

5

随机信号统计特征的实验研究方法

5.1 统计特征实验研究的基础

1. 随机信号的采样定理

如果平稳随机信号 $X(t)$ 的功率谱 $G_X(\omega)$ 满足： $G_X(\omega) = 0, |\omega| \geq \Omega_C$ 则称 $X(t)$ 为“低通带限”随机信号，式中 Ω_C 表示功率谱的最高截止频率。若以采样间隔 T 对“低通带限”的平稳随机信号 $X(t)$ 进行采样，采样后获得随机序列 $\{X(n)\}$ 。只要采样频率 f_s 满足

$$\omega_s = 2\pi f_s \geq 2\Omega_C \quad \text{或} \quad T_s \leq \frac{1}{2f_c} = \frac{\pi}{\Omega_c} \quad (5.1)$$

采样插值为

$$\hat{X}(t) = \sum_{n=-\infty}^{\infty} X(nT_s) \frac{\sin \Omega_c (t - nT_s)}{\Omega_c (t - nT_s)} \quad (5.2)$$

则可以证明

$$E [|X(t) - \hat{X}(t)|^2] = 0. \quad (5.3)$$

表明 $\hat{X}(t)$ 在均方意义上等于 $X(t)$ 。

2. 随机信号估计的准则

统计实验研究的理论是建立在假设随机信号具有各态历经性基础上的。在这个假设的前提下, 方可根据一个样本 (实现) 来“估计”出整个随机信号的有关统计参数。即根

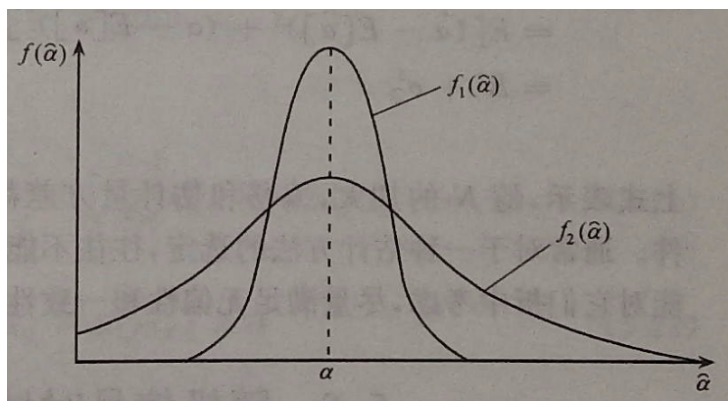


图 5-1 截取函数

据对随机信号 $X(t)$ 采样得到的随机序列 $\{X(n)\}$ 的一组样本 $x_0, x_1, x_2, \dots, x_{N-1}$, 来估计 $X(t)$ 的统计参数。

一般来说参数的点估计, 就是根据观测到的数据 $x_0, x_1, x_2, \dots, x_{N-1}$, 构造一个用来估计未知参数 α 的估计量 $\hat{\alpha}_a$

$$\hat{\alpha} = h(x_0, x_1, x_2, \dots, x_{N-1}). \quad (5.4)$$

由于观测到的数据 $x_0, x_1, x_2, \dots, x_{N-1}$ 是随机的, 因此由观测数据构造的估计量 $\hat{\alpha}$ 也是随机的, 但被估计的参数 α 是确定量, 因此构造的估计量 $\hat{\alpha}$ 与真实参数 α 之间存在随机的估计误差, 这个误差的大小则是判定估计方法好坏的标准。

估计误差用 $\tilde{\alpha}$ 表示

$$\tilde{\alpha} = \hat{\alpha} - \alpha \quad (5.5)$$

$\tilde{\alpha}$ 和 $\hat{\alpha}$ 都是随机变量, 而随机变量则存在一定的统计分布规律。设 $\hat{\alpha}$ 的概率密度曲线如图 5-1 所示, 图中 α 是要估计的参数。若估计值 $\hat{\alpha}$ 接近 α 的概率很大, 则说明这是一种比较好的估计方法, 如图中 $f_1(\hat{\alpha})$ 比 $f_2(\hat{\alpha})$ 要好。一般来说, 一个好的估计值 $\hat{\alpha}$ 的概率密度函数分布应比较集中在其估计值的真值 α 附近, 概率密度曲线必须窄。

通常, 评价估计性能好坏的标准有以下三种:

1) 偏移性。令估计量的统计平均值与真值之间的差值为偏移 B , 其公式为

$$B = \alpha - E[\hat{\alpha}]. \quad (5.6)$$

如果 $B = 0$, 称为无偏估计。无偏估计表示估计量仅在它真值附近摆动, 这是我们希望有的估计特性。如果 $B \neq 0$ 则称为有偏估计。如果随着观测次数 N 的加大, 能满足

$$\lim_{N \rightarrow \infty} E[\hat{\alpha}] = \alpha \quad (5.7)$$

则称为渐近无偏估计。

2) 估计量的方差。如果两个估计量的观察次数相同，又都是无偏估计，哪一个估计量在真值附近摆动更小一些，即估计量的方差更小一些，就说这一个估计量的估计更有效。比如 $\hat{\alpha}$ 和 $\hat{\alpha}'$ 是 X 的两个无偏估计量

$$\sigma_{\hat{\alpha}}^2 = E[(\hat{\alpha} - E[\hat{\alpha}])^2] \quad \sigma_{\hat{\alpha}'}^2 = E[(\hat{\alpha}' - E[\hat{\alpha}'])^2]. \quad (5.8)$$

若对任意 N ，它们的方差满足

$$\sigma_{\hat{\alpha}}^2 < \sigma_{\hat{\alpha}'}^2 \quad (5.9)$$

则称 $\hat{\alpha}$ 比 $\hat{\alpha}'$ 更有效。一般希望当 $N \rightarrow \infty$ 时， $\sigma_{\hat{\alpha}}^2 \rightarrow 0$ 。

3) 一致性——均方误差。

在许多情况下，比较两个有偏估计值是麻烦的。偏移较小的估计值，可能有较大的方差，而方差较小的估计值可能有较大的偏移，此时使用与估计值有关的均方误差会更方便。估计量的均方误差用下式表示

$$E[\tilde{\alpha}^2] = E[(\hat{\alpha} - \alpha)^2]. \quad (5.10)$$

如果估计量的均方误差随着观察次数的增加趋于 0，即估计量随着 N 的加大，在均方意义上趋于它的真值，则称该估计是一致估计。

估计量的均方误差、估计量的方差和偏移的关系推导如下

$$\begin{aligned} E[\tilde{\alpha}^2] &= E[(\hat{\alpha} - \alpha)^2] = E\{[(\hat{\alpha} - E[\hat{\alpha}]) - (\alpha - E[\hat{\alpha}])]^2\} \\ &= E[(\hat{\alpha} - E[\hat{\alpha}])^2 + (\alpha - E[\hat{\alpha}])^2 - 2E[(\hat{\alpha} - E[\hat{\alpha}]) \cdot (\alpha - E[\hat{\alpha}])] \\ &= B^2 + \sigma_{\hat{\alpha}}^2. \end{aligned} \quad (5.11)$$

