非线性统计预报方法及其应用

云南省学术著作
出版基金资助项目

严华生 谢应齐 曹杰 著

FEIXIANXING TONGJI
YUBAO FANGFA JIQI YINGYONG

云 南 科 技 出 版 社

云南省学术著作出版基金资助项目

非线性统计预报方法及其应用

严华生　谢应齐　曹　杰　著

云 南 科 技 出 版 社

责任编辑:肖　娅
封面设计:周　文

非线性统计预报方法及其应用
严华生　谢应齐　曹　杰　著

云南科技出版社出版发行(昆明市书林街 100 号)
昆明新星印刷厂印装

开本:787×1092　1/16　印张:8.5　字数:20 万
1998 年 7 月第 1 版　1998 年 7 月第 1 次印刷
印数:1-700 册

ISBN7-5416-1170-0/O·47　定价:15.00 元
若发现印装错误请与承印厂联系

前 言

统计预报方法早已在气象、地质、地震、农业、自动控制、经济预测等许多学科领域中有了很广泛的应用。最近几年来，在统计方法和应用两个方面又都有了新的进展。就统计学方法应用研究而言，它的广泛应用及与其他学科的结合，推动了交叉学科的产生与发展。例如动力统计预报的研究、海气相互作用的统计分析研究、地震与洪灾的遥相关研究等都有了一批成果。仅从气象统计预报来看，其内容已涉及到气候可预报性估算、预报区域划分、预报因子筛选、预报方法集成、预报效果评估、气候灾害统计分析、大气环境质量的区域划分、卫星云图资料统计处理、人工影响天气统计分析等。就统计学理论与方法的研究而言，也在统计预报模型、时间序列分析、统计参数检验方法、极值统计推断与设计等方面有了可喜的进展。

我们知道，自然界大量存在的相互作用是非线性的，线性作用其实只不过是非线性相互作用在一定条件下的近似。因而，很多学科领域的研究中，从“线性”向“非线性”的过度与发展，是研究深化的必然趋势。特别是近 20 年来，作为正在迅速发展的基础研究领域—非线性科学，在探求非线性现象的普遍规律和发展处理它的普适方法两方面已取得了明显的成就。这一切也必然对统计预报的理论和方法产生了强烈的冲击和影响。显然，过程本质的非线性，就决定了预报模型的非线性。所以，从线性统计分析向非线性统计分析发展，从而逐步建立非线性统计预报方法，已逐渐得到了较多的关注，并已成为统计学方法发展的方向之一。近些年来，作者顺应这一发展趋势，在发展统计预报方法及其在大气科学中的应用做了一系列较为系统的工作，取得了一些成果。这就为本书的出版提供了前提和条件。

虽然非线性统计预报这些年来有了较大发展，但却缺少系统介绍的专著。因此，作者将自己的研究成果及应用实例加以整理，编成此书奉献给广大读者，以期填补这方面的空白，为推动我国非线性统计预报理论和方法研究的不断发展，尽自己微薄之力。全书共十一章，第一至第七章是方法的介绍，对参数计算基本方法，统计性质评估，一个自变量时的统计预报模型，拟线性、半线性和曲线性预报模式，分段逼近模式，插值、距离、相似预报方法等作了系统介绍。后四章则是各种非线性统计预报模型在气象预报中的具体应用实例。这些应用均是从作者近几年研究工作中挑选出来的。

本书可作为高等院校大气、海洋、地质、地震、经济等相关专业的研究生或高年级本科生的教材或参考书，也可供从事涉及统计分析、统计预报、数据处理等研究与实际业务工作的科技人员学习参考。由于作者水平有限，错误和不当之处在所难免，敬请读者提出批评指正。

最后，作者还要感谢中国科学院大气物理研究所研究员周家斌先生、北京大学地球物理系教授黄嘉佑先生对本书出版给予的大力支持和帮助。

作　者

1998 年 6 月于云南大学

目　　录

第一章 概 论

§1.1 问题的提出

非线性统计预报问题在实际统计预报中是常见的。

例1. 某地降雨量多少与农作物产量有密切关系，因而可用降雨量来预报农作物产量。但预报实践发现，当降雨量比正常年份偏少很多、出现干旱时，常引起减产；当降雨量在正常年景范围内时，对农作物生长有利而增产；当降雨量比正常年份偏多很多，以至出现洪涝时，又会引起农作物减产。于是雨量与农作物产量的关系呈非线性关系，如下图(1-1)。

