

# 通用鲁棒的放松估计方法

原理和应用

Principles and Applications of RELAX

A Robust and Universal Estimator

吴仁彪 贾琼琼 杨磊 冯青 © 著



科学出版社

# 通用鲁棒的放松估计方法 原理和应用

吴仁彪 贾琼琼 著  
杨 磊 冯 青

科学出版社

北京

## 内 容 简 介

在噪声和干扰背景中多个混叠在一起的信号解混和参数估计问题是一个在雷达、声呐、通信、导航等领域经常遇到的问题。如何提高参数估计方法在低信噪比下的估计性能和存在模型误差时的鲁棒性一直是信号处理和控制学界的研究焦点。本书针对上述问题，给出了一种通用和鲁棒的放松估计方法，介绍了它的基本原理和在许多方面获得的应用。

全书共分7章。第1章介绍了参数估计的基本理论，第2章介绍了放松估计方法的基本原理，第3~5章分别介绍了放松估计方法在经典的线谱估计、时延估计、波达方向估计问题中的应用，第6章介绍了放松估计方法在雷达目标成像中的应用，第7章简要介绍了放松估计方法在其他方面的典型应用情况。

本书内容全面丰富，用恋爱语言生动和形象地讲述了参数估计的基本问题以及5种经典参数估计方法的基本原理和彼此之间的关系，它既可以作为信号处理和控制领域广大科技工作者的参考读物，也可以作为相关领域研究生的教材。

---

### 图书在版编目(CIP)数据

通用鲁棒的放松估计方法原理和应用/吴仁彪等著. —北京: 科学出版社, 2017

ISBN 978-7-03-052802-5

I. ①通… II. ①吴… III. ①参数估计 IV. ①O211.67

中国版本图书馆CIP数据核字(2017)第107253号

---

责任编辑: 赵丽欣 / 责任校对: 王万红  
责任印制: 吕春珉 / 封面设计: 东方人华

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>

北京中科印刷有限公司印刷

科学出版社发行 各地新华书店经销

\*

2017年5月第一版 开本: 787×1092 1/16

2017年5月第一次印刷 印张: 16 1/4 插页: 1

字数: 385 000

定价: 82.00 元

(如有印装质量问题, 我社负责调换〈中科〉)

销售部电话 010-62136230 编辑部电话 010-62134021

版权所有, 侵权必究

举报电话: 010-64030229; 010-64034315; 13501151303



吴仁彪 1966年2月生于武汉市。1991年在西北工业大学首届教改试点班（五年半本硕连读）毕业，1994年在西安电子科技大学获博士学位。先后四次以博士后、访问教授、国家首批高级研究学者的身份在美国佛罗里达大学和英国帝国理工大学工作近五年。现任中国民航大学智能信号与图像处理天津市重点实验室主任，天津市重点领域“民航安全信息处理”科技创新团队负责人，天津市首批三位特聘教授之一，中国民航总局首批特聘专家，交通运输专家委员会委员，国家安全生产专家组成员，天津大学和西安电子科技大学博士生导师，武汉大学和深圳大学兼职教授，IEEE高级会员，中国电子学会理事、学术工作委员会委员、信号处理分会常务委员，天津市信号处理专业委员会主任，3个英文刊物区域编辑和编委。研究方向为自适应信号处理、阵列信号处理和现代谱估计及其在雷达、卫星导航、通信和空中交通管理中的应用。共发表SCI和EI检索学术论文300余篇，其中50余篇发表在IEEE和IEE/IET会刊上。曾获省部级科技成果奖励10余项，国家发明专利50余项，出版学术专著3部（其中英文1部），英文著作特邀章节6部，译著2部，科普图书1部。1999年入选国家人事部百千万人才工程第一、二层次培养对象，2003年荣获国家杰出青年基金，2009年当选为天津市十大杰出留学人员。



著者吴仁彪（左）与 RELAX 的提出者、IEEE 会士、美国 Florida 大学 Jian Li 教授（右）在美国 Florida 州 Orlando 市出席 2002 年 ICASSP 会议期间合影



著者吴仁彪（左五）2002 年在美国 Florida 大学做访问教授期间合影照，左六为 RELAX 的共同作者、IEEE 会士和信号处理协会技术成就奖获得者、瑞典 Uppsala 大学 Petre Stoica 教授，右五为 IEEE 会士和信号处理协会技术成就奖获得者、美国 Minnesota 大学 Georgios B. Giannakis 教授，左七是 Jian Li 教授，其他人是她负责的实验室的几个中国博士生和博士后，很多与著者有论文合作关系。

