

3.1 信号参数估计的基本概念

信号处理的很多问题以参数估计的形式存在,例如在雷达及声呐信号处理中,需要估计目标(飞机或船只)的距离和到达角度,这些参数可以通过对发射信号和回波信号的处理获得。又如在现代的高保真录制系统中,需要精确保持录音基准电平,这一般通过电平自动控制系统实现,具体流程是周期性地估计该基准电平,并根据估计结果做出调整。

信号处理中的参数估计一般基于观测信号与待估计参数之间的依赖关系,在不同的问题模型中,这种依赖关系可能是确定性的或随机性的。例如在通信中的信道估计问题中,发送信号 $x(n)$ 经过冲激响应为 $h(n)$ 的信道输出为

$$y(n) = h(n) \otimes x(n) \quad (3-1)$$

如果系统采用基于训练序列的信道估计方法, $x(n)$ 为已知,这时信道输出 $y(n)$ 与信道参数 $h(n)$ 存在确定性依赖关系;如果系统信道估计不采用训练序列,而是基于信道输出的统计特性,则 $y(n)$ 与信道参数 $h(n)$ 存在随机性依赖关系,具体而言,就是 $y(n)$ 的概率分布函数与 $h(n)$ 存在依赖关系。随机性依赖关系与确定性依赖关系相比,更贴近实际环境。这首先是因为实际环境中观测信号的获取总是受到干扰,如环境干扰以及观测设备自身的热噪声干扰等。考虑噪声影响后的信道接收信号一般表示为

$$y(n) = h(n) \otimes x(n) + v(n) \quad (3-2)$$

其中, $v(n)$ 表示系统环境干扰或设备热噪声产生的加性干扰,一般情况下用零均值、独立同分布的高斯随机信号描述。这时 $y(n)$ 也体现为具有高斯分布特性的随机信号, $h(n)$ 对 $y(n)$ 的影响也体现为对其概率密度分布和统计特性的影响。

随机性依赖关系更具重要性的另一原因是实际系统中待估计参数经常具有时变性,例如雷达信号处理中目标距离就是不断变化的参数;移动通信中,发送端或接收端的移动会造成信道环境的改变。待估计参数的变化不是毫无规律,而是服从一定的概率密度分布,例如非视距移动信道的时变性服从瑞利分布。在式(3-1)中,如果 $h(n)$ 是具有一定概率结构的随机信号,则无论 $x(n)$ 是确定性信号还是随机信号, $h(n)$ 对 $y(n)$ 的影响都体现为概率密度分布和统计特性的影响。

下面以一个简单的例子说明参数估计的过程。

例 3-1 在电平自动控制系统的电平估计问题中,对某个电平值 θ 进行估计,由于环境干扰及测量设备的不完善,测量总会存在误差,测量误差可归结为噪声,因此,实际得到的测量值为

$$x = \theta + w \quad (3-3)$$

其中, w 一般服从零均值高斯分布,方差为 σ^2 。问题是,如何根据测量值 x 估计 θ 的值。

一种最简单的方法是将多个观测结果进行平均。首先定义一个时间窗口,认为实际电平值在这个时间窗口中是恒定的。在这个时间窗口中得到电平值的 N 个含噪测量结果记为 $x(0), x(1), \dots, x(N-1)$ 。这些测量结果中

$$x(n) = \theta + w(n), \quad n = 0, 1, \dots, N-1 \quad (3-4)$$

根据 $x(n)$ 得到 θ 的估计为

$$\hat{\theta} = \frac{x(0) + x(1) + \dots + x(N-1)}{N} \quad (3-5)$$

式(3-5)的一个直观效果是通过对噪声求均值,以降低噪声的影响。一般情况下,可以认为估计结果与观测数据由一个映射关系确定:

$$\hat{\theta} = f(x(0), x(1), \dots, x(N-1)) \quad (3-6)$$

上面提到,由于噪声的影响, $x(n)$ 更多时候体现为随机信号的特征,因此式(3-6)中 $x(0), x(1), \dots, x(N-1)$ 可看成随机变量。 $\hat{\theta}$ 与 $x(0), x(1), \dots, x(N-1)$ 存在映射关系,可认为 $\hat{\theta}$ 也是随机变量(不管 θ 是常量还是随机变量)。一次估计的结果可以看成 $\hat{\theta}$ 的一个样本,因此,验证估计算法的性能往往需基于多次估计结果进行评估。实际应用中往往很难得到类似于式(3-5)的闭式表达式,参数估计可能包含复杂的步骤,但包含的要素是一致的,从上面这个例子可以看出构造一个估计问题的基本要素如下:

(1) 待估计参数:指估计的目标参数,可以是单参量,也可以是多参量。待估计的参数可能是常量,也可能是随机变量,如果是随机变量一般要求其概率密度函数已知。

(2) 信号模型:指观测信号的产生模型,其中必须体现待估计参数与观测信号的依赖关系。观测信号可以是单参量,也可以是多参量。

(3) 误差准则:指估计算法的性能衡量指标,很多时候误差准则可指导估计算法的设计。误差准则的确定也有助于不同估计算法间的比较,特别是针对同一个估计问题的不同估计算法。

(4) 估计算法:对同一估计问题,估计算法不是唯一的,不同的估计算法可以得到不同的估计性能。如何得到误差小、计算量少的估计算法是信号处理的核心问题。

3.2 估计算法的性能指标

依据不同的估计准则,可以得到不同的估计量,这些估计量的性能需要有评价的指标。估计量是观测的函数,而观测的是随机变量(或向量),因此,估计量也是随机变量,随着观测数据的增多,希望估计量能逐渐逼近真值,因此,对估计量的均值和方差应有一定的要求。

3.2.1 性能指标

估计量的好坏可以从无偏性、有效性、一致性加以评价。

1. 无偏性

待估计量为常数 θ , 它的估计值为 $\hat{\theta}$, 如果 $E(\hat{\theta}) = \theta$, 则称 $\hat{\theta}$ 为 θ 的无偏估计, 否则称 $\hat{\theta}$ 为有偏估计。定义估计的偏差为

$$\tilde{\theta} = E(\hat{\theta}) - \theta \quad (3-7)$$

通常希望估计量的均值趋于被估计量的真值或被估计量的均值, 即估计应该是无偏的。如果估计 $\hat{\theta}$ 不是无偏估计, 但随着样本数目的增加, 其数学期望趋近于真的估计量, 即

$$\lim_{N \rightarrow \infty} [E(\hat{\theta}) - \theta] = 0 \quad (3-8)$$

则称估计 $\hat{\theta}$ 是渐近无偏估计。

当被估计量 θ 是随机变量时, 如果估计量的均值等于被估计量的均值, 即

$$E[\hat{\theta}] = E[\theta] \quad (3-9)$$

则称 $\hat{\theta}$ 为无偏估计。对于有偏估计, 如果观测数据越多, 估计的性能越好, 即

$$\lim_{N \rightarrow \infty} [E(\hat{\theta}) - E(\theta)] = 0 \quad (3-10)$$

则称 $\hat{\theta}$ 为随机变量的渐近无偏估计。

2. 有效性

估计量具有无偏性并不表明已经保证了估计的品质。当被估计量为未知常数时, 不仅希望估计量的均值等于真值, 而且希望估计量的取值集中在真值附近, 这一品质可以通过估计的方差来描述, 估计的方差为

$$\text{Var}(\hat{\theta}) = E\{[\hat{\theta} - E(\hat{\theta})]^2\} \quad (3-11)$$

对于无偏估计, 方差越小, 表明估计量的取值越集中, 估计的性能越好, 估计也越有效。

对于有偏估计, 估计的方差小并不能说明估计是好的, 因为如果估计有偏差, 方差小的估计仍然可能有较大的估计误差, 这时用均方误差加以描述更合适。估计的均方误差定义为

$$\text{Mse}(\hat{\theta}) = E\{|\hat{\theta} - \theta|^2\} \quad (3-12)$$

均方误差越小, 表明估计越有效。

3. 一致性

当用 N 个观测值估计参量时, 一般来说, 观测值越多, 估计越趋于真值, 如果

$$\lim_{N \rightarrow \infty} P\{|\theta - \hat{\theta}| < \epsilon\} = 1 \quad (3-13)$$

则称 $\hat{\theta}$ 为一致估计(式(3-13)中, ϵ 是任意小的正数)。

3.2.2 随机信号均值及自相关函数的估计

随机信号均值和自相关函数的估计是最常用的参数估计。这一节将讨论常用均值、自相关估计算法的无偏性、有效性和一致性, 这里只讨论实随机信号。

1. 均值的估计

对遍历平稳随机序列, 设 $x(0), x(1), \dots, x(N-1)$ 是观察到的 N 个样本, 则可利用式(3-14)进行均值的估计

$$\hat{m}_x = \frac{1}{N} \sum_{n=0}^{N-1} x(n) \quad (3-14)$$

此估计是无偏的,且当各子样本 $x(0), x(1), \dots, x(N-1)$ 互不相关时,此估计是一致估计,现证明这两个结论。

证明: 估计的均值为

$$E[\hat{m}_x] = E\left[\frac{1}{N} \sum_{n=0}^{N-1} x(n)\right] = \frac{1}{N} \sum_{n=0}^{N-1} E[x(n)] = \frac{1}{N} \sum_{n=0}^{N-1} m_x = m_x \quad (3-15)$$

可见估计是无偏的。

估计的均方值为

$$\begin{aligned} E[\hat{m}_x^2] &= E\left[\left(\frac{1}{N} \sum_{n=0}^{N-1} x(n)\right) \left(\frac{1}{N} \sum_{m=0}^{N-1} x(m)\right)\right] \\ &= \frac{1}{N^2} \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} E[x(n)x(m)] \\ &= \frac{1}{N^2} \left\{ \sum_{n=0}^{N-1} E[x^2(n)] + \sum_{n=0}^{N-1} \sum_{m=0, m \neq n}^{N-1} E[x(n)x(m)] \right\} \\ &= \frac{1}{N^2} \left\{ \sum_{n=0}^{N-1} E[x^2(n)] + \sum_{n=0}^{N-1} \sum_{m=0, m \neq n}^{N-1} E[x(n)]E[x(m)] \right\} \end{aligned} \quad (3-16)$$

记 $E[x^2(n)] = D_x$, $E[x(n)] = m_x$, 所以

$$\begin{aligned} E[\hat{m}_x^2] &= \frac{1}{N} D_x + \frac{N(N-1)}{N^2} m_x^2 \\ &= \frac{1}{N} D_x + \frac{N-1}{N} m_x^2 \end{aligned} \quad (3-17)$$

所以,估计的方差为

$$\begin{aligned} \text{Var}(\hat{m}_x) &= E(\hat{m}_x^2) - [E(\hat{m}_x)]^2 \\ &= \frac{1}{N} D_x + \frac{N-1}{N} m_x^2 - m_x^2 = \frac{1}{N} \sigma_x^2 \end{aligned} \quad (3-18)$$

可见,当 $N \rightarrow \infty$ 时,估计的方差趋于零,所以 $\hat{m}_x$ 是一致估计。

2. 自相关函数的均值

在实际工作中,对于平稳随机序列 $x(n)$,所能得到的 $x(n)$ 的 N 个观察值为 $x(0), x(1), \dots, x(N-1)$, 记为 $x_N(n)$ 。将 $n \geq N$ 时的 $x(n)$ 的值假设为零。现在的任务是如何由这 N 个观察值估计出 $x(n)$ 的自相关函数 $R_x(m)$ 。对 $R_x(m)$ 的估计方法通常有两种:一是直接估计法,二是利用 FFT 求解。下面以自相关函数的直接估计法为例分析其性能。

如果观察值的点数 N 为有限值,则通过直接估计法求 $R(m)$ 估计值为

$$\hat{R}(m) = \frac{1}{N} \sum_{n=0}^{N-1} x_N(n) x_N(n-m) \quad (3-19)$$

由于 $x(n)$ 只有 N 个观察值,因此,对于每一个固定的迟延 m ,在 $(0, N-1)$ 内可以利用的数据只有 $(N-|m|)$ 个,所以在实际计算 $\hat{R}(m)$ 时,式(3-19)变为

$$\hat{R}(m) = \frac{1}{N} \sum_{n=0}^{N-1-|m|} x(n) x(n-m) \quad (3-20)$$

值得注意的是, $\hat{R}(m)$ 的长度为 $2N-1$, 对于实信号, 它是以 m 为对称中心的偶对称函数。所以在式(3-20)的计算中, 可先计算 $m \leq 0$ 的部分, 再根据对称性求 $m > 0$ 的部分。

现在讨论 $\hat{R}(m)$ 估计的质量。

1) 估计的偏差

由偏差的定义, 有

$$\text{Bia}[\hat{R}(m)] = E[\hat{R}(m)] - R(m) \quad (3-21)$$

式中

$$\begin{aligned} E[\hat{R}(m)] &= E\left[\frac{1}{N} \sum_{n=0}^{N-1-|m|} x(n)x(n-m)\right] \\ &= \frac{1}{N} \sum_{n=0}^{N-1-|m|} E[x(n)x(n-m)] \\ &= \frac{1}{N} \sum_{n=0}^{N-1-|m|} R(m) \end{aligned}$$

即

$$E[\hat{R}(m)] = \frac{N-|m|}{N} R(m) \quad (3-22)$$

所以估计的偏差为

$$\text{Bia}[\hat{R}(m)] = -\frac{|m|}{N} R(m) \quad (3-23)$$

分析式(3-22)和式(3-23)可以看出, 估计不是无偏估计, 而且当 $m \rightarrow N$ 时, 估计的偏差很大。对于有限的 m , 当 $N \rightarrow \infty$ 时, 有 $E[\hat{R}_x(m)] = R_x(m)$, 所以估计是渐进无偏的。由此, 可以得到如下结论:

(1) 对于一个固定的延迟 $|m|$, 当 $N \rightarrow \infty$ 时, $\text{Bia}[\hat{R}(m)] \rightarrow 0$, 因此, $\hat{R}(m)$ 是对 $R(m)$ 的渐近无偏估计;

(2) 对于一个固定的观察值 N , 只有当 $|m| \ll N$ 时, $\hat{R}(m)$ 的均值才接近于真值 $R(m)$, 即当 $|m|$ 越接近于 N 时, 估计的偏差越大;

(3) 由式(3-22)可以看出, $\hat{R}(m)$ 的均值是真值 $R(m)$ 和三角窗函数 $w(m)$

$$w(m) = \begin{cases} \frac{N-|m|}{N}, & 0 \leq |m| \leq N-1 \\ 0, & |m| \geq N \end{cases} \quad (3-24)$$

的乘积, $w(m)$ 的长度是 $2N-1$ (如图 3-1 所示)。式(3-24)的三角窗函数又称 Bartlett 窗函数, 由于它对 $R(m)$ 进行了加权, 致使 $\hat{R}(m)$ 产生了偏差。显然, 由于加权窗函数的存在, 产生了上述第(2)点结论。

该窗函数实际上是由于对真实数据的截短而产生的。因为 $x_N(n)$ 可以看作真实数据 $x(n)$ 和一矩形窗函数 $d(n)$ 相乘的结果, 即

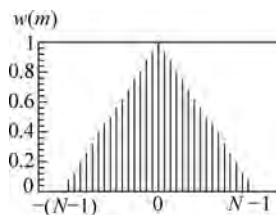


图 3-1 三角窗函数 $w(m)$

$$x_N(n) = x(n)d(n) \quad (3-25)$$

式中

$$d(n) = \begin{cases} 1, & 0 \leq n \leq N-1 \\ 0, & n < 0, n \geq N \end{cases} \quad (3-26)$$

根据式(3-20)和式(3-25),可得

$$\begin{aligned} \hat{R}(m) &= \frac{1}{N} \sum_{n=0}^{N-1-|m|} x_N(n)x_N(n-m) \\ &= \frac{1}{N} \sum_{n=0}^{N-1-|m|} x(n)d(n)x(n-m)d(n-m) \end{aligned} \quad (3-27)$$

$$\begin{aligned} E\{\hat{R}(m)\} &= \frac{1}{N} \sum_{n=0}^{N-1-|m|} E\{x(n)x(n-m)\}d(n)d(n-m) \\ &= \frac{R(m)}{N} \sum_{n=0}^{N-1-|m|} d(n)d(n-m) \\ &= R(m)\omega(m) \end{aligned} \quad (3-28)$$

可见, $\omega(m)$ 正是对矩形数据窗函数 $d(n)$ 作自相关的结果。

由上面的讨论可知, 当对一个信号做自然截短时, 就不可避免地对该数据施加了一个矩形窗口, 由此数据窗口就产生了加在自相关函数上的三角窗口。三角窗口影响了 $\hat{R}(m)$ 对 $R(m)$ 估计的质量。加在数据上的窗口一般称为数据窗, 加在自相关函数上的窗口一般称为延迟窗, 后面将看到, 这些窗函数也直接影响了功率谱估计的质量。

2) 估计的方差

根据方差的定义, 有

$$\begin{aligned} \text{Var}[\hat{R}(m)] &= E[\hat{R}(m) - E[\hat{R}(m)]]^2 \\ &= E[\hat{R}^2(m)] - \{E[\hat{R}(m)]\}^2 \end{aligned} \quad (3-29)$$

由式(3-22)可得

$$\{E[\hat{R}(m)]\}^2 = \left[\frac{N-|m|}{N} R(m) \right]^2 \quad (3-30)$$

同时可得

$$\begin{aligned} E[\hat{R}^2(m)] &= E \left[\frac{1}{N^2} \sum_{n=0}^{N-1-|m|} x(n)x(n-m) \sum_{k=0}^{N-1-|m|} x(k)x(k-m) \right] \\ &= \frac{1}{N^2} \sum_n \sum_k E[x(n)x(k)x(n-m)x(k-m)] \end{aligned} \quad (3-31)$$

这里要计算随机信号 $x(n)$ 的四阶矩, 但一般平稳信号的四阶矩很难计算, 因此假定 $x(n)$ 是零均值的高斯随机信号, 可以证明, 其四阶矩满足如下关系式:

$$\begin{aligned} E[x(n)x(k)x(n-m)x(k-m)] &= E[x(n)x(k)]E[x(n-m)x(k-m)] + \\ &\quad E[x(n)x(n-m)]E[x(k)x(k-m)] + \\ &\quad E[x(n)x(k-m)]E[x(k)x(n-m)] \\ &= R^2(n-k) + R^2(m) + \\ &\quad R(n-k+m)R(k-n+m) \end{aligned} \quad (3-32)$$

所以

$$\begin{aligned} E[\hat{R}^2(m)] &= \frac{1}{N^2} \sum_n \sum_k [R^2(n-k) + R^2(m) + R(n-k+m)R(k-n+m)] \\ &= \left[\frac{N-|m|}{N} R(m) \right]^2 + \frac{1}{N^2} \sum_n \sum_k [R^2(n-k) + \\ &\quad R(n-k+m)R(k-n+m)] \end{aligned} \quad (3-33)$$

将式(3-30)和式(3-33)代入式(3-29),可得

$$\text{Var}[\hat{R}(m)] = \frac{1}{N^2} \sum_{n=0}^{N-1-|m|} \sum_{k=0}^{N-1-|m|} [R^2(n-k) + R(n-k+m)R(k-n+m)] \quad (3-34)$$

对于任意序列 $g(n)$, 式(3-35)成立

$$\sum_{n=0}^{N-1-|m|} \sum_{k=0}^{N-1-|m|} g(n-k) = \sum_{i=-(N-1-|m|)}^{N-1-|m|} (N-|m|-|i|)g(i) \quad (3-35)$$

因此,令 $n-k=i$, 可把式(3-34)的双求和变成单求和,即对于一个确定的 i , $n-k=i$ 的组合数为 $N-|m|-|i|$ 。

$$\text{Var}[\hat{R}(m)] = \frac{1}{N} \sum_{i=-(N-1-|m|)}^{N-1-|m|} \left[1 - \frac{|m|+|i|}{N} \right] [R^2(i) + R(i+m)R(i-m)] \quad (3-36)$$

