控制科学与工程学科
研究生教学用书

最优估计理论

Theory of Optimal Estimation

周凤岐　卢晓东　编

高等教育出版社

控制科学与工程学科
研究生教学用书

最优估计理论

Theory of Optimal Estimation

周凤岐　卢晓东　编

高等教育出版社

内容提要

本书基本内容包括参数估计方法，最优线性预测与滤波的基本方程，最优线性平滑的基本方程，滤波的稳定性，滤波的发散及其克服方法，非线性滤波，卡尔曼滤波的应用实例。

本书最大特色是内容集中在最优估计的基本理论和方法，并尽可能与工程应用实例相结合。全书内容由浅入深，通俗易懂，可作为通信、导航、自动控制和电子等学科的硕士研究生教学用书，也可作为工程技术人员的参考书。

图书在版编目(CIP)数据

最优估计理论/周凤岐，卢晓东编. —北京：高等教育出版社，2009.7

ISBN 978-7-04-026479-1

Ⅰ.最… Ⅱ.①周…②卢… Ⅲ.参数估计-最佳化 Ⅳ.O211.67

中国版本图书馆 CIP 数据核字(2009)第 077729 号

策划编辑 许怀镕 **责任编辑** 许海平 **封面设计** 李卫青
责任绘图 尹 莉 **版式设计** 范晓红 **责任校对** 刘 莉
责任印制 尤 静

出版发行	高等教育出版社	**购书热线**	010-58581118
社　　址	北京市西城区德外大街 4 号	**咨询电话**	400-810-0598
邮政编码	100120	**网　　址**	http://www.hep.edu.cn
总　　机	010-58581000		http://www.hep.com.cn
		网上订购	http://www.landraco.com
经　　销	蓝色畅想图书发行有限公司		http://www.landraco.com.cn
印　　刷	北京市南方印刷厂	**畅想教育**	http://www.widedu.com
开　　本	787×960　1/16	**版　　次**	2009 年 7 月第 1 版
印　　张	12.5	**印　　次**	2009 年 7 月第 1 次印刷
字　　数	220 000	**定　　价**	20.30 元

本书如有缺页、倒页、脱页等质量问题，请到所购图书销售部门联系调换。

物料号　26479-00

前　言

在自动控制、航空、航天、航海、导航、通信、机器人、工业生产过程控制等领域中，都会遇到“测量”和“估计”问题。简单地讲，测量是通过仪器、传感器等测量出系统中的某些数据，这些数据通常是含有测量噪声的随机变量或随机过程；估计是从测量(或观测)数据中滤去各种随机噪声，提取所需信息。本书不讨论测量问题，只讨论在测量数据已经得到的情况下，如何从含有随机噪声的测量数据中估计出所需信息——参数或状态。

十几年来，编者所在学科组承担了多项国家重要科研项目，这些项目涉及的领域较广，如卫星等航天器的姿态和轨道确定与控制，导弹等飞行器制导、控制与导航，雷达和红外探测系统的目标检测、图像识别和目标跟踪，运动平台的惯性导航和卫星组合导航等。这些复杂的工程系统都是为了实现对观测数据的精确测量和对系统状态的精确估计，但由于实际系统经常受到各种随机干扰的影响，系统观测信息中夹杂有各种随机噪声，因此如何将系统中有用信号与随机噪声有效正确分离，以及从含有噪声的观测数据中准确估计出所需要的系统信息等都是我们不可回避的实际问题。最优估计理论正是解决这一问题的有效工具，其提供了随机系统信息处理的理论基础和实现算法，使受到随机干扰的系统数据测量和信息估计得以实现。本书将全面、系统地阐述最优估计的基本理论和方法，为读者从事科学研究、阅读相关参考文献和运用其他估计理论知识打下坚实的基础。

本书共分七章，第一章参数估计方法，讨论随机变量的参数估计方法，属静态估计问题；第二章最优线性预测与滤波的基本方程，第三章最优线性平滑的基本方程，主要讨论随机过程的状态估计问题，属动态估计问题；第四章滤波的稳定性，讨论滤波稳定性概念和滤波稳定性定理；第五章滤波的发散及其克服方法，讨论滤波发散的原因及几种常用的克服发散的方法；第六章非线性滤波，主要讨论经典的非线性系统线性化方法，Unscented 卡尔曼滤波和粒子滤波；第七章卡尔曼滤波的应用实例，具体推演了几个工程问题的最优线性滤波和非线性滤波的计算过程。

本书可作为工科院校自动控制、工业自动化、导航制导、通信等学科的研究生教学用书，教学参考学时为 40～60 学时。本书在取材和阐述方式上吸取了参考文献[1]、[2]、[3]的优点，并注重详细的理论分析和数学推导，注意与工程

