DCMMC

A NOTE FOR COURSES IN THUSZ

Table of Contents

List of Figures

List of Tables

1 | 工程硕士数学 (a.k.a. 数值分析)

主要内容是讲解数值问题和求解方程 – 工程硕士数学 @Winter 2020

Learning Objectives:

- 误差，有效数值
- 数值稳定性

1.1 绪论

1.1.1 计算机上的数

采用**浮点数**表示，尾数和阶数都具有限精度的位数，超过其位数会被**截断**。

病态问题：原始数据的微小变化会引起计算结果的巨大变化。

1.1.2 误差与有效数字

误差来源：

1. 数学模型
2. 观测误差
3. 截断误差
4. 舍入误差

Dependencies: 先修课程：线性代数

设 x, x^* 分别为准确值和近似值，对 $\epsilon = |x - x^*| \leq \sigma(x^*), \epsilon_r = |x - x^*|/|x^*| \leq \sigma_r(x^*) = \sigma(x^*)/|x^*|, \epsilon$ 和 $\sigma(x^*)$ 分别为**绝对误差**和**绝对误差界**，ϵ_r, σ_r 为**相对误差**和**相对误差界**。

将 x^* 转化为类似于科学技术法的形式：

$$|x - x^*| = \pm 10^k \times 0.a_1 a_2 \cdots c_n \cdots \leq 0.5 \times 10^{k-n}$$
$$s.t.\ a_1 \neq 0, 1 \leq a_i \leq 9, a_i \in \mathbb{N}, k \in \mathbb{Z}, n \in \mathbb{N}$$

其中 $\mathbb{Z}$ 表示整数集，n 表示**有效数字位数**。

函数误差：

$$|f(x) - f(x_A)| \leq |f'(x_A)||x - x_A|$$

其中 x_A 是 x 的近似值。

数值计算中常见误差的解决方法：

1. 两个相近的数相减：使用分子有理化将减转化为加

2. 避免大数吃小数

3. 避免除数的绝对值远小于被除数的绝对值

4. 简化运算次数

数值稳定性:

如果算法的初始值有误差，在运算中误差无限增加，不能控制，则该算法是数值不稳定的，反之则是数值稳定的。

1.2 线性代数复习

对角阵，三角阵的乘法和逆矩阵还是对角阵，三角阵。

正定阵：

矩阵的特征值：方程 $(\lambda I - A)x = \mathbf{0}$ 的非零解 $X \neq \mathbf{0}$ 就是 A 的特征向量。

如果特征值均大于 0，则矩阵为正定阵。

Guass 消去法解方程的公式 2.3 不需要记.

2 | 随机过程

主要内容是讲解随机过程 – 随机过程 @Winter 2020

Learning Objectives:
- 概率论复习
- 随机过程的定义

2.1 概率论复习

数学期望:

$$\mu = EX = \int_{-\infty}^{\infty} x \mathrm{d}F(x)$$
$$E[g(X)] = \int g(x) \mathrm{d}F_X(x)$$

实用公式:

1. $E[aX] = aE[X]$，对于常数 a。
2. X, Y 相互独立 $\Rightarrow P(X,Y) = P(X)P(Y) \Rightarrow E[XY] = E[X]E[Y]$。

Dependencies: 先修课程：线性代数，概率论

方差(二阶矩):

$$\sigma = DX := E[(X - EX)^2] = EX^2 - (EX)^2$$

协方差:

$$C(X,Y) := E[(X - EX)(Y - EY)] = E[XY] - E[X]E[Y]$$

用于多个 r.v. 之和的方差计算。

相关系数:

$$R(X,Y) := \rho(X,Y) := \frac{C(X,Y)}{\sigma_X \sigma_Y}$$

复 r.v., 对实 r.v. η, ζ:

$$\xi := \eta + j\zeta, j := \sqrt{-1}$$
$$E\xi := E\eta + jE\zeta, D\xi := E|\xi - E\xi| = E(\xi - E\xi)\overline{(\xi - E\xi)}$$

