工程硕士数学主干课程系列教材

数理统计及其应用

何春雄 朱锋峰 龙卫江 编著

SHULI TONGJI JIQI YINGYONG

华南理工大学出版社
SOUTH CHINA UNIVERSITY OF TECHNOLOGY PRESS

数理统计及其应用

何春雄　朱锋峰　龙卫江　编著

华南理工大学出版社
·广州·

内 容 简 介

本书是高等学校工程硕士各专业“数理统计”课程的教材，以微积分学、线性代数和初等概率论为基础，介绍数理统计的基础知识，包括抽样分布、参数估计、假设检验、回归分析、方差分析与试验设计，以及 R 软件简介. 书中尽量略去繁复的推导和计算，重点介绍数理统计的基本概念、基本思想和基本方法，同时给出每种统计方法的应用实例及 R 软件实现，各章配有习题并在书末附有习题参考答案.

本书可作为高等学校理工科、管理、经济、金融等专业本科或专业学位硕士“数理统计”课程的教材. 读者只要具有初等微积分学、线性代数、初等概率论以及简单的计算机知识，就可以阅读和学会用 R 软件进行数据处理. 课堂教学中可不专门讲解 R 软件，教师只在例题计算中做简单演示即可.

图书在版编目(CIP)数据

数理统计及其应用/何春雄，朱锋峰，龙卫江编著. —广州：华南理工大学出版社，2012.2
工程硕士数学主干课程系列教材
ISBN 978-7-5623-3569-6

Ⅰ. ①数… Ⅱ. ①何… ②朱… ③龙… Ⅲ. ①数理统计-研究生-教材 Ⅳ. ①O212

中国版本图书馆 CIP 数据核字(2012)第 011937 号

总 发 行： 华南理工大学出版社(广州五山华南理工大学 17 号楼，邮编 510640)
营销部电话：020-87113487 87110964 87111048(传真)
E-mail：scutc13@scut.edu.cn http://www.scutpress.com.cn
责任编辑： 詹志青
印 刷 者： 广州市穗彩彩印厂
开 本： 787mm×1092mm 1/16 **印张：** 11.5 **字数：** 287 千
版 次： 2012 年 2 月第 1 版 2012 年 2 月第 1 次印刷
印 数： 1～1000 册
定 价： 23.00 元

前　言

自20世纪90年代末开始，教育部陆续批准在一些较高层次的工科院校设立“工程硕士”学位授权点。此举的目的是为高新企业中第一线工作的科技人员提供一个在职培训和提高的机会，从而为企业的转型升级注入活力。多年来，教育部工程硕士课程教学指导小组希望工程硕士课程的教学内容、教学方法以及考核方式等方面与工学硕士课程有所区别，特别是强调学以致用。然而，目前在大多数院校中，工程硕士的数学课程基本上都是工学硕士课程的“缩水版”，既不能较深入地介绍基本理论和方法，又不能详细讲解如何应用，更谈不上介绍如何借助计算机软件工具来解决具体实际问题。针对这种现状，我们在华南理工大学研究生院的资助下，筹划编写“工程硕士数学主干课程系列教材”，力争使每门课程都能着重介绍模型的应用背景和应用实例，并介绍一种数学软件工具，力争每个实例都能以算法实现，以期全面改善教学效果，使学生能够自觉地、得心应手地运用该课程的基本理论、方法解决实际问题，大幅度提高运用数学工具和技术的能力与素质。

本教材突出了数理统计基本思想的介绍，略去了繁复的证明。对每章介绍的统计方法都配备了应用实例，并演示如何运用R软件来实现。期望读者在领会每种统计方法的基本思想的基础上，能够借助R软件来解决自己在工作中遇到的各种实际问题。

本教材的编写，理论和方法介绍部分由何春雄执笔，习题和习题解答由朱锋峰负责完成，应用实例和R软件部分由龙卫江负责实现。前5章介绍常用统计方法的基本思想和方法，其中第1章是为后续的4章做概率论基础方面的准备工作，第2、3、4和5章分别介绍估计理论、假设检验、回归分析、方差分析及试验设计等的基本思想和方法，基本上略去了证明(所有结果的证明可查阅参考文献[1]～[3])。在每章的最后一节给出了应用实例及R软件实现。第6章对R软件作了简单介绍。

本教材的写作和出版得到了华南理工大学研究生院研究生重点课程建设项目基金、华南理工大学出版基金的资助，特此表示感谢。

由于编者知识和能力所限，谬误与不足之处在所难免，有待在教学实践中加以更正和补充，也恳请读者批评指正。作者邮件地址为 machxhe@scut. edu. cn(何春雄)， ffzhu@scut. edu. cn(朱锋峰)，wjlong@scut. edu. cn(龙卫江)。

编　者

2012 年 1 月于华南理工大学理学院

目　　录

1　数理统计的基本概念 …… 1
　1.1　概率空间与随机变量 …… 1
　1.2　总体、样本和统计量 …… 8
　1.3　抽样分布 …… 9
　1.4　抽样分布中的应用实例及 R 语言实现 …… 12
　复习思考题 1 …… 19
　习题 1 …… 19