应用相结合,并附有例题、习题和附录。全书通俗易懂,便于自学,也适合于有关工程技术人员参考和使用。

本书由西北工业大学航天学院周凤岐教授和卢晓东博士合作编写,周凤岐任主编并编写第一至五章;卢晓东编写第六章和第七章。由于编者水平所限,书中可能存在不妥和错误之处,敬请读者批评指正。

编者 于西北工业大学

2009年5月

目　　录

概　述

在科学和技术领域中，经常遇到“估计”问题。所谓“估计”，就是对受到随机干扰和随机测量误差作用的物理系统，以某种性能指标为最优的原则，从具有随机误差的测量数据中提取信息，估计出系统的某些参数和状态变量。这就提出了参数和状态估计问题，这些被估参数或被估状态可统称为被估量。

一般，估计问题分两大类：参数估计和状态估计。

一、参数估计

参数估计属于曲线拟合问题。例如做完某项试验之后，得到若干个观测值 z_i 与相应时间 t_i 的关系 (z_i, t_i) $(i=1,2,\cdots,m)$。我们希望以一条曲线来表示 z 和 t 的关系，设

$$z(t)=x_1h_1(t)+x_2h_2(t)+\cdots+x_nh_n(t)$$

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

通常根据 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、…、x_n。按照什么准则来估计这些参数呢？这将是第一章讨论的主要问题。

二、状态估计

设系统的状态方程和观测方程分别为

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

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

式中，$\boldsymbol{x}(t)$ 为状态变量，它是随时间而变的随机过程，$\boldsymbol{u}(t)$ 为控制变量，$\boldsymbol{w}(t)$ 为系统噪声，$\boldsymbol{v}(t)$ 为测量噪声，$\boldsymbol{z}(t)$ 为观测值。现要根据观测值来估计状态变量 $\boldsymbol{x}(t)$，这就是状态估计问题，这将是第二章及其后各章讨论的问题。

为了正确地解决参数估计和状态估计问题，首先要研究估计方法。在参数估计方面，高斯(Gauss K. F.)于 1795 年在《天体运动理论》一书中提出了最小二乘法。1912 年费歇(Fisher R. A.)提出了极大似然估计方法，从概率密度出发来考虑估计问题。在状态估计方面，1940 年美国学者维纳(Wiener N.)提出在频域中设计统计最优滤波器的方法，称为维纳滤波。同一时期，苏联学者哥尔莫郭洛夫(КОЛМОГОРОВ А. Н.)提出并初次解决了离散平稳随机序列的预测和

外推问题。维纳滤波和哥尔莫郭洛夫滤波方法，局限于处理平稳随机过程，只能提供稳态的最优估值，因此在实际应用上受到很大的限制。1960 年美国学者卡尔曼(Kalman R. E.)和布西(Bucy R. S.)提出了最优递推滤波方法，称为卡尔曼滤波。在这一滤波方法中，考虑了被估量和观测值的统计特性，可用数字计算机来实现。卡尔曼滤波既适用于平稳随机过程，又适用于非平稳随机过程，因此卡尔曼滤波方法得到广泛的应用。

在自动控制中，为了实现最优控制和自适应控制，遇到许多参数估计或状态变量估计问题，这促使估计理论的发展。卡尔曼滤波方法推广到某些类型的非线性系统中，提出了非线性滤波问题。另外，由于数字计算机的迅速发展和广泛使用，使得许多复杂估计问题的解决成为可能，所以近几十年来最优估计理论及其实际应用得到迅速的发展。

第一章　参数估计方法

本章讨论参数估计准则和估计方法，根据对被估值统计特性的掌握程度不同，可提出不同的估计准则。依据不同的准则，就有相应的估计方法，即最小方差估计、线性最小方差估计、极大似然估计、极大验后估计、最小二乘法估计等，本章将对这些估计方法分别进行讨论。

§1.1　最小方差估计与线性最小方差估计

一、最小方差估计

最小方差估计，要求误差的方差为最小，它是一种最古典的估计方法。这种估计方法需要知道被估随机变量 x 的概率密度函数 $p(x)$ 和数学期望 $E(x)$。这种苛刻的先验条件，使此方法在工程上的应用受到很大限制。这里只以一维随机变量的估计为例，介绍最小方差估计方法。

