统计信号分析

第1章 统计信号处理中的数学知识复习

本章内容

1.1 概率论概要

确定性现象与随机现象

- **确定性现象**：在一定条件下必然会出现某一结果（或发生某一事件）的现象。
- **随机现象**：在一定条件下可能出现不同结果（或发生不同事件），且不能准确预言究竟出现哪一种结果的现象。

  如：掷硬币、抽牌、掷骰子

1.1.1 随机事件及其概率

- 基本事件，记为 ω 。
- 样本空间，记为 Ω 。
- 事件，记为 $A,B,C,\cdots$ 。

- 古典概率
- 几何概率
- 统计概率

事件的概率：

设随机试验E的样本空间为Ω，对于随机试验E的每一随机事件A，都赋予唯一确定的实数$P(A)$，并且集合函数$P(\cdot)$满足下列条件：

(1) 非负性：对每一个事件$A\subset\Omega$，都有 $P(A)\geq 0$；

(2) 规范性：$P(\Omega)=1$；

(3) 可列可加性：对任意互不相容的事件 $A_1,A_2,\cdots,A_i,\cdots$，有

$$P\left(\bigcup_{i=1}^{+\infty}A_i\right)=\sum_{i=1}^{+\infty}P(A_i)$$

条件概率：

设A和B为任意两个随机事件，且$P(B)>0$，称

$$P(A|B)=\frac{P(AB)}{P(B)}$$

为事件B发生条件下事件A发生的**条件概率**，也称A对B的概率。

全概率公式：

设某随机试验的样本空间Ω中的事件$A_1,A_2,\cdots,A_i,\cdots$（有限个或可列个）构成一个完备事件组，且$P(A_i)>0,\ i=1,2,\cdots$，则对任一事件$B$，有

$$P(B)=\sum_i P(A_i)P(B|A_i)$$

贝叶斯公式：

设某随机试验的样本空间Ω中的事件$A_1,A_2,\cdots,A_i,\cdots$（有限个或可列个）构成一个完备事件组，且$P(A_i)>0,\ i=1,2,\cdots$，则对任一事件$B$，$P(B)>0$，有

$$P(A_m|B)=\frac{P(A_m)P(B|A_m)}{\sum_i P(A_i)P(B|A_i)},\quad m=1,2,\cdots$$

事件的独立性：

如果随机事件A与B满足关系$P(AB)=P(A)P(B)$，则称事件A与B是相互独立的。

性质：若事件$A_1,A_2,\cdots,A_n$相互独立，则有

$$P\left(\bigcup_{i=1}^{n}\bar{A}_i\right)=1-\prod_{i=1}^{n}P(A_i)$$

1.1.2　随机变量及其分布

随机变量：

设某随机试验的样本空间为$\Omega=\{\omega\}$，对于每一个样本点$\omega\in\Omega$都有惟一的实数$X(\omega)$与之对应，这样就得到一个定义在Ω上的单值实函数$X=X(\omega)$。如果对任意实数x，“$X(\omega)\le x$”都是一个随机事件，并有其确定的概率，则称$X=X(\omega)$为随机变量。

$$\text{随机变量}\begin{cases}\text{离散型随机变量}\\ \text{连续型随机变量}\end{cases}$$

离散型随机变量：

随机变量X的全部可能取值为有限个或可列个。随机变量X所有可能的取值为$x_i\ (i=1,2,\cdots)$，事件$X=x_i$的概率为$P(X=x_i)=p_i,\quad i=1,2,\cdots$
上式称为离散型随机变量X的**概率分布或分布率**。

设X是一个随机变量，对任意实数$x(-\infty<x<+\infty)$，令$F(x)=P(X\le x)$，则称$F(x)$为随机变量X的**概率分布函数**。它的定义域是$(-\infty,+\infty)$，值域是$[0,1]$。

常见的离散型随机变量

0–1分布： 当一个随机试验只有两个可能结果 A与 $\bar{A}$ 时，随机变量

$$X=\begin{cases}0, & 当\bar{A}出现时\\ 1, & 当A出现时\end{cases}$$

X 表示在试验中事件A出现的次数，并设$P(A)=p(0<p<1)$，则X的概率分布为

$$P(X=k)=p^k(1-p)^{1-k},\quad k=0,1$$

这时称 X 服从参数为 p 的0–1分布，记为 $X\sim B(1,p)$。

常见的离散型随机变量

泊松分布： 如果随机变量X的分布为：

$$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}$$

其中 $\lambda>0$为常数，则称X 服从参数为 λ 的泊松分布，记为 $X\sim P(\lambda)$。

连续型随机变量：

如果对于随机变量X 的分布函数$F(x)$，存在一个非负可积函数 $f(x)(-\infty<x<+\infty)$，有

$$F(x)=\int_{-\infty}^{x}f(t)dt$$

则称X为**连续型随机变量**，函数 $f(x)$为其**概率密度函数**。

常见的连续型随机变量

均匀分布： 如果连续型随机变量X 的概率密度函数为

$$f(x)=\begin{cases}\dfrac{1}{b-a}, & a\le x\le b\\ 0 & 其它\end{cases}$$

其中 a和 b 为常数，则称X 在区间 $[a,b]$ 上服从均匀分布，记为$X\sim U(a,b)$。

常见的连续型随机变量

高斯分布： 如果连续型随机变量X 的概率密度函数为

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma_X}\exp\left\{-\frac{(x-m_X)^2}{2\sigma_X^2}\right\},\quad -\infty<x<+\infty$$

其中 m_X 和 σ_X 为常数，且 $\sigma_X>0$，则称X 服从参数为 m_X 和 σ_X 的高斯分布（又称正态分布），记为 $X\sim N(m_X,\sigma_X^2)$ 。

常见的连续型随机变量

瑞利分布： 如果连续型随机变量X的概率密度函数为

$$f(x)=\begin{cases}\dfrac{x}{\mu^2}e^{-\frac{x^2}{2\mu^2}}, & x\ge 0\\ 0, & x<0\end{cases}$$

