普通高等教育地震类专业“九五”省部级重点教材

反演理论及其应用

徐果明　编著

地震出版社

普通高等教育地震类专业“九五”省部级重点教材

反演理论及其应用

徐果明　编著

地震出版社

图书在版编目（CIP）数据

反演理论及其应用/徐果明编著．—北京：地震出版社，2003.6
ISBN 7－5028－2216－X

Ⅰ．反…　Ⅱ．徐…　Ⅲ．地球物理学－重力反演问题　Ⅳ．P312

中国版本图书馆 CIP 数据核字（2003）第 003429 号

内 容 简 介

本书中所讨论的内容，其核心部分是基于概率反演的概念，也包括反演理论的其他方面。在举例中较多地涉及地球物理学，但其基本精髓可适用于其他学科。本书内容并不涉及针对特殊专门情况下的反演方法，而着重讨论带有普遍意义的反演理论。全书共分八章。第一章为概论；第二章到第四章以概率反演为线索，讨论概率反演的概念和方法；第五章研究了直接反演的方法，即对于非线性问题的反演方法；第六章讨论了连续模型空间的反演；第七章介绍了线性系统反演的一些方法；第八章讨论了反演的一个重要应用方面：层析成像。有关某些数学推导的细节放在附录中。本书可作为高等院校相关专业的研究生教材，也可作为相关的科研和产业部门科技人员的参考书。

反演理论及其应用
徐果明　编著
责任编辑：薛广盈
责任校对：王花芝

出版发行：地震出版社

北京民族学院南路 9 号　　邮编：100081
发行部：68423031　68467993　　传真：88421706
门市部：68467991　　传真：68467972
总编室：68462709　68423029　　传真：68467972
E－mail：seis@ht.rol.cn.net

经销：全国各地新华书店
印刷：北京地大彩印厂

版（印）次：2003 年 6 月第一版　2003 年 6 月第一次印刷
开本：787×1092　1/16
字数：461 千字
印张：18
印数：0001～1000
书号：ISBN 7－5028－2216－X/P·1156（2784）
定价：50.00 元

目　　录

第一章 概　论

人类为了更好地生存，创造一个与自然界协调发展的生存环境，首先要认识自然环境，了解自然界中各种事物之间的关系。在长期的生产实践和科学实验中，人类得到了一系列规律性的认识，例如牛顿定律、相对论理论等。这些研究物质世界的一系列基本规律便构成了科学研究的一个分支——物理学。

在已经掌握了事物的发展规律和事物之间相互作用规律的基础上，便有可能由已知参数——源参数及介质参数，根据物理规律来推测及计算观测（主要是仪器观测）得到的资料与数据，或推测源本身参数的变化，我们称此种情况为正演。天气预报便是一种十分典型的正演，根据已有的观测资料（这也可算做是源参数），利用大气动力学的理论来预测各种气象参数在时间上的变化，属时间尺度上的正演；由地下介质的密度异常，根据万有引力定律计算重力场的异常（这种重力异常是可以通过仪器观测得到的），是空间尺度上的正演；由地震的震源参数，根据弹性动力学规律，计算各地震台应该记录到的地震动，是既有空间尺度又有时间尺度上的正演；天气预报事实上也包含有时间尺度上的正演和空间尺度上的正演，如预报台风中心在未来数天内的移动路线。由此可知，正演是由原因推测结果，即由因到果。

而反演则正相反，是由结果推测原因，即由果到因。而结果应该是可以观测到的结果，称之为观测资料。一般由果推测因可分为两种情况：一种是用于建立理论模型，如人们由天文台的观测资料来推测星系的结构及演变规律，并建立星系形成、演变的理论框架；又如在地震震源的研究中，人们由各地震台的地震记录、大地测量资料及野外的直接观测来研究地震震源机制；另一种情况是假定已经建立了一定的理论模型框架，则可以由观测资料来推测理论模型中的若干个参数，如根据地震台的观测资料来测定震中、发震时间、震源断层的走向、断层面的倾角和断裂错动的滑动角等参数。本书讨论的反演便属于后者，而建立理论模型的研究工作则属于各专门学科的课题。

