现代控制理论

张洪钺 主编

（第 三 册）

最佳估计理论

陈新海 编

北京航空學院出版社

现代控制理论

张洪钺　主编

（第三册）

最佳估计理论

陈新海　编

北京航空学院出版社

内 容 简 介

本书比较全面地介绍了现代控制理论的基本内容，附有大量的例题和习题，切合工程实际，便于自学。全书分为四篇，第一篇线性系统理论，第二篇最优控制理论，第三篇最佳估计理论，第四篇系统辨识。分四册出版，每册一篇。

本书是第三册，主要介绍最优估计理论的估计方法，线性最优预测、滤波和平滑、滤波和稳定性、滤波的发散及其克服、非线性滤波、次优滤波以及随机最优控制等。

本书主要作为高等院校自动控制专业研究生课的教材，也可供从事自动控制工作的科技人员参考。

Z K01/36

现代控制理论　第三册
——最佳估计理论
张洪钺 主编　陈新海 编
责任编辑　郭维烈 宋淑乔 严文璇
北京航空学院出版社出版
新华书店北京发行所发行　各地新华书店经售
昌平振兴胶印厂印装
787×1092　1/16　印张：14.125　字数：361千字
1987年12月第一版　1987年12月第一次印刷　印数：6000 册
统一书号：15432·081　定价：2.4元
ISBN 7-81012-028-X TP·006

前　言

目前，自动控制技术已广泛地应用于工农业生产，交通运输和国防建设。指导自动控制系统的分析和设计的控制理论也有了很大的发展。在四十和五十年代中发展起来的经典控制理论被成功地应用于单输入—单输出定常系统的分析和设计。在五十年代末、六十年代初，首先由航天、航空技术需要而发展起来的现代控制理论具有更广泛的适用性，它可用于多输入—多输出、定常或时变系统的分析和设计，它所讨论的问题也更复杂和深入。在系统的性能指标上，现代控制理论提出了最优的概念，因此产生了庞特里亚金的极大値原理，贝尔曼的动态规划以及二次型指标线性反馈控制等最优控制的理论。在对系统的描述上，现代控制理论应用了状态空间的概念，揭示了系统的内部特性以及和外部特性的关系，由此发现了可控性和可观测性这样的系统的基本特性，为系统的分析设计打下了深刻的理论基础。现代控制理论考虑了干扰所引起的系统状态的不确定性，提出了对状态进行最佳估计的理论和方法。建立系统的数学模型，根据模型进行最佳估计和控制是现代控制理论的一个重要特点，固此系统辨识和建模也就成了现代控制理论的一个重要内容。

在我国社会主义四个现代化的进程中，现代控制理论会起到它应有的作用。为了适应这种需要，我们编写了这套现代控制理论的教科书。本书由四篇组成，分四册出版，每册一篇。第一篇是线性系统理论，主要介绍了系统的可控性、可观测性和稳定性，以及系统的标准形、实现、极点配置、解耦、观测器、动态补偿器等。第二篇是最优控制理论，主要介绍了最优控制中的变分法、庞特里亚金的极大（小）原理、贝尔曼的动态规划以及二次型指标最优线性反馈控制等。第三篇是最优估计理论，主要介绍了估计方法，线性最优预测、滤波和平滑、滤波的稳定性、滤波的发散及其克服、非线性滤波、次优滤波以及随机最优控制等。第四篇是系统辨识，主要介绍了脉冲响应的辨识，各种类型的最小二乘法、辅助变量法、极大似然法、随机逼近法以及多变量系统的辨识、闭环系统的辨识和辨识试验中的一些问题。

本书为工科院校自动控制专业研究生教材，在取材和阐述方式上注意了工程性，附有大量例题和习题，推导也比较详细，便于自学。本书第一篇由程鹏编写，第二篇由孟宪仲编写，第三篇由陈新海编写，第四篇由胡干耀编写。主编张洪钺负责全书编写的组织讨论工作和统一审阅，修改定稿。徐缤昌、林道垣、王纪文、胡春松、林其璈、沈正华、陈天良等对全书提出了不少宝贵的意见。全书最后由高为炳、邬学礼负责审查。对这些同志，作者谨表示深切的感谢。

