周先波 主编

中山大学经济管理实验教学示范中心（国家级实验教学示范中心）系列教材

计量经济学

非参数估计及GAUSS应用

清华大学出版社

周先波 主编

中山大学经济管理实验教学示范中心（国家级实验教学示范中心）系列教材

计量经济学

非参数估计及GAUSS应用

清华大学出版社

北京

内 容 简 介

本书是一本理论和实验高度融合的教材，将介绍非参数估计的经典计量方法（密度函数、回归函数）、部分线性回归模型和变系数回归模型的估计方法、单指数模型、加法模型面板数据模型的半参数和非参数估计方法。在每一部分，将给出大量的模拟实验范例以说明相应估计量的小样本表现，同时给出一些实证应用例子，说明半参数和非参数估计方法的应用。相应地，这些模拟和实证的例子都配以GAUSS程序，允许读者结合相应模型重复模拟和实证例子的结果，更好地复习和消化各章内容。

图书在版编目（CIP）数据

计量经济学：非参数估计及 GAUSS 应用/周先波主编. —北京：清华大学出版社，2017
（中山大学经济管理实验教学示范中心（国家级实验教学示范中心）系列教材）
ISBN 978-7-302-45319-2

Ⅰ. ①计…　Ⅱ. ①周…　Ⅲ. ①计量经济学－高等学校－教材　Ⅳ. ①F224.0

中国版本图书馆 CIP 数据核字(2016)第 260840 号

责任编辑：陆浥晨
封面设计：单　良
责任校对：王凤芝
责任印制：何　芊

出版发行：清华大学出版社
　　网　　址：http://www.tup.com.cn，http://www.wqbook.com
　　地　　址：北京清华大学学研大厦 A 座　　**邮　　编**：100084
　　社 总 机：010-62770175　　**邮　　购**：010-62786544
　　投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn
　　质 量 反 馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
印 装 者：三河市少明印务有限公司
经　　销：全国新华书店
开　　本：185mm×260mm　　**印　　张**：14.5　　**字　　数**：297 千字
版　　次：2017 年 1 月第 1 版　　**印　　次**：2017 年 1 月第 1 次印刷
印　　数：1～3000
定　　价：39.00 元

产品编号：059221-01

前言

非参数计量经济学理论和应用是当前计量经济学的重要研究领域。写本书的动机是为研究生提供一本学习“非参数计量经济学”的中文教材或参考书。

自2008年9月至今，我一直承担中山大学岭南（大学）学院博士研究生“高级计量经济学Ⅱ：非参数估计”课程的教学，也一直在思考，如何让学生更好地领会非参数和半参数估计的动因。教学中发现，非参数估计的动因讲起来容易，即让模型设定形式更富灵活性，但真正使学生理解、感兴趣和恰当应用非参数估计方法真不是一件轻松活。

长期以来，同学们习惯于参数回归模型的设定、估计和检验，熟悉OLS、工具变量、GMM等参数估计和检验方法；而且，相当喜欢EViews、Stata等计量经济软件的命令操作，因为固定的线性回归模型形式的设定配以高效的计量软件操作可以产生高效率的实证研究文章。不过，别人（评审人）常从研究的严谨性角度，质疑研究文章中所用计量方法的合理性，所得结论和推断的可靠性。国内计量经济学者早在2008年就对计量经济学应用研究中的总体回归模型设定问题作了系统阐述（李子奈，2008），即使是国内具有较高学术水平和重要应用价值的已发表论文，从计量经济学模型总体设定的角度看，也是值得讨论的。尽管这些总体回归模型设定的原则（如唯一性、一般性、现实性和统计检验必要性等）是从参数模型的角度提出，但其逻辑性思想同样适合于非参数计量经济学建模。

模型总体设定的合理性和可信性是计量经济实证研究的关键点。正如李子奈教授指出的，实证研究中问题多多、错误普遍的重要原因主要是我们缺乏对计量经济学模型方法论基础的研究和理解。计量经济学作为一种方法论，其哲学基础、经济学基础、数学基础和统计学基础应受到足够的重视。教学中，教师如何让学生更好地领会计量模型总体设定的合理性和不足点是一个重要的问题。同学们在模型选择时经常产生诸多困惑：为什么选择这种设定而不用那种设定？为什么一种估计结果比另一种估计更合适？在什么情况下需怀疑所用模型的设定形式，进而使用非参数或半参数估计？为什么、怎样进行总体回归模型的设定检验？

本书希望从非参数和半参数回归模型的角度介绍模型总体设定的估计

方法，阐述非参数和半参数估计的基本内容，力图使不同内容之间的联系富有逻辑性；同时希望学生和读者在阅读本书后，可以在计量实证分析方面做出更好的研究成果。

本书共有七章。第1章给出本书的基础准备知识，并阐述非参数和半参数估计的动因，配以数值模拟说明。第2章介绍密度函数的非参数核估计方法，阐述密度函数核估计的有限样本和大样本性质。第3章介绍回归函数非参数核估计方法及模型设定检验。第4章介绍部分线性回归模型和变系数回归模型的半参数估计方法，同时也涉及模型的设定检验。第5章介绍单指数模型的半参数估计方法，并对二元选择模型等受限因变量模型的半参数估计方法作一简单介绍。第6章介绍加法模型的估计方法。第7章介绍面板数据模型的非参数估计和检验方法。