其中 μ 为常数，且 μ>0，则称X 服从参数为 μ 的瑞利分布。

常见的连续型随机变量

χ^2**分布：** 如果连续型随机变量X的概率密度函数为

$$f(x)=\begin{cases}\dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}x^{\frac{n}{2}-1}e^{-\frac{x}{2}}, & x>0\\ 0, & x\le 0\end{cases}$$

其中n为正整数，则称 X 服从自由度为n的 χ^2分布。

常见的连续型随机变量

莱斯分布： 如果连续型随机变量X 的概率密度函数为

$$f(x)=\begin{cases}\dfrac{x}{\sigma^2}e^{-\frac{(x^2+v^2)}{2\sigma^2}}I_0\left(\dfrac{xv}{\sigma^2}\right), & x>0\\ 0, & x\le 0\end{cases}$$

其中 σ 为常数，且 $\sigma>0$ ，$I_0(z)$ 是零阶第一类贝塞尔(Bessel)函数，则称X服从莱斯分布。

1.1.3 多维随机变量

定义：

设某随机试验的样本空间为 $\Omega=\{\omega\}$ ，如果每一个样本点 $\omega\in\Omega$ ，$X_1=X_1(\omega)$ 、$X_2=X_2(\omega)$ … $X_n=X_n(\omega)$ ，是定义在同一个样本空间 Ω 上的n个随机变量，则 $(X_1,X_2,\cdots,X_n)$ 称为***n* 维随机变量**，其矢量形式 $[X_1,X_2,\cdots,X_n]^T$ 也称为**随机矢量**。

下面以二维为例进行说明，更高维以此类推。

1.1.3 多维随机变量

联合分布函数：

设 (X,Y)是二维随机变量，$F(x,y)=P(X\le x,Y\le y)$ 称为二维随机变量(X,Y)的分布函数，或称随机变量 X 和Y 的联合分布函数。

多维离散型随机变量：

二维随机变量(X,Y)所有可能的取值是有限对或可列无限多对，并且以确定的概率取各个不同的数对

$$P(X=x_i,Y=y_j)=p_{ij},\quad i,j=1,2,\cdots$$

称为(X,Y)的**联合概率分布**。

多维连续型随机变量：

二维随机变量(X,Y)的分布函数是$F(x,y)$，存在非负函数 $f(x,y)$，使得对任意实数 x,y 有

$$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}f(u,v)dudv$$

称 $f(x,y)$ 为二维连续型随机变量(X,Y)的**联合概率密度函数**。

边缘分布函数：

二维随机变量(X,Y)的分布函数为$F(x,y)$，随机变量X或Y的分布即为二维随机变量的边缘分布。它与二维变量的分布函数具有如下关系：

$$F_X(x)=P(X\le x,Y<+\infty)=F(x,+\infty)$$

$$F_Y(y)=P(X<+\infty,Y\le y)=F(+\infty,y)$$

条件概率与条件分布函数：

对二维离散随机变量，在$Y=y_i$条件下$X=x_i$的条件概率为

$$P(X=x_i|Y=y_i)=\frac{P(X=x_i,Y=y_i)}{P(Y=y_i)}$$

其中$P(Y=y_i)$表示边缘分布函数。

条件概率与条件分布函数：

对二维连续型随机变量，设y是定值，任意$\Delta y>0$，$P(y-\Delta y<Y\le y+\Delta y)>0$，若对任意实数$x$，极限

$$\lim_{\Delta y\to 0}P(X\le x|y-\Delta y<Y\le y+\Delta y)$$

存在，则称此极限为在$Y=y$条件下X的条件分布函数，记为$P(X\le x|Y=y)$。

独立性：

若对任意实数x和y，有$F(x,y)=F_X(x)F_Y(y)$，则称随机变量X和Y是独立的。

若X与Y独立，显然有

$$P(X=x_i|Y=y_i)=\frac{P(X=x_i,Y=y_i)}{P(Y=y_i)}=P(X=x_i)$$

1.1.4 随机变量的数字特征

随机变量X的数学期望（均值）：

$$E\{X\}=\int_{-\infty}^{+\infty}xf(x)dx$$

基本性质：

（1）常量的数学期望等于常量本身。

（2）对于常数a_i $(i=1,2,\cdots,n)$ 则

$$E\left\{\sum_{i=1}^{n}a_iX_i\right\}=\sum_{i=1}^{n}a_iE\{X_i\}$$

（3）若X_i $(i=1,2,\cdots,n)$彼此统计独立，则

$$E\{X_1X_2\cdots X_n\}=\prod_{i=1}^{n}E\{X_i\}$$

随机变量X的方差：

$$\begin{aligned}Var\{X\}&=E\left\{[X-E\{X\}]^2\right\}\\&=\int_{-\infty}^{+\infty}(x-E\{X\})^2f(x)dx\\&=E\{X^2\}-(E\{X\})^2\end{aligned}$$

通常称$\sqrt{Var\{X\}}$为随机变量X的均方差或标准差，习惯上用σ_X^2表示$Var\{X\}$。

方差基本性质：

（1）$Var\{C\}=0, C$为常数。

（2）对于常数 a_i $(i=1,2,\cdots,n)$，则

$$Var\left\{\sum_{i=1}^{n}a_iX_i\right\}=\sum_{i=1}^{n}\sum_{j=1}^{n}a_ia_jE\left\{\left(X_i-E\{X_i\}\right)\left(X_j-E\{X_j\}\right)\right\}$$

（3）若 X_i $(i=1,2,\cdots,n)$ 彼此统计独立，那么

$$Var\left\{\sum_{i=1}^{n}a_iX_i\right\}=\sum_{i=1}^{n}a_i^2Var\{X_i\}$$

（4）$E\left\{(X-C)^2\right\}\geq Var\{X\}$，$C$为任意常数。

协方差：

对多维随机变量 $X=(X_1,X_2,\cdots,X_n)$，X_i 与 X_j 的协方差：

$$\begin{aligned}C_{ij}&=Cov\{X_i,X_j\}\\&=E\left\{\left(X_i-E\{X_i\}\right)\left(X_j-E\{X_j\}\right)\right\}\\&\qquad i,j=1,2,\cdots,n\end{aligned}$$

