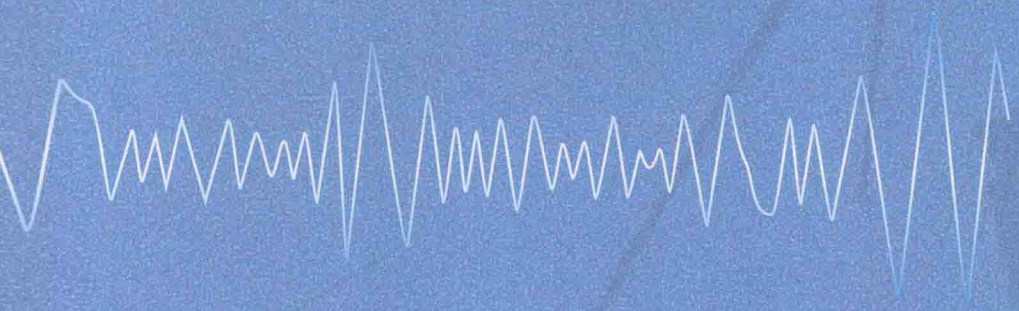
云南大学“211工程”项目资助

GAOJI XINHAO CHULI
YUANLI JI YINGYONG



高级信号处理
原理及应用

柏正尧 主 编
丁洪伟 余 映 副主编

科学出版社

云南大学“211 工程”项目资助

高级信号处理原理及应用

柏正尧　主　编
丁洪伟　余　映　副主编

科学出版社
北　京

内 容 简 介

本书是信号处理的高级教程，详细介绍了信号处理的若干高级专题的基本理论，并提供了丰富的应用实例。全书共分10章，包括贝叶斯推理、隐马尔可夫模型、维纳滤波器、自适应滤波器、功率谱分析、主成分分析、独立成分分析、小波变换、Hilbert-Huang变换、盲解卷积和信道均衡。

本书适合于信号与信息处理、通信与信息系统、电路与系统、控制科学与工程、模式识别与智能系统、生物医学工程等专业的硕士研究生和博士研究生阅读，也可供相关领域的教师和广大科技工作者参考。

图书在版编目(CIP)数据

高级信号处理原理及应用 / 柏正尧主编. —北京：科学出版社，2013.10

ISBN 978-7-03-038693-9

Ⅰ.①高… Ⅱ.①柏… Ⅲ.①信号处理-研究 Ⅳ. ①TN911.7

中国版本图书馆 CIP 数据核字（2013）第 227997 号

责任编辑：杨 岭 黄 桥 / 责任校对：葛茂香

责任印制：邝志强 / 封面设计：墨创文化

科学出版社出版

北京东黄城根北街16号

邮政编码：100717

http://www.sciencep.com

成都创新包装印刷厂印刷

科学出版社发行 各地新华书店经销

*

2013 年 10 月第 一 版 开本：787*1092 1/16

2013 年 10 月第一次印刷 印张：17.25

字数：410 千字

定价：39.00 元

(如有印装质量问题，我社负责调换)

前　言

随着计算机技术和微电子技术的发展，信号处理技术已经发展为一门主流的技术。信号处理的应用范围非常广泛，包括多媒体技术、音频信号处理、视频信号处理、蜂窝移动通信、雷达系统、模式分析与识别、医学信号处理、人工智能和控制系统等。信号处理的理论与应用涉及信号过程模式与结构的辨识、建模和运用。

为了适应信号与信息处理等专业硕士研究生和博士研究生进一步学习信号处理理论、方法和算法的要求，我们编写了这本信号处理的高级教程。学习本书，一般要求硕士研究生和博士研究生在本科阶段已学习过信号与系统、数字信号处理等课程，具有一定的信号处理基础理论。

本书主要介绍高级信号处理的理论及应用，全书共分为 10 章。其中，第 1～4 章及第 9 章由柏正尧教授编写，第 5、8、10 章由丁洪伟副教授编写，第 6、7 章由余映博士编写。全书由柏正尧教授统稿。各章内容简介如下：

第 1 章介绍贝叶斯理论、经典估计方法、估计最大化方法、最小估计方差的 Cramér-Rao 界、贝叶斯分类、随机信号空间建模。

第 2 章介绍用于非平稳信号建模的隐马尔可夫模型理论、模型训练方法以及模型在解码和分类中的应用。

第 3 章介绍维纳滤波器理论、维纳滤波器的向量空间解释及维纳滤波器的频率响应。

第 4 章介绍自适应滤波器理论，包括卡尔曼滤波器的状态空间方程、非线性卡尔曼滤波器——扩展卡尔曼滤波器和无迹卡尔曼滤波器、递归最小二乘误差滤波器和最速下降法以及最小均方误差滤波器。

