史忠科／著

最优估计的计算方法

科学出版社

国家自然科学基金及“211”工程资助

最优估计的计算方法

史忠科 著

科学出版社

2001

内 容 简 介

本书系统地总结了著者和国内外学者对最优估计计算方法的研究成果。重点介绍平方根滤波、平滑等算法的数值稳定性和计算效率，以及著者提出的U-D分解滤波、平滑新方法、状态与参数联合估计、鲁棒估计、非线性系统状态估计等研究成果。

著者建立的各类平滑新算法使计算效率有了突破性进展，有关的新算法已在实践中得到成功应用，取得了明显的社会、经济效益。

本书可供从事航空、航天、航海、工业过程控制、系统工程、信号处理、数据分析、计算数学等领域的科技人员学习，也可供有关专业的大学师生学习参考。

图书在版编目(CIP)数据

最优估计的计算方法/史忠科著.-北京:科学出版社,2001.
ISBN 7-03-009058-6

Ⅰ.最… Ⅱ.史… Ⅲ.估计-最优化算法 Ⅳ.O211.67

中国版本图书馆CIP数据核字(2000)第81893号

科学出版社 出版
北京东黄城根北街16号
邮政编码:100717
源海印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2001年5月第 一 版 开本:787×1092 1/16
2001年5月第一次印刷 印张:13 ½
印数:1—3 000 字数:307 000

定价:27.00元

(如有印装质量问题,我社负责调换〈北燕〉)

前　言

最优估计在航空、航天、工业过程控制等领域有着广泛的应用。它可用于统计图像的处理、交通密度估计、化工过程控制、河水流量估计、发电站负荷预测、卫星轨道估计、目标跟踪、飞行数据处理和其他实验数据处理等。这些处理大都采用卡尔曼滤波，推广卡尔曼滤波，固定点、固定区间、固定滞后平滑等方法。然而，理论研究和实际应用的结果表明，这些方法有两个明显的缺点：一是算法的数值稳定性差，在小型或微型机上实现时常常出现计算发散现象；二是平滑所需计算量较大，处理速度太慢。为了解决这些问题，国际上很多学者一直致力于研究滤波和平滑的具体算法，以提高数值稳定性和计算效率。近年来，著者对飞行状态估计问题进行了一系列的研究，提出了 U-D 分解滤波，固定点、固定区间、固定滞后平滑等新的计算方法和有关的鲁棒滤波算法，这些新算法使计算效率有了突破性进展，并在实践中得到成功应用，取得了明显的社会、经济效益。本书是对国内外 40 年来最优估计计算方法的系统总结，并讨论了状态与参数的联合估计问题和非线性系统的状态估计问题。

全书共十四章。第一章简单介绍最优估计理论的有关问题；第二章介绍线性系统的最优滤波和预测；第三章分析滤波的稳定性问题；第四章介绍线性离散系统的最优平滑估计理论和著者提出的固定点平滑估计的新方法；第五章介绍线性连续系统的平滑估计方法；第六章介绍线性离散系统的平方根滤波方法；第七章介绍线性离散系统的平方根平滑估计方法和著者对平滑问题的研究结果；第八章介绍连续系统的滤波和平滑计算方法；第九章介绍非线性系统的最优估计方法；第十章介绍大系统估计理论和著者给出的多级分解的卡尔曼滤波方法；第十一章介绍卡尔曼滤波器应用中的有关问题；第十二章介绍鲁棒估计理论；第十三章介绍状态和未知参数的联合估计及发展；第十四章给出了飞行状态估计的实际应用。

由于著者水平所限，书中难免有不成熟之处，敬请读者批评指正。

目　录

第一章　估计理论简介 ………… **1**
1.1　绪论 ………… 1
1.2　最小方差估计 ………… 3
1.3　极大似然估计 ………… 7
1.4　极大验后估计 ………… 9
1.5　线性最小方差估计 ………… 12
1.6　最小二乘估计 ………… 16
1.7　加权最小二乘法 ………… 17
第二章　线性系统的最优滤波和预测 ………… **19**
2.1　线性离散系统的最优滤波 ………… 19
2.2　线性离散系统的最优预测 ………… 22
2.3　线性连续系统的最优滤波 ………… 24
2.4　线性连续系统的最优预测 ………… 27
第三章　滤波的稳定性 ………… **29**
3.1　滤波的稳定性概念 ………… 29
3.2　随机线性系统的可控性和可观测性 ………… 31
3.3　滤波误差的界 ………… 35
第四章　线性离散系统的最优平滑估计 ………… **36**
4.1　固定区间最优平滑估计 ………… 36
4.2　固定点最优平滑估计 ………… 40
4.3　固定滞后最优平滑估计 ………… 43
4.4　固定点平滑估计的新方法 ………… 45
第五章　线性连续系统的平滑估计 ………… **47**
5.1　线性连续系统的最优平滑问题 ………… 47
5.2　固定区间平滑估计 ………… 50
5.3　固定点平滑估计 ………… 53
5.4　固定滞后平滑 ………… 54
第六章　线性离散系统的平方根滤波方法 ………… **56**
6.1　计算发散和 Joseph 方法 ………… 56
6.2　协方差平方根滤波方法 ………… 57
6.3　信息平方根滤波方法 ………… 61
6.4　序列平方根滤波方法 ………… 66
6.5　时间更新平方根滤波方法 ………… 68