显然

$$Var\{X_i\}=C_{ii}$$

相关系数：

“归一化”的协方差，称为相关系数：

$$\begin{aligned}\rho_{X_iX_j}&=\frac{E\left\{\left(X_i-E\{X_i\}\right)\left(X_j-E\{X_j\}\right)\right\}}{\sqrt{E\left\{\left(X_i-E\{X_i\}\right)^2\right\}\cdot E\left\{\left(X_j-E\{X_j\}\right)^2\right\}}}\\&=\frac{C_{ij}}{\sqrt{Var\{X_i\}}\cdot\sqrt{Var\{X_j\}}}\end{aligned}$$

协方差和相关系数的性质：

（1）$Cov\{X,Y\}=Cov\{Y,X\}$。

（2）$Cov\{aX,bY\}=abCov\{X,Y\}$，其中$a$ 和 b 为常数。

（3）$Cov\{X_1+X_2,Y\}=Cov\{X_1,Y\}+Cov\{X_2,Y\}$。

（4）$|\rho_{XY}|\leq 1$ 。

（5）若X与Y相互独立，则X与Y不相关。反之不一定成立。只有 X 与Y 服从高斯分布时，统计独立和互不相关才是等价的。

联合分布函数：

对于n 维随机变量 $(X_1,X_2,\cdots,X_n)$，

$$F(x_1,x_2,\cdots,x_n)=P(X_1\leq x_1,X_2\leq x_2,\cdots,X_n\leq x_n)$$

关于 X_i 的边缘分布函数：

$$F_{X_i}(x_i)=$$
$$P(X_1<+\infty,\cdots,X_{i-1}<+\infty,X_i\leq x_i,X_{i+1}<+\infty,\cdots,X_n<+\infty)$$

随机变量 $X_1,X_2,\cdots,X_n$是相互独立的，若对于所有的 $x_1,x_2,\cdots,x_n$ ，有

$$F(x_1,x_2,\cdots,x_n)=F_{X_1}(x_1)F_{X_2}(x_2)\cdots F_{X_n}(x_n)$$

特征函数：

若X 是随机变量，则称复随机变量 $e^{j\omega X}$ 的均值

$$\varphi(\omega)=E\left\{e^{j\omega X}\right\}$$

为X 的特征函数。

（1）若X 是离散型随机变量，其可能取值为$x_1,x_2,\cdots$ ，且 $P(X=x_k)=p_k$ ，则 $\varphi(\omega)=\sum_k e^{j\omega x_k}p_k$

（2）若X 是连续型随机变量，其概率密度函数为 $f(x)$，则 $\varphi(\omega)=\int_{-\infty}^{+\infty}e^{j\omega x}f(x)dx$

特征函数:

同理，我们可推导出多维随机变量$(X_1, X_2, \cdots, X_n)$的特征函数：

$$\varphi(\omega_1, \omega_2, \cdots, \omega_n) = E\left\{e^{j(\omega_1 X_1 + \omega_2 X_2 + \cdots + \omega_n X_n)}\right\}$$

特征函数是概率密度函数的傅立叶变换。已知随机变量的特征函数，可以求出它的各阶矩，特征函数也被称为矩生成函数。

1.1.5 高斯随机变量

1. 一维高斯随机变量

一维高斯随机变量X 的概率密度函数：

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma_X} \exp\left[-\frac{(x - m_X)^2}{2\sigma_X^2}\right], \quad -\infty < x < +\infty$$

如果 $m_X = 0$，$\sigma_X^2 = 1$，则

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{x^2}{2}\right], \quad -\infty < x < +\infty$$

2. 多维高斯随机变量

$X_1, X_2, \cdots, X_n$ 组成n维高斯随机变量，令 $\mathbf{x} = [X_1, X_2, \cdots, X_n]^T$

均值矢量：

$$\mathbf{m}_\mathbf{x} = [E\{X_1\}, E\{X_2\}, \cdots, E\{X_n\}]^T = [m_{X_1}, m_{X_2}, \cdots, m_{X_n}]^T$$

协方差矩阵：

$$\mathbf{C}_\mathbf{x} = \begin{bmatrix} E\left\{(X_1 - m_{X_1})^2\right\} & \cdots & \cdots & E\left\{(X_1 - m_{X_1})(X_n - m_{X_n})\right\} \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ E\left\{(X_n - m_{X_n})(X_1 - m_{X_1})\right\} & \cdots & \cdots & E\left\{(X_n - m_{X_n})^2\right\} \end{bmatrix}$$

n 维高斯随机变量的联合高斯概率密度函数：

$$f(\mathbf{X}) = \frac{1}{(2\pi)^{n/2} |\mathbf{C}_\mathbf{x}|^{1/2}} \exp\left[-\frac{(\mathbf{X} - \mathbf{m}_\mathbf{x})^T \mathbf{C}_\mathbf{x}^{-1} (\mathbf{X} - \mathbf{m}_\mathbf{x})}{2}\right]$$

性质：

(1) n维高斯随机变量经过线性变换后仍是高斯随机变量。

(2) n维互不相关的高斯随机变量一定是彼此统计独立的。

(3) 一般的n维随机变量，若它们彼此统计独立，则必然互不相关，反之，则不一定成立。

1.1.6 随机变量函数的分布

1. 一维随机变量函数的分布

X是连续型随机变量，概率密度函数为$f(x), -\infty < x < +\infty$，$g(x)$ 处处可导且有 $g'(x) > 0$（或恒有$g'(x) < 0$），则 $Y = g(X)$ 也是一个连续型随机变量，其概率密度函数为

$$f(y) = \begin{cases} f(h(y))\,|h'(y)|, & \alpha < y < \beta \\ 0, & \text{其他} \end{cases}$$

