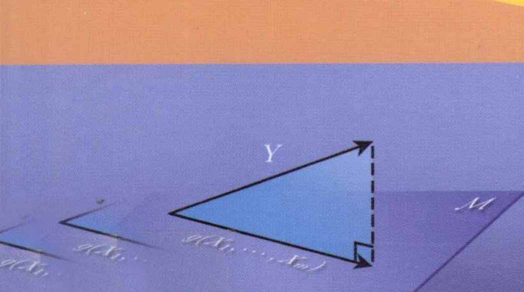
中央财经大学学术专著
出版基金资助出版

# 统计回归分析

## ——回归方程引论

陈乃辉 著




$$\|Y - g(X_1,\cdots,X_m)\| = \inf_{h(X_1,\cdots,X_m)\in\mathcal{M}} \|Y - h(X_1,\cdots,X_m)\|$$

科学出版社

中央财经大学学术专著出版基金资助出版

# 统计回归分析

## ——回归方程引论

陈乃辉　著

科学出版社

北　京

## 内容简介

统计回归分析的本体是回归方程理论，前者乃后者在统计层面上的推绎．回归方程是对一个变量于一组变量某类函数集合中的最佳逼近元的刻画与规定．

本书内容分两大部分．一部分是1~4章及附录A~C，另一部分是5~12章．其中，5~12章是主体，研讨了八大类回归方程，从统计观点而言，即八大类统计回归模型，分别为线性回归方程、Gauss-Markov线性回归方程、非参数回归方程与半参数回归方程、随机向量密度函数、函数系数回归方程、随机过程回归方程、微分回归方程、逆回归方程．前4章及3个附录是对主体部分的理论支撑与辅助，内容包括：概率论、数理统计学、泛函分析、Fourier分析、矩阵代数、测度论及模拟实验SAS软件程序编制等．

本书可作为学习统计回归分析及相关学科(如物理、生物、经济、金融与管理等)的高年级本科生和研究生教材，也可供教师及科研人员参考．

**图书在版编目(CIP)数据**

统计回归分析：回归方程引论/陈乃辉著.—北京：科学出版社，2012

ISBN 978-7-03-033110-6

Ⅰ.①统… Ⅱ.①陈… Ⅲ.①回归分析 Ⅳ.①O212.1

中国版本图书馆CIP数据核字(2011)第270982号

责任编辑：李 欣 赵彦超／责任校对：李 影

责任印制：钱玉芬／封面设计：王 浩

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

北京佳艺恒彩印刷有限公司 印刷

科学出版社发行 各地新华书店经销

*

2012年1月第一版 开本：B5(720×1000)

2012年1月第一次印刷 印张：19 1/2

字数：370 000

定价：68.00元

# 前　　言

## (一)

1805 年 Legendre、1809 年 Gauss 分别在大地测量领域与天文测量领域中相继提出所谓线性回归模型问题, 即带有随机误差的线性方程组之求解问题, 并创立最小二乘方法而予以解决, 这可以视为统计回归分析理论的渊源.

二百年来, 统计回归分析一直处于数理统计学的中心地位, 已有丰厚的积累. 然而, 与概率论相较, 其归宿于纯粹数学的进化, 可谓蛹在茧中.

20 世纪初叶, 借助于实变函数论及测度论, 概率论 (与数理统计学相对而言) 得以完成其数学化的蜕变. 就此, 苏联概率论大家 A.H.Kolmogrov 曾论述到: "要把一向显得特异的概率论基本概念很自然地归入近代数学一般概念的行列中去, 在测度与 Lebesgue 积分理论未产生以前, 这一问题几乎是没有解决希望的. 有了 Lebesgue 的研究以后, 集合测度与事件概率之间的相似性, 以及函数的积分与随机变量的数学期望之间的相似性就了如指掌. 这相似性还可更进一步, 例如独立随机变量的许多性质与正交函数的相应性质是完全相似的 ······"① 作者感觉, 现在统计回归分析所处之历史发展阶段, 相若于当年的概率论.