## 自序

什么是信号处理？“信号处理是一个隐身技术，只有真正搞信号处理的人才会思考，我们是谁？我们是干什么的？”（Signal processing is a “stealth technology” that only the people who actually work in signal processing think about who we are and what we do）这是 IEEE 信号处理协会前会长 Don Johnson 博士对信号处理的一句非常经典的解释，道出了信号处理的本质。而微弱信号的检测与参量估计是信号处理最基础的理论之一。近年来，我国航空航天事业取得了让世人瞩目的成就：神舟上天、嫦娥奔月、北斗导航，天宫号空间站虽然正在建设过程之中，但已经出现在美国科幻电影“Gravity”的惊险故事情节当中了。我国预警飞机、反隐身雷达、航空母舰上的航空吊放声呐探潜水艇系统等相继投入使用，大大提升了我国的防御实力，但在它们背后隐身的核心技术却都是微弱目标信号的检测与参量估计。2015 年，美国科学家通过激光干涉仪引力波天文台（LIGO）的两个相距较远的探测器成功探测到了引力波，验证了爱因斯坦的世纪预言，这更涉及极其微弱信号的检测与参量估计问题。

信号检测与参量估计的理论基础是信息论，其数理基础涉及概率、随机过程、数理统计、线性代数和矩阵分析、最优化理论和算法等，20 世纪 40 年代雷达和声呐的相继投入应用极大地推动了其理论的发展，目前已经成为通信、雷达、声呐、导航、自动控制、医学、天体观测、地震勘探等领域的共性基础。参数估计方法是依据建模技术对于信号和系统进行辨识，当模型与实际情况较吻合时，参数化方法往往具有比非参数化方法更好的估计性能，但是它对于各种模型误差较敏感，包括信号、噪声或干扰模型误差、传感器阵列校准误差（位置误差、通道失配、互耦等）、信号源个数不确定性、观察样本数有限导致的估计误差等。如何使参数化估计方法在各种模型误差存在时依然保持较好的性能一直是信号处理界的研究焦点和永恒的话题，这种性能测度在不同领域有不同的提法，控制界通常称“鲁棒性”，在雷达、通信和导航界通常称“稳健性”，声呐界通常称“宽容性”，在日常生活中常称“皮实”，本书中我们用“鲁棒性”。

1996~1998 年，2001 年和 2002 年，著者曾经 3 次以博士后和访问教授的身份到美国 Florida 大学电子与计算机工程系的谱分析实验室工作，与该实验室主任、年轻的 IEEE 会士 Jian Li 教授开展了卓有成效的合作。2004~2007 年，Jian Li 教授又在国家自然科学基金委海外杰出青年合作研究基金项目的资助下，每年到笔者负责的中国民航大学天津市智能信号与图像处理重点实验室工作两个月。1996 年，她提出了一种称之为 RELAX 的多正弦波信号估计方法，这是一个基于非线性最小二乘（Nonlinear Least Squares, NLS）准则和松弛（RELAXation-based）的参数化循环优化算法，该论文被原国际科技信息所（ISI）评为工程领域论文发表当年引用率最高的前 1%。后来我们合作将 RELAX 推广应用到时延估计、雷达目标成像、机载雷达地面动目标高距离分辨率成像、探地雷达、水下超空泡飞行器空泡形状控制等方面，合作发表 20 余篇 IEEE/IEE 刊物论文。笔者所在中国民航大学的科研团队在

国内同时开展了 RELAX 在卫星导航自适应抗干扰、机载气象雷达等方面的应用研究，发表了数十篇相关论文。

对于多个“信号”（此处“信号”是广义的，不一定是正弦波）混叠在噪声或干扰中的参数估计问题，非线性最小二乘（NLS）方法是解决这类问题的一种常用方法，在白的或色的高斯噪声甚至非高斯噪声背景下，NLS 方法均具有与最大似然方法相同或类似的估计性能，也比较鲁棒。由于它不估计噪声中的参数，所以运算量比最大似然方法要小，但 NLS 方法无法回避信号参数空间的多维搜索，通常运算量仍然较大，也难以保证收敛到全局最优解。

本书将 RELAX 的基本思想加以推广，用于解决这类问题，我们称之为通用的“放松估计方法”，英文简称为 GRELAX（Generalized RELAX）。和 RELAX 类似，它也是将多信号参数估计问题通过信号分离估计方法或特殊结构的循环优化方法转变成一系列单信号参数估计问题，不仅可以大大减少运算量，而且具有很好的全局收敛特性，也不需要单独的初始化过程。放松估计方法是一种通用的多信号参量估计方法，具有对于各种模型误差不敏感的特点，因而也是一种鲁棒的估计方法。“放松”一词取自“RELAX”算法的中文直接翻译，而 RELAX 又出自“RELAXation”（“松弛”）。“放松估计方法”也隐含着 RELAX 方法性能好且比较鲁棒，大家可以放心大胆地使用它，因为它是性能优越和鲁棒的 NLS 估计方法的高效实现形式。