尽量满足无偏性和一致性。

5.2 随机信号时域特征的估计

1) 最大似然估计

由于各态历经性的假设，实验测得的一组样本 $x_0, x_1, x_2, \dots, x_{N-1}$ 可看作是随机序列 $\{X(n)\}$ 中的一组随机变量。定义在参量 α 条件下 $\{X(n)\}$ 的多维条件概率密度 $f(x_0, x_1, \dots, x_{N-1}/\alpha)$ 为“似然函数”。若估计量 $\hat{\alpha}$ 能使似然函数 $f(x_0, x_1, \dots, x_{N-1}/\alpha)$ 在 $\alpha = \hat{\alpha}$ 时

$$f(x_0, x_1, \dots, x_{N-1}/\hat{\alpha}) = \max \{f(x_0, x_1, \dots, x_{N-1}/\alpha)\}, \quad (5.12)$$

则估计量 \hat{a} 被称为最大似然估计量。

2) 均值的估计

(1) 均值的最大似然估计

若对高斯信号 $X(t)$ 的采样是独立的, 则 $\{X(n)\}$ 就是独立高斯随机序列, 实验测得的一组样本 $x_0, x_1, x_2, \dots, x_{N-1}$ 也可看作是相互独立的高斯随机变量。那么可得均值的最大似然估计量为

$$\hat{m}_X = \frac{1}{N} \sum_{i=0}^{N-1} x_i. \quad (5.13)$$

(2) 估计量的评价

1° \hat{m}_X 是无偏估计量。因为

$$E[\hat{m}_X] = E\left[\frac{1}{N} \sum_{i=0}^{N-1} x_i\right] = \frac{1}{N} \sum_{i=0}^{N-1} E[x_i] = E[x_i] = m_X. \quad (5.14)$$

所以

$$B = m_X - E[\hat{m}_X] = 0, \quad (5.15)$$

这个估计量是无偏估计量。

2° \hat{m}_X 是一致估计量 (假设样本数据 $(x_0, x_1, \dots, x_{N-1})$ 之间不存在相关性)。由

$$\begin{aligned} E[\hat{m}_X^2] &= \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} E[x_i x_j] \\ &= \frac{1}{N^2} \left[\sum_{i=0}^{N-1} \sum_{j=0, j \neq i}^{N-1} E[x_i x_j] + \sum_{i=0}^{N-1} \sum_{j=0, j=i}^{N-1} E[x_i x_j] \right] \\ &= \frac{1}{N^2} \left[\sum_{i=0}^{N-1} E[x_i^2] + \sum_{i=0}^{N-1} \sum_{j=0, j \neq i}^{N-1} E[x_i] E[x_j] \right] \\ &= \frac{1}{N^2} [NE[x_i^2] + N(N-1)E[x_i] E[x_i]] \\ &= \frac{1}{N} E[x_i^2] + \frac{N-1}{N} m_X^2. \end{aligned} \quad (5.16)$$

可得估计的方差为

$$\begin{aligned} \sigma_{m_X}^2 &= E[(\hat{m}_X - E[\hat{m}_X])^2] = E[\hat{m}_X^2] - E^2[\hat{m}_X] \\ &= E[\hat{m}_X^2] - m_X^2 = \frac{1}{N} E[x_i^2] - \frac{1}{N} m_X^2 = \frac{1}{N} \sigma_X^2. \end{aligned} \quad (5.17)$$

则其极限

$$\lim_{N \rightarrow \infty} \sigma_{m_X}^2 = \lim_{N \rightarrow \infty} \frac{1}{N} \sigma_X^2 = 0. \quad (5.18)$$

估计量的均方误差

$$E [\tilde{m}_X^2] = E [(\hat{m}_X - m_X)^2] = B^2 + \sigma_{m_X}^2 = 0 + \sigma_{m_X}^2. \quad (5.19)$$

则其极限

$$\lim_{N \rightarrow \infty} E [\tilde{m}_X^2] = \lim_{N \rightarrow \infty} \sigma_{m_X}^2 = 0. \quad (5.20)$$

所以该估计量也是一致估计量。

结论: 当样本数据内部不相关时, 按照式 (5.13) 估计均值是一种无偏的一致估计, 是一种好的估计方法。但如果数据内部存在关联性, 会使一致性的效果下降, 估计量的方差比数据内部不存在相关情况的方差要大, 达不到信号方差的 $1/N$ 。

3) 方差的估计

(1) 方差的最大似然估计

假设采样得的 $\{X(n)\}$ 为独立的高斯随机序列, 用实验测得的一组样本数据 x_0, x_1, \dots, x_{N-1} 来估计方差。解出方差的最大似然估计量为

1° 若均值 m_X 已知

$$\hat{\sigma}_X^2 = \frac{1}{N} \sum_{i=0}^{N-1} (x_i - m_X)^2. \quad (5.21)$$

2° 若均值 m_X 未知

$$\hat{\sigma}_X^2 = \frac{1}{N} \sum_{i=0}^{N-1} (x_i - \hat{m}_X)^2. \quad (5.22)$$

(2) 估计值的评价 (假设样本数据 $(x_0, x_1, \dots, x_{N-1})$ 之间不存在相关性)

1° $\hat{\sigma}_X^2$ 是有偏估计量。

$$\begin{aligned} E [\hat{\sigma}_X^2] &= \frac{1}{N} \sum_{i=0}^{N-1} \langle E [x_i^2] + E [\hat{m}_X^2] - 2E [x_i \hat{m}_X] \rangle \\ &= \frac{1}{N} \sum_{i=0}^{N-1} E [x_i^2] + \frac{1}{N} \sum_{i=0}^{N-1} E [\hat{m}_X^2] - \frac{2}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} E [x_i x_j] \\ &= E [x_i^2] + E [\hat{m}_X^2] - 2E [\hat{m}_X^2]. \end{aligned} \quad (5.23)$$

代入式 (5.16) 结论可得

$$E [\hat{\sigma}_X^2] = \frac{N-1}{N} [E [x_i^2] - m_X^2] = \frac{N-1}{N} \sigma_X^2. \quad (5.24)$$

即 $\hat{\sigma}_X^2$ 为有偏估计量。但当 $N \rightarrow \infty$ 时, $\hat{\sigma}_X^2$ 为渐进无偏的。

另外, 为了得到无偏估计, 可以用下式计算:

$$\hat{\sigma}_X'^2 = \frac{1}{N-1} \sum_{i=0}^{N-1} (x_i - \hat{m}_X)^2. \quad (5.25)$$