由于水平所限，书中可能存在许多不妥和错误之处，敬請读者批评指正。

目　录

第三篇　最优估计理论

此为试读，需要完整PDF请访问：www.ertongbook.com

第三篇 最优估计理论

概 述

在自动控制、航空与航天、通讯、导航和工业生产等领域中，愈来愈多地遇到“估计”问题。所谓“估计”，简单地说，就是从观测数据中提取信息。例如，在做实验时，为了便于说明问题，常把实验结果用曲线的形式表示，需要根据观测数据来估计描述该曲线的方程中的某些参数，这一过程叫做参数估计，这些被估计的参数都是随机变量。再举一个例子，在飞行器导航中，要从带有随机干扰的观测数据中，估计出飞行器的位置，速度和加速度等运动状态变量，这就遇到状态变量的估计问题，这些状态变量都是随 机 过 程。因此，“估计”的任务就是从带有随机误差的观测数据中估计出某些参数或某些状态变量，这些被估参数或状态变量可统称为被估量。

我们希望估计出来的参数或状态变量愈接近实际值愈好。为了衡量估计的好坏，必须要有一个衡量估计的准则，衡量估计的准则可能是各式各样的。我们总是希望估计是最好的，因此提出最优估计问题。所谓最优估计，是指在某一确定的准则条件下，按照某种统计意义上来说，估计达到最优。很显然，最优估计不是唯一的，它随着准则不同而不同，因此在估计时，要恰当选择衡量估计的准则。

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。最早的估计方法是高斯（*Gauss*，*K.F.*）于1795 年在他的关于《天体运动理论》一书中提出的最小二乘法。最小二乘法沒有考虑被估参数和观测数据的统计特性，因此这种方法不是最优估计方法。另一方面由于最小二乘法在计算上比较简单，因此它是一种用得最广泛的估计方法。在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，对估计理论作出了重大贡献。对于随机过程的估计问题，到本世纪三十年代才积极开展起来。主要成果为1940年美国学者维纳(*Wiener*，*N*）所提出的在频域中设计统计最优滤波器的方法，这一方法称为维纳滤波。同一时期，苏联学者哥尔莫郭洛夫(*А.Н.КОЛМОГОРОВ*）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，并只能提供稳态的最优估值。这些滤波方法在工程实践上也遇到许多困难，因此在实际应用上受到一定的限制。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，这种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。另外，由于电子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，这又促进了估计理论的发展。所以近二十年来最优估计理论及其实际应用得到快速的发展。

本篇的重点是讨论基本的估计方法、卡尔曼滤波和随机线性系统的最优控制。全篇共分九章，主要内容如下：

第一章 估计方法

主要介绍最小方差估计，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等方法，为以后学习系统辨识和卡尔曼滤波作准备。

