高等数学部分

极限定义:

f(x)在 x_0 的某一**去心**邻域内有定义，存在常数 A 使得，给定任意小正数 ε ,总存在 δ 使得 x 满足，

$0<|x-x_0|<\delta$，对应的 f(x)满足不等式 $|f(x)-A|<\varepsilon$.

去心邻域说明了极限定义没有必要在 x0 处有定义．并且由于 x0 邻域必然包括左极限和右极限，因此可以简单的通过左右极限是否相等来判断极限是否存在.(必要条件)

极限存在的判定准则?

准则 1：夹逼准则：若 $n>n_0$ 时总有 $y_n \le x_n \le z_n$，且 $\lim\limits_{n\to\infty} y_n = a$, $\lim\limits_{n\to\infty} z_n = a$ ，则 $\lim\limits_{n\to\infty} x_n = a$

准则 2：单调有界序列极限存在.

准则 3：收敛序列并不一定单调，但是存在极限，因此利用柯西极限定理可证明其存在极限.

(充分必要条件)存在正数 N,M，使得当 n>N,m>M 时满足：$|x_n - x_m|<\varepsilon$

连续定义:

连续的定义是通过极限来得到的．极限中包含了去心邻域，而连续的定义为非去心邻域.

函数 $y=f(x)$ 在 x_0 的某一邻域内有定义．满足 $\lim\limits_{x\to x_0} f(x)=f(x_0)$，则说明函数连续.

判断函数连续的简单方法：只需在极限存在的基础上，在判断 x_0 处的值即可：左右极限存在且相等，且左极限值等于 $f(x_0)$.

导数

一元函数导数定义：$f'(x_0)=\lim\limits_{\Delta x\to 0}\dfrac{\Delta y}{\Delta x}=\lim\limits_{\Delta x\to 0}\dfrac{f(x_0+\Delta x)-f(x_0)}{\Delta x}$

由于 $f(x_0)$ 包含在定义中，因此可导必然连续．判断导数存在的简单方法仍然可以通过左右极限是否相等来判断，上式中左右极限给了新的名词 左导数右导数．因此判断左右导数是否相等可以简单判断导数是否存在.

微分

微分定义：函数 $y=f(x)$ 在 x_0 的某一邻域内有定义，在此区间内，**函数增量**

$$\Delta y = f(x_0+\Delta x) - f(x_0) = A\times\Delta x + o(\Delta x)$$

其中 A 不依赖 Δx，则称函数可微. 其中增量关键部分(舍去了高阶无穷小) $A\Delta x = dy$，给定了一个新名词微分 dy. 对于一元函数，可微必然可导，两者互为**充分必要**条件. 因为:

$$\Delta y = A\Delta x + o(\Delta x) \Rightarrow \frac{\Delta y}{\Delta x} = A + \frac{o(\Delta x)}{\Delta x} \Rightarrow A = \lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x} = f'(x_0)$$

因此微分又可以写成 $dy = f'(x_0)\Delta x$. 由于 $dx = 1\cdot\Delta x$，因此都写成微分的表达式:

$$dy = f'(x_0)dx \Rightarrow \frac{dy}{dx} = f'(x_0)$$

微分中值定理

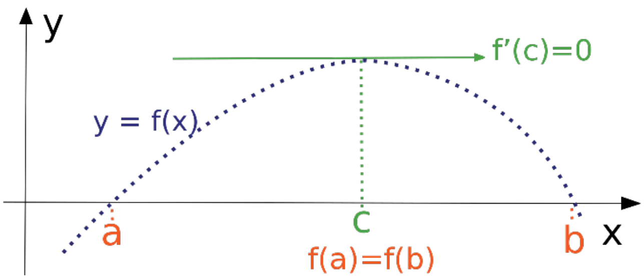
费马引理: 函数 $y = f(x)$ 在 x_0 的某一邻域 $U(x_0)$ 内有定义，且在 x_0 处可导，对任意 $x\in U(x_0)$ 满足 $f(x_0)\le f(x)$，则 $f'(x_0)=0$. 直观物理意义就是**局部极小值点**的导数必然为 0. 当然，改成局部极大值点也可以，只需对任意 $x\in U(x_0)$ 满足 $f(x_0)\ge f(x)$ 即可;

相关证明可通过导数的定义出来，求出左极限>=0，右极限<=0，因此导数=0;

四个定理: 罗尔定理，拉格朗日中值定理，柯西中值定理，泰勒中值定理

罗尔定理: (1)闭区间[a,b]连续. (2)开区间(a,b)可导 (3)端点值相等 $f(a) = f(b)$.

则开区间内存在 $f'(x)=0$. 如图，很形象的一个图. 条件(2)很重要，但是还需要条件(1)，因为开区间可导，并不能保证边缘点连续. 因此需要条件(1)来补充. 罗尔定义的证明可从费马引理出发，证明此区间内存在局部最大值点或者局部最小值即可.



拉格朗日中值定理: 罗尔定义有点严格，释放第三个条件可得到: 拉格朗日中值定理.

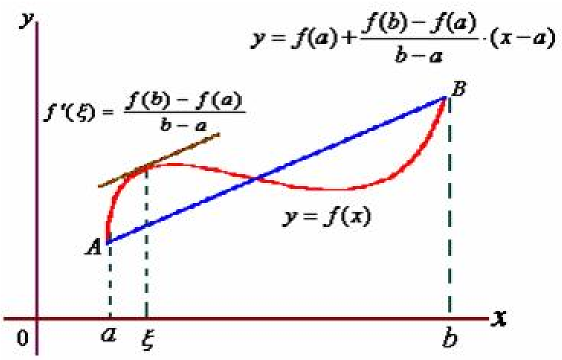
(1)闭区间[a,b]连续. (2)闭区间(a,b)可导

则区间内存在一点满足 $f(b)-f(a)=f'(x)(b-a)$. 物理含义如图.

证明方法也很简单, 把 f(x)掰平, 然后利用罗尔定理证明即可. 构造辅助函数:

$$g(x)=f(x)-\frac{f(b)-f(a)}{b-a}(x-a)+f(a)$$

利用罗尔定理: $g'(x)=0 \Rightarrow \frac{f(b)-f(a)}{(b-a)}=f'(x)$ 得证.

