高等学校研究生教材

电子信息系列

最优状态估计与系统辨识

王志贤 编著

Electronic Information

西北工业大学出版社

高等学校研究生教材

最优状态估计与系统辨识

王志贤　编著

西北工业大学出版社

【内容简介】 本书系统地阐述了最优状态估计与系统辨识的基本概念、基本理论和基本方法。全书共分两篇14章:第一篇(1～6章)为最优状态估计,分别介绍了最优估计的基本概念、线性系统的卡尔曼滤波、最优线性平滑、卡尔曼滤波的稳定性、滤波的发散及其克服方法、非线性滤波。第二篇(7～14章)为系统辨识,分别介绍了系统辨识的一般概念、脉冲响应法和相关函数法、最小二乘类辨识方法、极大似然法和预报误差法、时间序列模型和随机逼近法、多输入多输出线性系统辨识、闭环系统辨识。附录给出了学习本课程中用到的矩阵分析等一些数学工具。

本书结构清楚,重点突出,既注重理论的系统性和完整性,又突出了工程实用性。本书既可作为高等工科院校自动控制及相近专业的研究生教材,也可供从事自动控制类专业的科研工作者、工程技术人员参考。

图书在版编目 (CIP) 数据

最优状态估计与系统辨识/王志贤主编．—西安：西北工业大学出版社，2004.6
ISBN 7-5612-1777-3

Ⅰ．最…　Ⅱ．王…　Ⅲ．系统科学　Ⅳ．N94

中国版本图书馆 CIP 数据核字（2003）第 047477 号

出版发行：西北工业大学出版社
通信地址：西安市友谊西路 127 号　**邮编**：710072　**电话**：(029)8493844
网　　址：www.nwpup.com
印 刷 者：陕西友盛印务有限责任公司
开　　本：787 mm×1 092 mm　1/16
印　　张：24.75
字　　数：604 千字
版　　次：2004 年 6 月第 1 版　　2004 年 6 月第 1 次印刷
印　　数：1～2 000 册
定　　价：32.00 元

前 言

“最优状态估计与系统辨识”是导航、制导与控制类硕士研究生一门重要的理论课程。

状态估计、系统辨识和控制理论是现代控制理论中的三个相互独立又相互渗透的主要分支。状态估计和系统辨识离不开控制理论的支持，而最优控制和自适应控制又几乎不能没有状态估计和系统辨识技术。

任何自动控制系统都是在随机干扰的作用下工作的，所有控制系统的测量装置测出的量都程度不同的含有随机误差。这些随机因素的存在，都给实际的控制系统的分析和设计带来一系列的新问题。随机系统理论就是在解决这些新问题的过程中逐渐形成的。

研究随机系统的目的，往往在于对系统施加控制，使之按照人们预期的目标运行。随机最优控制理论则是研究如何选择控制策略或控制律，使某个目标函数在统计意义下达到最优。

由于随机系统固有的不确定性，系统的状态和输出都表现为具有某种统计特性的随机过程。因此，在一般情况下，企图测量系统某个时刻的状态，或精确预报系统的状态和输出在未来时刻的变化，都是不可能的。所以，随机系统理论必须借助于数理统计中的估计理论，来研究对系统的状态或输出的估计。

对实际过程的估计或控制，都有赖于获得对该过程正确描述的数学模型。对于一些极其简单的实际过程，其数学模型可以根据物理定律来建立，但大多数实际过程的作用机理却不能为人们所确切了解，这就需要通过实验数据来构造其数学模型。系统辨识和参数估计则是随机系统理论中研究如何建立系统数学模型的又一个重要分支。

本书是编者在多年讲授“最优状态估计与系统辨识”课程的讲稿基础上，经多次修改完善写成的。全书分两篇 14 章，其中第一篇最优状态估计（第 1～6 章），主要介绍卡尔曼滤波的基本理论；第二篇系统辨识（第 7～14 章），主要介绍系统辨识和参数估计的基本理论和方法。

本书力求贯彻理论和实践相结合的原则。考虑到工科研究生的知识结构特点，在取材和阐述方式上，既注意理论的系统性和完整性，又突出了工程实用性；公式推导比较详细，但又不片面追求数学上的严格性。问题表述力求浅显易懂。书中还附有较多的例题和习题，为读者自学提供了方便条件。

本书的编写工作得到了院、系领导的支持和关怀，导航、制导与控制学科的邓方林教授、许化龙教授以及很多同事也曾给予热情的鼓励、支持和帮助。牟建华副教授（博士）和胡昌华教授（博士）分别详细审阅了书稿的第一篇和第二篇，并提出不少宝贵的修改意见。更不能忘记教保处的同志们为本书的出版所付出的辛勤劳动。在此，一并向他们表示衷心的感谢！