2　参数估计 …… 23
　2.1　参数的点估计 …… 23
　2.2　估计量优劣性的评价 …… 28
　2.3　参数的区间估计 …… 30
　2.4　参数估计中的应用实例及 R 语言实现 …… 34
　复习思考题 2 …… 42
　习题 2 …… 43

3　假设检验 …… 45
　3.1　假设检验与两类错误 …… 45
　3.2　正态总体参数的假设检验 …… 48
　3.3　非正态总体均值的假设检验 …… 51
　3.4　非参数假设检验 …… 52
　3.5　假设检验应用实例及 R 语言实现 …… 55
　复习思考题 3 …… 62
　习题 3 …… 62

4　回归分析 …… 66
　4.1　一元线性回归 …… 66
　4.2　多元线性回归 …… 72
　4.3　可化为线性回归的非线性回归 …… 76
　4.4　回归分析中的应用实例及 R 语言实现 …… 77
　习题 4 …… 92

5 方差分析 …… 96
5.1 单因素方差分析 …… 96
5.2 两个因素的方差分析 …… 98
5.3 正交试验设计与方差分析 …… 102
5.4 方差分析与正交试验设计中的应用实例及 R 语言实现 …… 109
习题 5 …… 123

6 R 语言简介 …… 128
6.1 软件 R 简介 …… 128
6.2 基本操作 …… 129
6.3 数据类型和对象 …… 130
6.4 数据读写 …… 142
6.5 简单编程 …… 144
6.6 作图 …… 147

习题参考答案 …… 152
附表：数理统计常用表 …… 156
参考文献 …… 177

1 数理统计的基本概念

本章简单复习随机变量的分布和数字特征，并介绍数理统计的基本概念，包括总体、样本、统计量等，着重介绍三种重要的分布，即χ^2分布、t分布和F分布，并介绍正态总体的样本均值、样本方差及其相关统计量的分布.

1.1 概率空间与随机变量

数理统计是统计学的重要分支，具有广泛的应用．大家知道，现场统计是流水账式的统计，它记录所有发生的结果，比如一个生产班组的领料件数或数量、产品件数及各项指标等都一一作记录；再比如一批产品中各个产品的每项技术指标都作记录．数理统计有别于现场统计的最本质特征可以概括为“用局部推断整体”，这就使得数理统计所作推断的结论不可避免地存在偏差或错误，而刻画或把握这种偏差的有效方法则是概率论．概率论通过给出各种各样的统计量所服从的分布或数字特征，来演绎评价各种统计方法的优劣或置信程度．一般来说，数理统计的方法大多来自简单的直观想法或概率论的有关基本事实（比如大数定律、中心极限定理或某些渐近分布），而要评价这些方法则需要比较艰难的概率论推导或证明.

概率论的核心问题是回答某种随机现象各个结果(或称随机事件)发生的可能性的大小，具体刻画这种可能性的大小的度量就是一个事件发生的概率，这也是用确定的数学科学研究随机现象的切入点.

为方便地运用数学学科的各种工具，人们想到人为地或自然地将随机试验的结果与数值对应起来，这个对应的结果就称为随机变量，其含义是在试验结果出来之前无法预知该变量取何值，也就是说该变量的取值是随机的，进而通过刻画随机变量的分布来描述一个随机事件发生的概率．这样一来，概率论可以称为研究随机变量分布的科学．初等概率论研究的是独立随机变量，所以经典的数理统计的讨论大多基于独立随机变量，也就是数理统计中的简单随机样本.

1.1.1 概率的定义

概率论是研究随机现象的统计规律的科学．为用确定性的数学研究随机现象这种非确定性现象，人们在建立数学模型时，引进了三个基本的概念，它们是随机试验、随机事件和概率(测度).

所谓**随机试验**，是指对随机现象的观测或实验，其基本意义是：

1）试验可以在基本相同的条件下大量重复；

2）试验会出现的那些结果是预知的；

3) 每次试验将出现哪一个结果是无法预知的.

所谓**随机事件**，则指随机试验的结果，简称为**事件**. 而**概率**(**测度**)则是随机事件发生的可能性大小的一种度量，这是用概率论学科研究随机现象的关键切入点. 然而如何规定或定义这个度量却非易事，下面进行简单总结.

例 1.1.1 掷一枚骰子，观察其落定后朝上的面出现的点数，这就是一个随机试验. 而 $A=\{$出现的点数为奇数$\}$、$B=\{$出现的点数为偶数$\}$、$C=\{$出现的点数小于3$\}$等都是事件. 若骰子均匀，则 A、B、C 的概率应分别为 $P(A)=\frac{1}{2},P(B)=\frac{1}{2},P(C)=\frac{1}{3}$.