全书共分 7 章。第 1 章介绍了参数估计的基本理论，包括最大似然估计、贝叶斯估计、线性最小均方误差估计的基本原理，评价估计器性能好坏的标准，给出了用于高斯背景下矢量参数估计克拉美-劳界（CRB）计算的紧凑和通用的 Slepian-Bangs 公式。对于经典的贝叶斯估计理论给予了它新的内涵，从“贝叶斯公式”讲到了当前的“大数据背景下的贝叶斯革命”。第 2 章先提出了一种具有广泛代表性的通用的多“信号”参数估计问题的数据模型，介绍了另一类参数估计的著名方法——最小二乘法，包括线性和非线性最小二乘估计的基本理论，介绍了 NLS 的直接求解方法和循环优化方法，在此基础上给出了通用的放松估计方法（GRELAX）的基本原理及实现框架，指出了 GRELAX 与目前较热的压缩感知中匹配追踪（Matching Pursuit, MP）类贪婪算法的关系。第 3 章介绍了 GRELAX 在线谱估计方面的应用，包括一维和二维混合谱估计、指数衰减和任意包络正弦信号参数估计。第 4 章介绍了 GRELAX 在时延估计方面的应用，包括一般时延估计方法 WRELAX，以及相关函数高度振荡、时延间隔很近及色噪声背景下利用多快拍数据的估计方法。第 5 章介绍了 GRELAX 在波达方向估计（Direction of Arrival, DOA）问题上的应用，包括窄带和宽带信号源的 DOA 和波形估计。第 6 章专门讨论了 GRELAX 在雷达目标成像领域的应用，包括合成孔径雷达（Synthetic Aperture Radar, SAR）自聚焦和半参数化成像方法、曲线 SAR 自聚焦和三维成像、逆合成孔径雷达（Inverse Synthetic Aperture Radar, ISAR）成像和机动目标 ISAR 成像。第 7 章则简要介绍了 GRELAX 在其他各方面的典型应用，包括机载相控阵预警雷达空中机动目标检测、机载雷达地面动目标高距离分辨率成像、机载气象雷达、探地雷达、卫星导航自适应抗干扰、水下超空泡航行器空泡形状控制、压缩感知中的稀疏阵列信号处理、生物医学信号处理等。

笔者 1999 年回国后，一直给中国民航大学和天津大学信息与通信工程学科的研究生讲

授“Advanced signal processing”双语课程，它包括“Spectral analysis”和“Adaptive filtering”两大部分。它们都属于信号检测与估计的范畴，其中估计则包括“参量估计”和“波形估计”。“参量估计”主要属于“Spectral analysis”的范畴，而“波形估计”主要属于“Adaptive filtering”的范畴。在讲授此课的过程中，笔者先翻译出版了《现代信号谱分析》（已再版），并于2015年出版了国内首本同类专著《卫星导航自适应抗干扰技术》，本书算是它们的姊妹篇，主要讨论信号检测与参量估计问题。为了把复杂信号处理的理论和概念讲清楚，笔者在给研究生上课的过程中突发灵感，采用恋爱的语言介绍了最大似然估计、贝叶斯估计、线性最小均方误差估计、普通最小二乘估计、整体最小二乘估计的基本原理和彼此关系，很受学生欢迎。2015年全国信号处理大会在天津市中国民航大学召开，笔者应邀做了题为“信号处理教育与空管智能信息处理”的大会报告，其中关于用恋爱语言讲信号处理理论的方式受到了与会专家的好评，纷纷索要报告的PPT。所以，在本书第1章和第2章介绍参量估计基本理论和放松估计方法基本原理时保留了这种表述方式。

感谢美国Florida大学Jian Li教授给笔者提供了多次去她实验室工作与学习的机会，在她实验室工作期间，很幸运得到了瑞典Uppsala大学国际著名信号处理大师Petre Stoica教授给予的许多指导和帮助。感谢Jian Li教授和谱分析实验室同事Zhengshe Liu博士、Guosui Liu博士、Xi Li博士、Zhaoqiang Bi博士、Nanzhi Jiang博士、Jianhua Liu博士和Kunlong Gu与笔者的长期合作与交流，本书含有许多他们与笔者合作的研究成果。在本书撰写过程中，Jian Li教授给笔者提供了她们实验室在RELAX方面最新的研究成果。此外，Jian Li教授强烈建议在本书合适的地方，指出RELAX和目前较活跃的压缩感知中匹配追踪（Matching Pursuit, MP）类贪婪算法的关系。她在最近给笔者的来信中指出：“RELAX was first published in 1996, well before the emergence of “compressive sensing” and matching pursuit (MP) related algorithms, including MP, Orthogonal MP(OMP), CoSeMP(Compressive Sensing MP), Least-Squares MP, etc. These are the so-called greedy methods in the compressive sensing literature, which have been used in diverse applications due to the compressive sensing topic being a red hot topic for over a decade. Yet, MP is just CLEAN. RELAX is Super CLEAN, and hence should outperform most of the these MP variations in most applications.”我们采纳了她的建议，并在第2章专门进行了讨论。