设有一维随机变量 x，它的概率密度函数 $p(x)$ 和数学期望 $E(x)=m_x$ 都是已知的，求 x 的估值 $\hat{x}$。评价估计优劣的准则是 $\hat{x}$ 与 x 的误差的方差为最小，即

$$J = E[(x-\hat{x})^2] = \int_{-\infty}^{+\infty}(x-\hat{x})^2 p(x)\,\mathrm{d}x = \min \tag{1-1}$$

将上式展开，得

$$J=E[(x-\hat{x})^2]=E(x^2)-2\hat{x}E(x)+\hat{x}^2$$

求上式对 $\hat{x}$ 的偏导数，令偏导数等于零，得

$$\frac{\partial J}{\partial \hat{x}}=2\hat{x}-2E(x)=0$$

则 x 的最优估值为

$$\hat{x} = E(x) = \int_{-\infty}^{+\infty} xp(x)\,\mathrm{d}x = m_x \tag{1-2}$$

因此 x 的最小方差估值为 m_x，估计误差为

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

$$E(\tilde{x})=E(x)-E(\hat{x})=E(x)-m_x=m_x-m_x=0$$

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，则最小方差估计是无偏的。因此 x 的估计是无偏估计。

估计误差 $\tilde{x}$ 的方差为

$$E[(x-m_x)^2]=\int_{-\infty}^{+\infty}(x-m_x)^2p(x)\mathrm{d}x=\sigma_x^2 \tag{1-3}$$

所以数学期望 m_x 是 x 的最小方差估计。

这种方法可以推广到多维随机变量的估值，这里不再叙述。

二、线性最小方差估计

线性最小方差估计就是估计值为观测值的线性函数，估计误差的方差为最小。在使用这种方法时，需要知道观测值和被估值的一、二阶矩，即数学期望 $E(z)$ 和 $E(x)$、方差 $\mathrm{Var}(z)$ 和 $\mathrm{Var}(x)$ 及协方差 $\mathrm{Cov}(x,z)$ 和 $\mathrm{Cov}(z,x)$。

先讨论被估值 x 和观测值 z 都是一维随机变量的情况。线性最小方差估计是把 x 的估值 $\hat{x}$ 表示成 z 的线性函数，即

$$\hat{x}=az+b \tag{1-4}$$

式中，a 和 b 为两个待定常数。根据估计误差的方差

$$J=E[(x-\hat{x})^2]=E\{[x-(az+b)]^2\}=\min \tag{1-5}$$

的条件来确定系数 a 和 b。

求式(1-5)对 a 和 b 的偏导数，令偏导数等于零，可求得 a 和 b 两个系数。

$$\frac{\partial J}{\partial a}=-2E\{[x-(az+b)]z\}=0 \tag{1-6}$$

$$\frac{\partial J}{\partial b}=-2E[x-(az+b)]=0 \tag{1-7}$$

从式(1-7)可得

$$m_x-am_z-b=0$$

式中，m_x 和 m_z 为 x 和 z 的数学期望，从此式可得

$$b=m_x-am_z \tag{1-8}$$

将式(1-8)代入式(1-6)，得

$$E[(x-az-m_x+am_z)z]=0$$

把上式改写成

$$E\{[(x-m_x)-a(z-m_z)](z-m_z+m_z)\}=0$$

展开上式得

$$E[(x-m_x)(z-m_z)]+m_zE(x-m_x)-aE[(z-m_z)^2]-am_zE(z-m_z)=0$$

求上式的数学期望值，可得

$$
\begin{aligned}
&\mathrm{Cov}(x,z)-a\sigma_z^2=0\\
&a=\frac{\mathrm{Cov}(x,z)}{\sigma_z^2}=\frac{\gamma_{xz}\sigma_x\sigma_z}{\sigma_z^2}=\frac{\gamma_{xz}\sigma_x}{\sigma_z}
\end{aligned}
\tag{1-9}
$$

式中，σ_x、σ_z 分别为随机变量 x 和 z 的均方根差，γ_{xz}为 x 和 z 的相关系数

$$\gamma_{xz}=\mathrm{Cov}(x,z)/\sigma_x\sigma_z$$

于是 x 的估值为

$$\hat{x}=az+b=m_x+\frac{\mathrm{Cov}(x,z)}{\sigma_z^2}(z-m_z) \tag{1-10}$$

估计误差为

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

$$E(\tilde{x})=E(x)-E(m_x)-\frac{\gamma_{xz}\sigma_x}{\sigma_z}E(z-m_z)=m_x-m_x-\frac{\gamma_{xz}\sigma_x}{\sigma_z}(m_z-m_z)=0$$