在初等概率论教材中见到的古典概型和几何概型中，关于概率的定义都是基于某类特殊随机现象的，不能作为概率的数学定义或公理化定义，从而不能用来作一般的演绎推理. 经过数学家多年艰苦的探索，直到 20 世纪 30 年代，以俄国柯尔莫哥洛夫(A. H. Колмогоров)为代表的数学家引入概率的公理化定义后，才使概率论建立在坚实的、严密的数学基础之上. 概率的公理化定义的基本思想是把随机事件看作集合，从而事件的和、积、对立及差等运算分别对应集合的并、交、求余和差等集合运算，而把概率定义为集合的测度，从而把概率论建立在测度论的基础之上，也使所有的讨论都在概率空间的框架之下来进行. 在概率的公理化定义中，相应于前述随机试验、随机事件和概率等三个直观概念的，分别是基本事件空间 Ω(或称样本空间)、事件域 $\mathscr{F}$(或称事件 σ - 代数)和概率测度 P.

基本事件空间(Ω)

定义 1.1.1 如果若干个事件构成的集合 Ω 中，每次试验有且仅有其中的一个事件发生，则称集合 Ω 为**基本事件空间**(或**样本空间**)，而称 Ω 中的元素为**基本事件**(或**样本点**)，称 Ω 为**必然事件**，$\varnothing$ 为**不可能事件**.

例 1.1.2(例 1.1.1 续) 记 $\omega_i=\{$出现的点数为 $i\}$，$i=1,2,\cdots,6$，则 $\Omega_1=\{\omega_1,\omega_2,\cdots,\omega_6\}$ 为一基本事件空间，而 $\Omega_2=\{A,B\}$ 也是基本事件空间.

事件的运算

设 A 和 B 为任意二事件，则 A 和 B 的**积**(记作 $A\cap B$ 或 AB)表示 A 和 B 都出现这样的试验结果；A 和 B 的**和**(记作 $A\cup B$)表示 A 和 B 至少出现一个这样的结果；A 和 B 的**差**(记作 $A\setminus B$ 或 $A-B$)表示 A 出现但 B 不出现这样的结果；A 的**对立**(记作 $\bar{A}$)表示 A 不出现这样的结果.

显然，$A\setminus B=A\bar{B}$. 另外，若 $AB=\varnothing$，则称 A 与 B **互不相容**；若 A 出现则 B 必然出现，则称 B **包含** A，记作 $A\subset B$.

例 1.1.3(例 1.1.1 续) 沿用例 1.1.1 的有关记号，我们知道，$A\cap C=\{$出现的点数为 1$\}$，$A\cup C=\{$出现的点数为 1,2,3,5$\}$，$A\setminus C=\{$出现的点数为 3,5$\}$，$C\setminus A=\{$出现的点数为 2$\}$，$\bar{A}=B$，$\bar{B}=A$.

从例 1.1.3 中可以看出，如果将事件看成集合，那么事件的和、积、差和对立等运算，分别对应于集合的并、交、差和求余等运算.

正如实数经加、减、乘、除等运算之后仍为实数(即实数关于这几个运算封闭)，事件经运算后也是事件，即如果把一个试验的结果看成一个集合类，那么该集合类应当对事件

运算封闭，这就导致下面关于事件域的定义.

事件域($\mathscr{F}$)

定义 1.1.2 设Ω是基本事件空间，$\mathscr{F}$是由Ω的一些子集(即事件)为元素所组成的集合，如果满足下列条件：

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

(2) 若 $A \in \mathscr{F}$, 则 $\bar{A} \in \mathscr{F}$.

(3) 若 $A_n \in \mathscr{F}, n = 1,2,\cdots$, 则 $\bigcup\limits_{n=1}^{\infty} A_n \in \mathscr{F}$.

则称$\mathscr{F}$为**事件域**，$\mathscr{F}$中的元素称为事件，Ω称为必然事件.

事件域之所以这样定义，一方面它抽象出了一个由事件构成的集合关于事件运算封闭所需的条件，另一方面也为灵活运用概率论研究各种随机现象提供了一个基本框架.

例 1.1.4(例 1.1.2 续) 沿用例 1.1.2 的记号，则 $\mathscr{F}_1 = \{A \mid A \subset \Omega_1\}$ (通常记作 2^{Ω_1}) 和 $\mathscr{F}_2 = \{\Omega_2, \varnothing, A, B\}$ 以及 $\mathscr{F}_3 = \{\Omega_1, \varnothing, A, B\}$ 都为事件域.

概率测度(P)

定义 1.1.3 设Ω是随机试验的基本事件空间，P 为定义在事件域$\mathscr{F}$到实数集 $\mathbf{R}$ 的映射，满足：

(1) (非负性) 对任一事件 A ，有 $0 \leqslant P(A) \leqslant 1$.

(2) (规范性) $P(\Omega) = 1$.

(3) (可列可加性) 若 $A_1, A_2, \cdots$ ，互不相容，则

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

则称 P 为事件域$\mathscr{F}$上的**概率测度**，而称 $P(A)$ 为事件 A 的概率.