$\hat{\sigma}_X^2$ 和 $\hat{\sigma}_X'^2$ 之间的关系为

$$\hat{\sigma}_X'^2 = \frac{N}{N-1} \hat{\sigma}_X^2. \quad (5.26)$$

$$E [\hat{\sigma}_X'] = \sigma_X^2. \quad (5.27)$$

所以, σ_X^2 是个无偏估计量。

2° $\hat{\sigma}_X^2$ 是个一致估计量。由 $E [\tilde{\alpha}^2] = E [(\hat{\alpha} - \alpha)^2] = B + \sigma_{\hat{\alpha}}^2$ 知, 估计距 $\hat{\sigma}_X^2$ 的均方误差为

$$E [(\tilde{\sigma}_X^2)^2] = E [(\hat{\sigma}_X^2 - \sigma_X^2)^2] = B + D [\hat{\sigma}_X^2]. \quad (5.28)$$

其中

$$D [\hat{\sigma}_X^2] = E [(\hat{\sigma}_X^2)^2] - (E [\hat{\sigma}_X^2])^2. \quad (5.29)$$

令

$$\hat{\sigma}_X^2 = \frac{1}{N} \sum_{i=0}^{N-1} (x_i - \hat{m}_X)^2 = \frac{1}{N} \sum_{i=0}^{N-1} (y_i)^2. \quad (5.30)$$

有

$$\begin{aligned} E [(\hat{\sigma}_X^2)^2] &= \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (E [y_i^2 y_j^2]) \\ &= \frac{1}{N^2} \{ N E [y_i^4] + N(N-1) (E [y_i^2])^2 \} \\ &= \frac{1}{N} \{ E [y_i^4] + (N-1) (E [y_i^2])^2 \} \\ E [\hat{\sigma}_X^2] &= E \left[\frac{1}{N} \sum_{i=0}^{N-1} y_i^2 \right] = E [y_i^2]. \end{aligned} \quad (5.31)$$

则

$$D[\hat{\sigma}_X^2] = E[(\hat{\sigma}_X^2)^2] - (E[\hat{\sigma}_X^2])^2 = \frac{1}{N} \{E[y_i^4] - (E[y_i^2])^2\}. \quad (5.32)$$

估计量 $\hat{\sigma}_X^2$ 均方误差为

$$E[(\hat{\sigma}_X^2)^2] = E[(\hat{\sigma}_X^2 - \sigma_X^2)^2] = B + \frac{1}{N} \{E[y_i^4] - (E[y_i^2])^2\}. \quad (5.33)$$

因为其极限

$$\lim_{N \rightarrow \infty} E[(\hat{\sigma}_X^2)^2] = \lim_{N \rightarrow \infty} B + \lim_{N \rightarrow \infty} \frac{1}{N} \{E[y_i^4] - (E[y_i^2])^2\} = 0. \quad (5.34)$$

所以 $\hat{\sigma}_X^2$ 是个一致估计量。以上由关于 $\{X(n)\}$ 为高斯分布的假设, 导出了均值和方差的最大似然估计量。当未知 $\{X(n)\}$ 的密度函数形式, 或不是高斯分布时, 也常用式 (5.13) 和式 (5.21) 的估计方法。此时均值估计量仍为无偏一致估计量, 而方差估计量仍为渐近无偏一致估计量。但对有限样本来讲, 它们不再是最大似然估计, 从而不能保证是最佳的了。

4) 自相关函数的估计

设对零均值平稳过程 $X(t)$ 采样得随机序列 $\{X(n)\}(n = 0, \pm 1, \pm 2, \dots, \pm \infty)$, 其自相关函数为

$$R_X(m) = E[X(n)X(n+m)]. \quad (5.35)$$

(1) 自相关函数估计量

由实验手段测得 $\{X(n)\}$ 的一组样本数据 $x_0, x_1, x_2, \dots, x_{N-1}$ 估计的方差 $\hat{\sigma}_X^2$, 也就是零滞后自相关函数 $R_X(0)$ 的估计量

$$\hat{R}_X(0) = \hat{\sigma}_X^2 = \frac{1}{N} \sum_{i=0}^{N-1} x_i^2 = \frac{x_0^2 + x_1^2 + \dots + x_{N-1}^2}{N}. \quad (5.36)$$

$$\hat{R}_X(m) = \begin{cases} \frac{1}{N} \sum_{i=0}^{N-|m|-1} x_i \cdot x_{i+m}, & m = 0, 1, \dots, N-1 \\ 0, & |m| \geq N \end{cases} \quad (5.37)$$

注意:

- (1) 当原过程 $X(t)$ 为独立高斯过程时, 该估计量 $\hat{R}_X(0)$ 为最大似然估计量。
- (2) 当 $\{X(n)\}$ 中各 X_i 独立但非高斯时, $\hat{R}_X(0)$ 为渐近无偏一致估计量。
- (3) 当 $\{X(n)\}$ 中各 X_i 之间不独立时, 往往也采用这样的估计量, 严格来说这时的估计量已不再是最佳的, 甚至可能不再是渐近无偏一致估计量了。

若对 $R_X(0)$ 的估计方法推广至非零滞后自相关函数 $R_X(m)$ 的估计

$$\begin{aligned}\hat{R}_X(m) &= \frac{x_0x_0 + |m| + x_1x_{1+|m|} + \cdots + x_{N-1} - |m|x_{N-1-|m|+|m|}}{N} \\ &= \frac{1}{N} \sum_{i=0}^{N-|m|-1} x_i x_{i+|m|}, \quad m = 0, 1, \cdots, N-1\end{aligned}\quad (5.38)$$

估计量 $\hat{R}_X(m)$ 求和的项数一般小于 N . 特别是当 m 很大时, 参加求和的乘积项将非常少。当 $m \geq N$ 时, $x_i x_{i+|m|} = 0$ 。因而, 令 $m \geq N$ 时的自相关函数为 0, 即

$$\hat{R}_X(m) = 0, \quad |m| \geq N. \quad (5.39)$$

因此, 自相关函数的估计量为

$$\hat{R}_X(m) = \begin{cases} \frac{1}{N} \sum_{i=0}^{N-|m|-1} x_i x_{i+|m|}, & m = 0, 1, \cdots, N-1 \\ 0, & |m| \geq N. \end{cases} \quad (5.40)$$

(2) 估计量的评价

1° $\hat{R}_X(m)$ 是渐近无偏估计。由

$$E[\hat{R}_X(m)] = \frac{1}{N} \sum_{i=0}^{N-|m|-1} E[x_i x_{i+|m|}] = \left(1 - \frac{|m|}{N}\right) R_X(m). \quad (5.41)$$

其极限为

$$\lim_{N \rightarrow \infty} E[\hat{R}_X(m)] = \lim_{N \rightarrow \infty} \left(1 - \frac{|m|}{N}\right) R_X(m) = R_X(m). \quad (5.42)$$

偏移

$$B(m) = R_X(m) - E[\hat{R}_X(m)] = \frac{|m|}{N} R_X(m). \quad (5.43)$$

其极限为

$$\lim_{N \rightarrow \infty} B(m) = \lim_{N \rightarrow \infty} \frac{|m|}{N} R_X(m) = 0, \quad (5.44)$$