与气象观测不同，地球内部的各种参数和物理过程是难于直接观测到的，人们只有通过地面的观测（如用地震仪、重力仪、地电及地磁仪器），或通过在空间中的观测（如航空重力、航磁观测等），甚至通过对人造卫星轨道的观测，来推断地球内部介质的地震波速、密度、电导率等参数分布，从而得到地球内部介质分布的二维或三维结构图像。对地球物理学而言，绝大多数情况都是进行反演。反演不仅在地球物理学中占有极其重要的地位，在其它学科中也是十分重要的。如在物理学中对原子核结构的研究，人们目前并不能直接观测到原子核内部结构，而是要用粒子去轰击，通过观测粒子的衍射图像来推断原子核的结构，这就是一种反演。在医学中，CT 诊断也是一种反演，通过从各个方向对人体照射 X 射线得到的观测资料来推断人体内部的三维图像（实际工作中是一层一层的二维切片图），从而使医生得到病灶区的图像。在天文学中，用航天器中的仪器对太阳日冕的观测来推断日冕内部的结构，也是一种反演。当然，这并不说明正演不重要，正相反，在地球物理学中正演是反演的基础，实际的反演计算所含的正演部分占去大部分的计算机机时。

总之，在科学实践中，反演是一种相当普遍的手段。本书中所讨论的内容，其核心部分是

基于概率反演的概念,在举例中较多地涉及地球物理学,但其基本精髓可适用于其它学科。在地球物理学中,对于各种观测的物理量,如地震波、地磁、地电等,针对各种特别的观测手段和观测系统的布局及所要反演的内容,还有许多具体的反演计算方法,这些构成了地球物理学中各分支学科的核心内容。本书内容并不涉及针对特殊情况下的具体反演方法,而只讨论带有普遍意义的反演理论。

在用直线或曲线对观测数据进行数据拟合时,人们常用最小二乘法,这实际上是进行直线或曲线参数的反演,后又提出了阻尼最小二乘法以解决反演中的发散问题。利用数学中矩阵的奇异值分解理论,又发展了广义反演方法。以上各种方法都基于要求理论数据与观测数据的偏差最小,来求取反演中理论模型的参数。广义反演的实质略有不同,这在后面会详细讨论。20世纪60年代末,Backus和Gilbert求取最佳分辨率为出发点,发展了一套反演方法,人们称之为BG方法。在BG方法中,模型是连续函数,因此可以有无穷多维,而观测数据的数目是有限的,BG方法通过引入分辨率函数 $A(r_0,r)$ 来对模型加以限制,要求取得最佳分辨率来得到反演向量函数 $\boldsymbol{a}(r_0)$。到20世纪80年代,法国数学家Tarantola以概率论为基础发展了概率反演的一整套理论和方法,而以前的最小二乘法、阻尼最小二乘法等都成了概率反演的特例。在概率反演的实现中,要用到数学中的最优化理论。最优化理论中的许多方法,如最速下降法、牛顿法、共轭梯度法、变尺度法等在概率反演中起着十分重要的作用,因此本书将涉及最优化理论中的一些内容。

当数据集有大残差的数据甚至是错误数据时,如果反演理论仅建立在高斯误差概率分布的基础上,这些大残差的数据会对结果产生很大的影响。为了减小它们的影响,发展了建立在 l_1 范数基础上的反演理论和方法,Cleabout等发展了先进先出(FIFO)的反演方法,而建立在 l_1 范数基础上另一个有效方法是以线性规划理论为基础。本书对涉及的线性规划理论也作了相应的介绍。

如果反演的对象是一个连续或分段(片、块)连续的函数分布时,有两种处理方法,一种是一开始就将函数离散化,化为一组离散的未知模型参数,再对这些参数进行反演,这就是离散模型参数的反演;另一种思路是将反演理论直接建立在对连续函数进行反演的基础上,即直接对连续函数分布进行反演,这时离散反演中的求和变成了积分。在实际情况中,数据往往还是离散的,即仅模型是连续函数,而数据仍是离散的,涉及模型时离散反演中的求和号变成了积分,而涉及数据的求和号仍保留。当然在实际反演中还是会对连续分布的模型参数逐点计算,并且在积分计算时才将模型离散化通过插值进行积分,以便计算机进行数值积分。还有一种做法是将连续模型函数对正交函数基进行分解,反演连续模型在正交函数基上的投影系数,这样在反演中,仍可沿用离散反演的方法,这时反演的是模型参数函数分布的投影系数,而不是模型参数本身。医学层析成像中雷当变换及雷当逆变换的理论属于连续函数的反演,它通过一个二重积分可以直接得到模型参数的连续分布,而通过专门布局得到的数据构成二重积分中的被积函数,当然在具体计算中还是要对模型参数逐点反演。但是在地球物理的实践中,往往不能满足雷当变换对观测系统的布局要求,并且如果是地震波射线,则它们是弯曲的,也会使理论复杂化。因此往往一开始就将模型参数离散化,这时地球物理中的层析成像仍属于离散参数的反演,或者采用基于连续模型参数的概率反演等方法,这与医学中的基于雷当变换的层析成像方法有本质的不同。

