第1章

金融计量学介绍

1.1 金融计量学的含义及建模步骤

1.1.1 金融计量学的含义

要了解什么是金融计量学,先要了解什么是计量经济学。计量经济学从字面意思来看,是指经济学中的测量。实质上,计量经济学起源于经济学,是经济学的一个分支学科,是以揭示经济活动中客观存在的数量关系为内容的分支学科。

计量经济学不仅是检验经济理论的一门科学,而且也是预测经济变量未来值的一套工具,如预测股票价格走势、预测公司的销售额等。计量经济学可以利用经济数学模型拟合实际数据的过程,是基于历史数据给予政府和企业以数值化或定量化政策建议的一门科学和艺术。

一般认为金融计量学是计量经济学的一个分支,是计量经济学在金融学中的一种应用。在西方经济中,一般认为金融计量学是指金融市场的计量分析,特别是统计技术在处理金融问题中的应用。而本书所指的金融计量学除了包括以上计量经济学基本理论外,还涵盖了其在金融市场的应用,例如检验金融市场信息是否有效,检验资本资产定价模型(CAPM)是否是决定风险资产收益率的优良模型,测量和预测债券收益率的波动性,检验关于变量相互关系的假设,考察经济状况变化对金融市场的影响,等等,后面相应章节将做详细介绍。

1.1.2 金融计量建模步骤

通过各种不同的方法对金融市场进行描述和模拟就形成了各种各样的金融计量模型,这些模型都具有一些相同的主要步骤。根据布鲁克斯(Brooks)提出的金融模型步骤,对金融计量建模的基本步骤总结如图 1-1 所示。

步骤一:问题的概述。该步骤主要涉及金融或经济理论的形成,一般来自某种理论的认识或对某种理论的假设,根据理论建立模型用数学公式表示出来。具体包括:选择变量、确定变量之间的数学关系、拟定模型估计参数的数值范围。选择变量需要注意以下几点:①正确理解和把握所研究的经济现象中暗含的经济学理论和经济行为规律;②处在不同的环境中,要研究不同的行业,选择的变量也不同;③要考虑数据的可得性;④选择变量要考虑所有入选变量之间的关系,使得每一个解释变量都是独立的。

步骤二：收集样本数据。样本数据的收集与整理，是建立金融计量学模型过程中最基础的工作，也是对模型质量影响较大的一项工作。从工作程序上讲，它是在模型建立之后进行，但实际上经常是同时进行的，因为能否收集到合适的样本观测值是决定变量取舍的主要因素之一。对于金融数据的类型、特点、来源将在下一节详细介绍。

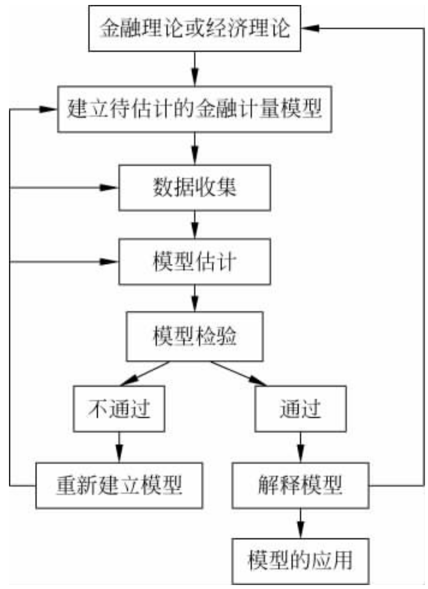

图 1-1　金融计量建模的基本步骤

步骤三：选择合适的估计方法。根据模型提出的假设和建立的数学表达式，数据的类型选择一元回归还是多元回归，选择单一方程还是联立方程。

步骤四：模型的检验。在步骤三之后初步得到估计结果，但需要进一步检验估计的结果是否满足我们的需要，是否合理地描述数据，是否具有经济学上的意义。一般检验包括三个方面：统计检验、计量经济学检验以及经济、金融意义检验。统计检验的目的在于检验模型参数估计值的可靠性，包括模型拟合优度检验、变量显著性检验、方程显著性检验等；计量经济学检验是否符合计量经济学理论知识，包括序列相关性检验、异方差检验、多重共线性检验以及协整检验等。经济、金融意义检验是将计量检验的结果与相应的经济理论或金融理论相比较，确定两者是否相符。估计的参数通过检验则可进行模型的解释，并应用在实际中，否则回到步骤一到步骤三，要么放宽假设条件，要么改变计算方法，要么重新建立模型，收集更多数据。