试读结束：需要全本请在线购买： www.ertongbook.com

6.6 U-D 分解的滤波方法 …… 70
6.7 奇异值分解最优滤波方法 …… 72
6.8 平方根类滤波方法计算量比较 …… 73
第七章 线性离散系统的平方根平滑估计方法 …… 75
7.1 平滑估计的稳定性和计算效率 …… 75
7.2 固定区间平滑估计的有效方法 …… 75
7.3 固定点平滑估计的有效方法 …… 81
7.4 固定滞后平滑估计的有效方法 …… 87
7.5 前向平滑估计方法 …… 98
7.6 信息平方根平滑估计方法 …… 108
第八章 线性连续系统的滤波和平滑计算方法 …… 117
8.1 黎卡提方程的求解方法 …… 117
8.2 李亚普诺夫方程求解方法 …… 120
8.3 基于奇异值分解的黎卡提方程求解 …… 121
第九章 非线性系统的最优估计 …… 124
9.1 围绕标称轨道线性化方法 …… 124
9.2 围绕滤波值 $\mathbf{x}(t|t)$ 线性化的滤波方法 …… 127
9.3 近似条件均值滤波及二阶滤波 …… 129
9.4 迭代滤波 …… 133
第十章 大系统估计理论 …… 135
10.1 极大验后估计方法 …… 135
10.2 连续系统的分割滤波方法 …… 139
10.3 多级分解的卡尔曼滤波方法 …… 141
第十一章 卡尔曼滤波器应用中的有关问题 …… 144
11.1 滤波的发散及其克服方法 …… 144
11.2 自适应滤波 …… 147
11.3 次优滤波器 …… 155
第十二章 鲁棒估计理论 …… 156
12.1 噪声统计特性问题 …… 156
12.2 模型不确定问题 …… 158
12.3 H_∞滤波方法 …… 162
12.4 连续系统的最小方差鲁棒滤波 …… 164
12.5 离散系统的最小方差鲁棒滤波 …… 166
第十三章 状态和未知参数的联合估计 …… 171
13.1 适应性分割估计方法 …… 171
13.2 线性系统的状态和参数分离算法 …… 175
13.3 非线性系统的分离算法 …… 178
13.4 状态和参数同时估计的极大似然方法 …… 179

第十四章 飞行状态估计与监视……………………………………………… **182**
14.1 飞机运动方程和估计模型………………………………………………… 182
14.2 两段序列滤波方法……………………………………………………… 186
14.3 U-D 分解的两段序列滤波方法 ……………………………………… 188
14.4 U-D 分解的序列平滑算法 …………………………………………… 189
14.5 结果分析…………………………………………………………………… 191
附录 A 矩阵运算的一些公式………………………………………………… **198**
附录 B 矩阵求逆引理 ………………………………………………………… **203**
附录 C 矩阵许瓦茨不等式……………………………………………………… **204**
附录 D 正交定理……………………………………………………………… **204**
参考文献………………………………………………………………………… **206**

第一章　估计理论简介

1.1　绪　　论

1.1.1　基本概念

系统状态和参数估计具有广泛的应用，它可用于统计图像处理、交通密度估计、化学过程估计、河水流量估计、发电站负荷预测、卫星轨道估计、飞行器气动特性估计、最优导航、验后数据分析等。以验后数据分析为例，当一个试验完成后，为了判断试验的结果，必须对试验过程中所得到的数据进行处理。由于测量仪器本身含有系统偏差、刻度量因子误差和随机噪声，因此，在进行结果分析之前，必须将这些误差确定出来，否则会导致错误的结果。估计理论实际上是根据状态方程和观测方程，按概率统计和优化的原理对系统的状态或参数作出估计。以下，仅举一个简单的例子说明最优估计问题。

例 1.1　假定用两个相同的仪表对某一恒定量 x 进行观测，观测值为 $z_i(i=1,2)$，观测时存在随机的、相互独立的、均值为零的观测误差 $v_i(i=1,2)$。试设计一种数据处理方法，组合两个仪表观测值，以便得到 x 的无偏最优估值。

解　根据题意，观测值 z_1,z_2 应为

$$z_1 = x + v_1,\quad z_2 = x + v_2 \tag{1.1}$$

在没有其他信息情况下，我们可以假定 x 的估计值应为两个观测值的线性函数，并以 $\hat{x}$ 表示 x 的估计值，则

$$\hat{x} = k_1 z_1 + k_2 z_2 \tag{1.2}$$

式中，k_1,k_2 为待定系数。估值误差定义为 $\tilde{x}=x-\hat{x}$，我们取 $\tilde{x}$ 的均方差为最小作为最优判据，而且选择 k_1,k_2 与 x 值无关。根据估计值为无偏的要求可得

$$E(\tilde{x}) = E[x - k_1(x + v_1) - k_2(x + v_2)] = 0 \tag{1.3}$$

由于 $E(v_1)=E(v_2)=0$，$E(x)=E(\hat{x})$，由(1.3)式得

$$k_2 = 1 - k_1 \tag{1.4}$$

以(1.4)式代入 $\tilde{x}=x-\hat{x}$，求 $E(\tilde{x}^2)$ 得

$$\begin{aligned} E(\tilde{x}^2) &= E\{[x - k_1(x + v_1) - (1 - k_1)(x + v_2)]^2\} \\ &= E\{[k_1 v_1 + (1 - k_1)v_2]^2\} \\ &= E[(k_1^2 v_1^2 + (1 - k_1)^2 v_2^2 + 2k_1(1 - k_1)v_1 v_2] \end{aligned}$$

假定 v_1,v_2 的方差为 $E(v_1^2)=\sigma_1^2$，$E(v_2^2)=\sigma_2^2$，根据题意 v_1,v_2 相互独立，即 $E(v_1v_2)=0$，故得

$$E(\tilde{x}^2) = k_1^2\sigma_1^2 + (1 - k_1)^2\sigma_1^2$$

适当地选择 k_1，可使 $E(\tilde{x}^2)$ 为最小。为此令 $E(\tilde{x}^2)$ 对 k_1 的导数为零，即

$$\frac{\partial E(\tilde{x}^2)}{\partial k_1} = 2k_1\sigma_1^2 - 2(1 - k_1)\sigma_2^2 = 0$$

由此求得 $k_1=\sigma_2^2/(\sigma_1^2+\sigma_2^2)$。于是 x 的最优估值为

$$\hat{x} = \left(\frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}\right)z_1 + \left(\frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}\right)z_2$$

对应的最小估值误差均方值为

$$E(\tilde{x}^2) = \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right)^{-1}$$