Schwarz 不等式(**重点**):

$$(E[XY])^2 \le E[X^2]E[Y^2]$$

实用复数公式：$j^2 = -1, a \in \mathbb{R}, b \in \mathbb{R}, z = a + jb, |z| = \sqrt{a^2 + b^2}, \bar{z} = a - jb, z \cdot \bar{z} = |z|^2, \exp(jx) = \cos(x) + j\sin(x), \frac{\partial jx}{\partial x} = j$

Pf.

$$E[(X-\alpha Y)^2] = E[X^2] - 2\alpha E[XY] + \alpha^2 E[Y^2] >= 0$$
$$\text{let } \alpha = \frac{E[XY]}{E[Y^2]}$$
$$\Rightarrow (E[XY])^2 \le E[X^2]E[Y^2]$$

母函数和特征函数这样定义就是为了方便计算，它们能够提供良好的性质：e.g. 母函数可以结合泰勒展开，简化计算均值和方差的计算。

母函数$G(S)$(重点)：

$$r.v.\ \xi, p_k := P\xi = k, k = 0,1,2,\cdots$$
$$G_\xi(s) := Es^\xi = \sum_{k=0}^{\infty} p_k s^k$$

性质：

1. $p_k = G^{(k)}(0)/k!$，k 阶导
2. $E\xi = G'(1), D\xi = G^{(2)}(1) + G'(1) - (G'(1))^2$
3. $\xi_1,\cdots,\xi_n$ 相互独立,$\eta = \sum_{k=1}^n \xi_k, \Rightarrow G_\eta(s) = \prod_{k=1}^n G_{\xi_k}$

常用 Taylor 公式：

$$\exp(x) = \sum_{k=0}^{+\infty} \frac{x^k}{k!}$$

母函数适用于离散型 r.v., 特征函数适用于连续型

特征函数$\Phi(t)$(核心就是 Fourier 变换)：

$$\Phi_\xi(t) := G_\xi(\exp(jt)) = \int_{-\infty}^{\infty} \exp(jt\xi)\mathrm{d}F(x)$$

性质：

1. 共轭对称：$\Phi(-t) = \overline{\Phi(t)}$
2. 若 ξ 的 n 阶矩存在 $\Rightarrow \Phi_\xi^{(k)}(0) = j^k E\xi^k$
3. $EX = j^{-1}\Phi'(0), DX = -\phi^{(2)}(0) - (EX)^2$
4. **非负定性(重点)**：对 $\forall \lambda_i \in \mathbb{C}, \forall t_1,\cdots,t_n \in \mathbb{R}, \Lambda = (\lambda_1,\cdots,\lambda_n)^\top \in \mathbb{C}^n, R_{ij} := \Phi_\xi(t_i - t_j)$,有 $\Lambda R \overline{\Lambda} = \sum_i \sum_k \lambda_i \int \exp(j(t_i - t_j)\xi)\mathrm{d}F(x)\overline{\lambda_j} = \int \sum_i \lambda_i \Phi(t_i) \sum_j \overline{\Phi(t_j)\lambda_j} = \int \left(\sum_i \lambda_i \Phi(t_i)\right)^2 \geq 0$(i.e., $R \succeq 0$)。

多维 r.v. 的特征函数：

$$\Phi(t_1,\cdots,t_n) := E\exp(j(t_1\xi_1,\cdots,t_n\xi_n)) \tag{2.1}$$

2.2 随机过程 (r.p.)