因此 $E(\hat{x})=E(x)$。所以估计是无偏的。

下面讨论 $\boldsymbol{x}$ 和 $\boldsymbol{z}$ 都是多维随机变量的估计问题。设 $\boldsymbol{x}$ 为 n 维，$\boldsymbol{z}$ 为 q 维，已知 $\boldsymbol{x}$ 和 $\boldsymbol{z}$ 的一、二阶矩，即 $E(\boldsymbol{x})$、$E(\boldsymbol{z})$、$\mathrm{Var}(\boldsymbol{x})$、$\mathrm{Var}(\boldsymbol{z})$、$\mathrm{Cov}(\boldsymbol{x},\boldsymbol{z})$和 $\mathrm{Cov}(\boldsymbol{z},\boldsymbol{x})$。

假定 $\boldsymbol{x}$ 的估值 $\hat{\boldsymbol{x}}$ 是 $\boldsymbol{z}$ 的线性函数

$$\hat{\boldsymbol{x}}(\boldsymbol{z})=\boldsymbol{b}+\boldsymbol{A}\boldsymbol{z} \tag{1-11}$$

式中，$\boldsymbol{b}$ 是 n 维非随机常数向量，$\boldsymbol{A}$ 是 $n\times q$ 非随机常数矩阵。

估计误差方差矩阵为

$$\boldsymbol{J}=E\{[\boldsymbol{x}-\hat{\boldsymbol{x}}(\boldsymbol{z})][\boldsymbol{x}-\hat{\boldsymbol{x}}(\boldsymbol{z})]^{\mathrm{T}}\}=E[(\boldsymbol{x}-\boldsymbol{b}-\boldsymbol{A}\boldsymbol{z})(\boldsymbol{x}-\boldsymbol{b}-\boldsymbol{A}\boldsymbol{z})^{\mathrm{T}}] \tag{1-12}$$

估计准则是方差矩阵 $\boldsymbol{J}$ 为最小，也可等价为方差矩阵 $\boldsymbol{J}$ 的迹 J_t 为最小，即 $\tilde{\boldsymbol{x}}$ 的各分量的方差之和为最小

$$
\begin{aligned}
J_t&=\mathrm{Trace}\ E[(\boldsymbol{x}-\boldsymbol{b}-\boldsymbol{A}\boldsymbol{z})(\boldsymbol{x}-\boldsymbol{b}-\boldsymbol{A}\boldsymbol{z})^{\mathrm{T}}]\\
&=E[(\boldsymbol{x}-\boldsymbol{b}-\boldsymbol{A}\boldsymbol{z})^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{b}-\boldsymbol{A}\boldsymbol{z})]=\min
\end{aligned}
\tag{1-13}
$$

用标量函数对矩阵的微分法则（附录一），求$\frac{\partial J_t}{\partial \boldsymbol{b}}$和$\frac{\partial J_t}{\partial \boldsymbol{A}}$，令$\frac{\partial J_t}{\partial \boldsymbol{b}}=\boldsymbol{0}$ 和$\frac{\partial J_t}{\partial \boldsymbol{A}}=\boldsymbol{0}$，联立求解可得 $\boldsymbol{b}$ 和 $\boldsymbol{A}$。

$$
\begin{aligned}
&\frac{\partial J_t}{\partial \boldsymbol{b}}=-2E(\boldsymbol{x}-\boldsymbol{b}-\boldsymbol{A}\boldsymbol{z})=2(\boldsymbol{b}+\boldsymbol{A}\boldsymbol{m}_z-\boldsymbol{m}_x)=\boldsymbol{0}\\
&\boldsymbol{b}=\boldsymbol{m}_x-\boldsymbol{A}\ \boldsymbol{m}_z=E(\boldsymbol{x})-\boldsymbol{A}E(\boldsymbol{z})
\end{aligned}
\tag{1-14}
$$

$$
\begin{aligned}
\frac{\partial J_t}{\partial \boldsymbol{A}}&=\frac{\partial}{\partial \boldsymbol{A}}E[(\boldsymbol{x}-\boldsymbol{b}-\boldsymbol{A}\boldsymbol{z})^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{b}-\boldsymbol{A}\boldsymbol{z})]\\
&=-2E(\boldsymbol{x}-\boldsymbol{b}-\boldsymbol{A}\boldsymbol{z})\boldsymbol{z}^{\mathrm{T}}=-2E[(\boldsymbol{x}-\boldsymbol{b})]\boldsymbol{z}^{\mathrm{T}}+2E(\boldsymbol{A}\boldsymbol{z}\ \boldsymbol{z}^{\mathrm{T}})\\
&=2[\boldsymbol{A}E(\boldsymbol{z}\ \boldsymbol{z}^{\mathrm{T}})+\boldsymbol{b}E(\boldsymbol{z}^{\mathrm{T}})-E(\boldsymbol{x}\ \boldsymbol{z}^{\mathrm{T}})]=\boldsymbol{0}
\end{aligned}
\tag{1-15}
$$

