



量化投资方法丛书

中国经图

刘振亚◎主编



经济计量 分析基础



刘振亚

—著—

Basic Econometrics



中国经济出版社
CHINA ECONOMIC PUBLISHING HOUSE



中国经图

量化投资方法丛书

刘振亚◎主编

014032653

F224.0
245



经济计量 分析基础

刘振亚
著



中国经济出版社
CHINA ECONOMIC PUBLISHING HOUSE



北航

C1720961

F224.0
245

图书在版编目 (CIP) 数据

经济计量分析基础 / 刘振亚著.

北京: 中国经济出版社, 2014. 1

ISBN 978 - 7 - 5136 - 2830 - 3

I. ①经… II. ①刘… III. ①经济计量分析 IV. ①F224. 0

中国版本图书馆 CIP 数据核字 (2013) 第 242794 号

选题策划 张晓楹
责任编辑 李博 李珂
责任审读 贺静
责任印制 张江虹
封面设计 任燕飞

出版发行 中国经济出版社
印刷者 三河市腾飞印务有限公司
经销者 各地新华书店
开本 800 × 1000mm 1/16
印张 15
字数 260 千字
版次 2014 年 1 月第 1 版
印次 2014 年 1 月第 1 次印刷
书号 978 - 7 - 5136 - 2830 - 3/F · 9905
定 价 36.00 元

中国经济出版社 网址 www.economyph.com 社址 北京市西城区百万庄北街 3 号 邮编 100037
本版图书如存在印装质量问题, 请与本社发行中心联系调换 (联系电话: 010 - 68319116)

版权所有 盗版必究 (举报电话: 010 - 68359418 010 - 68319282)

国家版权局反盗版举报中心 (举报电话: 12390)

服务热线: 010 - 68344225 88386794

总 序

量化投资与现代科技

量化投资被广泛应用在国际对冲基金中。在过去的三十年里，由于计算机技术和统计分析技术的进步，量化投资得到了迅猛的发展。

作为量化投资基金中的杰出代表，数学家西蒙斯（Jim Simons）所领导的复兴技术（Renaissance Technology）可谓独树一帜——他的大奖章基金在1988—2008年的20年时间里创造了年均收益35.6%的奇迹。1958年，20岁的西蒙斯从麻省理工大学数学专业本科毕业后，转入加州大学伯克利分校攻读数学博士，1961年博士毕业后回母校麻省理工大学任教。一年后，他跳槽到哈佛大学任教，1964年进入美国国防分析研究院工作。1967年，西蒙斯出任纽约大学石溪分校数学系系主任。在此期间，他与著名华裔数学家陈省身合作，创造了著名的陈-西蒙斯理论。1976年，他获得美国数学学会的威布伦奖。1978年，他离开石溪大学，成为职业投资人。1988年3月，西蒙斯成立复兴技术公司。

除西蒙斯外，复兴技术的三位元老：Leonard Baum、Henry Laufer 和 James Ax 都是一流的数学家，对复兴技术的长期发展产生了很大的影响——Baum 是西蒙斯在国防分析研究院的同事，统计学中著名的 Baum - Welsh 算法的发明者，该算法被广泛应用于隐蔽的马尔科夫链、语音识别、生物和应用统计中；James Ax 也曾任石溪大学数学系主任，在数论和几何学造诣颇深，曾在1967年获得美国数学学会数论方面的科伦奖；Henry Laufer 也曾是普林斯顿大学数学教授，退休后在石溪大学创办了量化投资专业，专业课程包括：概率论与统计方法、线性规划、线性几何、数据分析、随机微积分、金融计量、最优化算法、资产定价、投资组合、金融市场、衍生品定价、固定收益产品等。

当然，成功的量化投资基金，除数学家西蒙斯所领导的复兴技术外，还有计算机教授

肖尔 (David Shaw) 领导的肖尔公司, 物理学家哈丁 (David Harding) 领导的元胜资本 (Winton Capital), 经济学家格里芬 (Kenneth Griffin) 领导的大本营投资 (Citedal) 等。