步骤五：模型的应用。当模型通过检验，结果获得合理的解释，步骤一的理论得到支撑，就可以将模型用来检验步骤一提出的理论、进行预测或者提出建议。金融计量学模型的应用很广，主要分为以下几个方面：①结构分析。对经济现象中变量之间相互关系进行研究，如研究某一因素变化对经济状况的影响。②金融、经济预测。根据模型变量的设定，预测金融经济变量，如一元回归模型中被解释变量的预测、未来资产价值以及波动率的预测。③政策评价。研究不同的经济政策产生的不同结果，评估这些结果，或是从不同的政策结果中选择最优的政策方案。④检验与发展经济理论。实践是检验真理的唯一标准，任何经济理论都是要通过检验，金融计量学模型是一种很好的科学方法，对于探索经济规律提供了很大的帮助。

1.2　金融数据的主要类型、特点和来源

1.2.1　金融数据的主要类型

在金融问题的分析中，主要有三类数据可供使用，即时间序列数据、截面数据和面板数据。

时间序列数据(time series data)是指一个实体在不同时期内的观测数据。例如中国各年的通货通胀率，一家公司当年每个季度的营业额，每天的股票价格，等等。时间序列数据进行回归时应注意数据的一致性，确保各变量数据的频率相同，应注意模型随机误差项有可

能产生序列相关，还应注意数据的平稳性问题，同时也要注意数据的可比性，消除不同时点的数据受通货膨胀因素的影响。时间序列数据可用于研究变量随时间推移的发展变化，并预测这些变量的未来值。如研究某个国家股票指数价格如何随着国家宏观经济基础变量的变化而变化、贸易赤字上升对该国家汇率的影响、预测某股票的波动率等。

截面数据（cross-sectional data）是指多个实体在某一时点上的观测数据。例如亚洲各个国家 2015 年的人均 GDP（国内生产总值），某一时点上深圳交易所所有股票的收益率，中国各个省份 2015 年一年税收情况，等等。实体可以是国家、城市，也可以是公司、单位和个人。分析截面数据时，可能出现异方差情况，整理数据时应注意消除异方差。这类数据主要用于研究某一时期内不同的人、公司或者其他经济实体之间的差异，来了解变量之间的关系。例如研究公司规模对其进行股票投资的回报率之间的关系、一国 GDP 水平与其主权债务违约率之间的关系。

面板数据（panel data）是指时间序列数据和截面数据相结合的数据，即多个实体在多个时期内的观测数据。例如中国国内所有银行过去 3 年的贷款数据、所有蓝筹股 2010—2015 年每日收盘价等。可以应用面板数据研究不同实体的经历和根据每个实体的变量随时间变化的发展来了解经济关系。

对于以上数据的统计和建模分析大多采用的是点值模型（point-valued model）。随着大数据时代的到来，统计学家和计量经济学家开始采用区间模型（interval-valued model）对相关数据进行分析。本书不对区间模型进行介绍，有兴趣的读者可参阅相关文献。

1.2.2　金融数据的特点

金融计量学主要研究计量经济学在金融市场中的应用。相比宏观经济数据，金融数据在频率、准确性、周期性等方面具有自己特有的性质。下面选择金融数据关注的两个重要指标，即抽样频率和时间跨度逐一介绍。

金融数据可以是低频的、高频的和超高频的，随着技术的进步，研究的数据由以前的月、周、日到现在的 10 分钟、5 分钟、1 分钟，可用于计量的数据十分巨大，这是宏观经济数据不能比拟的。

与宏观经济数据相比，金融数据统计错误和数据修正问题会较少，宏观经济数据通常是测算和估计出来的，难免会有差错，而且新公布的统计数据也有可能出错，都会给计量分析带来困难。而金融数据虽然形式多样，但一般来说价格和其他金融变量都是在交易时准确记录下来的，当然也存在数据测量错误、记录错误的可能性。但总体上，金融学中的误差和修正问题不像经济学中的那么严重。