将式(1-14)代入式(1-15),得

$$\boldsymbol{A}E(\boldsymbol{z}\boldsymbol{z}^{\mathrm{T}})+E(\boldsymbol{x})E(\boldsymbol{z}^{\mathrm{T}})-\boldsymbol{A}E(\boldsymbol{z})E(\boldsymbol{z}^{\mathrm{T}})-E(\boldsymbol{x}\boldsymbol{z}^{\mathrm{T}})=\boldsymbol{0}$$

$$\boldsymbol{A}[E(\boldsymbol{z}\boldsymbol{z}^{\mathrm{T}})-E(\boldsymbol{z})E(\boldsymbol{z}^{\mathrm{T}})]-[E(\boldsymbol{x}\boldsymbol{z}^{\mathrm{T}})-E(\boldsymbol{x})E(\boldsymbol{z}^{\mathrm{T}})]=\boldsymbol{0}$$

$$\boldsymbol{A}E\{[\boldsymbol{z}-E(\boldsymbol{z})][\boldsymbol{z}-E(\boldsymbol{z})]^{\mathrm{T}}\}-E\{[\boldsymbol{x}-E(\boldsymbol{x})][\boldsymbol{z}-E(\boldsymbol{z})]^{\mathrm{T}}\}=\boldsymbol{0}$$

$$\boldsymbol{A}\mathrm{Var}(\boldsymbol{z})-\mathrm{Cov}(\boldsymbol{x},\boldsymbol{z})=\boldsymbol{0}$$

因此

$$\boldsymbol{A}=\mathrm{Cov}(\boldsymbol{x},\boldsymbol{z})[\mathrm{Var}(\boldsymbol{z})]^{-1} \tag{1-16}$$

将式(1-16)代入式(1-14),可得

$$\boldsymbol{b}=E(\boldsymbol{x})-\mathrm{Cov}(\boldsymbol{x},\boldsymbol{z})[\mathrm{Var}(\boldsymbol{z})]^{-1}E(\boldsymbol{z}) \tag{1-17}$$

根据式(1-16)和式(1-17)求得 $\boldsymbol{A}$ 和 $\boldsymbol{b}$,代入式(1-11),得

$$\begin{aligned}\hat{\boldsymbol{x}}&=E(\boldsymbol{x})+\mathrm{Cov}(\boldsymbol{x},\boldsymbol{z})[\mathrm{Var}(\boldsymbol{z})]^{-1}[\boldsymbol{z}-E(\boldsymbol{z})]\\&=\boldsymbol{m}_x+\mathrm{Cov}(\boldsymbol{x},\boldsymbol{z})[\mathrm{Var}(\boldsymbol{z})]^{-1}(\boldsymbol{z}-\boldsymbol{m}_z)\end{aligned} \tag{1-18}$$

由式(1-18)可得

$$E(\hat{\boldsymbol{x}})=\boldsymbol{m}_x+\mathrm{Cov}(\boldsymbol{x},\boldsymbol{z})[\mathrm{Var}(\boldsymbol{z})]^{-1}E(\boldsymbol{z}-\boldsymbol{m}_z)=\boldsymbol{m}_x=E(\boldsymbol{x})$$

所以估计是无偏的。

估计误差的方差矩阵为

$$\boldsymbol{J}=\mathrm{Var}(\boldsymbol{x})-\mathrm{Cov}(\boldsymbol{x},\boldsymbol{z})[\mathrm{Var}(\boldsymbol{z})]^{-1}\mathrm{Cov}(\boldsymbol{z},\boldsymbol{x}) \tag{1-19}$$

§1.2 极大似然法估计与极大验后法估计

一、极大似然法估计

极大似然法估计是以观测值出现的概率为最大作为估计准则,它是一种普通的参数估计方法。

设 z 是一维连续随机变量,其概率密度函数为 $p(z,\theta_1,\theta_2,\cdots,\theta_n)$,含有 n 个未知参数 $\theta_1,\theta_2,\cdots,\theta_n$。把 k 个独立观测值 $z_1,z_2,\cdots,z_k$ 分别代入 $p(z,\theta_1,\theta_2,\cdots,\theta_n)$ 中的 z,则得