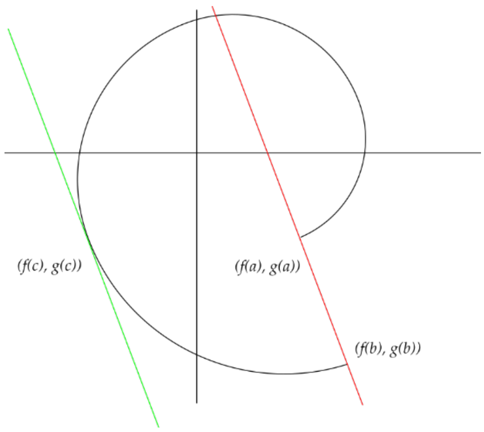


柯西中值定理: 扩展到函数之间的导数关系. f(x), g(x)

(1)闭区间[a,b]连续 (2)开区间(a,b)可导 (3)对任意 $x\in(a,b)$, $f'(x)\neq 0$.

则存在 $\varepsilon\in(a,b)$ 使得 $\frac{g(b)-g(a)}{f(b)-f(a)}=\frac{g'(\varepsilon)}{f'(\varepsilon)}$

物理含义不太直观, 但是也可以理解, 只需将 $f(x)$ 当成原本的 x 轴即可. 坐标为 $(f(x),g(x))$



泰勒中值定理

洛必达法则

洛必达法则由柯西中值定理得到. 它可以处理 $0/0$, ∞/∞ 的问题.

(1)当 $x \to a$ 时，$f(x)$，$g(x)$ 都趋向于 0;

(2) 在点 a 的某去心邻域内，$f'(x)$，$g'(x)$ 存在且 $f'(x) \neq 0$

(3) $\lim\limits_{x \to a} \dfrac{g'(x)}{f'(x)}$ 存在.

满足上述三个条件，则可以证明. $\lim\limits_{x \to a} \dfrac{g(x)}{f(x)} = \dfrac{g'(x)}{f'(x)}$

证明：由柯西中值定理可得：$\dfrac{g(x)}{f(x)} = \dfrac{g(x)-g(a)}{f(x)-f(a)} = \dfrac{g'(\varepsilon)}{f'(\varepsilon)}$，其中 $f(a) = g(a) = 0$

两边取极限 $x \to a$，则 $\varepsilon \to a$，得证.

泰勒公式:

函数在某点的多项式逼近. 因为多项式有很强的拟合能力，因此利用导数信息得到多项式的系数

利用函数值和导数信息求得：拟合函数

$$p(x) = f(x_0) + f'(x_0)(x-x_0) + \frac{f''(x_0)}{2!}(x-x_0)^2 + \ldots + \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n$$

原函数在某一邻域内为拟合函数+高阶无穷小：$f(x) = p_n(x) + o\left((x-x_0)^n\right)$，

但拟合是在邻域内才有较高的准确度，不保证远处也有相同精度.

泰勒中值定理:

泰勒拟合公式的最后一项高阶无穷小可以通过柯西中值定理得到**准确表达式**.

高阶无穷小：$R_n(x) = o\left((x-x_0)^n\right) = \dfrac{f^{(n+1)}(\varepsilon)}{(n+1)!}(x-x_0)^{n+1}$，其中 ε 介于 x 与 x_0 之间. 这也是为什么叫做**中值定理**的原因，通过**中间点确定边界点**的值.

当 n=0 时，泰勒中值退化成拉格朗日中值:

$$f(x) = p_n(x) + R_n(x) \Rightarrow f(x_0) + f'(x_0)(x-x_0)$$

多元函数

微分关注的是增量，求导关注的是导数值. 对于一元函数有:

$$\text{可微 } dy \Leftrightarrow \text{可导 } f'(x) \Rightarrow \text{连续} \Rightarrow \text{极限存在}$$

对于多元函数上述条件要发生改变：没画双向的都是单向条件，仅有充分性.

$$偏导连续 \Rightarrow 可微\ dz \Rightarrow 连续 \Rightarrow 极限存在的$$

值得注意的是：偏导存在 $\not\Leftrightarrow$ 连续. 因为偏导是一维连续，多元函数的连续是一个区域连续.

下面详细说明多元函数的偏导和微分的关系.

全微分与偏导

全增量：$\Delta z = f(x+\Delta x, y+\Delta y) - f(x,y)$，其中 $x+\Delta x$ 保证在 x 的邻域内.

如果全增量能化简到此形式，其中 A, B 为常数. 则可称函数 f 在**点(x,y)处可微**，

$$\Delta z = A\Delta x + B\Delta y + o(\sqrt{(\Delta x)^2 + (\Delta y)^2}$$

全微分：将可微的全增量的关键部分提取出来，称其为全微分：$dz = A\Delta x + B\Delta y$

偏增量：$\Delta_x z = f(x+\Delta x, y) - f(x,y)$，利用一元函数的求导定义得到偏导数

偏导数：$\dfrac{\partial z}{\partial x} = \lim\limits_{\Delta x \to 0} \dfrac{\Delta_x z}{\Delta x} = \lim\limits_{\Delta x \to 0} \dfrac{1}{\Delta x} f(x+\Delta x, y) - f(x,y)$

可以证明：全微分的 A,B 就是偏导数 $A = \dfrac{\partial z}{\partial x}, B = \dfrac{\partial z}{\partial y}$. 利用全微分和一元求导定义即可证明.

可微和偏导的关系：

从可微的定义出发，$\Rightarrow$ 偏导数一定存在，$\Rightarrow$ 函数在区域内连续. $\Rightarrow$ 极限存在.