在实际问题中,绝大多数情况下模型参数与观测数据都构成非线性的关系。这时常用的方

法是将非线性方程线性化，然后，利用线性反演的理论进行反演。前面所简述的各种方法都是基于线性反演，而对非线性方程则是将其线性化后进行反演迭代计算。在对非线性的数据与模型的理论关系式进行线性化时都涉及偏导数的计算，即求数据参数对模型参数的偏导数矩阵，所涉及的计算量往往非常大。另外，当非线性较强时，要求迭代的初值比较接近于真值，否则迭代结果得到的有可能不是模型参数的真值。如果非线性问题由于非线性太强或者计算量太大不能用上述利用偏导数矩阵进行反演的方法，而希望只进行正演计算，这时所涉及的方法统称为直接反演方法。直接反演方法中最简单而直接明了的是试错法，或称为尝试法，它适用于反演参数数目不大的情况。之后又发展了单纯形法、样板搜索法等方法；而蒙特卡洛方法，则是用生成随机数的方法来逐个生成模型参数组，对每一组参数进行正演，选取这样的参数组，使其通过正演计算得到的理论数据与观测数据之间的偏差最小。Kirkpatrick 等(1983) 提出了将模拟退火的方法用于多参数情况下的最优化问题。其理论基础是基于液态物质在慢慢冷却过程中的能态分布特征，称为 Gipps 分布。而冷却过程中能态分布的基本算法是由 Metropolise(1953) 提出来的。Rothman 等(1985,1986) 首先将模拟退火方法应用于地球物理领域，他们研究了地震勘探数据的大静校正的估计问题。以后 Mosegoard 等(1991) 将此技术用于一维地震波形反演。Martinez 等(2000) 将模拟退火应用于用瑞利波相、群速度反演地层结构。近年来，随着生物基因工程学的发展，人们利用基因工程学中的概念发展了遗传算法。遗传算法的初期研究于20世纪60年代，1975年美国芝加哥大学的 Holland 教授出版了《自然界中的自适应和人工系统》一书，系统地阐述了遗传算法的理论和方法。遗传算法是基于遗传、选择和变异三种要素，对一群模型参数组(称为一个种群) 进行选择、交叉和变异操作，从而生成新的一群模型参数组。通过选择使具有优良特性的个体(模型参数组) 得到更多的机会参加交叉或直接传代。交叉产生的后代一般会继承上一代父母双方的特征，但也允许种群中少数后代的性能发生变异，即个体的某个或某几个模型参数发生突变，突变的概率不能太大，否则种群特性不稳定，但突变在上述遗传算法中起着十分重要的作用，它能使种群中产生新的特性，再通过选择得以保留具有优秀特性的个体。遗传算法在工程领域和自然科学中多参数的最优化选取中得到了应用。而反演问题往往可以化为多参数的最优化选取问题。

第二章　反演问题的概率方法

§2.1　模型空间和数据空间

2.1.1　物理系统

反演问题中的研究对象是一个物理系统，研究过程大致可分为三个方面：

(1) 系统参数化：用最少的参数来合理地描述一个物理系统．用最少的参数反映该系统的基本特征，反映其物理本质．如用矩形断层面的位错滑动模型来描述一个地震震源，选取断层面中心的深度、断层面的长和宽、断层面的走向及倾角、断层的滑动方向、断层的平均滑距（或滑动的空间分布及滑动的时间函数）及断层面两侧介质的弹性常数等参数来描述这个地震震源．通过这些断层模型参数，能反映地震震源的基本物理特征．对强震则要用更复杂的几何形状来描述断层面，并且可能要将断层面分为几个区域，各区域发生地震位错的时间和位错的大小和方向可以各不相同，并且位错在断层面上的扩展有一个时间过程，这样使问题的研究不断深化．最少数量的参数是要求这些参数是必不可少的，且其中某个参数不能通过其它参数的直接推导得到．如矩形断层面顶部的深度或底部的深度是多余的，因为它们可以通过断层面中心的深度、断层面的长和宽、断层面的走向及倾角得到．

(2) 物理系统的正演：研究物理系统的物理规律，进行归纳总结，并由物理系统的参数通过物理规律来预测可测参数．如给定了地震震源的参数后，通过地震波的激发和传播规律，可推断出可观测量 —— 地震波所激发的地面运动的特征，这就是物理系统的正演．又如给定物理参数 —— 地下的物质密啡分布，通过物理规律 —— 万有引力定律，可推导出可观测量 —— 空间重力值的分布，这也是物理系统的正演．