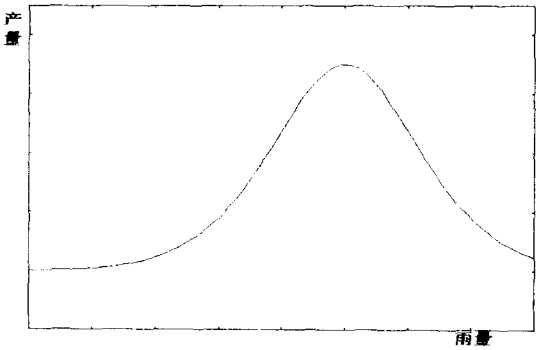

图1-1

若我们把雨量作为预报因子，用线性回归来预报产量，则难以正确反映雨量与产量之间的关系，常造成预报失败。

例2. 在气象统计预报中，常听到预报员说，统计预报在正常气候年景预报效果还可以，但遇到转折性、异常性情况，往往预报失败。究其原因，我们认为，这是因为目前气象统计预报所使用的绝大多数统计模型为线性模型，这样的模型一是把有些预报因子和预报对象的非线性关系当作线性关系来处理，二是模型本身难以筛选出与预报对象非线性关系密切的预报因子。另一个重要原因是预报关系随时间变化。第三个重要原因是一些强影响因子的异常变化使预报关系产生突变跳跃等现象。这些都是非线性统计预报要解决的问题。

大量事实表明，非线性是一切复杂现象的本质，是一切物质运动的普遍规律，非线性系统已逐渐成为自然科学的主要研究对象。可以毫不夸张地说，在分析高度非线性系统总体的统计性质所用的某些最先进的方法往往是在大气科学领域里产生并发展起来的。大气运动是一个具有耗散结构特征的非线性相互作用的极其复杂的动力时变系统。当代大气动力学正在从线性理论走向非线性理论。目前国内外气象工作者用非线性理论方法解释指数循环、阻塞形势、环流突变、气候变迁及跃变，以及讨论大气运动的可预报性等方面，已取得一系列成果。

纵观统计预报在农业、气象、经济、地震、地质等各行各业的应用情况，我们可以说，自然界的现象往往是很复杂的，变量间的关系不可能仅仅只是简单的线性关系，因而线性统计预报没有能力解决实践中遇到的各种复杂问题，于是很有必要发展非线性统计预报。

§1.2　非线性统计预报模型概述

1.2.1 线性和非线性模型之区别

把线性模型与非线性模型相比，尤如空间直线和平面与无数千变万化形状的曲线和曲面相比，非线性模型数量之多和形状之复杂，绝对不可能像线性模型那样，可以用一种模型形式完全表达出来。因而，在学习研究非线性模型时，我们绝不可把它想像成线性模型那样，以为模型只有一种形式，即线性形式。还有另一种不正确的想法，就是以为非线性模型仅是几种有限的基本初等函数的各种代数运算的组合，在具体运算时，无非是把自变量作一些初等变换，化为线性回归的形式即可。但实际上，很多曲线关系几乎不能用初等函数来表示。我们知道，曲线和曲面的类型无穷多，因而非线性统计模型几乎可以说有无穷多种类型。在很多情况下，是难以用几种简单的初等函数的复合把它表示出来的。有时即使能表示出来，形式也较为复杂，很难把它通过变量替换化为线性回归的形式。

可以说，实际工作中遇到的统计预报问题绝大多数是非线性的，只不过人们常用线性统计模型来近似地研究它。在非线性统计预报中，我们首先遇到的是模型的选择问题，然后才是模型的拟合和预报问题。下面我们先来看看非线性统计模型的特点：模型中含有非线性待估参数。