是否能从偏导数存在来得到可微的条件？

$$\Delta z = f(x+\Delta x, y+\Delta y) - f(x,y)$$

$$\Rightarrow \Delta z = f(x+\Delta x, y+\Delta y) - f(x, y+\Delta y) + f(x, y+\Delta y) - f(x,y)$$

$$\Rightarrow \Delta z = \left.\frac{\partial f}{\partial x}\right|_{(x,y+\Delta y)} + \left.\frac{\partial f}{\partial y}\right|_{(x,y)}$$

也就是说已知函数 f 在点(x,y)的偏导存在是不够的，因为需要的是在 $(x, y+\Delta y)$ 点的偏导存在，同理按照上述方法推导，需要在 $(x+\Delta x, y)$ 点偏导也存在. 假如已知偏导在点 (x,y) 存在，只需增加一个条件：偏导连续. 即：偏导存在且连续可得函数在点 $(x+\Delta x, y)$ 可微.

方向导数与梯度:

导数的概念是函数 f 的微分与 x 的微分相除而得. 对于多元函数情况类似, 但是对分母的选择却多了一些变化空间, $\Delta l=\sqrt{(\Delta x)^2+(\Delta y)^2}$, 在保证 Δl 为单位长度的前提下, 可以选择不同比例系数的 Δx 和 Δy. 由此求出来的导数也不一样, 因此为多元函数的求导定义了方向导数.

假设以 $\vec{l}$ 方向作为增量, 函数可微. 则全微分表达式为:

$$\Rightarrow \Delta z=f(x+\Delta l\cos\alpha, y+\Delta l\cos\beta)-f(x,y)$$

$$\Rightarrow \Delta z=\frac{\partial f}{\partial x}\cdot\Delta l\cos\alpha+\frac{\partial f}{\partial y}\cdot\Delta l\cos\beta+\mathrm{o}(\Delta\vec{l})$$

$$\Rightarrow dz=\frac{\partial f}{\partial x}\cdot\Delta l\cos\alpha+\frac{\partial f}{\partial y}\cdot\Delta l\cos\beta$$

$$\Rightarrow \frac{dz}{d\vec{l}}=\frac{\partial f}{\partial x}\cos\alpha+\frac{\partial f}{\partial y}\cos\beta$$

由于 Δl 为单位值, α, β 由 $\vec{l}$ 控制, 因此不同方向的选择会得到不同的导数值.

有没有一个最大的方向向量? --- 梯度的方向可以得到最大的方向导数.

因为方向导数等于两个向量的点乘, 同向时点乘最大,如下:

$$\frac{dz}{d\vec{l}}=\left(\frac{\partial f}{\partial x},\frac{\partial f}{\partial y}\right)\cdot(\cos\alpha,\cos\beta)$$

因此设梯度为: $\vec{G}=\frac{\partial f}{\partial x}\vec{e}_x+\frac{\partial f}{\partial y}\vec{e}_y$. 注意梯度是个向量!! 但长度不一定为 1.

多元积分:

了解即可, 详情参考电磁学.

高斯公式: 通量的面积分 = 散度的体积分

斯托克斯公式: 环量的线积分 = 旋度的面积分.

概率论部分

贝叶斯公式:

$$P(w_i \mid D)=\frac{p(D \mid w_j)p(w_j)}{\sum_j P(D \mid w_j)p(w_j)}$$

其中 $P(w_i \mid D)$ 为后验概率, $p(D \mid w_j)$ 为似然概率, $p(w_j)$ 为先验概率. 详情参见机器学习贝叶斯部分.

常见离散分布

0-1 分布: (伯努利分布, 两点分布): $p(x=1)=p$, $p(x=0)=1-p$

二项分布: 名字来自牛顿二项式, $p(X=k)=C_n^k p^k(1-p)^{n-k}$

二项分布相当于多次掷硬币, 每次结果满足 0-1 分布, 各次实验之间独立, 结果累加.

泊松分布: $p(X=k)=\frac{\lambda^k e^{-\lambda}}{k!}$, 其中 $\lambda>0$

泊松分布的期望和方差都为 $E[X]=\lambda, Var[X]=\lambda$. 推导也很简单, 就按期望定义式即可.

实际上当二项分布的 p 很小时, 可以用泊松分布来逼近而二项分布:

证明: 将 $p=\lambda/n$ 代入二项分布, 得到:

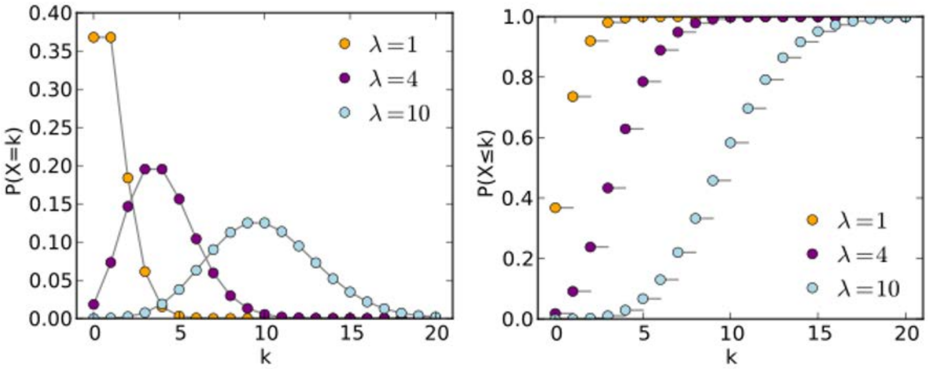
$$p(X=k)=\frac{n!}{k!(n-k)!}(\frac{\lambda}{n})^k(1-\frac{\lambda}{n})^{n-k}$$

由于 $\lim\limits_{n\to\infty}(1-\frac{\lambda}{n})^n=e^{-\lambda}$, 代入上式得到:

$$\lim_{n\to\infty}\frac{n!}{k!(n-k)!}(\frac{\lambda}{n})^k(1-\frac{\lambda}{n})^{n-k}\Rightarrow\lim_{n\to\infty}\frac{n!}{n^k(n-k)!}(1-\frac{\lambda}{n})^{-k}\frac{\lambda^k}{k!}e^{-\lambda}=\frac{\lambda^k}{k!}e^{-\lambda}$$

其中由于 $n\to\infty$, 因此 $(1-\frac{\lambda}{n})^{-k}=1$, $\frac{n!}{n^k(n-k)!}=\frac{n(n-1)...(n-k+1)}{n^k}=1$.

因此 $p(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}$ 得证.