(3) 物理系统的反演：由物理系统产生的可观测参数的观测值，根据已掌握的物理规律，来推断物理模型参数值，这就是物另系统的反演．这正是本书要研究的内容．如由空间的重力分布观测值来推断地下物质密度的分布，由地震波的观测记录来推断地震震源参数或地下介质弹性参数的分布图像等等．

2.1.2　模型空间

假设有抽象点集合的空间，空间中的每一个点代表系统一个假想的模型，此空间称为**模型空间**，以 $\mathfrak{M}$ 来表示．为区别和醒目，下面将用英文字母的哥德体来表示这种抽象空间．此种抽象空间并不依赖于特定的参数化方法．但为了对系统进行定量化的讨论，就必须选定一种特定的参数化方法，即用一组特定的参数来定量地描述物理系统．一旦选定一种参数化方法，就可以将此组参数的数值与模型空间中的点相联系．这时的模型空间记为 $\boldsymbol{M}$ 空间．假设 $\boldsymbol{M}$ 空间有 n 维，则 $\boldsymbol{M}$ 空间与 $\boldsymbol{R}_n$ 空间同构，两者的区别在于：$\boldsymbol{R}_n$ 空间中的每一个点对应 n 个实数，是无量纲的量，而 $\boldsymbol{M}$ 空间中的每一个点对应的 n 个参数的数值都带有一定的量纲，例如密度、温度的数值都是带有相应物理量单位的量．

反演并非一定要与某个特定的参数化方法相联系.事实上我们可以研究模型空间 $\boldsymbol{M}$ 中的测度,测度是一种映射,它把 $\boldsymbol{M}$ 空间中的任一子集 $\boldsymbol{A}$ 与一正实数相联系.即有

$$P(\boldsymbol{A}) \in \boldsymbol{R}.$$

如果 $P(\boldsymbol{A})$ 为有限正实数,则可以将 $P(\boldsymbol{A})$ 归一化,得

$$P(\boldsymbol{M}) = 1,$$

这时可将测度称为概率.从概率论的角度来理解,$P(\boldsymbol{A})$ 为模型空间中真实模型 $\in$ 子集 $\boldsymbol{A}$ 的概率,而 $\boldsymbol{M}$ 中任一点对应的P相当于概率密度.反演是根据观测的资料求 $\boldsymbol{M}$ 空间中这样的点,其概率密度为最大.在实际的反演计算中,人们总是把模型空间参数化,这样便于计算.

2.1.3 数据空间

为了要得到模型空间中的有关信息,必须进行一系列的实验.例如为了了解地下地质构造的分布,在地震勘探的工作中进行一系列的人工爆炸,通过检波器阵列获得大量地震记录.也就是对可测参数 —— 地面的运动进行了一系列的测量,得到了大量的观测数据.这些观测数据的集合所构成的空间称为数据空间.数据空间中测量数据的获取与要想由此得到的物理模型有密切的联系.如要想研究地层介质电磁参数的分布特征,就要通过电磁的方法进行实验和观测,用地震的方法一般就无效果.如要了解岩层的弹性参数,则要用地震的方法,用电磁的方法一般就没有效果.因为对于观测量而言,它带有模型空间的某些参数的信息量比较大,而对于模型空间的某些参数其信息量又特别小,甚至无信息量.不能用带有对模型空间的某种参数信息量特别小的观测量来反演.如用瑞利面波的相速度或群速度频散(速度随频率而变化的变化关系)来反演地下介质的弹性速度随深度的分布.由于瑞利面波的频散对地层内横波速度比较敏感,因此用面波频散的观测资料能很好地反演地下横波速度的结构.但是瑞利面波的频散对地层内纵波速度的敏感程度相比横波而言要差得多,因此用面波频散资料反演纵波速度效果就很差.

当选定被直接观测的参数后,所有可能的观测值所组成的空间称为**数据空间** $\boldsymbol{D}$.测量中的某一个特定的实现称为 $\boldsymbol{D}$ 中的一个点,记为 $\boldsymbol{d}$.它是一组观测参数的数值,其中每一参数值称为 $\boldsymbol{d}$ 的某一个分量,记为 d_i.

$$\boldsymbol{d} = \{d_i\}, \qquad i \in I_{\mathrm{D}}.$$

如果满足下列关系式

$$\left.\begin{aligned} (\boldsymbol{d}^{(1)} + \boldsymbol{d}^{(2)})_i &= d_i^{(1)} + d_i^{(2)}, \qquad && i \in I_{\mathrm{D}}. \\ (r\boldsymbol{d})_i &= r d_i, && i \in I_{\mathrm{D}},\ r \in R. \end{aligned}\right\}$$

则数据空间 $\boldsymbol{D}$ 称为**线性数据空间**,其中每一个点 $\boldsymbol{d}$ 称为**数据向量**.