Def:

设 $\{\Omega,\mathcal{F},P\}$ 为概率空间，T 为参数集，若对 $\forall t\in T$，$\xi(t)$ 是一个 r.v., 则称 r.v. 族 $\{\xi(t),t\in T\}$ 为该概率空间上的**随机过程**。

固定样本点 w_0, $\xi_t(w_0)$ 为一个关于 t 的确定性函数，称**样本函数**。

做题的时候我们可以固定 t 来理解题目

对 $\{\xi(t),t\in T\},\forall t_1,\cdots,t_n\in T,$,其 n 维分布函数:$F_{t_1,\cdots,t_n}(x_1,\cdots,x_n):=F(x_1,\cdots,x_n;t_1,\cdots,t_n):=P(\xi(t_1)\le x_1,\cdots,\xi(t_n)\le x_n)$。

分类:

1. T 离散集：随机序列/时间序列

2. T 连续集：随机过程

对 r.p. $\xi(t)$ 定义一些数字特征 (期望，二阶矩，自相关函数，自协方差函数，相关系数):

$$
\begin{aligned}
\mu(t) &:= \int x\mathrm{d}F(x,t)\\
\sigma(t) &:= E\xi^2(t)-(E\xi(t))^2\\
R(t_1,t_2) &:= E\xi(t_1)\xi(t_2)\\
C(t_1,t_2) &:= E[(\xi(t_1)-\mu(t_1))(\xi(t_2)-\mu(t_2))]=R(t_1,t_2)-\mu(t_1)\mu(t_2)\\
\rho(t_1,t_2) &:= \frac{C(t_1,t_2)}{\sigma(t_1)\sigma(t_2)}
\end{aligned}
$$

例题：对二项过程 $\{Y(t)=X_1+\cdots+X_t,t\in T,X_i\sim B(1,p)\}$，有:

r.p. 的均值和方差都是 t 的确定性函数而不是一个值。

$$
\begin{aligned}
\mu_Y(t) &= E[X_1+\cdots+X_t]=tp\\
\sigma_Y(m,n) &= E[Y(m)Y(n)]-EY(m)EY(n)\\
&= E\left[\sum_{i=1}^{m}X_i^2+\sum_{i\neq j}X_iX_j\right]-mnp^2 \quad (\text{w.l.o.g. Let } m=\min(m,n))\\
&= mp-m(n-1)p^2-mnp^2\\
&= \min(m,n)p(1-p)
\end{aligned}
$$

核心思想（**重点**）就是明确 EX_i^2 和 EX_iX_j 是不一样的，并且他们分别有 m 和 $m(n-1)$ 个。

复随机过程:

$\{X(t),t\in T\},\{T(t),t\in T\}$ 两个实过程具有相同的参数集 T 和概率空间，称 $Z(t)=X(t)+jY(t)$ 为复过程。

$$
\begin{aligned}
\mu(t) &:= EZ(t)=EX(t)+jEY(t)\\
C(t_1,t_2) &:= E|Z(t)|^2-|EZ(t)|^2
\end{aligned}
$$

复 r.p. $Z(t) = X(t) + jY(t)$, 其中 X, Y 为实过程:

$$\begin{aligned} \mu(t) &:= EX(t) + jEY(t) \\ R(t_1, t_2) &:= EZ(t_1)\overline{Z(t_2)} \\ DZ(t) &:= E|Z(t) - \mu(t)|^2 \\ C(t_1, t_2) &:= E[(Z(t_1) - \mu(t_1))\overline{(Z(t_2) - \mu(t_2))}] \end{aligned}$$

同样的对复 r.p. 的自相关函数也有非负定性 $R_{ij} := R(t_i, t_j), R \succeq 0$.

2.2.1 二阶矩过程

Def. 对 r.p. $\xi(t)$, $\forall t \in T, E|\xi(t)|^2 \leq \infty$.