从上面例子可以看出，当缺乏其他信息并只能从观测值确定 x 时，x 的最优估值 $\hat{x}$ 应为观测值 z_1,z_2 的加权平均，加权系数 k_1,k_2 分别与观测误差的均方值成反比，即观测误差的均方值愈大，该观测值的加权系数愈小。反之，观测误差的均方值愈小，加权系数愈大。如果 $\sigma_1^2=0$，则 $\hat{x}=z_1$；如果 $\sigma_1^2=\sigma_2^2$，则 $\hat{x}=(z_1+z_2)/2$。在以后的讨论中，我们将看到对所获得的系统状态各种信息进行加权处理，可以求得系统状态的最优估值，这是最优估计理论的基本思想。

1.1.2 估计问题提法和估计准则

实际中，人们根据对不同仪器仪表及被估值的统计特性的掌握程度提出了下列一些估计准则：最小方差准则、极大似然准则、极大验后准则、线性最小方差准则及最小二乘准则等。依据不同的准则，可得各种不同的估计方法，例如最小方差估计、极大似然估计、极大验后估计、线性最小方差估计等。

为了方便读者自学，本书将一维随机变量及其取值统一为小写字母表示，多维随机变量及其取值统一为小写黑体字母表示。

1. 估计问题的分类

估计问题可分成两大类，即参数估计和状态估计，分别讨论如下。

(1) 参数估计

例如，在做完试验之后，会得到一系列不同时刻 z 的观测值，我们希望用一条曲线来表示 z 与 t 的关系，设

$$z(t) = x_1h_1(t) + x_2h_2(t) + \cdots + x_nh_n(t) \tag{1.5}$$

式中，$h_1(t),h_2(t),\cdots,h_n(t)$ 为已知的时间函数，一般是 t 的幂函数、指数函数或正余弦函数等；$x_1,x_2,\cdots,x_n$ 为 n 个未知数，它们不随时间而变。

要求根据 m 对观测值 $(z_i,t_i)(i=1,2,\cdots,m;\ m>n)$ 来估计未知参数 $x_1,x_2,\cdots,x_n$。这就是前面提到过的曲线拟合问题。按照什么准则来拟合呢？一般采用 $z(t)$ 与各观测值 z_i 之间的差的平方和为最小作为估计参数的准则。

(2) 状态估计

如果被估计的量是系统的状态变量，则称这种估计为状态估计。状态变量是随时间而变的随机过程。例如，系统的状态变量 $\mathbf{x}(t)$ 满足下列的微分方程：

$$\dot{\mathbf{x}}(t) = A(t)\mathbf{x}(t) + B(t)\mathbf{u}(t) + F(t)\mathbf{w}(t)$$

要求根据观测值

$$\mathbf{z}(t) = H(t)\mathbf{x}(t) + \mathbf{v}(t)$$

来估计状态变量 $\mathbf{x}(t)$。在上述两式中，$\mathbf{w}(t)$ 和 $\mathbf{v}(t)$ 都为随机干扰，$\mathbf{u}(t)$ 为控制量。$\mathbf{x}(t)$ 是随时间而变的随机变量，即为随机过程。这种估计称为状态估计。

2. 估计准则

前面已提到过对估计的要求，估值愈接近真值愈好，这是一种不严格的说法。为了进行估计，必须有估计准则。所谓最优估计是指在某一估计准则条件下求得的最优估值，如果换了一个估计准则，则这一估值就不一定是最优的了。估计准则可能是多种多样的，选取不同的估计准则，就有不同的估计方法，估计方法与估计准则是紧密相关的。

(1) 最小方差准则(最小方差估计)

最小方差准则是以估计误差的方差阵达到最小作为估计的准则。按这种准则求得的最优估值叫做最小方差估计。为了进行最小方差估计，需要知道被估值 $\mathbf{x}$ 和观测值 $\mathbf{z}$ 的条件概率分布密度 $p(\mathbf{x}|\mathbf{z})$ 或 $p(\mathbf{z})$，以及它们的联合概率分布密度 $p(\mathbf{x},\mathbf{z})$。

(2) 极大似然准则(极大似然估计)

极大似然准则是使条件概率分布密度 $p(\mathbf{z}|\mathbf{x})$ 达到极大的那个 $\mathbf{x}$ 值作为估值。按这种估计准则求得的 $\mathbf{x}$ 的最优估值的方法称为极大似然估计。为了求出极大似然估计，需要知道条件概率分布密度 $p(\mathbf{z}|\mathbf{x})$。

(3) 极大验后准则(极大验后估计)

极大验后准则是使条件概率分布密度 $p(\mathbf{x}|\mathbf{z})$ 达到极大的那个 $\mathbf{x}$ 值作为估值。按这种估计准则求得的 $\mathbf{x}$ 的最优估值的方法称为极大验后估计。为了求出极大验后估计，需要知道验后概率分布密度 $p(\mathbf{x}|\mathbf{z})$。

(4) 线性最小方差准则(线性最小方差估计)

为了进行最小方差估计和极大验后估计，需要知道 $p(\mathbf{x}|\mathbf{z})$；为了进行极大似然估计，需要知道 $p(\mathbf{z}|\mathbf{x})$。如果我们只知道观测值和被估值的一、二阶矩，即 $E[\mathbf{x}]$，$E[\mathbf{z}]$，$\mathrm{Var}(\mathbf{x})$，$\mathrm{Var}(\mathbf{z})$，$\mathrm{Cov}(\mathbf{x},\mathbf{z})$。在这种情况下，为了得到有用的结果，必须对估计量的函数形式加以限制。我们限定所求的估计量是观测值的线性函数，以估计误差阵达到最小作为最优估计的准则。按这种方式求得的最优估计值称为线性最小方差估计。

(5) 最小二乘准则(最小二乘估计)

如果我们不知道 $\mathbf{x}$ 和 $\mathbf{z}$ 的概率分布密度，也不知道它们的一、二阶矩。这时就只能用高斯提出的最小二乘法。最小二乘法是将残差平方最小作为最优估计准则的，详细内容在后面讨论。