金融数据特别是时间序列数据，一般都是不平稳的，较难区分是随机游走、趋势或其他特征。金融数据通常还有许多其他特征，特别是高频数据，包含太多的噪声，很难分离出背后的趋势；另外金融数据大部分不服从正态分布，一般回归方法很难适用，但它推动了金融经济方法的发展。

1.2.3　金融数据的来源

金融数据主要来源于试验或者对现实世界的实际观测，有以下三个来源渠道。

（1）政府部门和国际组织的出版物及网站，如可以从国家统计局网站（www. stats.

gov.cn)获得宏观金融数据,可以从世界银行网站(www.worldbank.org)或国际货币基金组织网站(www.Imf.org)获得世界各国的经济数据。

(2) 专业数据公司和信息公司。一些具体公司的市盈率、资产、负债数据需要到特定的公司网站上面下载,有些公司只有本公司网站才会提供这些数据。另外,还有一些信息公司通过收集某方面的数据,建立和维护专业型数据处理,通过有偿方式来满足客户的需要。如睿思公司。当然,也可以在股票软件,如同花顺、通达信等,下载不同股票的收盘价、最高(低)价、换手率等方面的交易数据。

(3) 抽样调查。如果分析人员需要一些特定的数据,往往只能通过抽样调查来获得。特别是大数据,很难对总体数据进行分析,或者无法获得总体数据,只能通过抽样来估计总体。例如要对中国股市的投资者信心进行建模,就必须通过设置调查问卷,对不同的投资群体进行数据采集。表1-1列出了几个常用的金融机构和数据库及其网址。

表1-1 几个常用的金融机构和数据库及其网址

机构或数据库名称	网址
纽约证券交易所(NYSE)	http://www.nysc.com
伦敦证券交易所(LSE)	http://www.londonstockexchange.com
东京证券交易所(TSE)	http://www.tse.or.jp
芝加哥期货交易所(CBOT)	http://www.cbot.com
上海证券交易所(SSE)	http://www.sse.com.cn
深证证券交易所(SZSE)	http://www.szse.cn
计量经济学会(Econometric Society)	http://www.econometricsociety.org/es
证券交易委员会(SEC)	http://www.sec.gov
中国金融经济数据库(CCER)	http://www.ccer.edu.cn

1.3 收益率计算

1.3.1 单期收益率

金融许多问题都是从价格的时间序列开始的,如股票每天的收盘价、黄金期货每日价格等。在金融计量上,用得较多的是把价格时间序列转化为收益率时间序列。因为收益率序列统计特性良好,而且收益率还具有无量纲单位的优点,如年收益率为10%,那么投资者投资100就会得到110,投资1 000就会得到1 100。计算收益率一般有两种方法:简单收益率和连续复合收益率,计算方法如下:

简单收益率(simple return)计算公式:

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} \times 100\% \tag{1.1}$$

连续复合收益率(continuous compounding return)计算公式:

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) \times 100\% \tag{1.2}$$

式中,R_t为在t时期的简单收益率;r_t为在t时期的连续复合收益率;P_t为在t时期的资产价格;ln为自然对数。

这里计算的收益率指的都是单期收益率，连续复合收益率也称对数收益率。

根据简单收益率和连续复合收益率的定义，可以根据泰勒公式一阶展开得出如下结论：当简单收益率的值接近零时，连续复合收益率与简单收益率几乎是相等的。也就是求极限时，若样本数据的时间间隔越来越小，简单收益率与连续复合收益率将趋于一致。

1.3.2　多期收益率

多期收益率指的是间隔多个时间段求得的收益率。

多期(假设有 k 期)的简单收益率计算公式如下：

$$R_t(k) = \frac{P_t - P_{t-k}}{P_{t-k}} \times 100\% \tag{1.3}$$

多期的连续复合收益率计算公式如下：

$$\begin{aligned} r_t(k) &= \ln\left(\frac{P_t}{P_{t-k}}\right) \\ &= \ln\left[\left(\frac{P_t}{P_{t-1}}\right)\left(\frac{P_{t-1}}{P_{t-2}}\right)\cdots\left(\frac{P_{t-(k-1)}}{P_{t-k}}\right)\right] \\ &= r_t + r_{t-1} + \cdots + r_{t-k+1} \end{aligned} \tag{1.4}$$