第 5 章介绍功率谱估计的方法，包括经典谱估计方法、参数谱估计方法以及非参数谱估计方法。

第 6 章介绍主成分分析方法，包括主成分分析的特征结构及几种常见的主成分分析算法。

第 7 章介绍独立成分分析方法，包括独立成分分析的定义、统计独立性、独立成分分析的估计原理及常见算法。

第 8 章介绍小波变换的理论，包括连续小波变换的定义和性质、离散小波变换、小波框架理论及性质、多分辨分析和 Mallat 算法。

第 9 章介绍 Hilbert-Huang 变换方法，包括瞬时频率的概念、经验模态分解、固有模态函数和 Hilbert 谱。

第 10 章介绍盲解卷积和信道均衡，包括盲解卷积的数学模型、准则及算法，信道均衡的原理。

本书的出版得到了云南大学“211 工程”项目资助，在此表示感谢。

云南大学信息学院原副院长赵东风教授生前对本书的编写给予了关心、支持和鼓励，谨以本书的出版来告慰赵副院长。

在本书编写过程中，研究生万娟、赵变芳、李娟、闫帅辉、尹立国、王峥、郑哲、南京、刘正纲等参与了资料收集、公式编辑、绘图等工作，对他们的工作表示感谢。

在本书编写过程中，作者参考了国内外专家和学者的论文、专著等文献资料，在此一并表示衷心感谢。

感谢科学出版社黄桥编辑及其他编辑同志的辛勤劳动，是他们认真细致的工作，才使得本书能付梓出版。

信号处理的理论、方法、算法仍在不断发展中，应用领域也在不断拓展，尽管编者努力尝试介绍信号处理的高级理论和方法，但书中难免存在不全面、不妥甚至错漏之处，请广大读者批评指正。

柏正尧

2013 年 5 月于云南大学东陆园

目　录

第 1 章　贝叶斯推理

在信号处理中，推理就是从一组观测值中寻求观测信号或参数的过程。由于观测值通常是有噪声或不完全的，所以，从观测值得出的结论，其确定性和准确性依赖于观测的质量和推理方法的有效性。贝叶斯推理是根据一组观测值对信号或参数的未知值进行预测、估计和分类的概率推理，它将观测信号中的证据及其似然函数、随机过程概率分布的先验知识和误差代价函数结合在一起，对一个信号随机过程进行预测或估计。

贝叶斯推理以贝叶斯风险函数最小化为基础，包括在给定的观测条件下未知参数的概率模型和误差代价函数。贝叶斯方法包括最大后验(MAP)估计、最大似然(ML)估计、最小均方误差(MMSE)估计，以及作为其特殊情况的最小平均绝对误差(MAVE)，如图 1.1 所示。

本章首先介绍贝叶斯估计理论的基本概念，然后介绍几种经典估计方法，并用统计学方法对其性能进行评价，最后讨论离散或有限状态信号的贝叶斯分类和 K 均值聚类方法。

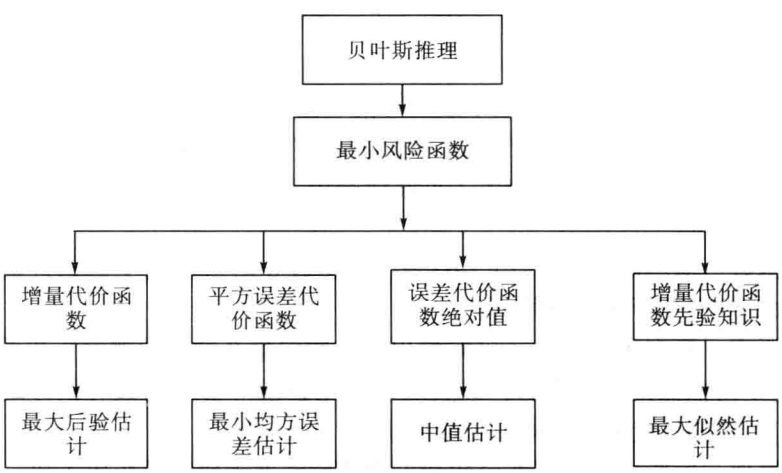


图 1.1　贝叶斯推理

1.1　贝叶斯估计理论

估计理论研究从一个观测信号确定未知参数向量的最优估计，或恢复一个受噪声污染或失真退化的信号。贝叶斯估计理论为统计估计方法的推导提供了一般的理论框架。统计估计法的各种形式可以看做贝叶斯估计的特殊情况。

1.1.1　贝叶斯定理

贝叶斯定理(或贝叶斯准则)：对于给定的观测向量 $\boldsymbol{y}$ 的随机参数向量 $\boldsymbol{\theta}$ 值的估计，已知 $\boldsymbol{y}$ 值时，参数向量 $\boldsymbol{\theta}$ 的后验概率密度函数可表示为

$$f_{\Theta|Y}(\boldsymbol{\theta}\mid\boldsymbol{y})=\frac{1}{f_Y(\boldsymbol{y})}f_{Y|\Theta}(\boldsymbol{y}\mid\boldsymbol{\theta})f_\Theta(\boldsymbol{\theta}) \tag{1.1}$$

