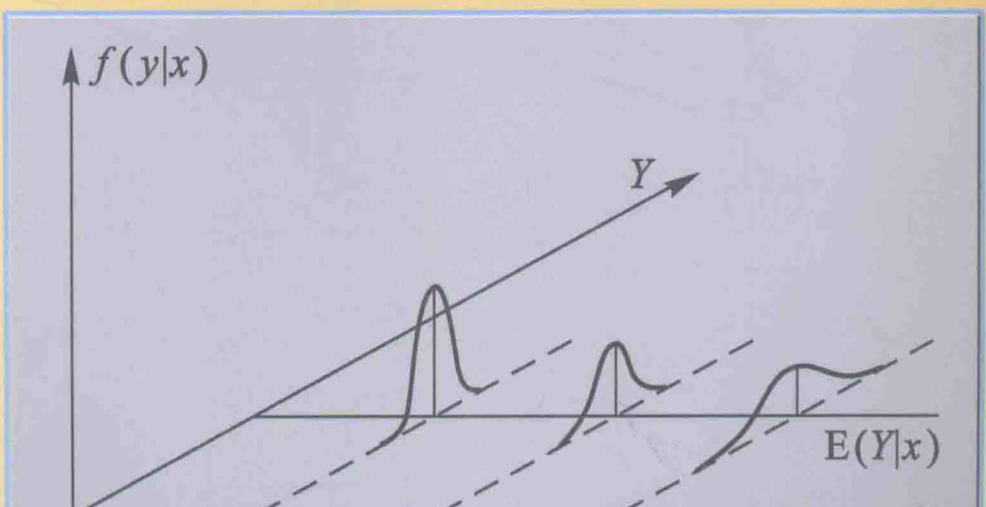
高等学校经济学类核心课程教材

高级计量经济学及Stata应用

陈　强　编著



高等教育出版社
HIGHER EDUCATION PRESS

高等学校经济学类核心课程教材

高级计量经济学及Stata应用

Gaoji Jiliang Jingjixue Ji Stata Yingyong

陈 强 编著

高等教育出版社·北京
HIGHER EDUCATION PRESS BEIJING

内容简介

本书较多地借鉴了现代计量经济学的最新发展，其内容全面，除了介绍传统的横截面数据外，对面板数据（含长面板、动态面板）、时间序列（含 VAR、单位根、协整）、自然实验、重复截面数据、GMM、蒙特卡罗法、自助法、分位数回归、门限回归、非参数估计、贝叶斯估计等方法均做了较深入的介绍。本书力图以生动的语言、较多的插图与经济意义来直观地解释计量方法，而又不失数学的严谨性。结合目前欧美最为流行的 Stata 计量软件，及时地介绍相应的 Stata 命令与实例，为读者提供“一站式”服务。

本书适合普通高等学校经济管理类或社科类硕士生、博士生与研究人员使用。为便于读者学习高级计量经济学，本书在内容安排上，假设读者已经学过微积分、线性代数与概率统计，但不要求学过本科阶段的计量经济学。

图书在版编目（CIP）数据

高级计量经济学及 Stata 应用 / 陈强编著. —北京：高等教育出版社，2010.10 （2011.8 重印）
ISBN 978-7-04-030181-6

Ⅰ.①高… Ⅱ.①陈… Ⅲ.①计量经济学-应用软件，Stata-高等学校-教材 Ⅳ.①F224.0-39

中国版本图书馆 CIP 数据核字（2010）第 172433 号

策划编辑 于 明　责任编辑 边晓娜　封面设计 杨立新　责任绘图 于 博
版式设计 王 莹　责任校对 王 超　责任印制 韩 刚

出版发行	高等教育出版社	咨询电话	400-810-0598
社　址	北京市西城区德外大街 4 号	网　址	http://www.hep.edu.cn
邮政编码	100120		http://www.hep.com.cn
印　刷	高等教育出版社印刷厂	网上订购	http://www.landraco.com
开　本	787×1092 1/16		http://www.landraco.com.cn
印　张	25.75	版　次	2010 年 10 月第 1 版
字　数	630 000	印　次	2011 年 8 月第 2 次印刷
购书热线	010-58581118	定　价	39.80 元

物 料 号　30181-00

前　言

本书是在山东大学经济学院硕士生、博士生《高级计量经济学》教案的基础上编著而成，适合高等学校经济管理类或社科类研究生与研究人员使用。

本书的主要特色如下：

（1）接轨现代计量经济学。本书较多地借鉴了 Baum (2006), Cameron and Trivedi (2005, 2009), Greene (2003), Hamilton (1994), Hayashi (2000), Kennedy (2003), Poirier(1995), Verbeek(2004), Wooldridge (2001)，其中尤以 Hayashi (2000)对本书的影响最深。

（2）内容全面。除了介绍传统的横截面数据外，本书对面板数据（含长面板、动态面板）、时间序列（含 VAR、单位根、协整）、自然实验、重复截面数据、GMM、蒙特卡罗法、自助法、分位数回归、门限回归、非参数估计、贝叶斯估计等方法均进行了较深入的介绍。

（3）计量理论与软件操作相结合。学习计量的学生，既需要了解计量原理，也需要知道如何在电脑上实现。为此，本书提供了“一站式”服务，在讲解每个估计方法后，随即介绍相应的 Stata 电脑操作及实例（Stata 为目前欧美最为流行的计量软件）。

（4）本书力图以生动的语言、较多的插图与经济意义来直观地解释计量方法，而不仅仅是从数学推导到数学推导；另一方面，又不失数学的严谨性（部分证明放在附录）。

（5）先修课不包括本科水平的计量经济学。在中国的国情下，不少经济类研究生并未学过本科阶段的计量经济学。因此，本书在内容安排上，假设读者已经学过微积分、线性代数与概率统计，但不要求学过本科阶段的计量经济学（当然，如果学过更好）。

学习计量经济学不是一件容易的事（我也经历过，以后还要经历），但回报却很丰厚（其为实证研究不可或缺之工具），可以说是“高投入、高产出”。对于许多初学者而言，或许计量经济学难就难在使用了较多的数学[①]。但数学只是一种语言，而任何数学符号原则上都可以“翻译”为汉语。事实上，看似复杂的数学公式后面，常常有着非常直观的道理。因此，只要渐渐地掌握数学这门语言，学会看数学符号背后的含义，学习计量也就不难了。