如何估计出参数 λ，第一种就是利用二项分布 $\lambda = np$，第二种可用 MLE 极大似然估计.

$$L(\lambda) = \ln\prod_i f(k_i \mid \lambda) = \sum_i \ln\frac{\lambda^{k_i}}{k_i!}e^{-\lambda} = n\lambda + \sum_i k_i \ln\lambda + \sum_i \ln(k_i!)$$

其中 n 为实验次数, ki 为每次实验得到的 k 值.

对 L 求导得到驻点，$\lambda = \sum_i k_i / n$ ，求二阶导发现驻点处二阶导<0，因此判断出是极大值点.

实际上假设每次实验满足二项分布，利用极大似然估计求出 $p = \left(\sum_i \frac{k_i}{m}\right) / n$ 其中 m 为每次实验的投掷次数, n 为实验次数. 而方法一得出的 $\lambda = mp$，因此方法一和方法二实际上是等价的.

几何分布: $p(X = k) = q^{k-1}p$，相当于 0-1 分布的实验中求**前 k-1 次**失败,第 **k** 次成功的概率.

这种分布的特点是无记忆性，每次抽取都是于 0-1 分布，不会因为抽取次数而改变.

超几何分布: 有放回抽样 $p(X = k) = \frac{C_M{}^k C_{N-M}^{n-k}}{C_N^n}$，其中 M 为次品总数, N 为全部产品总数.

n 为目标抽样数, k 为抽到的次品数量.

常见连续变量分布

连续分布用概率密度函数 $f(x)$ 和分布函数 $F(x = k) = p(x \le k)$ 来描述分布.

均匀分布: $f(x) = \frac{1}{b-a}$ ，其中 $a \le x \le b$，$F(x) = \frac{x-a}{b-a}$ ，其中 $a \le x \le b$，

指数分布: $f(x) = \lambda e^{-\lambda x}$ ，其中 $x > 0$. 期望为 $E[xf(x)] = \frac{1}{\lambda}$ ，$\mathrm{var}[X] = \frac{1}{\lambda^2}$

指数分布的特点是时间无关性，只和时间间隔有关，

泊松过程: $P(N(t+\tau)-N(t)=k)=\frac{(\lambda\tau)^k}{k!}e^{-\lambda\tau}$.泊松分布相当于单位时间的泊松过程, λ 被称为抵达率, 当时间间隔足够大时, k 值分布会整体往右移动. 当 k=0 时, 泊松分布变为: $p(k=0)=e^{-\lambda\tau}$, 非 0 概率为:

$$p(X(\tau)>0)=1-e^{-\lambda\tau}\Rightarrow f(X(\tau)>0)=\lambda e^{-\lambda\tau}$$

也就从泊松分布得到了指数分布. 指数分布就是有人来就好, 泊松分布还可以统计人数.

指数分布的时间无关性: 假设把 k=0 当成原件成功运行, k>0 表示器件损坏.

则旧原件和新原件的使用寿命仍然相等:

证明:

已知指数分布: 从 0 到 t 时刻不发生故障的概率 $p(t)=e^{-\lambda t}$

已知工作了 s 时间, 则未来 s 到 s+t 不发生故障的概率为

$$p(s+t\,|\,s)=\frac{p(s+t)}{p(s)}=\frac{e^{-\lambda(s+t)}}{e^{-\lambda s}}=e^{-\lambda t}$$

得证.

正态分布: 期望为 $E[X]=u$, 方差为 $Var[X]=\sigma^2$. 概率密度函数为:

$$f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{(x-u)^2}{2\sigma^2})$$

正态随机变量的线性函数仍为正态随机变量.

极大似然和最大后验概率等相关知识可参见传统机器学习篇章.

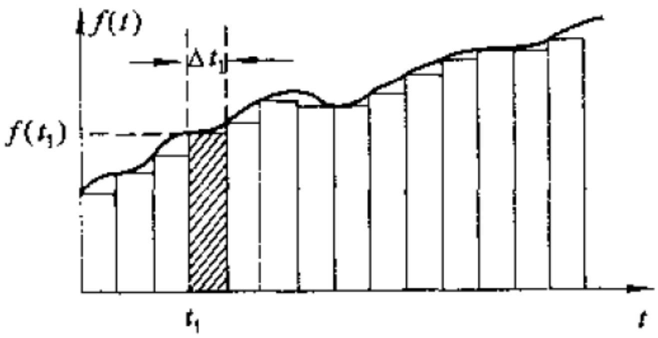
信号与系统

冲激定义: $\delta(t)=\frac{1}{\tau}\left(u(t+\tau/2)-u(t-\tau/2)\right)$. 物理意义是面积为 1 的信号.

宽度由阶跃函数 u 提供, 高度由$1/\tau$提供.

信号的冲激表达式: 利用微积分的思想, 将函数通过阶跃函数进行分段表达.

如: $f(t_1)[u(t-t_1)-u(t-t_1-\Delta t)]$, 表示在$\left(t_1,t_1+\Delta t\right)$这段时间内的函数值 $f(t_1)$ 的近似替代.



将分段信号累加可得: 可得信号的冲激表达式. 冲激表达式的好处是配合系统函数得到的.

$$f(t)=\sum_{t_1=-\infty}^{\infty} f(t_1)[u(t-t_1)-u(t-t_1-\Delta t)]$$

$$\Rightarrow f(t)=\sum_{t_1=-\infty}^{\infty} f(t_1)\frac{[u(t-t_1)-u(t-t_1-\Delta t)]}{\Delta t}\Delta t$$

$$\Rightarrow f(t)=\sum_{t_1=-\infty}^{\infty} f(t_1)\delta(t-t_1)\Delta t \Rightarrow \int_{-\infty}^{\infty} f(\tau)\delta(t-\tau)d\tau$$

冲激的系统响应: **冲激信号得到的零状态系响应** 等价于 **具有单位能量的零输入响应**.

这是由冲激函数的性质决定的, 因为信号在 t>0 时=0, 因此没有强迫响应, 瞬时的能量供应给了系统单位能量, 因此系统响应等效为零输入响应.