下面分别讨论最小方差估计、极大似然估计、极大验后估计、线性最小方差估计和最小二乘法估计。

1.2 最小方差估计

在参数估计中，经常采用最小方差准则，即要求估计误差的方差为最小。下面将会看到，被估量的数学期望或条件数学期望是最小方差估计。

下面分别讨论一维、二维和多维随机变量的最小方差估计。

1.2.1 一维随机变量的最小方差估计

设有一维随机变量 x，它的概率分布密度 $p(x)$ 是已知的，数学期望值为 m_x，常数 a 为 x 的估值 $\hat{x}$（通常在字母上方加上"^"记号表示估值），则评价估计优劣的准则是 $\hat{x}$ 与 x 的误差的方差为最小，即

$$J = E[(x-a)^2] = \int_{-\infty}^{+\infty}(x-a)^2 p(x)\mathrm{d}x = \min \tag{1.6}$$

将上式展开得

$$J = E[(x-a)^2] = E[x^2] - 2aE[x] + a^2$$

求上式对 a 的偏导数，令偏导数等于零，可得

$$\frac{\partial J}{\partial a} = 2a - 2E[x] = 0$$

所以 x 的最优估值为

$$\hat{x} = a = E[x] = \int_{-\infty}^{\infty} xp(x)\mathrm{d}x = m_x$$

（通常在字母上方加上"～"记号表示误差）

$$E[\tilde{x}] = E[x] - E[\hat{x}] = E[x] - E[m_x] = m_x - m_x = 0$$

即

$$E[\hat{x}] = E[x]$$

如果估值 $\hat{x}$ 的数学期望等于 x 的数学期望，或者估计误差 $\tilde{x}$ 的数学期望为零，则称这种估计为无偏估计。因而这里的 x 的估计是无偏估计。

估计误差 $\tilde{x}$ 的方差为

$$E[(x-m_x)^2] = \int_{-\infty}^{\infty}(x-m_x)^2 p(x)\mathrm{d}x = \sigma_x^2$$

所以数学期望 m_x 是 x 的最小方差估计。

1.2.2 多维随机变量的最小方差估计

设有两个随机变量 x 和 z，x 为被估随机变量，z 为观测值。x 与 z 没有明确的函数关系，只有概率上的联系。x 和 z 的概率分布密度分别为 $p(x)$ 和 $p(z)$，其联合概率分布密度为 $p(x,z)$。希望用随机变量 z 的函数 $g(z)$ 作为随机变量 x 的估值，并要求估计误差的方差为最小，即

$$E\{[x-g(z)]^2\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[x-g(z)]^2 p(x,z)\mathrm{d}x\mathrm{d}z = \min$$

可以证明，使上式为最小的函数 $g(z)$ 是 x 的条件数学期望，即

$$\hat{x} = g(z) = E[x|z]$$

x 和 z 的联合概率分布密度可用下式表示：

$$p(x,z) = p(x|z)p(z)$$

我们可得估计误差的方差

$$J = E\{[x-g(z)]^2\} = \int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty}[x-g(z)]^2 p(x|z)\mathrm{d}x\right\}p(z)\mathrm{d}z \tag{1.7}$$

试读结束：需要全本请在线购买： www.ertongbook.com

上面积分中的被积函数 $\int_{-\infty}^{\infty}[x-g(z)]^2 p(x|z)\mathrm{d}x$ 是非负的。因此为了使双重积分最小，只要对每个 z 值，使积分

$$\int_{-\infty}^{\infty}[x-g(z)]^2 p(x|z)\mathrm{d}x \tag{1.8}$$

为最小，就可以使 J 为最小。

对于给定的 z 值，随机变量 x 的条件概率密度为 $p(x|z)$，积分式(1.8)是 x 相对于常值 $g(z)$ 的二阶矩，参照(1.6)式可知，使这个二阶矩为最小的 $g(z)$ 值是 x 的条件数学期望

$$\hat{x} = g(z) = \int_{-\infty}^{\infty} x p(x|z)\mathrm{d}x = E[x|z]$$

对于两个随机变量 x 和 z 来说，如果已知其联合概率分布密度 $p(x,z)$，则 x 的最小方差估计为 x 的条件数学期望

$$\hat{x} = E[x|z]$$

因为 $g(z)$ 和 z 一般不成线性关系，所以最小方差估计一般称为非线性估计。

这一结果可推广到多维随机变量的估计问题。设 $\mathbf{x}$ 为需要估计的 n 维随机变量，它的可能取值为

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

观测值 $\mathbf{z}$ 为 q 维随机变量，其可能取值为

$$\mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_q \end{bmatrix}$$

假定 $\mathbf{x}$ 和 $\mathbf{z}$ 的联合概率分布密度为

$$p(\mathbf{x},\mathbf{z}) = p(x_1,x_2,\cdots,x_n;z_1,z_2,\cdots,z_q)$$

$\mathbf{x}$ 的概率分布密度为

$$p(\mathbf{x}) = p(x_1,x_2,\cdots,x_n)$$

$\mathbf{z}$ 的概率分布密度为

$$p(\mathbf{z}) = p(z_1,z_2,\cdots,z_q)$$

$p(\mathbf{x},\mathbf{z})$，$p(\mathbf{x})$和 $p(\mathbf{z})$都为标量函数。

现在要根据观测值 $\mathbf{z}$ 来估计 $\mathbf{x}$，设 $\mathbf{x}$ 的估值为

$$\hat{\mathbf{x}} = \hat{\mathbf{x}}(\mathbf{z})$$

则 $\mathbf{x}$ 的估计误差为

$$\tilde{\mathbf{x}} = \mathbf{x} - \hat{\mathbf{x}}$$

估计误差 $\tilde{\mathbf{x}}$ 是一个与 $\mathbf{x}$ 同维数的随机向量。希望估计误差的方差阵

$$J = E[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T] \tag{1.9}$$

为最小。

现在要求出 $\mathbf{x}$ 的最优估值$\hat{\mathbf{x}}(z)$，使得 J 为最小。关于矩阵大小的定义如下：设有两个 $n\times n$ 矩阵 A 和 B，若$(A-B)$正定(或非负定)，则称 $A>B$(或 $A\geqslant B$)。