本书附录部分给出各章模拟和应用例子的GAUSS程序，供读者学习时参考。虽然这部分内容较多，但作者觉得同学们学会编程是值得的。作者在这部分列出的是经测试运行相对稳定的程序，稍加修改可用于具体问题研究。除例5.3由单纯形法求解最小化问题的GAUSS段落程序AMOEB(见程序ex53)和例7.3面板数据非参数固定效应模型估计的GAUSS程序(见程序ex73)由作者根据D. J. Henderson提供的Matlab程序加以修改之外，其他各例的程序均由作者自己编写开发。作者希望这份艰辛能对同学们学习非参数估计方法有所帮助，当然，也希望读者在应用之后提出修改或完善程序的意见。

非参数计量经济学的内容相当丰富，本书只是介绍关于横截面数据和面板数据对应的部分内容，希望读者提出更好的建议和批评意见，以助作者进一步完善对非参数计量经济学内容的阐述和安排。

本书写作过程中参考了国内外许多非参数计量经济学方面的文献和教材，特别是李奇教授的书(Li and Racine, 2007)，衷心感谢这些学者。也感谢AMOEB程序的始编写者，感谢Henderson教授的帮助。

本书撰写是一个漫长的过程。在王美今教授的提议下岭南学院于2008年就开设了“非参数估计”课程，让我在这一教学平台接触许多优秀的同学，逐渐萌发了编写这本书的动机。感谢岭南学院各届领导和王美今教授的支持。

感谢李胜兰教授和中山大学经济管理实验教学中心“第二届实验课程建设项目”对本书出版的支持。感谢岭南学院信息与教育中心郭凌、罗霖等老师对软件运行平台AFS系统的维护，此平台对计量经济学上机教学和本书GAUSS程序运行起了重要作用。博士生潘哲文、窦智、欧阳梦倩、李赫扬等同学参与了书稿的校对工作，我的历届研究生们在本书编写过程中也给予了帮助，在此表示谢意。

最后，感谢清华大学出版社及陆浥晨编辑对本书出版提供的帮助，冀保存这份感激于以后的写作中。

周先波

2016年6月于中山大学岭南学院

E-mail：zhouxb@mail. sysu. edu. cn

目录

第1章

准备知识及非参数和半参数估计的动因

本章知识点

随机变量的总体特征；联合分布及密度函数；
条件数学期望及回归函数，迭代期望原理；
随机变量的统计特征；
大数定理和中心极限定理；
非参数与半参数估计的动因；
估计量小样本表现的模拟，检验统计量的 bootstrap；
GAUSS 编程语言。

本章以连续型随机变量为例，简单介绍数理统计的基础知识，从回归函数的角度说明非参数估计的动因。

1.1 随机变量及其分布

设 X 是随机变量，概率密度函数为 $f(x)$，则其分布函数是 $F(x)=\int_{-\infty}^{x} f(t)\mathrm{d}t$，概率 $P(a<X\leqslant b)=\int_{a}^{b} f(x)\mathrm{d}x$。随机变量的两个基本总体特征是数学期望 EX 和方差 $\operatorname{Var}(X)$，如果 $\int_{-\infty}^{\infty}|x|f(x)\mathrm{d}x<\infty$，它们分别定义为

$$EX\equiv\int_{-\infty}^{\infty} xf(x)\mathrm{d}x$$

$$\operatorname{Var}(X)\equiv\int_{-\infty}^{\infty}(x-EX)^2 f(x)\mathrm{d}x=E(X-EX)^2=E(X^2)-(EX)^2$$

数学期望是随机变量 X 取值的加权平均，说明 X 的取值总在此均值 EX 的周围波动，反映随机变量的平均水平，是消除变量随机性的主要手段。方差是 X 的取值 x 关于均值 EX 的波动幅度 $(x-EX)^2$ 的加权平均，刻画了 X 的取值在均值 EX 周围的波动程度，反映随机变量的集中度或分散性，也是消除变量随机性的手段。

由积分运算的线性性可知，数学期望是线性运算，即 $E(aX+b)=aEX+b$，其中 a、b 为常数。由此，方差运算满足 $\operatorname{Var}(aX+b)=a^2\operatorname{Var}(X)$。

由数学期望和方差运算，可得**切比雪夫(Chebyshev)不等式**：设 EX 和 $\operatorname{Var}(X)$ 都存

在，则 $\forall \varepsilon>0$，有

$$P(|X-EX|\geqslant\varepsilon)\leqslant\frac{\mathrm{Var}(X)}{\varepsilon^2}\quad \text{或}\quad P(|X-EX|<\varepsilon)\geqslant 1-\frac{\mathrm{Var}(X)}{\varepsilon^2}$$

事实上，

$$\begin{aligned}P(|X-EX|\geqslant\varepsilon)&=\int_{|x-EX|\geqslant\varepsilon}f(x)\mathrm{d}x\\&\leqslant\int_{|x-EX|\geqslant\varepsilon}\frac{(x-EX)^2}{\varepsilon^2}f(x)\mathrm{d}x\\&\leqslant\int_{-\infty}^{\infty}\frac{(x-EX)^2}{\varepsilon^2}f(x)\mathrm{d}x=\frac{\mathrm{Var}(X)}{\varepsilon^2}\end{aligned}$$

事件 $\{|X-EX|\geqslant\varepsilon\}$ 称为随机变量 X 的大偏差。切比雪夫不等式说明，随机变量 X 大偏差的发生概率存在一个与方差成正比的上界，方差越大，此上界也越大。