其中 $\alpha = \min\{g(-\infty), g(+\infty)\}$, $\beta = \max\{g(-\infty), g(+\infty)\}$, $h(y)$ 是 $g(x)$ 的反函数。

特别地，若随机变量$X \sim N(\mu, \sigma^2)$，令随机变量$Y = \frac{X - \mu}{\sigma}$，则 $Y \sim N(0,1)$。

2. 多维随机变量函数的分布

$(X_1, X_2, \cdots, X_n)$ 是具有概率密度函数 $f(x_1, x_2, \cdots, x_n)$ 的连续型n维随机变量，假设

(1) $Y_1 = g_1(X_1, X_2, \cdots, X_n), \cdots, Y_n = g_n(X_1, X_2, \cdots, X_n)$ 是n维实数空间到自身的一对一映射；

(2) 变换 g_i 和它的逆 h_i 都是连续的；

(3) 偏导数 $\frac{\partial h_i}{\partial y_j}, i = 1, 2, \cdots, n, j = 1, 2, \cdots, n$存在且连续；

(4) 逆变换的雅可比行列式

$$J(y_1, y_2, \cdots, y_n) = \begin{vmatrix} \frac{\partial h_1}{\partial y_1} & \frac{\partial h_1}{\partial y_2} & \cdots & \frac{\partial h_1}{\partial y_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial h_n}{\partial y_1} & \frac{\partial h_n}{\partial y_2} & \cdots & \frac{\partial h_n}{\partial y_n} \end{vmatrix} \neq 0$$

则$(Y_1, Y_2, \cdots, Y_n)$具有联合概率密度函数

$$f(y_1, y_2, \cdots, y_n) = |J|\, f\left(h_1(y_1, y_2, \cdots, y_n), \cdots, h_n(y_1, y_2, \cdots, y_n)\right)$$

1.1.7 复随机变量

复随机变量

$Z = X_1 + jX_2$，其中，X_1 和 X_2 为两个实随机变量。

概率密度函数

$f(z) = f(x_1, x_2)$， $f(z)$ 显然是一个实数。

均值

$$m_Z = E\{Z\} = E\{X_1 + jX_2\} = \int_{-\infty}^{+\infty} (x_1 + jx_2) f(x_1, x_2) dx_1 dx_2$$

方差

$$\sigma_Z^2 = E\left\{|Z - m_Z|^2\right\} = \sigma_{X_1}^2 + \sigma_{X_2}^2 = E\left\{|Z|^2\right\} - m_Z^2$$

对于两个复随机变量 $W = U_1 + jU_2, Z = X_1 + jX_2$，

$$f(w, z) = f(x_1, x_2, u_1, u_2)$$

其协方差为

$$C_{WZ} = E\left\{(W - m_W)(Z - m_Z)^*\right\}$$

相关系数为

$$\rho_{WZ} = \frac{C_{WZ}}{\sigma_W \sigma_Z}$$

对于复随机矢量 w 和 z，协方差矩阵和互协方差矩阵分别为

$$\mathbf{C}_Z = E\left\{(\mathbf{z} - \mathbf{m}_Z)(\mathbf{z} - \mathbf{m}_Z)^H\right\}$$

$$\mathbf{C}_{WZ} = E\left\{(\mathbf{w} - \mathbf{m}_W)(\mathbf{z} - \mathbf{m}_Z)^H\right\}$$

复高斯随机矢量的概率密度函数为

$$f(\mathbf{z}) = \frac{1}{\pi^n |\mathbf{C}_Z|} \exp\left[-(\mathbf{z} - \mathbf{m}_Z)^H \mathbf{C}_Z^{-1} (\mathbf{z} - \mathbf{m}_Z)\right]$$

高斯随机变量的特征函数为

$$\varphi(\omega) = \exp\left(j \operatorname{Re}\left\{\omega^H \mathbf{m}_Z\right\} - \frac{\omega^H \mathbf{C}_Z \omega}{4}\right)$$

1.2 随机过程基础

1.2.1 平稳与非平稳随机过程

1.2.2 随机过程的统计特性与维纳-辛钦定理

1.2.3 高斯随机过程

1.2.4 随机过程的积分微分特性

自然界存在一类随机现象，与之相联系的随机事件不能用一般的单维或多维随机变量去描述它。如：

- 用 $x(t)$ 表示 t 时刻以前某通信站接到的呼唤次数，对于某个固定的 t，$x(t)$ 是一个随机量，但 $x(t)$ 这类随机量将随着 t 的变化而变化；
- 电网电压可看作随时间变化的随机量 $V(t)$；
- 雷达接收机的噪声输出 $n(t)$ 。

用随机过程才能描述的随机现象，每做一次随机试验，随机试验的结果应是某一个随机现实，每一次随机试验之前，其试验结果究竟属于哪一种随机现实，事先不能预测。

设 $T \subset \mathbb{R}^1$，对于每一个 $t \in T$，$x_t(\omega)$ 为一随机变量。其中 ω 代表某概率空间 (Ω, F, P) 的元素，则 $\{x_t(\omega), t \in T\}$ 即为**随机过程**。简记为 $\{x(t)\}, t \in T$ 。

1.2.1 平稳与非平稳随机过程

定义：

如果由 $\{x(t)\}$ 所确定的n维概率密度函数 $f(x_1,x_2,\cdots,x_n;t_1,t_2,\cdots,t_n)$ 满足

$$f(x_1,x_2,\cdots,x_n;t_1,t_2,\cdots,t_n) = f(x_1,x_2,\cdots,x_n;t_1+\tau,t_2+\tau,\cdots,t_n+\tau)$$