另外，书后所列参考文献，为本书的基本内容提供了素材，有的还引用了其中的部分内容，在此，谨向有关作者表示谢意！

由于作者水平和实践经验有限，书中缺点和不足之处在所难免，热诚欢迎读者批评指正。

王志贤

2003 年 1 月于第二炮兵工程学院

目 录

第一篇 最优状态估计

第一篇

最优状态估计

第 1 章　最优估计的基本概念

§1.1　估计、最优估计和最优估计方法

1.1.1　估计

在自动控制、通讯、航空与航天等学科领域中，常常会遇到“估计”问题。所谓估计，就是从带有随机干扰的观测数据中，提取有用信息。估计问题可叙述为：

如果假设被估计量 $\boldsymbol{X}(t)$ 是一个 n 维向量，而 m 维向量 $\boldsymbol{Z}(t)$ 是其观测量，并且观测量与被估计量之间具有如下关系

$$\boldsymbol{Z}(t) = \boldsymbol{h}[\boldsymbol{X}(t), \boldsymbol{V}(t), t]$$

其中，$\boldsymbol{h}$ 是已知的 m 维向量函数，它由观测方法决定；$\boldsymbol{V}(t)$ 是观测误差向量，它通常是一个随机过程。那么，所谓估计问题，就是在时间区间 $[t_0, t]$ 内对 $\boldsymbol{X}(t)$ 进行观测，从而在得到观测数据 $\boldsymbol{Z} = \{\boldsymbol{Z}(\tau), t_0 \leqslant \tau \leqslant t\}$ 的情况下，要求构造一个观测数据的函数 $\hat{\boldsymbol{X}}(\boldsymbol{Z})$ 去估计 $\boldsymbol{X}(t)$ 的问题，并称 $\hat{\boldsymbol{X}}(\boldsymbol{Z})$ 是 $\boldsymbol{X}(t)$ 的一个估计量，或称 $\boldsymbol{X}(t)$ 的估计为 $\hat{\boldsymbol{X}}(\boldsymbol{Z})$。

估计理论是概率论和数理统计的一个分支。它所研究的对象是随机现象，它是根据受干扰的观测数据来估计关于随机变量、随机过程或系统的某些特性的一种数学方法。

估计问题大致可分为两类：状态估计和参数估计。状态和参数的基本差别在于，前者是随时间变化的随机过程，后者是不随时间变化或只随时间缓慢变化的随机变量。因此，可以说，状态估计是动态估计，而参数估计一般是静态估计。动态估计和静态估计是有联系的，可以这样说，把静态估计方法与动态随机过程或序列的内部规律性结合起来，就可得到动态估计方法。

1.1.2　估计准则和最优估计

如上所述，所谓估计问题，就是要构造一个观测数据 $\boldsymbol{Z}$ 的函数 $\hat{\boldsymbol{X}}(\boldsymbol{Z})$ 来作为被估计量 $\boldsymbol{X}(t)$ 的一个估计量。我们总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，必须要有一个衡量的标准，这个衡量标准就是估计准则。估计常常是以“使估计的性能指标达到极值”作为准则的。估计准则可以是多种多样的。常用的估计准则有：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、最小二乘准则等。

一个估计问题能否得到可行的明确解答，固然与随机过程、随机变量或系统的状态特点有关，但它与估计准则的选择关系也极大。可以说，估计准则在很大程度上将决定估计的性能、求

解估计问题所使用的估计方法及估计量的性质(是线性的还是非线性的)等。因此,要使估计问题得到好的结果,选择合理的估计准则是极其重要的。估计准则的选择在很大程度上取决于对被估计量的了解,对估计精度的要求,以及实现方便等。

所谓最优估计,是指在某一确定的估计准则条件下,按照某种统计意义,使估计达到最优。因此,最优估计是针对某一估计准则而言的。某一估计对某一估计准则为最优估计,但换一个估计准则,这一估计值就不一定是最优的了。这就是说,最优估计不是惟一的。

1.1.3 估计方法

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。根据观测 $\boldsymbol{Z}$ 与被估计值 $\boldsymbol{X}$ 的统计特性的掌握程度,可有下列一些估计方法:

1. *最小方差估计*

最小方差估计是以估计误差的方差阵达到最小为估计准则的。按照这种准则求得的最优估值叫最小方差估计。为了进行最小方差估计,需要知道被估计值 $\boldsymbol{X}$ 和观测值 $\boldsymbol{Z}$ 的条件概率密度值 $\boldsymbol{P}(\boldsymbol{X} \mid \boldsymbol{Z})$ 或 $\boldsymbol{P}(\boldsymbol{Z})$ 以及它们的联合概率分布密度 $\boldsymbol{P}(\boldsymbol{X},\boldsymbol{Z})$。

2. *极大似然估计*

极大似然准则是使条件概率分布密度 $\boldsymbol{P}(\boldsymbol{Z} \mid \boldsymbol{X})$ 达到极大的那个 $\boldsymbol{X}$ 值作为估值的。按照这种估计准则求得的 $\boldsymbol{X}$ 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 $\boldsymbol{P}(\boldsymbol{Z} \mid \boldsymbol{X})$。

3. *极大验后估计*

极大验后准则是使验后概率分布密度 $\boldsymbol{P}(\boldsymbol{X} \mid \boldsymbol{Z})$ 达到极大的那个 $\boldsymbol{X}$ 值作为估值的。按这种估计准则求得的 $\boldsymbol{X}$ 的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 $\boldsymbol{P}(\boldsymbol{X} \mid \boldsymbol{Z})$。