所以 $\hat{R}_X(m)$ 是渐近无偏估计。偏移 $B(m) = \frac{|m|}{N} R_X(m)$ 是 m 的函数, 如图 5-2 所示。

2° $\hat{R}_X(m)$ 是一致估计量。即需要证明, 当 $N \rightarrow \infty$ 时, $\hat{R}_X(m)$ 的方差趋向 0. 一般情况下这一结论的证明是比较复杂的。当原过程 $X(t)$ 是高斯过程时, 可以导出下列方差的近似公式:

$$D[\hat{R}_X(m)] \approx \frac{2}{N} \sum_{i=0}^{N-|m|-1} \left(1 - \frac{|m|+i}{N}\right) [R_X^2(m) + R_X(i+m)R_X(i-m)]. \quad (5.45)$$

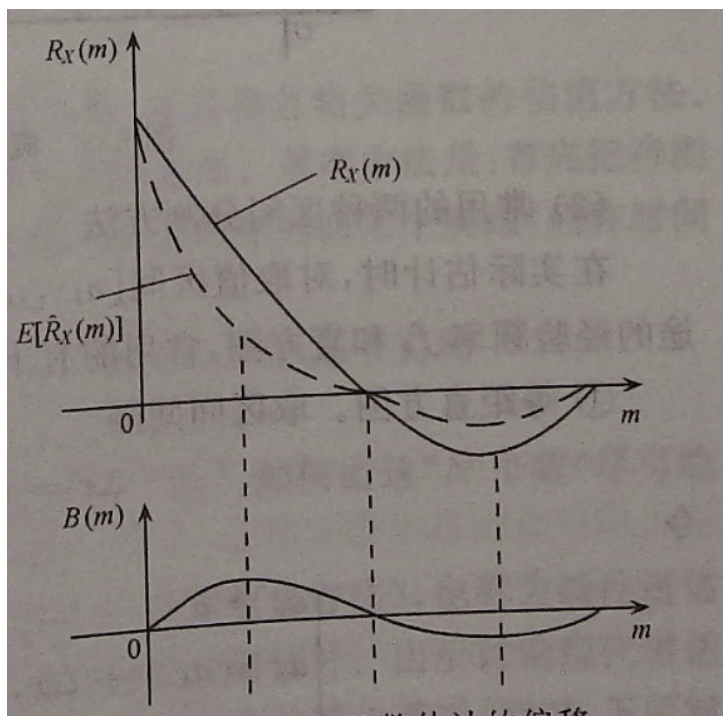


图 5-2 自相关函数估计的偏移

可以证明此式对一般的非高斯过程也是近似成立的。因为尽管 $N \rightarrow \infty$, 但是当 $i \rightarrow \pm\infty$ 时, $R_X(i) \rightarrow 0$ 。即 Σ 号内仅有有限项存在, 则

$$\lim_{N \rightarrow \infty} D[\hat{R}_X(m)] = 0. \quad (5.46)$$

5. 随机信号概率密度的估计

多维概率密度的估计十分复杂, 这里仅讨论一种最简单的一维概率密度的估计方法——“直方图法”。这种方法的合理性是建立在“分布函数具有各态历经性(略)”的基础上的。

(1) 直方图法

若实验测得随机过程 $X(t)$ 的一组样本数据 $x_0, x_1, x_2, \dots, x_{N-1}$, 由这组样本数据来计算 $X(t)$ 的“直方图”。

1) 首先求出其位置参量

极小值为

$$a = \min \{x_0, x_1, \dots, x_{N-1}\}. \quad (5.47)$$

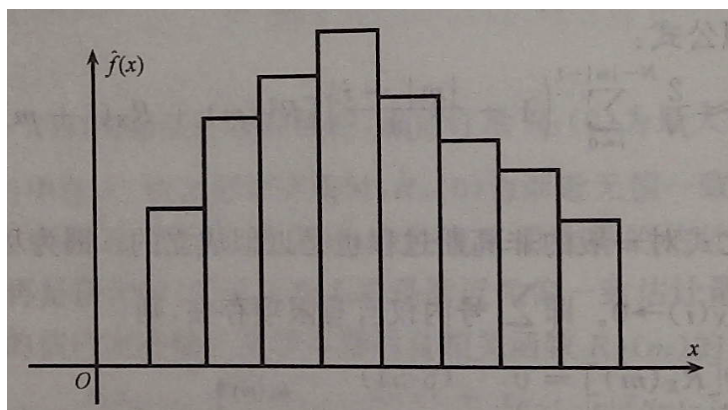


图 5-3 概率密度的直方图

极大值为

$$b = \max \{x_0, x_1, \dots, x_{N-1}\}. \quad (5.48)$$

(2) 分区间

将 x 的取值区间 $[a, b)$ 分成 K 个互不相交的分区间, 取分割点为 $a_0 < a_1 < \dots < a_k < \dots < a_K$ 且使 $a_0 \leq a, a_K > b$. 则子区间分别为 $[a_{k-1}, a_k)$ ($k = 1, 2, \dots, K$).

(3) 计算实验数据的经验频率

求出落入每个子区间 $[a_{k-1}, a_k)$ 的数据点的个数 N_k ($k = 1, 2, \dots, K$) 即计算 $a_{k-1} \leq x_n < a_k$, $n = 1, 2, \dots, K$ 的实验数据的点数。 N_k 称经验频数, 则经验频率为

$$f_k = \frac{N_k}{N}, \quad k = 1, 2, \dots, K. \quad (5.49)$$

显然, $0 \leq f_k \leq 1$, 且 $\sum_{k=1}^K f_k = 1$, 当 N 充分大时, f_k 可以近似表示随机变量 X 在区间 $[a_{k-1}, a_k)$ 上取值的概率。则概率密度在 $[a_{k-1}, a_k)$ 上的估值 $\hat{f}(x)$ 为

$$\hat{f}(x) = \frac{f_k}{a_k - a_{k-1}}, \quad k = 1, 2, \dots, K \quad x \in [a_{k-1}, a_k). \quad (5.50)$$

如图 5-3 所示。

5.3 随机信号功率谱密度的估计

前面, 我们已经讨论了有关随机序列的均值、方差和自相关函数的估值方法。本节将讨论概率谱密度的估计方法, 即功率谱的测量法。基本方法是: 首先把待测的随机过

程通过取样量化 (A/D 变换) 变换成一个随机序列的某个实现 (或称时间序列) 后, 再对它进行功率谱估值。

谱估值的基本问题是: 已知随机过程 $X(t)$ 或 $\{X(n)\}$ 的某个实现

$$\cdots, x_{-2}, x_{-1}, x_0, x_1, x_2, \cdots, x_{N-1}, \cdots \quad (5.51)$$

中的有限长序列段 $x_n (0 \leq n \leq N-1)$, 或者说 “ N 个数”, 如何由这 “ N 个数” 尽可能准确地得到 $X(t)$ 或 $\{X(n)\}$ 的功率谱密度 $G_X(\omega)$ 。