第二章 最优线性预测和滤波的基本方程

主要给出卡尔曼滤波基本方程的推导，在基本方程的基础上给出各种情况下的卡尔曼滤波方程。

第三章 最优线性平滑的基本方程

介绍固定区间最优平滑，固定点最优平滑和固定滞后最优平滑。

第四章 滤波的稳定性

扼要地介绍滤波的稳定性概念及与滤波稳定性定理的证明有关的随机线性系统的可控性和可观测性，详细地给出了稳定性定理的证明及定常线性系统滤波稳定性的判别。

第五章 滤波的发散及其克服方法

主要讨论产生滤波发散的原因及克服滤波发散的各种方法，如衰减记忆滤波，限定记忆滤波，平方根滤波和自适应滤波等。

第六章 非线性滤波

介绍围绕标称轨道线性化滤波方法和推广的卡尔曼滤波方法，以及其他的非线性滤波方法。

第七章 次优滤波器设计与灵敏度分析

讨论卡尔曼滤波在计算上对计算机速度和容量的要求，次优滤波和灵敏度分析。

第八章 随机线性系統的最优控制

本章给出了线性二次型问题的分离定理的证明。把线性系统的最优控制问题和优最滤波问题结合在一起。

第九章 卡尔曼滤波在航空中的应用

简单地介绍卡尔曼滤波在飞机和导弹控制中的应用。

本教材主要是为工科院校的研究生和有关工程技术人员而编写的，某些基本内容也可供本科生选修之用。学习本教材所需的数学基础，如线性代数，概率论和随机过程，都不超过工科院校的数学大纲范围。在编写中力求概念清楚，简明易懂。为了保持理论的系统性和严谨性，对估计理论的来龙去脉都作了说明，对基本公式和基本定理都用工科学生易于接受的方法作出了详细地推导或证明。在内容的选择上尽量联系实际，以适应工程实践的需要。

第一章 估计方法

本章先讨论估计问题的提法，以及估计时所依据的准则。根据对被估值统计特性的掌握程度不同，提出下列一些估计准则：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。对用得比较广泛的线性最小方差估计、极大似然估计和最小二乘估计都讨论得比较详细。

§1—1 估計問題的提法和估計准則

下面先讨论怎样把一个在实际中遇到的估计问题用合适的数学式子来表达，以便进行估计。一般，估计问题可分成二大类，即参数估计和状态估计，分别讨论如下：

一、参数估计

例如我们做完试验之后，得到观测值z与时间t的关系如图1—1所示。我们希望用一条曲线来表示z与t的关系，设

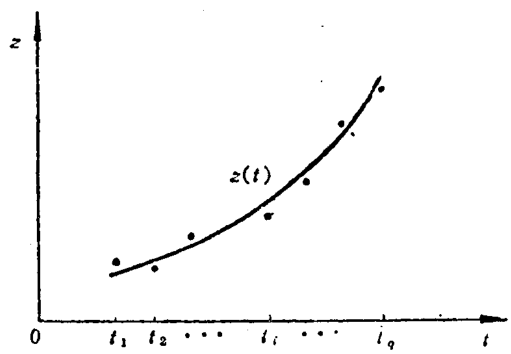

图1—1 拟合曲綫图

$$z(t)=X_1h_1(t)+X_2h_2(t)+\cdots+X_nh_n(t) \qquad (1—1)$$

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等。X_1、X_2、$\cdots$、X_n为n个未知参数，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。

在上述这类问题中，一般，被估参数X_1、X_2、$\cdots$、X_n是不随时间而变的随机变量。有的被估参数可能缓慢地随时间而变。本章主要讨论参数估计。

二、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

$$\dot{\mathbf{X}}(t)=\mathbf{A}(t)\mathbf{X}(t)+\mathbf{B}(t)\mathbf{U}(t)+\mathbf{F}(t)\mathbf{W}(t)$$

要求根据观测值

$$\mathbf{Z}(t)=\mathbf{H}(t)\mathbf{X}(t)+\mathbf{V}(t)$$

来估计状态变量$X(t)$。在上述二式中，$W(t)$和$V(t)$都为随机干扰，$U(t)$为控制量。$X(t)$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。卡尔曼滤波方法就是一种状态估计方法，将在以后各章中讨论。

现在来讨论估计准则问题。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。根据我们对观测值Z和被估值X的统计特性的掌握程度，可有下列的估计准则和估计方法。

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准则的。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X和观测值Z的条件概率分布密度$P(x|z)$或$P(z)$以及它们的联合概率分布密度$P(x, z)$。

2. 极大似然准则（极大似然估计）

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

3. 极大验后准则（极大验后估计）

极大验后准则是使验后概率分布密度$P(x|z)$达到极大的那个X值作为估值的。按这种估计准则求得的X最优估值称为极大验后估计。为了求出极大验后估计，需要知道验后概率分布密度。