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

将所得的 k 个概率密度函数相乘,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。极大似然法估计的实质就是求出使 L 达到极大时,$\theta_1,\theta_2,\cdots,\theta_n$ 的估值 $\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_n$。从式(1-20)可看到,$\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_n$ 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 $\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_n$，对式(1-20)取对数，则

$$\ln L=\sum_{i=1}^{k}\ln p(z_i,\theta_1,\theta_2,\cdots,\theta_n) \tag{1-21}$$

由于对数函数是单调递增函数，因此当 L 取极大值时，$\ln L$ 也同时取极大值，将式(1-21)分别对 $\theta_1,\theta_2,\cdots,\theta_n$ 求偏导数，令偏导数等于零，可得下列方程组

$$\begin{cases}\dfrac{\partial}{\partial\theta_1}\ln L=0\\ \quad\vdots\\ \dfrac{\partial}{\partial\theta_n}\ln L=0\end{cases} \tag{1-22}$$

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 $\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_n$，使 $z_1,z_2,\cdots,z_k$ 最有可能出现，并不需要 $\theta_1,\theta_2,\cdots,\theta_n$ 的验前知识，即不需要知道 $\theta_1,\theta_2,\cdots,\theta_n$ 的概率密度函数和一、二阶矩。

例 1-1　设有正态分布随机变量 z，给出 k 个观测值 $z_1,z_2,\cdots,z_k$。观测值相互独立，试根据这 k 个观测值，确定概率密度中的各参数。

解　z 的概率密度可用下式表示

$$p(z,m,\sigma)=\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left[-\frac{(z-m)^2}{2\sigma^2}\right]$$

式中的 m 和 σ 为未知参数。现用极大似然法来确定参数 m 和 σ。令似然函数为

$$L(z_1,z_2,\cdots,z_k;m,\sigma)=\prod_{i=1}^{k}\left\{\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left[-\frac{(z_i-m)^2}{2\sigma^2}\right]\right\}$$

对上式取对数，可得

$$\begin{aligned}\ln L(z_1,z_2,\cdots,z_k;m,\sigma)&=\sum_{i=1}^{k}\ln\left\{\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left[-\frac{(z_i-m)^2}{2\sigma^2}\right]\right\}\\&=-\sum_{i=1}^{k}\frac{(z_i-m)^2}{2\sigma^2}+k\ln\frac{1}{\sqrt{2\pi}}-k\ln\sigma\end{aligned}$$

将上式分别对 m 和 σ 求偏导数，令偏导数等于零，可得

$$\frac{\partial\ln L}{\partial m}=\sum_{i=1}^{k}\frac{z_i-m}{\sigma^2}=0$$

$$\frac{\partial\ln L}{\partial\sigma}=\frac{1}{\sigma^3}\sum_{i=1}^{k}(z_i-m)^2-\frac{k}{\sigma}=0$$

联立求解可得

$$\hat{m}=\frac{\sum_{i=1}^{k} z_i}{k},\qquad \hat{\sigma}^2=\frac{\sum_{i=1}^{k}(z_i-\hat{m})^2}{k}$$

上面介绍了极大似然法的基本概念。现在来讨论极大似然法估计参数的问题。

设 z 为 m 维随机变量，$\boldsymbol{x}$ 为 n 维未知参数，假定已知 z 的条件概率密度函数 $p(\boldsymbol{z}/\boldsymbol{x})$。现在得到 k 组 z 的观测值 $z_1, z_2, \cdots, z_k$。观测值相互独立。当参数 $\boldsymbol{x}$ 是何值时，$z_1, z_2, \cdots, z_k$ 出现的可能性最大？为此，确定似然函数

$$L(z,x)=p(z_1/\boldsymbol{x})p(z_2/\boldsymbol{x})\cdots p(z_k/\boldsymbol{x})=p(\boldsymbol{z}/\boldsymbol{x}) \tag{1-23}$$

或

$$\ln L(z,\boldsymbol{x})=\ln p(\boldsymbol{z}/\boldsymbol{x}) \tag{1-24}$$

求出使 L 为极大的 $\boldsymbol{x}$ 值，令

$$\frac{\partial L}{\partial \boldsymbol{x}}=\boldsymbol{0} \quad 或 \quad \frac{\partial \ln L}{\partial \boldsymbol{x}}=\boldsymbol{0} \tag{1-25}$$

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

L 取极大值的充分条件是

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

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 $\boldsymbol{x}$ 的估值 $\hat{\boldsymbol{x}}$。

设有一随机线性观测系统

$$\boldsymbol{z}=\boldsymbol{h}(\boldsymbol{x},\boldsymbol{v})=\boldsymbol{H}\boldsymbol{x}+\boldsymbol{v} \tag{1-26}$$