基于以往的积累, 援用泛函分析与 Fourier 分析的观点、方法, 欲成就统计回归分析之纯粹数学化的升华, 化臆测为逻辑, 化经验为先验, 化表象为本体, 化冗杂为简明, 化不美为美焉, 如是纯粹数学化的统计回归分析理论, 反过来再应用于物象世界, 将使本学科螺旋上升到一个新层次, 本书是这个意愿的结果. 或许本书之理论形态, 更接近统计回归分析的本来面目.

## (二)

从概率层次而观, 统计回归分析之本在于回归方程理论.

回归方程理论的意义, 可以借助法国哲学家、数学家笛卡儿的哲言来表达: "当我们不具备决定什么是真理的力量时, 我们应遵从什么是最可能的, 这是千真万确的真理."②

①赵荣侠, 崔群劳. 测度与积分. 西安: 西安电子科技大学出版社, 2002.

②C. R. 劳. 统计与真理: 怎样运用偶然性. 北京: 科学出版社, 2004.

当自变量组 $X_1,\cdots,X_m$ 包括了因变量 $Y$ 的所有影响因素，那么存在 $X_1,\cdots,X_m$ 的函数 $g(X_1,\cdots,X_m)$，使得下列方程成立

$$Y=g(X_1,\cdots,X_m)$$

而当自变量组 $X_1,\cdots,X_m$ 并未涵盖因变量 $Y$ 的所有影响因素时，$Y$ 不能表示成 $X_1,\cdots,X_m$ 的函数. 事实上，后者是更为普遍存在的情况，于此时，当退而求其次，就是求 $X_1,\cdots,X_m$ 的所有函数或某类函数 $\mathcal{M}(X_1,\cdots,X_m)$ 中最为逼近 $Y$ 的那一个 $g(X_1,\cdots,X_m)$，即下列方程的解

$$\|Y-g(X_1,\cdots,X_m)\|=\min_{h(X_1,\cdots,X_m)\in\mathcal{M}(X_1,\cdots,X_m)}\|Y-h(X_1,\cdots,X_m)\|$$

此方程称为 $Y$ 关于 $X_1,\cdots,X_m$ 的回归方程，而方程之解 $g(X_1,\cdots,X_m)$ 称为 $Y$ 关于 $X_1,\cdots,X_m$ 的回归函数. 回归方程是普通方程概念的自然延伸，方程是回归方程的特例，即当 $g(X_1,\cdots,X_m)\in\mathcal{M}(X_1,\cdots,X_m)$ 时，回归方程退化为方程. 换言之，回归函数是普通函数概念的自然延伸，函数是回归函数的特例. 回归方程理论就是求解回归方程之回归函数的理论.

回归方程的解有两种类型，一是概率解，一是统计解. 前者指在已知回归方程中自变量、因变量的联合概率分布的前提下所获之解，后者指在已知回归方程中自变量、因变量的联合样本的前提下所获之解. 统计回归分析就是求解回归方程之统计解的理论. 从理路上而言，概率解为本，统计解是末，末应由于本，囿之逐末未本，难免溺于茫然无归.

概率理论是一个纯粹逻辑体系，数理统计学 (统计推断理论) 不过为它的一个特殊部分，用陈希孺先生的话言之:“实际上，这种 (统计) 推断的意义正如在欧氏几何公理体系下所推演出的任何一个命题那样明确.”[4] 类此，统计回归分析乃为回归方程理论的一个特殊部分，其与欧几里得几何无异. 经验的、实验的、艺术的统计学，其当下的根据与最终的归宿是数理统计学，这类比于实验的物理与理论的物理之关系.

## (三)

回归方程论的理论基础来自三个学科: 概率理论 (包括数理统计学)、泛函分析与 (广义)Fourier 分析.