概率测度的这种定义是长度、面积和体积的抽象，也是我们日常生活中称重和量长度的一种抽象. 我们称 $(\Omega, \mathscr{F}, P)$ 为**概率空间**. 为方便应用，下面简单讨论概率测度的性质，其证明可在一般概率教材中找到.

概率的性质

(1) $P(\varnothing) = 0$;

(2) (有限可加性) 若 $A_1, A_2, \cdots, A_n$ 互不相容，则

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i);$$

(3) $P(\bar{A}) = 1 - P(A)$;

(4) (单调性和可减性) 若 $A \subset B$ ，则有

$$P(A) \leqslant P(B), \text{且 } P(B \setminus A) = P(B) - P(A);$$

(5) 加法公式：

$$P(A \cup B) = P(A) + P(B) - P(AB).$$

1.1.2 随机变量及其分布与数字特征

我们知道，概率测度 P 是定义在事件域 $\mathscr{F}$ 到实数集 $\mathbf{R}$ 的映射，它不是经典的函数. 为有效地应用分析数学工具来分析和研究随机现象，人们想到把基本事件 ω 变换成数，从

而把事件的概率用函数值来表达，这就是随机变量及其分布函数.

1.1.2.1 随机变量的定义及其分布函数

定义1.1.4 设 $(\Omega, \mathscr{F}, P)$ 为概率空间，记 $\mathbf{R}$ 为实数集. 称映射 $\xi: \Omega \to \mathbf{R}$ 为随机变量，如果对任意 $x \in \mathbf{R}$，有 $\{\omega \mid \xi(\omega) \leqslant x\} \in \mathscr{F}$. ξ 的**分布函数**定义为

$$F_\xi(x) = P(\{\omega : \xi(\omega) \leqslant x\}), \quad \forall x \in \mathbf{R}.$$

由定义1.1.4，显然有

$$P(a < \xi \leqslant b) = F_\xi(b) - F_\xi(a), \quad \forall a < b \in \mathbf{R}.$$

且容易证明下列事实：

$$P(\xi = a) = F_\xi(a) - F_\xi(a-0), \quad \forall a \in \mathbf{R};$$
$$P(a \leqslant \xi \leqslant b) = F_\xi(b) - F_\xi(a-0), \quad \forall a < b \in \mathbf{R};$$
$$P(a \leqslant \xi < b) = F_\xi(b-0) - F_\xi(a-0), \quad \forall a < b \in \mathbf{R};$$
$$P(a < \xi < b) = F_\xi(b-0) - F_\xi(a), \quad \forall a < b \in \mathbf{R}.$$

这些事实说明，引入随机变量的分布函数，人们所关心的有关事件的概率都可以用其分布函数来表达.

描述实际随机现象时，有两类重要的随机变量，一类是至多取可数多个不同的值，称为**离散型随机变量**；另一类是连续取值的，它的分布函数可以表示为 $F_\xi(x) = \int_{-\infty}^{x} f_\xi(t)\,\mathrm{d}t$，称为**连续型随机变量**，而称 f_ξ 为 ξ 的**分布密度函数**.

现实中为了刻画一个随机现象，随机试验的结果往往需要用维数高于1维的数组来表达，这就需要引入随机向量. 后文将要引入的样本就是随机向量.

定义1.1.5 设 $(\Omega, \mathscr{F}, P)$ 为概率空间，记称 $\mathbf{R}^n$ 为 n 维实数空间. 称向量 $(\xi_1, \xi_2, \cdots, \xi_n)$ 为随机向量，如果 ξ_i 为随机变量，$i = 1, 2, \cdots, n$. 其联合分布函数定义为

$$F_{\xi_1, \xi_2, \cdots, \xi_n}(x_1, x_2, \cdots, x_n) = P\left(\omega \in \Omega : \bigcap_{i=1}^{n} \{\xi_i(\omega) \leqslant x_i\}\right),$$
$$\forall (x_1, x_2, \cdots, x_n) \in \mathbf{R}^n.$$

与前述随机变量雷同，若分布函数可以表示为

$$F_{\xi_1, \xi_2, \cdots, \xi_n}(x_1, x_2, \cdots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_n} f_{\xi_1, \xi_2, \cdots, \xi_n}(t_1, t_2, \cdots, t_n)\,\mathrm{d}t_1 \mathrm{d}t_2 \cdots \mathrm{d}t_n,$$

则称 $(\xi_1, \xi_2, \cdots, \xi_n)$ 为**连续型随机向量**， 而称 $f_{\xi_1, \xi_2, \cdots, \xi_n}$ 为其**分布密度函数**. 若 $(\xi_1, \xi_2, \cdots, \xi_n)$ 的每个分量都至多取可数多个不同的值，则称之为**离散型随机向量**.

随机变量（向量）除了用其分布函数完整刻画之外，还可以用其数字特征来刻画其某些分布特性. 下面简单回顾一下.

1.1.2.2 随机变量的数字特征

数学期望是刻画随机变量取值的平均程度的一个数字特征.

我们知道，一个随机变量 ξ 取何值是无法预知的，它的取值因各次试验出现的结果而有所不同，但它平均取何值呢?