显然,当 $N \rightarrow \infty$ 时, $\text{Var}[\hat{R}(m)] \rightarrow 0$ 。又因为 $\lim_{N \rightarrow \infty} \text{Bia}[\hat{R}(m)] = 0$, 所以可得如下结论:

对于固定的延迟 $|m|$, $\hat{R}(m)$ 是 $R(m)$ 的一致估计。对 $R(m)$ 的另一种直接估计法是对式(3-20)稍做修改,即

$$\hat{R}(m) = \frac{1}{N-|m|} \sum_{n=0}^{N-1-|m|} x_N(n)x_N(n+m) \quad (3-37)$$

可以证明,式(3-37)的 $\hat{R}(m)$ 是对 $R(m)$ 的无偏估计,利用上述类似的推导,可分析估计子的性能,最终可得

$$\text{有偏估计子的方差: } \text{Var}[\hat{R}(m)] \propto O\left(\frac{1}{N}\right)$$

$$\text{无偏估计子的方差: } \text{Var}[\hat{R}(m)] \propto O\left(\frac{1}{N-|m|}\right)$$

可以看到无偏估计子的方差性能较差。实际应用一般采用有偏估计子,除了方差的原因外,还基于以下考虑: ①当 $|m|$ 取值较小时,有偏估计子的偏差不明显; ②对于大多数平稳随机信号,当 $|m|$ 值较大时, $R(m)$ 的值一般很小。式(3-23)显示偏差的值也和 $R(m)$ 的取值有关系,而 $|m|$ 较大时, $R(m)$ 的取值一般会变小,这时可以认为偏差也比较小。

3.3 估计性能界——CRB

不同的估计方法可以得到不同的估计量,估计量的性能可以通过前面介绍的无偏性、有效性和一致性评价,但在实际中,估计量可能比较复杂,很难评价估计量的有效性和一致性。此外,在得到一个估计量后,它的性能是否已经达到最佳? 是否还有更好的估计量? 克拉美

罗界(Cramér-Rao Bound, CRB)揭示了无偏估计量估计方差的最小值。

3.3.1 单参数实常量估计的 CRB

首先考虑只有一个待估计参数的情况,令 θ 为待估计参数,并且为实常量(非随机量)。观测信号 $x(n)$ 与 θ 的依赖关系用条件概率 $p(x|\theta)$ 表示[为描述方便,以下将 $x(n)$ 简记为 x_n],有如下定理。

定理 3-1 令 $\mathbf{x}=(x_0, x_1, \dots, x_{N-1})$ 为观测样本向量, $p(\mathbf{x}|\theta)$ 是 $\mathbf{x}$ 的条件概率密度。若 $\hat{\theta}$ 是 θ 的一个无偏估计子,且 $\partial f(\mathbf{x}|\theta)/\partial\theta$ 存在,则 $\hat{\theta}$ 的方差满足

$$\text{Var}(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] \geq I^{-1}(\theta) \quad (3-38)$$

式中

$$I(\theta) = E \left[\left(\frac{\partial \ln p(\mathbf{x}|\theta)}{\partial \theta} \right)^2 \right] = -E \left[\frac{\partial^2 \ln p(\mathbf{x}|\theta)}{\partial \theta^2} \right] \quad (3-39)$$

当且仅当

$$\frac{\partial \ln p(\mathbf{x}|\theta)}{\partial \theta} = K(\theta)(\hat{\theta} - \theta) \quad (3-40)$$

时,式(3-38)的等号成立。其中, $K(\theta)$ 是 θ 的正函数,且不包含 $\mathbf{x}$ 。

证明: 由假设条件知 $E(\hat{\theta} - \theta) = 0$, 因此

$$\begin{aligned} E[\hat{\theta} - \theta] &= \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} (\hat{\theta} - \theta) f(\mathbf{x}|\theta) dx_0 \dots dx_{N-1} \\ &\triangleq \int_{-\infty}^{+\infty} (\hat{\theta} - \theta) f(\mathbf{x}|\theta) d\mathbf{x} \\ &= 0 \end{aligned} \quad (3-41)$$

在式(3-41)两边对 θ 求偏微分,可得

$$\frac{\partial}{\partial \theta} E[\hat{\theta} - \theta] = \int_{-\infty}^{+\infty} \frac{\partial}{\partial \theta} [(\hat{\theta} - \theta) f(\mathbf{x}|\theta)] d\mathbf{x} = 0 \quad (3-42)$$

由式(3-42)得

$$\int_{-\infty}^{+\infty} -f(\mathbf{x}|\theta) d\mathbf{x} + \int_{-\infty}^{+\infty} (\hat{\theta} - \theta) \frac{\partial}{\partial \theta} f(\mathbf{x}|\theta) d\mathbf{x} = 0 \quad (3-43)$$

因为 $p(\mathbf{x}|\theta)$ 是概率密度函数,所以

$$\int_{-\infty}^{+\infty} f(\mathbf{x}|\theta) d\mathbf{x} = 1 \quad (3-44)$$

同时根据组合函数导数性质可得

$$\frac{\partial}{\partial \theta} f(\mathbf{x}|\theta) = \left[\frac{\partial}{\partial \theta} \ln f(\mathbf{x}|\theta) \right] \cdot f(\mathbf{x}|\theta) \quad (3-45)$$

式(3-43)重写为

$$\int_{-\infty}^{+\infty} \left[\frac{\partial}{\partial \theta} \ln f(\mathbf{x}|\theta) \right] \cdot (\hat{\theta} - \theta) f(\mathbf{x}|\theta) d\mathbf{x} = 1 \quad (3-46)$$

式(3-46)进一步重写为

$$\int_{-\infty}^{+\infty} \left[\frac{\partial}{\partial \theta} \ln f(\mathbf{x}|\theta) \sqrt{f(\mathbf{x}|\theta)} \right] \cdot (\hat{\theta} - \theta) \sqrt{f(\mathbf{x}|\theta)} d\mathbf{x} = 1 \quad (3-47)$$

根据泛函分析中的柯西不等式 $|\langle z_1, z_2 \rangle|^2 \leq \langle z_1, z_1 \rangle \cdot \langle z_2, z_2 \rangle$, 其中 z_1, z_2 是泛函空间的元素, $\langle \cdot \rangle$ 表示内积操作。对于式(3-47), 令

$$z_1 = f_1(\mathbf{x}) = \frac{\partial}{\partial \theta} \ln f(\mathbf{x} | \theta) \sqrt{f(\mathbf{x} | \theta)}$$

$$z_2 = f_2(\mathbf{x}) = (\hat{\theta} - \theta) \sqrt{f(\mathbf{x} | \theta)}$$

内积操作定义为 $\int_{-\infty}^{+\infty} f_1(\mathbf{x}) f_2(\mathbf{x}) d\mathbf{x}$ 。由式(3-47)得

$$\int_{-\infty}^{+\infty} \left[\frac{\partial}{\partial \theta} \ln f(\mathbf{x} | \theta) \right]^2 f(\mathbf{x} | \theta) d\mathbf{x} \cdot \int_{-\infty}^{+\infty} (\hat{\theta} - \theta)^2 f(\mathbf{x} | \theta) d\mathbf{x} \geq 1 \quad (3-48)$$

等价于

$$\int_{-\infty}^{+\infty} (\hat{\theta} - \theta)^2 f(\mathbf{x} | \theta) d\mathbf{x} \geq \frac{1}{\int_{-\infty}^{+\infty} \left[\frac{\partial}{\partial \theta} \ln f(\mathbf{x} | \theta) \right]^2 f(\mathbf{x} | \theta) d\mathbf{x}} \quad (3-49)$$

式(3-49)左边刚好就是无偏估计的方差计算结果。柯西不等式等号成立的条件是 z_1, z_2 成比例, 即式(3-40)成立。

证毕。

特别要指出的是, 式(3-42)积分和求导顺序可以互换, 这在实际中一般可以满足。为了得到更直观的描述, 将在式(3-44)两边对 θ 求导数, 并将积分和求导顺序互换可得

$$\int_{-\infty}^{+\infty} \frac{\partial}{\partial \theta} f(\mathbf{x} | \theta) d\mathbf{x} = 0 \quad (3-50)$$

基于式(3-45)可得

$$\int_{-\infty}^{+\infty} \left[\frac{\partial}{\partial \theta} \ln f(\mathbf{x} | \theta) \right] \cdot f(\mathbf{x} | \theta) d\mathbf{x} = E \left[\frac{\partial}{\partial \theta} \ln f(\mathbf{x} | \theta) \right] = 0 \quad (3-51)$$

式(3-51)称为定理 3-1 成立的规范化条件, 是积分和求导顺序可以互换的直观描述。在式(3-51)两边对 θ 求导数, 则可得式(3-39)。

CRB 给出了无偏估计量估计方差的下限。达到 CRB 的估计, 其估计的方差是最小的, 称这样的估计为有效估计。 $I(\theta)$ 称为数据的费希尔(Fisher)信息, CRB 是费希尔信息的倒数。直观理解是, 费希尔信息越多, CRB 越低。

费希尔信息具有信息度量的基本性质。首先, 由式(3-39)可以看出, 费希尔信息具有非负性; 其次, 对于独立观测, 费希尔信息具有可加性, 这是因为, 对于独立的观测, 有

$$\ln p(\mathbf{x} | \theta) = \sum_{n=0}^{N-1} \ln p(x_n | \theta) \quad (3-52)$$

所以

$$-E \left[\frac{\partial^2 \ln p(\mathbf{x} | \theta)}{\partial \theta^2} \right] = -\sum_{n=0}^{N-1} E \left[\frac{\partial^2 \ln p(x_n | \theta)}{\partial \theta^2} \right] \quad (3-53)$$

如果观测是独立同分布的, 那么

$$I(\theta) = Ni(\theta) \quad (3-54)$$

式中

$$i(\theta) = -E \left[\frac{\partial^2 \ln p(x_n | \theta)}{\partial \theta^2} \right] \quad (3-55)$$

对于非独立观测,费希尔信息要比独立观测低,例如,对于 $x_0 = x_1 = \cdots = x_{N-1}$ 这种完全相关的情况, $I(\theta) = i(\theta)$, 此时, CRB 并不随观测数据的增加而降低。

例 3-2 独立同分布平稳随机信号服从高斯分布, 用式(3-14)估计该随机信号的均值, 该估计子的方差是否达到 CRB?

解 平稳高斯信号的一维概率密度为

$$p(x | m_x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - m_x)^2}{2\sigma^2}\right) \quad (3-56)$$

多变量的联合概率密度为

$$p(\mathbf{x} | m_x) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{N} \prod_{n=0}^{N-1} \exp\left(-\frac{(x_n - m_x)^2}{2\sigma^2}\right) \quad (3-57)$$

对数联合概率密度的一阶、二阶偏导数为

$$\frac{\partial \ln p(\mathbf{x} | m_x)}{\partial m_x} = \sum_{n=0}^{N-1} \left(\frac{x_n - m_x}{\sigma^2} \right) \quad (3-58)$$

$$\frac{\partial^2 \ln p(\mathbf{x} | m_x)}{\partial m_x^2} = -\frac{N}{\sigma^2} \quad (3-59)$$

可得到其费希尔信息量为

$$I(m_x) = -E \left[\frac{\partial^2 \ln p(\mathbf{x} | m_x)}{\partial m_x^2} \right] = \frac{N}{\sigma^2} \quad (3-60)$$

对比 3.2.2 节中估计子的方差, 可以看到该均值估计可达到 CRB。注意到式(3-58)中的一阶导数可重写为

$$\frac{\partial \ln p(\mathbf{x} | m_x)}{\partial m_x} = \frac{N}{\sigma^2} (\hat{m}_x - m_x) \quad (3-61)$$

即一阶导数和 $(\hat{m}_x - m_x)$ 成正比, 显示该估计子满足式(3-40)的条件。

例 3-3 信号观测模型为 $x_n = A \cos(2\pi f_0 n + \phi) + w(n)$, 其中, f_0 为已知参数, $w(n) \sim N(0, \sigma^2)$, 相位 ϕ 为未知参数。求相位估计的 CRB。

解 观测信号的一维概率密度为

$$p(x | \phi) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_n - A \cos(2\pi f_0 n + \phi))^2}{2\sigma^2}\right) \quad (3-62)$$

假设参数估计基于 N 个观测, 记为 $\mathbf{x} = [x_0, x_1, \cdots, x_{N-1}]^T$, 其联合概率密度函数为

$$\begin{aligned} p(\mathbf{x} | \phi) &= p(x_0, x_1, \cdots, x_{N-1} | \phi) \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^N \prod_{n=0}^{N-1} \exp\left(-\frac{(x_n - A \cos(2\pi f_0 n + \phi))^2}{2\sigma^2}\right) \end{aligned} \quad (3-63)$$

可得到观测信号的对数似然函数为

$$\ln p(\mathbf{x} | \phi) = -\frac{N}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x_n - A \cos(2\pi f_0 n + \phi))^2 \quad (3-64)$$

可得

$$\begin{aligned}
\frac{\partial \ln p(\mathbf{x} | \phi)}{\partial \phi} &= -\frac{1}{\sigma^2} \sum_{n=0}^{N-1} [x_n - A \cos(2\pi f_0 n + \phi)] A \sin(2\pi f_0 n + \phi) \\
&= -\frac{A}{\sigma^2} \sum_{n=0}^{N-1} [x_n \sin(2\pi f_0 n + \phi) - \frac{A}{2} \sin(4\pi f_0 n + 2\phi)] \\
\frac{\partial^2 \ln p(\mathbf{x} | \phi)}{\partial \phi^2} &= -\frac{A}{\sigma^2} \sum_{n=0}^{N-1} [x_n \cos(2\pi f_0 n + \phi) - A \cos(4\pi f_0 n + 2\phi)] - E \left[\frac{\partial^2 \ln p(\mathbf{x} | \phi)}{\partial \phi^2} \right] \\
&= \frac{A}{\sigma^2} E \left[\sum_{n=0}^{N-1} [x_n \cos(2\pi f_0 n + \phi) - A \cos(4\pi f_0 n + 2\phi)] \right] \\
&= \frac{A}{\sigma^2} \sum_{n=0}^{N-1} [E[x_n] \cos(2\pi f_0 n + \phi) - A \cos(4\pi f_0 n + 2\phi)] \\
&= \frac{A}{\sigma^2} \sum_{n=0}^{N-1} [A \cos^2(2\pi f_0 n + \phi) - A \cos(4\pi f_0 n + 2\phi)] \\
&= \frac{A^2}{\sigma^2} \sum_{n=0}^{N-1} \left[\left(\frac{1}{2} + \frac{1}{2} \cos(4\pi f_0 n + 2\phi) - \cos(4\pi f_0 n + 2\phi) \right) \right] \\
&= \frac{A^2}{2\sigma^2} \sum_{n=0}^{N-1} [1 - \cos(4\pi f_0 n + 2\phi)] \\
&= \frac{NA^2}{2\sigma^2} - \frac{A^2}{2\sigma^2} \sum_{n=0}^{N-1} \cos(4\pi f_0 n + 2\phi) \tag{3-65}
\end{aligned}$$

式(3-65)中求和项为周期序列,有

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} \cos(4\pi f_0 n + 2\phi) = 0 \tag{3-66}$$

当 N 比较大时,可认为

$$\frac{1}{N} \sum_{n=0}^{N-1} \cos(4\pi f_0 n + 2\phi) \ll 1 \tag{3-67}$$

于是可认为式(3-65)中第二项远小于第一项,直接忽略第二项可得

$$-E \left[\frac{\partial^2 \ln p(\mathbf{x} | \phi)}{\partial \phi^2} \right] = \frac{NA^2}{2\sigma^2} \tag{3-68}$$

其倒数即为相位估计的 CRB。

3.3.2 多参量估计的 CRB

假设有 L 个未知参数,用向量表示为

$$\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_L]^T \tag{3-69}$$

参数估计的 Cramer-Rao 不等式表示为

$$E[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T] \geq \mathbf{I}^{-1}(\boldsymbol{\theta}) \tag{3-70}$$

其中,符号“ $\geq$ ”用于矩阵时有不同的含义。对于矩阵 $\mathbf{A}, \mathbf{B}$, $\mathbf{A} \geq \mathbf{B}$ 表示 $\mathbf{A} - \mathbf{B}$ 是半正定矩阵, $\mathbf{A} > \mathbf{B}$ 表示 $\mathbf{A} - \mathbf{B}$ 是正定矩阵。式(3-70)中左边为估计子的协方差矩阵,右边的 $\mathbf{I}(\boldsymbol{\theta})$ 称为费希尔信息矩阵,定义为

$$\mathbf{I}(\boldsymbol{\theta}) = -E \left[\frac{\partial^2 \ln p(\mathbf{x} | \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right] \quad (3-71)$$

式(3-71)中用到了向量函数的求导公式,下面给出其定义。设 $f(\boldsymbol{\theta})$ 是 $L \times 1$ 向量 $\boldsymbol{\theta}$ 的函数,函数值为标量,其导数表示为

$$\frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \frac{\partial f(\boldsymbol{\theta})}{\partial \theta_1} \\ \frac{\partial f(\boldsymbol{\theta})}{\partial \theta_2} \\ \vdots \\ \frac{\partial f(\boldsymbol{\theta})}{\partial \theta_L} \end{bmatrix}_{L \times 1} \quad (3-72)$$

根据该定义可得以下常用导数

$$\frac{\partial \mathbf{a}^T \boldsymbol{\theta}}{\partial \boldsymbol{\theta}} = \frac{\partial \boldsymbol{\theta}^T \mathbf{a}}{\partial \boldsymbol{\theta}} = \mathbf{a} \quad (3-73)$$

$$\frac{\partial \boldsymbol{\theta}^T \mathbf{A} \boldsymbol{\theta}}{\partial \boldsymbol{\theta}} = (\mathbf{A}^T + \mathbf{A}) \boldsymbol{\theta} \quad (3-74)$$

式(3-72)的定义可拓展到多个函数的组合,假设有 R 个关于 $\boldsymbol{\theta}$ 的函数构成如下向量:

$$\mathbf{f}(\boldsymbol{\theta}) = \begin{bmatrix} f_1(\boldsymbol{\theta}) \\ f_2(\boldsymbol{\theta}) \\ \vdots \\ f_R(\boldsymbol{\theta}) \end{bmatrix}_{R \times 1} \quad (3-75)$$

则其关于 $\boldsymbol{\theta}$ 的导数定义为

$$\frac{\partial \mathbf{f}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} = \begin{bmatrix} \frac{\partial f_1(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial f_1(\boldsymbol{\theta})}{\partial \theta_2} & \cdots & \frac{\partial f_1(\boldsymbol{\theta})}{\partial \theta_L} \\ \frac{\partial f_2(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial f_2(\boldsymbol{\theta})}{\partial \theta_2} & \cdots & \frac{\partial f_2(\boldsymbol{\theta})}{\partial \theta_L} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_R(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial f_R(\boldsymbol{\theta})}{\partial \theta_2} & \cdots & \frac{\partial f_R(\boldsymbol{\theta})}{\partial \theta_L} \end{bmatrix}_{R \times L} \quad (3-76)$$

根据上述定义,可得矩阵 $\mathbf{I}(\boldsymbol{\theta})$ 中第 $\{i, j\}$ 个元素为

$$[\mathbf{I}(\boldsymbol{\theta})]_{ij} = -E \left[\frac{\partial^2 \ln p(\mathbf{x} | \boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right] \quad (3-77)$$