4. *线性最小方差估计*

如上所述,为了进行最小方差估计和极大验后估计,需要知道条件概率分布密度 $\boldsymbol{P}(\boldsymbol{X} \mid \boldsymbol{Z})$;为了进行极大似然估计,需要知道 $\boldsymbol{P}(\boldsymbol{Z} \mid \boldsymbol{X})$。如果我们放松对概率分布密度知识的要求,只要求知道观测值和被估计值的一、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 $\mathrm{Cov}[\boldsymbol{Z},\boldsymbol{X}]$。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。若我们限定所求的估计量是观测值的线性函数,并以估计误差的方差阵达到最小作为最优估计准则。则按这种方式求得的最优估值称为线性最小方差估计。

5. *最小二乘估计*

当我们不知道 $\boldsymbol{X}$ 和 $\boldsymbol{Z}$ 的概率分布密度,也不知道它们的一、二阶矩时,就只能采用高斯提出的最小二乘法进行估计。最小二乘估计是以残差的平方和为最小作为估计准则的。

下文将分别介绍上述估计方法。

§1.2 最小方差估计

1.2.1 最小方差估计

设被估计量 $\boldsymbol{X}$ 是一个 n 维随机向量,观测值 $\boldsymbol{Z}$ 为 m 维向量,$\boldsymbol{X}$ 和 $\boldsymbol{Z}$ 没有明确的函数关系,

只有概率上的联系。$\boldsymbol{X}$ 和 $\boldsymbol{Z}$ 的概率分布密度分别为 $\boldsymbol{P}_1(\boldsymbol{X})$ 和 $\boldsymbol{P}_2(\boldsymbol{Z})$，其联合概率分布密度为 $\boldsymbol{P}(\boldsymbol{X},\boldsymbol{Z})$。选择估计误差 $\tilde{\boldsymbol{X}} = \boldsymbol{X} - \hat{\boldsymbol{X}}(\boldsymbol{Z})$ 的二次型函数为代价函数

$$f[\boldsymbol{X} - \hat{\boldsymbol{X}}(\boldsymbol{Z})] = [\boldsymbol{X} - \hat{\boldsymbol{X}}(\boldsymbol{Z})]^{\mathrm{T}}\boldsymbol{S}[\boldsymbol{X} - \hat{\boldsymbol{X}}(\boldsymbol{Z})] \tag{1-2-1}$$

式中　$\boldsymbol{S}$ 为 $n \times n$ 维对称非负定加权矩阵。

若有估计量 $\hat{\boldsymbol{X}}_{MV}(\boldsymbol{Z})$，使得贝叶斯风险最小，即

$$\beta[\hat{\boldsymbol{X}}(\boldsymbol{Z})]\Big|_{\hat{\boldsymbol{X}}(\boldsymbol{Z})=\hat{\boldsymbol{X}}_{MV}(\boldsymbol{Z})} = E\{[\boldsymbol{X} - \hat{\boldsymbol{X}}(\boldsymbol{Z})]^{\mathrm{T}}\boldsymbol{S}[\boldsymbol{X} - \hat{\boldsymbol{X}}(\boldsymbol{Z})]\}\Big|_{\hat{\boldsymbol{X}}(\boldsymbol{Z})=\hat{\boldsymbol{X}}_{MV}(\boldsymbol{Z})} = \min \tag{1-2-2}$$

则称 $\hat{\boldsymbol{X}}_{MV}(\boldsymbol{Z})$ 为最小方差估计。

下面讨论怎样求最小方差估计 $\hat{\boldsymbol{X}}_{MV}(\boldsymbol{Z})$。

按最小方差估计的定义，则有当 $\hat{\boldsymbol{X}}(\boldsymbol{Z}) = \hat{\boldsymbol{X}}_{MV}(\boldsymbol{Z})$ 时

$$\beta[\hat{\boldsymbol{X}}(\boldsymbol{Z})]\Big|_{\hat{\boldsymbol{X}}(\boldsymbol{Z})=\hat{\boldsymbol{X}}_{MV}(\boldsymbol{Z})} = \min$$

即

$$E\{[\boldsymbol{X} - \hat{\boldsymbol{X}}(\boldsymbol{Z})]^{\mathrm{T}}\boldsymbol{S}[\boldsymbol{X} - \hat{\boldsymbol{X}}(\boldsymbol{Z})]\}\Big|_{\hat{\boldsymbol{X}}(\boldsymbol{Z})=\hat{\boldsymbol{X}}_{MV}(\boldsymbol{Z})} = \min$$

或

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}[\boldsymbol{X} - \hat{\boldsymbol{X}}(\boldsymbol{Z})]^{\mathrm{T}}\boldsymbol{S}[\boldsymbol{X} - \hat{\boldsymbol{X}}(\boldsymbol{Z})]\boldsymbol{P}(\boldsymbol{X},\boldsymbol{Z})\mathrm{d}\boldsymbol{X}\ \mathrm{d}\boldsymbol{Z}\Big|_{\hat{\boldsymbol{X}}(\boldsymbol{Z})=\hat{\boldsymbol{X}}_{MV}(\boldsymbol{Z})} = \min$$