如果 ξ 为离散型随机变量，

ξ	a_1	a_2	$\cdots$	a_n	$\cdots$
P	p_1	p_2	$\cdots$	p_n	$\cdots$

则考虑到 ξ 取值的可能性的大小，该均值应为加权平均，所以此时定义随机变量 ξ 的数学期望 $E[\xi]$ 为

$$E[\xi] = \sum_{i=1}^{\infty} a_i p_i.$$

由于 ξ 的统计特性由其分布函数 F_ξ 确定，并考虑到 ξ 的平均值应当与其取值 $a_i, i = 1, 2, \cdots$ 的排列顺序无关，而有如下较为抽象的定义.

定义 1.1.6 设 $(\Omega, \mathscr{F}, P)$ 为概率空间，ξ 为随机变量，其分布函数为 F_ξ，如果 $\int_{-\infty}^{\infty} |x| \mathrm{d}F_\xi(x) < \infty$，则称

$$E[\xi] = \int_{-\infty}^{\infty} x \mathrm{d}F_\xi(x) \tag{1.1.1}$$

为 ξ 的**数学期望**(或称**均值**).

可以证明，若 g 为定义在实直线上的实值函数，ξ 为随机变量，$\eta = g(\xi)$，并且 $\int_{-\infty}^{\infty} |y| \mathrm{d}F_\eta(y)$，则

$$E[\eta] = \int_{-\infty}^{\infty} g(x) \mathrm{d}F_\xi(x). \tag{1.1.2}$$

此时，若 ξ 为离散型随机变量，取值为 $a_i, i = 1, 2, \cdots$，$p_i = P(\xi = a_i), i = 1, 2, \cdots$，则

$$E[\eta] = \sum_{i=1}^{\infty} g(a_i) p_i; \tag{1.1.3}$$

若 ξ 为连续型随机变量，分布密度函数为 f_ξ，则

$$E[\eta] = \int_{-\infty}^{\infty} g(x) f_\xi(x) \mathrm{d}x. \tag{1.1.4}$$

方差是刻画随机变量取值分散程度的又一个重要数字特征，它定义为随机变量与其均值差距平方的平均值，即

$$\mathrm{Var}[\xi] = E[(\xi - E[\xi])^2]. \tag{1.1.5}$$

另外，设 $m \geqslant 1$ 为整数，若 $E[|\xi|^m] < \infty$，则称 $E[\xi^m]$ 为 ξ 的 m 阶**原点矩**，而称 $E[(\xi - E[\xi])^m]$ 为 ξ 的 m 阶**中心矩**.

对于随机向量 $(\xi_1, \xi_2, \cdots, \xi_n)'$，定义其数学期望(向量)和协方差(矩阵)为

$$E[(\xi_1, \xi_2, \cdots, \xi_n)'] = (E[\xi_1], E[\xi_2], \cdots, E[\xi_n])'; \tag{1.1.6}$$

$$\mathrm{Var}[(\xi_1, \xi_2, \cdots, \xi_n)'] = (\mathrm{Cov}(\xi_i, \xi_j))_{1 \leqslant i, j \leqslant n}. \tag{1.1.7}$$

其中，

$$\mathrm{Cov}(\xi_i, \xi_j) = E[(\xi_i - E[\xi_i])(\xi_j - E[\xi_j])], \tag{1.1.8}$$

称为 ξ_i 与 ξ_j 的**协方差**，而称

$$r(\xi_i, \xi_j) = \frac{\mathrm{Cov}(\xi_i, \xi_j)}{\sqrt{\mathrm{Var}[\xi_i]} \cdot \sqrt{\mathrm{Var}[\xi_j]}}$$

为 ξ_i 与 ξ_j 的相关系数.

若 $\xi_1, \xi_2, \cdots, \xi_n$ 都有有限的数学期望，则对任意实数 $k_1, k_2, \cdots, k_n$，有

$$E[k_1\xi_1 + k_2\xi_2 + \cdots + k_n\xi_n] = k_1E[\xi_1] + k_2E[\xi_2] + \cdots + k_nE[\xi_n]. \tag{1.1.9}$$

1.1.2.3 常用随机变量的分布及其数字特征

为了方便读者阅读后文，将数理统计中常见的随机变量列于表 1.1.