常用不等式:

1. $E|\xi + \eta| \leq E|\xi| + E|\eta|$
2. $E\xi \leq |E\xi| \leq E|\xi|$
3. $E|\xi\eta| \leq \sqrt{E|\xi|^2 E|\eta|^2}$ (Schwartz 不等式)
4. $\sqrt{E|\xi + \eta|^2} \leq \sqrt{E|\xi|^2} + \sqrt{E|\eta|^2}$ (三角不等式)

   Pf. of 三角不等式:

$$\begin{aligned} E|\xi + \eta|^2 &= E\xi\overline{\xi} + E\xi\overline{\eta} + E\overline{\xi}\eta + E\eta\overline{\eta} \\ &\leq E|\xi|^2 + 2E|\xi\overline{\eta}| + E|\eta|^2 && (\xi\overline{\xi} = |\xi|^2, \text{ for } a = x + jy, \\ & && a + \overline{a} = 2x \leq 2|a| = 2\sqrt{x^2 + y^2}) \\ &\leq E|\xi|^2 + E|\eta|^2 + 2\sqrt{E|\xi|^2 E|\eta|^2} && (\text{Schwartz 不等式}) \\ &= \sqrt{E|\xi|^2} + \sqrt{E|\eta|^2} \end{aligned}$$

两个 r.p. 之间:

1. 互相关函数 $R_{\xi\eta} := E\xi(t_1)\eta(t_2) = \int\int xy \mathrm{d}F_{\xi\eta}(x, t; t_1, t_2)$
2. 互协方差 $C_{\xi\eta}(t_1, t_2) := E[(\xi(t_1) - \mu_\xi(t_1))(\eta(t_2) - \mu_\eta(t_2))]$

复二阶矩 r.p.
二阶矩空间: $H := \{\xi : E|\xi|^2 \leq \infty\}$ 即二阶矩存在的 r.v. 全体。
性质:

1. 线性空间, 有三角不等式证明
2. 定义范数 (距离空间): $\|\xi\| := \sqrt{E|\xi|^2} := R_\xi(t, t), d(\xi, \eta) := \|\xi - \eta\|$,
   范数算子满足:
   (a) $\|\xi\| \geq 0$
   (b) $\|c\xi\| = |c|\|\xi\|$
   (c) $\|\xi + \eta\| \leq \|\xi\| + \|\eta\|$, 三角不等式

3. 定义内积（内积空间/Hilbert 空间）：$\langle \xi, \eta \rangle := E\xi\overline{\eta} = R(\xi, \eta)$，特性（大多可直接按定义证明）：

   (a) $\langle \eta, \xi \rangle = \overline{\langle \xi, \eta \rangle}$, 注意顺序互换了
   (b) $\langle c\xi, \eta \rangle = c\langle \xi, \eta \rangle, c \in \mathbb{R}$
   (c) $\langle \xi, \xi \rangle = E|\xi|^2 \geq 0$ 且 $\langle \xi, \xi \rangle = 0 \Leftrightarrow \xi = 0, a.e.$ (a.e. 表示**几乎处处**)
   (d) 柯西不等式（a.k.a. Schwartz 不等式）：$|\langle \xi, \eta \rangle| \leq E|\xi\overline{\eta}| \leq \|\xi\|\|\eta\|$

为了满足相应概率上的性质，复数值的乘积一般写为 $a\bar{b}$ 而不是 ab。

r.v. 的收敛性, 对 $\{\xi(t) \in H\}, \exists \xi \in H$:

1. 概率 1 收敛 (强收敛) $\xi_t \xrightarrow{a.s.} \xi : P(\lim_{t\to\infty} = \xi) = 1$
2. **均方收敛**(强收敛, **重点**, 记作 $\text{l.i.m.}_{t\to\infty}\xi_t = \xi$) $\xi_t \xrightarrow{m.s} \xi$: $\lim_{t\to\infty} E|\xi_t - \xi|^2 = \lim_{t\to\infty}\|\xi_t - \xi\| = 0$
3. 依概率收敛 (弱收敛) $\xi_t \xrightarrow{P} \xi : \forall\epsilon > 0, \lim_{t\to\infty} P(|\xi_t - \xi| > \epsilon) = 0$
4. 依分布收敛 (弱收敛) $\xi_t \xrightarrow{P} \xi : \lim_{n\to\infty} F(x,t) = F(x)$

注意 $\xi_t i, \xi$ 分别是 r.p. 和 r.v., 并且上面四个收敛的强弱程度从上往下递减 (**强收敛的 r.p. 当然也是弱收敛**), 不过因为概率 1 收敛条件比较苛刻所以我们一般使用均方收敛.