在一个统计模型中，若待估模型参数与自变量在线性空间中呈线性表示，则称之为线性参数，记为 $\mathbf{c}$ 。它可以用线性最小二乘法来估计，否则称之为非线性参数，记为 $\mathbf{d}$ 。一般说来，$\mathbf{d}$ 难以用线性最小二乘法来估计。例如：设非线性模型为：

$$y = f(x,c,d) = c_0 + c_1(x-d_1)^{d_2} \tag{1-1}$$

待估参数：$\mathbf{c} = (c_0, c_1), \mathbf{d} = (d_1, d_2)$ ，可以验证，$\mathbf{d}$ 不能用线性统计中的最小二乘法来估计，只有当用某种非线性参数估计方法估计出 $\mathbf{d}$ 后，令 $x' = (x-d_1)^{d_2}$ ，才能将模型1-1化为：

$$y = c_0 + c_1 x' \tag{1-2}$$

此时才可用线性最小二乘法来估计 $\mathbf{c}$ 。

根据以上讨论可知，线性模型与非线性模型的区别之一是在线性模型中，所有参数都是自变量的线性函数，而在非线性模型中，至少存在一个待估参数不是自变量的线性函数，它们不能用线性最小二乘法来估计；区别之二是在线性模型中，参数与自变量是一一对应的，例如 n 元线性回归方程：

$$y = \sum_{i=1}^{n} b_i x_i \tag{1-3}$$

参数 b_i 对应自变量 x_i ，而在非线性模型中，参数与自变量可能不是一一对应关系，例如非线性模型：

$$y = \frac{1}{(c_1 x_1 + c_2 x_2)^d} \tag{1-4}$$

参数 d 是对整个模型的所有自变量而言的，不是对应某个自变量的；区别之三是线性模

型是连续函数，而非线性模型却不一定具备这个性质；区别之四是线性模型只用一个线性函数表示即可，而在非线性模型中有可能是多个函数组成，有时甚至难以用函数形式来表达。

1.2.2 曲线拟合与非线性统计预报之区别

从所得到的变量的离散数据中，构造一些拟合函数去逼近数据点，称之为曲线拟合问题。由此可见，曲线拟合的基本方法就是在所得到的变量观测值区间范围内，求出拟合函数。适合用曲线拟合方法解决的问题是：自变量与因变量之间已存在确定的物理联系，因、自变量观测值所反映出来的变化规律为较精确的一一对应变化，因而曲线拟合要求在所观测到的变量变化区间内，拟合函数能尽最大可能逼近数据点，而不考虑区间外推，它属于计算数学学科中逼近论的范畴。

非线性统计预报则假设预报模式中的每一个自变量与因变量都是随机变量，而且它们之间存在显著的统计联系，即对应一个变量的观测值，可能存在多个因变量的不同值，在给定自变量观测值的条件下，因变量的条件数学期望随自变量的变化而变化。因而非线性统计预报所要解决的问题就是：在给定模型、变量及参数的概率分布置信度的前提下，最大可能地拟合出因变量随自变量变化的统计规律。它所适合的问题是：因、自变量间的物理联系不清，假设均为随机变量，于是根据自变量的变化，统计推断因变量条件数学期望的变化，非线性统计预报属于统计数学范畴。

总的说来，它们之间的区别是：

(1) 对象不同。曲线拟合的对象是变量间有必然的物理联系，观测值足够准确地反映这种联系，而非线性统计预报的对象则假设为随机变量，分析研究其统计规律。

(2) 目标不一样。曲线拟合的目标是使在观测区间内的拟合误差足够小，而非线性统计预报的目标则是要求所反映出来的统计规律的概率尽可能地大。

上面，我们简单介绍了曲线拟合与非线性统计预报的差别，但是在非线性统计预报中，我们也常常吸取曲线拟合中一些对非线性统计预报有意义的思路和方法，加以改造运用。例如第四章中的样条回归，第七章中的Shepard插值等。若读者熟悉曲线拟合方面的知识，这对研究非线性统计预报，拓宽思路是很有益的。

1.2.3 非线性统计预报模型分类

根据非线性统计预报模型的性质，对其进行分类：

(1) 拟线性模式类

$$y = d_1 x + d_2 \ln x + d_3 e^x + d_4 x^2 + \cdots + d_k f(x) \tag{1-5}$$

特点：模式由自变量的各种初等函数的线性组合而成，只要作适当变量替换，实质上就不含非线性参数。

事实上，令 $x_1 = x$，$x_2 = \ln x$，$x_3 = e^x, \cdots, x_k = f(x)$ 则可把(1-5)式化为多元线性回归模式求解。