称 $\{x(t)\}$ 为**严格平稳随机过程**，又称为**强平稳随机过程**或**狭义平稳随机过程**。

1.2.1 平稳与非平稳随机过程

定义：

当 $E\{x^2(t)\}<+\infty$ 时，若满足：

$$E\{x(t)\}=E\{x(t+\tau)\}$$

$$E\{x(t_1)x(t_2)\}=E\{x(t_1+\tau)x(t_2+\tau)\}$$

称 $\{x(t)\}$ 为**广义平稳随机过程**，又称为**弱平稳随机过程**或**宽平稳随机过程**。

讨论：

当$E\{x^2(t)\}<+\infty$时，严格平稳随机过程也是广义平稳的，但广义平稳随机过程不一定是严格平稳的。

有一类随机过程，其本质是随机过程，但其表示形式却类似确定过程，称为**准随机过程**。

例： $\left\{x(t)=\sum_{k=0}^{\tau}\xi_k t^k\right\}$，$\xi_k$ $(k=1,2,\cdots,\tau)$ 为随机变量；

$\{x(t)=A\sin(\omega_0 t+\theta)\}$， A,θ 为随机变量。

复随机过程：

$$\{z(t)=x(t)+jy(t)\}$$

其概率密度函数为

$$f(x_1,x_2,\cdots,x_n;y_1,y_2,\cdots,y_n;t_1,t_2,\cdots,t_n)$$

高斯随机过程

任意时刻 $t_1,t_2,\cdots,t_n$，随机过程 $\{x(t)\}$ 所形成的n维随机变量，其概率密度函数为高斯分布

$$f(x_1,x_2,\cdots,x_n;t_1,t_2,\cdots,t_n) = \frac{1}{(2\pi)^{n/2}|\mathbf{C_x}|^{1/2}}\exp\left[-\frac{1}{2}(\mathbf{x}-\mathbf{m_x})^T\mathbf{C_x^{-1}}(\mathbf{x}-\mathbf{m_x})\right]$$

这时称随机过程 $\{x(t)\}$ 为**高斯过程**。

性质：

- 若高斯随机过程是广义平稳的，则一定是严格平稳的。
- 高斯随机过程经过线性变换，其输出仍是高斯随机过程。

1.2.2 随机过程的统计特性与维纳-辛钦定理

1. 随机过程的统计特性

时间平均：

随机过程$\{x(t)\}$的第 k 个现实为$x^{(k)}(t)$，相应有：

时间平均均值：

$$\langle x^{(k)}(t)\rangle = \lim_{T\to+\infty}\frac{1}{2T}\int_{-T}^{T}x^{(k)}(t)\,dt$$

时间平均二阶矩：

$$\left\langle\left[x^{(k)}(t)\right]^2\right\rangle = \lim_{T\to+\infty}\frac{1}{2T}\int_{-T}^{T}\left[x^{(k)}(t)\right]^2 dt$$

时间平均相关函数：

$$\langle x^{(k)}(t)x^{(k)}(t-\tau)\rangle = \lim_{T\to+\infty}\frac{1}{2T}\int_{-T}^{T}x^{(k)}(t)x^{(k)}(t-\tau)\,dt$$

时间平均互相关：

$$\langle x^{(k)}(t)y^{(k)}(t-\tau)\rangle = \lim_{T\to+\infty}\frac{1}{2T}\int_{-T}^{T}x^{(k)}(t)y^{(k)}(t-\tau)\,dt$$

随机过程统计平均（集平均）

随机过程的均值：

$$E\{x(t)\}=\int_{-\infty}^{+\infty}xf(x;t)dx=m_X(t)$$

随机过程的方差：

$$E\left\{\left[x(t)-m_X(t)\right]^2\right\}=\int_{-\infty}^{+\infty}\left[x-m_X(t)\right]^2 f(x;t)dx=\sigma_X^2(t)$$

随机过程的自相关函数：

$$E\{x(t_1)x(t_2)\}=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}x_1x_2 f(x_1,x_2;t_1,t_2)dx_1dx_2=R_X(t_1,t_2)$$

讨论：随机过程$\{x(t)\}$ 的自相关函数满足对称性，即

$$R_X(t_1,t_2)=R_X(t_2,t_1)$$

对于非平稳随机过程，其均值、方差及自相关函数都是t的函数，其自相关函数可用 $R_X(t,\tau)$ 表示。

- 对于严格平稳随机过程，

$$f(x;t)=f(x;t+\tau)=f(x)$$

$$f(x_1,x_2;t_1,t_2)=f(x_1,x_2;t_1-t_2=\tau)=f(x_1,x_2;\tau)$$

值与t 无关，$R_X(t_1,t_2)=R_X(\tau)$ 只与$\tau=t_1-t_2$ 有关。

- 对于广义平稳随机过程，均值与t无关，自相关函数是$\tau=t_1-t_2$的函数。

随机过程的相关系数：

$$\rho_X(t_1,t_2)=\frac{E\{[x(t_1)-m(t_1)][x(t_2)-m(t_2)]\}}{\sigma_1\sigma_2}$$

对于平稳随机过程，有$\rho_X(\tau)=\frac{R_X(\tau)-m^2}{\sigma^2}$

随机过程的互相关函数：

随机过程$\{x(t)\},\{y(t)\}$，前者对于后者的互相关函数：

$$R_{XY}(t_1,t_2)=E\{x(t_1)y(t_2)\}=\iint x_1y_2 f(x_1,y_2;t_1,t_2)dx_1dy_2$$

后者对于前者的互相关函数：

$$R_{YX}(t_1,t_2)=E\{y(t_1)x(t_2)\}=\iint y_1x_2 f(y_1,x_2;t_1,t_2)dy_1dx_2$$

随机过程的协方差函数：

$$C_X(t_1,t_2)=E\{(x(t_1)-m_1)(x(t_2)-m_2)\}=R_X(t_1,t_2)-m_1m_2$$

讨论：协方差函数亦具有对称性。

对于平稳随机过程，

$$C_X(t_1,t_2)=C_X(t_1-t_2)=C_X(\tau)=R_X(\tau)-m^2$$

对于联合随机过程，互协方差函数为：

$$C_{XY}(t_1,t_2)=E\{[x(t_1)-m_X(t_1)][y(t_2)-m_Y(t_2)]\}=R_{XY}(t_1,t_2)-m_X(t_1)m_Y(t_2)$$

$$C_{YX}(t_1,t_2)=E\{[y(t_1)-m_Y(t_1)][x(t_2)-m_X(t_2)]\}=R_{YX}(t_1,t_2)-m_Y(t_1)m_X(t_2)$$