套用一个参禅的故事，学习计量大致可以分为三个境界。第一境界是“见山是山，见水是水”，第二境界是“见山不是山，见水不是水”，第三境界是“见山又是山，见水又是水”。在第一阶段，以为计量就是作最小二乘回归而已，自然不在话下。在第二阶段，开始体会到计量的精妙之处，心中时时产生疑问。在第三阶段，通过考前复习及实践应用，对计量的理论与方法逐渐融会贯通，进而内化为熟练掌握的工具。其中，尤以第二阶段最为漫长。“取法乎上，仅得其中”。貌似难懂之处，其实正是取得进步的地方。这是一个“痛并快乐着”的过程，时常伴有顿悟之喜悦。我曾为学生们写了一首打油诗，收录在此，以博一笑。

① 对于从理科转学经济学的同学来说，可能面临另一问题，即如何更快地建立经济学的直觉，加深对经济意义的理解。

计量啊计量

辛苦读研学计量，推来导去费思量。
只因成绩盼优良，折腾数据叫爹娘。
爱恨交加为那般，实证研究是桥梁。
此情可待成佳酿，奈何当下心已凉。

在本书出版之际，特别要感激以下曾教授过我统计学或计量经济学的授业恩师们（以时间先后为序）：范培华、胡健颖、靳云汇、陈良焜（北京大学）；Dale Poirier （University of California, Irvine）；Susan Porter-Hudak, Nader Ebrahimi, Mohsen Pourahmadi（Northern Illinois University）。没有他们的谆谆教诲，本书是绝不可能完成的。

山东大学经济学院的领导与同事们对本书的写作给予了大力支持与鼓励，另外，2008 级与 2009 级硕士与博士生在听课过程中提出了很多好建议，在此一并感谢。山东大学经济学院王永老师、博士生韩青、硕士生李欢、李晶、林兴兰、戚传萍、吴振华、张甜等参与了校对，陈丽云同学协助制作了部分插图，在此表示衷心感谢（当然，文责自负）。最后，要特别感谢高等教育出版社的于明编辑、边晓娜编辑及其同仁们，为本书的撰写提出了许多宝贵意见，并付出了辛勤的劳动。

正如苏格拉底所说，学习是学生自我发现的过程，而教师不过是助产婆，但愿这本教科书能起到这个作用。

当然，由于本人知识有限，对于本书中的错误与不足之处，恳请各位老师与同学及时指出，以便在本书的网站上公布勘误表，并在未来的版本中更新。邮箱为 qiang2chen2@gmail.com。

陈 强
2010 年 2 月于济南

目 录

第1章 绪 论

1

1.1 什么是计量经济学

顾名思义，“计量经济学”(Econometrics，也译为“经济计量学”)就是运用概率统计的方法对经济变量之间的(因果)关系进行定量分析的科学。之所以把“因果”两个字加括号，是因为计量经济学常常不足以确定经济变量之间的因果关系(由于实验数据的缺乏)①，另一方面，大多数实证分析的目的恰恰正是要确定变量之间的因果关系(即是否 X 导致 Y)，而非仅仅是相关关系②。因此，在学习与应用计量经济学的过程中，很有必要时时以“因果关系”作为思考的框架与指引。

比如，你看到街上人们带伞，于是预测今天要下雨。这是一种相关关系。然而，“人们带伞”并不是造成“下雨”的原因。因此，计量分析必须建立在经济理论的基础之上。然而，即使有理论基础，因果关系常常依然不好分辨。首先，可能存在“逆向因果关系”(reverse causality)。比如，FDI(外商直接投资)能促进经济增长，但也可能是 FDI 被吸引到增长潜力高的国家或地区。其次，也可能是被遗漏的第三个变量(Z)对这两个变量(X，Y)同时产生了作用，参见图 1.1。

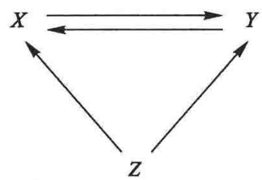


图 1.1 可能的因果关系

作为一个例子，考虑决定教育投资回报率(returns to schooling)的因素：

$$\ln W_i = \alpha + \beta S_i + \varepsilon_i \tag{1.1}$$

其中，“被解释变量”(dependent variable) $\ln W$ 为工资收入的自然对数，“解释变量”(explanatory variable or regressor) S 为受教育年限，ε 为“随机扰动项”(stochastic disturbance)或“误差项”(error term)，而下标 i 表示第 i 个观测值(即第 i 个个体)。

如果用数据估计这个简单的一元回归，其结果一般会显示，对数工资收入与受教育年限显著正相关，而且教育投资回报率 β 还挺高。然而，一个人的工资收入也与能力有关，但能

---

① 计量经济学家 G. Orcutt 曾说过，“做计量经济学就像试图通过播收音机来研究电的规律(Doing econometrics is like trying to learn the laws of electricity by playing the radio)”，可见其难度。

② 如果使用计量经济学做预测，则只需要相关关系，不必顾及因果关系。

力不能直接观测，而能力高的人通常选择接受更多教育。因此，在这个简单的回归中，教育的高回报率其实包含了对能力的回报。

另外，影响工资收入的因素还可能包括工作经验、毕业学校、人种、性别、外貌等。因此，需要尽可能多地引入“控制变量”(control variables)，也就是多元回归的方法，才能较准确地估计我们“感兴趣的参数”(parameters of interest)，即本例中的教育投资回报率 β。然而，现实中总有某些相关的变量无法观测，即存在“遗漏变量”(omitted variables)，而这些遗漏变量统统被纳入到随机扰动项 ε_i 中了。