(2) 半线性模式类

$$y = f(x,c,d) = \sum_{i=0}^{r} c_i \cdot f_i(x,d_i) \tag{1-6}$$

特点：可化为r个含有非线性待估参数的非线性函数的线性组合，例如：

$$y = c_1 d_1^{\ x} + c_2 d_2^{x} + \cdots + c_r d_r^{x} \tag{1-7}$$

$$y = c_1 x^{d_1} + c_2 x^{d_2} + \cdots + c_r x^{d_r} \tag{1-8}$$

就属于这一类，此时首先用非线性参数计算方法求得$\mathbf{d}$后，再作：$x_i = \delta_i(x, d_i)$的变换，就可化为多元线性回归模型求解参数$\mathbf{c}$。

(3) 其它非线性模式类

$$y = f(x, c, d) \tag{1-9}$$

特点：所有不能视为前两类的非线性模型都可归为这一类，此类模式显然不能简单化为线性模式用线性最小二乘法求解。对这类模式，还可细分为复杂函数类、分段函数类、无函数形式类等等来研究，从第五章起，将详细讨论这三类非线性模式。

由于非线性函数形式无穷多，当我们用非线性统计预报方法来解决一个实际预报问题时，常见的情况是我们根本不知道预报因子与预报对象是何种非线性函数关系，因此从统计预报应用问题和计算角度来对非线性统计预报进行划分，还可以分为如下两类：

(1) 预报对象与预报因子的函数关系模型已知时的非线性统计预报；

(2) 预报对象与预报因子间函数关系模型未知时的非线性统计预报。

若根据目前对非线性统计预报模式的分析研究方法来划分，又可分为如下三类：

(1) 非线性函数型模式类；

(2) 逐段线性化逼近模式类；

(3) 非函数模式类。

§1.3 非线性模型参数估计性质简述

在模型参数估计性质方面，非线性模型与线性模型的重大差别在于：在规定独立、同分布、正态随机项的通常假设后，线性模型产生无偏的、正态分布的、极小方差的参数估计，但非线性模型参数估计却没有这三个性质，而是产生有偏的、非正态的、超过极小方差的偏差估计。

就拿最简单的拟线性模型来说，当对自变量x进行变量替换，换为x'时，已是非线性变换，于是对x'进行线性回归后所得到的回归系数与原变量x之间的回归系数，一般说来是有差别的。并且若x服从正态分布，则通过非线性变换后的x'不一定服从正态分布，相应所产生的回归系数的分布也就不同。如果x'与x间的非线性变换不是单调函数特性的，则参数估计还将进一步丧失最小方差性和无偏性。对y与x'的模型拟合残差和y与x的模型拟合残差更不能相提并论。把x非线性变换变为x'，使变量的变化量级和变化区间发生很大变化，于是对模型拟合残差产生影响。若对因变量y也作变换，那么模型估计性质的变化还要大得多。例如经常用到的幂函数模型$y = ax^b$，为了用线性最小二乘法求解各参数，需要将其线性化，为此取对数，有：$\ln y = \ln a + b\ln x$，令$y' = \ln y$，$a' = \ln a$，$x' = \ln x$，便有$y' = a' + bx'$，其中$a' = \overline{y}' - b\overline{x}'$，式中$\overline{y}' = \dfrac{1}{n}\sum_{i=1}^{n}\ln y$，

$\bar{x}' = \frac{1}{n}\sum_{i=1}^{n}\ln x$。很明显，$\bar{y}'$ 的反对数值(即真值的几何平均数)不等于各真数值 y_i 的算术平均数 $\bar{y}$，其间就存在一个对数改正系数 $f_c = \exp(\frac{1}{2}\delta^2)$，($\delta^2$ 为对数计算的回归剩余方差)。因此，对于非线性模型 $y = f(x,d)$，由于 d 不是 y 的线性组合，所以 $\hat{d}$ 不服从正态分布，并且 $\hat{d}$ 不是 d 的无偏估计和最小方差估计。

非线性模型中的参数包括线性参数和非线性参数，并且至少含有一个非线性参数。一般地说，对非线性模型中所有参数的估计，不论是线性的还是非线性的。都是有偏的。

迄今为止，在估计非线性模型参数的方法研究中，还没有找到具有像估计线性模型参数那样估计性质优良的方法。这是因为在理论上，非线性问题比线性问题难处理得多，即使是估计参数，要得到一个简单的表达通式是不容易的。

但是，由于对线性模型及其参数估计已研究得较为清楚，于是人们总是试图借用线性模型已有的理论成果。根据目前研究，在下列情况下，非线性模型参数估计的性质可接近无偏性、正态性和最小方差性这三个优良性质：