将概率空间 $(\Omega,\mathscr{F},P)$ 上所有平方可积随机变量看做一个整体而得到空间 $L^2(\Omega,\mathscr{F},P)$，其成为一个 Hilbert 空间，如是参数型回归方程的概率解可援用正交投影定理得到解答，缘于此解是方程变量的诸概率矩，因而利用大数定律与矩方法,

可获概率 1 意义下的统计解. 至于非参数型回归方程, 在投影空间 $\mathcal{M}(X_1, \cdots, X_m)$ 存在至多可列维基的情形下, 借助 (广义)Fourier 级数理论, 将非参数型回归函数化成参数型回归函数之逼近序列, 而解决之.

上述回归方程论的理论基础, 非皆现成, 如 Hilbert 空间 $L^2(\Omega, \mathscr{F}, P)$ 中幂函数型 (准 Schauder) 基的存在条件, 本书第 4 章属意之于此问题, 其核心工作是推广 Luzin(卢津) 定理.

本回归方程论, 与以往的统计回归分析相较, 主要不同还有:

(1) 从概率层次着眼, 而非浮于统计的样本层面;

(2) 以最佳逼近定理 (正交投影定理的推论) 替换最小二乘方法, 后者是表象, 前者是本质;

(3) 关于非参数型回归方程所求之解, 乃意在全局解析的, 而非局部数值的;

(4) 较大地拓广了回归方程的谱系, 特别地, 其中微分回归方程已将数学中的大家族——微分方程纳入囊中. 进一步值得关注者, 乃变分方程与回归方程的关系如何? 而物象世界中的基本原理——最小作用原理之表现, 可用变分方程予以数理刻画和演绎.

## (四)

本书的后 8 章 (第 5 章 ~ 第 12 章) 为该书的主体部分, 研讨了八大类回归方程; 前 4 章及附录 (附录 A~ 附录 C) 为主体部分的理论基础与辅助. 书中有相当部分简单的或易于模仿的定理证明, 着意留给学生读者完成, 以此权作习题.

本书稿于中央财经大学全校研究生公共选修课上连续讲授三年 (2009~ 2011 年), 获得学生们的积极影响与应求, 此乃书稿成长的重要因素. 该书出版获中央财经大学学术专著出版基金资助, 在此表示感谢.

鉴于作者德能莫逮, 且所作欠于岁月磨砺, 书稿中尚需改进与琢磨者, 一定不少, 甚或错漏之处, 在所不免, 诚望专家、读者体察, 并予以批评指正, 以利修葺与完善, 庶几臻于有功.

作　者　谨识

2011 年 10 月于北京清河

# 符 号 表

| | |
|---|---|
| $C(F)$ | 函数 $F(x)$ 的连续点集合 |
| $\mathrm{Col}(A)$ | 矩阵 $A$ 列向量的张成空间 |
| $e_j$ | $e_j \in \mathbb{R}^n$, 其第 $j$ 项为 1, 余者皆为 0 |
| $\inf\{M\}$ | 实数集 $M$ 的最大下界 |
| $\limsup\limits_{n\to\infty} x_n$ | 数列 $x_n$ 的最大之部分极限 |
| $\liminf\limits_{n\to\infty} x_n$ | 数列 $x_n$ 的最小之部分极限 |
| $I_m$ | $m$ 阶单位阵 |
| $\mathcal{M}_1 \oplus \mathcal{M}_2$ | 子空间 $\mathcal{M}_1$ 与 $\mathcal{M}_2$ 的正交和 |
| $\max\{M\}$ | 实数集 $M$ 的最大值 |
| $\min\{M\}$ | 实数集 $M$ 的最小值 |
| $\mathcal{M}_1 + \mathcal{M}_2$ | 子空间 $\mathcal{M}_1$ 与 $\mathcal{M}_2$ 的和 |
| $\mathbb{N}$ | 自然数集合 |
| $\mathrm{rank}(A)$ | 矩阵 $A$ 的秩 |
| $\mathbb{R}$ | 实数集合 |
| $\mathbb{R}_+$ | 正实数集合 |
| $\mathbb{R}^{m\times n}$ | 实 $m \times n$ 型矩阵集合 |
| $\mathbb{R}^n$ | 实数集合的 $n$ 次直积 $\mathbb{R} \times \cdots \times \mathbb{R}$, 即 $n$ 维实数列向量 |
| $\sup\{M\}$ | 实数集 $M$ 的最小上界 |
| $\mathbb{Z}$ | 整数集合 |
| $\mathbb{Z}^+$ | 非负整数集合 |
| $\mathbf{1}_m$ | $m$ 阶列向量, 其元素皆是 1 |
| $\#(M)$ | 集合 $M$ 元素的个数 |