对式(3-71)这里不做详细证明,下面以一个例子介绍多参量 CRB 的计算。

例 3-4 对独立同分布平稳高斯随机信号,计算均值估计和方差估计的 CRB。

解 由于待估计向量 $\boldsymbol{\theta} = [m_x \quad \sigma^2]^T$, 因此费希尔信息矩阵是 2×2 的矩阵,即

$$\mathbf{I}(\boldsymbol{\theta}) = -E \left[\frac{\partial^2 \ln p(\mathbf{x} | \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right] = \begin{bmatrix} -E \left[\frac{\partial^2 \ln p(\mathbf{x} | \boldsymbol{\theta})}{\partial m_x^2} \right] & -E \left[\frac{\partial^2 \ln p(\mathbf{x} | \boldsymbol{\theta})}{\partial m_x \partial \sigma^2} \right] \\ -E \left[\frac{\partial^2 \ln p(\mathbf{x} | \boldsymbol{\theta})}{\partial \sigma^2 \partial m_x} \right] & -E \left[\frac{\partial^2 \ln p(\mathbf{x} | \boldsymbol{\theta})}{\partial (\sigma^2)^2} \right] \end{bmatrix} \quad (3-78)$$

由式(3-56)得到对数似然函数为

$$\ln p(\mathbf{x} | \boldsymbol{\theta}) = -\frac{N}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x_n - m_x)^2 \quad (3-79)$$

所以

$$\frac{\partial \ln p(\mathbf{x} | \boldsymbol{\theta})}{\partial m_x} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} (x_n - m_x) \quad (3-80)$$

$$\frac{\partial \ln p(\mathbf{x} | \boldsymbol{\theta})}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{n=0}^{N-1} (x_n - m_x)^2 \quad (3-81)$$

$$\frac{\partial^2 \ln p(\mathbf{x} | \boldsymbol{\theta})}{\partial m_x^2} = -\frac{N}{\sigma^2} \quad (3-82)$$

$$\frac{\partial^2 \ln p(\mathbf{x} | \boldsymbol{\theta})}{\partial (\sigma^2)^2} = \frac{N}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{n=0}^{N-1} (x_n - m_x)^2 \quad (3-83)$$

$$\frac{\partial^2 \ln p(\mathbf{x} | \boldsymbol{\theta})}{\partial m_x \partial \sigma^2} = -\frac{1}{\sigma^4} \sum_{n=0}^{N-1} (x_n - m_x) \quad (3-84)$$

$$\frac{\partial^2 \ln p(\mathbf{x} | \boldsymbol{\theta})}{\partial \sigma^2 \partial m_x} = \frac{\partial^2 \ln p(\mathbf{x} | \boldsymbol{\theta})}{\partial m_x \partial \sigma^2} \quad (3-85)$$

费希尔信息矩阵为

$$\mathbf{I}(\boldsymbol{\theta}) = \begin{bmatrix} -E\left(-\frac{N}{\sigma^2}\right) & -E\left[-\frac{1}{\sigma^4} \sum_{n=0}^{N-1} (x_n - m_x)\right] \\ -E\left[-\frac{1}{\sigma^4} \sum_{n=0}^{N-1} (x_n - m_x)\right] & -E\left[\frac{N}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{n=0}^{N-1} (x_n - m_x)^2\right] \end{bmatrix} = \begin{bmatrix} \frac{N}{\sigma^2} & 0 \\ 0 & \frac{N}{2\sigma^4} \end{bmatrix} \quad (3-86)$$

因此,容易得出参数估计的 CRB 为

$$\text{Var}(\hat{m}_x) \geq \frac{\sigma^2}{N} \quad (3-87)$$

$$\text{Var}(\hat{\sigma}^2) \geq \frac{2\sigma^4}{N} \quad (3-88)$$

3.3.3 参数变换的 CRB

在实际中,经常遇到希望估计的参数是某个参数的函数,例如,在电平估计中,有时感兴趣的不是信号电平 A 的估计,而是信号的功率 A^2 。假设估计 A 的 CRB 已知,如何得到 A^2 估计的 CRB 呢? 下面回答这一问题。

假定可以直接计算 CRB 的参数为 θ , 希望估计的参数为 α , 两者有如下函数关系: $\alpha = T(\theta)$ 。那么可以证明 $\hat{\alpha}$ 的 CRB 为

$$\text{Var}(\hat{\alpha}) \geq \frac{\left(\frac{\partial T(\theta)}{\partial \theta}\right)^2}{-E\left[\frac{\partial^2 \ln p(\mathbf{x} | \theta)}{\partial \theta^2}\right]} \quad (3-89)$$

例如,基于例 3-4 中式(3-87)的结果,求 $\alpha = m_x^2$ 的 CRB,则

$$\text{Var}(\hat{\alpha}) \geq \frac{(2m_x)^2}{N/\sigma^2} = \frac{4m_x^2\sigma^2}{N} \quad (3-90)$$

实际应用中经常碰到多变量的映射关系。例如,独立同分布高斯随机信号的平均功率 $D = m_x^2 + \sigma^2$ 。前面已经得到估计 m_x 、 σ^2 的 CRB,在此基础上可采用以下方法得到估计 D 的 CRB。

假设 $P \times 1$ 维向量 θ 包含模型中的可估计参数,有多个待估计参数,可表示为 θ 的函数,记为 $\alpha = T(\theta)$,其中 $T(\theta)$ 是 $R \times 1$ 维的函数向量,即

$$\alpha = [T_1(\theta) \quad T_2(\theta) \quad \cdots \quad T_R(\theta)]^T \quad (3-91)$$

则

$$E[(\hat{\alpha} - \alpha)(\hat{\alpha} - \alpha)^T] \geq \frac{\partial T(\theta)}{\partial \theta^T} \mathbf{I}^{-1}(\theta) \left[\frac{\partial T(\theta)}{\partial \theta^T} \right]^T \quad (3-92)$$

式中, $\partial T(\theta)/\partial \theta^T$ 的定义见式(3-76)。

例 3-5 对高斯白噪声中恒定电平,求信噪比估计的 CRB。

解 高斯白噪声中恒定电平观测的数学模型为 $x_n = A + w_n$,其中 A 为电平值, $w_n \sim (0, \sigma^2)$ 为零均值高斯白噪声,信噪比定义为 $\alpha = A^2/\sigma^2$,因此可先计算估计 A 、 σ^2 的 CRB。注意到观测数据可以看成非零均值高斯信号, A 、 σ^2 的估计对应均值和方差的估计,可采用例 3-4 的结果。设 $\theta = [A \quad \sigma^2]^T$, $T(\theta) = A^2/\sigma^2$,由例 3-4 可知

$$\mathbf{I}(\theta) = \begin{bmatrix} \frac{N}{\sigma^2} & 0 \\ 0 & \frac{N}{2\sigma^4} \end{bmatrix} \quad (3-93)$$

又因为

$$\frac{\partial T(\theta)}{\partial \theta^T} = \begin{bmatrix} \frac{\partial T(\theta)}{\partial A} & \frac{\partial T(\theta)}{\partial \sigma^2} \end{bmatrix} = \begin{bmatrix} \frac{2A}{\sigma^2} & -\frac{A^2}{\sigma^4} \end{bmatrix} \quad (3-94)$$

所以

$$\frac{\partial T(\theta)}{\partial \theta^T} \mathbf{I}^{-1}(\theta) \left[\frac{\partial T(\theta)}{\partial \theta^T} \right]^T = \begin{bmatrix} \frac{2A}{\sigma^2} & -\frac{1}{\sigma^4} \end{bmatrix} \begin{bmatrix} \frac{\sigma^2}{N} & 0 \\ 0 & \frac{2\sigma^4}{N} \end{bmatrix} \begin{bmatrix} \frac{2A}{\sigma^2} \\ -\frac{1}{\sigma^4} \end{bmatrix} \quad (3-95)$$

$$= \frac{4A^2}{N\sigma^2} + \frac{2}{N\sigma^4} \quad (3-96)$$

可以证明,对向量参数的有效估计量经过线性变换后仍然是有效估计量,而向量参数的有效估计量经过非线性变换后是渐近有效估计量。

3.3.4 复参数估计的 CRB

当待估计参数是复数时,推导涉及复变函数的微分操作,比实参数估计复杂得多。这一节先给出复变函数的基本概念,再介绍一种复参数估计的 CRB 的推导方法。

令复自变量为 $z = x + jy$, $x, y \in \mathbf{R}$ 。定义在复数域的函数为

$$f(z) = u(z, z^*) + jv(z, z^*) = u(x, y) + jv(x, y) \quad (3-97)$$

其中, $u(x, y), v(x, y)$ 为实二元函数。 $f(z)$ 的导数定义为

$$\begin{aligned} \frac{\partial f}{\partial z} &= \frac{1}{2} \left(\frac{\partial f}{\partial x} - j \frac{\partial f}{\partial y} \right) \\ \frac{\partial f}{\partial z^*} &= \frac{1}{2} \left(\frac{\partial f}{\partial x} + j \frac{\partial f}{\partial y} \right) \end{aligned} \quad (3-98)$$

复变函数可导的充分必要条件可参考相关复变函数文献。这里假设所涉及的复变函数均可导, 以下是一些常用的性质。

假设 $g(z), f(z)$ 在区域 D 内可导, 则其导数满足以下性质:

$$\frac{d}{dz}(f(z) + g(z)) = \frac{df(z)}{dz} + \frac{dg(z)}{dz} \quad (3-99)$$

$$\frac{d}{dz}(f(z) \cdot g(z)) = \frac{df(z)}{dz} g(z) + \frac{dg(z)}{dz} f(z) \quad (3-100)$$

$$\frac{d}{dz} \left(\frac{f(z)}{g(z)} \right) = \frac{1}{g^2(z)} \left(\frac{df(z)}{dz} g(z) - \frac{dg(z)}{dz} f(z) \right) \quad (3-101)$$

$$\frac{d}{dz}(f(g(z))) = \frac{df(g)}{dg} \frac{dg(z)}{dz} \quad (3-102)$$

对多变量函数, 如 $f(\mathbf{z})$, 其中 $\mathbf{z} = [z_1, z_2, \dots, z_N]^T$ 为包含所有自变量的向量, 其导数定义为

$$\frac{\partial f(\mathbf{z})}{\partial \mathbf{z}} = \left[\frac{\partial f(\mathbf{z})}{\partial z_1}, \frac{\partial f(\mathbf{z})}{\partial z_2}, \dots, \frac{\partial f(\mathbf{z})}{\partial z_N} \right]^T \quad (3-103)$$

对多维函数, 可定义与式(3-70)相同的偏导数。根据该定义可得以下对复向量的常用导数。

$$\frac{\partial \mathbf{a}^H \mathbf{z}}{\partial \mathbf{z}} = \mathbf{a}^* \quad (3-104)$$

$$\frac{\partial \mathbf{z}^H \mathbf{A} \mathbf{z}}{\partial \mathbf{z}} = \mathbf{A}^T \mathbf{z}^* \quad (3-105)$$

接下来考虑复参数估计的 CRB, 假设观测数据为 $\mathbf{Z} = [z_0, z_1, \dots, z_{N-1}]$, 观测数据的信号产生模型已知, 但其中有 L 个未知参数, 记为 $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_L]^T$ 。待估计的目标参数为 $\boldsymbol{\theta}$ 的函数, 记为

$$\mathbf{p} = [p_1(\boldsymbol{\theta}), p_2(\boldsymbol{\theta}), \dots, p_K(\boldsymbol{\theta})]^T \quad (3-106)$$

记估计子为 $\mathbf{r}(\mathbf{Z})$, 考虑无偏估计子, 有

$$\int_{\Omega} \mathbf{r}(\mathbf{Z}) f(\mathbf{Z} | \boldsymbol{\theta}) d\mathbf{Z} = \mathbf{p} \quad (3-107)$$

其中, Ω 表示观测空间, 在式(3-107)两边对 $\boldsymbol{\theta}$ 求偏导可得

$$\int_{\Omega} \mathbf{r}(\mathbf{Z}) \frac{\partial f(\mathbf{Z} | \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^H} d\mathbf{Z} = \frac{\partial \mathbf{p}}{\partial \boldsymbol{\theta}^H} \quad (3-108)$$

式(3-108)左边重写为

$$\begin{aligned} \int_{\Omega} \mathbf{r}(\mathbf{Z}) \frac{\partial f(\mathbf{Z} | \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^H} d\mathbf{Z} &= \int_{\Omega} \mathbf{r}(\mathbf{Z}) \frac{\partial \ln f(\mathbf{Z} | \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^H} f(\mathbf{Z} | \boldsymbol{\theta}) d\mathbf{Z} \\ &= E \left[\mathbf{r}(\mathbf{Z}) \frac{\partial \ln f(\mathbf{Z} | \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^H} \right] \end{aligned} \quad (3-109)$$

于是

$$\frac{\partial \mathbf{p}}{\partial \boldsymbol{\theta}^H} = E \left[\mathbf{r}(\mathbf{Z}) \frac{\partial \ln f(\mathbf{Z} | \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^H} \right] \quad (3-110)$$

简记 $\mathbf{r}(\mathbf{Z})$ 为 $\mathbf{r}$, $\ln f(\mathbf{Z} | \boldsymbol{\theta})$ 为 $\ln f$, 构造如下向量

$$\begin{bmatrix} \mathbf{r} \\ \left[\frac{\partial \ln f}{\partial \boldsymbol{\theta}^H} \right]^H \end{bmatrix}_{(K+L) \times 1} \quad (3-111)$$

可算出其自协方差矩阵,并由自协方差矩阵的半正定特性可得

$$\begin{bmatrix} \text{Cov}(\mathbf{r}, \mathbf{r}) & \frac{\partial \mathbf{p}}{\partial \boldsymbol{\theta}^H} \\ \left[\frac{\partial \mathbf{p}}{\partial \boldsymbol{\theta}^H} \right]^H & E \left[\left[\frac{\partial \ln f}{\partial \boldsymbol{\theta}^H} \right]^H \left[\frac{\partial \ln f}{\partial \boldsymbol{\theta}^H} \right] \right] \end{bmatrix} \geq 0 \quad (3-112)$$

其中已经用到式(3-110)的结论。根据自协方差矩阵特性,式(3-113)成立

$$\begin{bmatrix} \mathbf{D} & -\frac{\partial \mathbf{p}}{\partial \boldsymbol{\theta}^H} \mathbf{I}^{-1}(\boldsymbol{\theta}) \end{bmatrix} \begin{bmatrix} \text{Cov}(\mathbf{r}, \mathbf{r}) & \frac{\partial \mathbf{p}}{\partial \boldsymbol{\theta}^H} \\ \left[\frac{\partial \mathbf{p}}{\partial \boldsymbol{\theta}^H} \right]^H & \mathbf{I}(\boldsymbol{\theta}) \end{bmatrix} \begin{bmatrix} \mathbf{D} & -\frac{\partial \mathbf{p}}{\partial \boldsymbol{\theta}^H} \mathbf{I}^{-1}(\boldsymbol{\theta}) \end{bmatrix}^H \geq 0 \quad (3-113)$$

其中, $\mathbf{D}$ 为单位矩阵,进一步将费希尔信息量矩阵定义为

$$\mathbf{I}(\boldsymbol{\theta}) = E \left[\left[\frac{\partial \ln f}{\partial \boldsymbol{\theta}^H} \right]^H \left[\frac{\partial \ln f}{\partial \boldsymbol{\theta}^H} \right] \right] \quad (3-114)$$

经简单计算,式(3-113)矩阵可化简为

$$\text{Cov}(\mathbf{r}, \mathbf{r}) - \frac{\partial \mathbf{p}}{\partial \boldsymbol{\theta}^H} \mathbf{I}^{-1}(\boldsymbol{\theta}) \left[\frac{\partial \mathbf{p}}{\partial \boldsymbol{\theta}^H} \right]^H \geq 0 \quad (3-115)$$

可以得到估计子的误差界为

$$\text{Cov}(\mathbf{r}, \mathbf{r}) \geq \frac{\partial \mathbf{p}}{\partial \boldsymbol{\theta}^H} \mathbf{I}^{-1}(\boldsymbol{\theta}) \left[\frac{\partial \mathbf{p}}{\partial \boldsymbol{\theta}^H} \right]^H \quad (3-116)$$

式(3-116)和式(3-92)实参量估计的 CRB 是统一的。下面以一个例子介绍 CRB 的具体计算过程。

例 3-6 通信中常用以下信号观测模型:

$$x_n = A s_n + w_n, \quad n = 0, 1, \dots, N-1 \quad (3-117)$$

其中, A 为复值信道响应, s_n 为正交调制发送信号, w_n 是零均值复高斯白噪声序列, 方差为 σ^2 , 假设发送符号有归一化功率, 即 $|s_n|^2 = 1$, 求信道估计的 CRB。

解 根据式(2-95)给出的复高斯概率密度分布, 关于 A 的一维条件概率密度为

$$p(x | A) = \frac{1}{\pi \sigma^2} \exp \left(-\frac{|x_n - A s_n|^2}{\sigma^2} \right) \quad (3-118)$$

多变量的联合概率密度为

$$p(\mathbf{x} | A) = \left(\frac{1}{\pi \sigma^2} \right)^N \prod_{n=0}^{N-1} \exp \left(-\frac{|x_n - A s_n|^2}{\sigma^2} \right) \quad (3-119)$$

基于式(3-114)费希尔信息量矩阵的定义及式(3-50)所示的规范化条件, 可以得到与式(3-39)

类似的结论。

$$I(\theta) = E \left[\frac{\partial \ln p(\mathbf{x} | \theta)}{\partial \theta} \frac{\partial \ln p(\mathbf{x} | \theta)}{\partial \theta^*} \right] = -E \left[\frac{\partial^2 \ln p(\mathbf{x} | \theta)}{\partial \theta \partial \theta^*} \right] \quad (3-120)$$

对数联合概率密度的一、二阶偏导数为

$$\frac{\partial \ln p(\mathbf{x} | A)}{\partial A} = \sum_{n=0}^{N-1} \left[\frac{s_n (x_n - A s_n)^*}{\sigma^2} \right] \quad (3-121)$$

$$\frac{\partial^2 \ln p(\mathbf{x} | A)}{\partial A \partial A^*} = -\frac{1}{\sigma^2} \sum_{n=0}^{N-1} |s_n|^2 = -\frac{N}{\sigma^2} \quad (3-122)$$

可得到估计信道响应的性能界为

$$\text{CRB}_A = \frac{\sigma^2}{N} \quad (3-123)$$

3.4 最大似然估计

当被估计量为未知常量时,可以采用比较简单的最大似然估计。最大似然估计可以简便地完成对复杂估计问题的求解,而且当观测数据足够多时,其性能也是非常好的。因此,最大似然估计在实际中得到了广泛采用。本节只考虑实参数的估计。

3.4.1 最大似然估计的基本原理

设观测向量 $\mathbf{x} = [x_0 \ x_1 \ \cdots \ x_{N-1}]^T$, 待估计参量为 θ 。观测数据与待估计量的概率依赖关系以概率密度 $p(\mathbf{x} | \theta)$ 描述。概率密度是以 θ 为参量的函数,在观测给定的条件下,如 $\mathbf{x} = \mathbf{x}_0$, $p(\mathbf{x} = \mathbf{x}_0 | \theta)$ 反映了 θ 取各个值的可能性大小,称 $p(\mathbf{x} = \mathbf{x}_0 | \theta)$ 为 θ 的似然函数, $p(\mathbf{x} = \mathbf{x}_0 | \theta)$ 简记为 $p(\mathbf{x} | \theta)$ 。估计问题本质上就是根据观测求出未知量,也就是说,在得到某个观测值 $\mathbf{x} = \mathbf{x}_0$ 后,如何根据这个观测确定未知量 θ 的值。