在给定 $\mathbf{z}=\mathbf{z}$ 的条件下，$\mathbf{x}$ 的条件概率密度为 $p((\mathbf{x}|\mathbf{z})$，由贝叶斯公式有

$$p(\mathbf{x},\mathbf{z}) = p(\mathbf{x}|\mathbf{z})p(\mathbf{z})$$

估计误差的方差阵为

$$\begin{aligned}J &= E[\tilde{\mathbf{x}}\,\tilde{\mathbf{x}}^T] = E\{[\mathbf{x}-\hat{\mathbf{x}}(\mathbf{z})][\mathbf{x}-\hat{\mathbf{x}}(\mathbf{z})]^T\} \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[\mathbf{x}-\hat{\mathbf{x}}(\mathbf{z})][\mathbf{x}-\hat{\mathbf{x}}(\mathbf{z})]^T p(\mathbf{x},\mathbf{z})\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{z} \\ &= \int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty}[\mathbf{x}-\hat{\mathbf{x}}(\mathbf{z})][\mathbf{x}-\hat{\mathbf{x}}(\mathbf{z})]^T p(\mathbf{x}|\mathbf{z})\mathrm{d}\mathbf{x}\right\}p(\mathbf{z})\mathrm{d}\mathbf{z} \qquad (1.10)\end{aligned}$$

因为 $\mathbf{x}$ 和 $\mathbf{z}$ 都是向量，所以这里的 $\mathrm{d}\mathbf{x}=\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_n$，$\mathrm{d}\mathbf{z}=\mathrm{d}z_1\mathrm{d}z_2\cdots\mathrm{d}z_q$。我们要选择$\hat{\mathbf{x}}(\mathbf{z})$，使得(1.10)式的 J 是一个非负定的对称矩阵，参照(1.7)式可知，为了使 J 最小，只要对每个 z 值，使

$$\int_{-\infty}^{\infty}[\mathbf{x}-\hat{\mathbf{x}}(\mathbf{z})][\mathbf{x}-\hat{\mathbf{x}}(\mathbf{z})]^T p(\mathbf{x}|\mathbf{z})\mathrm{d}\mathbf{x} \qquad (1.11)$$

为最小即可。

对于给定的 $\mathbf{z}$ 值，随机变量 $\mathbf{x}$ 的条件概率密度为 $p(\mathbf{x}|\mathbf{z})$，积分式(1.11)是 $\mathbf{x}$ 相对于 $\mathbf{x}(\mathbf{z})$的二阶矩。参照(1.8)式可知，使得这个二阶矩为最小的 $\mathbf{x}(\mathbf{z})$值是 $\mathbf{x}$ 的条件数学期望：

$$\hat{\mathbf{x}}(\mathbf{z}) = \int_{-\infty}^{\infty}\mathbf{x}p(\mathbf{x}|\mathbf{z})\mathrm{d}\mathbf{x} = E[\mathbf{x}|\mathbf{z}] \qquad (1.12)$$

一般可写成

$$\hat{\mathbf{x}}(\mathbf{z}) = E[\mathbf{x}|\mathbf{z}] \qquad (1.13)$$

估计误差的方差阵为

$$J = E[\tilde{\mathbf{x}}\,\tilde{\mathbf{x}}^T] = E\{[\mathbf{x}-E(\mathbf{x}|\mathbf{z})][\mathbf{x}-E(\mathbf{x}|\mathbf{z})]^T\} \qquad (1.14)$$

下面证明，当 $\mathbf{x}$ 的估值$\hat{\mathbf{x}}=E[\mathbf{x}|\mathbf{z}]$时，确实使 J 为最小。设 $\mathbf{x}$ 的估值为 $\mathbf{z}$ 的任一向量函数 $\bar{\mathbf{x}}(\mathbf{z})$，则从(1.11)式出发，可写出

$$\begin{aligned}&\int_{-\infty}^{\infty}[\mathbf{x}-\bar{\mathbf{x}}(\mathbf{z})][\mathbf{x}-\bar{\mathbf{x}}(\mathbf{z})]^T p(\mathbf{x}|\mathbf{z})\mathrm{d}\mathbf{x} \\ =&\int_{-\infty}^{\infty}[\mathbf{x}-E(\mathbf{x}|\mathbf{z})+E(\mathbf{x}|\mathbf{z})-\bar{\mathbf{x}}(\mathbf{z})] \\ &\times[\mathbf{x}-E(\mathbf{x}|\mathbf{z})+E(\mathbf{x}|\mathbf{z})-\bar{\mathbf{x}}(\mathbf{z})]^T p(\mathbf{x}|\mathbf{z})\mathrm{d}\mathbf{x} \\ =&\int_{-\infty}^{\infty}[\mathbf{x}-E(\mathbf{x}|\mathbf{z})][\mathbf{x}-E(\mathbf{x}|\mathbf{z})]^T p(\mathbf{x}|\mathbf{z})\mathrm{d}\mathbf{x} \\ &+[E(\mathbf{x}|\mathbf{z})-\bar{\mathbf{x}}(\mathbf{z})][E(\mathbf{x}|\mathbf{z})-\bar{\mathbf{x}}(\mathbf{z})]^T\int_{-\infty}^{\infty}p(\mathbf{x}|\mathbf{z})\mathrm{d}\mathbf{x} \\ &+\int_{-\infty}^{\infty}[\mathbf{x}-E(\mathbf{x}|\mathbf{z})]p(\mathbf{x}|\mathbf{z})\mathrm{d}\mathbf{x}\cdot[E(\mathbf{x}|\mathbf{z})-\bar{\mathbf{x}}(\mathbf{z})]^T \\ &+[E(\mathbf{x}|\mathbf{z})-\bar{\mathbf{x}}(\mathbf{z})]\int_{-\infty}^{\infty}[\mathbf{x}-E(\mathbf{x}|\mathbf{z})]^T p(\mathbf{x}|\mathbf{z})\mathrm{d}\mathbf{x} \qquad (1.15)\end{aligned}$$