随机扰动项 ε_i 中还可能包含哪些其他因素呢？如果真实模型(true model)为

$$\ln W_i = \alpha + \beta S_i + \gamma S_i^2 + \varepsilon_i \tag{1.2}$$

那么 γS_i^2 也被纳入到扰动项中了(可以视为广义的遗漏变量)。如果变量测量得不准确，则测量误差也被放入扰动项中了。总之，扰动项就像是一个“垃圾桶”，所有你不想要、无法把握的东西都往里面扔。另一方面，我们又希望扰动项有很好的性质。在很多情况下，这是自相矛盾的。西方有个谚语“The devil is in the details.”，意即“魔鬼就在细节中”[①]。套用到计量经济学上来，或许可以说“The devil is in the error term.”，意即魔鬼就在扰动项中。计量经济学的很多玄妙之处就在于扰动项。如果真正理解了扰动项，也就加深了对计量经济学的理解。

1.2 经济数据的特点与类型

由于在经济学中通常无法像自然科学那样做“控制实验”(controlled experiment)，故经济数据一般不是“实验数据”(experimental data)[②]，而是自然发生的“观测数据”(observational data)。由于个人行为的随机性，所有经济变量原则上都是随机变量[③]。

在计量经济学的本科课程中，为了简单起见，有时假设解释变量是非随机的、固定的。这只是为了教学法上的方便，却给更深入的理论探讨带来了不便。比如，如果解释变量为非随机，则无法考虑其与扰动项的相关性。因此，在这本研究生水平的教材中，所有变量都是随机的(即便非随机的常数，也可以视为退化的随机变量)。

经济数据按照其性质，可大致分成以下三种类型。

(1) 横截面数据(cross-sectional data，简称截面数据)：指的是多个经济个体的变量在同一时点上的取值。比如，2009 年中国各省的 GDP。

(2) 时间序列数据(time series data)：指的是某个经济个体的变量在不同时点上的取值。比如，在 1978—2009 年山东省每年的 GDP。

(3) 面板数据(panel data)：指的是多个经济个体的变量在不同时点上的取值。比如，在 1978—2009 年中国各省每年的 GDP。

本书介绍的计量经济理论将包括以上三种数据类型，并使用国际上最为流行的 Stata 10 计量软件[④]进行数据处理。在此之前，我们首先要对概率统计进行简短的回顾，并引入一些新概念(比如，均值独立、迭代期望定律)。

---

① 比如，对于一份冠冕堂皇的合同，可能让你以后吃尽苦头的正是那些合同附录中的小字部分。

② 第 15 章将讨论随机实验与自然实验数据。

③ 你能举出哪些经济数据(变量)不是随机变量吗？

④ Stata 11 已于 2009 年推出。鉴于 Stata 11 在国内仍未普及，本书将以 Stata 10 为主。

第 2 章　概率统计回顾

2.1　概率与条件概率

1. 概率

假如街上有个老太太问你：“什么是概率”，那么你会怎么回答呢？若回答：“事情发生的可能性”，老太太可能反问你：“说‘可能性’不就行了，为什么又造了一个新词‘概率’”。也许她会问你一个更具体的问题：“天气预报说明天 70%概率下雨。这是啥意思？”。也许你想说，这表明“明天 70%的时间会下雨”，但更好的答案则是：“如果有 100 天的天气预报都报了 70%的概率明天降雨，则大约有 70 天会下雨”。

总之，可以将“概率”理解为在大量重复实验下，事件发生的频率趋于的某个稳定值。记事件“下雨”为 A，其发生的“概率”(probability)为 $\mathrm{P}(A)$[①]。

2. 条件概率

例　已知明天会出太阳，下雨的概率有多大？

记事件“出太阳”为 B，则在出太阳的前提条件下降雨的“条件概率”(conditional probability)为

$$\mathrm{P}(A|B) \equiv \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)} \tag{2.1}$$

其中，“$\cap$”表示事件的交集(intersection)，故 $\mathrm{P}(A \cap B)$ 为“太阳雨”的概率，参见图 2.1。条件概率是计量经济学的重要概念之一。

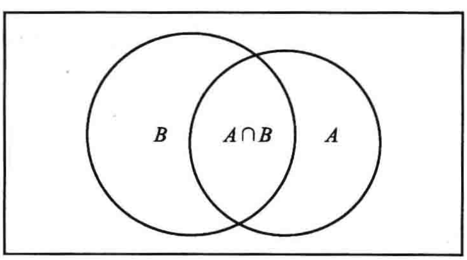


图 2.1　条件概率示意图

3. 独立事件

如果条件概率等于无条件概率，即 $\mathrm{P}(A|B) = \mathrm{P}(A)$，即 B 是否发生不影响 A 的发生，则称

① 这里不讨论对概率的严格公理化定义。

A, B 为相互独立的随机事件。此时，$\mathrm{P}(A|B) \equiv \dfrac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)} = \mathrm{P}(A)$，故

$$\mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(B) \tag{2.2}$$

也可以将此式作为独立事件的定义。

4．全概率公式

如果事件组 $\{B_1, B_2, \cdots, B_n\}$ $(n \geqslant 2)$ 两两互不相容，$P(B_i) > 0$ $(\forall i = 1, \cdots, n)$，且 $B_1 \cup B_2 \cup \cdots \cup B_n$ 为必然事件(即在 $B_1, B_2, \cdots, B_n$ 中必然有某个 B_i 发生，“$\cup$”表示事件的并集，union)，则对任何事件 A 都有(无论 A 与 $\{B_1, B_2, \cdots, B_n\}$ 是否有任何关系)，

$$\mathrm{P}(A) = \sum\nolimits_{i=1}^{n} \mathrm{P}(B_i)\mathrm{P}(A|B_i) \tag{2.3}$$

全概率公式把世界分成了 n 个可能的情形，再把每种情况下的条件概率“加权平均”而汇总成无条件概率(权重为每种情形发生的概率)。该公式有助于理解后面的迭代期望定律。

2.2 分布与条件分布

1．离散型概率分布

假设随机变量 X 的可能取值为 $\{x_1, x_2, \cdots, x_k, \cdots\}$，其对应的概率为 $\{p_1, p_2, \cdots, p_k, \cdots\}$，即 $p_k \equiv \mathrm{P}(X = x_k)$，则称 X 为离散型随机变量，其分布律可以表示为

$$\begin{array}{cccccc} X & x_1 & x_2 & \cdots & x_k & \cdots \\ p & p_1 & p_2 & \cdots & p_k & \cdots \end{array} \tag{2.4}$$