# 目　　录

# 第 1 章　概　率　论

## 1.1　随机向量

### 1.1.1　测度空间

随机变量是一个泛函, 且其定义域空间上设置了一个测度结构, 即其定义域空间是一个测度空间.

**定义 1.1.1**　设集合 $\Omega$ 的子集族 $\mathscr{F}$ 满足

(1) $\Omega \in \mathscr{F}$;

(2) 若 $A \in \mathscr{F}$, 则 $A^c \in \mathscr{F}$;

(3) 若 $A_n \in \mathscr{F}(n \in \mathbb{N})$, 则 $\bigcup\limits_{n=1}^{\infty} A_n \in \mathscr{F}$.

称 $\mathscr{F}$ 是 $\Omega$ 的 $\sigma$ 代数 ($\sigma$- 域).

**例 1.1.1**　$\mathbb{R}$ 的一个通常 $\sigma$ 代数为 Borel 集

$$\mathscr{B}(\mathbb{R}) \triangleq \text{包含}\{(a,b] \subset \mathbb{R} : -\infty \leqslant a \leqslant b \leqslant +\infty\} \text{ 的最小 } \sigma \text{ 代数}$$

**定义 1.1.2**　设 $\mathscr{F}$ 为集合 $\Omega$ 的 $\sigma$ 代数, 称 $(\Omega, \mathscr{F})$ 为可测空间.

**例 1.1.2**　$(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ 为可测空间.

**定义 1.1.3**　设 $(\Omega, \mathscr{F})$ 是可测空间, 集函数 $\mu : \mathscr{F} \to [0, +\infty]$ 满足对两两不交的集合序列 $A_n \in \mathscr{F}(n \in \mathbb{N})$, 有

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$$

称 $\mu$ 为可测空间 $(\Omega, \mathscr{F})$ 上的 $\sigma$ 可加测度, 简称测度; 若 $\mu$ 又满足 $\mu(\Omega) = 1$, 则称 $\mu$ 为可测空间 $(\Omega, \mathscr{F})$ 上的概率测度, 简称概率, 记为 $P$.

**定义 1.1.4**　设 $\mu$ 是可测空间 $(\Omega, \mathscr{F})$ 上的测度, 称 $(\Omega, \mathscr{F}, \mu)$ 为测度空间. 设 $P$ 是可测空间 $(\Omega, \mathscr{F})$ 的概率, 称 $(\Omega, \mathscr{F}, P)$ 为概率空间.

**例 1.1.3**　Lebesgue 测度 $\mu_L$ 是可测空间 $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ 上的常见测度, 其是区间长度概念的自然延拓. 而 $(\mathbb{R}, \mathscr{B}(\mathbb{R}), \mu_L)$ 为测度空间.

### 1.1.2　概率分布

随机数学中的随机变量 (或称随机数) 相当于非随机数学中的数.

**定义 1.1.5**　设 $(\Omega,\mathscr{F},P)$ 是概率空间, $X(\omega)$ 是定义在 $\Omega$ 上的单值实函数, $\forall x\in\mathbb{R}$, 有

$$\{\omega : X(\omega)\leqslant x\}\in\mathscr{F}$$

称 $X(\omega)$ 为 $(\Omega,\mathscr{F},P)$ 上的一个随机变量, 简记为 $X$. 称

$$P\{X(\omega)\in B\}\quad (B\in\mathscr{B}(\mathbb{R}))$$

为随机变量 $X(\omega)$ 的概率分布.