式中，$f_{Y|\Theta}(\boldsymbol{y}\mid\boldsymbol{\theta})$为似然函数；$f_\Theta(\boldsymbol{\theta})$为参数向量的先验概率密度函数。

对于给定的观测值，$f_Y(\boldsymbol{y})=\int f_{Y|\Theta}(\boldsymbol{y}\mid\boldsymbol{\theta})f_\Theta(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}$ 是一个常数。贝叶斯定理说明，后验概率与似然函数和先验概率的乘积成正比。

似然函数和先验概率密度函数对后验概率密度函数的影响程度取决于它们的函数形状。一般来说，概率密度函数的形状越尖，它对估计结果的影响就越大。而均匀分布的先验概率密度函数除那些超出概率密度函数范围的值外，对后验概率密度函数没有什么影响。

1.1.2　贝叶斯推理的定义

贝叶斯推理可以由风险函数最小化和未知参数向量值的误差代价平均定义，即

$$\mathrm{Risk}(\hat{\boldsymbol{\theta}})=\iint_{\boldsymbol{\theta}\boldsymbol{y}}\mathrm{Cost}(\hat{\boldsymbol{\theta}},\boldsymbol{\theta})f_{Y|\Theta}(\boldsymbol{y}\mid\boldsymbol{\theta})f_\Theta(\boldsymbol{\theta})\mathrm{d}\boldsymbol{y}\mathrm{d}\boldsymbol{\theta} \tag{1.2}$$

贝叶斯推理包括以下几个要素：

(1)贝叶斯风险。估计未知参数向量 $\boldsymbol{\theta}$ 的估计值 $\hat{\boldsymbol{\theta}}$ 的风险与误差代价函数的概率有关。由于参数向量 $\boldsymbol{\theta}$ 的真实值是未知的，因此需要对 $\boldsymbol{\theta}$ 的所有可能取值的代价求平均。

(2)代价函数。误差代价函数决定了贝叶斯解决方案的类型。最常用的一个代价函数是最小均方误差代价。

(3)似然函数。似然函数 $f_{Y|\Theta}(\boldsymbol{y}\mid\boldsymbol{\theta})$是给定参数 $\boldsymbol{\theta}$ 时观测值 $\boldsymbol{y}$ 的条件概率，它给出了由参数 $\boldsymbol{\theta}$ 决定观测信号 $\boldsymbol{y}$ 的可能性。

(4)先验概率。概率密度函数 $f_\Theta(\boldsymbol{\theta})$给出了参数向量 $\boldsymbol{\theta}$ 的先验概率。先验概率函数降低了参数向量 $\boldsymbol{\theta}$ 对似然函数的影响。似然函数和先验概率密度函数对估计的影响程度取决于信噪比、观测值长度和先验概率函数形状等。

(5)后验概率。后验概率与似然函数和先验概率密度函数的乘积成正比。

1.1.3　动态概率模型估计方法

优化估计算法，如卡尔曼滤波器使用了观测信号的动态概率模型。动态预测模型刻画了信号的相关结构，是对信号当前值和未来值与信号过去的变化轨迹及输入依赖关系的建模。概率统计模型用统计特性，如均值和方差，刻画一个随机信号的不同实现构成的空间。条件概率模型除对随机波动建模外，也可以对信号与其过去值和其他相关参数值或过程的关系进行建模。有限状态模型和卡尔曼滤波器是动态概率模型的两个例子。

对受噪声污染的观测向量 $\boldsymbol{y}=[y(0),\ y(1),\ \cdots,\ y(N-1)]$，其 P 维参数向量 $\boldsymbol{\theta}=[\theta_0,\ \theta_1,\ \cdots,\ \theta_{P-1}]$ 的估计可以用如下模型表示：

$$\boldsymbol{y}=\boldsymbol{x}+\boldsymbol{n}=h(\boldsymbol{\theta},\boldsymbol{e})+\boldsymbol{n} \tag{1.3}$$

如图 1.2 所示，假设纯净信号 $\boldsymbol{x}$ 是预测模型 $h(\cdot)$ 的输出，$\boldsymbol{e}$ 为预测模型的随机输入，$\boldsymbol{\theta}$ 为其参数向量，$\boldsymbol{n}$ 是加性随机噪声向量，则随机输入 $\boldsymbol{e}$、参数向量 $\boldsymbol{\theta}$ 和加性随机噪声向量 $\boldsymbol{n}$ 的概率密度函数分别为 $f_E(\cdot)$、$f_\Theta(\cdot)$ 和 $f_N(\cdot)$。概率密度函数模型常采用高斯模型。