其中，$p_k \geqslant 0$，$\sum\limits_k p_k = 1$。常见的离散分布有“两点分布”(Bernoulli)、“二项分布”(Binomial)、“泊松分布”(Poisson)与“负二项分布”(Negative Binomial)等(参见标准的概率统计本科教材)。

2．连续型概率分布

连续型随机变量可以取任意实数，其“概率密度函数”(probability density function，简记pdf) $f(x)$ 满足，

(i) $f(x) \geqslant 0,\ \forall x$；

(ii) $\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = 1$；

(iii) X 落入区间 $[a,b]$ 的概率为 $P(a \leqslant X \leqslant b) = \int_a^b f(x)\,\mathrm{d}x$。

定义“累积分布函数”(cumulative distribution function，简记 cdf)为

$$F(x) \equiv \mathrm{P}(-\infty < X \leqslant x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t \tag{2.5}$$

其中，t 为积分变量。直观来看，$F(x)$ 度量的是，从 $-\infty$ 至 x 为止，概念密度函数 $f(t)$ 曲线下的面积。

3. 多维随机向量的概率分布

为了研究经济变量之间的关系，常需要同时考虑两个或多个随机变量，即“随机向量”(random vector)。二维连续型随机向量 (X, Y) 的“联合密度函数”(joint pdf) $f(x, y)$ 满足，

(i) $f(x, y) \geqslant 0, \ \forall x, y$；

(ii) $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x, y)\,\mathrm{d}x\,\mathrm{d}y = 1$；

(iii) (X, Y) 落入平面某区域 D 的概率为 $\mathrm{P}\{(X, Y) \in D\} = \iint_D f(x, y)\,\mathrm{d}x\,\mathrm{d}y$。

直观来看，二维随机向量的联合密度函数就像是一顶倒扣的草帽，而落入平面某区域 D 的概率就是这顶草帽下在区域 D 之上的体积。更一般地，n 维连续型随机向量 $(X_1, X_2, \cdots, X_n)$ 可以由联合密度函数 $f(x_1, x_2, \cdots, x_n)$ 来描述。

从二维联合密度函数 $f(x, y)$，可以计算 X 的(一维)边缘密度函数(marginal pdf):

$$f_x(x) = \int_{-\infty}^{+\infty} f(x, y)\,\mathrm{d}y \tag{2.6}$$

这个公式的直观含义与“全概率公式”相似，即给定 x，把所有 y 取值的可能性都“加”起来(积分的本质就是加总)。类似地，可以计算 Y 的(一维)边缘密度函数:

$$f_y(y) = \int_{-\infty}^{+\infty} f(x, y)\,\mathrm{d}x \tag{2.7}$$

定义二维随机向量 (X,Y) 的累积分布函数为

$$F(x, y) \equiv \mathrm{P}(-\infty < X \leqslant x; -\infty < Y \leqslant y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(t, s)\,\mathrm{d}t\,\mathrm{d}s \tag{2.8}$$

4. 条件分布

“条件分布”(conditional distribution)的概念对于计量经济学至关重要。考虑在 $X = x$ 条件下 Y 的条件分布，记为 $Y|X = x$。然而，如果 X 为连续型随机变量，事件 $\{X = x\}$ 发生的概率为 0，该如何计算 $Y|X = x$ 的“条件概率密度”(conditional pdf)呢？为此，考虑 $X \in [x-\varepsilon, x+\varepsilon]$，然后让 $\varepsilon \to 0^+$，可以证明条件密度函数为(参见附录)，

$$f(y \mid x) = \frac{f(x, y)}{f_x(x)} \tag{2.9}$$

直观上，这个公式与条件概率的公式十分类似。

2.3 随机变量的数字特征

虽然随机变量的密度函数或累积分布函数能够完整地描述随机变量，但我们常希望用少数几个常数来刻画其主要特征(称为“数字特征”)，比如该随机变量平均来说在什么位置(即平均值，或集中趋势)，波动幅度(即离散趋势)有多大，与其他变量是否存在“协动”(即相关性)。为此，下面引入期望、方差、协方差、相关系数、原点矩、中心矩、偏度、峰度、协方差矩阵、条件期望、条件方差等概念。

定义 对于分布律为 $p_k \equiv \mathrm{P}(X = x_k)$ 的离散型随机变量 X，其“**期望**”(expectation)[①]为

$$\mathrm{E}(X) \equiv \mu \equiv \sum_{k=1}^{\infty} x_k p_k \tag{2.10}$$

定义 对于概率密度函数为 $f(x)$ 的连续型随机变量 X，其“**期望**”为

$$\mathrm{E}(X) \equiv \mu \equiv \int_{-\infty}^{+\infty} x f(x)\,\mathrm{d}x \tag{2.11}$$

容易证明，对随机变量求期望的这种运算(即“期望算子”，expectation operator)满足“线性性”(linearity)，即 $\mathrm{E}(X+Y) = \mathrm{E}(X) + \mathrm{E}(Y)$，$\mathrm{E}(kX) = k\,\mathrm{E}(X)$，其中 k 为任意常数。

定义 随机变量 X 的“**方差**”(variance)为

$$\mathrm{Var}(X) \equiv \sigma^2 \equiv \mathrm{E}\big[X - \mathrm{E}(X)\big]^2 \tag{2.12}$$

方差越大则随机变量取值的波动幅度越大。称方差的算术平方根为“**标准差**”(standard deviation)，通常记为 σ。在计算方差时，常使用以下简便公式。

命题 $\mathrm{Var}(X) = \mathrm{E}[X^2] - \big[\mathrm{E}(X)\big]^2$

证明： $\mathrm{Var}(X) \equiv \mathrm{E}\big[X - \mathrm{E}(X)\big]^2 = \mathrm{E}\Big\{X^2 - 2\,\mathrm{E}(X)X + \big[\mathrm{E}(X)\big]^2\Big\}$

$$= \mathrm{E}[X^2] - 2\big[\mathrm{E}(X)\big]^2 + \big[\mathrm{E}(X)\big]^2 = \mathrm{E}[X^2] - \big[\mathrm{E}(X)\big]^2$$

