# 相依线性回归模型的统计推断

胡宏昌　秦永松　黄收友　著

科学出版社

# 相依线性回归模型的统计推断

胡宏昌　秦永松　黄收友　著



科学出版社
北京

## 内 容 简 介

本书重点讨论相依（包括函数系数的自回归序列、一阶自回归序列、渐近几乎负相依、负超可加相依、正负相协误差、一般相依误差等）线性回归模型的极大似然估计（包括拟极大似然估计及 $L_q$ 极大似然估计）、M 估计和经验似然方法，也涉及与它们密切相关的模型和方法，如广义线性回归模型、部分线性回归模型和非线性回归模型以及 HD 估计和随机投影方法。

本书可供高等院校数学和统计学专业的研究生、教师及相关科研工作者阅读参考。

**图书在版编目（CIP）数据**

相依线性回归模型的统计推断/胡宏昌，秦永松，黄收友著. —北京：科学出版社，2017.11

ISBN 978-7-03-055623-3

Ⅰ.①相… Ⅱ.①胡… ②秦… ③黄… Ⅲ.①统计推断-研究 Ⅳ.①O212

中国版本图书馆 CIP 数据核字（2017）第 290761 号

责任编辑：吉正霞 王 晶/责任校对：董艳辉
责任印制：彭 超/封面设计：苏 波

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
武汉市首壹印务有限公司印刷
科学出版社发行 各地新华书店经销
*
开本：787×1092 1/16
2017 年 11 月第 一 版 印张：14
2017 年 11 月第一次印刷 字数：328 000

**定价：52.00 元**

（如有印装质量问题，我社负责调换）

# 前　言

线性回归模型是个古老而又非常重要的统计模型。独立误差的线性回归模型已经取得了非常多的研究成果，其相关理论已经基本成熟；但相依误差的线性回归模型还有很多值得探讨的研究课题。作者根据多年来对线性回归模型的教学和科研体会，对相依线性回归模型的研究成果进行了整理，汇总成本书。由于这一课题涉及领域太多，需要的知识很多，所以不可能涉及所有问题，而只重点讨论相依线性回归模型的(拟、$L_q$)极大似然估计、M 估计和经验似然方法，也涉及与它们密切相关的模型和方法，如广义线性回归模型、部分线性回归模型和非线性回归模型，以及 HD 估计和随机投影方法。

本书以相依线性回归模型为主线，重点讨论该类模型的极大似然估计、M 估计和经验似然方法。第 1 章由胡宏昌、秦永松执笔，第 2～3 章，第 5 章第 5.1 节、5.3 节由胡宏昌执笔，第 4 章、第 5 章 5.2 节由黄收友执笔，全书由胡宏昌负责统稿。

感谢所有从事线性回归模型及相关工作研究的同仁，正是他们深入研究的成果为作者提供了丰富的研究资料。特别是陈希孺院士、赵林城教授、孙海燕教授、M. J. Silvapulle 教授、R. A. Maller 教授、W. B. Wu 教授、D. Ferrari 教授和 Y. Yang 教授等。

感谢为本书提供宝贵建议和其他帮助支持的教师！感谢徐侃教授审阅全书，并进行了仔细修改！感谢宋蕾、夏雨荷、张宇、曾珍、余云彩等同学参与了本书的一些研究和计算工作。

本书得到了国家自然科学基金(No. 11471105，11471223)以及科学出版社的支持，在此一并感谢！

相依线性回归模型的研究涉及众多的内容，本书只是作者一孔之见，疏漏之处在所难免，恳请广大读者批评指正！

作　者

2017 年 5 月 25 日

# 目　录

# 第1章 绪　论

线性回归模型不仅是现代统计学中应用最为广泛的模型之一，而且是许多其他统计模型的研究或应用基础，主要原因如下(参见文献[1])：

第一，在现实世界中，许多变量之间具有线性或近似线性的依赖关系；

第二，尽管现实世界本质上是非线性的，但是许多变量经过适当变换后，新变量之间具有线性或近似线性的依赖关系；

第三，线性关系是数学中最基本的关系，相对比较容易处理.

本章主要介绍独立和相依误差的线性回归模型的研究进展，并简要介绍与线性回归模型密切相关的几种回归模型.

## 1.1 独立误差的线性回归模型

在人类活动的各种领域中，常常需要研究某些变量之间的关系. 要全面考察两个变量 $X$ 和 $Y$ 之间的相关关系，就要研究 $Y$ 的条件分布 $F(Y|X=x)$ 随 $X$ 取值 $x$ 的变化情况，这样做比较复杂，作为一个近似，我们可以考察分布 $F(Y|X=x)$ 的某个有代表性的数值，例如其期望值. 这个期望值当然与 $x$ 有关，记为 $f(x)$，称为 $Y$ 对 $X$ 的回归函数，而

$$y=f(x) \tag{1.1.1}$$

称为 $Y$ 对 $X$ 的回归方程. 对于身高和体重的例子，$f(x)$ 就是具有身高 $x$ 的所有人的平均体重. 回归方程是一个确定性的关系，而原变量 $X,Y$ 之间是相关关系. 回归方程的作用，正在于近似地代替这个相关关系.

在实际问题中，回归函数是不知道的，需要由试验或观察数据去估计. 假如做了 $n$ 次试验，得到 $n$ 组数据 $(x_i,y_i)$，$i=1,2,\cdots,n$. 根据回归函数的意义有

$$y=E(Y|X=x)=f(x). \tag{1.1.2}$$

可以写成

$$y_i=f(x_i)+e_i,\quad i=1,2,\cdots,n, \tag{1.1.3}$$

其中 $Ee_i=0$. 以后称 $X$ 为自变量，$Y$ 为因变量或响应变量.