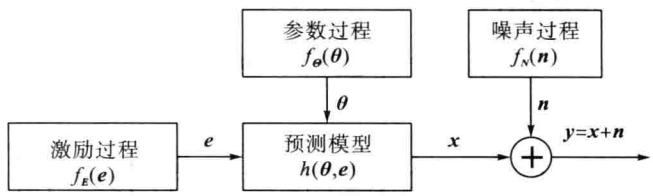


图 1.2　用预测模型 $h(\cdot)$ 表示随机过程 $\boldsymbol{y}$

随机过程的预测统计模型决定了未知参数的估计，这些未知参数与模型参数和观测值的先验分布是最一致的。一般来说，如果模型准确地刻画了观测值和参数过程，则估计过程中可利用的信息越多，估计结果就越可靠。反之，如果模型不准确，则估计结果不可靠。

1.1.4　估计的性能指标

通过 N 个观测样本 $\boldsymbol{y}$ 估计参数向量 $\boldsymbol{\theta}$ 时，需要对不同估计值的性能进行量化和比较。一般来说，参数向量是观测向量的函数，参数估计的结果取决于估计方法、观测值和先验知识。常用估计误差的均值和方差来表征估计的性能。评价估计性能最常用的指标包括：①估计期望值 $E[\hat{\boldsymbol{\theta}}]$；②估计偏差 $E[\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}]=E[\hat{\boldsymbol{\theta}}]-\boldsymbol{\theta}$；③估计协方差 $\mathrm{Cov}(\hat{\boldsymbol{\theta}})=E[(\hat{\boldsymbol{\theta}}-E[\hat{\boldsymbol{\theta}}])(\hat{\boldsymbol{\theta}}-E[\hat{\boldsymbol{\theta}}])^{\mathrm{T}}]$

一个最优估计的偏差为零，估计误差的协方差最小，它应该具有如下优良特性：

(1)无偏估计。如果估计期望等于真实参数值，则 $\boldsymbol{\theta}$ 的估计值是无偏的，即

$$E[\hat{\boldsymbol{\theta}}]=\boldsymbol{\theta} \tag{1.4}$$

如果随着观测长度 N 增加有

$$\lim_{N\to\infty}E[\hat{\boldsymbol{\theta}}]=\boldsymbol{\theta} \tag{1.5}$$

则估计结果是渐近无偏的。

(2)有效估计。如果与所有其他无偏估计 $\hat{\boldsymbol{\theta}}$ 相比具有最小协方差矩阵，则 $\boldsymbol{\theta}$ 的无偏估计是有效估计，即

$$\mathrm{Cov}[\hat{\boldsymbol{\theta}}_{Efficient}]\leqslant\mathrm{Cov}[\hat{\boldsymbol{\theta}}] \tag{1.6}$$

其中，$\hat{\boldsymbol{\theta}}$ 是 $\boldsymbol{\theta}$ 的其他任何估计量。

(3)一致估计。如果观测样本数 N 增加能改善估计值，当 N 变为无穷大时估计值 $\hat{\boldsymbol{\theta}}$ 依概率收敛于真实值 $\boldsymbol{\theta}$，则估计是一致的，即

$$\lim_{N\to\infty}P[\ |\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}|>\varepsilon]=0 \tag{1.7}$$

其中，ε 是任意小的正数。

例 1.1　信号均值和方差的估计。遍历随机过程 $\boldsymbol{y}$ 的 N 个观测样本为$[y(0), \cdots, y(N-1)]$，均值 μ_y 和方差 σ_y^2 的时间平均估计偏差可以计算如下：

$$\hat{\mu}_y = \frac{1}{N}\sum_{m=0}^{N-1} y(m) \tag{1.8}$$

$$\hat{\sigma}_y^2 = \frac{1}{N}\sum_{m=0}^{N-1}[y(m) - \hat{\mu}_y^2]^2 \tag{1.9}$$

容易证明，$\hat{\mu}_y$ 是一个无偏估计，即

$$E[\hat{\mu}_y] = \frac{1}{N}\sum_{m=0}^{N-1} E[y(m)] = \mu_y \tag{1.10}$$

方差 $\hat{\sigma}_y^2$ 估计的期望为

$$\begin{aligned} E[\hat{\sigma}_y^2] &= E\left[\frac{1}{N}\sum_{m=0}^{N-1}\left(y(m) - \frac{1}{N}\sum_{k=0}^{N-1} y(k)\right)^2\right] \\ &= \sigma_y^2 - \frac{2}{N}\sigma_y^2 + \frac{1}{N}\sigma_y^2 \\ &= \sigma_y^2 - \frac{1}{N}\sigma_y^2 \end{aligned} \tag{1.11}$$