功率谱估计分为两大类: 一类是经典谱估计 (非参数化方法), 也称为线性谱估计; 另一类是现代谱估计 (参数化方法), 也称为非线性谱估计。由于讨论现代谱估计所需要具备的理论基础超出了学生学习本书时所具备的理论基础, 所以, 下面对常用的几种经典谱估计的方法作原理性介绍, 而对现代谱估计仅作一般性介绍。有兴趣的读者可以参看有关现代谱估计的著作。

5.3.1 经典谱估计

功率谱估计分为两大类:

- ① 一类是经典谱估计 (非参数化方法), 也称为线性谱估计;
- ② 另一类是现代谱估计 (参数化方法), 也称为非线性谱估计。

下面仅对常用的两种经典谱估计的方法做原理性介绍, 而对现代谱估计仅做一般性介绍。

5.3.2 经典谱估计

1. BT 法 (Blackman – Tukey)

(1) 随机序列的维纳-辛钦定理

由于随机序列 $\{X(n)\}$ 的自相关函数 $R_X(m) = E[X(n)X(n+m)]$ 定义在离散点 m 上, 设取样间隔为 T_s , 则可将随机序列的自相关函数可用连续时间函数表示为

$$R_X^D(\tau) = \sum_{m=-\infty}^{\infty} R_X(\tau) \cdot \delta(\tau - mT_s) = \sum_{m=-\infty}^{\infty} R_X(mT_s) \cdot \delta(\tau - mT_s). \quad (5.52)$$

等式两端取傅立叶变换, 则随机序列的功率谱密度

$$\begin{aligned}
 G_X(\omega) &= \int_{-\infty}^{\infty} R_X^D(\tau) e^{-j\omega\tau} d\tau \\
 &= \sum_{m=-\infty}^{\infty} \int_{-\infty}^{\infty} R_X(mT_s) \delta(\tau - mT_s) e^{-j\omega mT_s} d\tau \\
 &= \sum_{m=-\infty}^{\infty} R_X(mT_s) e^{-j\omega mT_s}.
 \end{aligned} \tag{5.53}$$

注意: $G_X(\omega)$ 为 ω 的连续函数, 而且在频率轴上以 $2\pi/T_s$ 周期重复。我们感兴趣的仅仅是它在所谓的奈奎斯特间隔 $(-\pi/T_s, \pi/T_s)$ 上的值。因此, 已知 $G_X(\omega)$ 而要求 $R_X(mT_s)$, 仅需在一个周期 $(-\pi/T_s, \pi/T_s)$ 内积分即可

$$R_X(mT_s) = \frac{T_s}{2\pi} \int_{-\pi/T_s}^{\pi/T_s} G_X(\omega) e^{j\omega mT_s} d\omega. \tag{5.54}$$

令 $T_s = 1$, 则上述两式可写成

$$\begin{aligned}
 G_X(\omega) &= \sum_{m=-\infty}^{\infty} R_X(m) \cdot e^{-j\omega m} \\
 R_X(m) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} G_X(\omega) e^{j\omega m} d\omega.
 \end{aligned} \tag{5.55}$$

以上两式为随机序列的维纳-辛钦定理。

(2) 谱估计

BT 法是先估计自相关函数 $\hat{R}_X(m)$ ($|m| = 0, 1, 2, \dots, N-1$), 然后再经过离散傅立叶变换求得功率谱密度的估值 $\hat{G}_X(\omega)$ 。即

$$\hat{G}_X(\omega) = \sum_{m=-N+1}^{N-1} \hat{R}_X(m) e^{-j\omega m}. \tag{5.56}$$

用 **BT 法**估计功率谱的流程为

$$\left\{ \frac{1}{N} \sum_{i=0}^{N-|m|-1} x_i \cdot x_{i+|m|} \right\} = \left\{ \hat{R}_X(m) \right\} \xrightarrow{FFT} \left\{ \hat{G}_X(\omega) \right\}. \tag{5.57}$$

其中 **FFT** 表示快速傅立叶变换。

2. 周期图法

周期图法是根据各态历经随机过程功率谱的定义来进行谱估计的。在第三章中已知, 各态历经的连续随机过程的功率谱密度满足

$$G_X(\omega) = G_i(\omega) = \lim_{T \rightarrow \infty} \frac{1}{2T} |\mathcal{X}_{iT}(\omega)|^2, \tag{5.58}$$

式中 $X_{iT}(\omega)$ 是连续随机过程第 i 个样本的截取函数 $x_{iT}(t)$ 的频谱。对应于随机序列中则有

$$G_X(\omega) = G_i(\omega) = \lim_{N \rightarrow \infty} \frac{1}{2N+1} |X_N(\omega)|^2. \quad (5.59)$$

由于随机序列中观测数据 x_n 仅在 $0 \leq n \leq N-1$ 的点上存在, 则 x_n 的 N 点离散傅立叶变换 (DFT) 为

$$X_N(\omega) = \sum_{n=0}^{N-1} x_n e^{-j\omega n}. \quad (5.60)$$

因此有随机信号的观测数据 x_n 的功率谱估计值称“周期图”如下

$$\hat{G}_X(\omega) = \frac{1}{N} |X_N(\omega)|^2 = \frac{1}{N} \left| \sum_{n=0}^{N-1} x_n e^{-j\omega n} \right|^2. \quad (5.61)$$

由于上式中的离散傅立叶变换 (DFT) 可以用快速傅立叶变换 (FFT) 计算, 因此, 用周期图法估计功率谱的流程如下

$$\{x_n\} \xrightarrow{FFT} \{X_N(\omega)\} \rightarrow \left\{ \frac{1}{N} |X_N(\omega)|^2 \right\} = \hat{G}_X(\omega). \quad (5.62)$$

无论采用哪种方法, 估计量 $\hat{G}_X(\omega)$ 的性质是一样的。可以证明它们均为有偏估值, 但是渐进无偏的。直接采用上面两个公式的估值方法, 最大的问题是这个估计量不是一致估计量, 即当 N 很大时, $\hat{G}_X(\omega)$ 的方差也不小。理论分析可以证明

$$\lim_{N \rightarrow \infty} D \left[\hat{G}_X(\omega) \right] \approx G_X^2(\omega). \quad (5.63)$$

即功率谱估值的方差近似等于功率谱真值的平方。真实谱越大的地方 (通尚是我们于这两点结论的证明超出了本书讨论的范围, 有兴趣的读者可以参看有关谱估计的经典著作。因而在实际使用时, 必须将周期图法的谱估值加以修正, 下西就介绍两种卓有成效的修正方法。

5.3.3 经典谱估计的改进

从上面的分析可知, 周期图法作为功率谱的估计不满足一致估计的条件, 必须进行改进。改进的措施主要是将周期图进行平滑 (平均属于平滑的一种主要方法), 使估计方差减小, 从而得到一致谱估计。

平滑的方法主要有两种:

一种是常用的窗口处理法“窗函数法选择适当的窗口函数作为加权函数,对自相关函数估计值进行加权平均;另一种是周期图“平均法”先将数据分段,对各段数据分别求其周期图,最后将所有段周期图的平均作为功率谱的估计。由于“平均法”采用快速傅立叶变换(FFT),运算速度快,所以是当前用得较多的一种平滑方法。Welch法是前两种方法的结合,因此是用得最多的一种方法。