假设 h(t)为系统的冲激响应, 由于系统为 LTI(linear Time Invariance)系统, 响应可以累加, 因此

激励信号	响应信号	理论依据
$\delta(t)$	$h(t)$	定义
$\delta(t-\tau)$	$h(t-\tau)$	时不变特性
$\sum_{t_1=-\infty}^{\infty} f(t_1)\Delta t\delta(t-t_1)$	$\sum_{t_1=-\infty}^{\infty} f(t_1)\Delta t h(t-t_1)$	叠加性

并且由于信号的现实意义，积分变量$t_1 > 0$,且$t > t_1$，因此上述式子可以继续化简得到系统响应:

$$r(t) = \int_0^t f(\tau)h(t-\tau)d\tau$$

此式的数学形式又被称为卷积. 其物理意义很明确: 历史信号对 t 时刻的响应的累计.

比如说 t 时刻，最新的信号为 $f(t)$，系统对应的响应应为 $h(0)$，而前一秒信号 $f(t-1)$ 的响应

遗留到 t 时刻的影响为 $h(1)$，结合所有对 t 时刻还有影响的响应，得到的表达式就是卷积表达式.

傅里叶级数

上节讲了可以将信号表达为冲激函数的积分，这是时域的分解，接下来介绍一下频域的分解. 类似线性代数中，将向量转换到矩阵的特征空间，在矩阵的特征空间内的计算比较简单，然后再将结果转化到普通空间得到计算结果，这样的变换需要一个关键部分: 特征空间的正交基. 傅里叶变换也找到了正交基，$1, \cos(t), \cos(2), \ldots \cos(nt), \sin(t), \sin(2), \ldots \sin(nt)$

正交的定义:　当 $m \neq n$ 时

$$\int_{\pi}^{-\pi} \cos(nt)\sin(mt)dt = \frac{1}{2}\int_{\pi}^{-\pi} \cos[(n-m)t] - \cos[(n+m)t]dt = 0$$

$$\int_{\pi}^{-\pi} \cos(nt)\cos(mt)dt = \frac{1}{2}\int_{\pi}^{-\pi} \cos[(n+m)t] + \cos[(n-m)t]dt = 0$$

将函数 f(t)表达为正交基的线性组合:

$$f(t) = a_0 \cdot 1 + a_1\cos(t) + b_1\sin(t) + \ldots. + a_n\cos(nt) + b_n\sin(nt)$$

为保证上式成立，需要 $\cos(t)$ 与 $f(t)$ 的周期一致. 如 T 为 $f(t)$ 的周期，则第一项为: $\cos(\frac{2\pi}{T}t)$

为方便讨论 $f(t)$ 的周期 T 均设为 1.

如何求出系数? :

仅需左右两边乘以对应的正交基即可求出系数: $a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\cos(nt)dt$

当然对 a_0 情况特殊一些，两边同时乘 1 积分: $a_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)dt$

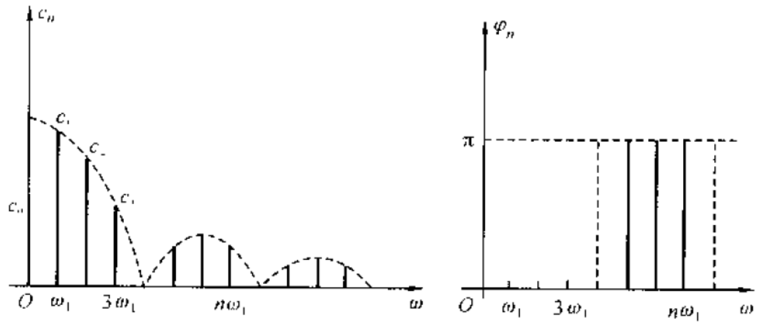
用 $\cos$, $\sin$ 作为正交基比较直观，实际上 n<0 也是满足正交基的定义，但由于历史原因(负频率没意义)，并不使用 n<0 的正交基.

使用条件? 狄利克雷条件

(1) 周期内间断点数目有限. (无穷间断点就不用玩了)

(2) 周期内极大值极小值数量有限. (参考连续不可导函数)

(3) 周期内信号可积. (积分值有限)



(a) 幅度谱　　(b) 相位谱

经过发展, 欧拉公式给了我们一个新途径: 用 e^{jnw} 作为正交基, 并且引入 n<0 的正交基.

$$e^{jw} = \cos w + j\sin w$$

由上面的正交基的积分结果, 很容易证明 e^{jnw} 也是正交基. 当 $m \neq -n$ 时, **注意是**-n.

$$\int_{-\pi}^{\pi} e^{jnt} e^{jmt} dt = \int_{-\pi}^{\pi} (\cos nt + j\sin nt)(\cos mt + j\sin mt) dt = 0$$

将函数 f(t)化为正交基的组合: 其中 $F(nw_1)$ 表示系数, w_1 表示基频.

$$f(t) = \sum_{n=-\infty}^{\infty} F(n) e^{jnt}$$

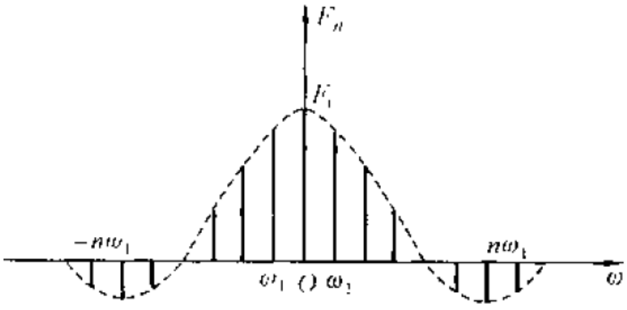
如何求出系数? : 仍然是乘正交基后积分.

$$\int_{-\pi}^{\pi} f(t) e^{-jmt} dt = \int_{-\pi}^{\pi} e^{jmt} \cdot \sum_{n=-\infty}^{\infty} F(n) e^{jnt} dt$$

$$\Rightarrow F(m) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) e^{-jmt} dt$$

由于指数正交基用到了负频率 因此图形为双边频谱, 且幅值大小为 cos 正交基的 1/2, 但是 0 点处是等高的. 如下图

帕塞瓦尔定理: 时域的平均功率等于 $P = \sum_{n=-\infty}^{\infty} |F_n|^2$. 说明: **时域转频域**空间也是符合**能量守恒**.