首先将模型(1.1.3)中的自变量推广到 $p-1$ 个，即 $X_1,X_2,\cdots,X_{p-1}$，然后考虑其最简单的情形，即线性函数，得到如下线性回归模型

$$Y=\beta_0+\beta_1X_1+\cdots+\beta_{p-1}X_{p-1}+e, \tag{1.1.4}$$

其中 $e$ 为误差项，它表示除了 $X_1,X_2,\cdots,X_{p-1}$ 之外其他因素对 $Y$ 的影响以及试验或测量误差，$\beta_0,\beta_1,\cdots,\beta_{p-1}$ 是待估计的未知参数. 假定因变量 $Y$ 和自变量 $X_1,X_2,\cdots,X_{p-1}$ 的 $n$ 组观测值为 $(y_i;x_{i1},\cdots,x_{i,p-1})$，$i=1,2,\cdots,n$，它们满足

$$y_i=\beta_0+\beta_1x_{i1}+\cdots+\beta_{p-1}x_{i,p-1}+e_i,\quad i=1,2,\cdots,n, \tag{1.1.5}$$

其中误差常常假定满足 Gauss-Markov 假设，即

$$E(e_i)=0,\quad \mathrm{Var}(e_i)=\sigma^2,\quad \mathrm{Cov}(e_i,e_j)=0\quad (i\neq j). \tag{1.1.6}$$

应用适当的统计方法可以得到未知参数的估计值 $\hat{\beta}_0,\hat{\beta}_1,\cdots,\hat{\beta}_{p-1}$，将它们代入模型(1.1.4)，并略去误差项得到

$$\hat{Y}=\hat{\beta}_0+\hat{\beta}_1X_1+\cdots+\hat{\beta}_{p-1}X_{p-1}, \tag{1.1.7}$$

称为(经验) 回归方程.

若用矩阵形式表示，则模型(1.1.5) 变形为

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}=\begin{pmatrix} 1 & x_{11} & \cdots & x_{1,p-1} \\ 1 & x_{21} & \cdots & x_{2,p-1} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{n,p-1} \end{pmatrix}\begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix}+\begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

或

$$y=X\beta+e, \tag{1.1.8}$$

其中 $y$ 是 $n$ 维观测列向量，$X$ 为 $n\times p$ 的已知设计矩阵，$\beta$ 为 $p$ 维未知参数列向量，$e$ 为随机误差列向量. 用矩阵形式可将 Gauss-Markov 假设(1.1.6) 写成

$$E(e)=0,\quad \mathrm{Cov}(e)=\sigma^2 I_n. \tag{1.1.9}$$

模型(1.1.8) 和(1.1.9) 常常称为经典线性回归模型(参见文献[2]). 当然该模型的适当推广到 $X$ 为随机设计(参见文献[3]—[4] 等) 和误差项具有不等方差(研究成果参见文献[5]—[15] 等)，但仍保留误差独立性的假设. 可以说，独立误差的线性回归模型的研究已经非常完善，研究成果很多，如：最小二乘估计方法见文献[16]，M 估计方法见文献[17]，稳健估计方法见文献[18]，Bootstrap 方法见文献[19]，其他方法可参见文献[20]—[22] 等.

## 1.2 相依误差的线性回归模型

尽管独立误差的线性回归模型是一种最重要的统计模型，应用广泛，但独立性的假设在一些实际应用中遇到了麻烦. 因此，近年来相依误差线性回归模型的研究已成为学术界研究的热点课题.

误差项存在相关性，即 $\mathrm{Cov}(e_i,e_j)=0,i\neq j$ 的假设不成立. 误差 $\{e_i,i=1,2,\cdots,n\}$ 相关性的形式有很多，如鞅差[23]、各种混合序列[24-25](如 $\alpha,\beta,\rho,\varphi,\psi,\tilde{\rho},\tilde{\varphi}$ 负相关序列等) 以及它们的线性组合，甚至假定误差为自回归(autoregressive，AR)、滑动平均(moving average，MA)、自回归滑动平均(autoregressive-moving average，ARMA)、自回归求和滑动平均 (autoregressive integrated moving average，ARIMA)、自回归条件异方差 (autoregressive corditionally heteroscedastic，ARCH)、广义自回归条件异方差 (generalized autoregressive conditionally heteroscedastic，GARCH) 等时间序列(参见文献[24]) 以及它们的线性组合(参见文献[26])，还可以假定误差为分形时间序列(参见文献[27]). 尽管这种情形下的线性回归模型的研究结果很多(参见文献[28]—[49])，但由于该领域内容极为丰富，仍还有非常大的研究空间，所以本书主要讨论相依误差的线性回

归模型的统计性质.

另外,也可以从误差的分布角度来体现误差的相依性和独立性,如假设误差为椭圆分布(定义参见文献[50]),这类分布包含正态分布、$t$ 分布等对称分布在内的一类分布. 在20世纪70年代来,基于独立椭圆分布的线性回归模型也取得了很多丰硕的研究成果(参见文献[51]—[52]等),而基于相依椭圆分布的线性回归模型的研究成果不多(参见文献[53]).

## 1.3 相关回归模型

与线性回归模型相关的回归模型有很多,本节我们简要列出如下几种.

情形 1. 涉及多个回归方程,具有如下形式

$$y_{ti}=\beta_{0i}+\beta_{1i}x_{ti}+e_{ti},\quad i=1,\cdots,M;\quad t=1,2,\cdots,n. \tag{1.3.1}$$

记 $y_i=(y_{1i},\cdots,y_{ni})^{\mathrm{T}}$, $\beta_i=(\beta_{0i},\beta_{1i})^{\mathrm{T}}$, $e_i=(e_{1i},\cdots,e_{ni})^{\mathrm{T}}$, $i=1,2,\cdots,M$, $X_i=\begin{pmatrix}1 & x_{1i}\\ \vdots & \vdots\\ 1 & x_{ni}\end{pmatrix}$,则式(1.3.1)可写成

$$\begin{pmatrix}y_1\\ y_2\\ \vdots\\ y_M\end{pmatrix}=\begin{pmatrix}X_1 & & & 0\\ & X_2 & & \\ & & \ddots & \\ 0 & & & X_M\end{pmatrix}\begin{pmatrix}\beta_1\\ \beta_2\\ \vdots\\ \beta_M\end{pmatrix}+\begin{pmatrix}e_1\\ e_2\\ \vdots\\ e_M\end{pmatrix}, \tag{1.3.2}$$

其误差项假定为

$$Ee_i=0,\quad E(e_ie_j^{\mathrm{T}})=\sigma_{ij}I_n,\quad i,j=1,2,\cdots,M, \tag{1.3.3}$$

称满足式(1.3.1)和式(1.3.3)的回归方程为半相依回归模型(有关结果参见文献[54]—[61]).

情形 2. 时变系数线性模型

$$Y(t)=X^{\mathrm{T}}(t)\beta(t)+\varepsilon(t), \tag{1.3.4}$$

其中 $t$ 为时间指标,$X(\cdot)$, $\beta(\cdot)$, $\varepsilon(\cdot)$, $Y(\cdot)$ 分别为 $p\times 1$ 维协变量过程、时变回归系数、残差过程和响应过程(有关结果参见文献[62]—[65]).