1. 窗函数法

从前面的讨论得知,BT法实质是对自相关函数的估值求傅里叶变换的一种方法。由于作自相关函数估计时,滞后量 m 增大,参与求和的项就会应少,因而平均的效果削弱。当 $m = N - 1$ 时,求和项只有一项,也就是说, m 较大的那些自相关估值 $\hat{R}_X(m)$ 可靠性差,估计的方差会较大。由于BT法用了从 $m = 0 \sim (N - 1)$ 的全部自相关函数估值,则方差大的自相关函数估值造成了谱估计较大的方差。如果对此的影响,就有可能提高谱估计的性能。

为减少谱估计的方差,经常用窗函数 $\tilde{\omega}(m)$ 对自相关函数进行加权,此时谱估计公式为

$$\hat{G}_x(\omega) = \sum_{m=-M-1}^{M-1} \hat{R}_X(m) \tilde{\omega}(m) e^{-j\omega m}, \quad M \leq N. \quad (5.64)$$

式中窗函数

$$\tilde{\omega}(m) = \begin{cases} \tilde{\omega}(m), & -(M-1) \leq m \leq (M-1) \\ 0, & \text{其他} \end{cases} \quad (5.65)$$

由于功率谱是非负的,则要求加窗后的功率谱也是非负的,因此窗函数 $\tilde{\omega}(m)$ 的选择必须满足此原则,即它的傅里叶变换必须是非负的(此时, $\tilde{\omega}(m)$ 也应是一个偶序列)。例如三角(Bartlett)窗和Parzen窗是满足此条件的,而海宁(Hanning)窗和汉明(Hamming)窗就不满足此条件。其中应用较多的Bartlett窗为

$$\tilde{\omega}(m) = \begin{cases} 1 - \frac{|m|}{M}, & |m| \leq M, M < N \\ 0, & \text{其他} \end{cases} \quad (5.66)$$

2. 平均法

为了了解周期图经过平均后会使其的方差减小,达到一致估计的目的,我们先来看一个定理:如果 x_1, x_2, \dots, x_L 是不相关的随机变量,且每一个都有均值 μ 及方差 σ^2 ,则可以证明它们的算术平均 $\bar{x} \triangleq (x_1 + x_2 + \dots + x_L) / L$ 的均值等于 μ ,方差为 $D[\bar{x}] = \sigma^2 / L$ 。

由定理可见:具有 L 个独立同分布随机变量平均的方差,是单个随机变量方估计值经过算术平均后得到的估计量的方差也是原估计量的方差的 $1/L$ 。

平均周期图方法即是将数据 x_0, x_1, \dots, x_{N-1} 分段求周期图后再平均。比如,给定 $N = 1000$ 个数据样本(平均法适用于数据量大的场合),则可以将它分成10个长度为100

的小段, 分别计算每一小段的周期图

$$G_{100,l}(\omega) = \frac{1}{100} \left| \sum_{n=100(l-1)}^{100\alpha-1} x_n e^{-jm\omega} \right|^2, \quad l = 1, 2, \dots, 1C. \quad (5.67)$$

然后将这 10 个周期图加以平均, 得谱估值为

$$\widehat{G}_{100}^{10}(\omega) = \frac{1}{10} \sum_{i=1}^{10} G_{100,i}(\omega). \quad (5.68)$$

由于这 10 小段的周期图取决于同一过程, 因而其均值应该相同。若这 10 个小段的周期图是统计独立的, 则这 10 个小段平均之后的方差却是单段方差的 1/10。

$$D \left[\widehat{G}_{100}^{10}(\omega) \right] = \frac{1}{10} D \left[G_{100,i}(\omega) \right]. \quad (5.69)$$

即: 平均法将 x_0, x_1, \dots, x_{N-1} 的 N 个数据分成 L 段 ($N = LM$), 若各段数据相互立、则平均后估计量的方差是原来不分段估计量方差的 $1/L$ 。所以当 $L \rightarrow \infty$ 时, 计置的方差趋于零, 达到了一致估计的目的。但是, 随着分段数 L 的增加, M 点数减少, 虽然偏差减小, 但方差增大。况且此时各组数据之间因存在一定的相关性而非相坦独立, 也会使方差增大。所以, 在实际应用时必须兼顾分辨力与方差的要求, 来适当选择 L 的值。

3. Welch

Welch 法又称修正周期图法, 它将“窗函数法”与“平均法”结合使用, 其步骤为:

(1) 先将 N 个数据分成 L 段, 每段 M 个数据, $N = ML$ 。

(2) 选择适当的窗函数 $\tilde{\omega}(n)$, 并用该 $\tilde{\omega}(n)$ 依次对每段数据作相应的加权, 然后确定每一段的周期图

$$G_{N,l}(\omega) = \frac{1}{MU} \left| \sum_{n=M(l-1)}^{Ml-1} x_n \tilde{\omega}(n) e^{-ik\omega} \right|^2, \quad 1 \leq l \leq L. \quad (5.70)$$

这里 U 为归一化因子

$$U = \frac{1}{M} \sum_{n=0}^{M-1} \tilde{\omega}^2(n). \quad (5.71)$$

对分段周期图进行平均得到功率谱估计

$$\widehat{G}_M^L(\omega) = \frac{1}{L} \sum_{l=1}^L G_{M,i}(\omega). \quad (5.72)$$

由前面的分析, 当数据量一定时, 若分段数 L 增加, M 点数减少, 则分辨力下降; 若分段数 L 增减少, 虽 M 增加, 但方差增大。数据量一定时, 由于 $N = ML$, 不可能 L 与 M

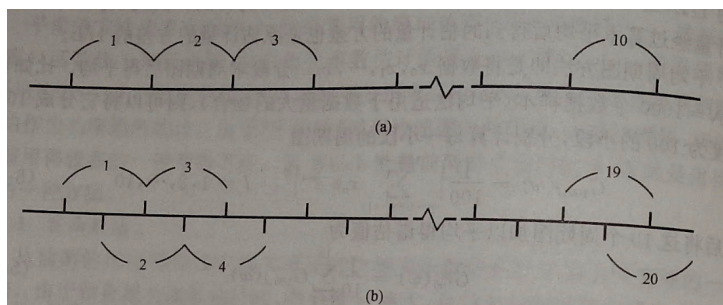


图 5-4 Welch 法

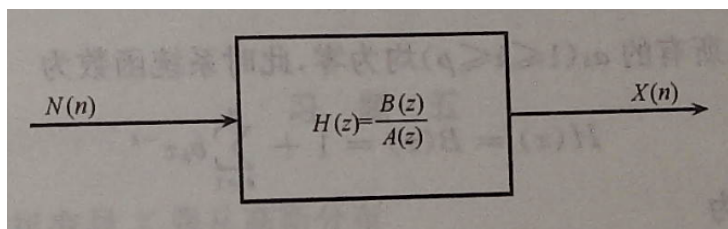
两个都大。解决这一矛盾的方法是, 让数据段之间适当“重造”, 如图 5-4(b) 所示。如果, 数据共有 1000 个点, 每段为 100 个点, 若重复 50%, 则可以分成 20 段。这样的分法, 由于段间相关性变大, 显然不能使方差减少到原来的 $1/20$, 但肯定小于 $1/10$, 而分辨力却没变, 这是 Welch 法的一个优点。Welch 法的另一个优点是, 对窗函数 $\tilde{w}(n)$ 没有要求, 因为在式 (5-83) 中可以看到, 无论什么样的窗函数都可以使谱估计非负。