记住一定不要把 l.i.m. 混淆成 lim, 考试写错直接没分.

均方收敛的重要性质 $\xi_n \xrightarrow{m.s.} \xi, \eta_n \xrightarrow{m.s.} \eta$:

1. 均值收敛 (**积分极限可互换**) $\lim_{n\to\infty} E\xi_n = E\xi = E[\text{l.i.m.}_{n\to\infty}\xi_n]$
2. 范数收敛 $\|\xi_n\| \to \|\xi\|$
3. 内积收敛 $\langle \xi_m, \eta_n \rangle \to \langle \xi, \eta \rangle$ (也就是 $\lim_{m,n\to\infty} E\xi_m\overline{\eta_n} = E\xi\overline{\eta}$)
4. 线性 $a\xi_n + b\eta_n \xrightarrow{m.s.} a\xi + b\eta$

Pf. of 均值收敛:

$$
\begin{aligned}
|E\xi_n - E\xi| &= |E[\xi_n - \xi]| \\
&\leq E|\xi_n - \xi| \\
&\leq \sqrt{E|\xi_n - \xi|^2} && \text{(Schwartz 不等式, 当}P\{Y=1\}=1\text{的特例)} \\
&= \sqrt{\|\xi_n - \xi\|} \\
\xi_n \xrightarrow{m.s.} \xi &\Rightarrow \lim_{n\to\infty} \|\xi_n - \xi\| = 0 \\
&\Rightarrow \lim_{n\to\infty} \sqrt{\|\xi_n - \xi\|} && (\lim\sqrt{f(x)} = \sqrt{\lim f(x)}) = 0 \\
&\Rightarrow \lim_{n\to\infty} |E\xi_n - \xi| = 0
\end{aligned}
$$

Pf. of 内积收敛:

$$
\begin{aligned}
E\xi_m\overline{\eta_n} - E\xi\overline{\eta} &= E[\xi_m\overline{\eta_n} - \xi\overline{\eta}] \\
&= E[(\xi_m - \xi)(\overline{\eta_n - \eta}) + (\xi_m - \xi)\overline{\eta} + \xi(\overline{\eta_n - \eta})] \\
&\leq \sqrt{E|\xi_m - \xi|^2 E|\eta_n - \eta|^2} + \sqrt{E|\xi|^2 E|\eta_n - \eta|^2} + \sqrt{E|\xi_m - \xi|^2 E|\eta|^2} \\
&\xrightarrow[m\to\infty]{n\to\infty} 0
\end{aligned}
$$

还有一种思路就是 $\xi_m(\overline{\eta_n - \eta}) + (\xi_m - \xi)\overline{\eta}$.

Pf. of 范数收敛:

$$
\begin{aligned}
E|\xi_n|^2 - E|\xi|^2 &= E[(|\xi_n| + |\xi|)(|\xi_n| - |\xi|)] \\
&\leq \sqrt{E(|\xi_n| + |\xi|)^2}\sqrt{E(|\xi_n| - |\xi|)^2} \\
&\leq \sqrt{E|\xi_n|^2 E|\xi|^2}\sqrt{E(|\xi_n| - |\xi|)^2}
\end{aligned}
$$

按照师兄经验, 考试基本考作业原题, 不考推导, 切勿落入推导公司的深坑!

证明上述这几个定理的核心思想就是凑 $\lim\|\xi_n - \xi\|$.