因为

$$\boldsymbol{P}(\boldsymbol{X},\boldsymbol{Z}) = \boldsymbol{P}[\boldsymbol{X} \mid \boldsymbol{Z})\boldsymbol{P}(\boldsymbol{Z}) \tag{1-2-3}$$

所以

$$\int_{-\infty}^{+\infty}\boldsymbol{P}(\boldsymbol{Z})\{\int_{-\infty}^{+\infty}[\boldsymbol{X} - \hat{\boldsymbol{X}}(\boldsymbol{Z})]^{\mathrm{T}}\boldsymbol{S}[\boldsymbol{X} - \hat{\boldsymbol{X}}(\boldsymbol{Z})]\boldsymbol{P}(\boldsymbol{X} \mid \boldsymbol{Z})\mathrm{d}\boldsymbol{X}\}\mathrm{d}\boldsymbol{Z}\Big|_{\hat{\boldsymbol{X}}(\boldsymbol{Z})=\hat{\boldsymbol{X}}_{MV}(\boldsymbol{Z})} = \min \tag{1-2-4}$$

由于 $\boldsymbol{S}$ 非负定，所以 $[\boldsymbol{X}-\hat{\boldsymbol{X}}(\boldsymbol{Z})]^{\mathrm{T}}\boldsymbol{S}[\boldsymbol{X}-\hat{\boldsymbol{X}}(\boldsymbol{Z})]$ 非负定，又因为 $\boldsymbol{P}(\boldsymbol{X} \mid \boldsymbol{Z})$ 和 $\boldsymbol{P}(\boldsymbol{Z})$ 非负，而 $\hat{\boldsymbol{X}}(\boldsymbol{Z})$ 只出现在内积分号内，所以只要使内积分号内积分为极小，即

$$\int_{-\infty}^{+\infty}[\boldsymbol{X} - \hat{\boldsymbol{X}}(\boldsymbol{Z})]^{\mathrm{T}}\boldsymbol{S}[\boldsymbol{X} - \hat{\boldsymbol{X}}(\boldsymbol{Z})]\boldsymbol{P}(\boldsymbol{X} \mid \boldsymbol{Z})\mathrm{d}\boldsymbol{X}\Big|_{\hat{\boldsymbol{X}}(\boldsymbol{Z})=\hat{\boldsymbol{X}}_{MV}(\boldsymbol{Z})} = \min \tag{1-2-5}$$

就可使贝叶斯风险为极小。这就是说，使贝叶斯风险为极小，即

$$\beta[\hat{\boldsymbol{X}}(\boldsymbol{Z})]\Big|_{\hat{\boldsymbol{X}}(\boldsymbol{Z})=\hat{\boldsymbol{X}}_{MV}(\boldsymbol{Z})} = \min$$

等价于在 $\boldsymbol{Z} = \boldsymbol{Z}$ 条件下，贝叶斯的条件风险为极小，即

$$\beta[\hat{\boldsymbol{X}}(\boldsymbol{Z}) \mid \boldsymbol{Z}]\Big|_{\hat{\boldsymbol{X}}(\boldsymbol{Z})=\hat{\boldsymbol{X}}_{MV}(\boldsymbol{Z})} = \int_{-\infty}^{+\infty}[\boldsymbol{X} - \hat{\boldsymbol{X}}(\boldsymbol{Z})]^{\mathrm{T}}\boldsymbol{S}[\boldsymbol{X} - \hat{\boldsymbol{X}}(\boldsymbol{Z})]\boldsymbol{P}(\boldsymbol{X} \mid \boldsymbol{Z})\mathrm{d}\boldsymbol{X}\Big|_{\hat{\boldsymbol{X}}(\boldsymbol{Z})=\hat{\boldsymbol{X}}_{MV}(\boldsymbol{Z})} = \min$$

这一等价的价值在于求贝叶斯风险最小时的 $(n+m)$ 重积分，简化成贝叶斯条件风险最小时的 n 重积分，从而简化了积分运算。

我们知道，$\hat{\boldsymbol{X}}(\boldsymbol{Z}) = \hat{\boldsymbol{X}}_{MV}(\boldsymbol{Z})$ 能使 $\beta[\hat{\boldsymbol{X}}(\boldsymbol{Z}) \mid \boldsymbol{Z}] = \min$ 的必要条件是它的梯度为零，即

$$\frac{\partial\beta[\hat{\boldsymbol{X}}(\boldsymbol{Z}) \mid \boldsymbol{Z}]}{\partial\hat{\boldsymbol{X}}(\boldsymbol{Z})}\Bigg|_{\hat{\boldsymbol{X}}(\boldsymbol{Z})=\hat{\boldsymbol{X}}_{MV}(\boldsymbol{Z})} = \boldsymbol{0} \tag{1-2-6}$$