**定义 1.1.6**　设 $X_1,\cdots,X_n$ 是概率空间 $(\Omega,\mathscr{F},P)$ 上的随机变量, 称有序随机变量组

$$(X_i)^{1\leqslant i\leqslant n}=(X_1,\cdots,X_n)^{\mathrm{T}}$$

为 $(\Omega,\mathscr{F},P)$ 上的 $n$ 维随机向量. 设 $X_{ij}(1\leqslant i\leqslant m,1\leqslant j\leqslant n)$ 皆是概率空间 $(\Omega,\mathscr{F},P)$ 上的随机变量, 称随机变量的矩阵

$$(X_{ij})^{1\leqslant i\leqslant m}_{1\leqslant j\leqslant n}=\begin{pmatrix} X_{11} & \cdots & X_{1n}\\ \vdots & & \vdots\\ X_{m1} & \cdots & X_{mn}\end{pmatrix}$$

为 $(\Omega,\mathscr{F},P)$ 上的随机矩阵.

下面三个定义是对随机向量完全刻画的三种手法.

**定义 1.1.7**　设 $(X_1,\cdots,X_n)^{\mathrm{T}}$ 是概率空间 $(\Omega,\mathscr{F},P)$ 上的随机向量, 称 $n$ 元函数

$$F(x_1,\cdots,x_n)=P\{X_1\leqslant x_1,\cdots,X_n\leqslant x_n\}\quad ((x_1,\cdots,x_n)^{\mathrm{T}}\in\mathbb{R}^n)$$

为随机向量 $(X_1,\cdots,X_n)^{\mathrm{T}}$ 的分布函数, 记作 $F_{(X_1,\cdots,X_n)}(x_1,\cdots,x_n)$.

**定义 1.1.8**　设 $(X_1,\cdots,X_n)^{\mathrm{T}}$ 是概率空间 $(\Omega,\mathscr{F},P)$ 上的随机向量, 存在非负可积函数 $f(x_1,\cdots,x_n)$, 使

$$F(x_1,\cdots,x_n)=\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_n}f(u_1,\cdots,u_n)\mathrm{d}u_1\cdots\mathrm{d}u_n\quad ((x_1,\cdots,x_n)^{\mathrm{T}}\in\mathbb{R}^n)$$

其中的积分为 Lebesgue 积分, 称 $(X_1,\cdots,X_n)^{\mathrm{T}}$ 为连续型随机向量, $f(x_1,\cdots,x_n)$ 为 $(X_1,\cdots,X_n)^{\mathrm{T}}$ 的密度函数, 记作 $f_{(X_1,\cdots,X_n)}(x_1,\cdots,x_n)$.

**例 1.1.4**　设 $\boldsymbol{X}=(X_1,\cdots,X_n)^{\mathrm{T}}$ 是概率空间 $(\Omega,\mathscr{F},P)$ 上的随机向量, 其密度函数为

$$f(x_1,\cdots,x_n)=\begin{cases}\dfrac{1}{\prod\limits_{i=1}^{n}(b_i-a_i)}, & (x_1,\cdots,x_n)^{\mathrm{T}}\in\prod\limits_{i=1}^{n}[a_i,b_i],\\ 0, & 其他\end{cases}$$

称 $(X_1,\cdots,X_n)^{\mathrm{T}}$ 服从 $\prod\limits_{i=1}^{n}[a_i,b_i]$ 上的 $n$ 维均匀分布, 记作 $\boldsymbol{X}\sim U\left(\prod\limits_{i=1}^{n}[a_i,b_i]\right)$.