(1) 当样本量充分大时，以往已有人证明，当样本量趋于无穷时，对非线性模型的参数估计趋近于线性模型参数估计的优良性质。

(2) 尽可能地选择与实际应用问题最相适应的非线性模型，以使模型拟合残差方差达到最小。这是因为最小二乘估计的偏差与拟合残差方差成正比，精心挑选模型和调整参数使所得残差方差越小，则它就越接近线性性态。

非线性模型参数估计偏差幅度大小、非正态的程度、超出最小方差界的多少，将因不同的非线性模型而异，不同模型间差别很大。如果非线性模型参数估计仅仅是略为有偏，其分布接近正态分布，并且方差仅少许超过最小方差界，那么由于估计量的性质接近线性模型的性质，于是可用线性模型的理论来近似地研究它。如果参数估计是相当有偏的，其分布远非正态分布，并且方差大大超过最小方差界，这就很明显是非线性模型的性态，与线性模型相差甚远，不能用线性模型的方法来研究。第三章将专门讨论非线性统计问题。

§1.4 非线性统计预报模型建立及参数计算方法概述

1.4.1 非线性统计预报模型建立概述

当给定一个实际预报问题后，首先遇到的是统计预报模型选择问题，这可分为两种情况来讨论：

第一种情况是，设预报模型已知，则预报员主要考虑计算模型参数、模型拟合和预报。例如已知实际问题的物理模型，根据观测数据来求解统计参数的动力—统计预报。把预报模型设为拟线性、半线性、曲线型模型，用计算机根据拟合残差自动挑选模型函数的自适应拟合。

第二种情况是，当给定一个实际预报问题，但我们并不知道预报因子与预报对象之间究竟是什么预报关系，设想它是比较复杂的非线性关系，甚至预报因子与预报对象间的函

数关系都不是连续函数、有突变、跳跃等间断点，且预报关系随不同情况还会发生变化。对这种情况的统计预报处理思路有两个，一个是把预报对象分割为k个小区间(可以有交叉重叠区间)，在每个小区间内用一个线性模式来逼近它，这就是分段线性化的思想；另一个是在预报关系函数模型难以建立的情况下，可直接根据预报因子与预报对象间的量值关系，采用插值、距离、相似等方法来预报，这就是非模型化的思路。

1.4.2 非线性模型参数计算概述

非线性模型参数计算分为模型已知和未知两种情况，在模型已知时求参数的方法有非线性最小二乘法、参数寻优等方法；在模型未知时有分段线性最小二乘逼近、非模式化方法等。现分别简介如下：

(1)非线性最小二乘法(高斯—牛顿—麦夸脱法)

设已给定一个非线性模型的预报因子、预报对象样本资料，要计算模型参数，现用非线性最小二乘法求解。它的思路为：设模型函数为$y=f(x,d)$，则首先给参数d赋一个初值，记为$d^{(0)}$，设其与真实值之差为$\Delta=d-d^{(0)}$对模型函数$f(x,d)$在$d^{(0)}$附近作Taylor级数展开，略去二次以上的项得：

$$f\left(x,d^{(0)}+\Delta\right)\approx f\left(x,d^{(0)}\right)+\frac{\partial f}{\partial d}\Delta \tag{1-10}$$

然后用线性最小二乘法确定出Δ，把它作为参数初值$d^{(0)}$的修正项，对$d^{(0)}$进行修正。记为：

$$d^{(1)}=d^{(0)}+\Delta \tag{1-11}$$

经过修正后的$d^{(1)}$比$d^{(0)}$更加接近d的真值，以把$d^{(1)}$作为初值，重复上述过程，循环叠代以逐步逼近参数d的真值。

这一方法的效果好坏关键在于经过多次反复叠代后是否收敛于非线性参数的真值以及收敛速度是否快。麦夸脱从算法上对高斯—牛顿法作了改进，以加快收敛速度，关于它的详细讨论，详见第二章。

(2)寻优法

设已知非线性模型函数及样本资料，给非线性参数赋以不同的值，比较模型拟合残差，然后不断调整非线性参数赋值，优选使模型拟合残差达到最小的非线性参数赋值。详见第二章。

(3)分段线性最小二乘逼近

若预报模型未知，则把预报对象分为几段，每一段用线性最小二乘法来逼近。第六章将详细介绍这方面的内容，它可以处理突变、跳跃等不连续现象及预报关系随时间变化的问题。

(4)非模型化方法

不去建立函数预报模型，而是应用距离、相似、插值方法对已给定的预报因子外推预报实时资料与历史预报因子样本资料逐个进行分析比较计算来作预报，由于没有待估模型参数，故称为非模型化方法，第七章将介绍这方面的内容。