由于

$$\int_{-\infty}^{\infty} p(\mathbf{x}|\mathbf{z})\mathrm{d}\mathbf{x} = 1$$

$$\int_{-\infty}^{\infty} [\mathbf{x} - E(\mathbf{x}|\mathbf{z})]p(\mathbf{x}|\mathbf{z})\mathrm{d}\mathbf{x} = 0$$

$$\int_{-\infty}^{\infty} [\mathbf{x} - E(\mathbf{x}|\mathbf{z})]^T p(\mathbf{x}|\mathbf{z})\mathrm{d}\mathbf{x} = 0$$

(1.15)式变成

$$\begin{aligned}&\int_{-\infty}^{\infty} [\mathbf{x} - \bar{\mathbf{x}}(\mathbf{z})][\mathbf{x} - \bar{\mathbf{x}}(\mathbf{z})]^T p(\mathbf{x}|\mathbf{z})\mathrm{d}\mathbf{x} \\ =&\int_{-\infty}^{\infty} [\mathbf{x} - E(\mathbf{x}|\mathbf{z})][\mathbf{x} - E(\mathbf{x}|\mathbf{z})]^T p(\mathbf{x}|\mathbf{z})\mathrm{d}\mathbf{x} \\ &+ [E(\mathbf{x}|\mathbf{z}) - \bar{\mathbf{x}}(\mathbf{z})][E(\mathbf{x}|\mathbf{z}) - \bar{\mathbf{x}}(\mathbf{z})]^T\end{aligned} \tag{1.16}$$

(1.16)式等号左边为非负定矩阵，等号右边的两个矩阵也分别为非负定矩阵，所以

$$\begin{aligned}&\int_{-\infty}^{\infty} [\mathbf{x} - \bar{\mathbf{x}}(\mathbf{z})][\mathbf{x} - \bar{\mathbf{x}}(\mathbf{z})]^T p(\mathbf{x}|\mathbf{z})\mathrm{d}\mathbf{x} \\ \geqslant&\int_{-\infty}^{\infty} [\mathbf{x} - E(\mathbf{x}|\mathbf{z})][\mathbf{x} - E(\mathbf{x}|\mathbf{z})]^T p(\mathbf{x}|\mathbf{z})\mathrm{d}\mathbf{x}\end{aligned} \tag{1.17}$$

这就证明了 $\mathbf{x}$ 的最优估计为

$$\hat{\mathbf{x}} = E[\mathbf{x}|\mathbf{z}]$$

下面讨论估计是否无偏的问题。由(1.12)式可得

$$\begin{aligned}E\{E[\mathbf{x}|\mathbf{z}]\} &= \int_{-\infty}^{\infty} E[\mathbf{x}|\mathbf{z}]p(\mathbf{z})\mathrm{d}\mathbf{z} \\ &= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} \mathbf{x}p(\mathbf{x}|\mathbf{z})\mathrm{d}\mathbf{x}\right]p(\mathbf{z})\mathrm{d}\mathbf{z} \\ &= \int_{-\infty}^{\infty}\mathbf{x}\left[\int_{-\infty}^{\infty} p(\mathbf{x}|\mathbf{z})p(\mathbf{z})\mathrm{d}\mathbf{z}\right]\mathrm{d}\mathbf{x} \\ &= \int_{-\infty}^{\infty}\mathbf{x}p(\mathbf{x})\mathrm{d}\mathbf{x} = E[\mathbf{x}]\end{aligned} \tag{1.18}$$

因此估计是无偏的。

估计误差的方差阵为

$$\begin{aligned}\mathrm{Var}\tilde{\mathbf{x}} &= E\{\tilde{\mathbf{x}}\,\tilde{\mathbf{x}}^T\} = E\{[\mathbf{x} - E[\mathbf{x}|\mathbf{z}]][\mathbf{x} - E[\mathbf{x}|\mathbf{z}]]^T\} \\ &= \int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty}[\mathbf{x} - E[\mathbf{x}|\mathbf{z}]][\mathbf{x} - E[\mathbf{x}|z]]^T p(\mathbf{x}|\mathbf{z})\mathrm{d}\mathbf{x}\right\}\cdot p(\mathbf{z})\mathrm{d}\mathbf{z} \\ &= \int_{-\infty}^{\infty}\mathrm{Var}(\mathbf{x}|\mathbf{z})p(\mathbf{z})\mathrm{d}\mathbf{z}\end{aligned} \tag{1.19}$$

1.3 极大似然估计

极大似然估计是以观测值出现的概率为最大作为准则的，这是一种很普通的参数估计方法。

设 z 是连续随机变量，其分布密度为 $p(z,\theta_1,\theta_2,\cdots,\theta_n)$，含有 n 个未知参数 θ_1，

$\theta_2,\cdots,\theta_n$。把 k 个独立观测值 $z_1,z_2,\cdots,z_k$ 分别代入 $p(z,\theta_1,\theta_2,\cdots,\theta_n)$ 中，可得

$$p(z_i,\theta_1,\theta_2,\cdots,\theta_n); \quad i=1,2,\cdots,k$$

将所得的 k 个函数相乘得

$$L(z_1,z_2,\cdots z_k;\theta_1,\theta_2,\cdots,\theta_n)=\prod_{i=1}^{k}p(z_i,\theta_1,\theta_2,\cdots,\theta_n) \tag{1.20}$$

称函数 L 为似然函数。当 $z_1,z_2,\cdots,z_k$ 固定时，L 是 $\theta_1,\theta_2,\cdots,\theta_n$ 的函数。极大似然法的实质，就是求出使 L 达到极大时，$\theta_1,\theta_2,\cdots,\theta_n$ 的估值 $\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_n$。从(1.20)式可以看到 $\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_n$ 是观测值 $z_1,z_2,\cdots,z_k$ 的函数。