由切比雪夫不等式可以证明：若 $\mathrm{Var}(X)$ 存在，则 $\mathrm{Var}(X)=0$ 当且仅当 X 几乎处处等于某常数 a，即 $P(X=a)=1$。事实上，由 $P(X=a)=1$，显然 $\mathrm{Var}(X)=0$；反之，令 $a=EX$，则有

$$\begin{aligned}0&\leqslant P(|X-a|>0)\\&=P\left(\bigcup_{n=1}^{\infty}\left\{|X-a|\geqslant\frac{1}{n}\right\}\right)\\&\leqslant\sum_{n=1}^{\infty}P\left(|X-a|\geqslant\frac{1}{n}\right)\\&\leqslant\sum_{n=1}^{\infty}\frac{\mathrm{Var}(X)}{1/n^2}=0\end{aligned}$$

从而 $P(|X-a|=0)=1-P(|X-a|>0)=1$，即 $P(X=a)=1$。

随机变量函数 $g(X)$ 的数学期望是 $E[g(X)]=\int_{-\infty}^{\infty}g(x)f(x)\mathrm{d}x$。由 $E(\cdot)$ 的线性性，有 $E[g_1(X)\pm g_2(X)]=E[g_1(X)]\pm E[g_2(X)]$，但一般而言，

$$E[g_1(X)\cdot g_2(X)]\neq E[g_1(X)]\cdot E[g_2(X)]$$
$$E[g_1(X)/g_2(X)]\neq E[g_1(X)]/E[g_2(X)]$$

其中 $g_2(X)\neq0$，$E[g_2(X)]\neq0$。

设函数 $y=g(x)$ 是严格单调函数，值域是 (c,d)，其反函数 $h(y)$ 有连续导函数 $h'(y)$，则随机变量函数 $g(X)$ 仍为随机变量，其密度函数是

$$f_Y(y)=\begin{cases}f_X(h(y))\cdot|h'(y)|, & c<y<d\\0, & \text{其他}\end{cases}$$

若 $X\sim N(\mu,\sigma^2)$，则对于 $a\neq0$，有 $(aX+b)\sim N(a\mu+b,a^2\sigma^2)$，即正态分布的线性变换仍为正态分布；$\mathrm{e}^X$ 的密度函数是

$$f_Y(y)=\begin{cases}\dfrac{1}{\sigma y}\phi\left(\dfrac{\ln y-\mu}{\sigma}\right), & y>0\\0, & y\leqslant0\end{cases}$$

这是对数正态分布的密度函数，其分布记为 $\mathrm{e}^X\sim LN(\mu,\sigma^2)$，即正态分布的指数变换为对

数正态分布。有些变量的分布不是正态分布，但近似于对数正态分布，如个体收入、绝缘材料寿命等随机变量。

除常用特征(数学期望和方差)外，随机变量还有其他重要的特征，如

k 阶矩 $E(X^k)=\int_{-\infty}^{\infty} x^k f(x)\mathrm{d}x$

k 阶中心矩 $E(X-EX)^k=\int_{-\infty}^{\infty}(x-EX)^k f(x)\mathrm{d}x$

变异系数 $C(X)\equiv\sqrt{\mathrm{Var}(X)}/E(X)$

p 分位数 x_p，其中 x_p 满足 $\int_{-\infty}^{x_p} f(x)\mathrm{d}x=p,\quad p\in(0,1)$

偏度系数 $\beta_s\equiv\dfrac{E(X-EX)^3}{[\mathrm{Var}(X)]^{3/2}}$

峰度系数 $\beta_k\equiv\dfrac{E(X-EX)^4}{[\mathrm{Var}(X)]^2}-3$

多维随机变量：设 Y 是另一个随机变量，则 X 和 Y 的**联合分布函数**是 $F(x,y)=P(X\leqslant x,Y\leqslant y)$，**联合密度函数**是 $f(x,y)=\partial^2 F(x,y)/\partial x\partial y$；$X$ 和 Y 的**边际密度函数**分别是 $f_X(x)=\int_{-\infty}^{\infty} f(x,y)\mathrm{d}y$ 和 $f_Y(y)=\int_{-\infty}^{\infty} f(x,y)\mathrm{d}x$。

对于 $f_Y(y)>0$ 的点 y，给定 $Y=y$ 的条件下 X 的**条件密度函数**是 $f(x|y)=f(x,y)/f_Y(y)$；同理，对于 $f_X(x)>0$ 的点 x，给定 $X=x$ 的条件下 Y 的**条件密度函数**是 $f(y|x)=f(x,y)/f_X(x)$。

X 和 Y 是**相互独立**的，如果 $F(x,y)=F_X(x)F_Y(y)$ 或 $f(x,y)=f_X(x)f_Y(y)$。若 X 和 Y 相互独立，则 $E(XY)=EX\cdot EY$ 及 $\mathrm{Var}(X\pm Y)=\mathrm{Var}(X)\pm\mathrm{Var}(Y)$。

设变换 $u=g_1(x,y)$，$v=g_2(x,y)$ 有连续偏导数，且存在唯一反函数 $x=x(u,v)$，$y=y(u,v)$，其变换的 Jacobi 行列式

$$J\equiv\frac{\partial(x,y)}{\partial(u,v)}\equiv\begin{vmatrix}\partial x/\partial u & \partial y/\partial u\\ \partial x/\partial v & \partial y/\partial v\end{vmatrix}=\begin{vmatrix}\partial u/\partial x & \partial u/\partial y\\ \partial v/\partial x & \partial v/\partial y\end{vmatrix}^{-1}\neq 0$$