一些国际知名的对冲基金对人才的要求也可归纳为以下内容: 很强的电脑编程能力 (C, C++; API, FIX; R, Matlab, SAS 等); 很强的数学或统计学分析技能 (线性和非线性时序分析, 数据挖掘, 隐蔽马尔科夫链, 随机分析等); 很强的大型数据库处理能力; 对衍生品、资产定价、市场微结构等有深入了解。

由此可以看出, 量化投资是现代金融理论、现代数理方法和现代信息技术的综合体。所涉及的金融理论主要包括: 资产定价, 投资组合, 衍生品定价, 市场微结构, 行为金融学等; 信息技术主要包括: 编程技术 (C, C++), 数据库技术, 交易底层通讯技术 (API/FIX); 数理方法主要包括: 经济计量分析基础, 线性和非线性时序分析, 数据挖掘, 隐蔽马尔科夫链, 随机分析等。

现代金融理论为量化投资提供了科学的理论基础。资产定价、衍生品定价、行为金融学和 市场微结构是量化投资策略设计的科学基础; 资产组合理论是量化投资降低和控制风险的有利工具。

现代信息技术和计算机技术为量化投资提供了坚实可靠的工具。软件工程和数据库技术是量化投资策略程序化的基础; API/FIX 是交易底层自动化的基础。以上两者的结合, 使得量化投资能够实现完全自动化交易。有人曾说, 即使西蒙斯将其所用策略公诸于世, 能够把公式变成钱的人在全球范围仍然是屈指可数。可见电脑技术和通讯技术在量化投资中的重要作用。

数理方法是量化投资最为核心的部分, 也是数据分析和交易策略设计以及评估的基础。其所涉及的领域也是五花八门、仁者见仁、智者见智, 主要包括: 经济计量分析基础, 线性和非线性时序分析, 数据挖掘, 隐蔽马尔科夫链, 随机分析, 小波分析等。西蒙斯曾经说过, 如果抛弃了数学模型, 将对我们没有任何好处。

在过去近十年的时间里, 作者本人作为摩根大通期货公司 (JP Morgan Futures Co.) 的董事, 有机会接触到一些国际知名的对冲基金, 并对其进行深入的研究和合作, 元胜资本 (Winton Capital) 就是其中的一个。作为全球最大的管理期货 (期货对冲基金), 元胜资本管理着约 300 亿美金的资产, 在其创始人科学家哈丁先生的领导下, 在 1987 年至今这 25 年的时间里创造了 16%—17% 的年均收益。仅从年均收益的数字来看, 元胜资本不如复兴技术可观, 但大家知道, 复兴技术的大奖章基金的规模只有约 50 亿美元, 而元胜资本的规模则接近 300 亿美元。



元胜资本现有员工 200 多人，近 50% 为研究人员，远离热闹非凡的一线业务，专心从事研究工作。若简单的按照国际对冲基金 2%—20% 的收费标准，元胜资本一线人员创造了人均产出近亿元人民币的奇迹。可见，当今现代科技与金融的结合能够创造出巨大的财富。量化投资行业绝对是高科技产业中的佼佼者。

在过去的三十多年里，我一直从事计量经济方法和信息技术方面的学习和研究，本系列丛书也算是我和我的学生们过去多年学习和研究的小结。这里，首先要感谢我的中国老师们：中国人民大学财金学院的黄达教授（我的博士生导师，前校长），陈共教授（前财政系主任），王传伦教授，经济学院的杜厚文教授（前国际经济系主任，前副校长），王景新教授（我的硕士导师），信息学院的陈禹教授（前院长），方美琪教授（我的本科导师），魏权龄教授，张怡兰教授，严颖教授，北京大学数学系的王萼芳教授。感谢我的美国老师们：George Horwich 教授（美国普渡大学，我的博士后导师），Roger Gordon 教授夫妇（加州大学圣地亚哥分校），Mark Machina 教授（加州大学圣地亚哥分校），Kajal Lahiri 教授（纽约大学阿尔巴尼分校），Harold Watts 教授（哥伦比亚大学）。他们的教育使我领略了数学的美妙、计算机的精巧、计量经济的严谨以及金融市场的奥妙，为日后的量化投资研究奠定了坚实的金融、经济、数理和计算机基础。