亦即

$$\frac{\partial}{\partial\hat{\boldsymbol{X}}(\boldsymbol{Z})}\int_{-\infty}^{+\infty}[\boldsymbol{X} - \hat{\boldsymbol{X}}(\boldsymbol{Z})]^{\mathrm{T}}\boldsymbol{S}[\boldsymbol{X} - \hat{\boldsymbol{X}}(\boldsymbol{Z})]\boldsymbol{P}(\boldsymbol{X} \mid \boldsymbol{Z})\mathrm{d}\boldsymbol{X}\Bigg|_{\hat{\boldsymbol{X}}(\boldsymbol{Z})=\hat{\boldsymbol{X}}_{MV}(\boldsymbol{Z})} =$$

$$\int_{-\infty}^{+\infty}\frac{\partial}{\partial\hat{\boldsymbol{X}}(\boldsymbol{Z})}[\boldsymbol{X} - \hat{\boldsymbol{X}}(\boldsymbol{Z})]^{\mathrm{T}}\boldsymbol{S}[\boldsymbol{X} - \hat{\boldsymbol{X}}(\boldsymbol{Z})]\boldsymbol{P}(\boldsymbol{X} \mid \boldsymbol{Z})\mathrm{d}\boldsymbol{X}\Bigg|_{\hat{\boldsymbol{X}}(\boldsymbol{Z})=\hat{\boldsymbol{X}}_{MV}(\boldsymbol{Z})} =$$

$$-2S\int_{-\infty}^{+\infty}[X-\hat{X}(Z)]P(X \mid Z)\mathrm{d}X\Big|_{\hat{X}(Z)=\hat{X}_{MV}(Z)}=0$$

因为 $S>0$，于是有

$$\hat{X}_{MV}(Z)\int_{-\infty}^{+\infty}P(X \mid Z)\mathrm{d}X=\int_{-\infty}^{+\infty}XP(X \mid Z)\mathrm{d}X$$

注意到

$$\int_{-\infty}^{+\infty}P(X \mid Z)\mathrm{d}X=1$$

所以有

$$\hat{X}_{MV}(Z)=\int_{-\infty}^{+\infty}XP(X \mid Z)\mathrm{d}X=E\{X \mid Z\} \tag{1-2-7}$$

又由于

$$\frac{\partial^2\beta[\hat{X}(Z) \mid Z]}{\partial\hat{X}(Z)\partial\hat{X}^{\mathrm{T}}(Z)}\Big|_{\hat{X}(Z)=\hat{X}_{MV}(Z)}=2S>0 \tag{1-2-8}$$

所以，当 $\hat{X}_{MV}(Z)=\int_{-\infty}^{+\infty}XP(X \mid Z)\mathrm{d}X=E\{X \mid Z\}$ 时，确实使 $\beta[\hat{X}(Z) \mid Z]$ 具有最小值。

由此可见，随机向量 X 的最小方差估计 $\hat{X}_{MV}(Z)$ 是在观测向量为 Z 的条件下 X 的条件数学期望 $E\{X \mid Z\}$。因此，有时又称最小方差估计为条件期望估计。

1.2.1 最小方差估计的几点说明

为了加深对最小方差估计的理解，再作以下几点说明。

(1) 最小方差估计量 $\hat{X}_{mv}(Z)$ 是无偏估计。这是因为

$$\begin{aligned}E\{\hat{X}_{mv}(Z)\}=&E\{E\{X \mid Z\}\}=\\&\int_{-\infty}^{+\infty}\left\{\int_{-\infty}^{+\infty}XP(X \mid Z)\mathrm{d}X\right\}P(Z)\mathrm{d}Z=\\&\int_{-\infty}^{+\infty}X\int_{-\infty}^{+\infty}P(X,Z)\mathrm{d}Z\,\mathrm{d}X=\\&\int_{-\infty}^{+\infty}XP(X)\mathrm{d}X=E\{X\}\end{aligned} \tag{1-2-9}$$

(2) 最小方差估计 $\hat{X}_{MV}(Z)=E\{X \mid Z\}$ 这个结果，只要求加权阵 S 是非负定的，而与其具体形式无关，因此，它可以选为任意非负定阵，一般可选 S 为单位阵。

(3) 由于 $\hat{X}_{MV}(Z)$ 是 X 的无偏估计，因此，估计的均方误差阵 $E\{\tilde{X}_{MV}\tilde{X}_{MV}^{\mathrm{T}}\}=E\{[X-\hat{X}_{MV}(Z)][X-\hat{X}_{MV}(Z)]^{\mathrm{T}}\}$ 就是估计误差的方差阵 $\mathrm{Var}\{\tilde{X}_{MV}\}$，其表达式为

$$\begin{aligned}\mathrm{Var}\{\tilde{X}_{MV}\}=&\mathrm{Var}\{X-\hat{X}_{MV}(Z)\}=E\{[X-\hat{X}_{MV}(Z)][X-\hat{X}_{MV}(Z)]^{\mathrm{T}}\}=\\&\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}[X-\hat{X}_{MV}(Z)][X-\hat{X}_{MV}(Z)]^{\mathrm{T}}P(X,Z)\mathrm{d}X\,\mathrm{d}Z=\\&\int_{-\infty}^{+\infty}\left\{\int_{-\infty}^{+\infty}[X-E\{X \mid Z\}][X-E\{X \mid Z\}]^{\mathrm{T}}P(X \mid Z)\mathrm{d}X\right\}P(Z)\mathrm{d}Z=\\&\int_{-\infty}^{+\infty}\mathrm{Var}\{X \mid Z\}P(Z)\mathrm{d}Z\end{aligned} \tag{1-2-10}$$