(c)

傅里叶变换

周期函数可以通过傅里叶级数展开，而非周期函数则不能. 如果将非周期函数想象成周期 T 无穷大的周期函数则可进行傅里叶级数展开，此时会出现如下情况：$F(w_1), F(2w_1), .., F(nw_1)$ 的间隔 $w_1 = \frac{2\pi}{T} \to 0$. 各个幅值在 w 为 x 坐标的图中会无限接近. 但幅值也会趋近 0.

$$F(mw_1) = \frac{1}{T}\int_{-T/2}^{T/2} f(t)e^{-jmw_1 t}dt \Rightarrow 0$$

由于帕塞瓦尔定理，频域得到的能量和时域是一样的，因此引入**频谱密度函数**来替代幅值

$$F(w) = \int_{-\infty}^{\infty} f(t)e^{-jwt}dt$$

有了概率密度函数后，看看之前的时域函数表达式.

$$f(t) = \sum_{n=-\infty}^{\infty} F(n)e^{-jnt} \Rightarrow f(t) = \frac{1}{T}\sum_{n=-\infty}^{\infty} F(w)e^{-jwt} = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(w)e^{-jwt}dw$$

上述两式也叫做**傅里叶变换**和**傅里叶逆变换**.

主要性质:

若 f(t)为实偶函数，则 F(w)为实偶函数. (可以将 exp(jw)拆开得到 cos 和 sin 积分得到)

若 f(t)为实奇函数，则 F(w)为虚奇函数. (可以将 exp(jw)拆开得到 cos 和 sin 积分得到)

时域卷积 = 频域相乘: $\left[f_1(t) * f_2(t)\right] \Rightarrow F_1(w)\cdot F_2(w)$

频域卷积 => 时域相乘: $\frac{1}{2\pi}F_1(t) * F_2(w) \Rightarrow f_1(t)\cdot f_2(t)$

由此，时域卷积可以求出系统响应，可以通过傅里叶变换得到频域的表达式，仅需相乘即可，最后再进行傅里叶逆变换即可得到时域响应.

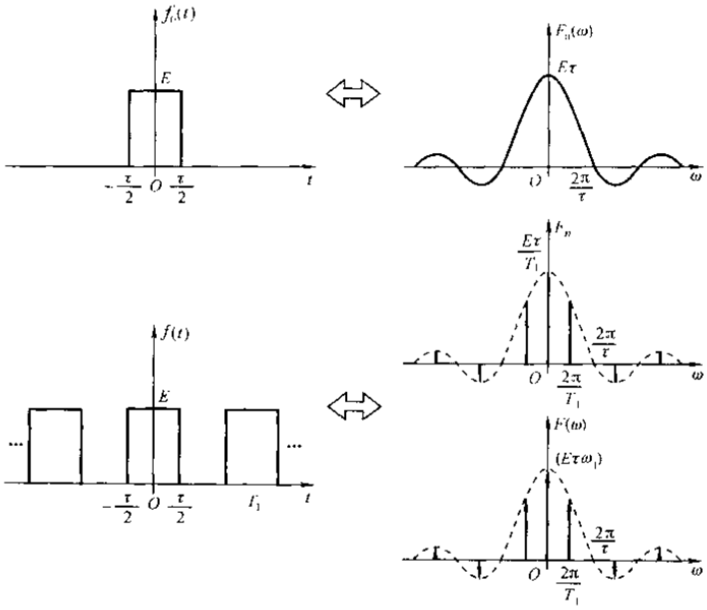
自相关函数的傅里叶变换为$|F(w)|^2$，由此可以得到功率谱或者能量谱.

周期函数的傅里叶变换:

周期函数可以进行傅里叶级数展开，非周期函数有傅里叶变换，那么周期函数能否进行傅里叶变换？可以--> 频谱为离散谱，但是幅度为**冲激函数**. 这样的原因在于 傅里叶变换得到的是频谱的**密度** $F(w)dw$ 会等于 $F(nw_1)$,假设 $F(nw_1)=C$，则频谱密度 $F(w)=2\pi C\delta(w)$.

由于 $\left[e^{jw_0t}\right]=2\pi\delta(w-w_0)$，其中[]表示傅里叶变换.

于是: $f(t)=F_n e^{jnw_0t} \Rightarrow \left[f(t)\right]=2\pi\sum_{-\infty}^{\infty}F_n\delta(w-w_0)$. 傅里叶变换为傅里叶级数的冲激序列.



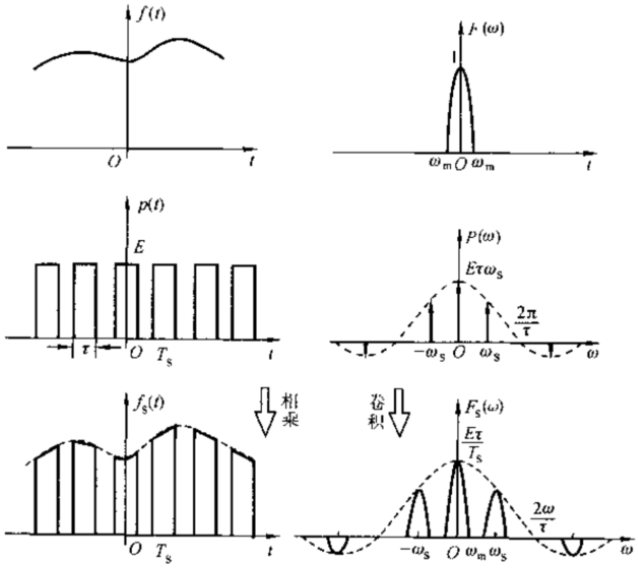
抽样

综合运用了卷积，傅里叶级数与傅里叶变换的关系 :

假设用矩形脉冲进行抽样: $f_{sp}(t)=f(t)p(t)$

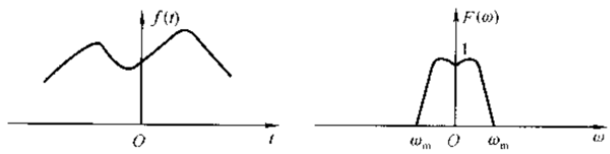
由于 $p(t)$ 为周期函数，因此其傅里叶变换为级数的冲激 $P(w)=2\pi\sum_{-\infty}^{\infty}P_n\delta(w-nw_s)$

时域相乘 = 频域卷积/2pi. $F_s(w)=\frac{1}{2\pi}F(w)*P(w)$. 如图:

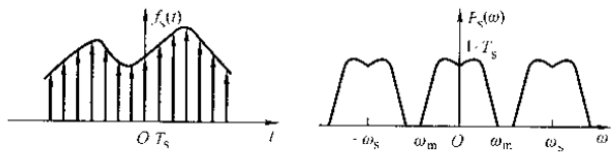


如果抽样性能足够好—冲激序列，则整体上可以完全复制频谱，并且以采样频率 w_s 搬移.