式中，$R_t(k)$为 k 期的多期简单收益率；$r_t(k)$为 k 期的多期连续复合收益率。

多期收益率可以由单期收益率计算得到，如 1 年有 12 个交易月，若采用简单收益率，则年总收益率可以由这 12 个月总收益率的乘积得到；若采用对数收益率，则可以由这 12 个月对数收益率求和计算得到。

单期与多期是相对的，例如，在用月收益率计算年收益率时，月收益率可被视作单期收益率，而在用日收益率计算月收益率时，月收益率被视作多期收益率。这表明单期与多期是相对的，既可以将多期视作一个大的单期，也可以将一个单期分解成包括若干个小单期的多期。

通过式(1.4)可以看出，对数收益率具有可加性，给计算带来了很大的方便；同时连续复合收益率有良好的性质，在比较资产间的收益率时，通过连续复合收益率的计算，可以消除不同频率的影响。所以在金融学术文献上面大都采用连续复利收益率公式计算收益率。

对于简单收益率而言，大部分金融资产以有限的负债形式表现，即投资者的最大损失仅仅是他的全部投资，这就意味着可获取的最小收益率是－100％，但是目前国际上常用的用来描述收益率序列特征的分布函数(如传统的正态分布)所需要的支撑，往往是整个实数轴，因此，简单收益率的有限负债性给金融建模带来了一定的不便。而对数收益率是对简单总收益率取对数得到，它的取值范围为整个实数轴，所以可以不受有限负债的限制。

近年来，年化收益率在金融市场中较为常见。例如，某个 6 个月的理财产品“年化收益率为 4.8％”，其意思是半年期到期收益率对应的是 2.4％。对于季节价格指数，年化收益率计算公式近似写成

$$100\% \times \ln\left(\frac{P_t}{P_{t-1}}\right)^4 = 400\% \times \ln\left(\frac{P_t}{P_{t-1}}\right) \tag{1.5}$$

以此类推，对于月度数据，年化收益率计算公式如下：

$$100\% \times \ln\left(\frac{P_t}{P_{t-1}}\right)^{12} = 1\,200\% \times \ln\left(\frac{P_t}{P_{t-1}}\right) \tag{1.6}$$

上述计算严格来说是一种近似计算,对于简单收益率而言,上述计算公式分别如下:

$$100\% \times \left[\left(\frac{P_t}{P_{t-1}} \right)^4 - 1 \right] \quad 和 \quad 100\% \times \left[\left(\frac{P_t}{P_{t-1}} \right)^{12} - 1 \right]$$

从基本的高数知识可知,在一般情况下,这两种计算方法近似相等。

1.4 常用的统计学与概率知识

1.4.1 随机变量

1. 随机变量的含义

随机变量是一个随机结果的数值概括。例如未来股票的价格、1 年 GDP 值、企业生产总值等,这些经济方面的观测值都是随机变量。随机变量根据其结果取值的特性,可以分为离散型随机变量和连续型随机变量,若所有可能的结果是有限的或可数无限的,则随机变量是离散型;若所有可能的结果充满一个或若干有限或无限区间,则随机变量是连续型的。

2. 随机变量的概率分布

为了描述随机变量的不确定性,往往将随机结果与概率联系起来。随机变量 X 所有可能取值为 x_i 的概率就是 X 的概率分布。对于离散型变量,其概率分布为

$$P(X = x_i) = p_i \quad (i = 1,2,3,\cdots,n) \tag{1.7}$$

对于连续型变量,所有可能的结果对应的概率为 0,度量该随机变量在某一特定范围或区间内的概率才有实际意义。概率密度函数(pdf)定义为 $f(x)$ 且满足:

$$P(a \leqslant x \leqslant b) = \int_a^b f(x)\mathrm{d}x \tag{1.8}$$

$$f(x) \geqslant 0$$

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

3. 随机变量的累积分布函数

随机变量小于或者等于某个特定值的概率记为 $F(x)$,就是累积分布函数(CDF),定义如下:

$$F(x) = P(X \leqslant x) \tag{1.9}$$

容易得出离散随机变量 X 的累积分布函数为

$$F(x) = \sum_{x_i \leqslant x} p_i \tag{1.10}$$

连续型随机变量 X 的累计分布函数为

$$F(x) = \int_{-\infty}^{x} f(x)\mathrm{d}x \tag{1.11}$$

累积分布函数和概率密度函数可以准确而全面地描述随机变量的分布情况。

4. 随机变量分布的矩条件

矩条件是金融计量建模中重要的内容，这里先回归一下随机变量分布矩条件的一些基本概念。假设 X 是一个密度函数为 $f(x)$的连续型随机变量，则矩条件分别定义为

P 阶原点矩定义为

$$E[X^p] = \int_{-\infty}^{\infty} x^p f(x)\mathrm{d}x \tag{1.12}$$

P 阶中心矩定义为

$$E\{[X-E(X)]^p\} = \int_{-\infty}^{\infty} [x-E(X)]^p f(x)\mathrm{d}x \tag{1.13}$$

1）随机变量的期望

期望就是随机变量平均值，度量了随机变量的集中趋势。用 $E(X)$来表示随机变量的数学期望，离散型随机变量期望定义为

$$E(X) = \sum_{i=1}^{n} x_i p_i \tag{1.14}$$

式中，x_i 为随机变量 X 所有可能取值；p_i 为 x_i 发生的概率。

连续型随机变量期望定义为

$$E(X) = \int_{-\infty}^{+\infty} x f(x)\mathrm{d}x \tag{1.15}$$

式中，$f(x)$为随机变量 X 的概率密度函数。

数学期望有一个重要的性质：

$$E(a+bx) = a+bE(x) \tag{1.16}$$

式中，a 和 b 都为常数。

期望也是随机变量的一阶矩，其在金融中的含义是金融资产平均收益或者期望收益，反映的是金融资产收益的一阶矩风险。

2）随机变量的方差和标准差

方差是刻画随机变量的偏离期望的程度，偏离的量 $X-E(X)$有正有负，为了不使正负偏离彼此抵消，一般那考虑$[X-E(X)]^2$，也是一个随机变量，取其均值就可以刻画随机变量 X 的波动程度。所以方差 $\sigma^2(X)$定义如下：

$$\mathrm{Var}(X) = \sigma^2(X) = \begin{cases} \sum_i [x_i - E(X)]^2 p_i & (X\text{ 是离散型随机变量}) \\ \int_{-\infty}^{+\infty} [x-E(X)]^2 f(x)\mathrm{d}x & (X\text{ 是连续型随机变量}) \end{cases} \tag{1.17}$$

由于期末的收益是不确定的，所以期末的资产回报率是随机变量，收益的方差度量收益率的分散趋势程度，经常用方差来衡量经济变量的波动性。标准差是方差的开方，通常用 σ 来表示：

$$\sigma = \sqrt{\sigma^2(X)} \tag{1.18}$$

方差的一个重要性质是

$$\sigma^2(a+bX) = b^2\sigma^2(X) \tag{1.19}$$

式中，a 和 b 都为常数。

方差或者标准差也是随机变量的二阶矩,其在金融中的含义是金融资产的一般波动(volatility),反映的是金融资产收益的二阶矩风险。

3) 随机变量的偏度和峰度

偏度是用于衡量分布的不对称程度或偏斜程度的指标,用 $E\{[X-E(X)]^3\}$来表示,也称三阶矩。实际中通常用偏度系数 S 表示对称性程度,当 $S=0$ 时,分布对称;当 $S>0$ 时,为正偏斜,有个较长的右尾部,均值大于中位数;当 $S<0$ 时,为负偏斜,有个较长的左尾部,均值小于中位数,其定义如下:

$$S=\frac{E\{[X-E(X)]^3\}}{\sigma^3} \tag{1.20}$$

对于离散型随机变量 X 的一系列观测值($x_1,x_1,\cdots,x_n$)组成的样本,其偏度系数 S 的计算方法如下:

$$S=\frac{1}{n}\sum_{i=1}^{n}\frac{(x_i-\bar{x})^3}{[\sigma^2(x_i)]^{3/2}} \tag{1.21}$$