从式(1.11)可以看出，方差估计的偏差反比于观测样本数 N，当 N 趋于无穷大时偏差消失，因此估计是渐近无偏的。在一般情况下，估计的偏差和方差随着观测样本数 N 的增加和模型的改进而降低。图 1.3 说明了渐进无偏估计的方差和偏差分布与观测样本数 N 的关系。

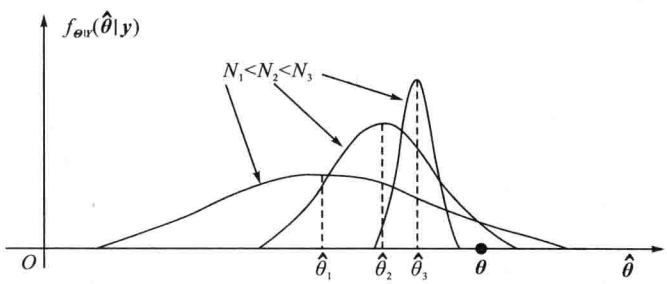


图 1.3　参数 $\boldsymbol{\theta}$ 的渐进无偏估计的偏差和方差随观测样本数的变化

1.2　贝叶斯估计

根据观测向量 $\boldsymbol{y}$ 对参数向量 $\boldsymbol{\theta}$ 进行贝叶斯估计，可以用贝叶斯风险最小平均代价的误差函数来定义：

$$\begin{aligned} R(\hat{\boldsymbol{\theta}}) = E[C(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})] &= \iint_{\boldsymbol{\theta}\,\boldsymbol{y}} C(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) f_{\boldsymbol{Y},\boldsymbol{\Theta}}(\boldsymbol{y}, \boldsymbol{\theta}) \mathrm{d}\boldsymbol{y}\mathrm{d}\boldsymbol{\theta} \\ &= \iint_{\boldsymbol{\theta}\,\boldsymbol{y}} C(\hat{\boldsymbol{\theta}},\ \boldsymbol{\theta}) f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y} \mid \boldsymbol{\theta}) f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{y}\mathrm{d}\boldsymbol{\theta} \end{aligned} \tag{1.12}$$

其中，误差代价函数 $C(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})$ 可对各种结果进行适当加权以得到所期望的估计。贝叶斯风险函数是在参数向量 $\boldsymbol{\theta}$ 和观测值 $\boldsymbol{y}$ 所有取值空间上进行平均。

对于给定的观测向量 $\boldsymbol{y}$，$f_{\boldsymbol{\Theta}|\boldsymbol{Y}}(\boldsymbol{\theta} \mid \boldsymbol{y})$ 是一个常数，对风险最小化过程没有影响。因

此，将 $f_{\boldsymbol{\Theta},\boldsymbol{Y}}(\boldsymbol{\theta},\boldsymbol{y})=f_{\boldsymbol{\Theta}\mid\boldsymbol{Y}}(\boldsymbol{\theta}\mid\boldsymbol{y})f_{\boldsymbol{Y}}(\boldsymbol{y})$代入式(1.12)，可以得到条件风险函数为

$$R(\hat{\boldsymbol{\theta}}\mid\boldsymbol{y})=\int_{\boldsymbol{\theta}}\mathbf{C}(\hat{\boldsymbol{\theta}},\boldsymbol{\theta})f_{\boldsymbol{\Theta}\mid\boldsymbol{Y}}(\boldsymbol{\theta}\mid\boldsymbol{y})\mathrm{d}\boldsymbol{\theta} \tag{1.13}$$

由此可得到 $\boldsymbol{\theta}$ 的贝叶斯估计 $\hat{\boldsymbol{\theta}}$（即最小风险参数向量）为

$$\hat{\boldsymbol{\theta}}_{\text{Bayes}}=\arg\min_{\hat{\boldsymbol{\theta}}}R(\hat{\boldsymbol{\theta}}\mid\boldsymbol{y})=\arg\min_{\hat{\boldsymbol{\theta}}}\left[\int_{\boldsymbol{\theta}}C(\hat{\boldsymbol{\theta}},\boldsymbol{\theta})f_{\boldsymbol{\Theta}\mid\boldsymbol{Y}}(\boldsymbol{\theta}\mid\boldsymbol{y})\mathrm{d}\boldsymbol{\theta}\right] \tag{1.14}$$

利用贝叶斯准则，式(1.14)可写成：

$$\hat{\boldsymbol{\theta}}_{\text{Bayes}}=\arg\min_{\hat{\boldsymbol{\theta}}}\left[\int_{\boldsymbol{\theta}}C(\hat{\boldsymbol{\theta}},\boldsymbol{\theta})f_{\boldsymbol{Y}\mid\boldsymbol{\Theta}}(\boldsymbol{y}\mid\boldsymbol{\theta})f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}\right] \tag{1.15}$$