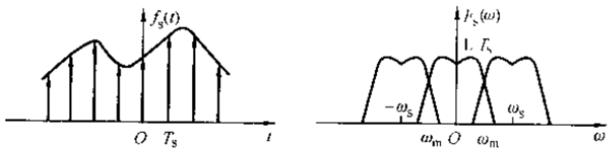
这也可以说明奈奎斯特采样频率要大于**信号最高频率的两倍**的来源.



(a) 连续信号的频谱



(b) 高抽样率时的抽样信号及频谱(不混叠)



(c) 低抽样率时的抽样信号及频谱(混叠)

离散序列傅里叶变换.

离散时间序列的傅里叶变换也是一样，仍用正交基来表示，但是通过频域的正交基的组合来表示

频域. 如: $X(e^{jw}) = FT[x(n)] = \sum_{n=-\infty}^{\infty} x(n)e^{-jnw}$.而正交基的系数就是时域表达式.

之前是时域周期连续，频域离散，现在是时域离散，频域周期连续.

逆变换: (求正交基的系数): $x(n) = IFT[X(jw)] = \int_{-\pi}^{\pi} X(jw)e^{jnw}dw$

因为正交基的特性，基频的周期为 $T_w = [-\pi, \pi]$，因此频谱是有周期的.

FT 的使用条件：绝对可和 $\sum_{n=-\infty}^{\infty} |x(n)| < \infty$，可令 X(jw)=0 得到该条件.

频域卷积定理：$y(n) = h(n)x(n) \Rightarrow FT[y(n)] = \frac{1}{2\pi} H(jw) * X(jw)$

帕萨瓦尔定理：$\sum_{n=-\infty}^{\infty} |x(n)|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |X(jw)| dw$

离散周期序列傅里叶变换 DFT.

上文的离散序列满足可和条件，因此频谱连续，但是如果是离散序列也是周期序列呢?

此时又可以用时域的正交基了 $e^{j\frac{2\pi}{N}kn}$. 正交性的简单证明:当 $k \neq m$ 时 $\sum_{n=0}^{N-1} e^{j\frac{2\pi}{N}(k-m)n} = 0$，

当 $k = m$ 时 $\sum_{n=0}^{N-1} e^{j\frac{2\pi}{N}(k-m)n} = N$，可用欧拉公式差分得证.

时域表达式为：$\tilde{x}(n) = \frac{1}{N} \sum_{k=0}^{N-1} \tilde{X}(k) e^{j\frac{2\pi}{N}kn}$

如何求系数？正交基的求和即可：$\tilde{X}(k) = \sum_{n=0}^{N-1} \tilde{x}(n) e^{-j\frac{2\pi}{N}kn}$

序列 $\tilde{x}(n)$ 与 $\tilde{X}(k)$ 都是周期为 N 的离散序列.

信息论

信息量与不确定度挂钩. 不确定度与概率挂钩. 于是信息量可以写成关于概率的函数.

自信息：$I(x_i) = \log \frac{1}{p(x_i)}$，概率越小，则自信息越大. 当然 100%的必然事件没有信息量.

平均信息量(熵)：$H(X) = E[I(X)] = -\sum_{i}^{n} p(x_i) \log p(x_i)$，表示平均每个事件种类的信息量.

条件自信息: $I(x_i\,|\,y_j) = -\log p(x_i\,|\,y_j)\ \Rightarrow I(x_i\,|\,y_j) + I(y_j) = I(x_i, y_j)$

联合熵: $H(X,Y) = \sum p(x_i, y_j) I(x_i, y_j)$. 注意熵的重点是平均,期望 Expectation.

条件熵: $H(X\,|\,Y) = \sum p(x_i, y_j)\log p(x_i\,|\,y_j)$.熵的重点是信息量的期望,因此使用 $p(x_i, y_j)$

平均互信息: $I(X;Y) = H(X) - H(X\,|\,Y)$,使用了总体符号 X 的变量一般都具有平均的含义

互信息表示已知 Y 的话，平均能够知道多少关于 X 的信息．$H(X)$ 表示 X 的信息量，后面

$H(X\,|\,Y)$ 表示知道 Y 以后, X 还剩多少信息量．因此两者之差就是 Y 能够挖掘的信息量.

互信息: 通过平均互信息的定义，得到互信息的定义 $I(x_i; y_j) = \log\dfrac{p(x_i\,|\,y_j)}{p(x_i)}$

通常情况下，知道 y 后 $p(x_i\,|\,y_j) > p(x_i)$,但这只是平均意义才行,因此互信息有正有负.

疑义度与噪声熵: 互信息衍生的产品．$H(X\,|\,Y)$ 表示接收到 y，对 x 的怀疑程度.

$H(Y\,|\,X)$ 表示发送 X，由于噪声产生的对 Y 的不确定度，叫做噪声熵.

条件互信息: $I(X;Y\,|\,Z) = I(X;Y,Z) - I(X;Y)$

香农第一定理: 可变长无失真信源**编码定理**．编码规则为尽量等概.

香农第二定理: 有噪信道编码定理，引入了信道的概念．著名公式:

$$C = B\log(1 + S/N)$$

香农第三定理: 保失真度准则下的**有失真信源编码**定理. ...

熵定义的唯一形式可以参考之前写的博客—>熵的唯一性证明．仅用三个条件便可证明.