Pf. of $\lim_{x\to a}\sqrt{f(x)} = \sqrt{\lim_{x\to a} f(x)}$ (是我太菜了...):

$$
\begin{aligned}
&\lim_{x\to a} f(x) = L \Leftrightarrow \\
&\forall \epsilon_0 > 0, \exists \sigma_0 > 0, \text{ s.t. } |x-a| \in (0,\sigma_0): \\
&\qquad |f(x) - L| < \sigma_0 \\
&\qquad \Downarrow \\
&\forall \epsilon_1 = \frac{\epsilon_0}{\sqrt{L}}, \exists \sigma_1 = \sigma_0, \text{ s.t. } |x-a| \in (0,\sigma_1):
\end{aligned}
$$

$$
\begin{aligned}
|\sqrt{f(x)} - \sqrt{L}| &= \frac{|f(x) - L|}{\sqrt{f(x)} + \sqrt{L}} \\
&\leq \frac{|f(x) - L|}{\sqrt{L}} \\
&< \frac{\epsilon_0}{\sqrt{L}} = \epsilon_1
\end{aligned}
$$

证明均方收敛的两个准则 (**重点**):

柯西准则: $\{\xi_n\}$ is Cauchy 列 $\Leftrightarrow \lim_{m,n\to\infty}\|\xi_m - \xi_n\|^2 = 0 \Leftrightarrow \exists \xi \in H, \xi_n \xrightarrow{m.s.} \xi$.

Loève 准则: $\xi_n \xrightarrow{m.s.} \xi \Leftrightarrow \lim_{m,n\to\infty}\langle \xi_m, \xi_n\rangle = c$, 其中 c 是一个**常数** (不是一定 0!)

对 $t_1 < \cdots t_n, t_i \in T, i \in [1,n]$，若增量

$$
X(t_1), X(t_2) - X(t_1), \cdots, X(t_n) - X(t_{n-1}) \tag{2.2}
$$

相对独立，则 $\{X(t), t \in T\}$ 为**独立增量过程**。若对一切 $0 \leq s < t$，增量 $X(t) - X(s)$ 的分布只依赖 $t - s$, 则 X_T 有**平稳增量**。

3 | 语音信号数字处理

语音信号数字处理 – Winter 2020

Learning Objectives:

- 绪论
- 语音学基础

3.1 绪论

声波的物理描述:

1. 音高 Pitch/音调: 由频率决定，单位：赫兹 Hz。
2. 音色 Timbre: 由声源本身的材料结构确定，说到底就是各个频率分量的振幅（能量）不同。
3. 音强 Loudness/响度: 由声音增幅及离声源的距离决定，单位：分贝 dB。

3.2 语音学基础

Dependencies: 无

语音产生:

1. 声道
2. 发音器官：肺，声带，软腭，阴腭，舌，牙齿，唇

声带（声门）⇒ 音高 ⇒ 基频（源）。

人的有些发音的时候声带是不震动的（清音，voiceless）。

声道（鼻腔，口腔，鼻腔，舌头，etc）⇒ 共振特性 ⇒ 音色/内容（调制，**滤波器**）。

肺部压缩空气力量的大小 ⇒ 响度。

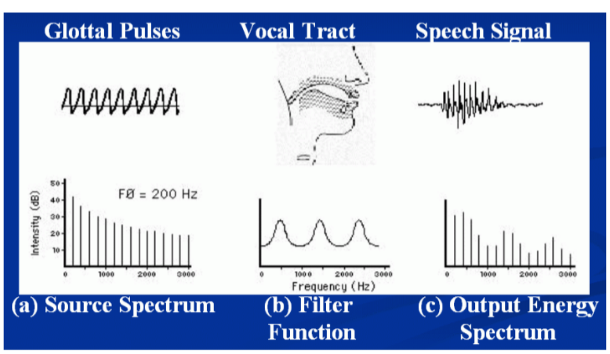

Figure 3.1: Source Filter Model.

Source-Filter Model:

Source Spectrum(谐波)⇒Filter Function(每个峰值代表一个共振频率，提高/降低不同频率的增幅，不会改变频率成分) ⇒Output Energy Spectrum.