表 1.1 常见随机变量的分布及其数字特征

分布名称	R 中的名称	分 布	记号	参 数	均值	方差
两点分布	binorm	$P(\xi=k)=p^k(1-p)^{1-k},\ k=0,1$	$B(1,p)$	$0<p<1$	p	$p(1-p)$
二项分布	binorm	$P(\xi=k)=\binom{n}{k}p^k(1-p)^{n-k}$, $k=0,1,2,\cdots,n$	$B(n,p)$	$n>0,0<p<1$	np	$np(1-p)$
Poisson 分布	pois	$P(\xi=k)=\frac{e^{-\lambda}\lambda^k}{k!},\ k=0,1,2,\cdots$	$\mathrm{Poi}(\lambda)$	$\lambda>0$	λ	λ
几何分布	geom	$P(\xi=k)=(1-p)^{k-1}p,\ k=1,2,\cdots$	$\mathrm{Geo}(p)$	$0<p<1$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
负二项分布	nbinorm	$P(\xi=k)=\binom{k-1}{l-1}p^l(1-p)^{k-l}$, $k=l,l+1,\cdots$	$\mathrm{Pas}(l,p)$	$l>0,0<p<1$	$\frac{l}{p}$	$\frac{l(1-p)}{p^2}$
均匀分布	unif	$f_\xi(x)=\begin{cases}\frac{1}{b-a} & x\in(a,b)\\ 0 & 其它\end{cases}$	$U(a,b)$	$a<b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
正态分布	norm	$f_\xi(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{\frac{-(x-\mu)2}{2\sigma^2}}$	$N(\mu,\sigma^2)$	$-\infty<\mu<\infty,\sigma>0$	μ	σ^2
指数分布	exp	$f_\xi(x)=\begin{cases}\lambda e^{-\lambda x} & x>0\\ 0 & 其它\end{cases}$	$\mathrm{Exp}(\lambda)$	$\lambda>0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Γ 分布	gamma	$f_\xi(x)=\begin{cases}\frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x} & x>0\\ 0 & 其它\end{cases}$	$\Gamma(\alpha,\lambda)$	$\alpha\geqslant1,\lambda>0$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$

1.1.3 条件概率与独立性

条件概率与独立性是概率论中的两个十分重要的概念，我们在此作简单复习.

1.1.3.1 条件概率

定义 1.1.7 设 $(\Omega,\mathscr{F},P)$ 为概率空间, $A,B\in\mathscr{F}$, 且 $P(B)>0$, 则定义

$$P(A|B)=\frac{P(AB)}{P(B)}. \tag{1.1.10}$$

为 B 发生的条件下 A 发生的**条件概率**.

显然，若取定 B 且 $P(B)>0$, 则 $P_B(\cdot)=P(\cdot|B)$ 为 $(\Omega,\mathscr{F})$ 上的概率测度，亦即 $(\Omega,\mathscr{F},P_B)$ 也是概率空间.

命题 1.1.1(乘法公式) 设 $(\Omega,\mathscr{F},P)$ 为概率空间,

$$A_i\in\mathscr{F},i=1,2,\cdots,n,\quad 且\ P(A_1A_2\cdots A_{n-1})>0,$$

则

$$P(A_1A_2\cdots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1A_2\cdots A_{n-1}). \tag{1.1.11}$$

命题1.1.2(全概率公式) 设 $(\Omega,\mathscr{F},P)$ 为概率空间, $B_i \in \mathscr{F}, P(B_i) > 0, i = 1,2,\cdots,$ $B_i \cap B_j = \varnothing, i \neq j$, 且 $A \subset \bigcup_{i=1}^{\infty} B_i$, 则

$$P(A) = \sum_{i=1}^{\infty} P(B_i)P(A|B_i). \tag{1.1.12}$$

命题1.1.3(Bayes 公式) 在命题1.1.2 的条件下, 若 $P(A) > 0$, 则

$$P(B_j|A) = \frac{P(B_j)P(A|B_j)}{\sum_{i=1}^{\infty} P(B_i)P(A|B_i)}.$$

1.1.3.2 独立性

关于随机事件和随机变量的独立性, 用一个定义给出.

定义1.1.8 设 $(\Omega,\mathscr{F},P)$ 为概率空间, 记 $I = \{1,2,\cdots,n\}$.

(1) 如果对 $\{1,2,\cdots,n\}$ 的任意非空子集 $\{i_1,i_2,\cdots,i_m\}$, 有

$$P\left(\bigcap_{k=1}^{m} A_{i_k}\right) = \prod_{k=1}^{m} P(A_{i_k}), \tag{1.1.13}$$

那么称事件 $A_i \in \mathscr{F}(i = 1,2,\cdots,n)$ 相互独立;

(2) 如果

$$F_{\xi_1,\xi_2,\cdots,\xi_n}(x_1,x_2,\cdots,x_n) = F_{\xi_1}(x_1)F_{\xi_2}(x_2)\cdots F_{\xi_n}(x_n), \tag{1.1.14}$$

那么称随机变量 $\xi_1,\xi_2,\cdots,\xi_n$ 相互独立.

对于独立随机变量, 有以下命题:

命题1.1.4 设 $(\Omega,\mathscr{F},P)$ 为概率空间, $\xi_1,\xi_2,\cdots,\xi_n$ 为其上的随机变量,

(1) 若 $\xi_1,\xi_2,\cdots,\xi_n$ 独立, 且都有有限的数学期望, 则

$$E\left[\prod_{i=1}^{n} \xi_i\right] = \prod_{i=1}^{n} E[\xi_i];$$

(2) 若 $\xi_1,\xi_2,\cdots,\xi_n$ 独立, 且都有有限的方差, 则

$$\mathrm{Var}\left[\sum_{i=1}^{n} \xi_i\right] = \sum_{i=1}^{n} \mathrm{Var}[\xi_i].$$

1.1.4 大数定律与中心极限定理

本节简单叙述大数定律和中心极限定理的有关主要结果, 数理统计的许多处理方法的思想都基于这些基本事实.