情形 3. 矩阵变量线性回归模型[66]

$$Y_i=\Theta X_i+U_i,\quad i=1,2,\cdots,n, \tag{1.3.5}$$

其中 $Y_i$ 为 $p\times r$ 维观测矩阵,$X_i$ 为 $q\times r$ 维因变量,$U_i$ 为 $p\times r$ 维随机误差,$\Theta$ 为 $p\times q$ 维待估参数.

情形 4. 分层分位回归模型[67]. 为方便起见,我们以具有两层数据的模型为例进行说明. 假定 $(X,W,Y)$ 的一组独立同分布观测值 $(x_i,w_i,y_i)$, $i=1,2,\cdots,n$,其中 $y_i$ 是实数响应变量的值,$x_i$ 为已知的 $1\times d$ 维第一层预测值向量,$w_i$ 为已知的 $d\times f$ 维第二层预测矩阵,满足第一层模型

$$y_i=x_i\beta_i+\varepsilon_i,\quad \varepsilon_i\sim N(0,\sigma^2), \tag{1.3.6}$$

其中 $\beta_i$ 是未知的 $d\times 1$ 维第一层系数向量,$\varepsilon_i$ 是独立同分布的不可观测的随机效应变量.

在第二层模型上，第一层模型中的系数成了输出结果

$$\beta_i = w_i\gamma + u_i, \quad u_i \sim N(0,T), \tag{1.3.7}$$

其中 $\gamma$ 为 $f\times 1$ 维固定效应向量，$u_i$ 为 $d\times 1$ 维第二层随机效应向量，假定它们与 $w_i$ 和 $\varepsilon_i$ 独立.

将式(1.3.7) 代入式(1.3.6) 得

$$y_i = x_i w_i\gamma + x_i u_i + \varepsilon_i, \quad \varepsilon_i \sim N(0,\sigma^2), \quad u_i \sim N(0,T). \tag{1.3.8}$$

若在给定 $X=x$ 和 $W=w$ 的条件下响应变量 $Y$ 的条件分布为 $F(y|x,w)$，则 $Y$ 的 $\tau$ 分位数为

$$q_\tau(x,w) = \inf\{t\in\mathbf{R}: F(t|x,w)\geqslant\tau\} = xw\gamma + (xTx^{\mathrm{T}}+\sigma^2)^{1/2}\Phi^{-1}(\tau), \tag{1.3.9}$$

其中 $0<\tau<1$，$\Phi(\cdot)$ 为标准正态分布函数. 模型(1.3.8) 和(1.3.9) 一起定义为分层分位回归模型.

另外，还有很多非线性模型可以认为是线性模型的推广，在此略.

上述几种情况都是针对模型的结构而言的，可视为经典回归模型的推广. 我们还可以考虑在自变量及因变量的限制条件推广模型.

情形 5. 由于人为或仪器或测量手段，自变量和因变量的观测值难于精确获得，而是含有随机误差. 以一元回归模型为例，可表示为

$$\begin{cases} y_i = \beta_0 + \beta_1 x_i + e_i, \\ X_i = x_i + \varepsilon_i, \end{cases} \tag{1.3.10}$$

其中 $x_i$ 不能被直接观测到，而只能观测到 $X_i$，$e_i$ 和 $\varepsilon_i$ 是随机误差. 统计上称这类模型为误差变量模型(errors-in-variables model)，这方面的研究成果可参见文献[68]—[87].

情形 6. 在模型(1.1.5) 中，若记 $\mu_i = Ey_i$，则

$$\mu_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_{p-1}x_{i,p-1}.$$

如果存在一个严格增的可微函数 $g$，使得

$$\begin{cases} g(\mu_i) = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_{p-1}x_{i,p-1}, \\ \mu_i = Ey_i, \end{cases} \quad i=1,2,\cdots,n, \tag{1.3.11}$$

则模型(1.3.11) 也是模型(1.1.5) 的一种推广，统计学上称这类模型为广义线性模型，其研究成果非常丰富，可参见文献[88]—[107]. 如果 $g(p)=\ln(p/(1-p))$，$0<p<1$，则模型(1.3.11) 为应用很广泛的一类模型 ——Logistic 模型.

## 参 考 文 献

[1] 王松桂，陈敏，陈立萍. 线性统计模型：线性回归与方差分析[M]. 北京：高等教育出版社，1999.

[2] MEERMEYER M. Weighted linear regression models with fixed weights and spherical disturbances[J]. Computational Statistics，2015，30：929-955.

[3] GUO H，ZOU C，WANG Z，et al. Empirical likelihood for high-dimensional linear regression models[J]. Metrika，2014，77：921-945.

[4] LAI T，WEI C. Least squares estimates in stochastic regression models with applications to identification and cont ol of dynamic systems[J]. The Annals of Statistics，1982，10(1)：154-166.

[5] CHENG T. Robust diagnostics for the heteroscedastic regression model[J]. Computational Statistics and Data

Analysis, 2011, 55: 1845-1866.

[6] COOK R D, WEISBERG S. Diagnostics for heteroscedasticity in regression[J]. Biometrika, 1983, 70: 1-10.

[7] WELSH A H, CARROLL R J, RUPPERT D. Fitting heteroscedastic regression models[J]. Journal of the American Statistical Association, 1994, 89: 100-116.

[8] HARVEY A C. Estimating regression models with multiplicative heteroscedasticity[J]. Econometrika, 1976, 38: 375-386.

[9] WEN M J, CHEN S Y, CHEN H J. On testing a subset of regression parameters under heteroscedasticity[J]. Computational Statistics and Data Analysis, 2007, 51: 5958-5976.

[10] MULLER H, STADTMULLER U. Estimation of heteroscedasticity in regression analysis[J]. The Annals of Statistics, 1987, 15: 610-625.

[11] DIXON S L, MCKEAN J W. Rank-based analysis of the heteroscedastic linear model[J]. Journal of the American Statistical Association, 1996, 91: 699-712.

[12] DUFOUR J, KHALAF L, BERNARD J, et al. Simulation-based finite-sample tests for heteroskedasticity and ARCH effects[J]. Journal of Econometrics, 2004, 122: 317-347.

[13] ÖZKALE M R. A jackknifed ridge estimator in the linear regression model with heteroscedastic or correlated errors[J]. Statistics and Probability Letters, 2008, 78: 3159-3169.

[14] CHENG T. On simultaneously identifying outliers and heteroscedasticity without specific form[J]. Computational Statistics and Data Analysis, 2012, 56: 2258-2272

[15] BIANCO A, BOENTE G, RIENZO J. Some results for robust GM-based estimators in heteroscedastic regression models[J]. Journal of Statistical Planning and Inference, 2000, 89: 215-242.