从另外一个角度,似然函数 $p(\mathbf{x} = \mathbf{x}_0 | \theta)$ 反映了对不同的 θ 值观测到 $\mathbf{x} = \mathbf{x}_0$ 的概率。因为当前的观测结果刚好就是 $\mathbf{x}_0$,于是可认为实际中观测到 $\mathbf{x} = \mathbf{x}_0$ 的概率应该尽可能大,即 θ 取值应该使得 $p(\mathbf{x} = \mathbf{x}_0 | \theta)$ 取得最大。使似然函数最大所对应的参数 θ 作为对 θ 的估计,称为最大似然(Maximum Likelihood, ML)估计,记为 $\hat{\theta}_{\text{ml}}$,即

$$\hat{\theta}_{\text{ml}} = \underset{\theta}{\operatorname{argmax}} p(\mathbf{x} | \theta) \quad \text{或} \quad \hat{\theta}_{\text{ml}} = \underset{\theta}{\operatorname{argmax}} \ln p(\mathbf{x} | \theta) \quad (3-124)$$

下面以两个例子说明最大似然准则的应用。

例 3-7 设实随机变量满足以下分布: $x \sim N(0, \sigma^2)$, 有 N 次独立观测 $x_0, \dots, x_{N-1}$, 求 σ^2 的最大似然估计。

解 单变量似然函数为

$$p(x | \sigma^2) = \left(\frac{1}{2\pi\sigma^2} \right)^{\frac{1}{2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \quad (3-125)$$

多变量联合似然函数为

$$p(\mathbf{x} | \sigma^2) = \left(\frac{1}{2\pi\sigma^2} \right)^{\frac{N}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} x_n^2\right) \quad (3-126)$$

对数联合似然函数为

$$\ln p(\mathbf{x} | \sigma^2) = -\frac{N}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{n=0}^{N-1} x_n^2 \quad (3-127)$$

令

$$\frac{\partial \ln p(\mathbf{x} | \sigma^2)}{\partial \sigma^2} = 0 \quad (3-128)$$

得到估计结果为

$$\hat{\sigma}_{\text{ml}}^2 = \frac{1}{N} \sum_{n=0}^{N-1} x_n^2 \quad (3-129)$$

很容易验证

$$\left. \frac{\partial^2 \ln p(\mathbf{x} | \sigma^2)}{\partial (\sigma^2)^2} \right|_{\sigma^2 = \frac{1}{N} \sum_{n=0}^{N-1} x_n^2} < 0 \quad (3-130)$$

所以式(3-129)求得的是最大值。

例 3-8 假定观测序列为

$$x_n = A \cos(2\pi f_0 n + \phi) + w_n, \quad n=0, 1, \dots, N-1 \quad (3-131)$$

幅度 A 和频率 f_0 是已知的, w_n 是零均值高斯白噪声序列, 方差为 σ^2 , 求相位 ϕ 的最大似然估计。

解 似然函数为

$$p(\mathbf{x} | \phi) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} [x_n - A \cos(2\pi f_0 n + \phi)]^2 \right\} \quad (3-132)$$

对数似然函数及其导数为

$$\ln p(\mathbf{x} | \phi) = -\frac{N}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{n=0}^{N-1} [x_n - A \cos(2\pi f_0 n + \phi)]^2 \quad (3-133)$$

$$\frac{\partial \ln p(\mathbf{x} | \phi)}{\partial \phi} = -\frac{1}{\sigma^2} \sum_{n=0}^{N-1} [x_n - A \cos(2\pi f_0 n + \phi)] A \sin(2\pi f_0 n + \phi) \quad (3-134)$$

令式(3-134)等于零, 可得

$$\begin{aligned} \sum_{n=0}^{N-1} x_n \sin(2\pi f_0 n + \hat{\phi}_{\text{ml}}) &= A \sum_{n=0}^{N-1} \cos(2\pi f_0 n + \hat{\phi}_{\text{ml}}) \sin(2\pi f_0 n + \hat{\phi}_{\text{ml}}) \\ &= 2A \sum_{n=0}^{N-1} \sin(4\pi f_0 n + 2\hat{\phi}_{\text{ml}}) \end{aligned} \quad (3-135)$$

当 f_0 不在 0 或 1/2 附近时, 式(3-135)右边是周期序列求和, 当 N 足够大时, 可认为其值近似为零。因此, 最大似然估计近似满足

$$\sum_{n=0}^{N-1} x_n \sin(2\pi f_0 n + \hat{\phi}_{\text{ml}}) = 0 \quad (3-136)$$

展开式(3-136), 得

$$\sum_{n=0}^{N-1} x_n \sin 2\pi f_0 n \cos \hat{\phi}_{\text{ml}} = - \sum_{n=0}^{N-1} x_n \cos 2\pi f_0 n \sin \hat{\phi}_{\text{ml}} \quad (3-137)$$

$$\hat{\phi}_{\text{ml}} = -\arctan \frac{\sum_{n=0}^{N-1} x_n \sin 2\pi f_0 n}{\sum_{n=0}^{N-1} x_n \cos 2\pi f_0 n} \quad (3-138)$$

前面讨论的是单参量的最大似然估计问题,在实际中经常需要同时估计多个参量,例如,在雷达信号的处理中,当检测到目标后,需要同时估计目标的位置、速度等参量,这样的问题称为多参量同时估计,或称为向量估计。式(3-124)所描述的最大似然估计很容易推广到向量估计的情形。

考虑实参数估计问题,设有 p 个参量 $\theta_0, \theta_1, \dots, \theta_{p-1}$ 需要同时估计,定义一个 p 维向量 $\boldsymbol{\theta} = [\theta_0, \theta_1, \dots, \theta_{p-1}]^T$,那么最大似然估计定义为

$$\hat{\boldsymbol{\theta}}_{\text{ml}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} p(\mathbf{x} | \boldsymbol{\theta}) \quad \text{或者} \quad \hat{\boldsymbol{\theta}}_{\text{ml}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \ln p(\mathbf{x} | \boldsymbol{\theta}) \quad (3-139)$$

类似地,最大似然估计的必要条件是

$$\left. \frac{\partial p(\mathbf{x} | \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\text{ml}}} = 0 \quad \text{或者} \quad \left. \frac{\partial \ln p(\mathbf{x} | \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\text{ml}}} = 0 \quad (3-140)$$

例 3-9 设有 N 次独立观测 $x_n = A + w_n$, 其中 $w_n \sim N(0, \sigma^2)$, A, σ^2 均未知,求 A, σ^2 的最大似然估计。

解 令 $\boldsymbol{\theta} = [A \quad \sigma^2]^T$, 则

$$p(\mathbf{x} | \boldsymbol{\theta}) = \left(\frac{1}{2\pi\sigma^2} \right)^{\frac{N}{2}} \exp \left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x_n - A)^2 \right] \quad (3-141)$$

$$\ln p(\mathbf{x} | \boldsymbol{\theta}) = -\frac{N}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x_n - A)^2 \quad (3-142)$$

$$\frac{\partial \ln p(\mathbf{x} | \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \frac{N}{\sigma^2} \left(\frac{1}{N} \sum_{n=0}^{N-1} x_n - A \right) \\ -\frac{N}{2\sigma^4} \left[\sigma^2 - \frac{1}{N} \sum_{n=0}^{N-1} (x_n - A)^2 \right] \end{bmatrix} \quad (3-143)$$

令 $\left. \frac{\partial \ln p(\mathbf{x} | \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\text{ml}}} = 0$, 可求得最大似然估计为

$$\hat{\boldsymbol{\theta}}_{\text{ml}} = \begin{bmatrix} \hat{A}_{\text{ml}} \\ \hat{\sigma}_{\text{ml}}^2 \end{bmatrix} = \begin{bmatrix} \bar{x} \\ \frac{1}{N} \sum_{n=0}^{N-1} (x_n - \bar{x})^2 \end{bmatrix} \quad (3-144)$$

其中, $\bar{x}$ 表示样本的直接平均, 定义为

$$\bar{x} = \frac{1}{N} \sum_{n=0}^{N-1} x_n \quad (3-145)$$

以下例子给出更一般化线性估计模型中的最大似然估计算法。

例 3-10 假定待估计向量为 $\boldsymbol{\theta} = [\theta_0, \theta_1, \dots, \theta_{p-1}]^T$, 观测数据 $\mathbf{x} = [x_0, x_1, \dots, x_{N-1}]^T$, 有如下的线性模型表示

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w} \quad (3-146)$$

其中, $\mathbf{H}$ 是 $N \times p$ 的矩阵, 且 $N > p$, $\mathbf{H}$ 的秩为 p , $\mathbf{w}$ 是多维高斯噪声, 其概率模型为 $\mathbf{w} \propto N(0, \mathbf{C})$, 其中 $\mathbf{C}$ 为 $\mathbf{w}$ 的协方差矩阵。求 θ 的最大似然估计。

解 由式(2-91)得似然函数为

$$p(\mathbf{x} | \theta) = \frac{1}{(2\pi)^{\frac{N}{2}} [\det(\mathbf{C})]^{\frac{1}{2}}} \exp \left[-\frac{1}{2} (\mathbf{x} - \mathbf{H}\theta)^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{H}\theta) \right] \quad (3-147)$$

要使似然函数最大, 只需使式(3-148)的表达式最小

$$J(\theta) = (\mathbf{x} - \mathbf{H}\theta)^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{H}\theta) \quad (3-148)$$

基于式(3-73)、式(3-74), 并利用协方差矩阵的对称性, 将式(3-148)对 θ 求导

$$\frac{\partial J(\theta)}{\partial \theta} = -2\mathbf{H}^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{H}\theta) \quad (3-149)$$

令导数等于零, 可求得最大似然估计, 即

$$\mathbf{H}^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{H}\hat{\theta}_{\text{ml}}) = 0 \quad (3-150)$$

解式(3-150)的方程式得最大似然估计为

$$\hat{\theta}_{\text{ml}} = (\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}^{-1} \mathbf{x} \quad (3-151)$$

由式(3-151)可以看出, 线性观测模型的最大似然估计是观测结果的线性函数, 计算简单, 因此在噪声协方差矩阵已知情况下是常用的估计算法。

3.4.2 变换参数的最大似然估计

在许多情况下, 希望得到估计参量 θ 的一个函数, 例如 $\alpha = T(\theta)$, 如果利用最大似然准则求得 $\hat{\theta}$, 如何进一步求解 $\hat{\alpha}$? 下面通过几个例子说明变换参数的最大似然估计的求法。

例 3-11 设有 N 次独立观测 $x_n = A + w_n, n = 0, 1, \dots, N-1$ 。其中, $w_n \propto N(0, \sigma^2)$, A 为未知参数, σ^2 为已知, 求 $\alpha = e^A$ 的最大似然估计。

解 α 的似然函数及其导数为

$$p(\mathbf{x} | \alpha) = \left(\frac{1}{2\pi\sigma^2} \right)^{\frac{N}{2}} \exp \left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x_n - \ln \alpha)^2 \right] \quad (3-152)$$

$$\frac{\partial \ln p(\mathbf{x} | \alpha)}{\partial \alpha} = \frac{1}{\alpha \sigma^2} \sum_{n=0}^{N-1} (x_n - \ln \alpha) = \frac{N}{\alpha \sigma^2} \left(\frac{1}{N} \sum_{n=0}^{N-1} x_n - \ln \alpha \right) \quad (3-153)$$

令式(3-153)等于零, 可解得

$$\hat{\alpha}_{\text{ml}} = \exp \left(\frac{1}{N} \sum_{i=0}^{N-1} x_i \right) = \exp(\bar{x}) \quad (3-154)$$

由于 $\bar{x}$ 刚好是 A 的最大似然估计, 所以

$$\hat{\alpha}_{\text{ml}} = \exp(\hat{A}_{\text{ml}}) \quad (3-155)$$

可见, α 的最大似然估计只需要用 A 的最大似然估计代入变换式 $\alpha = e^A$ 中就可以求得。因此, 如果变换 $\alpha = T(\theta)$ 是一一对应的, 那么变换参数后的最大似然估计可以直接由式(3-156)得到。

$$\hat{\alpha}_{\text{ml}} = T(\hat{\theta}_{\text{ml}}) \quad (3-156)$$

这一特性称为最大似然估计的不变性。

如果变换 $\alpha = T(\theta)$ 不是一一对应的, 则不能简单地应用式(3-156)。下面通过一个例子加以说明。

例 3-12 在例 3-11 中, 假定要求的估计为 $\alpha = A^2$, 求 α 的最大似然估计。

解 由于 $A = \pm\sqrt{\alpha}$, 所以本例的变换不是一一对应的, 似然函数 $p(\mathbf{x}|\alpha)$ 需要两个概率密度描述

$$p_1(\mathbf{x}|\alpha) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{N}{2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x_n - \sqrt{\alpha})^2\right], \quad A \geq 0 \quad (3-157)$$

$$p_2(\mathbf{x}|\alpha) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{N}{2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x_n + \sqrt{\alpha})^2\right], \quad A < 0 \quad (3-158)$$

那么, 最大似然估计为

$$\hat{\alpha}_{\text{ml}} = \arg\max_{\alpha} \{p_1(\mathbf{x}|\alpha), p_2(\mathbf{x}|\alpha)\} \quad (3-159)$$

式(3-159)的计算可以分为两步进行。

(1) 给定一个 α 值, 如 $\alpha = \alpha_0$, 比较 $p_1(\mathbf{x}|\alpha_0)$ 和 $p_2(\mathbf{x}|\alpha_0)$ 的大小, 如果

$$p_1(\mathbf{x}|\alpha_0) \geq p_2(\mathbf{x}|\alpha_0) \quad (3-160)$$

则 $p_T(\mathbf{x}|\alpha_0) = p_1(\mathbf{x}|\alpha_0)$, 对所有 α 的取值重复以上过程, 得到 $p_T(\mathbf{x}|\alpha)$, 称 $p_T(\mathbf{x}|\alpha)$ 为修正的似然函数。

(2) 求出使 $p_T(\mathbf{x}|\alpha)$ 最大的 α 值作为 α 的最大似然估计, 即

$$\hat{\alpha}_{\text{ml}} = \arg\max_{\alpha} p_T(\mathbf{x}|\alpha) \quad (3-161)$$

对于本例, 要比较 $p_1(\mathbf{x}|\alpha_0)$ 和 $p_2(\mathbf{x}|\alpha_0)$ 的大小, 只需要比较 $\sum_{n=0}^{N-1} (x_n - \sqrt{\alpha})^2$ 和 $\sum_{n=0}^{N-1} (x_n + \sqrt{\alpha})^2$ 的大小, 令

$$J(\mathbf{x}, \alpha) = \sum_{n=0}^{N-1} (x_n - \sqrt{\alpha})^2 - \sum_{n=0}^{N-1} (x_n + \sqrt{\alpha})^2 \quad (3-162)$$

式(3-162)经整理得

$$J(\mathbf{x}, \alpha) = -4 \sum_{n=0}^{N-1} x_n \sqrt{\alpha} = -4N\sqrt{\alpha}\bar{x} \quad (3-163)$$

可见, 如果 $\bar{x} \geq 0$, 则 $J(\mathbf{x}, \alpha) < 0$, $p_1(\mathbf{x}|\alpha_0) \geq p_2(\mathbf{x}|\alpha_0)$; 如果 $\bar{x} < 0$, 则 $J(\mathbf{x}, \alpha) > 0$, $p_2(\mathbf{x}|\alpha_0) > p_1(\mathbf{x}|\alpha_0)$, 所以修正的似然函数为

$$p_T(\mathbf{x}|\alpha) = \begin{cases} p_1(\mathbf{x}|\alpha), & \bar{x} \geq 0 \\ p_2(\mathbf{x}|\alpha), & \bar{x} < 0 \end{cases} \quad (3-164)$$

当 $\bar{x} \geq 0$ 时, 由 $p_T(\mathbf{x}|\alpha) = p_1(\mathbf{x}|\alpha)$ 可求得 $\hat{\alpha}_{\text{ml}} = \bar{x}^2$; 当 $\bar{x} < 0$ 时, 同样由 $p_T(\mathbf{x}|\alpha) = p_2(\mathbf{x}|\alpha)$ 可求得 $\hat{\alpha}_{\text{ml}} = \bar{x}^2$ 。综合两种情况可得

$$\hat{\alpha}_{\text{ml}} = \bar{x}^2 \quad (3-165)$$

需要注意的是, 如果将 A 的原始估计 $\hat{A}_{\text{ml}} = \bar{x}$ 直接代入变换式 $\alpha = A^2$ 中, 也可以得到式(3-165), 但这只是一种巧合。参数变换如果不是——对应的, 那么变换后参数的最大似

然估计不能简单地通过代入变换式得到。

3.5 贝叶斯估计

最大似然估计没有假定被估计量的先验信息,把待估计量看成一个常量。实际环境中待估计量可能是一个随机变量,并且概率密度已知。这时最大似然方法依然可用,但未能充分利用待估计量的信息。这一节介绍贝叶斯估计,它是一类能充分利用待估计量概率信息的估计方法。本节只考虑实参数的估计。

3.5.1 代价函数

在估计某个参量 θ 时,噪声的影响会使估计产生误差,估计误差是要付出代价的,这种代价可以用代价函数加以描述,记为 $c(\theta, \hat{\theta})$ 。一般而言,估计误差 $\Delta\theta = \theta - \hat{\theta}$ 越小,代价越小。代价函数可表示为 $c(\theta, \hat{\theta})$ 、 $c(\theta - \hat{\theta})$ 或 $c(\Delta\theta)$,典型的代价函数有以下几种。

(1) 平方代价函数

$$c(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2 \quad (3-166)$$

(2) 绝对值代价函数

$$c(\theta, \hat{\theta}) = |\theta - \hat{\theta}| \quad (3-167)$$

(3) 均匀代价函数

$$c(\theta, \hat{\theta}) = \begin{cases} 1, & |\theta - \hat{\theta}| \geq \frac{\epsilon}{2} \\ 0, & |\theta - \hat{\theta}| < \frac{\epsilon}{2} \end{cases} \quad (\epsilon \text{ 为一常数}) \quad (3-168)$$

(4) 二次型代价函数

如果被估计量是 p 维随机向量,令误差向量为 $\Delta\theta = \theta - \hat{\theta}$,则二次型代价函数定义为

$$c(\theta, \hat{\theta}) = \|\Delta\theta\|_S^2 = (\theta - \hat{\theta})^T \mathbf{S} (\theta - \hat{\theta}) \quad (3-169)$$

式中, $\|\Delta\theta\|$ 为误差向量的范数, $\mathbf{S}$ 为 $p \times p$ 维的对称非负定的加权矩阵。二次型代价函数实际上是估计误差的加权平方和,当 $\mathbf{S}$ 为单位矩阵时,二次型代价函数是估计误差的平方和。

代价函数也可以考虑其他函数,但实际上,最佳估计对代价函数的选择并不敏感。不失一般性,以下考虑单变量,代价函数确定后,可以计算平均代价

$$\bar{c} = E\{c(\theta, \hat{\theta}(x))\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} c(\theta, \hat{\theta}(x)) p(\theta, x) d\theta dx \quad (3-170)$$

式(3-170)中用 $\hat{\theta}(x)$ 代替 $\hat{\theta}$,强调 $\hat{\theta}$ 是从 x 计算得到,存在函数依赖关系。所谓贝叶斯估计就是使平均代价最小的估计。式(3-170)可以写成

$$\bar{c} = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} c(\theta, \hat{\theta}(x)) p(\theta | x) d\theta \right] p(x) dx = \int_{-\infty}^{\infty} \bar{c}(\theta | x) p(x) dx \quad (3-171)$$

实际应用中 $p(x)$ 可能是未知的或有复杂的表达式。在一次观测中,如果观测结果为 x ,则可认为其对应较大的概率。同时由于 $p(x)$ 是非负的,所以使平均代价最小近似等价于使条件平均代价 $\bar{c}(\theta | x)$ 最小,也就是使

$$\bar{c}(\theta | x) = \int_{-\infty}^{\infty} c(\theta, \hat{\theta}(x)) p(\theta | x) d\theta \quad (3-172)$$