如果风险函数是可微的，并有一个定义明确的最小值，则可以得到贝叶斯估计为

$$\hat{\boldsymbol{\theta}}_{\text{Bayes}}=\arg\operatorname{zero}_{\hat{\boldsymbol{\theta}}}\frac{\partial R(\hat{\boldsymbol{\theta}}\mid\boldsymbol{y})}{\partial\hat{\boldsymbol{\theta}}}=\arg\operatorname{zero}_{\hat{\boldsymbol{\theta}}}\left[\frac{\partial}{\partial\hat{\boldsymbol{\theta}}}\int_{\boldsymbol{\theta}}C(\hat{\boldsymbol{\theta}},\boldsymbol{\theta})f_{\boldsymbol{Y}\mid\boldsymbol{\Theta}}(\boldsymbol{y}\mid\boldsymbol{\theta})f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}\right] \tag{1.16}$$

1.2.1　最大后验估计

最大后验估计$\hat{\boldsymbol{\theta}}_{\text{MAP}}$是由参数向量后验概率密度函数的最大化得到的。它是具有均匀代价函数的贝叶斯估计，其代价函数是如图 1.4 所示的陷波形状，定义为

$$C(\hat{\boldsymbol{\theta}},\boldsymbol{\theta})=1-\delta(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}) \tag{1.17}$$

其中，$\delta(\hat{\boldsymbol{\theta}},\boldsymbol{\theta})$是克罗内克 δ 函数。

将代价函数代入贝叶斯风险公式中得到：

$$R_{\text{MAP}}(\hat{\boldsymbol{\theta}}\mid\boldsymbol{y})=\int_{\boldsymbol{\theta}}[1-\delta(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})]\,f_{\boldsymbol{\Theta}\mid\boldsymbol{Y}}(\boldsymbol{\theta}\mid\boldsymbol{y})\mathrm{d}\boldsymbol{\theta}=1-f_{\boldsymbol{\Theta}\mid\boldsymbol{Y}}(\hat{\boldsymbol{\theta}}\mid\boldsymbol{y}) \tag{1.18}$$

可以看出，参数向量 $\boldsymbol{\theta}$ 的最大后验估计是由贝叶斯风险最小化或后验函数的最大化得到的，即

$$\hat{\boldsymbol{\theta}}_{\text{MAP}}=\arg\max_{\boldsymbol{\theta}}f_{\boldsymbol{\Theta}\mid\boldsymbol{Y}}(\boldsymbol{\theta}\mid\boldsymbol{y})=\arg\max_{\boldsymbol{\theta}}[f_{\boldsymbol{Y}\mid\boldsymbol{\Theta}}(\boldsymbol{y}\mid\boldsymbol{\theta})f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})] \tag{1.19}$$

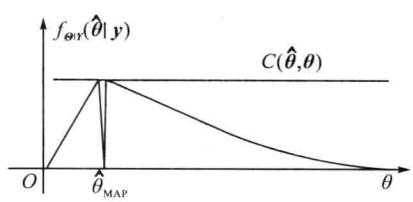


图 1.4　最大后验估计的贝叶斯代价函数

1.2.2　最大似然估计

最大似然估计$\hat{\boldsymbol{\theta}}_{\text{ML}}$是使似然函数 $f_{\boldsymbol{Y}\mid\boldsymbol{\Theta}}(\boldsymbol{y}\mid\boldsymbol{\theta})$最大的参数向量。其代价函数为陷波形状，参数先验概率密度函数为均匀分布时的贝叶斯估计：

$$R_{\text{ML}}(\hat{\boldsymbol{\theta}}\mid\boldsymbol{y})=\int_{\boldsymbol{\theta}}[1-\delta(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})]f_{\boldsymbol{Y}\mid\boldsymbol{\Theta}}(\boldsymbol{y}\mid\boldsymbol{\theta})f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}=\text{const.}[1-f_{\boldsymbol{Y}\mid\boldsymbol{\Theta}}(\boldsymbol{y}\mid\hat{\boldsymbol{\theta}})] \tag{1.20}$$

其中，先验概率密度函数 $f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$=常数。

最大似然估计和最大后验估计的主要区别在于，前者假定 $\boldsymbol{\theta}$ 的先验概率密度函数是均匀分布的。先验概率均匀分布除可用于真正均匀概率密度函数的建模外，还可用于先验概率密度函数的参数未知或参数是一个未知的常数时的建模。

由式(1.20)可知，风险函数最小化也可以由似然函数最大化实现，即

$$\hat{\boldsymbol{\theta}}_{\mathrm{ML}}=\arg\max_{\boldsymbol{\theta}} f_{\boldsymbol{Y}\mid\boldsymbol{\Theta}}(\boldsymbol{y}\mid\boldsymbol{\theta}) \tag{1.21}$$