非常感谢中国人民大学校长陈雨露教授，金融与证券研究所所长吴晓求教授（校长助理），财金学院院长郭庆旺教授，副院长赵锡军教授，他们多年的友情、宽容和理解使得我能够有充分的自由和时间，在中国和英国两地从事我所感兴趣的研究项目。我也非常感谢我的伯明翰大学商学院同事们：David Dickinson（前院长），Nicholas Horsewood，Robert Elliott，Toby Kendall，Alessandra Guariglia 和 William Pouliot。

感谢摩根大通期货公司董事长周小雄先生多年来无微不至的兄长般关照。感谢元胜资本的创始人 David Harding，元胜亚洲 CEO Charles Allard，Kurt Settle 和田野先生。与元胜资本的合作，尤其是元胜资本每年的年会使我眼界大开，受益匪浅。感谢 AHL——牛津量化金融研究院（AHL - Oxford Institute）使我有机会参加他们所举办的世界一流水平有关量化投资的研讨会。

最后，感谢我在中国人民大学和英国伯明翰大学所指导的学生们，尤其是中国人民大学财金学院的博士生、硕士生和实验班的学生们。他们根据我的讲稿帮助整理了这套丛书中的很多内容。尤其是我的博士生杨武（现任教于中央财经大学），唐涛（现任职于中国人民银行研究所），罗涛（现任职于国家审计署），邓磊（现任职于北京工商大学），陈宇，张庆雪，刘琳，葛静，李伟，还有我在伯明翰大学的博士生曹瑞玫，汪仕炫，张昆

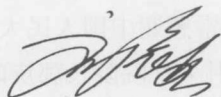
(阿斯顿大学),他们对量化投资的深入研究督促着我不断学习新的知识。这些学生们对新知识的追求,使得我多年来不敢懈怠,不断学习。从他们身上,我看到了中国量化投资业的未来。

出版量化投资系列丛书的主要目的是为国内的投资者系统地介绍有关量化投资理论、技术和方法,国际上成功的量化对冲基金公司,以及量化投资的研究成果和量化投资产业的发展动向。本系列丛书内容主要包括:经济计量分析基础,投资组合理论与实践,时序分析与神经网络,金融数据挖掘,解密复兴科技,随机分析,小波分析等。

我深切地希望,此套丛书能够为中国的证券、基金、期货、私募以及个人投资者提高量化投资水平起到抛砖引玉的作用。

写这个总序前后花了我三、四个月的时间,从北京写到香港,从香港写到英国,从英国写到法国,最后又从法国写回北京。每次提笔都是千言万语涌上心头,本人最大的心愿就是在未来20到30年时间里,中国能够出现复兴技术和元胜资本这样世界一流的对冲基金。

为此,我愿奉献一生!



2013年9月于北京中国人民大学

前 言

在量化投资策略设计中，基本的经济计量分析方法占有举足轻重的地位。经济计量分析方法以数理统计为基础、数学方法为手段、经济理论为指导，考察经济社会中各变量之间的相互量化关系。由于经济计量分析方法具有很大的实用性，国际各大投资银行和对冲基金都不同程度地采用经济计量方法来预测金融市场的未来趋势，设计投资策略，并检验投资策略的有效性。

经济计量学的内容随时间的推进而日新月异，尤其是 20 世纪 80 年代以来，由于计算机技术的飞跃发展、个人计算机的普及和计算成本的急剧降低，人们开始更多地强调经济计量分析的应用问题，量化投资在短短的二十多年时间里，得到了突飞猛进的进步。

目前，我国有关经济计量的书籍很多是举经济学的例子，很少涉及投资领域。更为突出的问题是，由于大多数经济计量分析书籍偏重于数理推导，以数学语言代替经济语言的地方很多，这对我国大部分证券期货分析研究人员而言，学习起来十分吃力。