显见，互协方差函数不具有对称性。

随机过程的各态历经性

平稳随机过程的所有各类集平均统计特征，以概率1等于由任意实现得到的相应的时间平均特征。

各态历经性的意义：

2. 维纳-辛钦定理

确定信号 $s(t)$ 如满足$E=\int_{-\infty}^{+\infty}s^2(t)dt<+\infty$，则为能量信号。满足 $P=\lim\limits_{T\to+\infty}\frac{1}{T}\int_{-T/2}^{T/2}s^2(t)dt<+\infty$ 的确定信号为功率信号。

维纳—辛钦定理：

平稳随机信号的功率谱密度是其相关函数的傅里叶变换，即：

$$S_X(\omega)=\int_{-\infty}^{+\infty}R_X(\tau)e^{-j\omega\tau}d\tau \quad R_X(\tau)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}S_X(\omega)e^{j\omega\tau}d\omega$$

以上两式成立的条件是

$$\int_{-\infty}^{+\infty}S_X(\omega)d\omega<+\infty，\int_{-\infty}^{+\infty}|R_X(\tau)|d\tau<+\infty$$

1.3 线性代数导论

1.3.1 矩阵的概念和基本运算

- 矩阵、矢量概念
- 基本运算
  - 转置
  - 共轭
  - 共轭转置
  - 内积
  - 矩阵的逆（广义逆）

把n阶方阵A中元素 a_{ij} 所在的第 i 行和第 j 列划去后，剩下的n-1阶方阵称为**元素 a_{ij} 的余子矩阵**，记为 $(-1)^{i+j}\det\{\mathbf{A}(i,j)\}$, $\mathbf{A}(i,j)$为元素 a_{ij} 的余子式。
$cof(\mathbf{A}_{ij})=\det\{\mathbf{A}(i,j)\}$ 称为元素 a_{ij} 的代数余子式，记为 $\mathbf{A}_{ij}$ 。

行列式

特别地，当 $\det\{\mathbf{A}\}=0$ 时，A 称为**奇异（退化）方阵**；否则称为**非奇异（非退化）方阵**，也称**正则矩阵**。

设 A 和 B 是 n 阶方阵，则

$$\det\{\mathbf{AB}\}=\det\{\mathbf{A}\}\det\{\mathbf{B}\}$$

设矩阵 $\mathbf{A}\in\mathbb{C}^{m\times n}$ 有一个k 阶子式不等于零，而 A 的所有k+1阶子式（若存在）都等于零，则称正整数k为矩阵 A 的**秩**。
若 m=n=k，则称之为**满秩矩阵**；
若k=0，则称为**零秩矩阵**。

设矩阵A为n阶方阵，如对于任意n阶非零列矢量x，有：
（1）当 $\mathrm{Re}\{\mathbf{x}^H\mathbf{A}\mathbf{x}\}>0$ 时，则称A为**正定矩阵**；
（2）当 $\mathrm{Re}\{\mathbf{x}^H\mathbf{A}\mathbf{x}\}\geq 0$ 时，则称A为**半正定（非负定）矩阵**；
（3）当 $\mathrm{Re}\{\mathbf{x}^H\mathbf{A}\mathbf{x}\}<0$ 时，则称A为**负定矩阵**。

如矩阵A的某个函数$\|\mathbf{A}\|$ ，满足：
（1）A为非零矩阵时，$\|\mathbf{A}\|>0$ ；$\mathbf{A}=\mathbf{O}$ 时，$\|\mathbf{A}\|=0$ ；
（2）对于任意复数c有 $\|c\mathbf{A}\|=|c|\|\mathbf{A}\|$ ；
（3）$\|\mathbf{A}+\mathbf{B}\|\leq\|\mathbf{A}\|+\|\mathbf{B}\|$ ；
（4）$\|\mathbf{AB}\|\leq\|\mathbf{A}\|\|\mathbf{B}\|$ 。
则称 $\|\mathbf{A}\|$ 是 $\mathbb{C}^{m\times n}$ 上的**一个矩阵范数**。

常用的矩阵范数：

（1）Frobenius范数

$$\|\mathbf{A}\|_F=\left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2\right)^{1/2}$$

（2）行和范数

$$\|\mathbf{A}\|_{row}=\max_{1\le i\le m}\left\{\sum_{j=1}^{n}|a_{ij}|\right\}$$

（3）列和范数

$$\|\mathbf{A}\|_{col}=\max_{1\le j\le n}\left\{\sum_{i=1}^{m}|a_{ij}|\right\}$$

施瓦兹不等式：

设 $\mathbf{A}\in\mathbb{C}^{m\times n},\mathbf{B}\in\mathbb{C}^{n\times l}$ ，$\mathbf{AA}^T$ 可逆，则

$$\mathbf{B}^T\mathbf{B}\ge(\mathbf{AB})^T\left(\mathbf{AA}^T\right)^{-1}(\mathbf{AB})$$

等号成立的充要条件是存在 $\mathbf{C}\in\mathbb{C}^{m\times l}$ ，使得 $\mathbf{B}=\mathbf{A}^T\mathbf{C}$ 。