[16] 陈希孺，王松桂. 线性模型中的最小二乘法[M]. 上海：上海科学技术出版社，2003.

[17] 陈希孺，赵林城. 线性模型中的 M 方法[M]. 上海：上海科学技术出版社，1996.

[18] HUBER P J, RONCHETTI E M. Robust Statistics[M]. 2nd ed. Hoboken, New Jersey: John Wiley & Sons, 2009.

[19] DAVISON A C, HINKLEY D V. Bootstrap Methods and Their Application[M]. Cambridge: Cambridge University Press, 1997.

[20] 胡宏昌，崔恒建，秦永松，等. 近代线性回归分析方法[M]. 北京：科学出版社，2013.

[21] 周勇. 广义估计方程估计方法[M]. 北京：科学出版社，2013.

[22] MONTGOMERY D C, PECK B A, VINING G G. Introduction to Linear Regression Analysis[M]. 4th ed. Hoboken, New Jersey: John Wiley & Sons, 2006.

[23] HALL P, HEYDE C C. Martingale Limit Theory and its Application[M]. New York: Academic Press, 1980.

[24] FAN J Q, YAO Q W. Nonlinear Time Series[M]. Berlin: Springer-Verlag, 2005.

[25] 吴群英. 混合序列的概率极限理论[M]. 北京：科学出版社，2006.

[26] KOUL H L, SURGAILIS D. Testing a sub-hypothesis in linear regression models with long memory covariates and errors[J]. Applications of Mathematics, 2008, 53(3): 235-248.

[27] ROBINSON P M. Efficiency improvements in inference on stationary and nonstationary fractional time series[J]. The Annals of Statistics, 2005, 33(4): 1800-1842.

[28] FAN J. Moderate deviations for m-estimators in linear models with $\varphi$-mixing errors[J]. Acta Mathematica Sinica, English Series, 2012, 28(6): 1275-1294

[29] BABU G J. Strong representations for LAD estimators in linear models[J]. Probability Theory and Related Fields, 1989, 83: 547-558

[30] WU Q Y. Strong consistency of M estimator in linear model for $\tilde{\rho}$-mixing samples[J]. Acta Mathematica Scientia, 2005, 25A(1): 41-46.

[31] WU Q Y. Further study strong consistency of M estimator in linear model for $\tilde{\rho}$-mixing random samples[J]. Journal of Systems Science & Complexity, 2011, 24: 969-980

[32] HU H C. QML estimators in linear regression models with functional coefficient autoregressive processes[J]. Mathematical Problems in Engineering, 2010, (12): 242-256.

[33] SONG L, HU H C, CHENG X S. Hypothesis testing in GLM with FCA[J]. Mathematical Problem in

Engineering,2012.

[34] MALLER R A. Asymptotics of regressions with stationary and nonstationary residuals[J]. Stochastic Processes and their Applications,2003,105:33-67.

[35] FULLER W A. Introduction to Statistical Time Series[M]. 2nd ed. New York:John Wiley & Sons,1996.

[36] PERE P. Adjusted estimates and Wald statistics for the AR(1)model with constant[J]. Journal of Econometrics,2000,98:335-363.

[37] YAJIMA Y. On estimation of a regression model with long-memory stationary errors[J]. The Annals of Statistics,1988,16:791-807.

[38] YAJIMA Y. Asymptotic properties of the LSE in a regression model with long-memory stationary errors[J]. The Annals of Statistics,1991,19:158-177.

[39] KOUL H L,MUKHERJEE K. Asymptotics of R-,MD-and LAD-estimators in linear regression with long range dependent errors[J]. Probability Theory and Related Fields,1993,95:538-553.

[40] DAHLHAUS R. Efficient location and regression estimation for long range dependent regression models[J]. The Annals of Statistics,1995,23:1029-1047.

[41] HALL P,LAHIRI S N,POLZEHL J. On bandwidth choice in nonparametric regression with both short and long range dependency errors[J]. The Annals of Statistics,1995,23:1921-1936.

[42] ROBINSON P M, HIDALGO F J. Time series regression with long-range dependence[J]. The Annals of Statistics,1997,25:77-104.

[43] IGLESIAS P,JORQUERA H,PALMA W. Data analysis using regression models with missing observations and long-memory:an application study[J]. Computational Statistics & Data Analysis,2006,50:2028-2043

[44] KLEIBER C. Finite sample efficiency of OLS in linear regression models with long-memory disturbances[J]. Economics Letters,2001,72:131-136

[45] KOUL H L,SURGAILIS D. Goodness-of-fit testing under long memory[J]. Journal of Statistical Planning and Inference,2010,140:3742-3753.

[46] ZHOU Z,WU W B. On linear models with long memory and heavy-tailed errors[J]. Journal of Multivariate Analysis,2011,102:349-362.

[47] WU W B. M-estimation of linear models with dependent errors[J]. The Annals of Statistics, 2007, 35(2): 495-521.

[48] WU R, WANG Q. Shrinkage estimation for linear regression with ARMA errors[J]. Journal of Statistical Planning and Inference,2012,142:2136-2148.

[49] BERA A K,ZUO X L. Specification test for a linear regression model with ARCH process[J]. Journal of Statistical Planning and Inference,1996,50:283-308.

[50] FANG K T,ANDERSON T W. Statistical Inferences in Elliptical Contoured and Related Distributions[M]. New York:Allerton,1990.

[51] GALEA M,PAULA G A,BOLFARINE H. Local influence in elliptical linear regression models[J]. The Statistician,1997,46(1):71-79.

[52] LIU S. On local influence for elliptical linear models[J]. Statistical Papers,2000,41:211-224.

[53] ARASHI M,GOLAM KIBRIA B M,NOROUZIRAD M,et al. Improved preliminary test and Stein-rule Liu estimators for the ill-conditioned elliptical linear regression model[J]. Journal of Multivariate Analysis,2014,126:53-74.

[54] ZELLNER A. An efficient method of estimating seemingly unrelated regression equations and test for aggregation bias[J]. Journal of American Statistical Association,1962,57:348-368.

[55] SRIVASTAVA V K,GILES D E A. Seemingly Unrelated Regression Equations Models[M]. New York:Marcel Dekker,1987.

[56] VELU R,RICHARDS J. Seemingly unrelated reduced-rank regression model[J]. Journal of Statistical Planning and Inference,2008,138:2837-2846.

[57] WANG H. Sparse seemingly unrelated regression modelling: applications in finance and econometrics[J].

Computational Statistics and Data Analysis,2010,54:2866-2877.