我们常需要考虑两个变量之间的相关性，即一个随机变量的取值会对另一随机变量的取值有多大影响。

定义 随机变量 X 与 Y 的“**协方差**”(covariance)为

$$\mathrm{Cov}(X,Y) \equiv \sigma_{XY} \equiv \mathrm{E}\Big[\big(X - \mathrm{E}(X)\big)\big(Y - \mathrm{E}(Y)\big)\Big] \tag{2.13}$$

如果当随机变量 X 的取值大于(小于)其期望 $\mathrm{E}(X)$ 时，随机变量 Y 的取值也倾向于大于(小于)其期望值 $\mathrm{E}(Y)$，则 $\mathrm{Cov}(X,Y) > 0$，二者存在正相关；反之，如果当随机变量 X 的取值大于(小于)其期望 $\mathrm{E}(X)$ 时，随机变量 Y 的取值反而倾向于小于(大于)其期望值 $\mathrm{E}(Y)$，则 $\mathrm{Cov}(X,Y) < 0$，二者存在负相关。如果 $\mathrm{Cov}(X,Y) = 0$，则说明二者“线性不相关”，但不一定“相互独立”，因为二者还可能存在非线性的相关关系。在计算协方差时，常使用以下简便公式：

$$\begin{aligned}\mathrm{Cov}(X,Y) &\equiv \mathrm{E}\big[(X - \mathrm{E}(X))(Y - \mathrm{E}(Y))\big] \\ &= \mathrm{E}\big[XY - X\,\mathrm{E}(Y) - \mathrm{E}(X)Y + \mathrm{E}(X)\,\mathrm{E}(Y)\big] \\ &= \mathrm{E}(XY) - \mathrm{E}(X)\,\mathrm{E}(Y)\end{aligned} \tag{2.14}$$

协方差的运算也满足线性性，可以证明(参见习题)

$$\mathrm{Cov}(X, Y+Z) = \mathrm{Cov}(X,Y) + \mathrm{Cov}(X,Z) \tag{2.15}$$

协方差的一个缺点是，它受 X 与 Y 的计量单位的影响。为了将其标准化，引入相关系数的定义。

定义 随机变量 X 与 Y 的“**相关系数**”(correlation)为

$$\rho \equiv \mathrm{Corr}(X,Y) \equiv \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \tag{2.16}$$

可以证明，$-1 \leqslant \rho \leqslant 1$。需要注意的是，如果以上各定义式中的积分不收敛，则随机变量的数字特征可能不存在。更一般地，对于随机变量 X，可以定义一系列的数字特征，即各阶

① 也称为“数学期望”(mathematical expectation)或“均值”(mean)。

“矩”(moment)的概念。

定义 一阶原点矩为 $\mathrm{E}(X)$ (即期望)，二阶原点矩为 $\mathrm{E}(X^2)$，三阶原点矩为 $\mathrm{E}(X^3)$，四阶原点矩为 $\mathrm{E}(X^4)$，等等。

定义 二阶中心矩为 $\mathrm{E}\left[X-\mathrm{E}(X)\right]^2$ (即方差)，三阶中心矩为 $\mathrm{E}\left[X-\mathrm{E}(X)\right]^3$，四阶中心矩为 $\mathrm{E}\left[X-\mathrm{E}(X)\right]^4$，等等。

其中，一阶原点矩(期望)表示随机变量的平均值，二阶中心矩(方差)表示随机变量的波动程度，三阶中心矩表示随机变量密度函数的不对称性(偏度)，而四阶中心矩表示随机变量密度函数的最高处(山峰)有多“尖”及尾部有多“厚”(峰度)。然而，三、四阶中心矩还取决于变量的单位。为此，首先将变量“标准化”(即减去其期望 μ，再除以其标准差 σ)，并引入以下定义。

定义 随机变量 X 的“**偏度**”(skewness)为 $\mathrm{E}\left[(X-\mu)/\sigma\right]^3$。

显然，如果随机变量为对称分布(比如，正态分布)，则其偏度为 0(奇函数在关于原点对称的区间上积分为 0)。

定义 随机变量 X 的“**峰度**”(kurtosis)为 $\mathrm{E}\left[(X-\mu)/\sigma\right]^4$。

对于正态分布，其峰度为 3。如果随机变量 X 的峰度大于 3(比如 t 分布)，则其密度函数的最高处(山峰)比正态分布更“尖”，而两侧尾部则更“厚”(参见 2.6 节的图 2.3)。

定义 随机变量 X 的“**超额峰度**”(excess kurtosis)为 $\mathrm{E}\left[(X-\mu)/\sigma\right]^4-3$。

第 6 章将使用正态分布的偏度与峰度性质来检验某个分布是否为正态分布。

更一般地，对于任意函数 $g(\cdot)$，称随机变量函数的期望 $\mathrm{E}\left[g(X)\right]=\int_{-\infty}^{+\infty}g(x)f(x)\mathrm{d}x$ 为“**矩**”(moment)。

定义 “**条件期望**”(conditional expectation)就是条件分布 $Y|x$ 的期望，即

$$\mathrm{E}(Y\,|\,X=x)\equiv\mathrm{E}(Y\,|\,x)=\int_{-\infty}^{+\infty}yf(y\,|\,x)\mathrm{d}y \tag{2.17}$$

在上式中，由于 y 已被积分积掉，故 $\mathrm{E}(Y|x)$ 只是 x 的函数，参见图 2.2。

定义 “**条件方差**”(conditional variance)就是条件分布 $Y|x$ 的方差，即

$$\mathrm{Var}(Y\,|\,X=x)\equiv\mathrm{Var}(Y\,|\,x)=\int_{-\infty}^{+\infty}\left[y-\mathrm{E}(Y)\right]^2f(y\,|\,x)\mathrm{d}y \tag{2.18}$$