则随机变量 X 和 Y 的变换 $U=g_1(X,Y)$，$V=g_2(X,Y)$ 的联合密度函数是

$$f(x(u,v),y(u,v))\,|J|$$

X 和 Y 之间相互关联的特征数是**协方差**，其定义为 X 的偏离 $X-EX$ 与 Y 的偏离 $Y-EY$ 乘积的数学期望：

$$\mathrm{Cov}(X,Y)\equiv E[(X-EX)(Y-EY)]=E(XY)-EX\cdot EY$$

协方差可正可负，当 $\mathrm{Cov}(X,Y)>0$ 时，称 X 和 Y 正相关；当 $\mathrm{Cov}(X,Y)<0$ 时，称 X 和 Y 负相关；当 $\mathrm{Cov}(X,Y)=0$ 时，称 X 和 Y 不相关。

关于 X 和 Y 的关联性，有下列结论：

(1) 若 X 和 Y 相互独立，则 $\mathrm{Cov}(X,Y)=0$，即 X 和 Y 不相关；反之不真。

(2) $\mathrm{Var}(X\pm Y)=\mathrm{Var}(X)+\mathrm{Var}(Y)\pm 2\mathrm{Cov}(X,Y)$

(3) $\mathrm{Cov}(X,Y)=\mathrm{Cov}(Y,X)$；$\mathrm{Cov}(aX+b,cY)=ac\mathrm{Cov}(X,Y)$，$a,b,c$ 为常数。

为消除协方差的量纲，引入相关系数。设 $\sigma_X^2\equiv\mathrm{Var}(X)>0$，$\sigma_Y^2\equiv\mathrm{Var}(Y)>0$，则 X 和

Y 的**相关系数**定义为

$$\rho \equiv \operatorname{Corr}(X,Y) \equiv \frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X)}\cdot\sqrt{\operatorname{Var}(Y)}} = \frac{\operatorname{Cov}(X,Y)}{\sigma_X\sigma_Y}$$

它是 X 和 Y 标准化变量 $X^* = \dfrac{X-EX}{\sigma_X}$ 和 $Y^* = \dfrac{Y-EY}{\sigma_Y}$ 的协方差，即

$$\operatorname{Corr}(X,Y) = \operatorname{Cov}(X^*,Y^*)$$

相关系数有两个基本性质：①$|\rho|\leqslant 1$；②$\rho=\pm 1$ 当且仅当 X 和 Y 间存在几乎处处线性关系，即存在常数 a,b 使 $P(Y=aX+b)=1$。

由此，相关系数刻画了随机变量之间线性关系的强弱，也称线性相关系数。若 $\rho=0$，则 X 和 Y 不相关，即 X 和 Y 之间没有线性关系，不过 X 和 Y 之间有可能存在其他非线性函数关系，如平方关系、对数关系等。若 $\rho=1$，则 X 和 Y 完全正相关；若 $\rho=-1$，则 X 和 Y 完全负相关。若 $0<|\rho|<1$，则 X 和 Y 有一定程度的线性关系，$|\rho|$ 越接近于 1，线性相关程度越强；$|\rho|$ 越接近于 0，线性相关程度越弱。

如果 X 和 Y 服从联合正态分布，即 $(X,Y)\sim N(\mu_x,\mu_y,\sigma_x^2,\sigma_y^2,\rho)$，其中 $\mu_x=E(X)$，$\mu_y=E(Y)$，$\sigma_x^2=\operatorname{Var}(X)$，$\sigma_y^2=\operatorname{Var}(Y)$，$\rho$ 为两者的相关系数，则 X 和 Y 的边际密度分别是 $f_X(x)=\dfrac{1}{\sigma_x}\phi\left(\dfrac{x-\mu_x}{\sigma_x}\right)$ 和 $f_Y(y)=\dfrac{1}{\sigma_y}\phi\left(\dfrac{y-\mu_y}{\sigma_y}\right)$，其中 $\phi(\cdot)$ 是标准正态分布的密度函数。给定 X 条件下 Y 的条件分布也是正态分布，且其条件密度函数是

$$f_{Y|X}(y \mid x) = \frac{f(x,y)}{f_X(x)} = \frac{1}{\sigma_y\sqrt{1-\rho^2}}\phi\left(\frac{y-(\mu_y+\rho\sigma_y\sigma_x^{-1}(x-\mu_x))}{\sigma_y\sqrt{1-\rho^2}}\right) \tag{1.1}$$

条件分布的数学期望称为**条件数学期望**。若 $E(|Y|)<\infty$，则给定 $X=x$ 条件下 Y 的**条件数学期望**定义为 x 的函数：

$$E(Y \mid x) \equiv \int_{-\infty}^{\infty} yf(y \mid x)\mathrm{d}y = \frac{\int_{-\infty}^{\infty} yf(x,y)\mathrm{d}y}{f_X(x)}$$

条件方差定义为 x 的函数：

$$\operatorname{Var}(Y \mid x) \equiv \int_{-\infty}^{\infty} (y-E(Y \mid x))^2 f(y \mid x)\mathrm{d}y = \frac{\int_{-\infty}^{\infty} (y-E(Y \mid x))^2 f(x,y)\mathrm{d}y}{f_X(x)}$$