(4) 如果设 X 的任一其他估计为 $\hat{X}(Z)$，则其相应的均方误差阵为

$$
\begin{aligned}
E\{\hat{\boldsymbol{X}}\hat{\boldsymbol{X}}^{\mathrm{T}}\} = & E\{[\boldsymbol{X}-\hat{\boldsymbol{X}}(\boldsymbol{Z})][\boldsymbol{X}-\hat{\boldsymbol{X}}(\boldsymbol{Z})]^{\mathrm{T}}\} = \\
& \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}[\boldsymbol{X}-\hat{\boldsymbol{X}}(\boldsymbol{Z})][\boldsymbol{X}-\hat{\boldsymbol{X}}(\boldsymbol{Z})]^{\mathrm{T}}\boldsymbol{P}(\boldsymbol{X},\boldsymbol{Z})\mathrm{d}\boldsymbol{X}\mathrm{d}\boldsymbol{Z} = \\
& \int_{-\infty}^{+\infty}\left\{\int_{-\infty}^{+\infty}[\boldsymbol{X}-\hat{\boldsymbol{X}}(\boldsymbol{Z})][\boldsymbol{X}-\hat{\boldsymbol{X}}(\boldsymbol{Z})]^{\mathrm{T}}\boldsymbol{P}(\boldsymbol{X}\mid\boldsymbol{Z})\mathrm{d}\boldsymbol{X}\right\}\boldsymbol{P}(\boldsymbol{Z})\mathrm{d}\boldsymbol{Z} = \\
& \int_{-\infty}^{+\infty}\left\{\int_{-\infty}^{+\infty}[\boldsymbol{X}-E\{\boldsymbol{X}\mid\boldsymbol{Z}\}+E\{\boldsymbol{X}\mid\boldsymbol{Z}\}-\hat{\boldsymbol{X}}(\boldsymbol{Z})][\boldsymbol{X}-E\{\boldsymbol{X}\mid\boldsymbol{Z}\}+\right. \\
& \qquad \left. E\{\boldsymbol{X}\mid\boldsymbol{Z}\}-\hat{\boldsymbol{X}}(\boldsymbol{Z})]^{\mathrm{T}}\boldsymbol{P}(\boldsymbol{X}\mid\boldsymbol{Z})\mathrm{d}\boldsymbol{X}\right\}\cdot\boldsymbol{P}(\boldsymbol{Z})\mathrm{d}\boldsymbol{Z} = \\
& \int_{-\infty}^{+\infty}\left\{\int_{-\infty}^{+\infty}[\boldsymbol{X}-E\{\boldsymbol{X}\mid\boldsymbol{Z}\}][\boldsymbol{X}-E\{\boldsymbol{X}\mid\boldsymbol{Z}\}]^{\mathrm{T}}\boldsymbol{P}(\boldsymbol{X}\mid\boldsymbol{Z})\mathrm{d}\boldsymbol{X}\right\}\boldsymbol{P}(\boldsymbol{Z})\mathrm{d}\boldsymbol{Z}+ \\
& \int_{-\infty}^{+\infty}\left\{\int_{-\infty}^{+\infty}[\boldsymbol{X}-E\{\boldsymbol{X}\mid\boldsymbol{Z}\}][E\{\boldsymbol{X}\mid\boldsymbol{Z}\}-\hat{\boldsymbol{X}}(\boldsymbol{Z})]^{\mathrm{T}}\boldsymbol{P}(\boldsymbol{X}\mid\boldsymbol{Z})\mathrm{d}\boldsymbol{X}\right\}\boldsymbol{P}(\boldsymbol{Z})\mathrm{d}\boldsymbol{Z}+ \\
& \int_{-\infty}^{+\infty}\left\{\int_{-\infty}^{+\infty}[E\{\boldsymbol{X}\mid\boldsymbol{Z}\}-\hat{\boldsymbol{X}}(\boldsymbol{Z})][\boldsymbol{X}-E\{\boldsymbol{X}\mid\boldsymbol{Z}\}]^{\mathrm{T}}\boldsymbol{P}(\boldsymbol{X}\mid\boldsymbol{Z})\mathrm{d}\boldsymbol{X}\right\}\boldsymbol{P}(\boldsymbol{Z})\mathrm{d}\boldsymbol{Z}+ \\
& \int_{-\infty}^{+\infty}\left\{\int_{-\infty}^{+\infty}[E\{\boldsymbol{X}\mid\boldsymbol{Z}\}-\hat{\boldsymbol{X}}(\boldsymbol{Z})][E\{\boldsymbol{X}\mid\boldsymbol{Z}\}-\hat{\boldsymbol{X}}(\boldsymbol{Z})]^{\mathrm{T}}\boldsymbol{P}(\boldsymbol{X}\mid\boldsymbol{Z})\mathrm{d}\boldsymbol{X}\right\}\boldsymbol{P}(\boldsymbol{Z})\mathrm{d}\boldsymbol{Z}
\end{aligned}
$$