为了便于求出使 L 达到极大的 $\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_n$，对(1.20)式取对数，把乘转变为连加，即

$$\ln L=\sum_{i=1}^{k}\ln p(z_i,\theta_1,\theta_2,\cdots,\theta_n) \tag{1.21}$$

由于对数函数是单调增加函数，当 L 取极大值时，$\ln L$ 也同时取极大值。将上式分别对 $\theta_1,\theta_2,\cdots,\theta_n$ 求偏导数，令偏导数等于零，可得下面的一组方程：

$$\left.\begin{array}{c}\dfrac{\partial}{\partial\theta_1}(\ln L)=0\\ \vdots\\ \dfrac{\partial}{\partial\theta_n}(\ln L)=0\end{array}\right\} \tag{1.22}$$

解上述方程组，可得使 L 达到极大值的 $\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_n$。

按极大似然法确定的 $\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_n$ 使 $z_1,z_2,\cdots,z_k$ 最有可能出现，而不需要 $\theta_1,\theta_2,\cdots,\theta_n$ 的验前知识，即不需要知道 $\theta_1,\theta_2,\cdots,\theta_n$ 的概率分布密度或一、二阶矩。因此，极大似然估计是最小方差估计的次最优估计。

下面讨论用极大似然法估计参数的问题。

设 $\mathbf{z}$ 为 m 维随机变量，$\mathbf{x}$ 为 n 维随机变量，假定已知 $\mathbf{z}$ 的条件概率密度为 $p(\mathbf{z}|\mathbf{x})$，现在得到 k 组 $\mathbf{z}$ 的观测值 $\mathbf{z}_1,\mathbf{z}_2,\cdots,\mathbf{z}_k$，观测值互相独立。问 $\mathbf{x}$ 是什么值时，$\mathbf{z}_1,\mathbf{z}_2,\cdots,\mathbf{z}_k$ 出现的可能性最大？为此，确定似然函数

$$L(\mathbf{z},\mathbf{x})=p(\mathbf{z}_1|\mathbf{x})p(\mathbf{z}_2|\mathbf{x})\cdots p(\mathbf{z}_k|\mathbf{x})=p(\mathbf{z}|\mathbf{x}) \tag{1.23}$$

或

$$\ln L(\mathbf{z},\mathbf{x})=\ln p(\mathbf{z}|\mathbf{x}) \tag{1.24}$$

求使 L 为极大的 $\mathbf{x}$ 值，令

$$\frac{\partial L}{\partial\mathbf{x}}=0\text{ 或}\frac{\partial\ln L}{\partial\mathbf{x}}=0 \tag{1.25}$$

解之，可得 $\mathbf{x}$ 的估值 $\hat{\mathbf{x}}$。而

$$\frac{\partial^2 L}{\partial\mathbf{x}^2}<0\quad\text{或}\quad\frac{\partial^2\ln L}{\partial\mathbf{x}^2}<0$$

则为 L 取极大值的充分条件。

因此用极大似然法时，应先求似然函数 L，然后用微分法求出使似然函数 L 为极大的 $\mathbf{x}$ 的估值 $\hat{\mathbf{x}}$。

设有一线性观测系统

$$\mathbf{z}=\mathbf{h}(\mathbf{x},\mathbf{v})=H\mathbf{x}+\mathbf{v} \tag{1.26}$$

式中，$\mathbf{z}$ 为 m 维观测值；$\mathbf{x}$ 为 n 维未知参数；$\mathbf{v}$ 为 m 维测量误差。设 $\mathbf{v}$ 与 $\mathbf{x}$ 独立，已知 $\mathbf{v}$ 的统计特性，求 $\mathbf{x}$ 的极大似然估计。由(1.26)式得

$$\mathbf{v} = \mathbf{h}^{-1}(\mathbf{x},\mathbf{z}) = \mathbf{z} - H\mathbf{x}$$

$$J = \frac{\partial \mathbf{h}^{-1}(\mathbf{x},\mathbf{z})}{\partial \mathbf{z}} = I \quad (m \times m \text{ 单位阵})$$

下面求出似然函数：

$$L(\mathbf{z},\mathbf{x}) = p(\mathbf{z}|\mathbf{x}) = \frac{p(\mathbf{x},\mathbf{z})}{p(\mathbf{x})}$$

根据不同随机变量的概率密度变换公式，并考虑到 $\mathbf{v}$ 与 $\mathbf{x}$ 独立，可得

$$p(\mathbf{x},\mathbf{z}) = p(\mathbf{x},\mathbf{v})\,|JJ^T|^{\frac{1}{2}} = p(\mathbf{x},\mathbf{v}) = p(\mathbf{x})p(\mathbf{v})$$

$$L(\mathbf{z},\mathbf{x}) = \frac{p(\mathbf{x})p(\mathbf{v})}{p(\mathbf{x})} = p(\mathbf{v}) = p(\mathbf{z} - H\mathbf{x})$$

令

$$\frac{\partial L(\mathbf{z},\mathbf{x})}{\partial \mathbf{x}} = \frac{\partial p(\mathbf{z} - H\mathbf{x})}{\partial \mathbf{x}} = 0$$

可得 $\mathbf{x}$ 的估值 $\hat{\mathbf{x}}$。

假定噪声 $\mathbf{v}$ 服从于零均值正态分布，其方差阵为

$$E[\mathbf{v}\mathbf{v}^T] = R$$

则

$$L(\mathbf{z},\mathbf{x}) = p(\mathbf{v}) = \frac{1}{(\sqrt{2\pi})^m |R|^{\frac{1}{2}}} e^{-\frac{1}{2}\mathbf{v}^T R^{-1}\mathbf{v}}$$

将 $\mathbf{v}=z-H\mathbf{x}$ 代入上式得

$$L(\mathbf{z},\mathbf{x}) = p(\mathbf{x} - H\mathbf{x}) = C\exp[(\mathbf{z} - H\mathbf{x})R^{-1}(\mathbf{z} - H\mathbf{x})]$$