与数学期望运算一样，$E(\cdot|X=x)$ 是线性运算：设 Y_1 和 Y_2 都是随机变量，则

$$E(aY_1+bY_2 \mid X=x) = aE(Y_1 \mid X=x) + bE(Y_2 \mid X=x)$$

或 $E(aY_1+bY_2|X)=aE(Y_1|X)+bE(Y_2|X)$。条件方差的运算也有与方差运算相类似的性质。

条件期望 $E(Y|X=x)$ 是 x 的函数，记为 $m(x)=E(Y|X=x)$。由此，$m(X)$ 是一个随机变量函数，表示 $E(Y|X)$，而 $E(Y|X=x)$ 看作是 $m(X)=E(Y|X)$ 在 $X=x$ 时的一个取值，也记为 $E(Y|x)$。我们总可以将随机变量 Y 写为它关于另一随机变量 X 的条件期望函数 $E(Y|X)$ 加上一个满足零条件均值假定的误差项 u，即有如下**简约式**关系成立：

$$Y = E(Y \mid X) + u, \quad 其中 E(u \mid X) = 0 \tag{1.2}$$

这是 Y 关于 X 的回归关系，$m(X)=E(Y|X)$ 也称为 Y 关于 X 的回归函数，u 称为随机扰

动项。在 $X=x$ 处 X 对 Y 的边际影响(marginal effect)或偏影响(partial effect)为 $m(x)$ 的偏导函数$\partial m(x)/\partial x$。如 $m(x)$具有线性参数形式 $m(x)=x'\beta$,则 X 对 Y 的边际影响为常数 β。

在什么情形下,Y 关于 X 具有线性回归关系呢?一个**充分条件**是:X 和 Y 服从联合正态分布$(X,Y)\sim N(\mu_x,\mu_y,\sigma_x^2,\sigma_y^2,\rho)$。由正态分布密度函数(1.1)的表达式,知,

$$E(Y\mid X)=\mu_y+\rho\sigma_y\sigma_x^{-1}(X-\mu_x)=\mu_y+\frac{\mathrm{Cov}(Y,X)}{\sigma_x^2}(X-\mu_x) \tag{1.3}$$

这说明,当 Y 和 X 不服从联合正态分布时,Y 关于 X 的回归关系有可能不是线性的;此时,如用线性关系 $Y=\beta_0+\beta_1 X+u$(其中 $E(u|X)=0$)来刻画 Y 关于 X 的回归关系,则它可能存在设定偏误。

$E(Y|X)$作为随机变量,其数学期望是

$$\begin{aligned}
E[E(Y\mid X)]&=\int_{\Omega_x}E(Y\mid x)f_X(x)\mathrm{d}x=\int_{\Omega_x}\frac{\int_{\Omega_y}yf(x,y)\mathrm{d}y}{f_X(x)}f_X(x)\mathrm{d}x\\
&=\int_{\Omega_x}\left[\int_{\Omega_y}yf(x,y)\mathrm{d}y\right]\mathrm{d}x=\int_{\Omega_y}\left[\int_{\Omega_x}yf(x,y)\mathrm{d}x\right]\mathrm{d}y\\
&=\int_{\Omega_y}y\left[\int_{\Omega_x}f(x,y)\mathrm{d}x\right]\mathrm{d}y=\int_{\Omega_y}yf_Y(y)\mathrm{d}y=EY
\end{aligned}$$

其中假设两个累次积分相等,故成立如下的**迭代期望公式(law of iterated expectations)**:

$$E[E(Y\mid X)]=EY \tag{1.4}$$

在实际中,有时计算取值于较大范围上的随机变量 Y 的均值 EY 较难,但可以换一种思维方式:找一个与 Y 有关的随机变量 X,用 X 的不同取值将大范围划分为若干个小区域,先计算(假设易求)小区域上 Y 的平均(即求条件期望),再对这些平均计算加权平均(即求期望),得到大范围上 Y 的平均 EY。例如,证明经典回归模型 $Y_i=X_i'\beta+\varepsilon_i$,$E[\varepsilon_i|X_1,\cdots,X_n]=0$ 参数 OLS 估计量 $\hat{\beta}$ 的无偏性时,我们先限定 X 区域,计算条件数学期望

$$\begin{aligned}
E[\hat{\beta}\mid X_1,\cdots,X_n]&=E\left[\left(\sum_{i=1}^n X_iX_i'\right)^{-1}\sum_{i=1}^n X_iY_i\,\middle|\,X_1,\cdots,X_n\right]\\
&=\beta+\left(\sum_{i=1}^n X_iX_i'\right)^{-1}\sum_{i=1}^n X_iE[\varepsilon_i\mid X_1,\cdots,X_n]=\beta
\end{aligned}$$

然后由迭代期望公式(1.4),$E(\hat{\beta})=E[E(\hat{\beta}|X_1,\cdots,X_n)]=E\beta=\beta$,证得 $\hat{\beta}$ 关于 β 的无偏性。

除上述**迭代期望原理**外,条件期望 $E(Y|X)$还满足如下一些性质(详见 Wooldridge 2002: Chapter 2):

(1) 设 W 为另一随机变量或向量,且 $X=f(W)$,则

$$E[E(Y\mid W)\mid X]=E[Y\mid X]$$
$$E[E(Y\mid X)\mid W]=E[Y\mid X]$$

此性质说明,较小信息的变量在多重条件期望中总起决定作用。其理解是:条件 $X=f(W)$意即知道 W 就知道 X,但知道 X 不一定知道 W,即 X 的信息比 W 的信息少;