在实际应用中，利用对数似然函数的最大化计算最大似然估计更方便，即

$$\hat{\boldsymbol{\theta}}_{\mathrm{ML}}=\arg\max_{\boldsymbol{\theta}} \log f_{\boldsymbol{Y}\mid\boldsymbol{\Theta}}(\boldsymbol{y}\mid\boldsymbol{\theta}) \tag{1.22}$$

对数似然函数具有良好的性质，如对数似然函数具有与似然函数相同的拐点，多个独立变量的联合对数似然函数等于各个变量似然函数之和，对数似然函数的动态范围不像似然函数那样会导致计算值溢出。

例 1.2　高斯过程的均值和方差的最大似然估计。设 P 维高斯向量过程为 $y(\mathrm{m})$，观测向量为$[y(0),\cdots,y(N-1)]$，现对该过程的均值向量和协方差矩阵进行最大似然估计。假设观测向量之间不相关，则观测向量序列的概率密度函数为

$$f_{\boldsymbol{Y}}(y(0),\cdots,y(N-1))=\prod_{m=0}^{N-1}\frac{1}{(2\pi)^{P/2}\mid\sum_{yy}\mid^{1/2}}\exp\left\{-\frac{1}{2}[y(m)-\mu_y]^{\mathrm{T}}\sum_{yy}^{-1}[y(m)-\mu_y]\right\} \tag{1.23}$$

相应的对数似然函数为

$$\ln f_{\boldsymbol{Y}}(y(0)\cdots,y(N-1))=\sum_{m=0}^{N-1}\left\{-\frac{P}{2}\ln(2\pi)-\frac{1}{2}\ln\mid\sum_{yy}\mid-\frac{1}{2}[y(m)-\mu_y]^{\mathrm{T}}\sum_{yy}^{-1}[y(m)-\mu_y]\right\} \tag{1.24}$$

求对数似然函数对均值向量的偏导数并令其等于零：

$$\frac{\partial\ln f_{\boldsymbol{Y}}(y(0),\cdots,y(N-1))}{\partial\mu_y}=\sum_{m=0}^{N-1}[-2\sum_{yy}^{-1}y(m)+2\sum_{yy}^{-1}\mu_y]=0 \tag{1.25}$$

由此可以得到均值的估计：

$$\hat{u}_y=\frac{1}{N}\sum_{m=0}^{N-1}y(m) \tag{1.26}$$

为了获得协方差矩阵的最大似然估计，取对数似然函数关于协方差矩阵的偏导数并令其等于零：

$$\frac{\partial\ln f_{\boldsymbol{Y}}(y(0),\cdots,y(N-1))}{\partial\sum_{yy}^{-1}}=\sum_{m=0}^{N-1}\left\{\frac{1}{2}\sum_{yy}-\frac{1}{2}[y(m)-\mu_y][y(m)-\mu_y]^{\mathrm{T}}\right\}=0 \tag{1.27}$$

得到协方差矩阵的估计值：

$$\hat{\sum}_{yy}=\frac{1}{N}\sum_{m=0}^{N-1}[y(m)-\hat{u}_y][y(m)-\hat{u}_y]^{\mathrm{T}} \tag{1.28}$$

例 1.3　一个随机高斯参数的最大似然估计和最大后验估计。利用 N 维观测向量 $\boldsymbol{y}$ 估计 P 维随机参数向量 $\boldsymbol{\theta}$。假设信号向量 $\boldsymbol{y}$ 和参数向量 $\boldsymbol{\theta}$ 的关系是一个线性模型，则

$$\boldsymbol{y}=\boldsymbol{G}\boldsymbol{\theta}+\boldsymbol{e} \tag{1.29}$$

其中，$\boldsymbol{e}$ 是一个随机激励信号输入；$\boldsymbol{G}$ 是一个数据矩阵。

例如，对于自回归过程，$\boldsymbol{G}$ 由 $\boldsymbol{y}$ 过去的取值组成。在已知观测向量 $\boldsymbol{y}$ 的情况下，参数向量 $\boldsymbol{\theta}$ 的概率密度函数可利用贝叶斯准则描述为

$$f_{\boldsymbol{\Theta}|\boldsymbol{Y}}(\boldsymbol{\theta}\mid\boldsymbol{y})=\frac{1}{f_{\boldsymbol{Y}}(\boldsymbol{y})}f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y}\mid\boldsymbol{\theta})f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) \tag{1.30}$$

假设式(1.29)中的 $\boldsymbol{G}$ 是已知的，则在给定参数向量 $\boldsymbol{\theta}$ 的条件下信号 $\boldsymbol{y}$ 的似然函数是随机向量的概率密度分布函数：

$$f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y}\mid\boldsymbol{\theta})=f_{\boldsymbol{E}}(\boldsymbol{e}=\boldsymbol{y}-\boldsymbol{G}\boldsymbol{\theta}) \tag{1.31}$$