第二章　模型已知时的参数计算基本方法

§2.1　概　述

若预报因子和预报对象分别取n个样本，共有m个预报因子，则可用矩阵形式来表示样本资料，记为$\underset{n\times 1}{\mathbf{Y}}$，$\underset{n\times m}{\mathbf{X}}$。设非线性模型为：

$$\mathbf{Y}=f\left(\mathbf{X},\mathbf{c},\mathbf{d}\right) \tag{2-1}$$

为了今后参数计算叙述方便，有时也把上式简记为：

$$\mathbf{Y}=f\left(\mathbf{X},\mathbf{d}\right) \tag{2-2}$$

其中，$\mathbf{X}$为向量，它包括m个分量，即$\mathbf{X}=\left(x_1,\cdots,x_m\right)$；$\mathbf{d}$为待估模型参数向量，设有$k$个待估参数，记为$\mathbf{d}=\left(d_1,\cdots,d_k\right)$，并且$\mathbf{d}$既包括非线性参数，也包括线性参数。当计算出参数估值$\hat{\mathbf{d}}$后，记模型拟合残差为：

$$E\left(\hat{\mathbf{d}}\right)=\left\|\mathbf{Y}-f\left(\mathbf{X},\hat{\mathbf{d}}\right)\right\|=\sum_{i=1}^{n}\left(y_i-f\left(x_i,\mathbf{d}\right)\right)^2 \tag{2-3}$$

即残差大小是模型估计参数$\hat{\mathbf{d}}$的函数。

参数计算的基本要求是：要找到这样的$\hat{\mathbf{d}}$，使残差函数$E\left(\hat{\mathbf{d}}\right)$取极小值。根据微积分原理，函数$E\left(\hat{\mathbf{d}}\right)$取得极小值的条件是：

必要条件：

$$\frac{\partial E}{\partial d_1}=\frac{\partial E}{\partial d_2}=\cdots=\frac{\partial E}{\partial d_k}=0 \tag{2-4}$$

或梯度：

$$grad\quad \Delta E(\mathbf{d})=(\frac{\partial E}{\partial d_1},\frac{\partial E}{\partial d_2},\cdots,\frac{\partial E}{\partial d_k})=0 \tag{2-5}$$

充分条件：

我们称函数$E(\mathbf{d})$的所有二阶偏导数组成的下列矩阵称为海赛矩阵，记为$H(\mathbf{d})$：

$$H(\mathbf{d})=\begin{bmatrix} \frac{\partial^2 E}{\partial d_1^2} & \frac{\partial^2 E}{\partial d_1 \partial d_2} & \cdots & \frac{\partial^2 E}{\partial d_1 \partial d_k} \\ \frac{\partial^2 E}{\partial f_2 \partial d_1} & \frac{\partial^2 E}{\partial d_2^2} & \cdots & \frac{\partial^2 E}{\partial d_2 \partial d_k} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial^2 E}{\partial d_k \partial d_1} & \frac{\partial^2 E}{\partial d_k \partial d_2} & \cdots & \frac{\partial^2 E}{\partial d_k^2} \end{bmatrix} \tag{2-6}$$

则当$E\left(\hat{\mathbf{d}}\right)$是极小值时，必有海赛(Hessen)矩阵正定。

若函数只有一个极小值点，称之为单谷函数，如果$E(\mathbf{d})$为单谷函数，则所求极值点为全局极小值，否则为局部极小值。

于是，从理论上说，参数计算的基本思路为：对n个统计样本，按下式求残差平方和：

$$E(\mathbf{d})=\sum_{i=1}^{n}\left(y_i-f\left(x_i,\mathbf{d}\right)\right)^2 \tag{2-7}$$

然后对$E(\mathbf{d})$求梯度，令其为零。解得$\hat{\mathbf{d}}$，再把$E\left(\hat{\mathbf{d}}\right)$的二阶导数值代入海赛矩阵进行判别，求得所有极小值点，从所有极小值中，找出全局最小值作为$\mathbf{d}$的计算值$\hat{\mathbf{d}}$，它就是使残差达到最小的估计参数。