感谢西安电子科技大学雷达信号处理国防科技重点实验室保铮院士（笔者的博士生导师）和邢孟道教授，在本书第6章分享了他们在机动目标ISAR成像方面的研究成果。

感谢王晓峰博士同意我们在第1章中以附录形式收录他在《赛先生》公众号发表的姊妹篇综述文章“大数据背后的神秘公式（上）：贝叶斯公式”和“大数据背后的神秘公式（下）：贝叶斯革命”。这是笔者很欣赏的两篇文章，很好地介绍了贝叶斯估计的基本原理、发展历史、成功的应用案例和在大数据时代可能迎来的爆炸式发展。

感谢我指导的硕士生王光力博士撰写了第7.6节，介绍他在上海交通大学攻读博士期间，如何巧妙地把在我们实验室学习到的GRELAX用于视网膜神经元活动多电极记录信号的处理。

感谢中国民航大学天津市智能信号与图像处理重点实验室的合著者贾琼琼、杨磊和冯青三位教师，还有为本书撰写提供了大量帮助的教师和研究生们，教师包括韩萍教授、李海教



授、贾桂敏博士、卢晓光博士、王文益教授、卢丹博士、何炜琨博士、胡铁桥副教授、苏志刚教授、王璐讲师，博士研究生包括毕葵和包莉娜，硕士研究生包括吴琛琛、张昊、任文、牛志华、朱威、湛蕾、马晨曦、刘超、张瑞华、陈蕾、赵岸斐、刘娟、李佳怡。

感谢中国民航工会 2016 年 11 月中旬安排笔者以全国劳模名义到云南工人俱乐部疗养，使笔者在昆明滇池边美丽和安静的环境中高效率地完成了本书的第三轮修改工作。

感谢国家自然科学基金委从 1996 年以来对笔者持续给予的支持，包括国家杰出青年基金、海外杰出青年合作研究基金、重点和面上项目共 12 项。本书得到了国家自然科学基金项目（61471363，61231017）的资助。

吴仁彪

初稿于 2016 年国庆节写于中国民航大学

2017 年春节定稿于美丽的珠海市

# 目 录

第 1 章 参数估计的基本理论	1
1.1 引言	1
1.2 最大似然估计	1
1.3 贝叶斯估计	3
1.3.1 随机参数估计模型	4
1.3.2 常用代价函数	4
1.3.3 风险评估	5
1.4 线性最小均方误差估计	8
1.4.1 估计准则	9
1.4.2 正交性原理	10
1.5 估计器的性能测度	11
1.6 克拉美-劳界 (CRB)	12
1.7 几种估计方法的比较	15
1.8 大数据时代的贝叶斯革命	17
本章小结	18
附录 1-A 一般分布下矢量参数估计的 CRB	19
附录 1-B 高斯分布下矢量参数估计的 CRB	19
附录 1-C 大数据背后的神秘公式 (上): 贝叶斯公式	21
附录 1-D 大数据背后的神秘公式 (下): 贝叶斯革命	27
参考文献	32
第 2 章 放松估计方法的基本原理	35
2.1 引言	35
2.2 线性最小二乘估计	36
2.2.1 普通最小二乘解	36
2.2.2 整体最小二乘解	38
2.3 非线性最小二乘估计	39
2.3.1 可化简的问题	40
2.3.2 常规迭代求解算法	41
2.3.3 循环最小化器	43
2.4 放松估计方法	44
2.4.1 多正弦信号参数估计的 RELAX 算法	44
2.4.2 通用多信号参数估计的 GRELAX 算法	51