作者几十年来一直从事经济计量的教学和研究工作。在本书写作的过程中，作者参考了国内外大量的书刊，并与国内外专家进行了多次探讨和研究，力图在运用最少的数学语言同时，更多地强调以文字形式对基本经济计量方法加以描述证明，以便使更多的读者能够了解和应用经济计量方法。

本书在吸收了国内外大量有关经济计量著作的基础上，系统地介绍了经济计量方法及其在量化投资中的应用。在结构安排上，本书可分为以下四部分：第一部分为经济计量的数理统计基础（第一章），第二部分为经济计量分析方法（第二至十二章），第三部分为时间序列软件包 EViews（第九章），第四部分详细介绍了经济计量方法在量化投资中的应用（第十三章）。

此外，为了更好地把理论和实践相结合，本书还加以大量例子配合证明，在内容的取舍上也注重吸收国际上较为成熟的最新成果。尤其是在最后一章，在系统地介绍了经济计量分析的基础上，结合国际著名的《金融经济学杂志》2011 年刊登的芝加哥大学商学院

T. J. Moskowitz 教授、对冲基金 AQR 资本管理公司的 Yao Hua Ooi 以及纽约大学商学院 Lasse Heje Pederson 教授的《时序动量》一文（获 2012 年度对冲基金白盒（Whitebox）研究奖），基于上证指数来阐述基本的经济计量分析方法在量化投资中的使用。

众所周知，一本数十万字的著作，它的写作过程并不是一个单纯的写作过程，期间还有大量繁重细致的劳动。从本书的第一章到最后一章，我的学生于晓筠博士（美国印地安那大学），杨武博士（中央财经大学），唐涛博士（中国人民银行研究所）等在资料收集和整理等诸多方面给了我很多的帮助，他们的友谊和帮助令我终生难忘。

希望本书能够使更多的读者应用经济计量方法去解决量化投资中的实际问题，进一步丰富量化投资的研究内容，加深研究深度，使中国的量化投资业早日腾飞。

最后应指出，由于作者水平所限，本书的缺点与错误在所难免，恳请广大读者批评指正。



于英国伯明翰大学 J. G. Smith 楼

2013 年 10 月

目 录

第一章 经济计量分析的统计基础	1
第一节 总体与样本	3
第二节 总体与样本的桥梁——随机变量	6
第三节 对总体的描述——随机变量的数字特征	11
第四节 对样本的描述——样本分布的数字特征	14
第五节 随机变量的分布——总体和样本的连接点	15
第六节 通过样本，估计总体（一）——估计量的特征	22
第七节 通过样本，估计总体（二）——估计方法	28
第八节 通过样本，估计总体（三）——假设检验	35
第九节 应用实例：最佳持股周期	42
第二章 数据与模型	44
第一节 数据、模型和变量	44
第二节 方程形式	46
第三节 相关系数	48
第三章 最小二乘法：一元线性回归	51
第一节 拟合直线性质	53
第二节 拟合优度的评价	56
第三节 一元线性回归中的计算问题	61
第四节 应用案例：对资产定价模型（CAPM）的简单检验	62



第四章 最小二乘法与多因素模型——含多个自变量的最小二乘法	66
第一节 含两个自变量的最小二乘法	66
第二节 二元回归最小二乘法解法又探	68
第三节 从另一个角度看估计系数	73
第四节 多重共线性	74
第五节 一般线性回归	75
第五章 古典线性回归模型	77
第一节 有关 ε_i 的古典模型假设	78
第二节 古典假设的一些内涵	81
第三节 高斯-马尔科夫定理	86
第四节 高斯-马尔科夫定理再探	91
第六章 正态条件下回归的推论	101
第一节 问题的引入	102
第二节 问题的预解决	102
第三节 派生的内容：自由度	103
第四节 回归系数的假设检验	105
第五节 预测	106
第六节 复习与提高	109
第七章 假设检验	110
第一节 t 检验：对单个参数检验的方法	110
第二节 F 检验的引入	112
第三节 一般假设检验： F 检验	113
第四节 F 检验的附加例子	115