但是，一般说来，当统计模型为非线性模型时，按上述求解过程形成的求解参数$\hat{\mathbf{d}}$的方程组是非线性的，解起来相当困难，有时甚至无法求解，因而在实际工作中很难使用这种求解参数的直接方法，只能考虑其它计算参数的方法，于是就产生了逐步逼近的叠代算法。它的基本思路是这样的：从给定$\mathbf{d}$的一个初值$\mathbf{d}^{(0)}$出发，先找到能使拟合残差小一点的$\mathbf{d}^{(1)}$，然后再进一步寻找能使拟合残差更小一点的$\mathbf{d}^{(2)}$，如此不断进行下去，最终希望能达到或接近残差极小点，这就是本章所要讨论的非线性模型参数计算的基本思路。从中可以看出，不论是线性还是非线性模型参数的估计准则都是一样的，即要找到这样的估计参数$\hat{\mathbf{d}}$，使模型拟合残差达到最小。

线性模型和非线性模型参数估计的区别在于计算参数所使用的方法不同，由于线性模型比较简单，在计算参数时，要满足使残差达到最小这一条件，用线性最小二乘算法就可以了；而非线性模型有时比较复杂，难以用线性最小二乘求解参数，所以对非线性模型的参数计算是采用给参数赋初值，不断比较试验，进行叠代以逐步逼近最小拟合残差这一反复循环过程来完成的。

这里还要指出：在模型参数计算中，拟合残差是估计参数的函数，记为$E=E(\hat{\mathbf{d}})$。于是可以证明，只有当其是单谷函数时，才能保证用叠代算法计算的参数$\hat{\mathbf{d}}$具有使残差函数收敛于最小的性质。若残差函数不是单谷函数，则用叠代算法难以求得残差最小点，这时可采用本章后面将要介绍的参数数值试验法来求得参数估值。理论上判别残差函数是否为单谷函数，可采用二阶导数海赛矩阵判别法。

本章所讨论的非线性模型参数计算的逐步叠代逼近算法的特点是：每次叠代使残差有所下降。这里的关键问题是要根据拟合残差的某些信息，确定一个由$\mathbf{d}^{(k)}$到下一点$\mathbf{d}^{(k+1)}$的规则，把这个规则记为：

$$\mathbf{d}^{(k+1)}=\alpha(\mathbf{d}^{(k)}) \tag{2-8}$$

于是可以构造出如下算法：

(1)取初始点$\mathbf{d}^{(0)}$，置$k=0$；

(2)置$\mathbf{d}^{(k+1)}=\alpha(\mathbf{d}^{(k)})$；

(3)若$\mathbf{d}^{(k+1)}\geq\alpha(\mathbf{d}^{(k)})$，就停止计算，否则令$k=k+1$，转(2)；

对有多个模型参数时：

(1)取初始点$d_1^{(0)}$，$d_2^{(0)},\cdots,d_l^{(0)}$，记$k=0$；

(2)选择适当的搜索方向；

(3)在此方向上取$d_1^{(k+1)}$，$d_2^{(k+1)},\cdots,d_l^{(k+1)}$，计算$E(d_1^{(k+1)}$，$d_2^{(k+1)},\cdots,d_l^{(k+1)})$；

(4) 若 $E(d_1^{(k+1)}$，$d_2^{(k+1)},\cdots,d_l^{(k+1)}) \geq E(d_1^{(k)}$，$d_2^{(k)},\cdots,d_l^{(k)})$，就停止，否则记$k=k+1$，转(2)。

显然，对有多参数的寻优问题，关键在于如何选择搜索方向。这里有如下计算实用收敛准则：

在许多统计预报计算的实际问题中，由于随机误差、资料误差、模型误差的存在，往往不可能完全精确地找到拟合残差极小点，所以只要达到一定精度就行了，于是可以考虑用下列较松的条件作为收敛准则。在给定误差限ε后，有：

(1)拟合残差值的下降量充分小，即：

$$\left\| E(\mathbf{d}^{(k)}) - E(\mathbf{d}^{(k+1)}) \right\| < \varepsilon \tag{2-9}$$

或者用：

$$\left\| \frac{E(\mathbf{d}^{(k)}) - E(\mathbf{d}^{(k+1)})}{E(\mathbf{d}^{(k)})} \right\| < \varepsilon \tag{2-10}$$