式中,n 为样本的个数;x_i 为样本的观测值;$\bar{x}$ 为观测值的平均值;$\sigma^2(x_i)$为观测值的方差。

峰度是随机变量的四阶矩,用 $E\{[X-E(X)]^4\}$来表示。峰度是用于衡量分布的集中程度或分布曲线的尖峭程度的指标,也就是反映分布函数尾巴厚度的指标。实际中通常用峰度系数 K 来表示,正态分布的 $K=3$,当 $K>3$ 时,分布呈尖峰状态。如金融时间序列多是这种分布,则具有"尖峰厚尾"的特征;$K<3$ 为扁峰分布,其定义如下:

$$K=\frac{E\{[X-E(X)]^4\}}{\sigma^4} \tag{1.22}$$

同样地,对于离散型随机变量 X 的一系列观测值($x_1,x_1,\cdots,x_n$)组成的样本,其偏度系数 K 的计算方法如下:

$$K=\frac{1}{n}\sum_{i=1}^{n}\frac{(x_i-\bar{x})^4}{[\sigma^2(x_i)]^2} \tag{1.23}$$

偏度在金融中的含义是衡量金融资产受信息冲击的非对称程度,反映的是金融资产收益的三阶矩风险(偏度风险)。峰度在金融中的含义是测度极端风险发生的可能性,反映的是金融资产收益的四阶矩风险(峰度风险),三阶矩风险和四阶矩风险统称高阶矩风险。

以上四个阶矩风险统称矩风险,它们在金融资产定价、投资组合、风险建模与管理等方面有着广泛的应用。

1.4.2 常用概率分布

每个随机变量都有其特定的概率分布特征,下面介绍金融计量学中常用的几种概率分布,即正态分布、χ^2 分布、t 分布、F 分布以及二项分布。

1. 正态分布

服从正态分布(normal distribution)的连续型随机变量的概率密度曲线如图 1-2 所示,其的概率密度为

$$f(x)=\frac{1}{\sqrt{2\pi\sigma}}\mathrm{e}^{-(x-\mu)^2/(2\sigma^2)} \quad (-\infty<x<+\infty) \tag{1.24}$$

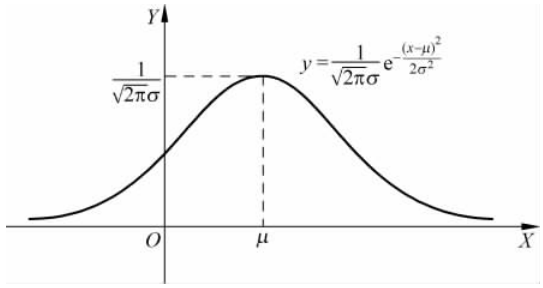

图 1-2　正态分布概率密度曲线

此时称随机变量 X 服从正态分布，记作 $X \sim N(\mu,\sigma^2)$，其中 $\mu,\sigma(\sigma>0)$ 分别是正态分布的期望和方差。正态分布也称高斯(Gauss)分布。

对于 $\mu=0,\sigma=1$ 的特殊情况，即如果 $X \sim N(0,1)$，则称 X 服从标准正态分布，它的概率密度有

$$\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} \tag{1.25}$$

从理论上看，正态分布具有很多良好的性质，许多概率分布可以用它来近似；还有一些常用的概率分布是由它直接导出的，下面介绍的三种分布都是由正态分布推导得到的。

2. χ^2 分布

随机变量 $X_1, X_2, \cdots, X_n$ 独立同分布于标准正态分布 $N(0,1)$，则 $\chi^2 = X_1^2 + X_2^2 + X_3^2 + \cdots + X_n^2$ 的分布称为自由度为 n 的 χ^2 分布(chisquare distribution)，记为 $\chi^2 \sim \chi^2(n)$。图 1-3 所示为不同的 n 值 χ^2 分布的密度函数。