第五节	调整后的相关系数 \bar{R}^2	117
第六节	因果关系检验	118
第七节	复习与提高	122
第八章	虚拟变量	124
第一节	运用虚拟变量改变回归直线的截距	126
第二节	运用虚拟变量改变回归直线的斜率	127
第三节	运用虚拟变量同时改变回归直线的斜率和截距	128
第四节	折线回归	129
第五节	复习与提高	131
第九章	计量经济学软件 EViews 简介	135
第一节	关于计算机的一些基本概念	135
第二节	EViews 概述	137
第三节	工作文件的建立及相关操作	140
第四节	基本回归模型	146
第五节	复习与提高	147
第十章	异方差	148
第一节	异方差概述	150
第二节	如何发现异方差	151
第三节	异方差的后果	156
第四节	异方差的解决方法	160
第十一章	自相关	168
第一节	自相关概述	168
第二节	Durbin-Watson 检验	169



第三节	自相关的处理方法	170
第四节	自相关误差下的回归过程	175
第五节	一阶自相关对最小二乘估计量的影响	178
第六节	对含滞后因变量的序列相关模型的检验	181
第七节	DW 检验的更为深入的结论	182
第八节	DW 显著时的策略	182
第十二章	联立方程模型	185
第一节	概 述	185
第二节	内生变量和外生变量	188
第三节	间接最小二乘法	188
第四节	识别问题	190
第五节	识别的充要条件	192
第六节	估计方法	194
第七节	运用最小二乘法估计联立方程组问题	202
第八节	复习与提高	204
第十三章	基于时序动量回报率 (Time - series Momentum Return) 的量化 交易策略	206
附表	统计表	211
参考书目	227



第一章

经济计量分析的统计基础

数理统计是进行经济计量分析的基本方法，即便是最为简单和实用的经济计量分析也涉及到诸多数理统计的基本概念。同时，数理统计在大学教学中属比较困难的部分。鉴于此，在进一步讨论如何进行经济计量分析之前，我们将首先对本书中经常使用到的数理统计的基本内容进行一些回顾。

事实上，不懂得数理统计就不可能学习和研究经济计量分析，数理统计是经济计量分析的基础，它为经济计量分析提供了唯一而有效的方法。这一点，读者们将在今后各章节的学习中深刻地体会到。从某种意义上来说，经济计量分析就是使数理统计在经济和投资中得以应用的一门学科。正因为如此，在世界各权威性的经济计量分析著作中，对数理统计基础内容的回顾和复习往往占据极大的篇幅，并以独立、完整而重要的章节出现。这一点从 William H. Greene 所著的《Econometric analysis》一书中可见一斑。

本书将以一个整章的篇幅对数理统计的基础原理给出一个概略的复习和回顾。应当指出的是，这一回顾是“概略的”，但不是“肤浅的”。通过本章的学习，读者将能从逻辑结构的层次上，比较完整地掌握数理统计的基本内容。无论读者事先学过还是没有学过数理统计，本章内容都会很有帮助，这不仅由于本章内容完整而自成体系，而且因为它注重于逻辑结构分析，注重于方法的阐述，注重于公式、定义和定理的内在涵义及其相互关系的揭示，同时，也注意了紧扣经济计量分析这一主题。

为了把注意力集中在数理统计的逻辑结构及其在量化投资研究的应用上，而不是数学推导的具体细节上，并且鉴于这不是一部数学书，我们对公式、定理往往强调它们的结论、涵义和它所给出的方法，而不强调定理、公式本身的数学证明。

在本章的讨论展开之前，首先给出这部分内容的逻辑结构。读者可以在以后各节的学习中随时参阅此表，以便了解自己所阅读的部分在整个数理统计体系中所处的地位，及其与其它部分的联系。在本章结束时，将再次仔细地讨论此表，并以此作为全章小结。