最小。贝叶斯估计与代价函数的选取有关,使用不同的代价函数将得到不同的贝叶斯估计。

3.5.2 最小均方误差估计

当代价函数为平方代价函数时,平均代价刚好等于估计误差的均方值。先考虑单个参数的估计问题,有

$$\bar{c} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [\theta - \hat{\theta}(x)]^2 p(\theta, x) d\theta dx = E\{[\theta - \hat{\theta}(x)]^2\} \quad (3-173)$$

使平均代价最小等价于使均方误差最小,这时的贝叶斯估计称为最小均方误差估计,记为 $\hat{\theta}_{\text{ms}}$ 。

把平方代价函数代入式(3-172)中,得

$$\bar{c}(\theta | x) = \int_{-\infty}^{\infty} [\theta - \hat{\theta}(x)]^2 p(\theta | x) d\theta \quad (3-174)$$

在式(3-174)两边对 $\hat{\theta}$ 求导,得

$$\frac{\partial \bar{c}(\theta | x)}{\partial \hat{\theta}} = -2 \int_{-\infty}^{\infty} [\theta - \hat{\theta}(x)] p(\theta | x) d\theta \quad (3-175)$$

并令导数在 $\hat{\theta} = \hat{\theta}_{\text{ms}}$ 处为零,得

$$\hat{\theta}_{\text{ms}} = \int_{-\infty}^{\infty} \theta p(\theta | x) d\theta = E(\theta | x) \quad (3-176)$$

由于条件平均代价对 $\hat{\theta}$ 的二阶导数为正,因此 $\hat{\theta}_{\text{ms}}$ 所对应的条件平均代价为极小值,即 $\hat{\theta}_{\text{ms}}$ 为最小均方估计。由式(3-176)可以看出,最小均方误差估计为被估计量 θ 的条件均值。

如果被估计量为向量 $\theta = [\theta_0, \theta_1, \dots, \theta_{p-1}]^T$,将二次型代价函数代入式(3-170),得

$$\begin{aligned} \bar{c} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [\theta - \hat{\theta}(x)]^T \mathbf{S} [\theta - \hat{\theta}(x)] p(\theta, x) d\theta dx \\ &= E\{[\theta - \hat{\theta}(x)]^T \mathbf{S} [\theta - \hat{\theta}(x)]\} \end{aligned} \quad (3-177)$$

将二次型代价函数代入式(3-172),得

$$\bar{c}(\theta | x) = \int_{-\infty}^{\infty} [\theta - \hat{\theta}(x)]^T \mathbf{S} [\theta - \hat{\theta}(x)] p(\theta | x) d\theta \quad (3-178)$$

对 $\hat{\theta}(x)$ 求导并令导数在 $\hat{\theta} = \hat{\theta}_{\text{ms}}$ 处为零,得

$$-\int_{-\infty}^{\infty} 2\mathbf{S} [\theta - \hat{\theta}(x)] p(\theta | x) d\theta = 0 \quad (3-179)$$

求解可得

$$\hat{\theta}_{\text{ms}} = \int_{-\infty}^{\infty} \theta p(\theta | x) d\theta = E(\theta | x) \quad (3-180)$$

由于 $\frac{\partial^2 \bar{c}(\theta | x)}{\partial \hat{\theta} \partial \hat{\theta}^T} = 2\mathbf{S}$,所以式(3-180)求得的是极小值, $\hat{\theta}_{\text{ms}}$ 称为最小均方误差估计。

3.5.3 条件中位数估计

当代价函数为绝对值代价函数时,条件平均代价为(考虑单变量情况)

$$\begin{aligned}\bar{c}(\theta | x) &= \int_{-\infty}^{\infty} |\theta - \hat{\theta}(x)| p(\theta | x) d\theta \\ &= \int_{-\infty}^{\hat{\theta}(x)} [\hat{\theta}(x) - \theta] p(\theta | x) d\theta + \int_{\hat{\theta}(x)}^{\infty} [\theta - \hat{\theta}(x)] p(\theta | x) d\theta\end{aligned}\quad (3-181)$$

在式(3-181)两边对 $\hat{\theta}$ 求导,并令导数在 $\hat{\theta} = \hat{\theta}_{\text{abs}}$ 处为零,得

$$\int_{-\infty}^{\hat{\theta}_{\text{abs}}} p(\theta | x) d\theta = \int_{\hat{\theta}_{\text{abs}}}^{\infty} p(\theta | x) d\theta \quad (3-182)$$

由式(3-182)可以看出,采用绝对值代价函数的贝叶斯估计 $\hat{\theta}_{\text{abs}}$ 刚好是条件概率密度 $p(\theta | x)$ 的中位数(Median),所以也称为条件中位数估计,记为 $\hat{\theta}_{\text{med}}$ 。

3.5.4 最大后验概率估计

当采用均匀代价函数时,条件平均代价为

$$\begin{aligned}\bar{c} = c(\theta | x) &= \int_{-\infty}^{\hat{\theta} - \frac{\Delta}{2}} p(\theta | x) d\theta + \int_{\hat{\theta} + \frac{\Delta}{2}}^{\infty} p(\theta | x) d\theta \\ &= 1 - \int_{\hat{\theta} - \frac{\Delta}{2}}^{\hat{\theta} + \frac{\Delta}{2}} p(\theta | x) d\theta\end{aligned}\quad (3-183)$$

很显然,当式(3-183)中后面的积分最大时,平均代价最小。要使所述积分达到最大,当 $p(\theta | x)$ 具有类似于高斯分布的特征时, $\hat{\theta}$ 可选为 $p(\theta | x)$ 的最大值所对应的 θ (如图 3-2(a) 所示),这时对应的贝叶斯估计就是最大后验概率(Maximum A Posteriori, MAP)估计,记为 $\hat{\theta}_{\text{map}}$ 。但是,当 $p(\theta | x)$ 不具备上述特征时,MAP 不能得到最优的估计。如图 3-2(b) 所示, $p(\theta | x)$ 具有指数分布的特征,这时 $\hat{\theta}_{\text{map}}$ 显然不是合理的选择。

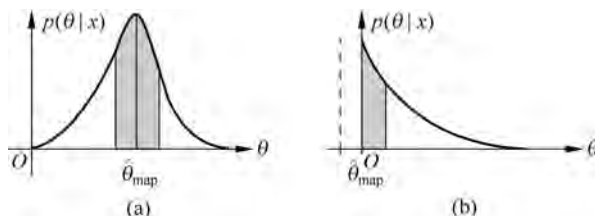


图 3-2 θ 条件概率的不同形式

高斯信号是实际应用中最常见的信号,再加上 MAP 估计思路直观,实现也比较简便,MAP 估计成为一种重要的参数估计准则。

下面进一步讨论最大后验概率估计与最大似然估计的联系。最大后验概率估计的数学描述为

$$\hat{\theta}_{\text{map}} = \operatorname{argmax}_{\theta} p(\theta | x) \quad \text{或} \quad \hat{\theta}_{\text{map}} = \operatorname{argmax}_{\theta} \ln p(\theta | x) \quad (3-184)$$

由于

$$p(\theta | x) = \frac{p(x | \theta) p(\theta)}{p(x)} \quad (3-185)$$

当难以得到 $p(x)$ 的表达式时,对于当前样本,认为 $p(x)$ 有固定值,这时后验概率密度最大和使 $p(x | \theta) p(\theta)$ 最大近似相同,因此,最大后验概率估计为

$$\hat{\theta}_{\text{map}} = \operatorname{argmax}_{\theta} p(x | \theta) p(\theta) \quad \text{或} \quad \hat{\theta}_{\text{map}} = \operatorname{argmax}_{\theta} [\ln p(x | \theta) + \ln p(\theta)] \quad (3-186)$$

式(3-186)直观给出了最大后验概率估计和最大似然估计的联系,可以看到最大后验概率估计实际上是在最大似然估计的基础上利用待估计参数的先验知识进行修正。当后验概率密度是可导的函数时,最大后验概率估计的必要条件是

$$\left. \frac{\partial p(\theta | x)}{\partial \theta} \right|_{\theta = \hat{\theta}_{\text{map}}} = 0 \quad \text{或} \quad \left. \frac{\partial \ln p(\theta | x)}{\partial \theta} \right|_{\theta = \hat{\theta}_{\text{map}}} = 0 \quad (3-187)$$

式(3-187)称为最大后验概率方程。

式(3-187)所描述的最大后验概率估计很容易推广到向量估计的情形。假定待估计量为 p 维向量 $\theta = [\theta_0, \theta_1, \dots, \theta_{p-1}]^T$, 后验概率密度为 $p(\theta | x)$, 那么向量最大后验概率估计为

$$\hat{\theta}_{\text{map}} = \operatorname{argmax}_{\theta} p(\theta | x) \quad \text{或} \quad \hat{\theta}_{\text{map}} = \operatorname{argmax}_{\theta} \ln p(\theta | x) \quad (3-188)$$

3.5.5 贝叶斯估计举例

例 3-13 设观测信号模型为 $x = A + w$, 其中被估计量 A 在 $[-A_0, A_0]$ 上均匀分布, 测量噪声 $w \propto N(0, \sigma^2)$, 求 A 的最大后验概率估计和最小均方误差估计。

解 先求最大后验概率估计, 因为只有一个观测值, 得到条件概率为

$$p(x | A) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{(x - A)^2}{2\sigma^2} \right]$$

$$p(A) = \begin{cases} \frac{1}{2A_0}, & -A_0 \leq A \leq A_0 \\ 0, & |A| > A_0 \end{cases} \quad (3-189)$$

根据贝叶斯公式

$$p(A | x) = \frac{p(x | A) p(A)}{p(x)} \quad (3-190)$$

这时难以得到 $p(x)$ 的表达式, 用式(3-191)进行最大后验概率估计

$$A_{\text{map}} = \operatorname{argmax}_A [p(x | A) p(A)] \quad (3-191)$$

当 $-A_0 \leq x \leq A_0$ 时, $p(x | A) p(A)$ 的最大值出现在 $A = x$ 处, 所以, $\hat{A}_{\text{map}} = x$; 当 $x > A_0$ 时, $p(x | A) p(A)$ 的最大值出现在 $A = A_0$ 处, $\hat{A}_{\text{map}} = A_0$; 当 $x < -A_0$ 时, $p(x | A) p(A)$ 的最大值出现在 $A = -A_0$ 处, $\hat{A}_{\text{map}} = -A_0$, 即

$$\hat{A}_{\text{map}} = \begin{cases} -A_0, & x < -A_0 \\ x, & -A_0 \leq x \leq A_0 \\ A_0, & x > A_0 \end{cases} \quad (3-192)$$

易证最大似然估计的结果为 $\hat{A}_{\text{ml}} = x$ 。 A 满足在 $[-A_0, A_0]$ 上均匀分布的假设下, 显然最大后验概率估计的结果更合理。图 3-3 给出了不同噪声强度下 MAP 和 ML 估计的均方误差性能比较, 可见在高噪声强度下, MAP 估计性能明显优于 ML 估计。在实际应用中, 如果能获得被估计参数的先验概率信息, 可改善估计结果。

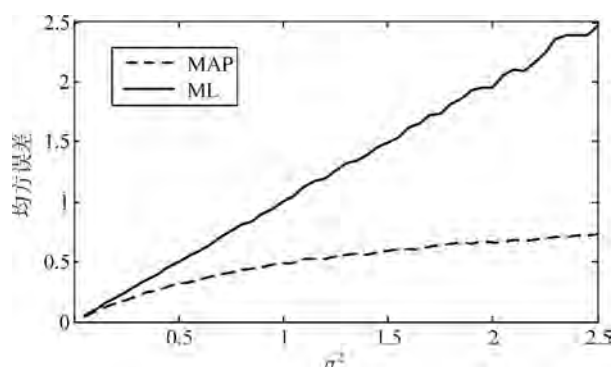


图 3-3 不同噪声强度下 MAP 和 ML 估计性能比较

下面求最小均方误差估计,由式(3-176),得

$$\begin{aligned}
 \hat{A}_{\text{ms}} &= \int_{-\infty}^{\infty} A p(A | x) dA \\
 &= \int_{-\infty}^{\infty} A \frac{p(x | A) p(A)}{p(x)} dA \\
 &= \frac{\int_{-\infty}^{\infty} A p(x | A) p(A) dA}{\int_{-\infty}^{\infty} p(x | A) p(A) dA} \\
 &= \frac{\int_{-A_0}^{A_0} \frac{A}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-A)^2}{2\sigma^2}\right] \cdot \frac{1}{2A_0} dA}{\int_{-A_0}^{A_0} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-A)^2}{2\sigma^2}\right] \cdot \frac{1}{2A_0} dA} \\
 &= \frac{\int_{x-A_0}^{x+A_0} (x-u) \cdot \exp\left(-\frac{u^2}{2\sigma^2}\right) du}{\int_{x-A_0}^{x+A_0} \exp\left(-\frac{u^2}{2\sigma^2}\right) du} \\
 &= x - \frac{2\sigma^2 \int_{(z-a)/\sqrt{2}}^{(z+a)/\sqrt{2}} u \cdot \exp(-u^2) du}{\sigma \int_{z-a}^{z+a} \exp(-u^2/2) du} \\
 &= x - \frac{\sigma \{\exp[-(z-a)^2/2] - \exp[-(z+a)^2/2]\}}{\sqrt{2\pi} [Q(z-a) - Q(z+a)]} \quad (3-193)
 \end{aligned}$$

式中, $a = A_0/\sigma$, $z = x/\sigma$ 表示归一化观测值, $Q(\cdot)$ 为标准正态概率密度函数的概率右尾函数, 如式(3-194)所示。

$$Q(v) = \frac{1}{\sqrt{2\pi}} \int_v^{+\infty} \exp(-x^2) dx \quad (3-194)$$

例 3-14 设信号有 N 次独立观测模型 $x_n = A + w_n$ ($n = 0, 1, \dots, N-1$), 其中 $w_n \sim N(0, \sigma^2)$, $A \sim N(\mu_A, \sigma_A^2)$, 求 A 的最大后验概率估计和最小均方误差估计。

解 先求后验概率密度

$$\begin{aligned}
p(A | \mathbf{x}) &= \frac{p(\mathbf{x} | A)p(A)}{\int_{-\infty}^{\infty} p(\mathbf{x} | A)p(A)dA} \\
&= \frac{\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x_n - A)^2\right] \frac{1}{\sqrt{2\pi\sigma_A^2}} \exp\left[-\frac{1}{2\sigma_A^2} (A - \mu_A)^2\right]}{\int_{-\infty}^{\infty} \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x_n - A)^2\right] \frac{1}{\sqrt{2\pi\sigma_A^2}} \exp\left[-\frac{1}{2\sigma_A^2} (A - \mu_A)^2\right] dA} \\
&= \frac{\exp\left\{-\frac{1}{2} \left[\frac{1}{\sigma^2} (NA^2 - 2NA\bar{x}) + \frac{1}{\sigma_A^2} (A - \mu_A)^2 \right]\right\}}{\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2} \left[\frac{1}{\sigma^2} (NA^2 - 2NA\bar{x}) + \frac{1}{\sigma_A^2} (A - \mu_A)^2 \right]\right\} dA} \\
&= \frac{\exp\left(-\frac{1}{2}W(A)\right)}{\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}W(A)\right) dA} \quad (3-195)
\end{aligned}$$

式中, $\bar{x} = \frac{1}{N} \sum_{n=0}^{N-1} x_n$ 为样本均值, 且

$$W(A) = \frac{1}{\sigma^2} (NA^2 - 2NA\bar{x}) + \frac{1}{\sigma_A^2} (A - \mu_A)^2 \quad (3-196)$$

注意, 式(3-195)的分母与 A 无关, $W(A)$ 是 A 的二次型, 经过配方可以把式(3-196)写成

$$W(A) = \frac{1}{\sigma_{A|x}^2} (A - \mu_{A|x})^2 - \frac{\mu_{A|x}^2}{\sigma_{A|x}^2} + \frac{\mu_A^2}{\sigma_A^2} \quad (3-197)$$

式中,

$$\sigma_{A|x}^2 = \left(\frac{N}{\sigma^2} + \frac{1}{\sigma_A^2} \right)^{-1} \quad (3-198)$$

$$\mu_{A|x} = \left(\frac{N}{\sigma^2} \bar{x} + \frac{\mu_A}{\sigma_A^2} \right) \sigma_{A|x}^2 \quad (3-199)$$

将式(3-197)代入式(3-195), 得

$$\begin{aligned}
p(A | \mathbf{x}) &= \frac{\exp\left[-\frac{1}{2\sigma_{A|x}^2} (A - \mu_{A|x})^2\right] \exp\left[-\frac{1}{2} \left(\frac{\mu_A^2}{\sigma_A^2} - \frac{\mu_{A|x}^2}{\sigma_{A|x}^2} \right)\right]}{\int_{-\infty}^{\infty} \exp\left[-\frac{1}{2\sigma_{A|x}^2} (A - \mu_{A|x})^2\right] \exp\left[-\frac{1}{2} \left(\frac{\mu_A^2}{\sigma_A^2} - \frac{\mu_{A|x}^2}{\sigma_{A|x}^2} \right)\right] dA} \\
&= \frac{1}{\sqrt{2\pi\sigma_{A|x}^2}} \exp\left[-\frac{1}{2\sigma_{A|x}^2} (A - \mu_{A|x})^2\right] \quad (3-200)
\end{aligned}$$

由式(3-200)可以看出, 后验概率密度分布满足高斯分布的。由于最小均方误差估计为被估计量的条件均值, 所以

$$\hat{A}_{\text{ms}} = \mu_{A|x} = \left(\frac{N}{\sigma^2} \bar{x} + \frac{\mu_A}{\sigma_A^2} \right) \sigma_{A|x}^2 = \frac{\frac{N}{\sigma^2} \bar{x} + \frac{\mu_A}{\sigma_A^2}}{\frac{N}{\sigma^2} + \frac{1}{\sigma_A^2}} \quad (3-201)$$

另外,由于最大后验概率估计是使后验概率最大对应的 A 值,因此,由式(3-183)可得

$$\hat{A}_{\text{map}} = \mu_{A|x} = \hat{A}_{\text{ms}} \quad (3-202)$$

3.6 线性最小均方误差估计

对于随机参量的估计,在 3.5.2 节中介绍了最小均方误差估计。最小均方误差估计是被估计量的条件均值,这个条件均值通常都是观测的非线性函数,估计器实现起来比较复杂。条件均值的计算需要用到被估计量 θ 的概率密度 $p(\theta)$,如果并不知道概率密度 $p(\theta)$,而只知道 θ 的一、二阶矩特性,并且希望估计器能用线性系统实现,这时可以采用线性最小均方误差估计。

3.6.1 随机参量的线性最小均方误差估计

线性最小均方误差(Linear Minimum Mean Square Error, LMMSE)估计是一种使均方误差最小的线性估计。假定观测为 $\{x_n\}$, $n=0,1,\dots,N-1$,那么估计结果定义为观测的线性组合

$$\hat{\theta} = \sum_{n=0}^{N-1} a_n x_n + b \quad (3-203)$$

考虑实信号,估计的均方误差为

$$\text{MSE}(\hat{\theta}) = E[(\theta - \hat{\theta})^2] = E\left[\left(\theta - \sum_{n=0}^{N-1} a_n x_n - b\right)^2\right] \quad (3-204)$$

选择一组最佳系数 a_n 和 b ,使式(3-204)的均方误差达到最小。注意,引入系数 b 是因为观测 x 和被估计量 θ 的均值可能不为零。如果观测 x 和被估计量 θ 的均值都为零,那么系数 b 就可以省略。

均方误差对系数 b 求导,并令导数等于零,得

$$\frac{\partial \text{MSE}(\hat{\theta})}{\partial b} = -2E\left[\left(\theta - \sum_{n=0}^{N-1} a_n x_n - b\right)\right] = 0 \quad (3-205)$$

经整理后得

$$b = E(\theta) - \sum_{n=0}^{N-1} a_n E(x_n) \quad (3-206)$$

将式(3-206)代入式(3-204),得

$$\text{MSE}(\hat{\theta}) = E\left[\left(\sum_{n=0}^{N-1} a_n [x_n - E(x_n)] - [\theta - E(\theta)]\right)^2\right] \quad (3-207)$$