**例 1.1.5** 设 $\boldsymbol{X}=(X_1,\cdots,X_p)^{\mathrm{T}}$ 是概率空间 $(\Omega,\mathscr{F},P)$ 上的随机向量, 其密度函数为

$$f(x)=\frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}}\exp\left\{-\frac{1}{2}(x-\mu)^{\mathrm{T}}\Sigma^{-1}(x-\mu)\right\}\quad(x\in\mathbb{R}^p)$$

其中 $\mu\in\mathbb{R}^p,\Sigma\in\mathbb{R}^{p\times p},\Sigma>0$, 称 $\boldsymbol{X}$ 服从 $p$ 元正态分布, 记作 $\boldsymbol{X}\sim N_p(\mu,\Sigma)$.

多元正态分布随机向量的线性函数仍服从多元正态分布.

**定理 1.1.1** 若随机向量 $\boldsymbol{X}\sim N_p(\mu,\Sigma)$, $B$ 是 $s\times p$ 型常数矩阵, $d$ 是 $s$ 维常数向量, 令

$$\boldsymbol{Z}=B\boldsymbol{X}+d$$

则随机向量 $\boldsymbol{Z}\sim N_s(B\mu+d,B\Sigma B^{\mathrm{T}})$.

**定义 1.1.9** 设 $(X_1,\cdots,X_n)^{\mathrm{T}}$ 是概率空间 $(\Omega,\mathscr{F},P)$ 上的随机向量, 存在 $\mathbb{R}^n$ 上的至多可列集 $\{(x_{k_1},\cdots,x_{k_n})\in\mathbb{R}^n:(k_1,\cdots,k_n)\in\Lambda\}$, 使

$$P\{(X_1,\cdots,X_n)=(x_{k_1},\cdots,x_{k_n}):(k_1,\cdots,k_n)\in\Lambda\}=1$$

称 $(X_1,\cdots,X_n)^{\mathrm{T}}$ 为离散型随机向量, 称

$$P\{(X_1,\cdots,X_n)=(x_{k_1},\cdots,x_{k_n})\}=p_{k_1,\cdots,k_n}\quad((k_1,\cdots,k_n)\in\Lambda)$$

为 $(X_1,\cdots,X_n)^{\mathrm{T}}$ 的分布律.

**例 1.1.6** 设 $(X_1,\cdots,X_n)^{\mathrm{T}}$ 是概率空间 $(\Omega,\mathscr{F},P)$ 上的随机向量, 其分布律为

$$P\{(X_1,\cdots,X_n)=(k_1,\cdots,k_n)\}=\mathrm{e}^{-(\lambda_1+\cdots+\lambda_n)}\frac{\lambda_1^{k_1}\cdots\lambda_n^{k_n}}{k_1!\cdots k_n!}\quad((k_1,\cdots,k_n)\in(\mathbb{Z}^+)^n)$$

其中 $\lambda_i>0(1\leqslant i\leqslant n)$, 称 $(X_1,\cdots,X_n)^{\mathrm{T}}$ 服从 $n$ 维 Poisson 分布, 记作 $(X_1,\cdots,X_n)^{\mathrm{T}}\sim P(\lambda_1,\cdots,\lambda_n)$.

随机向量关于子随机向量的边缘概率分布, 指随机向量的概率分布向该子随机向量所在子空间的垂直投置.

**定义 1.1.10** 称随机向量 $(X_1,\cdots,X_n)^{\mathrm{T}}$ 的子随机向量 $(X_{i_1},\cdots,X_{i_k})^{\mathrm{T}}$ 的分布函数

$$F_{(X_{i_1},\cdots,X_{i_k})}(x_{i_1},\cdots,x_{i_k})=P(X_{i_1}\leqslant x_{i_1},\cdots,X_{i_k}\leqslant x_{i_k})\quad((x_{i_1},\cdots,x_{i_k})^{\mathrm{T}}\in\mathbb{R}^k)$$

为 $(X_1,\cdots,X_n)^{\mathrm{T}}$ 关于 $(X_{i_1},\cdots,X_{i_k})^{\mathrm{T}}$ 的边缘分布函数.