注意: 段长 M 应满足频率分辨力的要求。若要求的频率分辨力是 $\Delta f(\text{Hz})$, 则归一化的频率分辨力要求 $\Delta f/f$ 。因此要求 $1/M \leq \Delta f/f_s$ 或 $M \geq f_s/\Delta f$ 。为了增加 FFT 后的谱线的密度, 可以在每段数据后适当补零, 一般是对 M 长的数据增添 m 个零 (或者使添零的结果让总数为 2 的方幂), 作 $2M$ 个点的 FFT。注意, 补零并不能提高频率分辨力, 只能增加功率谱的谱线密度, 克服栅栏效应。。

5.3.4 现代谱估计简介

经典谱估计方法可以利用 FFT 计算, 因而有计算效率高的优点, 在谱分辨力要求不是很高的地方常用这种方法。经典谱估计的一个主要缺点是频率分辨力低。这是由于 BT 法仅利用 N 个有限的观测数据作自相关函数估计, 也就隐含地建立了在已知数据之外的自相关函数均为零的假设, 或者周期图法在计算中把观测到的有限长的 N 个数据以外的数据认为是零。这显然与事实不符。把观测不到的或估计不到的值认为是零, 相当于在时域中乘上了一个矩形窗口函数, 这在频域里相当于引入了一个与之卷积的 sinc 函数。由于主瓣宽度反比于数据记录长度 N , 而在实际中一般又不可能获得很长的数据记录, 所以 sinc 函数主瓣不会很窄。如果原来真实的功率谱是很窄的, 那么与主瓣卷积后会使得功率向附近频域扩散, 使得信号模糊, 降低了分辨力。可见主瓣越宽分辨力越差, 严重时, 会使主瓣产生很大的失真, 甚至主瓣中的弱分量会被旁瓣中的强泄漏所掩盖。

为了克服以上缺点, 人们提出了平均、加窗平滑等方法, 在一定程度上改善了经典谱估计的性能。但是, 经典谱估计始终无法解决频率分辨力与谱估计稳定性之间的矛盾。

图 5-5 $X(n)$ 表示为白噪声通过线性系统的输出

特别是在数据记录很短的情况下, 这一矛盾显得尤为突出。

在第四章的讨论中我们知道, 任何具有有理功率谱密度的随机信号都可以看能克服经典谱估计的缺点。如由这个模型来求功率谱估计, 可望得到比较好的代统的输出, 如图 5-5 所示。其传递函数为

$$H(z) = \frac{B(z)}{A(z)} = \frac{\sum_{k=0}^q b_k z^{-k}}{\sum_{k=0}^p a_k z^{-k}} \quad (5.73)$$

当输入白噪声的功率谱密度 $G_N(\omega) = \sigma_N^2$ 时, 输出的功率谱密度为

$$G_X(\omega) = \sigma_N^2 |H(e^{j\omega})|^2 = \sigma_N^2 \left| \frac{B(e^{j\omega})}{A(e^{j\omega})} \right|^2. \quad (5.74)$$

如果能确定 σ_N^2 与 $a_k (0 \leq k \leq p)$, $b_k (0 \leq k \leq q)$ 值, 通过上式就可得到所需信号的功率谱密度。由于 $|H(e^{j\omega})|$ 的增益系数可并入 σ_N^2 进行考虑, 因此, 为不失一般性可假设 $a_0 = 1$, $b_0 = 1$ 。我们将这个线性系统分为三种情况。

1) AR 模型

除 $b_0 = 1$ 外所有的 $b_k (1 \leq k \leq q)$ 均为零, 此时系统函数为

$$H(z) = \frac{1}{A(z)} = \frac{1}{1 + \sum_{k=1}^p a_k z^{-k}}. \quad (5.75)$$

其差分方程为

$$X(n) = - \sum_{k=1}^p a_k X(n-k) + N(n). \quad (5.76)$$

其信号流图则如图 5-6 所示。

这种模型称为 p 阶自回归模型, 简称 AR (Autoregressive) 模型。其系统函数只有极点, 没有零点, 因此也称为全极点模型。模型输出功率谱为

$$G_X(\omega) = \frac{\sigma_N^2}{|A(e^{j\omega})|^2} = \frac{\sigma_N^2}{1 + \sum_{k=1}^p a_k e^{-j\omega k}}^2. \quad (5.77)$$

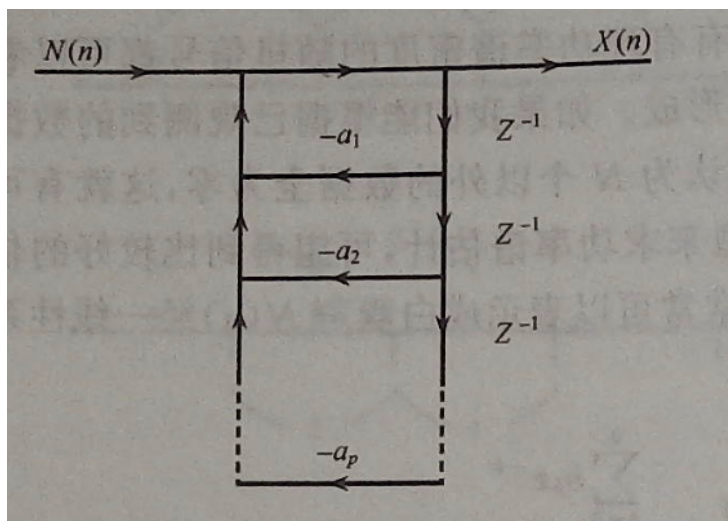


图 5-6 AR 模型

只要我们能求得 σ_N^2 及所有 a_k 值, 就可以求得随机信号 $X(n)$ 的功率谱。

2) MA 模型

除 $a_0 = 1$ 外所有的 $a_k (1 \leq k \leq p)$ 均为零, 此时系统函数为

$$H(z) = B(z) = 1 + \sum_{k=1}^q b_k z^{-k} \quad (5.78)$$

系统的差分方程为

$$X(n) = \sum_{k=0}^q b_k N(n-k). \quad (5.79)$$

其信号流图则如图 5-7 所示。

这种模型称为 q 阶滑动平均模型, 简称 MA (Moving Average) 模型。其系统函数只有零点, 没有极点, 因此也称为全零点模型。模型输出功率谱为

$$G_X(\omega) = \sigma_N^2 |B(e^{j\omega})|^2 = \sigma_N^2 \left| 1 + \sum_{k=1}^q b_k e^{-j\omega k} \right|^2. \quad (5.80)$$