结论 $E[E(Y|W)|X]=E[Y|X]=E[E(Y|X)|W]$是说,不论先 W 作条件的双重条件期望$E[E(Y|W)|X]$,还是先 X 作条件的双重条件期望 $E[E(Y|X)|W]$,它们都等于条件少的变量 X 为条件的单重条件期望$E[Y|X]$。

(2) 设 Z 为另一随机变量或向量,则

$$E[E(Y \mid X,Z) \mid X]= E[Y \mid X]$$

事实上,这是性质(1)的特例:令 $W=(X',Z')'$,即 X 是 W 的部分,X 的信息比 W 的少,则 X 起决定作用。可令 $f(W)=(I_K 0_{K \cdot L})W=X$,其中 K 和 L 分别是 X 和 Z 的维数,$\boldsymbol{I_K}$ 是 $\boldsymbol{K}$ 阶单位矩阵,$0_{\boldsymbol{K} \cdot \boldsymbol{L}}$是元素全 0 的 $\boldsymbol{K} \cdot \boldsymbol{L}$ 矩阵,再用性质(1)。

(3) 设 $f(x)$是 x 的函数,使 $E[Y|X]=g(f(X))$,则 $E[Y|f(X)]=E[Y|X]$。

此性质的一个特例是 $f(X)$具有参数单指数形式 $f(X)=X'\beta$。若 $E[Y \mid X]=g(X'\beta)$,则 $E[Y|X'\beta]=E[Y|X]$。例如,在二元选择模型、Tobit 模型的两步估计方法中将用到此性质(见第 5 章)。

(4) 若 $E(Y^2)<\infty$ 且 $E[Y \mid X]=m(X)$,则 $m(\cdot)$是函数最小化问题 $\min\limits_{\mu(\cdot)} E[(Y-\mu(X))^2]$的解。

此性质说明,回归函数按最小均方误准则总是可识别的,这是回归函数估计方法的基础。

1.2 随机变量序列及$O_p(\cdot)$和$o_p(\cdot)$的性质

设$\{X_n\}_{n=1}^{\infty}$为随机变量序列,称$\{X_n\}_{n=1}^{\infty}$**依概率收敛**于 X,记为$\underset{n\to\infty}{p\lim} X_n=X$ 或 $X_n \xrightarrow{p} X$,如果$\forall \varepsilon>0$,有$\lim\limits_{n\to\infty} P(|X_n-X|\geqslant\varepsilon)=0$。

按一贯做法,如果$\{X_n\}_{n=1}^{\infty}$是随机变量向量列,则它依概率收敛于向量 $\boldsymbol{X}$ 的定义与上面定义一样,只是将一维情形下的$|\cdot|$换成向量的模,仍用$|\cdot|$表示,即$|X_n-\boldsymbol{X}|\equiv[(X_n-\boldsymbol{X})'(X_n-\boldsymbol{X})]^{1/2}$。可以证明,$\{X_n\}_{n=1}^{\infty}$依概率收敛于向量 $\boldsymbol{X}$ 当且仅当它的每个分量序列分别依概率收敛于 X 的对应分量。如$\{X_n\}_{n=1}^{\infty}$为随机矩阵列,$X_n \xrightarrow{p} \boldsymbol{X}$ 的定义也一样,只是模改成$|X_n-\boldsymbol{X}|\equiv[\mathrm{tr}((X_n-\boldsymbol{X})'(X_n-\boldsymbol{X}))]^{1/2}$,其中 $\mathrm{tr}(\boldsymbol{A})$为矩阵 $\boldsymbol{A}$ 的迹。如注意$|\cdot|$在不同情形下的定义,以下阐述应没有异议。

Slutsky 定理:设$\underset{n\to\infty}{p\lim} X_n=\boldsymbol{X}$,$g(\cdot)$为 $\boldsymbol{X}$ 的连续函数,则

$$\underset{n\to\infty}{p\lim}\, g(X_n)= g(\underset{n\to\infty}{p\lim} X_n)= g(\boldsymbol{X})$$

即求概率极限运算和连续函数符号可交换顺序。

$O_p(\cdot)$和 $o_p(\cdot)$在估计量大样本性质推导中被广泛应用,现集中简述。

定义 假设$\{X_n\}$和$\{Y_n\}$是两个随机向量序列,

(1) 如果$\forall \varepsilon>0$,存在正数 M_ε 和正整数 N_ε,当 $n\geqslant N_\varepsilon$ 时,有 $P(|X_n|>M_\varepsilon)<\varepsilon$ 或 $P(|X_n|\leqslant M_\varepsilon)>1-\varepsilon$,则称序列$\{X_n\}$是**依概率有界的**,记为 $X_n=O_p(1)$。

(2) 如果 $X_n/Y_n=O_p(1)$,则称 $X_n=O_p(Y_n)$。

(3) 如果 $X_n \xrightarrow{p} 0$(即$\forall \delta>0$,有$\lim\limits_{n\to\infty} P(|X_n|>\delta)=0$),则称序列$\{X_n\}$是依概率收敛于 0(或**依概率无穷小量**),记为 $X_n=o_p(1)$。