令 $\mathbf{a} = [a_0, a_1, \dots, a_{N-1}]^T$, $\mathbf{x} = [x_0, x_1, \dots, x_{N-1}]^T$,那么,式(3-203)可表示为

$$\hat{\theta} = \mathbf{a}^T \mathbf{x} + b \quad (3-208)$$

而式(3-206)可改写为

$$b = E(\theta) - \mathbf{a}^T E(\mathbf{x}) \quad (3-209)$$

估计的均方误差可表示为

$$\begin{aligned} \text{MSE}(\hat{\theta}) &= E\{[\mathbf{a}^T(\mathbf{x} - E(\mathbf{x})) - (\theta - E(\theta))]^2\} \\ &= E[\mathbf{a}^T(\mathbf{x} - E(\mathbf{x}))(\mathbf{x} - E(\mathbf{x}))^T \mathbf{a}] - E[\mathbf{a}^T(\mathbf{x} - E(\mathbf{x}))(\theta - E(\theta))] - \\ &\quad E[(\theta - E(\theta))(\mathbf{x} - E(\mathbf{x}))^T \mathbf{a}] + E[(\theta - E(\theta))^2] \\ &= \mathbf{a}^T \mathbf{C}_x \mathbf{a} - \mathbf{a}^T \mathbf{c}_{\theta x} - \mathbf{c}_{\theta x}^T \mathbf{a} + C_\theta \end{aligned} \quad (3-210)$$

其中, $\mathbf{C}_x = E\{[\mathbf{x} - E(\mathbf{x})][\mathbf{x} - E(\mathbf{x})]^T\}$ 为观测的协方差矩阵, $\mathbf{c}_{\theta x} = E\{[\theta - E(\theta)][\mathbf{x} - E(\mathbf{x})]\}$ 是待估计量 θ 与观测 $\mathbf{x}$ 的协方差向量, 且 C_θ 是 θ 的方差。将估计的均方误差对 $\mathbf{a}$ 求导, 得

$$\frac{\partial \text{MSE}(\hat{\theta})}{\partial \mathbf{a}} = 2\mathbf{C}_x \mathbf{a} - 2\mathbf{c}_{\theta x} \quad (3-211)$$

令导数等于零, 可解得

$$\mathbf{a} = \mathbf{C}_x^{-1} \mathbf{c}_{\theta x} \quad (3-212)$$

将式(3-212)和式(3-209)代入式(3-208), 得

$$\hat{\theta}_{\text{lmse}} = \mathbf{c}_{\theta x}^T \mathbf{C}_x^{-1} \mathbf{x} + E(\theta) - \mathbf{c}_{\theta x}^T \mathbf{C}_x^{-1} E(\mathbf{x}) \quad (3-213)$$

式(3-213)经整理后得到线性最小均方误差估计为

$$\hat{\theta}_{\text{lmse}} = E(\theta) + \mathbf{c}_{\theta x}^T \mathbf{C}_x^{-1} [\mathbf{x} - E(\mathbf{x})] \quad (3-214)$$

如果 θ 和 $\mathbf{x}$ 的均值为零, 则

$$\hat{\theta}_{\text{lmse}} = \mathbf{c}_{\theta x}^T \mathbf{C}_x^{-1} \mathbf{x} \quad (3-215)$$

将式(3-215)代入式(3-210)中, 可以得到最小均方误差的表达式

$$\text{MSE}(\hat{\theta}_{\text{lmse}}) = C_\theta - \mathbf{c}_{\theta x}^T \mathbf{C}_x^{-1} \mathbf{c}_{\theta x} \quad (3-216)$$

线性最小均方误差能得到简洁的估计子, 并且可定量分析估计误差, 因此应用广泛。

例 3-15 设有 N 次独立观测信号

$$x_n = A + w_n, \quad n = 0, 1, \dots, N-1 \quad (3-217)$$

其中, A 在 $(-A_0, A_0)$ 上服从均匀分布, w_n 是零均值高斯白噪声, 方差为 σ^2 , 且 A 与 w_n 统计独立, 求 A 的线性最小均方误差估计。

解 可以把观测信号模型写成向量形式

$$\mathbf{x} = \mathbf{A} \mathbf{e}_{N \times 1} + \mathbf{w} \quad (3-218)$$

其中, $\mathbf{e}_{N \times 1}$ 是 N 维的全 1 向量。根据题意, $E(A) = 0$, $\sigma_A^2 = E(A^2) = (2A_0)^2/12$, $E(\mathbf{x}) = 0$ 且

$$\mathbf{C}_x = E(\mathbf{x}\mathbf{x}^T) = E[(\mathbf{A}\mathbf{e} + \mathbf{w})(\mathbf{A}\mathbf{e} + \mathbf{w})^T] = E(A^2)\mathbf{e}\mathbf{e}^T + \sigma^2 \mathbf{I} = \sigma_A^2 \mathbf{e}\mathbf{e}^T + \sigma^2 \mathbf{I} \quad (3-219)$$

$$\mathbf{c}_{Ax} = E(A\mathbf{x}) = E[A(\mathbf{A}\mathbf{e} + \mathbf{w})] = E(A^2)\mathbf{e} = \sigma_A^2 \mathbf{e} \quad (3-220)$$

其中, $\mathbf{I}$ 为单位矩阵, 由式(3-215), 得

$$\hat{A}_{\text{lmse}} = \mathbf{c}_{Ax}^T \mathbf{C}_x^{-1} \mathbf{x} = \sigma_A^2 \mathbf{e}^T [\mathbf{e}\mathbf{e}^T \sigma_A^2 + \sigma^2 \mathbf{I}]^{-1} \mathbf{x} = \frac{\sigma_A^2}{\sigma^2} \mathbf{e}^T \left[\mathbf{e}\mathbf{e}^T \frac{\sigma_A^2}{\sigma^2} + \mathbf{I} \right]^{-1} \mathbf{x} \quad (3-221)$$

根据矩阵求逆中常用的 Sherman-Morrison 公式

$$(\mathbf{B} + \mathbf{c}\mathbf{d}^T)^{-1} = \mathbf{B}^{-1} - \frac{\mathbf{B}^{-1}\mathbf{c}\mathbf{d}^T\mathbf{B}^{-1}}{1 + \mathbf{d}^T\mathbf{B}^{-1}\mathbf{c}} \quad (3-222)$$

式(3-221)经整理后,得

$$\hat{A}_{\text{lmse}} = \frac{\sigma_A^2}{\sigma_A^2 + \sigma^2/N} \bar{x} = \frac{A_0^2/3}{A_0^2/3 + \sigma^2/N} \bar{x} \quad (3-223)$$

3.6.2 线性最小均方误差估计的几何解释

可以用向量空间的概念解释线性最小均方误差估计。在以下的讨论中假定被估计量 θ 和观测 x 都是零均值的,如果不是零均值,则总可以定义零均值的随机变量 $\theta' = \theta - E(\theta)$, $x' = x - E(x)$ 。

考虑一个由所有随机变量构成的线性空间(见图 3-4), θ 和 $x_0, x_1, \dots, x_{N-1}$ 可以看作线性空间的点(每个点由一个向量描述)。定义两个向量的内积为 $(x, y) = E(xy)$, 向量的长度为随机变量方差的均方根, 即 $\|x\| = \sqrt{E(x)^2}$, 如果 $(x, y) = E(xy) = 0$, 称这两个向量是正交的。由于随机变量是零均值的, 因此两个向量正交和随机变量不相关等价。线性最小均方误差估计的目的是用 $x_0, x_1, \dots, x_{N-1}$ 的线性组合逼近 θ 。先考虑只有一个样本的情况, 即用 $\hat{\theta} = a_0 x_0$ 逼近 θ , 易见最优的估计就是 θ 到 x_0 的投影(如图 3-4(a)所示), 这时误差为 θ 到 x_0 的垂直距离, 从几何的角度易知这时误差向量的长度最短。当使用两个样本 x_0, x_1 时, $\hat{\theta} = a_0 x_0 + a_1 x_1$, 注意到 x_0, x_1 的线性组合张成一个平面, 这时最优估计就是 θ 到该平面的投影(如图 3-4(b)所示), 同时误差向量 ϵ 正交于由 $\{x_0, x_1\}$ 所张成的平面, 易证这时误差向量的长度最短。

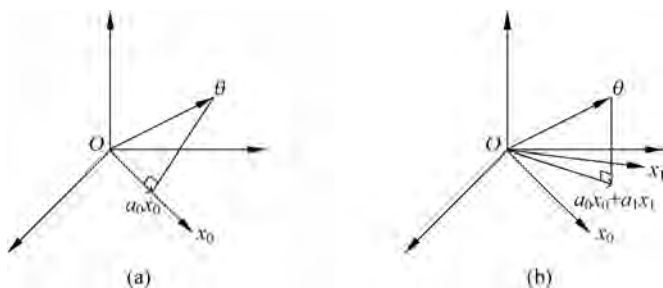


图 3-4 线性最小均方误差估计的几何解释

一般情况下, 线性估计子用 $x_0, x_1, \dots, x_{N-1}$ 的线性组合近似 θ , 相当于在 $x_0, x_1, \dots, x_{N-1}$ 所张成的子空间上找一个点, 使其与 θ 的距离最小。根据误差定义

$$\text{MSE}(\hat{\theta}) = E[(\theta - \hat{\theta})^2] = E\left[\left(\theta - \sum_{i=0}^{N-1} a_i x_i\right)^2\right] = \left\| \theta - \sum_{i=0}^{N-1} a_i x_i \right\|^2 \quad (3-224)$$

可见最小均方误差等价于误差向量 $\epsilon = \theta - \hat{\theta}$ 长度的平方最小。根据上面的论述, 这时误差向量将正交于 $x_0, x_1, \dots, x_{N-1}$ 所张成的子空间, 即

$$\epsilon \perp \text{span}\{x_0, x_1, \dots, x_{N-1}\} \quad (3-225)$$

根据正交的定义, 可以得出

$$E[(\theta - \hat{\theta})x_n] = 0, \quad n = 0, 1, \dots, N-1 \quad (3-226)$$

式(3-226)是线性最小均方误差估计非常重要的正交原理,也可由式(3-224)直接对加权系数求解得到。由正交原理可知,线性最小均方误差估计可以通过使估计误差和每个观测数据正交得到。利用正交原理,加权系数很容易求出

$$E \left[\left(\theta - \sum_{n=0}^{N-1} a_n x_n \right) x_j \right] = 0, \quad j = 0, 1, \dots, N-1 \quad (3-227)$$

或者写成

$$\sum_{n=0}^{N-1} a_n E(x_n x_j) = E(\theta x_j), \quad j = 0, 1, \dots, N-1 \quad (3-228)$$

用矩阵表示为

$$\begin{bmatrix} E(x_0^2) & E(x_0 x_1) & \cdots & E(x_0 x_{N-1}) \\ E(x_1 x_0) & E(x_1^2) & \cdots & E(x_1 x_{N-1}) \\ \vdots & \vdots & \ddots & \vdots \\ E(x_{N-1} x_0) & E(x_{N-1} x_1) & \cdots & E(x_{N-1}^2) \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{N-1} \end{bmatrix} = \begin{bmatrix} E(\theta x_0) \\ E(\theta x_1) \\ \vdots \\ E(\theta x_{N-1}) \end{bmatrix} \quad (3-229)$$

或者写成

$$\mathbf{C}_x \mathbf{a} = \mathbf{c}_{\theta x} \quad (3-230)$$

所以

$$\mathbf{a} = \mathbf{C}_x^{-1} \mathbf{c}_{\theta x} \quad (3-231)$$

θ 的线性最小均方误差估计为

$$\hat{\theta}_{\text{lmmse}} = \mathbf{a}^T \mathbf{x} = \mathbf{c}_{\theta x}^T \mathbf{C}_x^{-1} \mathbf{x} \quad (3-232)$$

式(3-232)与式(3-215)一致。

下面考虑随机向量的线性最小均方误差估计。设待估计量为 $\boldsymbol{\theta} = [\theta_0, \theta_1, \dots, \theta_{p-1}]^T$, 观测向量 $\mathbf{x} = [x_0, x_1, \dots, x_{N-1}]^T$, 线性估计可表示为

$$\hat{\boldsymbol{\theta}} = \mathbf{A} \mathbf{x} + \mathbf{b} \quad (3-233)$$

式中, $\mathbf{b}$ 是 $p \times 1$ 维的向量, $\mathbf{A}$ 是 $p \times N$ 维的矩阵, 所有估计的均方误差和可表示为

$$\text{MSE}(\hat{\boldsymbol{\theta}}) = E \left[\sum_{i=0}^{p-1} (\theta_i - \hat{\theta}_i)^2 \right] = E[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})] \quad (3-234)$$

将式(3-233)代入式(3-234), 得

$$\text{MSE}(\hat{\boldsymbol{\theta}}) = E[(\boldsymbol{\theta} - \mathbf{A} \mathbf{x} - \mathbf{b})^T (\boldsymbol{\theta} - \mathbf{A} \mathbf{x} - \mathbf{b})] \quad (3-235)$$

分别对 $\mathbf{A}$ 和 $\mathbf{b}$ 求导, 并令导数等于零, 得

$$\frac{\partial \text{MSE}(\hat{\boldsymbol{\theta}})}{\partial \mathbf{b}} = -2E[\boldsymbol{\theta} - \mathbf{A} \mathbf{x} - \mathbf{b}] = 0 \quad (3-236)$$

$$\frac{\partial \text{MSE}(\hat{\boldsymbol{\theta}})}{\partial \mathbf{A}} = -2E\{[\boldsymbol{\theta} - \mathbf{A} \mathbf{x} - \mathbf{b}] \mathbf{x}^T\} = 0 \quad (3-237)$$

式(3-237)中对矩阵的求导与式(3-72)相同, 定义为对矩阵每个元素求偏导数, 结果为矩阵, 且大小与 $\mathbf{A}$ 相同。由式(3-72)、式(3-237), 可得

$$\mathbf{b} = E(\boldsymbol{\theta}) - \mathbf{A} E(\mathbf{x}) \quad (3-238)$$

$$\mathbf{A} = \mathbf{C}_{\theta x} \mathbf{C}_x^{-1} \quad (3-239)$$

式中,

$$\mathbf{C}_{\theta x} = \text{Cov}(\boldsymbol{\theta}, \mathbf{x}) = E\{[\boldsymbol{\theta} - E(\boldsymbol{\theta})][\mathbf{x} - E(\mathbf{x})]^T\} \quad (3-240)$$

$$\mathbf{C}_x = \text{Cov}(\mathbf{x}) = E\{[\mathbf{x} - E(\mathbf{x})][\mathbf{x} - E(\mathbf{x})]^T\} \quad (3-241)$$

由此可得线性最小均方差估计为

$$\hat{\boldsymbol{\theta}}_{\text{lmmse}} = E(\boldsymbol{\theta}) + \mathbf{C}_{\theta x} \mathbf{C}_x^{-1} [\mathbf{x} - E(\mathbf{x})] \quad (3-242)$$

式(3-242)在表达形式上除 $\boldsymbol{\theta}$ 是向量外,其他与式(3-214)相同。线性最小均方差估计的性质总结如下(这里不做具体证明):

(1) 线性最小均方差估计是无偏估计。

(2) 线性最小均方差估计的均方误差阵要小于任何其他线性估计的均方误差阵,即

$$E[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{\text{lmmse}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{\text{lmmse}})^T] \leq E[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T] \quad (3-243)$$

式中, $\hat{\boldsymbol{\theta}}$ 是任意的线性估计。式(3-243)中符号 $\leq$ 作用于矩阵时,含义如下:如果 $\mathbf{B} \leq \mathbf{A}$,则 $\mathbf{A} - \mathbf{B}$ 是半正定矩阵。式(3-243)证明从略。

(3) 对于线性变换 $\boldsymbol{\alpha} = \mathbf{H}\boldsymbol{\theta} + \mathbf{b}$,线性最小均方差估计具有不变性,即

$$\hat{\boldsymbol{\alpha}}_{\text{lmmse}} = \mathbf{H}\hat{\boldsymbol{\theta}}_{\text{lmmse}} + \mathbf{b} \quad (3-244)$$

(4) 线性最小均方差估计具有叠加性,即如果 $\boldsymbol{\alpha} = \boldsymbol{\theta}_1 + \boldsymbol{\theta}_2$,则

$$\hat{\boldsymbol{\alpha}}_{\text{lmmse}} = \hat{\boldsymbol{\theta}}_{1, \text{lmmse}} + \hat{\boldsymbol{\theta}}_{2, \text{lmmse}} \quad (3-245)$$

3.7 最小二乘估计

在前面介绍的几种估计方法中,最小均方差估计、最大后验概率估计需要知道被估计量的先验概率密度,最大似然估计需要知道似然函数,线性最小均方差估计需要知道被估计量的一、二阶矩,如果这些概率密度或矩未知,就不能采用这些方法,这时可以采用最小二乘(Least Square, LS)估计。最小二乘估计对统计特性没有做任何假定,因此,它的应用非常广泛。

在线性最小均方差估计中,选择均方误差作为估计量性能好坏标准的度量,也就意味着估计与真实参数之间的差别平均来说达到最小。而在最小二乘估计中,则是使给定的观测数据与信号或者无噪声的观测数据之差的平方和达到最小。

采用例 3-10 的线性观测模型,待估计量为 $\boldsymbol{\theta} = [\theta_0, \theta_1, \dots, \theta_{p-1}]^T$,观测信号用向量和矩阵可表示为

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w} \quad (3-246)$$

式中,

$$\mathbf{x} = [x_0, x_1, \dots, x_{N-1}]^T, \quad \mathbf{w} = [w_0, w_1, \dots, w_{N-1}]^T$$

$$\mathbf{H} = \begin{bmatrix} h_{00} & h_{01} & \cdots & h_{0(p-1)} \\ h_{10} & h_{11} & \cdots & h_{1(p-1)} \\ \vdots & \vdots & \ddots & \vdots \\ h_{(N-1)0} & h_{(N-1)1} & \cdots & h_{(N-1)(p-1)} \end{bmatrix}$$

估计偏差的平方和可表示为

$$J(\hat{\theta}) = [\mathbf{x} - \mathbf{H}\hat{\theta}]^T [\mathbf{x} - \mathbf{H}\hat{\theta}] = \sum_{n=0}^{N-1} \left[x_n - \sum_{j=0}^{p-1} h_{nj} \hat{\theta}_j \right]^2 \quad (3-247)$$

在没有任何先验信息情况下,可以认为最优估计就是使得上述的平方和误差最小。当待估计参数的重要性存在差别时,可以对每个参数的误差进行加权,观测与估计偏差的加权平方和可表示为

$$J_w(\hat{\theta}) = [\mathbf{x} - \mathbf{H}\hat{\theta}]^T \mathbf{W} [\mathbf{x} - \mathbf{H}\hat{\theta}] = \sum_{j=0}^{N-1} \sum_{n=0}^{N-1} \left[x_n - \sum_{k=0}^{p-1} h_{nk} \hat{\theta}_k \right] W_{nj} \left[x_n - \sum_{k=0}^{p-1} h_{jk} \hat{\theta}_k \right] \quad (3-248)$$

其中 $\mathbf{W}$ 为加权矩阵。最小二乘估计就是使 $J(\hat{\theta})$ 最小的估计,记为 $\hat{\theta}_{ls}$; 加权最小二乘 (Weighted Least Square) 估计则是使 $J_w(\hat{\theta})$ 最小的估计,记为 $\hat{\theta}_{lsw}$ 。

接下来考虑如何求解,将式(3-247)重写为

$$J(\hat{\theta}) = \mathbf{x}^T \mathbf{x} - \mathbf{x}^T \mathbf{H} \hat{\theta} - \hat{\theta}^T \mathbf{H}^T \mathbf{x} + \hat{\theta}^T \mathbf{H}^T \mathbf{H} \hat{\theta} \quad (3-249)$$