[58] CADAVEZ V A P,HENNINGSEN A. The use of seemingly unrelated regression (SUR) to predict the carcass composition of lambs[J]. Meat Science,2012,92(4):548-553.

[59] ZELLNER A,ANDO T. Bayesian and non-Bayesian analysis of the seemingly unrelated regression model with Student-$t$ errors and its application for forecasting[J]. International Journal of Forecasting,2010,26:413-434.

[60] MA T,YE R. Efficient improved estimation of the parameters in two seemingly unrelated regression models[J]. Journal of Statistical Planning and Inference,2010,140:2749-2754.

[61] TIAN M,CHEN G. Hierarchical linear regression models for conditional quantiles[J]. Science in China(Series A) Mathematics,2006,49(12):1800-1815.

[62] ZHOU Z,WU W B. Simultaneous inference of linear models with time varying coefficients[J]. Journal of the Royal Statistical Society B,2010,72:513-531.

[63] RAMSAY J,SILVERMAN B W. Functional Data Analysis[M]. New York:Springer,2005.

[64] FAN J,ZHANG W Y. Simultaneous confidence bands and hypothesis testing in varying-coefficient models[J]. Scandinavian Journal of Statistics,2000,27:715-731.

[65] HONDA T. Quantile regression in varying coefficient models[J]. Journal of Statistical Planning and Inference,2004,121:113-125.

[66] VIROLI C. On matrix-variate regression analysis[J]. Journal of Multivariate Analysis,2012,111:296-309.

[67] 田茂再,陈歌迈. 条件分位中的分层线性回归模型[J]. 中国科学(A辑),2006,36(10):1103-1118.

[68] GLESER L J. Estimation in a multivariate "errors in variables" regression model:large sample results[J]. The Annals of Statistics,1981,9(1):24-44.

[69] AMEMIYA Y,FULLER W A. Estimation for the multivariate errors-in-variables model with estimated error covariance matrix[J]. The Annals of Statistics,1984,12(2):497-509.

[70] DEATON A. Panel data from a time series of cross-sections[J]. Journal of Econometrics,1985,30:109-126.

[71] FULLER W A. Measurement Error Models[M]. New York:Wiley,1987.

[72] MITTAG H J. Estimating parameters in a simple errors-in-variables model: A new approach based on finite sample distribution theory[J]. Statistical Papers,1989,30:133-140.

[73] CUI H J. Asymptotic normality of M-estimator in the EV model[J]. Journal of System Science and Mathematics,1997,10(3):225-236.

[74] CHENG C L,VAN NESS J W. Statistical Regression with Measurement Error[M]. London:Arnold,1999.

[75] CUI H J,CHEN S X. Empirical likelihood confidence region for parameter in the errors-in-variables models[J]. Journal of Multivariate Analysis,2003,84:101-115.

[76] BARAN S. A consistent estimator for linear models with dependent observations[J]. Communications in Statistics:Theory and Methods,2004,33(10):2469-2486.

[77] LIU J X,CHEN X R. Consistency of LS estimator in simple linear EV regression models[J]. Acta Mathematica Sinica,English Series,2005,25(1):50-58.

[78] MIAO Y,YANG G Y,SHEN L M. The central limit theorem for LS estimator in simple linear EV regression models[J]. Communications in Statistics:Theory and Methods,2007,36(12):2263-2272.

[79] MIAO Y,LIU W A. Moderate deviation for LS estimator in simple linear EV regression models[J]. Journal of Statistical Planning and Inference,2009,139(9):3122-3131.

[80] FAN G L,LIANG H Y,WANG J F,et al. Asymptotic properties for LS estimators in EV regression model with dependent errors[J]. ASTA-Advances in Statistical Analysis,2010,94:89-103.

[81] MIAO Y. Convergence rate for LS estimator in simple linear EV regression models[J]. Results in Mathematics,2010,58(1-2):93-104.

[82] MIAO Y,WANG K,ZHAO F. Some limit behaviors for the LS estimator in simple linear EV regression models[J]. Statistics and Probability Letters,2011,81(1):92-102.

[83] MIAO Y,YANG G Y. The loglog law for LS estimator in simple linear EV regression models[J]. Statistics,2011,45(2):155-162.

[84] JIANG R, YANG X, QIAN W. Random weighting M-estimation for linear errors-in-variables models[J]. Journal of the Korean Statistical Society, 2012, 41: 505-514.
[85] FEKRI M, RUIZ-GAZEN A. Robust estimation in the simple errors-in-variables model[J]. Statistics & Probability Letters, 2006, 76: 1741-1747.
[86] CARMICHAEL B, COëN A. Asset pricing models with errors-in-variables[J]. Journal of Empirical Finance, 2008, 15: 778-788.
[87] LAI T L, ROBBINS H, WEI C Z. Strong consistency of least squares estimates in multiple regression[J]. Journal of Multivariate Analysis, 1979, 9: 343-362.
[88] NELDER J A, WEDDERBURN R W M. Generalized linear models[J]. Journal of the Royal Statistical Society A, 1972, 135: 370-384.
[89] KOLASSA J E. Convergence and accuracy of Gibbs sampling for conditional distributions in generalized linear models[J]. The Annals of Statistics, 1999, 27(1): 129-142.
[90] FORSTERJ J, MCDONALDJ W, SMITHP W F. Monte Carlo exact conditional tests for log-linear and logistic models[J]. Journal of the Royal Statistical Society B, 1996, 58: 445-453.
[91] TSOU T S. Determining the mean-variance relationship in generalized linear models-a parametric robust way[J]. Journal of Statistical Planning and Inference, 2011, 141: 197-203.
[92] CROWDER M. On the use of a working correlation matrix in using generalised linear models for repeated measures[J]. Biometrika, 1995, 82: 407-410.
[93] SMYTH G K. Generalized linear models with varying dispersion[J]. Journal of the Royal Statistical Society B, 1989, 51: 47-60.
[94] FAN J, PENG H. Nonconcave penalized likelihood with diverging number of parameters[J]. The Annals of Statistics, 2004, 32: 928-961.
[95] WANG M, SONG L, WANG X. Bridge estimation for generalized linear models with a diverging number of parameters[J]. Statistics and Probability Letters, 2010, 80: 1584-1596.
[96] CHIOGNA M, GAETAN C. Dynamic generalized linear models with application to environmental epidemiology [J]. Applied Statistics, 2002, 51: 453-468.
[97] XUE D, XUE L, CHENG W. Empirical likelihood for generalized linear models with missing responses[J]. Journal of Statistical Planning and Inference, 2011, 141: 2007-2020.
[98] HWANG H, TOMIUK M A. Fuzzy clusterwise quasi-likelihood generalized linear models[J]. Advances in Data Analysis and Classification, 2010, 4: 255-270.
[99] HU B, SHAO J. Generalized linear model selection using $\mathbf{R}^2$[J]. Journal of Statistical Planning and Inference, 2008, 138: 3705 -3712.
[100] FAHRMEIR L, KAUFMANN H. Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models[J]. The Annals of Statistics, 1985, 13(4): 342-368.
[101] CHEN K, HU I, YING Z. Strong consistency of maximum likelihood estimators in generalized linear models with fixed and adaptive designs[J]. The Annals of Statistics, 1999, 27(4): 1155-1163.
[102] ZHAO L C, YIN C M. The strong consistency of the quasi maximum likelihood estimate in the generalized linear model[J]. Science in China A, 2005, 35(3): 312-317.
[103] ZHANG S G, LIAO Y. The weak consistency of maximum likelihood estimators in generalized linear models[J]. Science in China A, 2007, 37(11): 1368-1376.
[104] CHEN S X, CUI H. An extended empirical likelihood for generalized linear models[J]. Statistica Sinica, 2003, 13: 69-81.
[105] BAI Y, FUNG W, ZHU Z. Weighted empirical likelihood for generalized linear models with longitudinal data[J]. Journal of Statistical Planning and Inference, 2010, 140: 3446-3456.
[106] MCCULLAGH P, NELDER J A. Generalized Linear Models[M]. 2nd ed.. London: Chapman & Hall, 1989.
[107] CANTONI E, RONCHETTI E. Robust inference for generalized linear models[J]. Journal of the American Statistical Association, 2001, 96(455): 1022-1030.