命题1.1.5(弱大数定律) 设 $\xi_1,\xi_2,\cdots$ 为独立同分布随机变量序列, 具有有限的数学期望 μ 和方差 σ^2, 则 $n^{-1}\sum_{k=1}^{n} \xi_k$ 依概率收敛于 μ, 即对任意 $\varepsilon > 0$, 有

$$\lim_{n\to\infty} P\left(\left|n^{-1}\sum_{k=1}^{n} \xi_k - \mu\right| \geqslant \varepsilon\right) = 0. \tag{1.1.15}$$

命题1.1.6(强大数定律) 设 $\xi_1,\xi_2,\cdots$ 为独立同分布随机变量序列, 具有有限的数学

期望μ和方差σ^2，则$n^{-1}\sum_{k=1}^{n}\xi_k$几乎必然收敛于$\mu$，即

$$P\left(\lim_{n\to\infty}n^{-1}\sum_{k=1}^{n}\xi_k=\mu\right)=1.$$

命题1.1.7(中心极限定理) 设$\xi_1,\xi_2,\cdots$为独立同分布随机变量序列，具有有限的数学期望μ和方差σ^2，则有

$$\lim_{n\to\infty}P\left([\sigma\sqrt{n}]^{-1}\left[\sum_{k=1}^{n}\xi_k-n\mu\right]\leqslant x\right)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\mathrm{e}^{-\frac{u^2}{2}}\mathrm{d}u.$$

1.2 总体、样本和统计量

上节中已经提到，数理统计的基本特征是用局部推断整体．这个整体在数理统计中称为**总体**，也就是为了某一目的而要研究的对象的全体，而将每个对象称为个体或样本．但是，实际中我们往往关心的是研究对象某方面的数量特征，比如灯泡的寿命、一台机器正常工作的持续时间、某种药物的疗效，等等. 由于在对一个个体进行试验或观测结束之前，我们无法预知该数量的取值，这一点类似于上一节所述的随机变量. 所以，我们可以认为**总体就是一个随机变量**，每次试验后，获得的一个个体的具体取值就是该随机变量的一次观测值. 另外，若我们准备抽取n个个体进行试验或观测，在试验或观测结束之前，也无法预知各个个体的取值，所以从概率分析的角度来说，它们的取值也应该为随机变量.

总之，我们以后所说的总体，就是一个随机变量ξ，需要通过观测n个样本来推断它的分布或数字特征. 通常记ξ的分布为$F_\xi(x,\theta)$，其中$x\in(-\infty,\infty)$为实变元，$\theta\in\Theta$为参数，Θ称为参数空间，它可能包含多个参数. 样本为n维随机向量$(\xi_1,\xi_2,\cdots,\xi_n)$，$n$称为**样本容量**. 为处理方便，初等数理统计都假定$\xi_i(i=1,2,\cdots,n)$与ξ同分布，且相互独立. 此时称之为**简单随机样本**，简称为**样本**.

基于以上讨论和随机变量独立性的定义(见定义1.1.8)，有如下的命题：

命题1.2.1 设总体$\xi\sim F_\xi(x,\theta)$，$\xi_1,\xi_2,\cdots,\xi_n$为其样本，则

$$F_{\xi_1,\xi_2,\cdots,\xi_n}(x_1,x_2,\cdots,x_n;\theta)=F_\xi(x_1,\theta)F_\xi(x_2,\theta)\cdots F_\xi(x_n,\theta). \tag{1.2.1}$$

其中，$F_\xi(x,\theta)$可以是分布函数，也可以是分布密度函数(对于连续型随机变量)或分布律(对于离散型随机变量).

按上所述，总体ξ是我们研究的目标，而出发点是样本$(\xi_1,\xi_2,\cdots,\xi_n)$，那么研究的途径或手段就是所谓的**统计量**. 它的直观意思是要通过观测值的处理来回答或推断总体的分布、数字特征、参数等是怎样的. 因此，统计量就是样本$(\xi_1,\xi_2,\cdots,\xi_n)$的一个函数，但其中不能含未知参数，否则由观测值计算不出任何结果.