本章小结 .....	53
参考文献 .....	53
第3章 GRELAX 在线谱估计中的应用 .....	56
3.1 引言 .....	56
3.2 正弦信号参数估计 .....	56
3.2.1 一维正弦信号的混合谱估计 .....	57
3.2.2 二维正弦信号的混合谱估计 .....	60
3.2.3 实验结果 .....	63
3.3 指数衰减正弦信号参数估计 .....	71
3.3.1 数据模型 .....	71
3.3.2 DRELAX 算法 .....	71
3.3.3 实验结果 .....	74
3.4 任意包络正弦信号参数估计 .....	75
3.4.1 数据模型 .....	75
3.4.2 单信号源的参数估计 .....	76
3.4.3 多信号源的模糊问题 .....	78
3.4.4 实验结果 .....	79
本章小结 .....	81
附录 3-A 正弦信号参数估计的 CRB .....	81
附录 3-B 指数衰减正弦信号参数估计的 CRB .....	82
附录 3-C 任意包络正弦信号参数估计的 CRB .....	83
参考文献 .....	83
第4章 GRELAX 在时延估计中的应用 .....	87
4.1 引言 .....	87
4.2 数据模型 .....	87
4.3 WRELAX 算法 .....	88
4.3.1 基本原理 .....	88
4.3.2 实验结果 .....	91
4.4 相关函数高度振荡时的时延估计方法 .....	98
4.4.1 Hybrid-WRELAX 算法 .....	99
4.4.2 EXIP-WRELAX 算法 .....	101
4.4.3 实验结果 .....	102
4.5 超分辨率时延估计方法 .....	108
4.5.1 复值信号的 MODE-WRELAX 算法 .....	108
4.5.2 实值信号的 MODE-WRELAX 算法 .....	111
4.5.3 MODE-WRELAX 算法的高效实现 .....	112
4.5.4 实验结果 .....	115

4.6 色噪声背景下基于多次快拍数据的时延估计方法 .....	125
4.6.1 数据模型 .....	125
4.6.2 TWRELAX 算法的基本原理 .....	125
4.6.3 实验结果 .....	129
本章小结 .....	133
附录 4-A 复信号情况下时延估计的 CRB .....	133
附录 4-B 实信号情况下时延估计的 CRB .....	134
附录 4-C 色噪声背景下多快拍数据时延估计的 CRB .....	136
参考文献 .....	137
<b>第 5 章 GRELAX 在波达方向估计中的应用 .....</b>	<b>141</b>
5.1 引言 .....	141
5.2 窄带信号源 DOA 估计 .....	142
5.2.1 阵列基本概念 .....	142
5.2.2 数据模型 .....	144
5.2.3 阵列数据的统计特性 .....	145
5.2.4 NB-RELAX 算法 .....	147
5.2.5 实验结果 .....	149
5.3 宽带信号源 DOA 估计 .....	152
5.3.1 数据模型 .....	152
5.3.2 WB-RELAX 算法 .....	154
5.3.3 实验结果 .....	154
本章小结 .....	156
附录 5-A 窄带源 DOA 估计的 CRB .....	156
附录 5-B 宽带源 DOA 估计的 CRB .....	157
参考文献 .....	159
<b>第 6 章 GRELAX 在雷达目标成像中的应用 .....</b>	<b>163</b>
6.1 引言 .....	163
6.2 合成孔径雷达 (SAR) 成像 .....	164
6.2.1 数据模型 .....	164
6.2.2 自聚焦 MCCLEAN 算法及实验结果 .....	166
6.2.3 半参数化 SPAR 成像算法及实验结果 .....	175
6.3 曲线 SAR 三维成像 .....	190
6.3.1 数据模型 .....	190
6.3.2 自聚焦和三维成像方法 .....	192
6.3.3 实验结果 .....	194
6.4 逆合成孔径雷达 (ISAR) 成像 .....	196
6.4.1 数据模型 .....	198

6.4.2	基于单个特显点的 AUTOCLEAN 算法	200
6.4.3	基于多个特显点的 AUTOCLEAN 算法	201
6.4.4	实验结果	205
6.5	机动目标 ISAR 成像	209
6.5.1	三维运动目标二维成像模型	209
6.5.2	分距离单元基于波数谱函数重构目标图像	210
6.5.3	基于回波时频分布重构目标图像	212
6.5.4	机动目标 ISAR 成像算法	214
6.5.5	实验结果	216
	本章小结	218
	参考文献	218
<b>第 7 章</b>	<b>GRELAX 的其他典型应用</b>	<b>224</b>
7.1	引言	224
7.2	在雷达目标探测中的应用	224
7.2.1	相控阵机载预警雷达空中机动目标检测	224
7.2.2	机载雷达地面运动目标高距离分辨率成像	226
7.2.3	机载气象雷达	229
7.2.4	探地雷达	231
7.3	在卫星导航抗干扰中的应用	233
7.3.1	压制式干扰抑制	233
7.3.2	欺骗式干扰抑制	235
7.3.3	多径干扰抑制	236
7.4	在水下超空泡航行器空泡形状控制中的应用	237
7.5	在压缩感知 DOA 估计中的应用	239
7.6	在生物医学工程神经元信息解混中的应用	240
	本章小结	242
	参考文献	242