因为上式中第一项中{……}部分

$$
\int_{-\infty}^{+\infty}[\boldsymbol{X}-E\{\boldsymbol{X}\mid\boldsymbol{Z}\}][\boldsymbol{X}-E\{\boldsymbol{X}\mid\boldsymbol{Z}\}]^{\mathrm{T}}\boldsymbol{P}(\boldsymbol{X}\mid\boldsymbol{Z})\mathrm{d}\boldsymbol{X} = \mathrm{Var}\{\boldsymbol{X}\mid\boldsymbol{Z}\}
$$

第二项中{……}部分

$$
\begin{aligned}
& \int_{-\infty}^{+\infty}[\boldsymbol{X}-E\{\boldsymbol{X}\mid\boldsymbol{Z}\}][\boldsymbol{X}-E\{\boldsymbol{X}\mid\boldsymbol{Z}\}]^{\mathrm{T}}\boldsymbol{P}(\boldsymbol{X}\mid\boldsymbol{Z})\mathrm{d}\boldsymbol{X} = \\
& \left[\int_{-\infty}^{+\infty}\boldsymbol{X}\boldsymbol{P}(\boldsymbol{X}\mid\boldsymbol{Z})\mathrm{d}\boldsymbol{X}-E\{\boldsymbol{X}\mid\boldsymbol{Z}\}\int_{-\infty}^{+\infty}\boldsymbol{P}(\boldsymbol{X}\mid\boldsymbol{Z})\mathrm{d}\boldsymbol{X}][E\{\boldsymbol{X}\mid\boldsymbol{Z}\}\hat{\boldsymbol{X}}(\boldsymbol{Z})\right]= \\
& [E\{\boldsymbol{X}\mid\boldsymbol{Z}\}-E\{\boldsymbol{X}\mid\boldsymbol{Z}\}][E\{\boldsymbol{X}\mid\boldsymbol{Z}\}-\hat{\boldsymbol{X}}(\boldsymbol{Z})] = 0
\end{aligned}
$$

第三项中{……}部分，同样也等于零，即

$$
\int_{-\infty}^{+\infty}[E\{\boldsymbol{X}\mid\boldsymbol{Z}\}-\hat{\boldsymbol{X}}(\boldsymbol{Z})][\boldsymbol{X}-E\{\boldsymbol{X}\mid\boldsymbol{Z}\}]^{\mathrm{T}}\boldsymbol{P}(\boldsymbol{X}\mid\boldsymbol{Z})\mathrm{d}\boldsymbol{X} = 0
$$

第四项中{……}部分

$$
\begin{aligned}
& \int_{-\infty}^{+\infty}[E\{\boldsymbol{X}\mid\boldsymbol{Z}\}-\hat{\boldsymbol{X}}(\boldsymbol{Z})][E\{\boldsymbol{X}\mid\boldsymbol{Z}\}-\hat{\boldsymbol{X}}(\boldsymbol{Z})]^{\mathrm{T}}\boldsymbol{P}(\boldsymbol{X}\mid\boldsymbol{Z})\mathrm{d}\boldsymbol{X} = \\
& \quad [E\{\boldsymbol{X}\mid\boldsymbol{Z}\}-\hat{\boldsymbol{X}}(\boldsymbol{Z})][E\{\boldsymbol{X}\mid\boldsymbol{Z}\}-\hat{\boldsymbol{X}}(\boldsymbol{Z})]\int_{-\infty}^{+\infty}\boldsymbol{P}(\boldsymbol{X}\mid\boldsymbol{Z})\mathrm{d}\boldsymbol{X} = \\
& \quad [E\{\boldsymbol{X}\mid\boldsymbol{Z}\}-\hat{\boldsymbol{X}}(\boldsymbol{Z})][E\{\boldsymbol{X}\mid\boldsymbol{Z}\}-\hat{\boldsymbol{X}}(\boldsymbol{Z})]^{\mathrm{T}}
\end{aligned}
$$

所以，有

$$
\begin{aligned}
E\{\hat{\boldsymbol{X}}\hat{\boldsymbol{X}}^{\mathrm{T}}\} = & E\{[\boldsymbol{X}-\hat{\boldsymbol{X}}(\boldsymbol{Z})][\boldsymbol{X}-\hat{\boldsymbol{X}}(\boldsymbol{Z})]^{\mathrm{T}}\} = \\
& \int_{-\infty}^{+\infty}\mathrm{Var}\{\boldsymbol{X}\mid\boldsymbol{Z}\}\boldsymbol{P}(\boldsymbol{Z})\mathrm{d}\boldsymbol{Z}+ \\
& \int_{-\infty}^{+\infty}[E\{\boldsymbol{X}\mid\boldsymbol{Z}\}-\hat{\boldsymbol{X}}][E\{\boldsymbol{X}\mid\boldsymbol{Z}\}-\hat{\boldsymbol{X}}(\boldsymbol{Z})]^{\mathrm{T}}\boldsymbol{P}(\boldsymbol{Z})\mathrm{d}\boldsymbol{Z}
\end{aligned}
$$