# 第 2 章　极大似然估计

## 2.1　极大似然估计概述

极大似然估计法是求未知参数估计的一种方法，它最早是由高斯提出的，后来由 R. A. Fisher 在 1912 年的文章中重新提出，并且证明了这个方法的一些性质. 极大似然估计这一名称也是 R. A. Fisher 给的，目前仍然具有十分广泛的应用，它是建立在极大似然原理的基础上的一个统计方法. 极大似然的直观想法是：一个随机试验如有若干个可能的结果 $A,B,C,\cdots$. 若在一次试验中，结果 $A$ 出现，则一般认为试验条件对 $A$ 出现有利，也即 $A$ 出现的概率最大.

下面对连续型与离散型母体两种情况来阐述极大似然估计法. 设 $\xi_1,\xi_2,\cdots,\xi_n$ 为取自具有密度函数(或概率函数)$\{f(x;\theta):\theta\in\Theta\}$ 的母体 $\xi$ 的一个子样.

如果 $\xi$ 是离散型母体，子样 $\xi_1,\xi_2,\cdots,\xi_n$ 的联合密度函数在 $\xi_i$ 取已知观测值 $x_i$，$i=1,2,\cdots,n$ 时的概率

$$P(\xi_1=x_1,\cdots,\xi_n=x_n)=\prod P(\xi_i=x_i)=f(x_1;\theta)(x_2;\theta)\cdots(x_n;\theta)$$

是 $\theta$ 的函数，我们用 $L(\theta)=L(\theta;x_1,x_2,\cdots,x_n)$ 表示，称为这个子样的似然函数，于是

$$L(\theta)=L(\theta;x_1,x_2,\cdots,x_n)=f(x_1;\theta)f(x_2;\theta)\cdots f(x_n;\theta).\tag{2.1.1}$$

$L(\theta;x_1,x_2,\cdots,x_n)$ 给出了观测到$(x_1,x_2,\cdots,x_n)$的概率. 因此，可以把 $L(\theta;x_1,x_2,\cdots,x_n)$ 看成为了观测到$(x_1,x_2,\cdots,x_n)$时出现什么样 $\theta$ 的可能性的一个测度. 所以我们只要寻找这样的观测值$(x_1,x_2,\cdots,x_n)$的函数 $\hat{\theta}=\hat{\theta}(x_1,x_2,\cdots,x_n)$，以 $\hat{\theta}$ 代 $\theta$ 使

$$L(\hat{\theta};x_1,x_2,\cdots,x_n)=\sup_{\theta\in\Theta}L(\theta;x_1,x_2,\cdots,x_n)\tag{2.1.2}$$

成立. 满足式(2.1.2)的 $\hat{\theta}(x_1,x_2,\cdots,x_n)$ 就是最可能产生 $x_1,x_2,\cdots,x_n$ 的参数 $\theta$ 的值，称 $\hat{\theta}(x_1,x_2,\cdots,x_n)$ 为参数 $\theta$ 的极大似然估计值，其相应的统计量 $\hat{\theta}(\xi_1,\xi_2,\cdots,\xi_n)$ 称作参数 $\theta$ 的极大似然估计量.

如果 $\xi$ 是连续型的，则$\{f(x;\theta):\theta\in\Theta\}$ 表示密度函数. 于是子样$(\xi_1,\xi_2,\cdots,\xi_n)$落入点$(x_1,x_2,\cdots,x_n)$的邻域内的概率为 $\prod\limits_{i=1}^{n}f(x_i;\theta)\Delta x_i$，同样是 $\theta$ 的函数. 既然$(x_1,x_2,\cdots,x_n)$在一次抽样中出现，当然可以认为子样$(\xi_1,\xi_2,\cdots,\xi_n)$落入点$(x_1,x_2,\cdots,x_n)$的邻域内的概率达到最大. 所以我们只要找出使 $\prod\limits_{i=1}^{n}f(x_i;\theta)\Delta x_i$ 达到最大的 $\theta$ 的值 $\hat{\theta}(x_1,x_2,\cdots,x_n)$. 由于 $\Delta x_i$ 是不依赖于 $\theta$ 的增量，只需求出使得

$$L(\hat{\theta};x_1,x_2,\cdots,x_n)=\prod_{i=1}^{n}f(x_i;\theta)$$

达到极大的 $\hat{\theta}(x_1,x_2,\cdots,x_n)$，便可得到极大似然估计. 综上所述，连续型母体的参数的极大似然估计同样可以用式(2.1.1) 与式(2.1.2) 表示.

由于 $\ln x$ 是 $x$ 的单调函数，使

$$\ln L(\hat{\theta};x_1,x_2,\cdots,x_n)=\sup_{\theta\in\Theta}\ln L(\theta;x_1,x_2,\cdots,x_n) \tag{2.1.3}$$

成立的 $\hat{\theta}$ 也使式(2.1.2) 成立，所以有时我们只要从式(2.1.3) 中求 $\hat{\theta}$ 即可.