4. 线性最小方差准则（线性最小方差估计）

在上面已谈到，为了进行最小方差估计和极大验后估计，需要知道$P(x|z)$；为了进行极大似然估计，需要知道$P(z|x)$。如果我们放松对概率密度知识的要求，只要求知道观测值和被估值的一、二阶矩，即$E[X]$，$E[Z]$，$Var\,X$，$Var\,Z$，$COV(X, Z)$。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，以估计误差的方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X和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。最小二乘法是以残差的平方和为最小作为估计准则的，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 最小方差估計

在参数估计中，经常采用最小方差准则，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。

一、一维随机变量

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，现在要用一个常数a作为X的估值$\hat{X}$（以后在表示估值的字母上方加上记号∧）。评价估计优劣的准则是$\hat{X}$与X的误差的方差为最小，即

$$J=E[(X-a)^2]=\int_{-\infty}^{+\infty}(x-a)^2p(x)dx=min \tag{1-2}$$

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

$$\hat{X}=a=E[X]=\int_{-\infty}^{\infty}xp(x)dx=m_x \tag{1-3}$$

因此X的最小方差估值为m_x，估计误差为

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

（以后在表示误差的字母上方都加上记号～）

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

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

$$E[(X-m_x)^2]=\int_{-\infty}^{\infty}(x-m_x)^2p(x)dx=\sigma_x^2 \tag{1-4}$$

所以数学期望m_x是X的最小方差估计。

二、二维随机变量

设有两个随机变量X和Z。设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$p_1(x)$和$p_2(z)$，其联合概率分布密度为$p(x,z)$。想用随机变量Z的函数$g(Z)$作为随机变量X的估值，要求估计误差的方差为最小，即

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

下面可以证明，使上式为最小的函数$g(z)$是X的条件数学期望，即

$$\hat{X}=g(z)=E[X|z] \tag{1-5}$$

X和Z的联合概率分布密度可用下式表示

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

此为试读，需要完整PDF请访问：www.ertongbook.com

我们可得估计误差的方差

$$J=E\{[X-g(z)]^2\}=\int_{-\infty}^{\infty}p(z)\left\{\int_{-\infty}^{\infty}[x-g(z)]^2p(x|z)\,dx\right\}dz \qquad (1-6)$$

上面的积分中的被积函数$\int_{-\infty}^{\infty}[x-g(z)]^2p(x|z)dx$是非负的。因此为了使双重积分最小，只要对每一个$z$值，使积分

$$\int_{-\infty}^{\infty}[x-g(z)]^2p(x|z)dx \qquad (1-7)$$

为最小，就足以使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—7)是X相对于常值$g(z)$的二次矩，参照(1—2)式可知，使这个二次矩为最小的$g(z)$值是x的条件数学期望

$$\hat{X}=g(z)=\int_{-\infty}^{\infty}xp(x|z)dx=E[x|z] \qquad (1-8)$$

函数$g(z)=E[X|z]$就是通常所说的回归曲线，如图1—2所示。

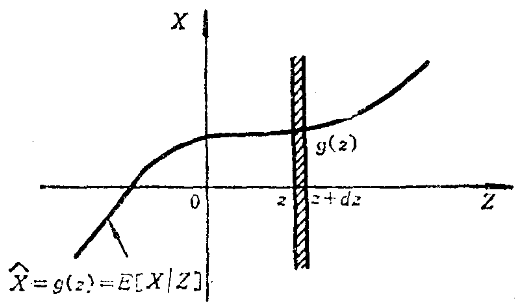

图1—2　条件数学期望曲綫图

如果把$p(x|z)$比拟为质量分布密度，则$g(z)$就是窄条$(z,z+dz)$的质量中心，$g(Z)$是质量中心的轨迹。如质量分布靠近$g(z)$，则估计误差的方差$E\{[X-E(x|z)]^2\}$就小。在这里得到一个很重要的结论，对于二个随机变量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