2.1.4 联合参数空间

定义积空间 $\boldsymbol{X} = \boldsymbol{D} \times \boldsymbol{M}$,其中元素 $\boldsymbol{x} = (\boldsymbol{d}, \boldsymbol{m})$,而 $\boldsymbol{d}$ 为观测参数,$\boldsymbol{m}$ 为模型参数.$\boldsymbol{X}$ 空间称为**联合参数空间**或简称为参数空间.可将 $\boldsymbol{X}$ 理解为包括测量仪器本身在内的更广泛的物理系统 $\boldsymbol{S}$.$\boldsymbol{x}$ 为参数空间 $\boldsymbol{X}$ 的一个点,$\boldsymbol{x}$ 的分量可看作参数空间的坐标,记为 x_i 或

$$\boldsymbol{x} = \{x_i\}, \qquad i \in I_{\mathrm{X}}.$$

§2.2 信　　息

2.2.1 测度、概率

1) 测度的定义

令 X 为一任意集合(空间),设对任何子集 $A \subset X$,均对应有正实数 $P(A)$,且 $P(A)$ 满足以下性质:

(1) 如 $\boldsymbol{\Phi}$ 为空集,则 $P(\boldsymbol{\Phi}) = 0$;

(2) 如 $\boldsymbol{A}_1, \boldsymbol{A}_2 \cdots$ 表示 $\boldsymbol{X}$ 的不相交子集,那么

$$P(\sum_i \boldsymbol{A}_i) = \sum_i P(\boldsymbol{A}_i),$$

则可称 $P(A)$ 为 A 的测度.

2) 概率的定义

如 $P(X)$ 一定为有限正实数,则称 P 为 X 的**概率**,一般可将概率 P 进行归一化,使

$$P(X) \equiv 1.$$

对概率可作出两种物理解释:

(1) 统计性质:对一个随机的物理过程(如抛硬币),导致一系列的实现(如硬币的正面或反面向上),对于观测到的大量实现可用概率来描述某种事实(如硬币正面向上)出现的可能性;

(2) 主观认知的性质:即对某一物理参数的真值主观上的认知程度,如所要反演的模型空间中某个模型参数的真值只有一个,但反演得到的只是一种概率分布,即反演结果可以是一个数据集,其中每一个子集 A 对应有出现的概率.

3) 测度密度函数和概率密度函数的定义

设 P 为参数空间 X 中的测度,设对 X 选定一特定的坐标系,$\{x_1, x_2, \cdots\}$ 为 X 中点的坐标.如存在函数 $f(\boldsymbol{x})$,使对任意的子集 $A \subset X$,均有

$$P(A) = \int \mathrm{d}x_1 \int \mathrm{d}x_2 \cdots \int f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}_N = \int f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}, \tag{2.2.1}$$

则 $f(\boldsymbol{x})$ 称为**测度密度函数**.

如果 $P(\boldsymbol{x})$ 为有限,则 $f(\boldsymbol{x})$ 称为**概率密度函数**.一般情况坐标分量 $x_1, x_2 \cdots$ 均是有量纲的量,因此概率密度函数也是有量纲的量.

4) 坐标变换时概率密度函数的变换关系

令 P 为 X 空间中定义的概率,$f(\boldsymbol{x})$ 为给定的坐标中对应 P 的概率密度函数,令有坐标变换 $\boldsymbol{x}^* = \boldsymbol{x}^*(\boldsymbol{x})$,则由概率的定义可知,对任一子集 $A \subset X$,

$$P(A) = \int_A f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = \int_A f(\boldsymbol{x}^*) \mathrm{d}\boldsymbol{x}^* .$$

另外有

$$\int_A f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = \int_A f(\boldsymbol{x}) \left| \frac{\partial \boldsymbol{x}}{\partial \boldsymbol{x}^*} \right| \mathrm{d}\boldsymbol{x}^* ,$$

其中 $\left|\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{x}^*}\right|$ 为坐标变换的雅可比行列式.由于 $\boldsymbol{A}$ 为任意子集,因此有

$$f(\boldsymbol{x}^*) = f(\boldsymbol{x})\left|\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{x}^*}\right|. \tag{2.2.2}$$