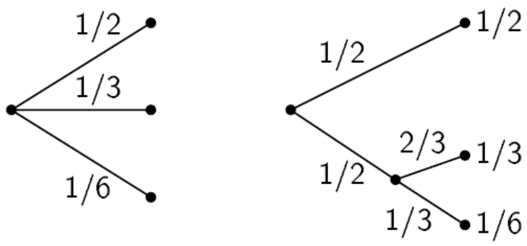
条件一: H 具有**连续性**.

条件二: $H(\frac{1}{n},...,\frac{1}{n})$ 随 n 单调递增.(因为可能情况增加,所以不确定度增加)

条件三: 熵的组合性质，表示为:

$$H_m(p_1, p_2, ..., p_m) = H_{m-1}(p_1 + p_2, p_3 ..., p_m) + (p_1 + p_2) H_2(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2})$$

条件三的物理含义可理解为将抽取过程分段进行,但总体的熵不变.利用图例来进一步说明:



将单次随机抽取过程转化为二次抽取过程,阶段一为 $H(\frac{1}{2},\frac{1}{2})$,同时有 $\frac{1}{2}$ 的概率会进入阶段二,而阶段二的不确定熵为 $H(\frac{2}{3},\frac{1}{3})$,但由于仅有 $\frac{1}{2}$ 的概率会进入阶段二,所以阶段二对总过程的不确定性的贡献为 $\frac{1}{2}H(\frac{2}{3},\frac{1}{3})$.最终得到如下形式 $H(\frac{1}{2},\frac{1}{3},\frac{1}{6})=H(\frac{1}{2},\frac{1}{2})+\frac{1}{2}H(\frac{2}{3},\frac{1}{3})$

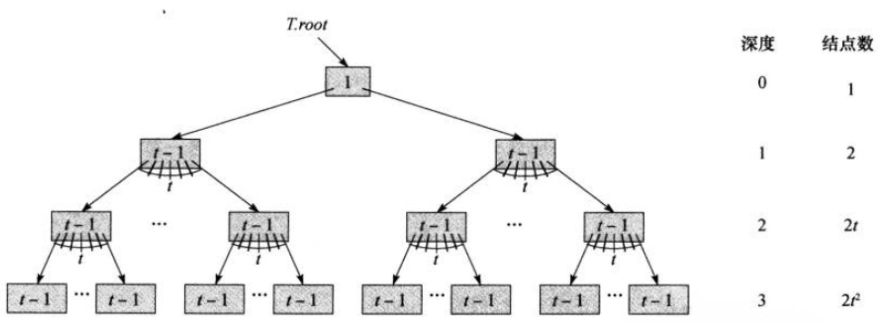
数据结构

二叉搜索树: 节点左大右小, 使用中序遍历能够从小到大输出节点. 最佳情况 $\log N$.

红黑树: 为了使二叉搜索树平衡而建立的, 加入了颜色信息. 配合相关定义能使得各个叶子节点的高度差不超过 2. 有以下几条规则: (1) root 节点为黑色. (2) 叶子节点为黑色 (3) 红色节点的子结点都为黑色. (4) **任一一条路径具有相同数目的黑色节点数量**.

B 树: 类似红黑树, 但可以有很多子结点, 是一种平衡搜索树, 广泛应用在磁盘等设备上.

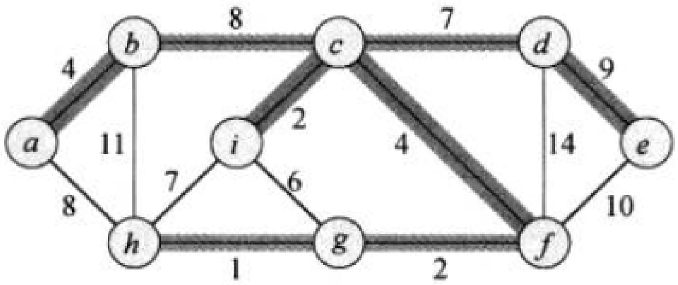
原因是二叉树的一次索引就需要访问磁盘一次, 如果同一个节点的子结点多了, 那树的高度就小多了, 因此访问 IO 的次数就减少了. B 树的每个节点关键字非降序排列, 因此对每个节点仅需比较两次便可知道索引点是否在该节点内. 并且非降序, 可用二分法快速索引内部的子结点.



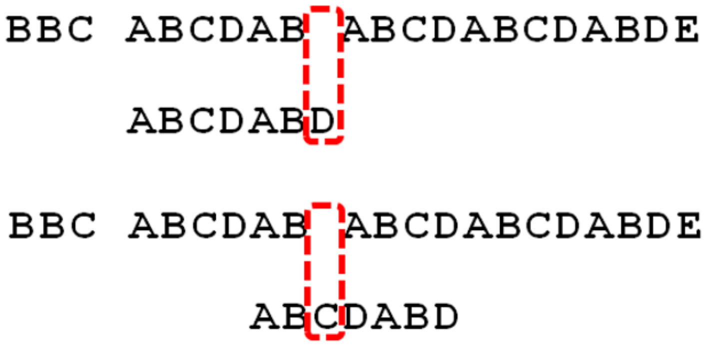
最小生成树: 在一个图中, 利用树结构, 用**最小的代价**连接所有节点.

Kruskal 算法: 贪心的加边法. 代价排序, 如果边的两端为不同树, 则选择边, 融合两个子树. 如果为相同树, 则不选择此边, 重复上述过程, 得到最小生成树.

Prim 算法: 加点法, 利用一个队列, 不断加入当前 cur 的子结点, 然后判断子结点的代价, 如果子结点没在代价表内(map), 则加入此节点, 值为当前代价, 如果存在 map 中, 则判断是否更新该代价.



KMP 算法: 字符串匹配算法, 关键点在于寻找**最长公共前后缀**. 比如前缀内容出现在后缀中,如下图的第二个 AB, 则下一次开始匹配的点就从 C 开始, C 之前是不可能匹配的.



线性代数部分

这部分其实在传统机器学习基础知识 和 PCA 与 SVD 部分已经总结的差不多了. 可自行查阅.