty}^{+\infty} \frac{e^{iu}}{u} du\right]$. Depois fazemos o cálculo da seguinte integral:

$$\oint \frac{e^{iz}}{z} dz = \int_{-\infty}^{-\varepsilon} \frac{e^{iz}}{z} dz + \oint_{semicirculo\,\inf\; R=\varepsilon} \frac{e^{iz}}{z} dz + \int_{\varepsilon}^{\infty} \frac{e^{iz}}{z} dz + \oint_{semicirculo\,\sup R} \frac{e^{iz}}{z} dz = 2\pi i \,\mathrm{Res}\left(z=0\right)$$

Agora $\oint_{semicirculo\,\sup R} \frac{e^{iz}}{z} dz = 0$ porque para $z = \mathrm{Re}^{i\theta} = R\cos\theta + iR\sin\theta$ então $e^{iz} = e^{iR\cos\theta} e^{-R\sin\theta}$ e o como $\sin\theta > 0$ para $0 < \theta < \pi$ o termo $e^{-R\sin\theta}$ anula tudo no limite $R \to \infty$.

Por outro lado:

$$\oint_{semicirculo\,\inf\; R=\varepsilon} \frac{e^{iz}}{z} dz = \lim_{\varepsilon \to 0} \int_{\pi}^{2\pi} \frac{e^{i\varepsilon\left(\cos\theta + i\sin\theta\right)}}{\varepsilon e^{i\theta}} i\varepsilon e^{i\theta} d\theta = i \lim_{\varepsilon \to 0} \int_{\pi}^{2\pi} e^{i\varepsilon\left(\cos\theta + i\sin\theta\right)} d\theta = i\int_{\pi}^{2\pi} d\theta = \pi i$$

O resíduo do polo simples em $z = 0$ vale $\mathrm{Res} = \lim_{z \to 0}\left[z \frac{e^{iz}}{z}\right] = 1$

Juntando tudo temos:

$$\oint \frac{e^{iz}}{z} dz = \int_{-\infty}^{-\varepsilon} \frac{e^{iz}}{z} dz + \pi i + \int_{\varepsilon}^{\infty} \frac{e^{iz}}{z} dz = 2\pi i$$ ou seja: $$\int_{-\infty}^{+\infty} \frac{e^{iz}}{z} dz = \lim_{\varepsilon \to 0}\left[\int_{-\infty}^{-\varepsilon} \frac{e^{iz}}{z} dz + \int_{\varepsilon}^{\infty} \frac{e^{iz}}{z} dz\right] = \pi i$$

Portanto: $\int_{-\infty}^{+\infty} \frac{\sin u}{u} du = \mathrm{Im}\left[\int_{-\infty}^{+\infty} \frac{e^{iu}}{u} du\right] = \mathrm{Im}\left[\pi i\right] = \pi$

Resultado final: $\int_{-\infty}^{+\infty} \frac{\sin u}{u} du = \pi$