5) 联合概率密度函数的定义

设 $\boldsymbol{x} = \{x_1, x_2, \cdots\}$ 上有概率密度函数 $f(\boldsymbol{x})$,可将坐标 $\{x_1, x_2, \cdots\}$ 分成两组,分别记为 $\boldsymbol{y}$ 和 z,得 $\boldsymbol{x} = \{\boldsymbol{y}, z\}$.可记为 $f(\boldsymbol{x}) = f_1(\boldsymbol{y}, z)$,称 $f_1(\boldsymbol{y}, z)$ 为**联合概率密度函数**.

6) 边际概率密度函数的定义

边际概率密度函数定义为

$$f_Y(\boldsymbol{y}) = \int_Z f(\boldsymbol{y}, z)\mathrm{d}z. \tag{2.2.3}$$

7) 条件概率密度函数的定义

联合概率密度函数 $f(\boldsymbol{y}, z)$,在 $\boldsymbol{y} = \boldsymbol{y}_0$ 时对 z 的概率密度分布称为**条件概率密度分布**,或**条件概率密度函数**,即有

$$f_{Z|Y}(z \mid \boldsymbol{y}_0) = \frac{f(\boldsymbol{y}_0, z)}{\int_Z f(\boldsymbol{y}_0, z)\mathrm{d}z}. \tag{2.2.4}$$

由式(2.2.3) 及(2.2.4) 可知,联合概率密度 $f(\boldsymbol{y}, z)$ 为条件概率密度乘边际概率密度,即

$$\left.\begin{aligned} f(\boldsymbol{y}, z) &= f_{Z|Y}(z \mid \boldsymbol{y}) f_Y(\boldsymbol{y}), \\ f(\boldsymbol{y}, z) &= f_{Y|Z}(\boldsymbol{y} \mid z) f_Z(z). \end{aligned}\right\} \tag{2.2.5}$$

8) 贝叶斯定理

贝叶斯定理可表达为

$$f_{Z|Y}(z \mid \boldsymbol{y}) = \frac{f_{Y|Z}(\boldsymbol{y} \mid z) f_Z(z)}{\int_Z f_{Y|Z}(\boldsymbol{y} \mid z) f_Z(z)\mathrm{d}z}. \tag{2.2.6}$$

证:

$$\begin{aligned} f_{Z|Y}(z \mid \boldsymbol{y}) &\overset{\text{式(2.2.4)}}{=} \frac{f(\boldsymbol{y}, z)}{\int_Z f(\boldsymbol{y}_0, z)\mathrm{d}z} \\ &\overset{\text{式(2.2.5)}}{=} \frac{f_{Y|Z}(\boldsymbol{y} \mid z) f_Z(z)}{\int_Z f_{Y|Z}(\boldsymbol{y} \mid z) f_Z(z)\mathrm{d}z}. \end{aligned}$$

9) 中心估计值和离差估计值

(1) 一维情况.

设有一维变量 x,其概率密度函数为 $f(x)$,定义函数

$$S_p(m) = \left[\int_{-\infty}^{\infty} |x - m|^p f(x)\mathrm{d}x\right]^{1/p}, \tag{2.2.7}$$

式中　p 为正实数.

对给定的 p 值,使 S_p 为极小的 m 称为在 l_p 范数意义下的变量 x 的**中心估计值**,记为 m_p.称 m_1 为**中位数**,满足

$$\int_{-\infty}^{\infty} |x - m_1| f(x) \mathrm{d}x \tag{2.2.8a}$$

为极小.它等价于

$$\int_{-\infty}^{m_1} f(x) \mathrm{d}x = \int_{m_1}^{\infty} f(x) \mathrm{d}x = \frac{1}{2}. \tag{2.2.8b}$$

称 m_2 为**均值**,它使

$$\left[\int_{-\infty}^{\infty} (x - m_2)^2 f(x) \mathrm{d}x\right]^{1/2} \tag{2.2.9a}$$

为极小.由式(2.2.9a)易证

$$m_2 = \int_{-\infty}^{\infty} x f(x) \mathrm{d}x. \tag{2.2.9b}$$

称 m_∞ 为**中列数**,它使

$$\lim_{p \to \infty}\left[\int_{-\infty}^{\infty} |x - m_\infty|^p f(x) \mathrm{d}x\right]^{1/p} \tag{2.2.10a}$$

为极小.由式(2.2.10a)可得

$$m_\infty = \frac{x_{\sup} + x_{\inf}}{2}, \tag{2.2.10b}$$

其中 $x_{\sup}$ 及 $x_{\inf}$ 分别为满足 $f(x) \neq 0$ 条件下的 x 的最大值和最小值.