由于

$$
\mathrm{Var}\{\boldsymbol{X}\mid\boldsymbol{Z}\}\geqslant 0
$$
$$
[E\{\boldsymbol{X}\mid\boldsymbol{Z}\}-\hat{\boldsymbol{X}}(\boldsymbol{Z})][E\{\boldsymbol{X}\mid\boldsymbol{Z}\}-\hat{\boldsymbol{X}}(\boldsymbol{Z})]^{\mathrm{T}}\geqslant 0
$$

$P(Z)$ 总是非负的,所以由上式可得

$$E\{\hat{X}\hat{X}^{\mathrm{T}}\} = E\{[X - \hat{X}(Z)][X - \hat{X}(Z)]^{\mathrm{T}}\} \geqslant \int_{-\infty}^{+\infty} \mathrm{Var}\{X \mid Z\} P(Z) \mathrm{d}Z = \mathrm{Var}\{\hat{X}_{MV}\} \tag{1-2-11}$$

并且当 $\hat{X}(Z) = E\{X \mid Z\} = \hat{X}_{MV}(Z)$ 时,上式取等号。

不等式(1-2-12)告诉我们:任何其他估计的均方误差阵或任何其他无偏估计的方差阵都将大于最小方差估计的误差方差阵。亦即,最优估计 $\hat{X}_{MV}(Z) = E\{X \mid Z\}$ 具有最小的估计误差方差阵。这就是为什么称它为最小方差估计的原因。

由于无偏估计的误差方差阵,亦即估计误差的二阶矩表示了误差分布在零附近的密集程度,因此,最小方差估计 $\hat{X}_{MV}(Z) = E\{X \mid Z\}$ 是一种最接近真值 X 的估计。

§1.3 极大似然估计

极大似然估计是以观测值出现的概率最大作为准则的,这是一种很普通的参数估计方法。1906 年费希尔(R. A. Fisher)首先使用这种方法,它是以似然函数概念为基础的。

设 X 为 n 维被估计量(它可以是未知的确定性量,也可以是随机变量),Z 为 X 的测量值,它是 m 维随机向量。为了估计 X,假设对它进行了 k 次观测,得到了 $\{Z(i); i = 1,2,\cdots,k\}$,如果对观测的总体 $Z = \{Z(i); i = 1,2,\cdots,k\}$,我们考虑其概率密度函数 $P(z)$(其中 z 是 Z 的具体取值),由于观测数据 Z 是在被估计量 $X = x$ 的条件下取得的,因此,概率密度函数 $P(z)$ 应该是一种条件概率密度函数,即 $P(z) = P(z \mid x)$,一般情况下,$P(z \mid x)$ 应该是 z 和 x 两者的函数,但是对于具体的观测值 z 来说,$P(z \mid x)$ 就可认为只是 x 的函数,并称它为似然函数,记为 $L = P(z \mid x)$。为什么取这个名字呢?因为 $P(z \mid x)$ 表示了 $X = x$ 的条件下,Z 的概率分布密度,因此,如果 $X = x_1$ 时的 $P(z \mid x_1)$ 要比 $X = x_2$ 时的 $P(z \mid x_2)$ 大,则这时 x_1 是准确值的可能性就要比 x_2 是准确值的可能性大。因此,如果对所有可能的 x 值,$P(z \mid \hat{X})$ 是 $P(z \mid x)$ 的最大值,那么,$\hat{X}$ 是准确值的可能性就最大,这时我们就称 $\hat{X}$ 是 X 的极大似然估计,并记为 $\hat{X}_{mL}(Z)$。由此可见,极大似然估计 $\hat{X}_{mL}(Z)$ 是使似然函数 $L = P(z \mid x)$ 达到极大值的一种最优估计。显然这里的最优估计准则是"使似然函数达到极大"。

由极大似然估计定义可知,如果已经得到观测向量 Z,估计值 $\hat{X}_{ML}(Z)$ 在一切 X 值中,有

$$L = P(Z \mid X)\mid_{X=\hat{X}_{ML}(Z)} = \max \tag{1-3-1}$$

为了便于求出极大似然估计,常对似然函数 $L = P(Z \mid X)$ 取自然对数,即

$$\ln L = \ln P(Z \mid X)$$

并称之为对数似然函数。由于对数函数是单调增加函数,因此,$\ln L = \ln P(Z \mid X)$ 与 $L = P(Z \mid X)$ 在相同的 X 值达到极大,即也有

$$\ln L = \ln P(Z \mid X)\mid_{X=\hat{X}_{ML}(Z)} = \max \tag{1-3-2}$$

我们知道,$X = \hat{X}_{ML}(Z)$ 能使

$$L = P(Z \mid X) = \max$$

或

$$\ln L = \ln P(Z \mid X) = \max$$

的必要条件为