根据式(3-73)和式(3-74)的向量函数求导公式,得到

$$\frac{\partial \mathbf{x}^T \mathbf{x}}{\partial \hat{\theta}} = 0 \quad (3-250)$$

$$\frac{\partial \mathbf{x}^T \mathbf{H} \hat{\theta}}{\partial \hat{\theta}} = \mathbf{H}^T \mathbf{x} \quad (3-251)$$

$$\frac{\partial \hat{\theta}^T \mathbf{H}^T \mathbf{x}}{\partial \hat{\theta}} = \mathbf{H}^T \mathbf{x} \quad (3-252)$$

$$\frac{\partial \hat{\theta}^T \mathbf{H}^T \mathbf{H} \hat{\theta}}{\partial \hat{\theta}} = 2 \mathbf{H}^T \mathbf{H} \hat{\theta} \quad (3-253)$$

求 $J(\hat{\theta})$ 对 $\hat{\theta}$ 的导数,并令导数等于零,得

$$\frac{\partial J(\hat{\theta})}{\partial \hat{\theta}} = -2 \mathbf{H}^T \mathbf{x} + 2 \mathbf{H}^T \mathbf{H} \hat{\theta} = 0 \quad (3-254)$$

由此可解得最小二乘估计为

$$\hat{\theta}_{ls} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x} \quad (3-255)$$

求 $J_w(\hat{\theta})$ 对 $\hat{\theta}$ 的导数,并令导数等于零,得

$$\frac{\partial J_w(\hat{\theta})}{\partial \hat{\theta}} = -2 \mathbf{H}^T \mathbf{W} [\mathbf{x} - \mathbf{H}\hat{\theta}] = 0 \quad (3-256)$$

由此可解得加权最小二乘估计为

$$\hat{\theta}_{lsw} = (\mathbf{H}^T \mathbf{W} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{W} \mathbf{x} \quad (3-257)$$

最小二乘估计具有如下特点:

(1) 对于线性的观测模型,最小二乘估计和加权最小二乘估计都是线性估计,对测量噪声的统计没有做任何假定,应用十分广泛。

(2) 当测量噪声的均值为零时,即 $E(w_i) = 0$ 时,最小二乘估计和加权最小二乘估计都

是无偏估计。

(3) 对于加权最小二乘估计,如果有一些模型的知识,如 $E(\mathbf{w})=0, E[\mathbf{w}\mathbf{w}^T]=\mathbf{R}$, 当 $\mathbf{W}=\mathbf{R}^{-1}$ 时,估计误差的方差达到最小,这时加权最小二乘估计和式(3-151)的最大似然估计是一致的。

3.8 信号检测基础

信号检测的目的是在干扰、噪声背景下判断目标信号是否存在。当存在多种类型目标信号时,还需要判断信号类型。根据目标信号模型的不同,信号检测分确定性信号检测和随机信号检测进行讨论,本节只讨论实信号。

3.8.1 确定性信号检测

这里以确定性信号为例介绍信号检测的基本概念和方法。假设目标信号 s_n 的波形已知,得到了接收信号 x_n 的 N 个样本,在此基础上判决其中是否包含目标信号,可定义如下两种事件(或称假设)。

$$\begin{aligned} H_0: x_n &= w_n, \quad n=0,1,\dots,N-1 \\ H_1: x_n &= s_n + w_n, \quad n=0,1,\dots,N-1 \end{aligned} \quad (3-258)$$

接下来考虑如何对上述事件进行检验,其过程可以看成:将观测空间划分成不同判决区域,观测样本 $\mathbf{x}=[x_0, x_1, \dots, x_{N-1}]$ 落在哪个区域就判决成哪个相应事件。一般可基于事件的概率来划分判决区域,定义两种事件的后验概率分别为 $p(H_1|\mathbf{x}), p(H_0|\mathbf{x})$, 后验概率给出了在得到观测样本后,对 H_0, H_1 事件发生概率的认知。一种直观的判决准则是,在观测样本的基础上,哪个事件发生的概率大就认为发生了该事件,即当 $p(H_1|\mathbf{x}) > p(H_0|\mathbf{x})$ 时,认为 H_1 事件(假设)为真;反之,若 $p(H_1|\mathbf{x}) < p(H_0|\mathbf{x})$ 时,则认为 H_0 为真。若 $p(H_1|\mathbf{x}) = p(H_0|\mathbf{x})$,则可随机认定某一事件为真。于是,可以把 H_1 事件的判决区域定义为所有使得 $p(H_1|\mathbf{x}) > p(H_0|\mathbf{x})$ 的观测空间。具体的数学描述如式(3-259)所示。

$$H_1: \frac{p(H_1|\mathbf{x})}{p(H_0|\mathbf{x})} > 1 \quad (3-259)$$

式(3-259)中的后验概率一般难以直接计算。基于贝叶斯公式有

$$p(H_1|\mathbf{x})p(\mathbf{x}) = p(\mathbf{x}|H_1)p(H_1) \quad (3-260)$$

式(3-259)等价于

$$H_1: L(\mathbf{x}) = \frac{p(\mathbf{x}|H_1)}{p(\mathbf{x}|H_0)} \geq \gamma \quad (3-261)$$

其中, $p(\mathbf{x}|H_0), p(\mathbf{x}|H_1)$ 是 3.4 节最大似然估计中提到的似然函数, $L(\mathbf{x})$ 称为似然比(Likelihood Ratio, LR), 有时为了计算便利,使用其对数形式 $\ln[L(\mathbf{x})]$, 称为对数似然比(Logarithm Likelihood Ratio, LLR)。 $\gamma = p(H_0)/p(H_1)$ 表示判决门限。

计算似然函数需要知道 x_n 的概率密度,实际中可通过假设噪声符合某种概率分布实现,这里以最常见的假设: w_n 是零均值、方差为 σ^2 的高斯白噪声为例,分析如何具体应用式(3-261)进行检验。根据假设可得到如下似然函数:

$$p(\mathbf{x} | H_1) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x_n - s_n)^2 \right] \quad (3-262)$$

$$p(\mathbf{x} | H_0) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} x_n^2 \right] \quad (3-263)$$

利用式(3-263)可计算如下似然比和对数似然比:

$$L(\mathbf{x}) = \exp \left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} ((x_n - s_n)^2 - x_n^2) \right] \quad (3-264)$$

$$l(\mathbf{x}) = \ln L(\mathbf{x}) = -\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} ((x_n - s_n)^2 - x_n^2) \quad (3-265)$$

采用与式(3-261)等价的检验准则: $\ln L(\mathbf{x}) > \ln \gamma$, 经简单计算, 可得式(3-261)等价于对式(3-266)中 $T(\mathbf{x})$ 的检验。

$$H_1: T(\mathbf{x}) = \sum_{n=0}^{N-1} x_n s_n = \mathbf{x}^T \mathbf{s} > \bar{\gamma} \quad (3-266)$$

其中

$$\bar{\gamma} = \sigma^2 \ln \gamma + \frac{1}{2} \sum_{n=0}^{N-1} s_n^2 \quad (3-267)$$

注意到 $T(\mathbf{x})$ 可以看作匹配滤波器的输出。确定性信号的匹配滤波器定义如下:

$$h(n) = \begin{cases} s(N-1-n), & n=0, 1, \dots, N-1 \\ 0, & \text{其他} \end{cases} \quad (3-268)$$

匹配滤波器的输出为 $y(n) = x(n) \otimes h(n)$, 并且可验证 $T(\mathbf{x}) = y(N-1)$ 。

在已知噪声概率特性情况下, 上述分析均可用, 下面将上述推导拓展到高斯有色噪声下的信号检测问题。记 $\mathbf{x} = [x_0, x_1, \dots, x_{N-1}]^T$, $\mathbf{s} = [s_0, s_1, \dots, s_{N-1}]^T$, $\mathbf{w} = [w_0, w_1, \dots, w_{N-1}]^T$, 检测问题可描述为

$$\begin{aligned} H_0: \mathbf{x} &= \mathbf{w}, \quad \mathbf{w} \propto N(0, \mathbf{C}) \\ H_1: \mathbf{x} &= \mathbf{s} + \mathbf{w} \end{aligned} \quad (3-269)$$

判决准则重写为

$$H_1: L(\mathbf{x}) = \frac{p(\mathbf{x} | H_1)}{p(\mathbf{x} | H_0)} > \gamma \quad (3-270)$$

可得到似然函数、似然比及对数似然比分别为

$$p(\mathbf{x} | H_1) = \frac{1}{(2\pi)^{\frac{N}{2}} \det^{\frac{1}{2}}(\mathbf{C})} \exp \left[-\frac{1}{2} (\mathbf{x} - \mathbf{s})^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{s}) \right] \quad (3-271)$$

$$p(\mathbf{x} | H_0) = \frac{1}{(2\pi)^{\frac{N}{2}} \det^{\frac{1}{2}}(\mathbf{C})} \exp \left[-\frac{1}{2} \mathbf{x}^T \mathbf{C}^{-1} \mathbf{x} \right] \quad (3-272)$$

$$l(\mathbf{x}) = \ln L(\mathbf{x}) = -\frac{1}{2} [(\mathbf{x} - \mathbf{s})^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{s}) - \mathbf{x}^T \mathbf{C}^{-1} \mathbf{x}] = \mathbf{x}^T \mathbf{C}^{-1} \mathbf{s} - \frac{1}{2} \mathbf{s}^T \mathbf{C}^{-1} \mathbf{s} \quad (3-273)$$

利用 $\ln L(\mathbf{x}) > \ln \gamma$ 准则进行判决, 得到

$$H_1: T(\mathbf{x}) = \mathbf{x}^T \mathbf{C}^{-1} \mathbf{s} > \gamma + \frac{1}{2} \mathbf{s}^T \mathbf{C}^{-1} \mathbf{s} = \bar{\gamma} \quad (3-274)$$

同理,可以由 $\mathbf{C}^{-1} \mathbf{s}$ 构造滤波器,而 $T(\mathbf{x})$ 是滤波器某个时刻的输出,这种滤波器称为广义匹配滤波器。

$T(\mathbf{x}) > \bar{\gamma}$ 定义了观测空间的一种划分规则,对所有满足 $T(\mathbf{x}) > \bar{\gamma}$ 的观测 $\mathbf{x}$ 计算积分可以得到 H_1 成功检测的概率。根据事件的不同,可定义 4 种不同概率,其定义和计算方式如表 3-1 所示。

表 3-1 二元信号检测中的事件和概率计算

事 件	定 义	标 号
H_1 为真,判 H_1 成立	H_1 检测概率	$P(H_1 H_1) = \int_{T(\mathbf{x}) > \bar{\gamma}} p(\mathbf{x} H_1) d\mathbf{x}$
H_0 为真,判 H_0 成立	H_0 检测概率	$1 - P(H_1 H_0)$
H_0 为真,判 H_1 成立	虚警概率	$P(H_1 H_0) = \int_{T(\mathbf{x}) > \bar{\gamma}} p(\mathbf{x} H_0) d\mathbf{x}$
H_1 为真,判 H_0 成立	漏检概率	$1 - P(H_1 H_1)$

在实际应用中表 3-1 中不同的事件会有不同的代价,例如在雷达信号处理中,漏检造成的代价远大于虚警造成的代价。考虑更一般化的情况,为表中 4 种不同事件定义各自的代价因子,并定义如下平均代价:

$$C = [C_{00}P(H_0 | H_0) + C_{01}P(H_1 | H_0)]P(H_0) + [C_{10}P(H_0 | H_1) + C_{11}P(H_1 | H_1)]P(H_1) \quad (3-275)$$

其中, $C_{00}, C_{01}, C_{10}, C_{11}$ 分别为各事件的代价因子,经简单计算可得

$$\begin{aligned} C &= [C_{00}(1 - P(H_1 | H_0)) + C_{01}P(H_1 | H_0)]P(H_0) + \\ &\quad [C_{10}(1 - P(H_1 | H_1)) + C_{11}P(H_1 | H_1)]P(H_1) \\ &= C_{00}P(H_0) + (C_{01} - C_{00})P(H_0)P(H_1 | H_0) + \\ &\quad C_{10}P(H_1) - (C_{10} - C_{11})P(H_1)P(H_1 | H_1) \\ &= C_{00}P(H_0) + C_{10}P(H_1) + \\ &\quad \int_{\mathbf{x} \in Z_1} [(C_{01} - C_{00})P(H_0)P(\mathbf{x} | H_0) - (C_{10} - C_{11})P(H_1)P(\mathbf{x} | H_1)] d\mathbf{x} \end{aligned} \quad (3-276)$$

其中, Z_1 表示判决 H_1 成立的判决区域。可见要使代价最小,在判决区域内的积分函数取值为负,即在判决区域内满足

$$(C_{01} - C_{00})P(H_0)P(\mathbf{x} | H_0) < (C_{10} - C_{11})P(H_1)P(\mathbf{x} | H_1) \quad (3-277)$$

$$\frac{P(\mathbf{x} | H_1)}{P(\mathbf{x} | H_0)} > \frac{(C_{01} - C_{00})P(H_0)}{(C_{10} - C_{11})P(H_1)} \quad (3-278)$$

式(3-278)称为信号检测的贝叶斯准则。从式(3-278)可以看到,如果认为正确的检测不产生额外代价,即式(3-275)中 $C_{00} = C_{11} = 0$,同时虚警和漏检有相同的代价,即式(3-275)

中 $C_{01}=C_{10}=1$, 则式(3-278)的检测准则等价于式(3-261)。

以上的贝叶斯准则需要确定代价因子和先验概率, 实际中往往很难确定, 特别是先验概率的确定。实际中往往对虚警概率和漏检概率有明确的要求, 可以从虚警概率和漏检概率的约束考虑信号检测。一种常见的准则是在虚警概率恒定的约束下最小化漏检概率, 称为纽曼-皮儿森准则(Neyman-Pearson, NP)。

令 P_F 、 P_M 分别表示虚警概率和漏检概率, α 表示允许的虚警概率上限, Z_0, Z_1 分别表示信号检测的判决区域, 问题模型可表示为

$$\min_{Z_0, Z_1} P_M$$

$$P_F = \alpha \quad (3-279)$$

即在满足虚警概率要求的条件下, 寻找观测空间的划分准则, 使得漏检概率尽量小。可利用拉格朗日乘子法求解上述问题, 构造如下目标函数

$$J = P_M + \lambda(P_F - \alpha) \quad (3-280)$$

其中, λ 为拉格朗日乘子。目的是要确定一种观测空间的划分, 使 J 最小, 将虚警概率和漏检概率的计算表达式代入, 得到

$$J = \int_{Z_0} p(\mathbf{x} | H_1) d\mathbf{x} + \lambda \left(\int_{Z_1} p(\mathbf{x} | H_0) d\mathbf{x} - \alpha \right)$$

$$= \lambda(1 - \alpha) + \int_{Z_0} [p(\mathbf{x} | H_1) - \lambda p(\mathbf{x} | H_0)] d\mathbf{x} \quad (3-281)$$

式(3-281)中第一项非负, 要使 J 尽可能小, 则式(3-281)积分结果尽可能小。一种直观的做法是将所有使 $p(\mathbf{x} | H_1) - \lambda p(\mathbf{x} | H_0) < 0$ 的观测空间归入 Z_0 , 或将使 $p(\mathbf{x} | H_1) > \lambda p(\mathbf{x} | H_0)$ 成立的观测空间定义为 Z_1 , 从而得到如下 H_1 的检测准则:

$$H_1: \frac{p(\mathbf{x} | H_1)}{p(\mathbf{x} | H_0)} > \lambda \quad (3-282)$$

式(3-282)中 λ 为未知量, 不能直接应用。 λ 可根据约束条件确定, 但难以得到闭式解。一般可采用迭代方式求解, 即从 λ 的一个初始值开始, 由式(3-282)得到观测空间的分割 Z_0 、 Z_1 , 再根据如下约束条件进行验证。如果满足, 则停止迭代; 否则, 调整 λ 取值重新验证。

$$\int_{Z_1} p(\mathbf{x} | H_0) d\mathbf{x} = P_F \quad (3-283)$$

3.8.2 随机信号检测

随机信号检测所采用的准则与确定性信号检测一致。但由于对随机信号只能知道其统计特征, 因此具体计算有所不同。下面看一个简单例子, 考虑以下两种假设。

$$H_0: \mathbf{x} = \mathbf{w}, \quad \mathbf{w} \sim N(0, \sigma^2 \mathbf{I})$$

$$H_1: \mathbf{x} = \mathbf{s} + \mathbf{w} \quad (3-284)$$

其中, s_n 是平稳随机信号。先考虑一个简单情况, s_n 是高斯信号, 均值为零, 方差为 σ_s^2 , 以上两种假设可描述为

$$H_0: \mathbf{x} \sim N(0, \sigma^2 \mathbf{I})$$

$$H_1: \mathbf{x} \sim N(0, (\sigma_s^2 + \sigma^2) \mathbf{I}) \quad (3-285)$$

采用似然比及对数似然比进行判断, 得到如下 H_1 的判决准则:

$$H_1: L(\mathbf{x}) = \frac{p(\mathbf{x} | H_1)}{p(\mathbf{x} | H_0)} > \gamma \quad (3-286)$$

得到似然函数及似然比如下:

$$P(\mathbf{x} | H_1) = \frac{1}{(2\pi(\sigma_s^2 + \sigma^2))^{\frac{N}{2}}} \exp\left\{-\frac{1}{2(\sigma_s^2 + \sigma^2)} \sum_{n=0}^{N-1} x_n^2\right\} \quad (3-287)$$

$$P(\mathbf{x} | H_0) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} x_n^2\right\} \quad (3-288)$$

$$l(\mathbf{x}) = \ln L(\mathbf{x}) = \frac{N}{2} \ln\left(\frac{\sigma^2}{\sigma_s^2 + \sigma^2}\right) + \frac{1}{2} \frac{\sigma_s^2}{\sigma^2(\sigma_s^2 + \sigma^2)} \sum_{n=0}^{N-1} x_n^2 \quad (3-289)$$

采用 $l(\mathbf{x}) > \ln \gamma$ 判决准则, 对上式略加整理, 可得到如下等效判决准则:

$$H_1: T(\mathbf{x}) = \sum_{n=0}^{N-1} x_n^2 = \mathbf{x}^T \mathbf{x} > \left[\ln \gamma - \frac{N}{2} \ln\left(\frac{\sigma^2}{\sigma_s^2 + \sigma^2}\right) \right] \frac{2\sigma^2(\sigma_s^2 + \sigma^2)}{\sigma_s^2} = \bar{\gamma} \quad (3-290)$$

可以看到 H_1 的判决依赖于采样信号的能量。当能量大于门限值时, 判断存在目标信号, 因此上述检测方法又称为能量检测器。能量检测器可推广到更一般的情况, 即检测信号具有任意的协方差矩阵 $\mathbf{C}_s$, 考虑以下两种假设

$$\begin{aligned} H_0: \mathbf{x} &\sim N(0, \sigma^2 \mathbf{I}) \\ H_1: \mathbf{x} &\sim N(0, \mathbf{C}_s + \sigma^2 \mathbf{I}) \end{aligned} \quad (3-291)$$

得到似然函数及似然比如下:

$$P(\mathbf{x} | H_1) = \frac{1}{(2\pi)^{\frac{N}{2}} \det^{\frac{1}{2}}(\mathbf{C}_s + \sigma^2 \mathbf{I})} \exp\left[-\frac{1}{2} \mathbf{x}^T (\mathbf{C}_s + \sigma^2 \mathbf{I})^{-1} \mathbf{x}\right] \quad (3-292)$$

$$P(\mathbf{x} | H_0) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2} \mathbf{x}^T \mathbf{x}\right] \quad (3-293)$$

$$l(\mathbf{x}) = \ln L(\mathbf{x}) = \frac{1}{2} \ln\left(\frac{(\sigma^2)^N}{\det(\mathbf{C}_s + \sigma^2 \mathbf{I})}\right) - \frac{1}{2} \mathbf{x}^T [(\mathbf{C}_s + \sigma^2 \mathbf{I})^{-1} - \frac{1}{\sigma^2} \mathbf{I}] \mathbf{x} \quad (3-294)$$