定义处于极小值的 $S_p(m)$ 值为 l_p 范数意义下的**离差**,记为 σ_p,即

$$\sigma_p = S_p(m_p). \tag{2.2.11}$$

称 σ_1 为**均差**,

$$\sigma_1 = \int_{-\infty}^{\infty} |x - m_1| f(x) \mathrm{d}x, \tag{2.2.12a}$$

或有

$$\sigma_1 = \int_{m_1}^{\infty} x f(x) \mathrm{d}x - \int_{-\infty}^{m_1} x f(x) \mathrm{d}x. \tag{2.2.12b}$$

称 σ_2 为**标准差**

$$\sigma_2 = \left[\int_{-\infty}^{\infty} (x - m_2)^2 f(x) \mathrm{d}x\right]^{1/2}, \tag{2.2.13a}$$

易证有

$$\sigma_2 = \left[\int_{-\infty}^{\infty} x^2 f(x) \mathrm{d}x - m_2^2\right]^{1/2}. \tag{2.2.13b}$$

称 σ_∞ 为**半幅**

$$\sigma_\infty = \lim_{p \to \infty}\left[\int_{-\infty}^{\infty} |x - m_\infty|^p f(x) \mathrm{d}x\right]^{1/p}, \tag{2.2.14a}$$

易证有

$$\sigma_\infty = \frac{x_{\sup} - x_{\inf}}{2}. \tag{2.2.14b}$$

(2) 多维情况.

设有变量 $\boldsymbol{x} = \{x_i, i \in I_x\}$,其概率密度函数为 $f(\boldsymbol{x})$.对多维情况,仅限于讨论 l_2 范数的情况.定义

$$C_{ij}(\boldsymbol{m}) = \int_{-\infty}^{\infty} (x_i - m_i)(x_j - m_j) f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}, \tag{2.2.15}$$

使 $C_{ij}(\boldsymbol{m})$ 的对角元素为极小的向量 $\boldsymbol{m}_2$,称为在 l_2 范数意义下变量 $\boldsymbol{x}$ 的**均值或数学期望**. $\boldsymbol{m}_2$ 满足,

$$\boldsymbol{m}_2 = \int_{-\infty}^{\infty} \boldsymbol{x} f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}. \tag{2.2.16}$$

$C_{ij}(\boldsymbol{m})$ 在 $\boldsymbol{m} = \boldsymbol{m}_2$ 处的值 $C_{ij}(\boldsymbol{m}_2)$ 称为在 l_2 范数意义下 $\boldsymbol{x}$ 的**协方差**,简记为 $\boldsymbol{C}$,

$$\boldsymbol{C} = \boldsymbol{C}(\boldsymbol{m}_2). \tag{2.2.17}$$

l_2 范数意义下的协方差算子 $\boldsymbol{C}$ 具有下列性质:

① 对称性:

$$C_{ij} = C_{ji}, \tag{2.2.18}$$

② 非负定性:对任一向量 $\boldsymbol{x}$,均有

$$\boldsymbol{x}^{\mathrm{T}} \boldsymbol{C}^{-1} \boldsymbol{x} \geqslant 0. \tag{2.2.19}$$

定义**相关函数**

$$\rho_{ij} = \frac{C_{ij}}{(\sigma_2)_i (\sigma_2)_j}, \tag{2.2.20a}$$

有

$$-1 \leqslant \rho_{ij} \leqslant 1. \tag{2.2.20b}$$

2.2.2 信息状态

设随机变量 $\boldsymbol{x}$ 有概率密度函数 $f(\boldsymbol{x})$, $f(\boldsymbol{x})$ 反映了对 $\boldsymbol{x}$ 的了解程度,或反映了 $\boldsymbol{x}$ 的信息,所对应的物理状态称为**信息状态**.