谐波就是不同频率的波叠加在一起，随着各个谐波分量频率的增加，其分量的分贝值也在下降。

浊音 Voiced Speech:

由声带震动引起，语音波形（声门波，EGG Signal）具有明显的周期性。声带振动的频率称为基频（F_0），人们可感受到稳定的音高存在。**基频就是谐波分量中频率最低/周期最长的分量的频率**。

清音 Voiceless/Unvoiced Speech：

声带不振动，波形类似白噪声，人们无法感受到稳定的音高存在。

All languages use pitch to express emotionaland other paralinguisticinformation(超语言学信息), and to convey emphasis, contrast, and other such features in what is called intonation(语调)。

Phoneme 音位/音素：发音时不可分割的、最小的音位学单位，国际音标中每一个音标就是一个音位。

Morpheme 语素：最小的、具有语义的结构单元，是最小的语法单位，是最小的语音语义结合体。例如英语的 un-break-able, 或者中文中的词语。这是 NLP 的东西，不是语音学的重点。

Viseme 视位/视素：A visemeis a representational unit used to classify speech sounds in the visual domain, corresponding to the phoneme in the aural domain. 其实就是嘴型这种。

Bimodal Processing 双模态处理：

声音最好同时结合听觉和视觉，因为有时候同一个发音，在只听声音，只看唇形和两者结合随着三种情况下在人的感知里是三种声音（MkGurkEffect）。

Spectrum 语谱：

The spectrum of a signal is a representation of each of its frequency components and their amplitudes（振幅）.

Spectrogram 语谱图：

A spectrogram is a way of envisioning how the different frequencies that make up a waveform change over time.

Wide-band Spectrogram 宽带语谱图：

频率分辨率取 300-400Hz，时间分辨率 2-5ms，良好的时间分辨率，频率分辨率较差。

Narrow-band Spectrogram 窄带语谱图：

频率分辨率取 50-100Hz，时间分辨率 5-10ms，良好的频率分辨率，时间分辨率较差。

共振峰：

是指在声音的频谱中能量相对集中的一些区域（语谱峰值）。声音在经过共振腔时，受到腔体的滤波作用，使得频域中不同频率的能量重新分配。

第一和第二共振峰（F_1 和 F_2）对于区分不同元音尤为重要。

3.3 音频数字化

1. 抽样: 采样频率（抽样周期）
2. 量化: 将无穷多个电压幅度用有限个数字表示电压幅度范围。
3. 编码: 压缩存储

周期信号采样后经 FFT 变换 (又叫频率响应特性) 为离散信号，非周期信号采样后经 FFT 变换为周期信号. 非周期信号可以复制变为周期信号, 这样就可以的到离散的频率响应特征, 离散的频率响应特征方便计算机处理和存储。

Nyquist 抽样定理:

要从抽样信号无失真地恢复信号，采样频率必须大于等于两倍信号谱的最高频率。

而以前的电话语音的采样率只有 $8k$ Hz, 这是因为该音频基本只保留谐波分量中的 F_0, F_1, F_2 功能峰。

4 | 大数据分析

吴志勇老师 – Winter 2020

Learning Objectives:

- 绪论
- 数据抽样和假设检验

4.1 数据抽样和假设检验

例：产品品控的检验：随机抽样来求次品率。

统计推断：Statistical inference is the act of generalizing from a sample (抽样) to a population (总体) with calculated degree of certainty (置信度). e.g. $\overline{x} \to \mu$.

Metric: Precision and Reliability.

Sampling Distribution of a Mean ($\overline{x}$) (SDM).

Central Limit Theorem, unbiasedness (无偏估计), seqaure root law (估计方差: $\sigma_{\overline{x}} = \frac{\sigma_{x_i}}{\sqrt(n)}$).

假设检验:

小概率推断: $1 < \alpha \leq 0.05$

α 显著性水平, Z 统计量, 统计假设 H_0, H_1.

Dependencies: 无

5 | Bibliography