极大似然估计具有许多优良、重要的性质，下面介绍几个常见性质.

**定理 2.1.1**[1] 设 $\hat{\theta}$ 为 $f(x;\theta)$ 中参数 $\theta$ 的极大似然估计，并且函数 $u=u(\theta)$ 具有单值反函数 $\theta=\theta(u)$，则 $\hat{u}=u(\hat{\theta})$ 是 $u(\theta)$ 的极大似然估计. 这里 $\theta\in\Theta$，$u$ 为 $u(\theta)$ 的值域.

在一些正则条件下，极大似然估计存在唯一，且具有相合性和渐近正态性，见文献[2]，至于极大似然估计的可测性参见文献[3].

**定理 2.1.2**[2] 设随机变量具有密度函数 $f(x;\theta)$，未知参数 $\theta$ 属于 $\mathbf{R}$ 的一个开区间，假定

(1) 对于每一个 $\theta\in\Theta$ 和任意 $x$，偏导数 $\dfrac{\partial\log f(x;\theta)}{\partial\theta}$，$\dfrac{\partial^2\log f(x;\theta)}{\partial\theta^2}$，$\dfrac{\partial^3\log f(x;\theta)}{\partial\theta^3}$ 存在；

(2) 对于每一个 $\theta_0\in\Theta$，存在函数 $g(x)$，$h(x)$ 和 $H(x)$(可能依赖于 $\theta_0$)，使得

$$\frac{\partial\log f(x;\theta)}{\partial\theta}\leqslant g(x),\quad \frac{\partial^2\log f(x;\theta)}{\partial\theta^2}\leqslant h(x),\quad \frac{\partial^3\log f(x;\theta)}{\partial\theta^3}\leqslant H(x),$$

对 $\theta$ 属于 $\theta_0$ 的一个邻域 $N(\theta_0)$ 和任意 $x$ 都成立，且

$$\int g(x)\mathrm{d}x<\infty,\quad \int h(x)\mathrm{d}x<\infty,\quad E_\theta(H(\xi))<\infty\quad(\theta\in N(\theta_0));$$

(3) 对于每一个 $\theta\in\Theta$，

$$0<E_\theta\left\{\left(\frac{\partial\log f(x;\theta)}{\partial\theta}\right)^2\right\}<\infty,$$

则似然方程 $\dfrac{\partial\log L}{\partial\theta}=0$ 以概率 1 有解 $\{\hat{\theta}_n\}$，且

① $\hat{\theta}_n\to\theta$, a.s., $n\to\infty$；

② $\hat{\theta}_n$ 渐近服从正态分布

$$N\left(\theta,\frac{1}{ne_\theta\{(\partial\log f(x;\theta)/\partial\theta)^2\}}\right).$$

**证** 由假设(1)，(2)，对于 $\lambda\in N(\theta)$，有

$$\frac{\partial\log f(x;\lambda)}{\partial\lambda}-\frac{\partial\log f(x;\lambda)}{\partial\lambda}\bigg|_{\lambda=\theta}=(\lambda-\theta)\frac{\partial^2\log f(x;\lambda)}{\partial\lambda^2}\bigg|_{\lambda=\theta}+\frac{1}{2}\eta(\lambda-\theta)^2H(x), \tag{2.1.4}$$

其中 $|\eta|<1$. 令

$$A_n = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial \log f(\xi_i;\lambda)}{\partial \lambda}\Big|_{\lambda=\theta},\quad B_n = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 \log f(\xi_i;\lambda)}{\partial \lambda^2}\Big|_{\lambda=\theta},\quad C_n = \frac{1}{n}\sum_{i=1}^{n} H(\xi_i),$$

则由式(2.1.4)得

$$\frac{1}{n}\frac{\partial \log L(\lambda)}{\partial \lambda} = A_n + B_n(\lambda-\theta) + \frac{1}{2}\eta C_n(\lambda-\theta)^2. \tag{2.1.5}$$

由假设(1)和(2)得

$$\int \frac{\partial f(x;\lambda)}{\partial \lambda}\mathrm{d}x = \frac{\partial}{\partial \lambda}\int f(x,\lambda)\mathrm{d}x = \frac{\partial}{\partial \lambda}(1) = 0, \tag{2.1.6}$$

从而有

$$\int \frac{\partial^2 f(x;\lambda)}{\partial \lambda^2}\mathrm{d}x = 0. \tag{2.1.7}$$

由式(2.1.6)得

$$E_\theta\left\{\frac{\partial f(\xi;\theta)}{\partial \theta}\right\} = \int \frac{1}{f(x;\theta)}\frac{\partial f(x;\theta)}{\partial \theta}f(x;\theta)\mathrm{d}x = \int \frac{\partial f(x;\theta)}{\partial \theta}\mathrm{d}x = 0. \tag{2.1.8}$$

由式(2.1.8)得

$$\begin{aligned} E_\theta\left\{\frac{\partial^2 f(\xi;\theta)}{\partial \theta^2}\right\} &= \int\left[\frac{1}{f(x;\theta)}\frac{\partial^2 f(x;\theta)}{\partial \theta^2} - \left(\frac{1}{f(x;\theta)}\frac{\partial f(x;\theta)}{\partial \theta}\right)^2\right]f(x;\theta)\mathrm{d}x \\ &= -E_\theta\left\{\left(\frac{\partial f(\xi;\theta)}{\partial \theta}\right)^2\right\}. \end{aligned} \tag{2.1.9}$$

由假设(3)知

$$0 < v_\theta = E_\theta\left\{\left(\frac{\partial \log f(\xi;\theta)}{\partial \theta}\right)^2\right\} < \infty. \tag{2.1.10}$$

由式(2.1.8)—式(2.1.10)得

① $A_n$ 是均值为 0 和方差为 $v_\theta$ 的独立同分布随机变量的平均；

② $B_n$ 是均值为 $-v_\theta$ 的独立同分布随机变量的平均；

③ $C_n$ 是均值为 $E_\theta\{H(\xi)\}$ 的独立同分布随机变量的平均.