(2)计算参数的改变量充分小，即：

$$\left\| \mathbf{d}^{(k)} - \mathbf{d}^{(k+1)} \right\| < \varepsilon \tag{2-11}$$

或者用：

$$\left\| \frac{\mathbf{d}^{(k)} - \mathbf{d}^{(k+1)}}{\mathbf{d}^{(k)}} \right\| < \varepsilon \tag{2-12}$$

对于半线性模型，即含有线性参数$\mathbf{c}$，又含有非线性参数$\mathbf{d}$的模型，它可以表示为一些非线性函数的线性组合，记为：

$$y = \sum_{i=0}^{r} c_i f_i(\mathbf{X}, \mathbf{d}_i) \tag{2-13}$$

对这类模型的参数计算基本方法为：对$f_i(\mathbf{X},\mathbf{d}_i)$，给$\mathbf{d}_i$赋初值，然后作变量替换：$\eta_1 = f_1(\mathbf{X},\mathbf{d}_1^{(k)})$，$\eta_2 = f_2(\mathbf{X},\mathbf{d}_2^{(k)}),\cdots,\eta_r = f_r(\mathbf{X},\mathbf{d}_r^{(k)})$，于是模型(2-13)化为以$\eta$为自变量的线性模型：

$$y = \sum_{i=0}^{r} c_i \eta_i \tag{2-14}$$

用线性最小二乘法求解(2-14)式的线性回归方程。记模型拟合残差仍然为$E\left(\hat{\mathbf{c}},\hat{\mathbf{d}}\right)$，应用多变量函数寻优法对$\mathbf{d}$取不同值，求得不同的$E\left(\hat{\mathbf{c}},\hat{\mathbf{d}}\right)$，进行比较，按$E\left(\hat{\mathbf{c}},\hat{\mathbf{d}}\right)$减少的方向逐步逼近$\mathbf{c}$，$\mathbf{d}$的估计值。例如对第一章(1-1)式所讨论的例子，应用下面将要介

绍的寻优法，对$\mathbf{d}$取不同值，然后用线性最小二乘法求线性参数$\mathbf{c}$，得出残差函数$E(\hat{\mathbf{c}},\hat{\mathbf{d}})$进行比较，使之达到最小。

§2.2 一维搜索

一维搜索主要用于只有一个非线性待估参数(可以有多个线性待估参数)的情形，它的基本做法是:给非线性待估参数赋一个初值，用线性最小二乘法估计线性参数，求得拟合残差，按照函数寻优方法，寻找使残差达到最小的非线性参数最优点。这里关键是非线性参数的寻优办法，下面我们就来讨论常用的黄金分割法和抛物线法。

2.2.1 黄金分割法

设有数列$\{F_n\}$，$n=0,\ 1,\ \cdots$。它具有这样的性质，即从第三项起，每一项都是它前面两项之和：$F_0=0,F_1=1,\cdots,F_n=F_{n-1}+F_{n-2}$，这种数列称为斐波纳契数列。若我们要在区间$[a,\ b]$中安排一种试验方案，能迅速找到最优点。首先我们可以根据斐波纳契数列来安排试验，设j为余下待做的测试次数，$n-j$为已做过的测试次数，区间缩减迭代关系是:

$$F_{n-j}=F_{\eta-j-1}+F_{n-j-2} \tag{2-15}$$

这就是说，每一不定区间之长是其后顺次所得到的两个一定区间的长度之和。但我们事先并不知道要做几次试验，即n是多少?这里，我们建立一个逐次长度比的常数，记这一比例常数为τ，有:

$$\frac{F_{n-j}}{F_{n-j-1}}=\tau=\frac{F_{n-j+1}}{F_{n-j}} \tag{2-16}$$

推得:

$$\frac{F_{n-j}}{F_{n-j-1}}.\frac{F_{n-j+1}}{F_{n-j}}=\tau^2 \tag{2-17}$$

于是有:

$$\frac{F_{n-j+1}}{F_{n-j-1}}=\tau^2 \tag{2-18}$$

对(2-17)式两边都除以F_{n-j-2}得:

$$\frac{F_{n-j}}{F_{n-j-2}}=\frac{F_{n-j+1}}{F_{n-j-2}}+1 \tag{2-19}$$

运用(2-18)和(2-19)式得:

$$\tau^2-\tau-1=0 \tag{2-20}$$