1.3.2 特殊矩阵

- 单位矩阵　　反向单位矩阵
- 厄米特矩阵　反厄米特矩阵
- 范德蒙矩阵
- 汉克矩阵
- 托普利兹矩阵
- 正规矩阵
- 酉矩阵
- 正交矩阵

1.3.3 矩阵分解

常用的矩阵分解方法：

- 特征值分解(EVD)
- 奇异值分解(SVD)
- Cholesky分解
- LU分解
- QR分解

1.特征值分解

特征值与特征矢量

若 $\lambda_1,\lambda_2,\cdots,\lambda_n$ 是 n 阶方阵A的特征值，$\mathbf{u}_1,\mathbf{u}_2,\cdots,\mathbf{u}_n$ 分别是与特征值对应的特征矢量，则

$$\mathbf{A}=\mathbf{E}\mathbf{\Lambda}\mathbf{E}^{-1}$$

其中 $\mathbf{E}=[\mathbf{u}_1,\mathbf{u}_2,\cdots,\mathbf{u}_n]$ ，$\mathbf{\Lambda}=diag\{\lambda_1,\lambda_2,\cdots,\lambda_n\}$ 。

上述分解称为**矩阵A的特征值分解**。

2. 奇异值分解

对于矩阵**A**，λ_i 为 $\mathbf{A}^H\mathbf{A}$ 的特征值，则称 $\sigma_i=\lambda_i^{1/2}$ 为A的**奇异值**。

若 $\sigma_1\ge\sigma_2\ge\cdots\ge\sigma_r>0$ 为矩阵A的奇异值，则存在酉阵U和V，使得 $\mathbf{A}=\mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$ ，式 $\Sigma=\begin{bmatrix}\Sigma_1 & \mathbf{O}\\ \mathbf{O} & \mathbf{O}\end{bmatrix}$ ，且 $\Sigma_1=diag\{\delta_1,\delta_2,\cdots,\delta_r\}$ ，$|\delta_i|=\sigma_i$ 。上述分解称为**矩阵A的奇异值分解**。

3. Cholesky分解

方阵 A 是正定矩阵，$\mathbf{A}=\mathbf{G}\mathbf{G}^H$ 称为矩阵 A 的**Cholesky分解**，其中，G是一个具有正的对角线元素的下三角矩阵，即

$$\mathbf{G}=\begin{bmatrix} g_{11} & & & 0 \\ g_{21} & g_{22} & & \\ \vdots & \vdots & \ddots & \\ g_{n1} & g_{n2} & \cdots & g_{nn} \end{bmatrix}$$

若A是正定矩阵，则其Cholesky分解是**唯一的**。

4. LU分解

$\mathbf{A}\in\mathbb{C}^{m\times n}$，则 $\mathbf{A}=\mathbf{LU}$ 称为**矩阵的LU分解**。其中，L为 $m\times m$ 单位下三角矩阵(对角线为1的下三角矩阵)，U是A的 $m\times n$ 上阶梯型矩阵。

如果 $\mathbf{A}\in\mathbb{C}^{n\times n}$ 非奇异，并且其LU分解存在的话，则A的LU分解是唯一的，且 $\det\{\mathbf{A}\}=u_{11}u_{22}\cdots u_{nn}$ 。

5. QR分解

$\mathbf{A}\in\mathbb{C}^{m\times n}$，且 $m\geq n$，则存在列正交阵 $\mathbf{Q}\in\mathbb{C}^{m\times n}$ 和上三角矩阵 $\mathbf{R}\in\mathbb{C}^{n\times n}$，使得 $\mathbf{A}=\mathbf{QR}$ 。

当 $m=n$ 时，**Q是正交矩阵**。若**A是非奇异的** $n\times n$ 矩阵，则R的所有对角线元素均为正，并且在这种情况下，**R和Q二者是唯一的**。

1.3.4 子空间

n维复矢量空间$\mathbb{C}^n$是所有n维复矢量的集合。令 $m<n$，则m个n维复矢量的子集合便构成 $\mathbb{C}^n$ 内的一个**矢量子空间**。若 $S=\{\mathbf{u}_1,\mathbf{u}_2,\cdots,\mathbf{u}_m\}$ 是矢量空间V 的矢量子集合，则 $\mathbf{u}_1,\mathbf{u}_2,\cdots,\mathbf{u}_m$的所有线性组合的集合$W$称为**由 $\mathbf{u}_1,\mathbf{u}_2,\cdots,\mathbf{u}_m$ 张成的子空间**，定义为

$$\begin{aligned} W &= Span\{\mathbf{u}_1,\mathbf{u}_2,\cdots,\mathbf{u}_m\} \\ &= \{\mathbf{u}:\mathbf{u}=a_1\mathbf{u}_1+a_2\mathbf{u}_2+\cdots+a_m\mathbf{u}_m\} \end{aligned}$$

只包含了一个零矢量的矢量子空间称为**平凡子空间**。

子空间W的任何一组基的矢量个数称为**W的维数**，用符号 $\dim(W)$ 表示。

若W 的任何一组基都不是由有限个线性无关的矢量组成时，则称W是**无限维矢量子空间**。

若某一向量与子空间W的所有向量都正交，则称该向量正交于子空间W。若 $\forall\mathbf{a}_i\in W_i,\ \mathbf{a}_j\in W_j(i\neq j)$，恒有 $\mathbf{a}_i\perp\mathbf{a}_j$，则称子空间$W_i,W_j$为**正交子空间**，记作 $W_i\perp W_j,(i\neq j)$ 。

特别地，与子空间W 正交的所有向量的集合组成的向量子空间，称为**W的正交补空间**，记作

$$W^{\perp}=\{\mathbf{x}\,|\,\mathbf{x}^T\mathbf{y}=0,\forall\mathbf{y}\in W\}$$

1.3.6 梯度分析

$n\times 1$ **矢量x的梯度算子**记作

$$\nabla_{\mathbf{x}}=\left[\frac{\partial}{\partial x_1},\frac{\partial}{\partial x_2},\cdots,\frac{\partial}{\partial x_n}\right]^T=\frac{\partial}{\partial\mathbf{x}}$$