数理统计学的逻辑结构

1. 总体 (Population) 和样本 (Sample)

如何用一个变量来描述总体——随机变量 (Random Variable) 的引入。

2. 对总体的描述: 随机变量的数字特征

数学期望 $\mu_x = E(X)$ 描述总体的一般水平

方差 $\sigma_x^2 = Var(X)$ 描述总体的离散程度

3. 对样本的描述: 样本分布的数字特征

样本平均数 \bar{X} 描述样本的一般水平

样本方差 S^2 描述样本的离散程度

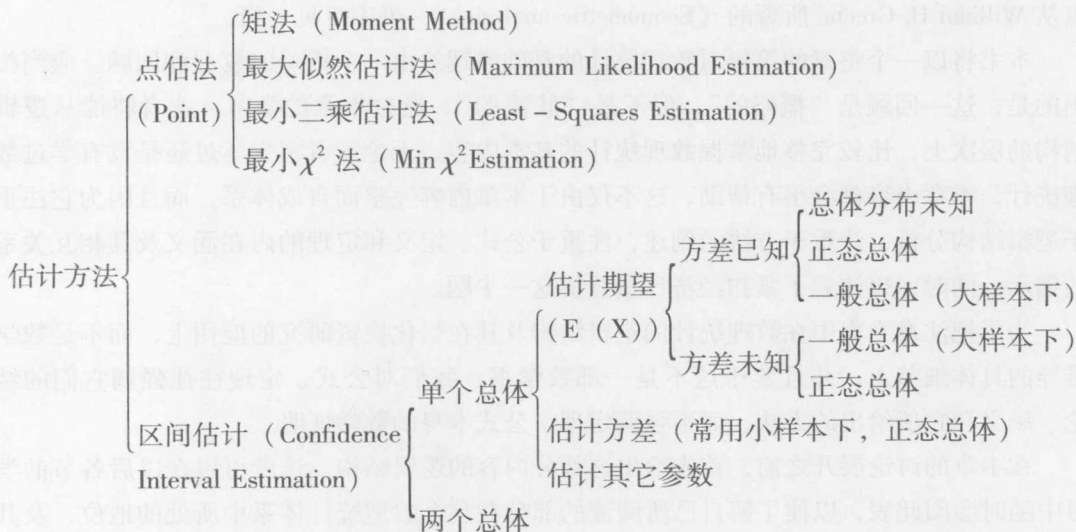
实际上, \bar{X} 和 S^2 是 $\mu_x = E(X)$ 和 $\sigma_x^2 = Var(X)$ 的无偏估计量。

4. 总体与样本的连接点: 随机变量的分布 (Distribution)