只要我们能求得 σ_N^2 及所有 b_k 值, 就可以求得随机信号 $X(n)$ 的功率谱。

3) ARMA 模型

设 $a_0 = 1$, 和 $b_0 = 1$, 其余所有的 a_k 和 b_k 不全为零。此时模型的传递函数和输出功率谱分别由式 (5.73) 和式 (5.74) 表示。这是一个“极点、零点”模型, 称为 ARMA(p, q) 模型, 或直接简称 ARMA 模型。

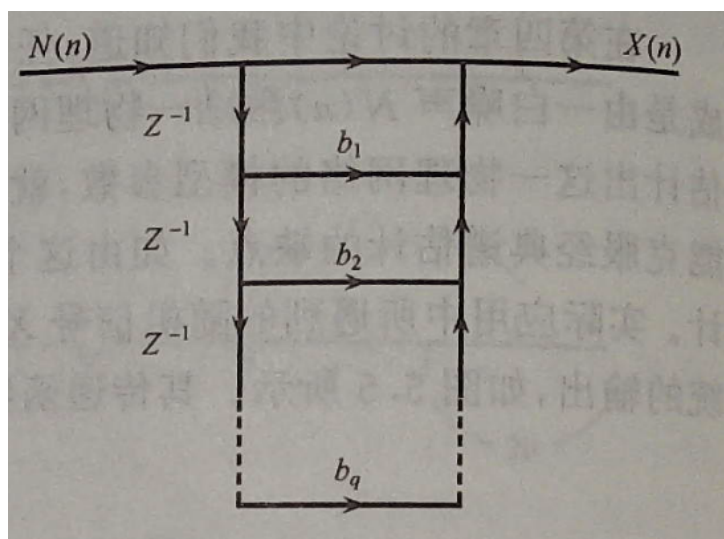


图 5-7 MA 模型

- (1) 选择一个合适的模型。
- (2) 用已观测到的数据估计模型参数。
- (3) 将模型参数代入功率谱的计算公式就, 得到功率谱估值。

由以上讨论可知, 用**模型法**作功率谱估计, 实际上要解决的是模型的参数估计问题, 所以这类谱估计方法又统称为参数化方法。

基于模型的功率谱估计方法首先要选择一个模型, 但在一般情况卜我们没有随机信号模型的先验知识. 若模型选择不当是否会对谱估计性能产生较大的影响? Word 分解定理告诉我们, 如果功率谱是连续的, 则任何 ARMA 过程或 AR 过程也可以用一个无限阶的 MA 过程表示。Kolmogorov 也提出, 任何 ARMA 或 MA 过程也可以用一个无限阶的 AR 过程表示, 所以如果选择了一个不合适的模型, 但只要模型的阶数足的高, 它仍然能够比较好地逼近被建模的随机过程。

在这三种参数模型中, AR 模型得到了普遍应用。其原因为

- (1) 任意 ARMA 或 MA 信号模型都可以用无限阶或阶数足够大的 AR 模型来表示。
- (2) AR 模型的参数计算是线性方程, 比较简便。同时适合表示很窄的频谱。
- (3) AR 模型在作谱估计时, 由于具有递推特性, 所以所需的数据较短, 很适合表示窄的功率谱:

而 MA 模型表示窄谱时, 一般需要数量很多的参数; ARMA 模型虽然所需的参数数量最少, 但参数估计的算法是非线性方程组, 其运算远比 AR 模型复杂。

5.4 习题五

练习 5.1 已知随机变量 X 服从高斯分布

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \quad (5.81)$$

求未知参数 μ 和 σ^2 的最大似然估计。

练习 5.2 已知平稳过程 $X(t)$ 的数学期望 μ 和方差 σ^2 存在, x_0, x_1, \dots, x_{N-1} 是过和的一个样本数据。证明:

- (1) $\hat{\mu} = \sum_{i=0}^{N-1} c_i x_i$ 是数学期望 μ 的无偏估计量, 其中 $\sum_{i=0}^{N-1} c_i = 1, c_i > 0$.
- (2) $\hat{\sigma}_X^2 = \frac{1}{N} \sum_{i=0}^{N-1} (x_i - m_X)^2$ 是 σ^2 的无偏估计量。
- (3) $\hat{\sigma}_X^2 = \frac{1}{N} \sum_{i=0}^{N-1} (x_i - \bar{m}_x)^2$ 不是 σ^2 的无偏估计量。

练习 5.3 已知 $\hat{\theta}$ 是参数 θ 的无偏估计, 且有 $D[\hat{\theta}] > 0$, 证明: $(\hat{\theta})^2$ 不是 θ^2 的无偏估计。

练习 5.4 证明: 用式 $\hat{R}_X(m) = \frac{1}{N} \sum_{n=0}^{N-|m|-1} x^*(n)x(n+m)$ 进行自相关函数的估计是渐进一致估计。

练习 5.5 用周期图法进行谱估计时

$$\hat{G}_X(\omega) = \frac{1}{N} \left| \sum_{n=0}^{N-1} x_n e^{-i\omega n} \right|^2. \quad (5.82)$$

说明为什么可用 FFT 进行计算。周期图法的谱分辨率较低, 且估计的方差也较大, 说明造成这两种缺点的原因, 并说明无论选取什么样的窗函数都难以从根本上解决问题的原因。

练习 5.6 说明白噪声过程的周期图是一个无偏估计器。

练习 5.7 采用下式给出的有偏自相关函数的定义, 并加窗 $\tilde{\omega}(k)$ 得到 BT 谱估计器

$$\begin{aligned} \hat{R}_X(k) &= \begin{cases} \frac{1}{N} \sum_{n=0}^{N-1} x^*(n)x(n+k), & k = 0, 1, \dots, N-1 \\ \hat{R}_X(-k), & k = -(N-1), -(N-2), \dots, -1 \end{cases} \\ \tilde{\omega}(k) &= \begin{cases} 1, & |k| \leq N-1 \\ 0, & \text{其他} \end{cases} \\ \bar{G}_X(\omega) &= \sum_{k=-N-1}^{N-1} \hat{R}_X(k) \tilde{\omega}(k) e^{-j\omega k} \end{aligned}$$

说明 BT 谱估计器与周期图相同。

5.5 知识点总结

- ▶ 估计的准则
 - 偏移性 (无偏估计、有偏估计、渐进无偏估计)
 - 估计量的方差
 - 均方误差 (一致估计)
- ▶ 时域信号的估计 (最大似然估计)
 - 均值的估计及其评价
 - 方差的估计及其评价
 - 自相关的估计及其评价
 - 概率密度的估计 (直方图法)
 - 频域特征的估计
 - 经典谱估计 (BT 法、周期图法、窗函数法、平均法、welch 法)