进一步假设输入信号向量 $\boldsymbol{e}$ 是具有零均值和对角协方差矩阵的高斯分布随机过程，参数向量 $\boldsymbol{\theta}$ 也是一个均值为 μ_θ，协方差矩阵为 $\sum_{\theta\theta}$ 的高斯过程。因此，似然函数可以写成：

$$f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y}\mid\boldsymbol{\theta})=f_{\boldsymbol{E}}(\boldsymbol{e})=\frac{1}{(2\pi\sigma_e^2)^{N/2}}\exp\left[-\frac{1}{2\sigma_e^2}(\boldsymbol{y}-\boldsymbol{G}\boldsymbol{\theta})^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{G}\boldsymbol{\theta})\right] \tag{1.32}$$

$$f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})=\frac{1}{(2\pi)^{P/2}\left|\sum_{\theta\theta}\right|^{1/2}}\exp\left[-\frac{1}{2}(\boldsymbol{\theta}-\mu_\theta)^{\mathrm{T}}\sum_{\theta\theta}^{-1}(\boldsymbol{\theta}-\mu_\theta)\right] \tag{1.33}$$

对数似然函数 $\ln[f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y}\mid\boldsymbol{\theta})]$ 取得最大值时的 $\boldsymbol{\theta}$ 值就是最大似然估计值：

$$\hat{\boldsymbol{\theta}}_{\mathrm{ML}}(\boldsymbol{y})=(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G})^{-1}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{y} \tag{1.34}$$

为了获得最大后验估计值，首先将式(1.32)和式(1.33)代入式(1.30)，得到后验概率密度函数：

$$f_{\boldsymbol{\Theta}|\boldsymbol{Y}}(\boldsymbol{\theta}\mid\boldsymbol{y})=\frac{1}{f_{\boldsymbol{Y}}(\boldsymbol{y})}\frac{1}{(2\pi\sigma_e^2)^{N/2}}\frac{1}{(2\pi)^{P/2}\left|\sum_{\theta\theta}\right|^{1/2}}$$
$$\times\exp\left(-\frac{1}{2\sigma_e^2}(\boldsymbol{y}-\boldsymbol{G}\boldsymbol{\theta})^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{G}\boldsymbol{\theta})-\frac{1}{2}(\boldsymbol{\theta}-\mu_\theta)^{\mathrm{T}}\sum_{\theta\theta}^{-1}(\boldsymbol{\theta}-\mu_\theta)\right) \tag{1.35}$$

将对数似然函数 $\ln f_{\boldsymbol{\Theta}|\boldsymbol{Y}}(\boldsymbol{\theta}\mid\boldsymbol{y})$ 求偏导数并令其为零，即可得到最大后验估计值：

$$\hat{\boldsymbol{\theta}}_{\mathrm{MAP}}(\boldsymbol{y})=(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}+\sigma_e^2\sum_{\theta\theta}^{-1})^{-1}(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{y}+\sigma_e^2\sum_{\theta\theta}^{-1}\mu_\theta) \tag{1.36}$$

可以看到，随着高斯分布参数的协方差的增加，即协方差矩阵的逆矩阵 $\sum_{\theta\theta}^{-1}\to 0$，高斯先验概率分布趋于均匀的先验概率分布，最大后验估计值趋近于最大似然估计值。反过来，若参数向量 $\boldsymbol{\theta}$ 的概率密度函数越尖锐，即 $\sum_{\theta\theta}\to 0$，则最大后验估计值趋近于先验概率密度函数的均值 μ_θ。

1.2.3　最小均方误差估计

均方误差代价函数定义为

$$R_{\mathrm{MSE}}(\hat{\boldsymbol{\theta}}\mid\boldsymbol{y})=E\left[(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^2\mid\boldsymbol{y}\right]=\int_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^2 f_{\boldsymbol{\Theta}|\boldsymbol{Y}}(\boldsymbol{\theta}\mid\boldsymbol{y})\mathrm{d}\boldsymbol{\theta} \tag{1.37}$$

贝叶斯最小均方误差估计是使 $R_{\mathrm{MSE}}(\hat{\boldsymbol{\theta}}\mid\boldsymbol{y})$ 最小化的参数向量，如图 1.5 所示。下面将证明，贝叶斯最小均方误差估计是后验概率密度函数的条件均值。假设均方误差风险函数是可微的，且有确定的最小值，则令均方误差风险函数的梯度值为零，可求得最小均方误差估计值，即

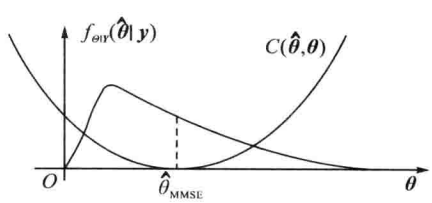


图 1.5　均方误差代价函数和估计