(4) 如果 $X_n/Y_n=o_p(1)$，则称 $X_n=o_p(Y_n)$。

注 1：设 Z 是一随机变量，则 $Z=O_p(1)$，$\sqrt{n}Z=O_p(\sqrt{n})$ 及 $\sqrt{n}Z=o_p(n^\alpha)$，$\forall\alpha>0.5$。如果 $X_n=o_p(1)$，则有 $X_n=O_p(1)$，但反之不然。

注 2：与数列的无穷小量与有界量一样，依概率无穷小量 $o_p(\cdot)$ 和依概率有界量 $O_p(\cdot)$ 也有性质：两无穷小量之和仍为无穷小量；两有界量之和仍为有界量；无穷小量与有界量之和为有界量；无穷小量与有界量之积为无穷小量。具体地，设 a,b 为任意常数，则有

$$a\cdot o_p(\cdot)\pm b\cdot o_p(\cdot)=o_p(\cdot)$$
$$a\cdot O_p(\cdot)\pm b\cdot O_p(\cdot)=O_p(\cdot)$$
$$a\cdot o_p(\cdot)\pm b\cdot O_p(\cdot)=O_p(\cdot)$$
$$o_p(\cdot)\cdot O_p(\cdot)=o_p(\cdot)$$
$$o_p(\cdot)\cdot o_p(\cdot)=o_p(\cdot)$$
$$O_p(\cdot)\cdot O_p(\cdot)=O_p(\cdot)$$

对于依概率无穷小向量 $o_p(\cdot)$ 和依概率有界向量 $O_p(\cdot)$（甚至矩阵序列），上述相加及乘积的性质也成立。

注 3：若 $X_n=O(1)$（有界），则有 $X_n=O_p(1)$；但反之不然。例如，若 $\{X_n\}$ 是从 $N(0,1)$ 随机抽取的一个样本，则 $X_n\neq O(1)$，但 $X_n=O_p(1)$。事实上，从良好定义的累积分布[①]任意抽取一个随机样本 $\{X_n\}$，它都是一个 $O_p(1)$ 随机变量序列，这是因为，对任意 $M>0$，有

$$P(|X_n|>M)=P(X_n>M\ \text{或}\ X_n<-M)=1-F(M)+F(-M)\to 0(M\to\infty)$$

注 4：设 a_n 和 b_n 是非随机的非负数列，如果 $E|X_n|=O(a_n)$，则 $X_n=O_p(a_n)$；如果 $E|X_n|^2=O(b_n)$，则 $X_n=O_p(b_n^{1/2})$。对于前者的证明，由 $E|X_n|=O(a_n)$，则存在常数 $M_0>0$，使 $E|X_n/a_n|\leqslant M_0$，故由切比雪夫不等式知，对于任何 $M>0$，有

$$P(|X_n/a_n|>M)\leqslant E|X_n/a_n|/M\leqslant M_0/M$$

因此，对于任何 $\varepsilon>0$，取 $M=M_0/\varepsilon$，则 $P(|X_n/a_n|>M)\leqslant\varepsilon$，即 $X_n=O_p(a_n)$。对于后者的证明，由 $E|X_n|^2=O(b_n)$，存在 $M_0>0$，使得 $E|X_n^2/b_n|\leqslant M_0$。由切比雪夫不等式知，对于任何 $M>0$，

$$P(|X_n/b_n^{1/2}|>M)\leqslant E|X_n^2/b_n|/M^2\leqslant M_0/M^2$$

因此，对于任意 $\varepsilon>0$，取 $M=\sqrt{M_0/\varepsilon}$，则有 $P(|X_n/b_n^{1/2}|>M)\leqslant\varepsilon$，即 $X_n=O_p(b_n^{1/2})$。

注 5：对于非随机的非负数列 a_n 和 b_n，因 $\sqrt{a_n+b_n}\leqslant\sqrt{a_n}+\sqrt{b_n}\leqslant 2\sqrt{a_n+b_n}$，故有 $O(\sqrt{a_n+b_n})=O(\sqrt{a_n}+\sqrt{b_n})=O(\sqrt{a_n})+O(\sqrt{b_n})$。

上述注中的性质将广泛用于密度函数和回归函数非参数估计量大样本性质（见第 2、第 3 章）以及其他半参数估计量性质的证明中。

---

① 一个良好定义的 CDF，是指 $\lim\limits_{x\to\infty}F(x)=1$，$\lim\limits_{x\to-\infty}F(x)=0$，$F(x)$ 非降，且右连续。

1.3 收敛定理

本书中一般假设样本是随机样本，或者是相互独立样本，故本节仅以相互独立的随机变量（或向量）序列介绍大数定律和中心极限定理。

大数定理是联系总体指标和样本统计量的桥梁。设$\{X_i\}_{i=1}^{n}$是随机变量X所对应总体的一个随机样本，且均值EX、方差$\mathrm{Var}(X)$和k阶矩$E(X^k)$分别存在，则当$n\to\infty$时，分别有

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i\xrightarrow{p}EX\text{（大数定律）}$$

$$s_X^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2\xrightarrow{p}\mathrm{Var}(X)$$

$$a_{n,k}=\frac{1}{n}\sum_{i=1}^{n}X_i^k\xrightarrow{p}E(X^k)$$

上述收敛均为依概率收敛，左边是样本统计量（样本形式指标），右边是总体指标。因为上述概率极限成立，我们称样本均值$\overline{X}$、样本方差s_X^2和样本k阶矩$a_{n,k}$分别是总体均值EX、总体方差$\mathrm{Var}(X)$和总体k阶矩$E(X^k)$的**一致估计量（consistent estimator）**。上述利用样本矩作为总体矩的一致估计量的原理称为**样本类比原理（sample analog principle）**。