$$\mathbf{X}=\begin{pmatrix}X_1\\X_2\\\vdots\\X_n\end{pmatrix}$$

$\mathbf{X}$的可能值为

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

观测值为q维随机变量Z

$$\mathbf{Z}=\begin{pmatrix}Z_1\\Z_2\\\vdots\\Z_q\end{pmatrix}$$

$\mathbf{Z}$的可能值为

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

假定已知$\mathbf{X}$和$\mathbf{Z}$的联合概率分布密度$p(\mathbf{x},\mathbf{z})$

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

$\mathbf{X}$的概率分布密度为$p(\mathbf{x})$

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

$\mathbf{Z}$的概率分布密度为$p(\mathbf{z})$

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

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

$$\hat{X}=\hat{X}(Z)$$

则$\mathbf{X}$的估计误差为

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

估计误差$\tilde{X}$是一个与$\mathbf{X}$同维数的随机向量

$$\tilde{\mathbf{X}}=\begin{pmatrix} \tilde{X}_1 \\ \tilde{X}_2 \\ \vdots \\ \tilde{X}_n \end{pmatrix}$$

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：设有二个$n\times n$矩阵$\mathbf{A}$和$\mathbf{B}$，若($\mathbf{A}-\mathbf{B}$)正定（或非负定），则称$\mathbf{A}>\mathbf{B}$（或$\mathbf{A}\geqslant\mathbf{B}$）。

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

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

估计误差的方差阵为

$$\begin{aligned} J=E[\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T]&=E\{[\mathbf{X}-\hat{\mathbf{X}}(\mathbf{Z})][\mathbf{X}-\hat{\mathbf{X}}(\mathbf{Z})]^T\} \\ &=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[\mathbf{x}-\hat{\mathbf{X}}(\mathbf{z})][\mathbf{x}-\hat{\mathbf{X}}(\mathbf{z})]^T p(\mathbf{x},\mathbf{z})\,d\mathbf{x}d\mathbf{z} \\ &=\int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty}[\mathbf{x}-\hat{\mathbf{X}}(z)][\mathbf{x}-\hat{\mathbf{X}}(\mathbf{z})]^T p(\mathbf{x}|\mathbf{z})d\mathbf{x}\right\}p(\mathbf{z})d\mathbf{z} \end{aligned} \tag{1—10}$$

因为$\mathbf{X}$和$\mathbf{Z}$都是向量，所以这里的$\int_{-\infty}^{\infty}$表示多重积分，$d\mathbf{x}=dx_1dx_2\cdots dx_n$，$d\mathbf{z}=dz_1dz_2\cdots dz_q$。

我们要选择$\hat{\mathbf{X}}(\mathbf{Z})$，使得(1－10)式的$J$达到极小。$J$是一个非负定的对称矩阵，参照(1—6)式可知，为了使J最小，只要对每一个z值，使

$$\int_{-\infty}^{\infty}[\mathbf{x}-\hat{\mathbf{X}}(\mathbf{z})][\mathbf{x}-\hat{\mathbf{X}}(\mathbf{z})]^T p(\mathbf{x}|\mathbf{z})dx \tag{1—11}$$

为最小。

对于给定的Z值，随机变量X的条件概率密度为 $p(\mathbf{x}|\mathbf{z})$，积分式（1—11）是X相对于 $\hat{\mathbf{X}}(\mathbf{z})$的二次矩。参照（1—7）式可知，使得这个二次矩为最小的 $\hat{\mathbf{X}}(\mathbf{z})$ 值是X的条件数学期望

$$\hat{\mathbf{X}}(\mathbf{z})=\int_{-\infty}^{\infty}\mathbf{x}p(\mathbf{x}|\mathbf{z})dx=E[\mathbf{X}|\mathbf{z}] \tag{1—12}$$

一般可写成

$$\hat{\mathbf{X}}(\mathbf{Z})=E[\mathbf{X}|\mathbf{Z}] \tag{1—13}$$

估计误差的方差阵为

$$J=E[\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T]=E\{[\mathbf{X}-E(\mathbf{X}|\mathbf{Z})][\mathbf{X}-E(\mathbf{X}|\mathbf{Z})]^T\} \tag{1—14}$$