# 第 1 章 参数估计的基本理论

## 1.1 引言

现代估计理论是很多提取有用信息的电子信号处理系统的核心<sup>[1]-[17]</sup>，包括雷达、导航、声呐、通信、控制、语音、图像、生物医学、地震勘探、物理实验、经济分析与处理系统等<sup>[18]-[41]</sup>。在这些应用中，往往需要对数据进行建模，然后再估计模型参数，这就是参数估计问题。感兴趣的物理量（或信号参数）通常只能从观测量来估计得到，例如信号的幅度、频率和相位。然而实际获得的观测量除包含感兴趣的信号之外，还不可避免地受到噪声和干扰的影响而表现出极强的随机性，因此，所能得到的只是对信号参数在某种意义上的最佳估计<sup>[1]-[17]</sup>。

参数估计有着悠久而辉煌的历史，最早可以追溯到 1795 年（此时间尚有争议）高斯用最小二乘法来预测行星的运动<sup>[42]</sup>。在参数估计理论中，通常把一个真实参数  $\theta$  的估计值或估计方法称为  $\theta$  的估计器。一个估计器是一个统计量，它在某种意义下“最接近”真实参数  $\theta$ 。那么，如何对真实参数进行估计？以及如何衡量或评价一个估计器与真实参数之间的“接近度”呢？这些问题形成了参数估计理论的两个核心内容。

从以上问题出发，并考虑到感兴趣的参数可以是确定性或随机性的未知参数，本章 1.2 节介绍了广泛用于确定性参数估计的最大似然估计方法；1.3 节则介绍了用于随机参数估计的贝叶斯方法；1.4 节介绍了线性最小均方误差估计；为了评价估计器的性能，1.5、1.6 节给出了估计器性能测度和克拉美-劳界；1.7 节对上述估计方法进行了比较，并用恋爱语言生动和形象地说明了其基本原理和差别；1.8 节简要介绍了贝叶斯方法与当下最受瞩目的大数据革命的渊源；最后对本章内容进行了小结。关于参数估计的最小二乘方法将在第 2 章及后续章节中重点介绍。

## 1.2 最大似然估计

最大似然（Maximum Likelihood, ML）估计是最常用和最有效的估计方法之一。最大似然估计的基本思想是，当被估计的未知参数没有任何先验信息的情况下，利用已知的若干观测数据估计该参数<sup>[3]</sup>。在使用最大似然估计方法时，被估计的参数  $\theta$  假定为未知但确定，而已知的观测数据则是随机变量。此时，参数估计的模型可用图 1.1 来描述，它由如下四部分组成<sup>[7]</sup>。

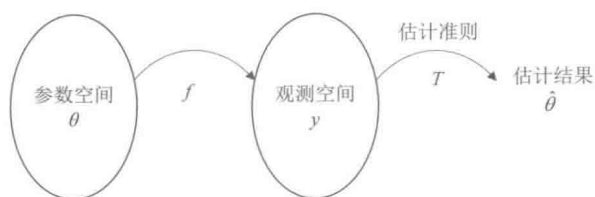


图 1.1 确定性参数估计的模型

### 1) 参数空间

参数空间表示未知参数  $\theta$  的取值范围，这里讨论的未知参数  $\theta$  是一个确定量，可以被视为参数空间中的一个点。参数  $\theta$  可以是一维的，也可以是多维的，为了表述简便起见，本章正文部分仅以一维实参数估计为例加以讨论，所有结论可直接扩展至多维实或复参数估计的情况。

### 2) 从参数空间到观测空间的映射

表示参数  $\theta$  与观测值  $y$  之间的函数关系， $y = f(\theta) + e$ ，其中  $e$  表示噪声、干扰等随机因素的影响。

### 3) 观测空间

观测值  $y$  可被视为是观测空间中的一个点。将可获得的  $N$  个观测样本表示为集合  $\{y_0, y_1, \dots, y_{N-1}\}$ ，通常还可用  $\mathbf{y} = [y_0, y_1, \dots, y_{N-1}]^T$  来表示多个样本值构成的样本矢量，此处  $(\cdot)^T$  表示转置算子。

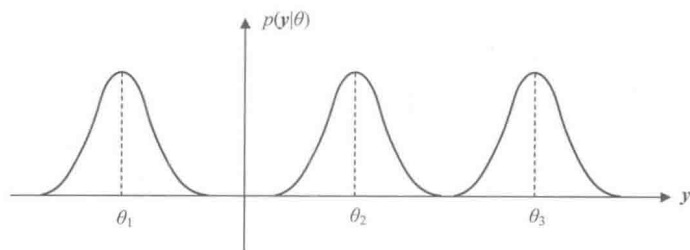
### 4) 估计准则

利用观测样本矢量  $\mathbf{y}$  获得参数  $\theta$  的估计值，表示为  $\hat{\theta} = T(\mathbf{y})$ ，称为估计器。观测空间至估计值所在的参数空间的映射被称为估计准则。