同样地，在上式中，y 已被积分积掉，故 $\mathrm{Var}(Y|x)$ 只是 x 的函数，参见图 2.2。

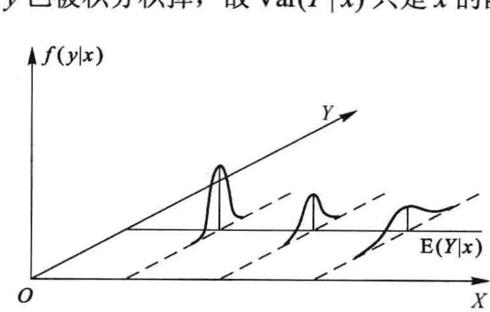


图 2.2 条件期望与条件方差示意图

为了引入随机向量的数字特征，首先需要回顾关于矩阵半正定与正定的概念。

定义 对于$n\times n$对称矩阵$\boldsymbol{A}$ (symmetric)，如果对于任意n维非零列向量$\boldsymbol{c}$，都有二次型$\boldsymbol{c}'\boldsymbol{A}\boldsymbol{c}\geqslant 0$，则称$\boldsymbol{A}$为**半正定矩阵**。

定义 对于$n\times n$对称矩阵$\boldsymbol{A}$，如果对于任意n维非零列向量$\boldsymbol{c}$，都有二次型$\boldsymbol{c}'\boldsymbol{A}\boldsymbol{c}>0$，则称$\boldsymbol{A}$为**正定矩阵**。

根据线性代数知识，正定矩阵的行列式一定不等于 0，故其逆矩阵一定存在。从几何意义上看，对于正定矩阵，可以通过坐标变换变为一个主对角线上元素全部为正数的对角矩阵(特征值全部为正)。特别地，在一维的情形下，正定矩阵就相当于正数。

类似地，可以定义半负定与负定矩阵。

命题 对于任意矩阵$\boldsymbol{D}$，$\boldsymbol{D}'\boldsymbol{D}$为半正定矩阵。

证明：首先，由于$\boldsymbol{D}'\boldsymbol{D}=(\boldsymbol{D}'\boldsymbol{D})'$，故$\boldsymbol{D}'\boldsymbol{D}$为对称矩阵。不失一般性，假设$\boldsymbol{D}'\boldsymbol{D}$为$n$级矩阵。其次，对于任意$n$维非零列向量$\boldsymbol{c}$，二次型

$$\boldsymbol{c}'(\boldsymbol{D}'\boldsymbol{D})\boldsymbol{c}=(\boldsymbol{c}'\boldsymbol{D}')(\boldsymbol{D}\boldsymbol{c})=\underbrace{(\boldsymbol{D}\boldsymbol{c})'\,\boldsymbol{D}\boldsymbol{c}}_{\text{平方和}}\geqslant 0 \tag{2.19}$$

因此，$\boldsymbol{D}'\boldsymbol{D}$为半正定矩阵。

定义 设$\boldsymbol{X}=(X_1\ X_2\cdots X_n)'$为$n$维随机向量，则其**“协方差矩阵”**(covariance matrix)为$n\times n$的对称半正定矩阵：

$$\begin{aligned}
\mathrm{Cov}(\boldsymbol{X})\equiv\mathrm{Var}(\boldsymbol{X})&\equiv\mathrm{E}\left[\big(\boldsymbol{X}-\mathrm{E}(\boldsymbol{X})\big)\big(\boldsymbol{X}-\mathrm{E}(\boldsymbol{X})\big)'\right]\\
&=\mathrm{E}\left[\begin{pmatrix}X_1-\mathrm{E}(X_1)\\ \vdots\\ X_n-\mathrm{E}(X_n)\end{pmatrix}\big(X_1-\mathrm{E}(X_1)\ \ \cdots\ \ X_n-\mathrm{E}(X_n)\big)\right]\\
&=\mathrm{E}\begin{pmatrix}[X_1-\mathrm{E}(X_1)]^2 & \cdots & [X_1-\mathrm{E}(X_1)][X_n-\mathrm{E}(X_n)]\\ \vdots & \ddots & \vdots\\ [X_1-\mathrm{E}(X_1)][X_n-\mathrm{E}(X_n)] & \cdots & [X_n-\mathrm{E}(X_n)]^2\end{pmatrix}\\
&=\begin{pmatrix}\sigma_{11} & \cdots & \sigma_{1n}\\ \vdots & \ddots & \vdots\\ \sigma_{n1} & \cdots & \sigma_{nn}\end{pmatrix}
\end{aligned} \tag{2.20}$$

其中，主对角线元素$\sigma_{ii}\equiv\mathrm{Var}(X_i)$，而非主对角线元素$\sigma_{ij}\equiv\mathrm{Cov}(X_i,X_j)$。

假设$\boldsymbol{A}$为$m\times n$常数矩阵(不含随机变量)，可以证明以下重要性质(参见习题)：

(i) $\mathrm{E}(\boldsymbol{A}\boldsymbol{X})=\boldsymbol{A}\,\mathrm{E}(\boldsymbol{X})$； (期望算子的线性性)

(ii) $\mathrm{Var}(\boldsymbol{X})=\mathrm{E}(\boldsymbol{X}\boldsymbol{X}')-\mathrm{E}(\boldsymbol{X})[\mathrm{E}(\boldsymbol{X})]'$； (一维随机变量公式的推广)

(iii) $\mathrm{Var}(\boldsymbol{A}\boldsymbol{X})=\boldsymbol{A}\,\mathrm{Var}(\boldsymbol{X})\boldsymbol{A}'$。

命题 n维随机向量$\boldsymbol{X}$的协方差矩阵$\mathrm{Var}(\boldsymbol{X})$为半正定矩阵。

证明：根据协方差矩阵的定义，$\mathrm{Var}(\boldsymbol{X})$为$n\times n$对称矩阵。对于$n$维非零列向量$\boldsymbol{c}$，随机变量$\boldsymbol{c}'\boldsymbol{X}$的方差必然大于或等于 0。因此，

$$\mathrm{Var}(\boldsymbol{c}'\boldsymbol{X})=\boldsymbol{c}'\,\mathrm{Var}(\boldsymbol{X})\boldsymbol{c}\geqslant 0 \tag{2.21}$$

根据定义，$\mathrm{Var}(\boldsymbol{X})$为半正定矩阵。进一步，如果希望任意线性组合$\boldsymbol{c}'\boldsymbol{X}$的方差为正数(不考