下面来证明，当X的估值 $\hat{\mathbf{X}}=E[\mathbf{X}|\mathbf{Z}]$，确实使$J$为最小。设X的估值为Z的任一向量函数$\overline{\mathbf{X}}(\mathbf{z})$，则从(1—11)式出发，可写出

$$\begin{aligned}
&\int_{-\infty}^{\infty}[\mathbf{x}-\bar{\mathbf{x}}(\mathbf{z})][\mathbf{x}-\overline{\mathbf{X}}(\mathbf{z})]^T p(\mathbf{z}|\mathbf{x})dx\\
&=\int_{-\infty}^{\infty}[\mathbf{x}-E(\mathbf{x}|\mathbf{z})+E(\mathbf{x}|\mathbf{z})-\overline{\mathbf{X}}(\mathbf{z})]\cdot\\
&\qquad[\mathbf{x}-E(\mathbf{X}|\mathbf{z})+E(\mathbf{X}|\mathbf{z})-\overline{\mathbf{X}}(\mathbf{z})]^T p(\mathbf{x}|\mathbf{z})dx\\
&=\int_{-\infty}^{\infty}[\mathbf{x}-E(\mathbf{X}|\mathbf{z})][\mathbf{x}-E(\mathbf{X}|\mathbf{z})]^T p(\mathbf{x}|\mathbf{z})dx\\
&\quad+[E(\mathbf{X}|\mathbf{z})-\overline{\mathbf{X}}(\mathbf{z})][E(\mathbf{x}|\mathbf{z})-\overline{\mathbf{X}}(\mathbf{z})]^T\int_{-\infty}^{\infty}p(\mathbf{x}|\mathbf{z})dx\\
&\quad+\int_{-\infty}^{\infty}[\mathbf{x}-E(\mathbf{x}|\mathbf{z})]p(\mathbf{x}|\mathbf{z})dx\cdot[E(\mathbf{X}|\mathbf{z})-\overline{\mathbf{X}}(\mathbf{z})]^T\\
&\quad+[E(\mathbf{x}|\mathbf{z})-\overline{\mathbf{X}}(\mathbf{z})]\int_{-\infty}^{\infty}[\mathbf{x}-E(\mathbf{X}|\mathbf{z})]^T p(\mathbf{z}|\mathbf{x})dx
\end{aligned} \tag{1—15}$$

由于

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

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

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

则(1—15)式变成

$$\begin{aligned}
&\int_{-\infty}^{\infty}[\mathbf{x}-\overline{\mathbf{X}}(\mathbf{z})][\mathbf{x}-\overline{\mathbf{X}}(\mathbf{z})]^T p(\mathbf{x}|\mathbf{z})dx\\
&=\int_{-\infty}^{\infty}[\mathbf{x}-E(\mathbf{X}|\mathbf{z})][\mathbf{x}-E(\mathbf{X}|\mathbf{z})]^T p(\mathbf{x}|\mathbf{z})dx
\end{aligned}$$

$$+[E(\mathbf{X}|\mathbf{z})-\mathbf{X}(\mathbf{z})][E(\mathbf{X}|\mathbf{z})-\overline{\mathbf{X}}(\mathbf{z})]^T \tag{1—16}$$

(1—16)式等号左边为非负定矩阵，等号右边的二个矩阵也分别为非负定矩阵，所以

$$\int_{-\infty}^{\infty}[\mathbf{x}-\overline{\mathbf{X}}(\mathbf{z})][\mathbf{x}-\overline{\mathbf{X}}(\mathbf{z})]^T p(\mathbf{x}|\mathbf{z})d\mathbf{x} \geqslant$$

$$\int_{-\infty}^{\infty}[\mathbf{x}-E(\mathbf{X}|\mathbf{z})][\mathbf{x}-E(\mathbf{X}|\mathbf{z})]^T p(\mathbf{x}|\mathbf{z})d\mathbf{x} \tag{1—17}$$