上述依概率收敛本质上是点收敛概念，意指随机变量序列收敛于一常数。很多情形下，我们需要随机变量序列收敛于一个具有已知极限分布的随机变量。由估计量的极限分布计算估计量标准误、构建置信区间、作统计推断等。这需要依分布收敛概念。序列$\{X_n\}_{n=1}^{\infty}$**依分布收敛**于一个连续型随机变量X，如果

$$\lim_{n\to\infty}F_n(x)=F(x)$$

其中$F_n(\cdot)$和$F(\cdot)$分别为X_n和连续型随机变量X的累积分布函数，记为$X_n\xrightarrow{d}X$。进一步，若此分布极限随机变量X服从正态分布，即$X\sim N(\mu,\sigma^2)$，则称序列X_n依分布收敛于正态分布，记为$X_n\xrightarrow{d}N(\mu,\sigma^2)$，称序列$X_n$为**渐近正态的（asymptotically normal）**。

显然，若$X_n\xrightarrow{d}X$，其中X是随机向量，则$X_n=O_p(1)$，即依分布收敛向量序列一定是依概率有界的。

连续映射定理：设$X_n\xrightarrow{d}X$，$g(\cdot)$为X的连续函数，则

$$g(X_n)\xrightarrow{d}g(X)$$

即求依分布收敛极限运算和连续函数符号可交换顺序。

中心极限定理：设$\{X_n\}_{n=1}^{\infty}$是独立同分布（i. i. d.）的随机向量序列，满足$E(X_n)=0$和$E(|X_n|^2)<\infty$，则$n^{-1/2}\sum_{i=1}^{n}X_i\xrightarrow{d}N(0,V)$，其中$V=E(X_iX_i')$。

在证明非参数估计量的渐近正态性时，常遇到上述和式的一般项序列具有下三角形

式(不一定同分布)情形,比如一般项是$\{Z_{n,i},i=1,2,\cdots,n;n=1,2,\cdots\}$,即

$$\begin{array}{llllll}
Z_{11} & & & & & \\
Z_{21} & Z_{22} & & & & \\
Z_{31} & Z_{32} & Z_{33} & & & \\
Z_{41} & Z_{42} & Z_{43} & Z_{44} & & \\
\vdots & \vdots & \vdots & \vdots & & \\
Z_{n1} & Z_{n2} & Z_{n3} & Z_{n4} & \cdots & Z_{nn} \\
\vdots & \vdots & \vdots & \vdots & \cdots & \ddots
\end{array}$$

Liapunov 中心极限定理:设$\{Z_{n,i}\}$是相互独立的随机向量序列,$i=1,2,\cdots,n$; $n=1,2,\cdots$满足$E(Z_{n,i})=\mu_{ni}$和$\mathrm{Var}(Z_{n,i})=\sigma_{ni}^2$,且对于某个$\delta>0$,有$E|Z_{n,i}|^{2+\delta}<\infty$。记$S_n=\sum_{i=1}^{n}Z_{n,i}$,$\sigma_n^2=\mathrm{Var}(S_n)=\sum_{i=1}^{n}\sigma_{ni}^2$。如果$\sigma_n^2=\sigma^2+o(1)$(其中$\sigma^2$是常量),且$\lim_{n\to\infty}\sum_{i=1}^{n}E|Z_{n,i}-\mu_{ni}|^{2+\delta}=0$,则当$n\to\infty$时,有

$$\sigma_n^{-1}(S_n-ES_n)=\sigma_n^{-1}\sum_{i=1}^{n}(Z_{n,i}-\mu_{n,i})\xrightarrow{d}N(0,1)$$

1.4 非参数与半参数估计的动因

本书涉及随机变量密度函数和随机变量之间的回归函数的估计。通常密度函数可由一些参数确定,例如,正态分布的密度函数由均值μ和方差σ^2两个参数而定:

$$f(x)=(2\pi\sigma^2)^{-1}\exp(-(x-\mu)^2/\sigma^2)$$

指数分布的密度函数由一个参数θ而定:

$$f(x)=\theta\exp(-\theta x)$$

Weibull 分布密度函数由两个参数θ和α而定:

$$f(x)=\alpha x^{\alpha-1}\cdot\theta\exp(-\theta x^{\alpha})$$

回归函数$g(x)\equiv E[Y|X=x]$也一样,如果其形式由参数完全确定,则它是参数回归函数,例如,线性回归函数$g(x)=x'\beta$由解释变量的系数β确定。由样本数据将这些参数估计出来,并根据参数估计对参数进行推断,这是**参数估计方法**。

参数估计方法事先设定具有未知参数的密度函数或回归函数形式,然后由变量的已观察信息估计这些参数,进而得到对分布或回归关系的刻画和推断。可以想象,参数估计结果的准确性依赖于函数参数形式设定的正确性;在对参数设定形式的正确性确定无疑的情况下,研究者所得的参数估计结果一般才是可信的。例如,由式(1.3)知,如果Y和X服从联合正态分布,则两者间的线性回归关系设定就不存在偏误;反之,就可能存在偏误。

不过,问题是,现实中我们并不知道或不能确定随机变量密度函数或随机变量间回归函数关系的参数形式,所以,参数方法必须通过人为的参数设定"中间环节"来实现对密度