虑方差为 0 的退化情形)，则一般假设 $\mathrm{Var}(\boldsymbol{X})$ 为正定矩阵。

对于两个随机向量，也可以类似地考虑两个随机向量之间的协方差。

定义 设 $\boldsymbol{X}=(X_1\ X_2\ \cdots X_n)'$ 为 n 维随机向量，$\boldsymbol{Y}=(Y_1\ Y_2\ \cdots Y_m)'$ 为 m 维随机向量，则这两个随机向量之间的协方差矩阵为

$$\mathrm{Cov}(\boldsymbol{X},\boldsymbol{Y})\equiv \mathrm{E}\left[\left(\boldsymbol{X}-\mathrm{E}(\boldsymbol{X})\right)\left(\boldsymbol{Y}-\mathrm{E}(\boldsymbol{Y})\right)'\right]=\mathrm{E}(\boldsymbol{XY}')-\mathrm{E}(\boldsymbol{X})\,\mathrm{E}(\boldsymbol{Y}') \tag{2.22}$$

2.4 迭代期望定律

对于条件期望的运算，有以下重要的"迭代期望定律"(Law of iterated expectation)，

$$\mathrm{E}(Y)=\mathrm{E}_X\left[\mathrm{E}(Y\,|\,x)\right] \tag{2.23}$$

上式表明，无条件期望 $\mathrm{E}(Y)$ 等于，对于给定 $X=x$ 情况下 Y 的条件期望 $\mathrm{E}(Y\,|\,x)$ 再对 X 求期望。下面以连续型变量为例来证明。

证明： $\mathrm{LHS}=\mathrm{E}(Y)=\int_{-\infty}^{+\infty} yf_y(y)\,\mathrm{d}y$

$$\mathrm{RHS}=\mathrm{E}_X\left[\int_{-\infty}^{+\infty} y\frac{f(x,y)}{f_x(x)}\mathrm{d}y\right]=\int_{-\infty}^{+\infty}\left[\int_{-\infty}^{+\infty} y\frac{f(x,y)}{f_x(x)}\mathrm{d}y\right]f_x(x)\,\mathrm{d}x$$

$$=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} yf(x,y)\,\mathrm{d}x\,\mathrm{d}y=\int_{-\infty}^{+\infty} y\left[\int_{-\infty}^{+\infty} f(x,y)\,\mathrm{d}x\right]\mathrm{d}y=\mathrm{LHS}$$

其中，LHS 指"方程左边"(Left Hand Side)，而 RHS 指"方程右边"(Right Hand Side)。从上面的证明可以看出，迭代期望定律很像全概率公式。直观来看，无条件期望等于条件期望之加权平均，而权重为条件"$X=x$"的概率密度(取值可能性)。在离散随机变量的情形下，可以看得更为清楚：

$$\mathrm{E}(Y)=\sum_{x_i}\mathrm{P}(X=x_i)\,\mathrm{E}(Y\,|\,x_i) \tag{2.24}$$

推而广之，对于任意函数 $g(\cdot)$，可以得到

$$\mathrm{E}\left[g(Y)\right]=\mathrm{E}_X\,\mathrm{E}[g(Y)\,|\,x] \tag{2.25}$$

有时期望算子 E_X 的下标被省去，此时需注意对什么变量求期望。

2.5 随机变量无关的三个层次概念

定义 对于连续型随机变量 X 与 Y，如果其联合密度等于边缘密度的乘积，即 $f(x,y)=f_x(x)f_y(y)$，则称 X 与 Y 相互独立。

直观来看，如果 X 与 Y 相互独立，则 X 的取值不对 Y 的取值产生任何影响，反之亦然。这是有关随机变量"无关"的最强概念。线性不相关的概念则更弱，仅要求协方差为 0，即 $\mathrm{Cov}(X,Y)=0$。显然，"相互独立"意味着"线性不相关"，但反之不然。事实上，在二者之间还有一个中间层次的无关概念，即"均值独立"(mean-independent)，在计量经济学中很有用。

定义 假设条件期望 $\mathrm{E}(Y\,|\,x)$ 存在。如果 $\mathrm{E}(Y\,|\,x)$ 不依赖于 X，则称"Y **均值独立**于 X"

(Y is mean-independent of X)。

注意：均值独立不是一种对称的关系，即“Y均值独立于X”并不意味着“X均值独立于Y”。

命题 “Y均值独立于X”当且仅当$\mathrm{E}(Y\,|\,x)=\mathrm{E}(Y)$(即条件期望等于无条件期望)。

证明：(1)假设“Y均值独立于X”，则$\mathrm{E}(Y\,|\,x)$不依赖于X，故$\mathrm{E}_X\left[\mathrm{E}(Y\,|\,x)\right]=\mathrm{E}(Y\,|\,x)$。根据迭代期望定律，$\mathrm{E}(Y)=\mathrm{E}_X\left[\mathrm{E}(Y\,|\,x)\right]=\mathrm{E}(Y\,|\,x)$。

(2) 假设$\mathrm{E}(Y\,|\,x)=\mathrm{E}(Y)$，则显然$\mathrm{E}(Y\,|\,x)$不依赖于$X$。

命题 如果X与Y相互独立，则Y均值独立于X，且X均值独立于Y。

证明： X与Y相互独立意味着，X与Y一点关系也没有，证明参见习题。

命题 如果Y均值独立于X或X均值独立于Y，则$\mathrm{Cov}(X,Y)=0$。

证明：

$$\begin{aligned}
\mathrm{Cov}(X,Y)&=\mathrm{E}\big[\big(X-\mathrm{E}(X)\big)\big(Y-\mathrm{E}(Y)\big)\big] && \text{(协方差的定义)}\\
&=\mathrm{E}_X\,\mathrm{E}_Y\big[\big(X-\mathrm{E}(X)\big)\big(Y-\mathrm{E}(Y)\big)\big|x\big] && \text{(迭代期望定律)}\\
&=\mathrm{E}_X\big[\big(X-\mathrm{E}(X)\big)\mathrm{E}_Y\big(Y-\mathrm{E}(Y)\big|x\big)\big] && \text{(将 } X-\mathrm{E}(X) \text{ 视为常数提出)}\\
&=\mathrm{E}_X\big[\big(X-\mathrm{E}(X)\big)\big(\mathrm{E}(Y\,|\,x)-\mathrm{E}(Y)\big)\big] && \text{(期望算子的线性性)}\\
&=\mathrm{E}_X\big[\big(X-\mathrm{E}(X)\big)\cdot 0\big]=0 && \text{(均值独立的定义)}
\end{aligned}$$