这就证明了$\mathbf{X}$的最佳估计为

$$\hat{\mathbf{X}}=E[\mathbf{X}|\mathbf{Z}]$$

下面讨论估计是否无偏的问题。由(1—12)式可得

$$E\{E(\mathbf{X}|\mathbf{Z})\}=\int_{-\infty}^{\infty}E(\mathbf{X}|\mathbf{z})p(\mathbf{z})d\mathbf{z}$$

$$=\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}\mathbf{x}p(\mathbf{x}|\mathbf{z})d\mathbf{x}\right]p(\mathbf{z})d\mathbf{z}$$

$$=\int_{-\infty}^{\infty}\mathbf{x}\left[\int_{-\infty}^{\infty}p(\mathbf{x}|\mathbf{z})p(\mathbf{z})d\mathbf{z}\right]d\mathbf{x}$$

$$=\int_{-\infty}^{\infty}\mathbf{x}p(\mathbf{x})d\mathbf{x}=E[\mathbf{X}] \tag{1—18}$$

就是说估计是无偏的。

估计误差的方差阵为

$$Var\tilde{\mathbf{X}}=E\{\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T\}=E\{[\mathbf{X}-E(\mathbf{X}|\mathbf{Z})][\mathbf{X}-E(\mathbf{X}|\mathbf{Z})]^T\}$$

$$=\int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty}[\mathbf{x}-E(\mathbf{X}|\mathbf{z})][\mathbf{x}-E(\mathbf{x}|\mathbf{z})]^T p(\mathbf{x}|\mathbf{z})d\mathbf{x}\right\}\cdot p(\mathbf{z})d\mathbf{z}$$

$$=\int_{-\infty}^{\infty}Var(\mathbf{X}|\mathbf{z})p(\mathbf{z})d\mathbf{z} \tag{1—19}$$

例1—1 设有两个随机变量X和Z，服从正态分布，联合概率分布密度为

$$p(x,z)=\frac{1}{2\pi\sigma_x\sigma_z\sqrt{1-r^2}}e^{-\frac{1}{2(1-r^2)}\left[\frac{(x-m_x)^2}{\sigma_x^2}-\frac{2r(x-m_x)(z-m_z)}{\sigma_x\sigma_z}+\frac{(z-m_z)^2}{\sigma_z^2}\right]}$$

式中

$m_x=E[X]$，$m_z=E[Z]$，σ_x^2和σ_z^2分别为X和Z的方差，r为X与Z的互相关系数，求X的估值。

解　X的条件概率分布密度为

$$p(x|z)=\frac{1}{\sigma_x\sqrt{2\pi(1-r^2)}}e^{-\frac{1}{2\sigma_x^2(1-r^2)}\left[x-m_x-\frac{r\sigma_x}{\sigma_z}(z-m_z)\right]^2}$$

经过比较复杂的推导后，可得在$Z=z$的条件下，X的条件数学期望为

$$E[X|z]=m_x+\frac{r\sigma_x}{\sigma_z}(z-m_z)$$

或

$$\hat{X}=E[X|Z]=m_x+\frac{r\sigma_x}{\sigma_z}(z-m_z) \tag{1—20}$$

从(1—20)式可看出X的估值(条件数学期望)是Z的线性函数。

§1—3 极大似然估計

极大似然估计是从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$\theta_1$，$\theta_2$，…，$\theta_n$。把$k$个独立观测值$z_1$，$z_2$，…，$z_k$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

$$L(z_1, z_2, \cdots, z_k; \theta_1, \theta_2, \cdots, \theta_n)=\prod_{i=1}^{k} p(z_i, \theta_1, \theta_2, \cdots, \theta_n) \qquad (1—21)$$

称函数L为似然函数。当z_1、z_2、…、z_k固定时，L是θ_1、θ_2、…、θ_n的函数。极大似然法的实质，就是求出使L达到极大时，θ_1、θ_2、…、θ_n的估值$\hat{\theta}_1$、$\hat{\theta}_2$、…、$\hat{\theta}_n$。从(1-21)式可以看到$\hat{\theta}_1$、$\hat{\theta}_2$、…、$\hat{\theta}_n$是观测值z_1、z_2、…、z_k的函数。