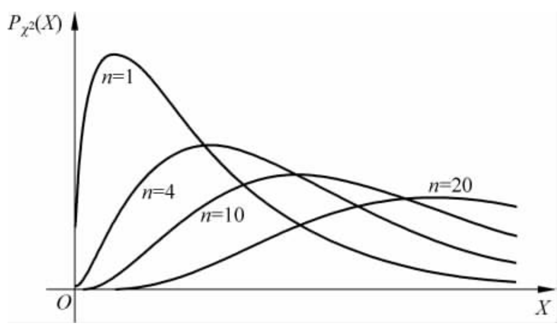

图 1-3　χ^2 分布的密度函数

χ^2 分布在第一象限内，呈正偏态，随着参数 n 的增大，χ^2 分布趋近于正态分布。自由度为 n 的 χ^2 分布，其期望值为 $E(\chi^2)=n$，方差 $\sigma(\chi^2)=2n$。如果来自方差为 σ^2 的一个正态分布的 N 个观测值的样本方差为 s^2，则可以得到

$$(N-1)s^2/\sigma^2 \sim \chi^2(N-1)$$

3. t 分布

t 分布(t distribution)又称学生的 t 分布,它与正态分布和 χ^2 分布密切相关,其定义如下,随机变量 X 和 Y 独立,且 $X\sim N(0,1)$,$Y\sim\chi^2(n)$,则称 $X/\sqrt{Y/n}$ 的分布为自由度为 n 的 t 分布,记为 $t=X/\sqrt{Y/n}\sim t(n)$。图 1-4 所示为不同自由度下 t 分布的密度函数曲线。

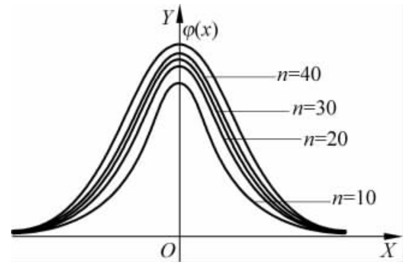

图 1-4 不同自由度下 t 分布密度函数曲线

t 分布是一簇曲线,以 O 为中心,左右对称的单峰分布;其形态变化与自由度 n 大小有关。自由度 n 越小,t 分布曲线越低平;自由度 n 越大,t 分布曲线越接近标准正态分布曲线。对应于每五个自由度 n,都有一条 t 分布曲线,每条曲线都有其曲线下统计量 t 的分布规律,计算较复杂。

t 分布在概率统计中计算置信区间估计、显著性检验等问题时发挥重要作用。还有一个很有用的结论,设 $x_1,x_2,\cdots,x_N$ 是来自正态分布 $N(\mu,\sigma^2)$ 的一个样本,N 个观测值的样本方差为 s^2,样本均值为 $\bar{x}$,则有

$$\frac{(\bar{x}-\mu)/(\sigma/\sqrt{N})}{\sqrt{(N-1)s^2/\sigma^2(N-1)}}=\frac{\sqrt{N}(\bar{x}-\mu)}{s}\sim t(N-1) \tag{1.26}$$

当然,这里 t 分布是对称的,其实也存在着偏 t 分布,其分布图是非对称的。

4. F 分布

设 X 与 Y 是相互独立的随机变量,且 $X\sim\chi^2(m)$,$Y\sim\chi^2(n)$,则称 $F=\dfrac{X/m}{Y/n}$ 的分布是自由度为 m 和 n 的 F 分布(F distribution),记为 $F\sim F(m,n)$,其中 m 称为分子自由度或第一自由度,n 称为分母自由度或第二自由度。图 1-5 给出了不同自由度的 F 分布密度函数。

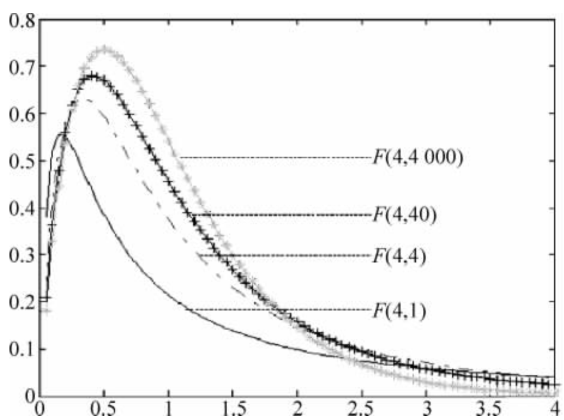

图 1-5 不同自由度的 F 分布密度函数曲线

F 分布是一种非对称分布,另外有两个结论:① 若随机变量 $F\sim F(m,n)$,则 $\dfrac{1}{F}\sim F(n,m)$;②若 $t\sim t(n)$,则 $t^2\sim F(1,n)$。