以 $n\times 1$ 实矢量x为变元的实标量函数 $f(\mathbf{x})$ 相对于x的梯度为

$$\nabla_{\mathbf{x}}f(\mathbf{x})=\left[\frac{\partial f(x)}{\partial x_1},\frac{\partial f(x)}{\partial x_2},\cdots,\frac{\partial f(x)}{\partial x_n}\right]^T=\frac{\partial f(\mathbf{x})}{\partial\mathbf{x}}$$

可见：

- 一个以矢量为变元的标量函数的梯度为一矢量。
- 梯度的每个分量给出了标量函数在该分量方向上的变化率。

类似地，**实值函数** $f(\mathbf{x})$ 相对于 $1\times n$ 行矢量 $\mathbf{x}^T$ 的梯度

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}^T}=\left[\frac{\partial f(\mathbf{x})}{\partial x_1},\frac{\partial f(\mathbf{x})}{\partial x_2},\cdots,\frac{\partial f(\mathbf{x})}{\partial x_n}\right]=\nabla_{\mathbf{x}^T} f(\mathbf{x})$$

$1\times m$ 行矢量函数 $\mathbf{f}(\mathbf{x})=\left[f_1(\mathbf{x}),f_2(\mathbf{x}),\cdots,f_m(\mathbf{x})\right]$相对于 $n\times 1$ 实矢量 $\mathbf{x}$的梯度为

$$\frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}}=\begin{bmatrix}\frac{\partial f_1(\mathbf{x})}{\partial x_1} & \frac{\partial f_2(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial f_m(\mathbf{x})}{\partial x_1}\\ \frac{\partial f_1(\mathbf{x})}{\partial x_2} & \frac{\partial f_2(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial f_m(\mathbf{x})}{\partial x_2}\\ \vdots & \vdots & & \vdots\\ \frac{\partial f_1(\mathbf{x})}{\partial x_n} & \frac{\partial f_2(\mathbf{x})}{\partial x_n} & \cdots & \frac{\partial f_m(\mathbf{x})}{\partial x_n}\end{bmatrix}=\nabla_{\mathbf{x}}\mathbf{f}(\mathbf{x})$$

复矢量函数的梯度

引入复数求导的定义，若 $f(w)$ 是复数 w 的函数，其中 $w=w_r+jw_i$，则

$$\frac{df(w)}{dw}=\frac{1}{2}\left(\frac{\partial f(w)}{\partial w_r}-j\frac{\partial f(w)}{\partial w_i}\right)$$

$$\frac{df(w)}{dw^*}=\frac{1}{2}\left(\frac{\partial f(w)}{\partial w_r}+j\frac{\partial f(w)}{\partial w_i}\right)$$

那么，目标函数相对于复矢量的梯度定义为

$$\frac{\partial f(\mathbf{w})}{\partial \mathbf{w}}=\nabla_{\mathbf{w}} f(\mathbf{w})=\left[\frac{\partial f(\mathbf{w})}{\partial w_1},\frac{\partial f(\mathbf{w})}{\partial w_2},\cdots,\frac{\partial f(\mathbf{w})}{\partial w_n}\right]^T$$

共轭梯度（矢量）

$$\frac{\partial f(\mathbf{w})}{\partial \mathbf{w}^*}=\nabla_{\mathbf{w}^*} f(\mathbf{w})=\left[\frac{\partial f(\mathbf{w})}{\partial w_1^*},\frac{\partial f(\mathbf{w})}{\partial w_2^*},\cdots,\frac{\partial f(\mathbf{w})}{\partial w_n^*}\right]^T$$

若复行矢量函数 $\mathbf{f}(\mathbf{w})=\left[f_1(\mathbf{w}),f_2(\mathbf{w}),\cdots,f_m(\mathbf{w})\right]$，则其相对于复列矢量的梯度为

$$\frac{\partial \mathbf{f}(\mathbf{w})}{\partial \mathbf{w}}=\begin{bmatrix}\frac{\partial f_1(\mathbf{w})}{\partial w_1} & \frac{\partial f_2(\mathbf{w})}{\partial w_1} & \cdots & \frac{\partial f_m(\mathbf{w})}{\partial w_1}\\ \frac{\partial f_1(\mathbf{w})}{\partial w_2} & \frac{\partial f_2(\mathbf{w})}{\partial w_2} & \cdots & \frac{\partial f_m(\mathbf{w})}{\partial w_2}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{\partial f_1(\mathbf{w})}{\partial w_n} & \frac{\partial f_2(\mathbf{w})}{\partial w_n} & \cdots & \frac{\partial f_m(\mathbf{w})}{\partial w_n}\end{bmatrix}$$

类似地，行矢量函数相对于复共轭列矢量的梯度称为共轭梯度矩阵，定义为

$$\frac{\partial \mathbf{f}(\mathbf{w})}{\partial \mathbf{w}^*}=\begin{bmatrix}\frac{\partial f_1(\mathbf{w})}{\partial w_1^*} & \frac{\partial f_2(\mathbf{w})}{\partial w_1^*} & \cdots & \frac{\partial f_m(\mathbf{w})}{\partial w_1^*}\\ \frac{\partial f_1(\mathbf{w})}{\partial w_2^*} & \frac{\partial f_2(\mathbf{w})}{\partial w_2^*} & \cdots & \frac{\partial f_m(\mathbf{w})}{\partial w_2^*}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{\partial f_1(\mathbf{w})}{\partial w_n^*} & \frac{\partial f_2(\mathbf{w})}{\partial w_n^*} & \cdots & \frac{\partial f_m(\mathbf{w})}{\partial w_n^*}\end{bmatrix}$$

结论：

任何一个复矢量 $\mathbf{w}$ 和它的共轭矢量 $\mathbf{w}^*$ 都可以当作两个独立的复变元处理，即在求梯度的过程中，复矢量相对于其共轭矢量可视为一常数；反之，$\mathbf{w}^*$相对于 $\mathbf{w}$ 也可视为一常数。

本章内容小结

1 概率论概要

2 随机过程基础

3 线性代数导论