总之，“相互独立”$\Rightarrow$“均值独立”$\Rightarrow$ “线性不相关”。

2.6 常用连续型统计分布

在计量经济学中常用的连续型统计分布包括正态分布、χ^2分布、t分布与F分布等。

1. 正态分布： 如果随机变量X的概率密度函数为

$$f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{\frac{-(x-\mu)^2}{2\sigma^2}\right\} \tag{2.26}$$

则称X服从正态分布，记为$X\sim N(\mu,\sigma^2)$，其中μ为期望，而σ^2为方差。将X进行标准化，定义$Z\equiv\dfrac{X-\mu}{\sigma}$，则$Z$服从标准正态分布，记为$Z\sim N(0,1)$。标准正态分布的概率密度以原点为对称，呈钟形(bell-shaped)(参见图 2.3)，通常记为$\phi(x)$；其累积分布函数则记为$\Phi(x)$。据说，当高斯(Gauss)发现标准正态分布的“核”(kernel) $\exp\left\{-x^2/2\right\}$时欣喜若狂，因为全世界只有一个标准正态分布，正态分布因此也叫“高斯分布”(Gaussian distribution)。

如果n维随机向量$\boldsymbol{X}=(X_1\ X_2\cdots X_n)'$的联合密度函数为

$$f(x_1,\cdots,x_n)=\frac{1}{(2\pi)^{n/2}\left|\boldsymbol{\Sigma}\right|^{1/2}}\exp\left\{-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})\right\} \tag{2.27}$$

则称$\boldsymbol{X}$服从期望为$\boldsymbol{\mu}$、协方差矩阵为$\boldsymbol{\Sigma}$的n维正态分布，记为$\boldsymbol{X}\sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})$。正态分布具有良好的性质。比如，多维正态的每个分量都是正态，其分量之任意线性组合仍然是正态。反之，每个分量均为一维正态并不足以保证其联合分布也是多维正态的。

命题 对于n个随机变量$X_1,X_2,\cdots,X_n$，如果任意线性组合$k_1X_1+\cdots+k_nX_n$(其中$k_1,\cdots,k_n$不全为0)都服从一维正态分布，则$(X_1\ X_2\cdots X_n)$服从n维正态分布。

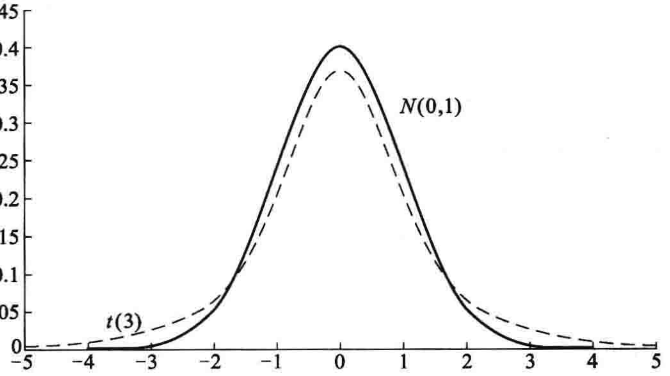


图 2.3　$N(0,1)$ 与 $t(3)$ 的概率密度

利用这个性质，可以证明如下命题。

命题　如果 $(X_1\ X_2 \cdots X_n)$ 服从 n 维正态分布，设 $Y_1, \cdots, Y_m$ 分别是 $(X_1\ X_2 \cdots X_n)$ 的线性函数，则 $(Y_1 \cdots Y_m)$ 也服从多维正态分布。

证明：考虑 $Y_1, \cdots, Y_m$ 的任意线性组合 $k_1 Y_1 + \cdots + k_m Y_m$，其中 $k_1, \cdots, k_m$ 不全为 0。由于 $Y_1, \cdots, Y_m$ 分别是 $(X_1\ X_2 \cdots X_n)$ 的线性函数，故 $k_1 Y_m + \cdots + k_m Y_m$ 也是 $(X_1\ X_2 \cdots X_n)$ 的线性函数。由于多维正态的线性组合仍为正态，故 $k_1 Y_m + \cdots + k_m Y_m$ 服从一维正态分布。根据 $k_1, \cdots, k_m$ 的任意性可知，$(Y_1 \cdots Y_m)$ 服从多维正态分布。

另外，如果 $(X_1\ X_2 \cdots X_n)$ 服从 n 维正态分布，则“$X_1, X_2, \cdots, X_n$ 相互独立”与“$X_1, X_2, \cdots, X_n$ 两两不相关”是等价的。换言之，对于正态分布，“不相关”就意味着“相互独立”。利用这个性质，有时可以很容易地证明正态变量的独立性。

2. χ^2 分布(卡方分布，Chi-square)：如果 $Z \sim N(0,1)$，则 $Z^2 \sim \chi^2(1)$，即自由度为 1 的 χ^2 分布。如果 $\{Z_1, \cdots, Z_k\}$ 为独立同分布的标准正态，则其平方和

$$\sum_{i=1}^{k} Z_i^2 \sim \chi^2(k) \tag{2.28}$$

其中，参数 k 为“自由度”(degree of freedom)。由于 χ^2 分布来自标准正态的平方和，故其取值只能为正数，参见图 2.4。

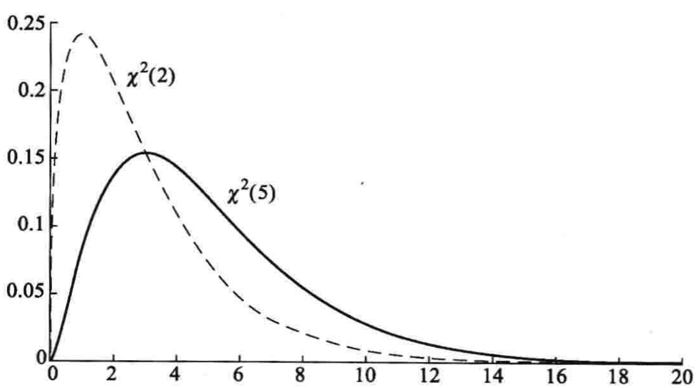


图 2.4　χ^2 分布的概率密度