可得到如下等价于 $l(\mathbf{x}) > \ln \gamma$ 的判决准则:

$$H_1: T(\mathbf{x}) = \sigma^2 \mathbf{x}^T \left[\frac{1}{\sigma^2} \mathbf{I} - (\mathbf{C}_s + \sigma^2 \mathbf{I})^{-1} \right] \mathbf{x} > \bar{\gamma} \quad (3-295)$$

利用矩阵求逆定理

$$(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{B} (\mathbf{DA}^{-1} \mathbf{B} + \mathbf{C}^{-1})^{-1} \mathbf{DA}^{-1} \quad (3-296)$$

在式(3-296)中, 令 $\mathbf{A} = \sigma^2 \mathbf{I}$, $\mathbf{B} = \mathbf{D} = \mathbf{I}$, $\mathbf{C} = \mathbf{C}_s$, 有

$$(\sigma^2 \mathbf{I} + \mathbf{C}_s)^{-1} = \frac{1}{\sigma^2} \mathbf{I} - \frac{1}{\sigma^4} \left(\frac{1}{\sigma^2} \mathbf{I} + \mathbf{C}_s^{-1} \right)^{-1} \quad (3-297)$$

代入式(3-295), 得到如下判决准则:

$$\begin{aligned} H_1: T(\mathbf{x}) &= \mathbf{x}^T \left[\frac{1}{\sigma^2} \left(\mathbf{C}_s^{-1} + \frac{1}{\sigma^2} \mathbf{I} \right)^{-1} \right] \mathbf{x} = \mathbf{x}^T (\sigma^2 \mathbf{C}_s^{-1} + \mathbf{I})^{-1} \mathbf{x} \\ &= \mathbf{x}^T [(\sigma^2 \mathbf{I} + \mathbf{C}_s) \mathbf{C}_s^{-1}]^{-1} \mathbf{x} = \mathbf{x}^T \mathbf{C}_s (\sigma^2 \mathbf{I} + \mathbf{C}_s)^{-1} \mathbf{x} > \bar{\gamma} \end{aligned} \quad (3-298)$$

和式(3-290)相比, 式(3-298)相当于在能量检测器基础上引入加权矩阵 $\mathbf{W} = \mathbf{C}_s (\sigma^2 \mathbf{I} +$

$\mathbf{C}_s)^{-1}$, 得到加权检测量 $T(\mathbf{x}) = \mathbf{x}^T \mathbf{W} \mathbf{x}$, 称为广义能量检测器。注意到式(3-298)可写成

$$\begin{aligned} H_1: T(\mathbf{x}) &= \mathbf{x}^T \hat{\mathbf{s}} > \bar{\gamma} \\ \hat{\mathbf{s}} &= \mathbf{C}_s (\sigma^2 \mathbf{I} + \mathbf{C}_s)^{-1} \mathbf{x} \end{aligned} \quad (3-299)$$

信号检测的过程可以看成先得到目标信号的一个估计, 然后再利用估计得到的信号构造一个如式(3-266)所示的匹配滤波器。注意到上式中的 $\hat{\mathbf{s}}$ 实际是目标信号的线性最小均方误差(LMMSE)估计。说明如下: 根据 3.6 节的结论, 基于观测 $\mathbf{x}$, 目标参数 $\boldsymbol{\theta}$ 的线性 MMSE 估计为

$$\hat{\boldsymbol{\theta}} = \mathbf{C}_{\theta x} \mathbf{C}_x^{-1} \mathbf{x} \quad (3-300)$$

将式(3-300)用于从 $\mathbf{x}$ 中估计 $\mathbf{s}$, 可得

$$\mathbf{C}_{sx} = E(\mathbf{s} \mathbf{x}^T) = E(\mathbf{s} \mathbf{s}^T + \mathbf{s} \mathbf{w}^T) = E(\mathbf{s} \mathbf{s}^T) = \mathbf{C}_s \quad (3-301)$$

$$\begin{aligned} \mathbf{C}_x &= E(\mathbf{x} \mathbf{x}^T) = E(\mathbf{s} \mathbf{s}^T + \mathbf{s} \mathbf{w}^T + \mathbf{w} \mathbf{s}^T + \mathbf{w} \mathbf{w}^T) \\ &= E(\mathbf{s} \mathbf{s}^T) + E(\mathbf{w} \mathbf{w}^T) \\ &= \mathbf{C}_s + \sigma^2 \mathbf{I} \end{aligned} \quad (3-302)$$

式(3-302)中假设目标信号和噪声不相关, 这在实际中一般可以满足。

以上讨论都基于假设在 H_0 和 H_1 条件下的似然函数是可知的, 但实际情况经常没那么理想。一个最简单的例子是描述噪声特性的参数, 如方差, 经常是未知的。目标信号参数也有可能是未知的, 例如在无源雷达和声呐系统中, 接收机对来的波信号的频率是未知的。估计存在未知参数时, 一般分两个步骤处理, 首先是解决 H_0 和 H_1 条件下的未知参数估计问题, 例如用最大似然准则

$$\hat{\boldsymbol{\theta}}_0 = \max_{\boldsymbol{\theta}_0} p(\mathbf{x} | \boldsymbol{\theta}_0, H_0) \quad \hat{\boldsymbol{\theta}}_1 = \max_{\boldsymbol{\theta}_1} p(\mathbf{x} | \boldsymbol{\theta}_1, H_1) \quad (3-303)$$

然后将估计结果用于似然比准则进行判决

$$H_1: L(\mathbf{x}) = \frac{p(\mathbf{x} | \hat{\boldsymbol{\theta}}_1, H_1)}{p(\mathbf{x} | \hat{\boldsymbol{\theta}}_0, H_0)} > \gamma \quad (3-304)$$

这种方法称为广义似然比准则。如果先验概率分布已知, 则可以使用贝叶斯准则计算似然比

$$H_1: L(\mathbf{x}) = \frac{p(\mathbf{x} | H_1)}{p(\mathbf{x} | H_0)} = \frac{\int p(\mathbf{x} | \boldsymbol{\theta}_1, H_1) d\boldsymbol{\theta}_1}{\int p(\mathbf{x} | \boldsymbol{\theta}_0, H_0) d\boldsymbol{\theta}_0} > \gamma \quad (3-305)$$

下面以一个例子进一步说明, 基于式(3-290)且假设目标信号具有非零均值

$$\begin{aligned} H_0: \mathbf{x} &\propto N(0, \sigma^2 \mathbf{I}) \\ H_1: \mathbf{x} &\propto N(\mathbf{u} \mathbf{e}, \mathbf{C}_s + \sigma^2 \mathbf{I}) \end{aligned} \quad (3-306)$$

其中, $\mathbf{e}$ 是 $N \times 1$ 全 1 向量, 得到似然函数如下:

$$P(\mathbf{x} | H_1) = \frac{1}{(2\pi)^{\frac{N}{2}} \det^{\frac{1}{2}}(\mathbf{C}_s + \sigma^2 \mathbf{I})} \exp \left[-\frac{1}{2} (\mathbf{x} - \mathbf{u} \mathbf{e})^T (\mathbf{C}_s + \sigma^2 \mathbf{I})^{-1} (\mathbf{x} - \mathbf{u} \mathbf{e}) \right] \quad (3-307)$$

$$P(\mathbf{x} | H_0) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma^2} \mathbf{x}^T \mathbf{x} \right] \quad (3-308)$$

u 的最大似然估计为

$$\hat{u} = \frac{1}{N} \mathbf{x}^T (\mathbf{C}_s + \sigma^2 \mathbf{I})^{-1} \mathbf{e} \quad (3-309)$$

和式(3-295)的推导相似,可得到以下判决准则

$$\begin{aligned} T(\mathbf{x}) &= -(\mathbf{x} - \hat{u}\mathbf{e})^T (\mathbf{C}_s + \sigma^2 \mathbf{I})^{-1} (\mathbf{x} - \hat{u}\mathbf{e}) + \frac{1}{\sigma^2} \mathbf{x}^T \mathbf{x} > \bar{\gamma} \\ &\Leftrightarrow \mathbf{x}^T \left[(\mathbf{C}_s + \sigma^2 \mathbf{I})^{-1} - \frac{1}{\sigma^2} \mathbf{I} \right] \mathbf{x} + N\hat{u}^2 > \bar{\gamma} \\ &\Leftrightarrow \mathbf{x}^T \mathbf{C}_s (\sigma^2 \mathbf{I} + \mathbf{C}_s)^{-1} \mathbf{x} + N\hat{u}^2 > \bar{\gamma} \end{aligned} \quad (3-310)$$

和式(3-295)相比,式(3-310)的不同之处在于引入了非零均值的影响。

上述讨论可推广到多种信号的检测,考虑如下模型:

$$\begin{aligned} H_0: \mathbf{x} &= \mathbf{w}, \quad \mathbf{w} \sim N(0, \mathbf{C}) \\ H_i: \mathbf{x} &= \mathbf{s}_i + \mathbf{w}, \quad i = 1, 2, \dots, K \end{aligned} \quad (3-311)$$

其中, K 表示有多少种不同的信号。为了实现最低的检测错误概率,可采用最大后验概率检测准则

$$H_i: p(H_i | \mathbf{x}) > p(H_j | \mathbf{x}), \quad j = 0, 1, \dots, K; j \neq i \quad (3-312)$$

假设各种假设具有相等概率,可采用最大似然检测准则

$$H_i: p(\mathbf{x} | H_i) > p(\mathbf{x} | H_j), \quad j = 0, 1, \dots, K, j \neq i \quad (3-313)$$

在实际应用中经常先计算各种假设对 H_0 的似然比

$$L_i(\mathbf{x}) = \frac{p(\mathbf{x} | H_i)}{p(\mathbf{x} | H_0)}, \quad i = 1, 2, \dots, K \quad (3-314)$$

如果所有 $L_i(\mathbf{x})$ 的取值均小于 1, 则判决 H_0 为真; 否则, 判决具有最大似然比的假设为真。特别要指出的是, 上述方法只能应用于各种假设具有相等概率的情况, 如果概率不相等, 则必须直接基于式(3-312)计算。

本章习题

1. 假定观测 $x_0, x_1, \dots, x_{N-1}$ 是独立同分布的, 且在 $(0, \theta)$ 上服从均匀分布, 求 θ 的最大似然估计。

2. 证明式(3-39)

$$I(\theta) = E \left[\left(\frac{\partial \ln p(\mathbf{x} | \theta)}{\partial \theta} \right)^2 \right] = -E \left[\frac{\partial^2 \ln p(\mathbf{x} | \theta)}{\partial \theta^2} \right]$$

3. 设 N 次观测为 $x_i = A + w_i, i = 0, 1, \dots, N-1$, 其中, A 为未知的确定信号, 噪声 $w_i (i = 0, 1, \dots, N-1)$ 相互独立并服从相同的分布 $N(0, \sigma^2)$, 噪声的方差已知。

(1) 求 A 的最大似然估计。

(2) A 的最大似然估计是否为无偏估计?

(3) A 的最大似然估计是否为有效估计? 估计的方差等于多少?

4. 在习题 3 中, 如果 A 和 σ^2 都是未知的, 试分别求其最大似然估计, 并讨论它们的无偏性。

5. 设观测模型为 $x_i = A + w_i$, 其中 $w_i \sim N(0, \sigma^2)$, 如果 A 为未知常量, 求 A 的最大后验概率估计和最小均方误差估计。

6. 设有 N 次独立同分布的观测 $x_0, x_1, \dots, x_{N-1}$, 且单个样本在 $(0, \theta)$ 内服从均匀分布, 证明:

$$E\left(\frac{\partial \ln p(\mathbf{x} | \theta)}{\partial \theta}\right) \neq 0, \quad \theta > 0$$

7. 平坦信道的信道估计问题可以用下式描述

$$x_i = A s_i + w_i, \quad i = 0, 1, \dots, N-1$$

其中, w_i 是零均值高斯白噪声, 方差为 σ^2 , s_i 是已知的训练序列, x_i 是接收信号, 求估计 A 的 CRB。证明有效估计量存在, 并求它的方差。当 $N \rightarrow \infty$ 时, 估计的方差会怎样变化?

8. 设有两次观测

$$x_0 = A s_0 + w_0$$

$$x_1 = A s_1 + w_1$$

其中, $\mathbf{w} = [w_0, w_1]^T$ 是零均值高斯随机向量, 协方差矩阵为

$$\mathbf{C} = \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

求估计 A 的 CRB, 并将它与 $\mathbf{w}$ 为高斯白噪声的情况 ($\rho = 0$) 进行比较, 解释当 $\rho = \pm 1$ 时会怎样?

9. 考虑高斯噪声下的单频频率估计问题

$$x_i = A e^{j\omega n + \varphi} + w_i, \quad i = 0, 1, \dots, N-1$$

其中, w_i 是复高斯白噪声, 方差为 σ^2 。 A, ω, ϕ 分别为幅度、频率和相位, 求:

(1) A, ϕ 为已知参数时频率估计的 CRB;

(2) A, ω, ϕ 都是未知参数时频率估计的 CRB。

10. 设观测的概率密度为 $p(x | \theta)$, θ 为未知常量, 待估计量为 $\alpha = g(\theta)$, 证明: 估计 α 的 CRB 为

$$\text{Var}(\alpha) \geq \frac{\left(\frac{\partial g(\theta)}{\partial \theta}\right)^2}{-E\left[\frac{\partial^2 \ln p(x | \theta)}{\partial \theta^2}\right]}$$

11. 假定 $\alpha = g(\theta) = A\theta + b$, 假设 $\bar{\theta}$ 是 θ 的有效估计, 即 $\text{Var}(\bar{\theta}) = I^{-1}(\theta)$, 证明: $\alpha = A\bar{\theta} + b$ 也是有效估计量, 即对于线性变换, 估计量的有效性得以保持。

12. 考虑一个正弦参数的估计问题, 假定观测为

$$x_i = A \cos(2\pi f_0 i + \varphi) + w_i, \quad i = 0, 1, \dots, N-1$$

其中, w_i 为零均值高斯白噪声, 方差为 σ^2 , A, f_0, ϕ 为未知参量, 且 $A > 0, 0 < f_0 < 1/2$, 设 $\theta = [A, f_0, \phi]^T$, 求其费希尔信息矩阵。

13. 设有 N 次独立的观测为

$$x_i = A + w_i, \quad i = 0, 1, \dots, N-1$$

其中, A 为未知常数, w_i 是拉普拉斯噪声, 分布为

$$p(w_i) = \frac{1}{2} e^{-|w_i|}$$

求 A 的最大似然估计子, 并分析当 $N \rightarrow \infty$ 时, 最大似然估计的方差是否达到 CRB。

14. 设观测信号模型为

$$x_i = A + w_i, \quad i = 0, 1, \dots, N-1$$

式中, 随机变量 A 为信号幅值, $E[A] = 0$, $E[A^2] = a$, A 是已知参数。 w_i 是零均值白噪声, 与 A 相互独立, 求 A 的线性最小均方误差估计。

15. 在平稳白噪声背景中, 对信号参量做线性最小均方误差估计, 两次观测数据为 $x_i = A + w_i$ ($i = 0, 1$), 待估计参数 A 为信号幅度, w_i 为零均值, 方差为 σ_n^2 的白噪声, 且信号幅值与噪声不相关, 求 A 的最佳线性估计。

16. 设观测信号模型为

$$x_i = A + w_i, \quad i = 0, 1, \dots, N-1$$

未知参数 A 具有如下的概率密度分布:

$$f(A) = \begin{cases} \lambda \exp(-\lambda A), & A > 0 \\ 0, & A < 0 \end{cases}$$

其中, $\lambda > 0$, w_i 是零均值。方差为 σ_n^2 的高斯白噪声与 A 相互独立, 试求 A 的最大后验概率估计。

17. 对于无偏估计子, 最小均方误差等价于最小方差。对线性估计子

$$\hat{\theta} = \sum_{i=0}^{N-1} a_i x_i$$

引入约束条件 $E[x_i] = s_i \theta$, 则最小方差无偏估计子参数如下:

$$a = \frac{\mathbf{C}^{-1} \mathbf{s}}{\mathbf{s}^H \mathbf{C}^{-1} \mathbf{s}}$$

其中, $\mathbf{C}$ 是 x_i 的协方差矩阵, 证明所得估计子是无偏估计子并求其估计误差。

18. 设 $x(t) = A \cos \Omega_0 t + w(t)$, 通过采样对幅度 A 做线性估计。在 $\Omega_0 t = 0, \Omega_0 t = \pi/3$ 处采样两次, 记为 x_0, x_1 , 并设 $E[A] = 0, E[A^2] = 2, E[w_0] = E[w_1] = 0, E[w_0^2] = E[w_1^2] = 1, E[w_0 w_1] = 0, E[A w_0] = E[A w_1] = 0$, 求:

(1) 线性最小均方误差估计 $\hat{A}_{\text{lmse}}$;

(2) 线性最小均方误差估计的误差。

19. 对于 $\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$ 的实系数线性观测模型, 假定测量噪声的均值为零, 证明最小二乘估计误差的均方误差阵为

$$\mathbf{P}_{\theta_{\text{ls}}} = E[(\boldsymbol{\theta} - \boldsymbol{\theta}_{\text{ls}})(\boldsymbol{\theta} - \boldsymbol{\theta}_{\text{ls}})^T] = (\mathbf{H}^T \mathbf{H}) \mathbf{H}^T \mathbf{R} \mathbf{H} (\mathbf{H}^T \mathbf{H})^{-1}$$

其中, $\mathbf{R} = E[\mathbf{w} \mathbf{w}^T]$ 。

20. 观测信号模型为 $y_i = \alpha + \beta x_i + w_i$ ($i = 0, 1, \dots, N-1$), 通过采样对幅度 a 做线性估计, 其中, x_i 已知, w_i 为未知干扰。求 α, β 的最小二乘估计。

21. 设有零均值实平稳随机过程 $x(t)$, t 是 $[0, T]$ 内的一点。若已知 $x(0)$ 和 $x(T)$, 利用 $x(0)$ 和 $x(T)$ 求 $x(t)$ 的最佳线性估计。

22. 设有如下两种假设, 观测次数为 N 次

$$\begin{cases} H_0, & z_k = n_k \\ H_1, & z_k = 2 + n_k \end{cases} \quad (k = 1, 2, \dots, N)$$

其中, n_k 是满足零均值, 方差为 σ_n^2 的正态分布, 假设 $p(H_0) = p(H_1) = 0.5$, 求:

- (1) 最小错误概率准则下的判决表达式;
- (2) 虚警概率和检测概率(结果由误差函数表示)。

23. 设有如下两种假设, 观测次数为 N 次

$$\begin{cases} H_0, & z_k = n_k \\ H_1, & z_k = 1 + n_k \end{cases} \quad (k = 1, 2, \dots, N)$$

其中, n_k 是满足零均值, 方差为 σ_n^2 的正态分布。试构造一个虚警概率为 0.1 的 NP(纽曼-皮儿森)检测器, 求相应的检测概率。

24. 考虑如下四元数字通信系统:

$$\begin{cases} H_0, & z_k = n_k \\ H_1, & z_k = 1 + n_k \\ H_2, & z_k = 2 + n_k \\ H_3, & z_k = 3 + n_k \end{cases}$$

其中, n_k 是满足零均值, 方差为 σ_n^2 的正态分布, 各个假设的先验概率相等。试设计一个四元信号的最佳检测系统, 由每一个接收符号 z_k 检测其发送符号。