5. 通过样本数据和样本分布特征来估计总体的数字特征, 即 μ_x 和 σ_x^2 , 以及总体中数据生成过程的各种参数。

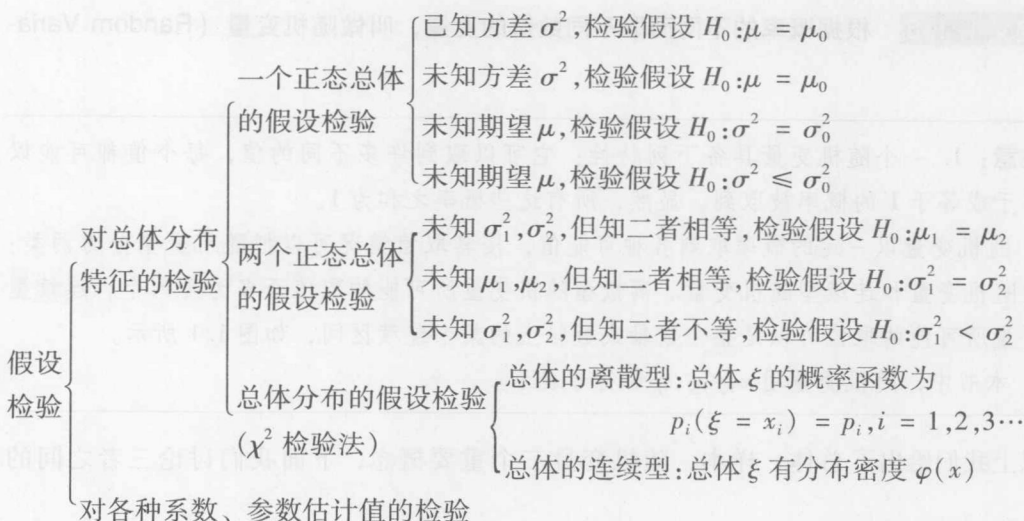
a. 估计量的特性

b. 估计的方法



c. 对估计量的检验——假设检验

它与区间估计本质相同, 只是所讨论的出发点不同:



第一节 总体与样本

首先, 我们给出三个定义:

定义 1.1.1 研究对象的全体称为总体或母体, 组成总体的每个基本单位称为个体。

注意: 1. 总体根据所包含的个体数目可分为: 有限总体和无限总体。

2. 总体中的每一个体, 具有共同的可观察特征, 也就是说, 有相同的数据产生过程。我们把这种特征作为不同总体的区别指标。

3. 量度同一对象所得到的数据, 也构成一种类型的总体。这种总体中的数据之所以不同, 是因为测量误差的存在。

4. 表现总体状况的某一个或某几个数量特征在总体中依不同概率取不同值。

定义 1.1.2 总体中抽出若干个体而组成的集合, 称为样本。样本中所含个体的个数, 称为样本容量。

注意: 在进行抽样时, 样本的选取必须是随机的, 即总体中每一个体有同样的机会被选入样本。



定义 1.1.3 根据概率的不同而取不同数值的变量，叫做随机变量 (Random Variable)。

注意：1. 一个随机变量具备下列特性：它可以取到许多不同的值，每个值都可能以一个小于或等于 1 的概率被取到。显然，所有这些概率之和为 1。

2. 随机变量以一定的概率取到各种可能值，按其取值情况可以把随机变量分为两类：离散型随机变量和连续型随机变量。离散型随机变量所可能的取值至多为可列个。连续型随机变量所可能的取值可以是整个数轴或数轴上的某个连续区间，如图 1.1 所示。

3. 本书中，随机变量用 x, y, ξ, η 等符号表示。

以上我们给出了总体、样本、随机变量三个重要概念，下面我们讨论三者之间的关系。

由上面我们已经了解到，表示总体状况的某一个或某几个数量特征在总体中往往随概率不同而取不同的值 (定义 1.1, 注意 4)。显然，对于这样的数量特征，我们用一般的变量是无法加以描述的，能够对之给予描述的是一类特殊的变量，即随机变量 (定义 1.3)。

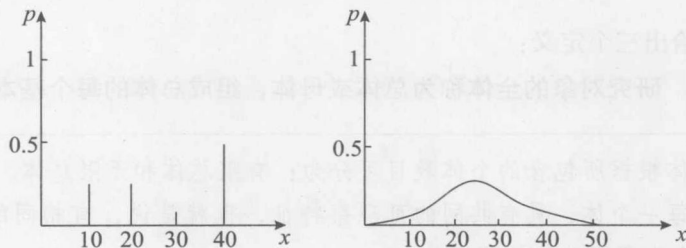


图 1.1

我们对于总体中某一个体所具有的特殊属性往往并不关心，真正感兴趣的是表示总体特征的数量指标。例如，基于股价计算出来的股票回报率等。就总体的某一数量特征 ξ 而言，如股票回报率分布，每个个体的取值不一定完全相同，但它是按照一定规律分布的，是基本稳定的。因此，对于一个总体来说，其每一个数量特征完全符合定义 1.3，是根据概率的不同而取不同值的变量，即总体中每一个数量特征就是一个随机变量。由于我们主要是研究总体的数量特征，所以我们把总体看成是具有若干数量特征的研究对象的全体，可直接用一个随机变量来表示。

可见，所谓总体就是一个随机变量，所谓样本就是 n 个 (样本容量为 n) 相互独立且