从而由强大数定律和中心极限定理分别得

$$A_n \to 0,\text{a.s.},\quad B_n \to -v_\theta,\text{a.s.},\quad C_n \to E_\theta\{H(\xi)\},\text{a.s.},\quad n\to\infty \tag{2.1.11}$$

和

$$A_n \to \mathrm{AN}(0, n^{-1}v_\theta)\quad (\text{AN 表示渐近正态分布}). \tag{2.1.12}$$

任意给定 $\varepsilon > 0$，使得 $\varepsilon < v_\theta / E_\theta\{H(\xi)\}$ 和 $\lambda_1 = \theta - \varepsilon \in N(\theta)$，$\lambda_2 = \theta + \varepsilon \in \boldsymbol{N}(\theta)$. 由式(2.1.5)得

$$\left|\frac{1}{n}\frac{\partial \log L(\lambda)}{\partial \lambda}\Big|_{\lambda_1} - v_\theta\varepsilon\right| \leqslant |A_n| + \varepsilon|B_n + v_\theta| + \frac{1}{2}\varepsilon^2|C_n| \tag{2.1.13}$$

和

$$\left|\frac{1}{n}\frac{\partial \log L(\lambda)}{\partial \lambda}\Big|_{\lambda_2} + v_\theta\varepsilon\right| \leqslant |A_n| + \varepsilon|B_n + v_\theta| + \frac{1}{2}\varepsilon^2|C_n|. \tag{2.1.14}$$

由式(2.1.11)知，对于充分大的 $n$，式(2.1.13)和式(2.1.14)右边均以概率 1 小于等于 $\frac{3v_\theta\varepsilon}{4}$. 对于如此 $n$，区间

$$\left[\frac{1}{n}\frac{\partial \log L(\lambda)}{\partial \lambda}\Big|_{\lambda_2}, \frac{1}{n}\frac{\partial \log L(\lambda)}{\partial \lambda}\Big|_{\lambda_1}\right]$$

包含 0 点,因此由 $\partial \log L(\lambda)/\partial\lambda$ 的连续性知,区间

$$[\theta-\varepsilon,\theta+\varepsilon]=[\lambda_1,\lambda_2]$$

包含极大似然估计方程的一个解.特别地,它包含解

$$\hat{\theta}_{n\varepsilon}=\inf\{\lambda:\theta-\varepsilon\leqslant\lambda\leqslant\theta+\varepsilon,\partial\log L(\lambda)/\partial\lambda=0\}. \tag{2.1.15}$$

显然 $\hat{\theta}_{n\varepsilon}$ 是可测随机变量.事实上,对 $\forall\, t\geqslant\theta-\varepsilon$,有

$$\{\hat{\theta}_{n\varepsilon}>t\}=\left\{\inf_{\theta-\varepsilon\leqslant\lambda\leqslant t}\frac{\partial\log L(\lambda)}{\partial\lambda}>0\right\}\cup\left\{\sup_{\theta-\varepsilon\leqslant\lambda\leqslant t}\frac{\partial\log L(\lambda)}{\partial\lambda}<0\right\}.$$

由 $\partial\log L(\lambda)/\partial\lambda$ 在 $[\theta-\varepsilon,\theta+\varepsilon]$ 的连续性得

$$\inf_{\theta-\varepsilon\leqslant\lambda\leqslant t}\frac{\partial\log L(\lambda)}{\partial\lambda}=\inf_{\theta-\varepsilon\leqslant\lambda\leqslant t,\lambda\in\mathbf{Q}}\frac{\partial\log L(\lambda)}{\partial\lambda},\quad \sup_{\theta-\varepsilon\leqslant\lambda\leqslant t}\frac{\partial\log L(\lambda)}{\partial\lambda}=\sup_{\theta-\varepsilon\leqslant\lambda\leqslant t,\lambda\in\mathbf{Q}}\frac{\partial\log L(\lambda)}{\partial\lambda}$$

(其中 **Q** 为有理数),所以 $\{\hat{\theta}_{n\varepsilon}>t\}$ 为可测集.

记 $(\Omega,A,P_\theta)$ 为潜在的概率空间,$\hat{\theta}_{n\varepsilon}=\hat{\theta}_{n\varepsilon}(\omega)$,充分大的 $n\geqslant N_\varepsilon(\omega)$,$P_\theta(\omega\in\Omega_\varepsilon)=1$. 令 $\Omega_0=\bigcap_{k=1}^{\infty}\Omega_{1/k}$,则 $P_\theta(\Omega_0)=1$. 取 $\omega\in\Omega_0$,不失一般性取

$$N_1(\omega)\leqslant N_{1/2}(\omega)\leqslant N_{1/3}(\omega)\leqslant\cdots.$$

于是定义

$$\hat{\theta}_n(\omega)=\begin{cases}\hat{\theta}_{n,1/k}(\omega), & N_{1/k}(\omega)\leqslant n<N_{1/(k+1)}(\omega), \quad k=1,2,\cdots,\omega\in\Omega_0,\\ 0, & n<N_1(\omega), \quad k=1,2,\cdots,\omega\in\Omega_0,\\ 0, & \forall\, n,\omega\notin\Omega_0.\end{cases}$$

所以随机序列 $\{\hat{\theta}_n\}$ 以概率 1 满足:对于足够大的 $n$,$\{\hat{\theta}_n\}$ 是似然方程 $\frac{\partial\log L}{\partial\theta}=0$ 的解,且

$$\hat{\theta}_n\to\theta,\text{a. s.},n\to\infty.$$

对于足够大的 $n$,以概率 1 有

$$0=\frac{1}{n}\frac{\partial\log L(\lambda)}{\partial\lambda}\Big|_{\lambda=\hat{\theta}_n}=A_n+B_n(\hat{\theta}_n-\theta)+\frac{1}{2}\eta C_n(\hat{\theta}_n-\theta)^2, \tag{2.1.16}$$

从而有

$$n^{1/2}(\hat{\theta}_n-\theta)^2-\frac{-n^{1/2}a_n}{B_n+1/2\cdot\eta C_n(\hat{\theta}_n-\theta)}\to 0,\text{a. s.} \tag{2.1.17}$$

由于 $\hat{\theta}_n\to\theta$,a. s. 和 $B_n\to-v_\theta$,a. s. ,所以

$$B_n+1/2\cdot\eta C_n(\hat{\theta}_n-\theta)\to-v_\theta,\text{a. s.} \tag{2.1.18}$$

由式(2.1.17),式(2.1.18) 及 Slutsky 定理,得

$$n^{1/2}(\hat{\theta}_n-\theta)\xrightarrow{D}N(0,v_\theta^{-1}),$$

从而证明了定理 2.1.2.

为了讨论线性回归模型的极大似然估计,我们首先考虑正态随机变量 $\xi\sim N(\mu,\sigma^2)$ 中的未知参数 $\mu,\sigma^2>0$ 的极大似然估计.

**例 2.1**　设 $\xi_1,\xi_2,\cdots,\xi_n$ 是取自正态母体 $N(\mu,\sigma^2)$ 的一个子样,其中 $\mu,\sigma^2>0$ 是未知