参数估计的第一步就是建立观测数据的数学模型。由于数据固有的随机性，通常用它的概率密度函数来描述它，即  $p(\mathbf{y}|\theta)$ 。当考虑有  $N$  个样本时，联合条件概率分布密度函数简记为

$$p(\mathbf{y}|\theta) = p(y_0, y_1, \dots, y_{N-1}|\theta) \quad (1-1)$$

概率密度函数以未知量  $\theta$  为参数，因此可以得到一族概率密度函数，其中每一个概率密度函数因为  $\theta$  的不同而不同。不同  $\theta$  对应的概率密度函数如图 1.2 所示，显然，由于  $\theta$  的取值影响  $\mathbf{y}$  的概率，因此可以从观测到的样本数据  $\mathbf{y}$  估计出  $\theta$ 。例如，如果  $\mathbf{y}$  的值是负的，那么  $\theta = \theta_2$  或  $\theta = \theta_3$  就值得怀疑，而  $\theta = \theta_1$  则更合理。

图 1.2 依赖于未知参数  $\theta$  的条件概率密度函数

条件概率密度函数  $p(\mathbf{y}|\theta)$  也称为似然函数, 就是包含未知参数  $\theta$  信息的可能性函数。最大似然估计就是使似然函数  $p(\mathbf{y}|\theta)$  最大化的估计器, 记为

$$\hat{\theta}_{\text{ML}} = \arg \max_{\theta} p(\mathbf{y}|\theta) \quad (1-2)$$

因此, 最大似然估计也可以看作是样本数据条件概率密度函数  $p(\mathbf{y}|\theta)$  的全局极大点。最大似然估计也可以用图 1.3 来加以解释: 对于某个给定的参数  $\theta$ ,  $\mathbf{y}$  落在一个小区域的概率为  $p(\mathbf{y}|\theta)d\mathbf{y}$ 。图 1.3 是在给定  $\mathbf{y}=\mathbf{y}_0$  时计算得到的 (这里矢量  $\mathbf{y}_0$  表示矢量  $\mathbf{y}$  的一次具体观测值)。对于每一个  $\theta$ ,  $p(\mathbf{y}=\mathbf{y}_0|\theta)d\mathbf{y}$  的值表明了对于给定  $\theta$  值,  $\mathbf{y}$  落在观测空间中以  $\mathbf{y}_0$  为中心的  $d\mathbf{y}$  区域内的概率。如果已经观测到  $\mathbf{y}=\mathbf{y}_0$ , 那么可推断  $\theta=\theta_1$  是不太可能的, 因为如果  $\theta=\theta_1$ , 那么实际上观测到  $\mathbf{y}=\mathbf{y}_0$  的概率应该非常小。看起来  $\theta=\theta_2$  更有可能是真值, 此时观测到  $\mathbf{y}=\mathbf{y}_0$  的概率应该很高。因此可以选择  $\hat{\theta}=\theta_2$  作为估计值, 即选择在  $\theta$  允许的范围, 使  $p(\mathbf{y}=\mathbf{y}_0|\theta)$  最大的  $\theta$  值作为估计值<sup>[3]</sup>。

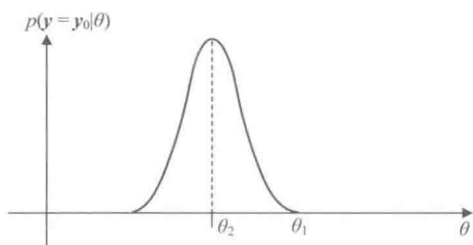


图 1.3 最大似然估计原理示意图

严格地说,  $p(\mathbf{y}|\theta)$  与观测样本  $\mathbf{y}$  的任意函数相乘的结果都可以看作一个似然函数, 而本书只称样本数据的条件概率密度函数  $p(\mathbf{y}|\theta)$  本身为似然函数, 显而易见, 随机变量  $\mathbf{y}$  的不同实现会给出不同的  $p(\mathbf{y}|\theta)$ , 所以似然函数的全局极大点与观测样本值有关, 即最大似然估计与观测样本有关。从这个意义上讲, 最大似然估计  $\hat{\theta}_{\text{ML}}$  是一个合理的估计器。