式中，$\boldsymbol{z}$ 是 m 维观测值，$\boldsymbol{x}$ 是 n 维未知参数，$\boldsymbol{v}$ 是 m 维测量误差。设 $\boldsymbol{v}$ 与 $\boldsymbol{x}$ 独立。给出 $\boldsymbol{v}$ 的统计特性，求 $\boldsymbol{x}$ 的极大似然估计。

下面求似然函数

$$L(\boldsymbol{z},\boldsymbol{x})=p(\boldsymbol{z}/\boldsymbol{x})=\frac{p(\boldsymbol{x},\boldsymbol{z})}{p(\boldsymbol{x})}$$

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

$$p(\boldsymbol{x},\boldsymbol{z})=p[\boldsymbol{x},(\boldsymbol{H}\boldsymbol{x}+\boldsymbol{v})]=p(\boldsymbol{x},\boldsymbol{v})=p(\boldsymbol{x})p(\boldsymbol{v})$$

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

令

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

解上式，可得 $\boldsymbol{x}$ 的估值 $\hat{\boldsymbol{x}}$。

假定噪声 $\boldsymbol{v}$ 是正态分布，其均值为零，方差矩阵为 $E(\boldsymbol{v}\boldsymbol{v}^{\mathrm{T}})=\boldsymbol{R}$，则

$$L(\boldsymbol{z},\boldsymbol{x})=p(\boldsymbol{v})=\frac{1}{(\sqrt{2\pi})^{m}|\boldsymbol{R}|^{1/2}}\exp\left(-\frac{1}{2}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{v}\right)$$

把 $\boldsymbol{v}=\boldsymbol{z}-\boldsymbol{H}\boldsymbol{x}$ 代入上式，得

$$L(\boldsymbol{z},\boldsymbol{x})=p(\boldsymbol{z}-\boldsymbol{H}\boldsymbol{x})=c\cdot\exp\left[-\frac{1}{2}(\boldsymbol{z}-\boldsymbol{H}\boldsymbol{x})^{\mathrm{T}}\boldsymbol{R}^{-1}(\boldsymbol{z}-\boldsymbol{H}\boldsymbol{x})\right]$$

式中

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

求出 $\boldsymbol{x}$，使 $L(\boldsymbol{z},\boldsymbol{x})=p(\boldsymbol{z}-\boldsymbol{H}\boldsymbol{x})$ 为最大，也就是使

$$\boldsymbol{J}=\frac{1}{2}(\boldsymbol{z}-\boldsymbol{H}\boldsymbol{x})^{\mathrm{T}}\boldsymbol{R}^{-1}(\boldsymbol{z}-\boldsymbol{H}\boldsymbol{x})=\min \tag{1-27}$$

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

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

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

$$\hat{\boldsymbol{x}}=(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{H})^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{z} \tag{1-28}$$

二、极大验后法估计

如果给出 n 维随机变量 $\boldsymbol{x}$ 的条件概率密度 $p(\boldsymbol{x}/\boldsymbol{z})$——也称验后概率密度，怎样求 $\boldsymbol{x}$ 的最优估值 $\hat{\boldsymbol{x}}$ 呢？极大验后法估计准则：使 $\boldsymbol{x}$ 的验后概率密度 $p(\boldsymbol{x}/\boldsymbol{z})$ 达到最大的那个 $\boldsymbol{x}$ 值为极大验后法估值 $\hat{\boldsymbol{x}}$。可见，极大验后法估计是已知 $p(\boldsymbol{x}/\boldsymbol{z})$ 求 $\boldsymbol{x}$ 的最优估值 $\hat{\boldsymbol{x}}$ 的一种有效方法。

极大验后法估计是以已知 $p(\boldsymbol{x}/\boldsymbol{z})$ 为前提的。如果只知道 $p(\boldsymbol{z}/\boldsymbol{x})$，可按下式计算 $p(\boldsymbol{x}/\boldsymbol{z})$。

$$p(\boldsymbol{x}/\boldsymbol{z})=\frac{p(\boldsymbol{z}/\boldsymbol{x})p(\boldsymbol{x})}{p(\boldsymbol{z})} \tag{1-29}$$

式中，$p(\boldsymbol{x})$ 是 $\boldsymbol{x}$ 的验前概率密度，$p(\boldsymbol{z})$ 是观测值 $\boldsymbol{z}$ 的概率密度，$p(\boldsymbol{z}/\boldsymbol{x})$ 可用计算方法或实验方法求得。为了计算 $p(\boldsymbol{x}/\boldsymbol{z})$ 需要知道 $p(\boldsymbol{x})$。在 $\boldsymbol{x}$ 没有验前知识可供利用时，可假定 $\boldsymbol{x}$ 在很大范围内变化。在这种情况下，可把 $\boldsymbol{x}$ 的验前概率密度 $p(\boldsymbol{x})$ 近似地看作方差矩阵趋于无限大的正态分布密度函数，即

$$p(\boldsymbol{x})=\frac{1}{(\sqrt{2\pi})^{n}|\boldsymbol{P}|^{1/2}}\exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{m}_x)^{\mathrm{T}}\boldsymbol{P}^{-1}(\boldsymbol{x}-\boldsymbol{m}_x)\right]$$