式中

$$C = \frac{1}{(\sqrt{2\pi})^m |R|^{\frac{1}{2}}}$$

求出 $\mathbf{x}$，使

$$J = \frac{1}{2}(\mathbf{z} - H\mathbf{x})R^{-1}(\mathbf{z} - H\mathbf{x}) = \min \tag{1.27}$$

求 J 对 $\mathbf{x}$ 的偏导数，令偏导数等于零，可得 $\mathbf{x}$ 的估值 $\hat{\mathbf{x}}$：

$$\frac{\partial J}{\partial \mathbf{x}} = - H^T R^{-1}\mathbf{z} + H^T R^{-1} H\mathbf{x} = 0$$

$$H^T R^{-1} H\mathbf{x} = H^T R^{-1}\mathbf{z}$$

$$\hat{\mathbf{x}} = (H^T R^{-1} H)^{-1} H^T R^{-1}\mathbf{z} \tag{1.28}$$

1.4 极大验后估计

1.4.1 极大验后估计

使 $\mathbf{x}$ 的验后概率密度 $p(\mathbf{x}|\mathbf{z})$ 达到最大的那个 $\mathbf{x}$ 值称为极大验后估计。$\hat{\mathbf{x}}=\mathbf{x}_a$ 称为 $\mathbf{x}$

的极大验后估计。$\mathbf{x}_a$ 为 $\mathbf{x}$ 的最可能出现值，称为"众数"或"最频数"。从直观上看很清楚，能使验后概率密度 $p(\mathbf{x}|\mathbf{z})$ 达到最大的那个 $\mathbf{x}$ 值就是 $\mathbf{x}_a$，随机变量 $\mathbf{x}$ 落在 $\mathbf{x}_a$ 附近的小邻域内的概率大于落在其他任何 $\mathbf{x}$ 值的同样邻域内的概率。由于对数函数是单调增加函数，所以 $\ln p(\mathbf{x}|\mathbf{z})$ 与 $p(\mathbf{x}|\mathbf{z})$ 在相同的 $\mathbf{x}$ 值达到极大值。由微分可知 $\mathbf{x}_a$ 应满足下列方程：

$$\left.\frac{\partial \ln p(\mathbf{x}|\mathbf{z})}{\partial \mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}_a} = 0 \tag{1.29}$$

上述方程叫做验后方程。

1.4.2 最小方差估计及估计误差

要进行极大验后估计，必须知道验后概率密度 $p(\mathbf{x}|\mathbf{z})$。如果知道 $p(\mathbf{z}|\mathbf{x})$，可按下式计算：

$$p(\mathbf{x}|\mathbf{z}) = \frac{p(\mathbf{z}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{z})} \tag{1.30}$$

式中，$p(\mathbf{x})$是 $\mathbf{x}$ 的验前概率密度；$p(\mathbf{z})$是 $\mathbf{z}$ 的概率密度；$p(\mathbf{z}|\mathbf{x})$可用计算方法或实验方法求得。为了计算 $p(\mathbf{x})$，我们在 $\mathbf{x}$ 没有验前知识可供利用时，可假定 $\mathbf{x}$ 在很大范围内变化。在这种情况下，可把 $\mathbf{x}$ 的验前概率密度 $p(\mathbf{x})$近似地看作是方差阵趋于无限大的正态分布密度

$$p(\mathbf{x}) = e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m}_\mathbf{x})^T P^{-1}(\mathbf{x}-\mathbf{m}_\mathbf{x})} / (2\pi)^{\frac{n}{2}} |P|^{\frac{1}{2}}$$

式中，P 为 $\mathbf{x}$ 的方差阵，$P \to \infty I$（I 为 $n \times n$ 单位阵），$P^{-1} \to 0$，于是

$$\ln p(\mathbf{x}) = -\ln\left[(2\pi)^{\frac{n}{2}} |P|^{\frac{1}{2}}\right] - \frac{1}{2}(\mathbf{x} - \mathbf{m}_\mathbf{x})^T P^{-1}(\mathbf{x} - \mathbf{m}_\mathbf{x})$$

$$\frac{\partial}{\partial \mathbf{x}} \ln p(\mathbf{x}) = -P^{-1}(\mathbf{x} - \mathbf{m}_\mathbf{x}) \tag{1.31}$$

当 $P^{-1} \to 0$ 时

$$\frac{\partial}{\partial \mathbf{x}} \ln p(\mathbf{x}) = 0 \tag{1.32}$$

当缺乏 $\mathbf{x}$ 的验前概率分布密度时，极大验后估计与极大似然估计是等同的，这可证明如下：

对于极大似然估计，为了求得 $\mathbf{x}$ 的最优估值$\hat{\mathbf{x}}$，应令

$$\frac{\partial \ln p(\mathbf{z}|\mathbf{x})}{\partial \mathbf{x}} = 0 \tag{1.33}$$

上述方程叫做极大似然方程。

对于极大验后估计，为了求得 $\mathbf{x}$ 的最优估值$\hat{\mathbf{x}}$，应令

$$\frac{\partial \ln p(\mathbf{x}|\mathbf{z})}{\partial \mathbf{x}} = 0 \tag{1.34}$$

根据(1.30)式得

$$\ln p(\mathbf{x}|\mathbf{z}) = \ln p(\mathbf{z}|\mathbf{x}) + \ln p(\mathbf{x}) - \ln p(\mathbf{z})$$

$$\frac{\partial \ln p(\mathbf{x}|\mathbf{z})}{\partial \mathbf{x}} = \frac{\partial \ln p(\mathbf{z}|\mathbf{x})}{\partial \mathbf{x}} + \frac{\partial \ln p(\mathbf{x})}{\partial \mathbf{x}} - \frac{\partial \ln p(\mathbf{z})}{\partial \mathbf{x}} = 0$$

考虑到 $p(\mathbf{z})$不是 $\mathbf{x}$ 的函数，同时考虑到(1.32)式，可得

试读结束：需要全本请在线购买： www.ertongbook.com