1) 全知信息状态

如果对 $\boldsymbol{x}$ 的真值 $\boldsymbol{x} = \boldsymbol{x}_0$ 有确切的了解,则此种物理状态称为**全知信息状态**.其相应的概率密度函数为

$$f(\boldsymbol{x}) = \delta(\boldsymbol{x} - \boldsymbol{x}_0).$$

2) 非信息状态

非信息状态又可称为参考信息状态. $\boldsymbol{x}$ 的所有的信息状态中,在某种意义下的最低信息状态或反映的信息量最小的信息状态称为**非信息状态**或**参考信息状态**.其概率密度函数通常用 $\mu(\boldsymbol{x})$ 来表示, $\mu(\boldsymbol{x})$ 称为**非信息概率密度函数**或**参考信息概率密度函数**.相应的概率记为 $\mu(\boldsymbol{A})$,即对任一子集 $\boldsymbol{A} \subset \boldsymbol{X}$,有

$$\mu(\boldsymbol{A}) = \int_{\boldsymbol{A}} \mu(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}.$$

如果参数集仅包含独立定义的参数,则非信息联合概率密度 $\mu(\boldsymbol{x})$ 满足

$$\mu(\boldsymbol{x}) = \prod_i \mu_i(x_i), \qquad i \in I_x,$$

其中　$\mu_i(x_i)$ 为参数 x_i 的非信息概率密度函数.

例:假设要估计事件在空间发生的位置,则非信息状态为事件在空间任何一点发生的可能性是均等的,即在空间任何一个相等的体积内事件发生的概率是相等的,即有

$$P(V) = \iiint_V \mu(x,y,z)\mathrm{d}x\mathrm{d}y\mathrm{d}z$$

正比于体积 V,其中 $\mu(x,y,z)$ 为非信息概率密度,在笛卡儿坐标系中非信息概率密度 $\mu(x,y,z)$ 为常数.但在球坐标系中

$$P(V) = \iiint_V \mu^*(r,\theta,\varphi)\mathrm{d}r\mathrm{d}\theta\mathrm{d}\varphi$$

$$= \iiint_V cr^2\sin\theta\mathrm{d}r\mathrm{d}\theta\mathrm{d}\varphi,$$

其中　c 为常数;非信息概率密度函数

$$\mu^*(r,\theta,\varphi) = cr^2\sin\theta$$

不再是常数.

3) 信息量

信息量是信息状态所含信息多少的一种量度.概率密度函数 $f(\boldsymbol{x})$ 反映了对参数 $\boldsymbol{x}$ 的所含有的信息,显然 $f(\boldsymbol{x})$ 越接近 δ 函数,对参数 $\boldsymbol{x}$ 所含有的信息越多.设有两个归一化的概率密度函数 $f_1(\boldsymbol{x})$ 和 $f_2(\boldsymbol{x})$,则 $f_1(\boldsymbol{x})$ 相对于 $f_2(\boldsymbol{x})$ 的**相对信息量**可定义为

$$I(f_1,f_2) = \int f_1(\boldsymbol{x})\log_a\left(\frac{f_1(\boldsymbol{x})}{f_2(\boldsymbol{x})}\right)\mathrm{d}\boldsymbol{x}. \tag{2.2.21}$$

当对数 $\log_a$ 以 2 为底时,信息量的单位为**比特**(bit),当对数 $\log_a$ 以 e 为底时,信息量的单位为**奈培**(nep),当对数 $\log_a$ 以 10 为底时,信息量的单位为数字(digit),概率密度函数 $f(\boldsymbol{x})$ 相对于非信息概率密度函数 $\mu(\boldsymbol{x})$ 的相对信息量

$$I(f,\mu) = \int f(\boldsymbol{x})\log_a\left(\frac{f(\boldsymbol{x})}{\mu(\boldsymbol{x})}\right)\mathrm{d}\boldsymbol{x} \tag{2.2.22}$$

称为 $f(\boldsymbol{x})$ 的信息量.

可证信息量总是正值,即

$$I(f,\mu) \geqslant 0. \tag{2.2.23}$$

证:当给定 $\mu(\boldsymbol{x})$,求约束条件$\int f(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = 1$ 下,

$$S(f) = \int f(\boldsymbol{x})\log_a\left(\frac{f(\boldsymbol{x})}{\mu(\boldsymbol{x})}\right)\mathrm{d}\boldsymbol{x} \tag{2.2.24}$$

的最小值.利用拉格朗日乘子法,引入泛函

$$S(f,\lambda) = \int f(\boldsymbol{x})\log_a\left(\frac{f(\boldsymbol{x})}{\mu(\boldsymbol{x})}\right)\mathrm{d}\boldsymbol{x} - \lambda\left[1 - \int f(\boldsymbol{x})\mathrm{d}\boldsymbol{x}\right].$$

对任意 δf,定义泛函的微商$\dfrac{\partial S}{\partial f}$ 为

$$S(f+\delta f,\lambda) - S(f,\lambda) = \frac{\partial S}{\partial f}\delta f + O(\|\delta f\|^2).$$