由于对数函数是单调函数, 所以  $\ln p(\mathbf{y}|\theta)$  和  $p(\mathbf{y}|\theta)$  将会在同个  $\theta$  值上达到极大值,  $\ln p(\mathbf{y}|\theta)$  称为对数似然函数, 常用以代替似然函数  $p(\mathbf{y}|\theta)$ 。为了书写方便, 记为

$$L(\theta) = \ln p(\mathbf{y}|\theta) \quad (1-3)$$

$\theta$  的最大似然估计可以通过令似然函数的偏导数等于零来求得, 这可表示为

$$\frac{\partial L(\theta)}{\partial \theta} = \frac{\partial \ln p(\mathbf{y}|\theta)}{\partial \theta} = 0 \quad (1-4)$$

### 1.3 贝叶斯估计

1.2 节讨论了当感兴趣的参数  $\theta$  是确定性未知参量的最大似然估计理论。本节研究  $\theta$  为随机变量时的参数估计问题, 事实上这里要估计的物理量是随机变量  $\theta$  的一次特定实现, 这就是贝叶斯方法 (Bayesian 估计)。运用贝叶斯方法代替最大似然估计的主要原因是, 如果已知  $\theta$  的一些先验知识, 那么就可以将先验知识应用到估计器中<sup>[3][4]</sup>, 因为任何先验知识的利用都将有助于提高估计器的性能。



### 1.3.1 随机参数估计模型

为了便于后文讨论，这里先给出几种常用的概率密度函数的定义，并介绍它们之间的关系。观测数据  $\mathbf{y}$  和未知参数  $\theta$  的联合概率密度函数可表示为

$$p(\mathbf{y}, \theta) = p(\mathbf{y}|\theta)p(\theta) = p(\theta|\mathbf{y})p(\mathbf{y}) \quad (1-5)$$

给定观测数据  $\mathbf{y}$  时， $\theta$  的条件概率密度函数称为后验概率密度函数，记作  $p(\theta|\mathbf{y})$ 。后验概率密度函数是指在得到观测数据  $\mathbf{y}$  后， $\theta$  的概率密度函数。与后验概率密度函数相对应，下式

$$p(\theta) = \int_{-\infty}^{\infty} p(\mathbf{y}, \theta) d\mathbf{y} \quad (1-6)$$

可以看作  $\theta$  的先验概率密度函数，它表示在数据被观测到之前的概率密度函数。

假设  $\theta$  是一个随机变量，其概率密度函数  $p(\theta)$  已知，那么贝叶斯估计过程可用图 1.4 来描述。其中参数的样本空间是指随机变量  $\theta$  的取值范围。如果样本空间是离散的，那么问题变成多重假设检验问题；如果样本空间是连续的，就需要分配代价函数来估计。参数的样本空间到观测空间之间的映射是参数  $\theta$  对观测值影响的概率规则，以  $p(\mathbf{y}|\theta)$  表示。

从前面的介绍已知，当  $\hat{\theta}$  作为  $\theta$  的估计时，估计误差通常不为零。因此估计值  $\hat{\theta}$  的质量取决于估计误差究竟多大，可以利用误差的范围作为估计性能的测度，这种测度称为代价函数，记为  $C(\bullet)$ 。在确定代价函数之后，通过使  $C(\bullet)$  对应的代价函数最小化从而得到参数的估计值。

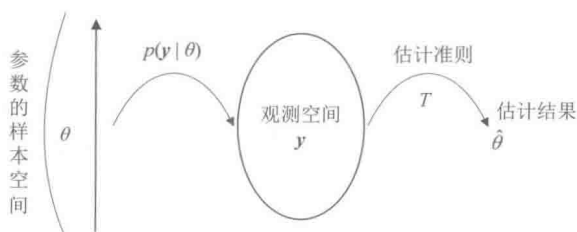


图 1.4 随机参数估计的模型

### 1.3.2 常用代价函数

性能测度是用以判断估计器性能优劣的标准，可以有很多形式，最常用的是代价函数  $C(\bullet)$  的期望值。这里代价函数等于真值和估计值之差（误差  $\varepsilon \triangleq \theta - \hat{\theta}$ ）的函数，因此性能测度可定义为

$$J_{\text{BAYES}} \triangleq E[C(\varepsilon)] \quad (1-7)$$

由  $\mathbf{y}$  对  $\theta$  的贝叶斯估计器记为  $\hat{\theta} = T(\mathbf{y})$ ，这里  $T(\mathbf{y})$  的选取应该是使  $J_{\text{BAYES}}$  最小的函数。如果选择如下平方误差代价函数

$$C_{\text{MMSE}}(\varepsilon) = \varepsilon^2 \quad (1-8)$$

则对应的性能测度被称为均方误差，此时  $T(\mathbf{y})$  是  $\theta$  的最小均方误差（Minimum Mean Squared Error, MMSE）估计。除了上述平方误差代价函数，一般较常用的代价函数还有绝对值误差（Absolute Error, ABS）和均匀误差（Uniform Error, UN），分别定义为