为了便于求出使L达到极大的$\hat{\theta}_1$，$\hat{\theta}_2$，…，$\hat{\theta}_n$，对(1—21)式取对数，则把连乘转变为连加，即

$$lnL=\sum_{i=1}^{k} lnp(z_i, \theta_1, \theta_2, \cdots \theta_n) \qquad (1—22)$$

由于对数函数是单调增加函数，当L取极大值时，lnL也同时取极大值。将上式分别对θ_1、θ_2、…、θ_n求偏导数，令偏导数等于零，可得下面的一组方程：

$$\left.\begin{array}{c}\dfrac{\partial}{\partial\theta_1}(lnL)=0 \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \dfrac{\partial}{\partial\theta_n}(lnL)=0\end{array}\right\} \qquad (1—23)$$

解上述方程组，可得使L达到极大值的$\hat{\theta}_1$、$\hat{\theta}_2$、…、$\hat{\theta}_n$。按极大似然法确定的$\hat{\theta}_1$、$\hat{\theta}_2$、…、$\hat{\theta}_n$，使z_1，z_2，…，z_k最有可能出现，而不需要θ_1、θ_2、…、θ_n的验前知识，即不需要知道θ_1、θ_2、…、θ_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。

例1—2 设有正态分布随机变量Z，给出k个观测值z_1，z_2，…，z_k。观测值互相独立。试根据这k个观测值，确定分布密度中的各参数。

解 z的分布密度可用下式表示

$$p(z, m, \sigma)=\frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(z-m)^2}{2\sigma^2}}$$

式中m和σ为未知参数。现用极大似然法来确定参数m和σ。作似然函数

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

对上式取对数，可得

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

$$=-\sum_{i=1}^{k}\frac{(z_i-m)^2}{2\sigma^2}+kln\frac{1}{\sqrt{2\pi}}-kln\sigma$$

将上式分别对m和σ求偏导数，令偏导数等于零，可得

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

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

解上述二式，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维未知参数，假定已知$\mathbf{Z}$的条件概率密度$p(\mathbf{z}|\mathbf{x})$。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。问参数$\mathbf{X}$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

$$L(\mathbf{z}, \mathbf{x})=p(\mathbf{z}_1|\mathbf{x})p(\mathbf{z}_2|\mathbf{x})\cdots p(\mathbf{z}_k|\mathbf{x})=p(\mathbf{z}|\mathbf{x}) \tag{1—24}$$

或

$$lnL(\mathbf{z}, \mathbf{x})=lnp(\mathbf{z}|\mathbf{x}) \tag{1—25}$$

求出使L为极大的X值，令

$$\frac{\partial L}{\partial \mathbf{X}}=0 \quad 或 \quad \frac{\partial lnL}{\partial \mathbf{X}}=0 \tag{1—26}$$

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

$$\frac{\partial^2 L}{\partial \mathbf{X}^2}<0 \quad 或 \quad \frac{\partial^2 lnL}{\partial \mathbf{X}^2}<0$$

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

$$\mathbf{Z}=\mathbf{h}(\mathbf{X}, \mathbf{V})=\mathbf{H}\mathbf{X}+\mathbf{V} \tag{1—27}$$

式中　$\mathbf{Z}$—m维观测值，$\mathbf{X}$—n维未知参数，$\mathbf{V}$—m维测量误差。设$\mathbf{V}$与$\mathbf{X}$独立。给出$\mathbf{V}$的统计特性，求$\mathbf{X}$的极大似然估计。由(1—27)式得

$$\mathbf{V}=\mathbf{h}^{-1}(\mathbf{X}, \mathbf{Z})=\mathbf{Z}-\mathbf{H}\mathbf{X}$$

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

下面求出似然函数

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

此为试读，需要完整PDF请访问：www.ertongbook.com