式中，$\boldsymbol{P}$ 为 $\boldsymbol{x}$ 的方差矩阵，$\boldsymbol{P}\to\infty\cdot\boldsymbol{I}$，$\boldsymbol{I}$ 为 $n\times n$ 单位矩阵，$\boldsymbol{P}^{-1}\to\boldsymbol{0}$，于是

$$\ln p(\boldsymbol{x})=-\ln[(2\pi)^{n/2}|\boldsymbol{P}|^{1/2}]-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{m}_x)^{\mathrm{T}}\boldsymbol{P}^{-1}(\boldsymbol{x}-\boldsymbol{m}_x)$$

$$\frac{\partial}{\partial \boldsymbol{x}}\ln p(\boldsymbol{x}) = -\boldsymbol{P}^{-1}(\boldsymbol{x}-\boldsymbol{m}_x) \tag{1-30}$$

当$\boldsymbol{P}^{-1}\to\boldsymbol{0}$时,有

$$\frac{\partial}{\partial \boldsymbol{x}}\ln p(\boldsymbol{x}) = \boldsymbol{0} \tag{1-31}$$

当缺乏 $\boldsymbol{x}$ 的验前概率密度时,极大验后法估计与极大似然法估计是等同的,现证明如下:

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

$$\frac{\partial \ln p(\boldsymbol{z}/\boldsymbol{x})}{\partial \boldsymbol{x}} = \boldsymbol{0} \tag{1-32}$$

根据式(1-29)得

$$\ln p(\boldsymbol{x}/\boldsymbol{z}) = \ln p(\boldsymbol{z}/\boldsymbol{x}) + \ln p(\boldsymbol{x}) - \ln p(\boldsymbol{z})$$

$$\frac{\partial \ln p(\boldsymbol{x}/\boldsymbol{z})}{\partial \boldsymbol{x}} = \frac{\partial \ln p(\boldsymbol{z}/\boldsymbol{x})}{\partial \boldsymbol{x}} + \frac{\partial \ln p(\boldsymbol{x})}{\partial \boldsymbol{x}} - \frac{\partial \ln p(\boldsymbol{z})}{\partial \boldsymbol{x}} = \boldsymbol{0} \tag{1-33}$$

考虑到 $p(\boldsymbol{z})$ 不是 $\boldsymbol{x}$ 的函数,同时考虑到式(1-31),可得

$$\frac{\partial \ln p(\boldsymbol{x}/\boldsymbol{z})}{\partial \boldsymbol{x}} = \frac{\partial \ln p(\boldsymbol{z}/\boldsymbol{x})}{\partial \boldsymbol{x}} \tag{1-34}$$

一般说来,由于计算似然函数比计算验后概率密度较为简单,极大似然法估计比极大验后法估计应用普遍。

§1.3 最小二乘法估计与加权最小二乘法估计

上面讨论的几种估计方法,分别对被估随机变量 $\boldsymbol{x}$ 的概率密度函数 $p(\boldsymbol{x})$、条件概率密度 $p(\boldsymbol{z}/\boldsymbol{x})$、$p(\boldsymbol{x}/\boldsymbol{z})$ 以及一、二阶矩等条件有着不同的要求。假定我们并不掌握上述任何条件,仍要估计随机变量 $\boldsymbol{x}$ 的最优估值 $\hat{\boldsymbol{x}}$,只有用高斯提出的最小二乘法。

一、最小二乘法估计

设 m 次独立试验,得到 m 对观测值:(z_1,t_1),(z_2,t_2),$\cdots$,(z_m,t_m)。这里 t_i 表示时间或其他物理量。现在的任务是:根据这些观测值,用最优的形式来表示 z 与 t 之间的函数关系。

通常,z 的未知函数可用 $f(t)$ 表示,$f(t)$ 的类型应根据这 m 对数据(m 个点)的分布情况或所研究问题的物理性质来确定。为了便于计算,可采用多项式

$$f(t) = x_1 + x_2 t + x_3 t^2 + \cdots + x_n t^{n-1} \tag{1-35}$$

来表示,也可以用更一般的形式表示