常用的统计量有以下几种：

(1) **样本均值**：$\bar{\xi}=\dfrac{1}{n}(\xi_1+\xi_2+\cdots+\xi_n)=\dfrac{1}{n}\sum_{i=1}^{n}\xi_i$；

(2) **样本方差**：$S_n^2=\dfrac{1}{n}\sum_{i=1}^{n}(\xi_i-\bar{\xi})^2$；

修正样本方差：$S_n^{*2}=\dfrac{1}{n-1}\sum_{i=1}^{n}(\xi_i-\overline{\xi})^2$；

(3) **样本 k 阶原点矩：**$\overline{\xi^k}=\dfrac{1}{n}\sum_{i=1}^{n}\xi_i^k$；

(4) **样本 k 阶中心矩：**$\dfrac{1}{n}\sum_{i=1}^{n}(\xi_i-\overline{\xi})^k$；

(5) **顺序统计量：**$\xi_{(1)}\leqslant\xi_{(2)}\leqslant\cdots\leqslant\xi_{(n)}$，其中 $\xi_{(1)}=\min\{\xi_1,\xi_2,\cdots,\xi_n\}$，$\xi_{(n)}=\max\{\xi_1,\xi_2,\cdots,\xi_n\}$，而 $\xi_{(k)}$ 是将 $\xi_1,\xi_2,\cdots,\xi_n$ 的取值从小到大排列第 k 位的值；

(6) **样本中位数：**

$$\tilde{\xi}=\begin{cases}\xi_{\left(\frac{n+1}{2}\right)} & \text{若 } n \text{ 为奇数}\\ \dfrac{1}{2}\left(\xi_{\left(\frac{n}{2}\right)}+\xi_{\left(\frac{n}{2}+1\right)}\right) & \text{若 } n \text{ 为偶数}\end{cases}；$$

(7) **样本极差：**$R_n^{\xi}=\xi_{(n)}-\xi_{(1)}$；

(8) **样本相关系数：**设另有总体 η 和样本(η_1，η_2，…，η_n)，样本相关系数定义为

$$r_{\xi\eta}=\frac{\sum_{i=1}^{n}(\xi_i-\overline{\xi})(\eta_i-\overline{\eta})}{\sqrt{\sum_{i=1}^{n}(\xi_i-\overline{\xi})^2}\cdot\sqrt{\sum_{j=1}^{n}(\eta_j-\overline{\eta})^2}}.$$

1.3　抽样分布

为了用概率方法来探讨一个统计量在推断总体时的性能或把握推断结论的置信程度，我们必须要知道统计量的分布或近似分布．因此，下面先介绍一些具体统计量的分布(通常称为抽样分布)．

例 1.3.1　设总体 $\xi\sim N(0,1)$，$\xi_1,\xi_2,\cdots,\xi_n$ 为其样本，试求 $\eta=\xi_1^2+\xi_2^2+\cdots+\xi_n^2$ 的分布．

解　先求 $\eta=\xi^2$ 的分布．显然，对 $y<0$，有 $P(\xi^2\leqslant y)=0$．对 $y\geqslant0$，有

$$F_\eta(y)=P(\xi^2\leqslant y)=P(-\sqrt{y}\leqslant\xi\leqslant\sqrt{y})=\int_{-\sqrt{y}}^{\sqrt{y}}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx.$$

从而，对 $y\geqslant0$，η 的分布密度函数为

$$\begin{aligned}f_\eta(y)&=\frac{1}{\sqrt{2\pi}}e^{-\frac{(\sqrt{y})^2}{2}}\times\left(\frac{1}{2\sqrt{y}}\right)-\frac{1}{\sqrt{2\pi}}e^{-\frac{(-\sqrt{y})^2}{2}}\times\left(-\frac{1}{2\sqrt{y}}\right)\\&=\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{y}}e^{-\frac{y}{2}}\\&=\frac{\left(\frac{1}{2}\right)^{\frac{1}{2}}}{\Gamma\left(\frac{1}{2}\right)}y^{-\frac{1}{2}}e^{-\frac{y}{2}},\end{aligned}$$

其中用到 $\Gamma\left(\frac{1}{2}\right)=\sqrt{\pi}$.

总之，
$$f_\eta(y)=\begin{cases}\dfrac{\left(\frac{1}{2}\right)^{\frac{1}{2}}}{\Gamma\left(\frac{1}{2}\right)}y^{-\frac{1}{2}}e^{-\frac{y}{2}} & 当\ y>0\\ 0 & 当\ y\leqslant 0\end{cases}.$$

这说明 ξ^2 服从表 1.1 中的 Γ 分布，其中的参数 $\alpha=\lambda=\frac{1}{2}$.

用初等概率论中求独立随机变量和的分布公式，容易证明，若 $\eta_1\sim\Gamma(\alpha_1,\lambda)$，$\eta_2\sim\Gamma(\alpha_2,\lambda)$，且 η_1 与 η_2 独立，则 $\eta_1+\eta_2\sim\Gamma(\alpha_1+\alpha_2,\lambda)$. 因此，可以得到

$$\eta=\xi_1^2+\xi_2^2+\cdots+\xi_n^2\sim\Gamma\left(\frac{n}{2},\frac{1}{2}\right).$$ [解毕].

【注 1.3.1】 (1) 在数理统计中，通常称 $\Gamma(\frac{n}{2},\frac{1}{2})$ 分布为**自由度为 n 的 χ^2 分布**，记为 $\chi^2(n)$.

(2) 若 $\xi\sim N(0,1)$，$\eta\sim\chi^2(n)$，且 ξ 与 η 独立，则称 $\frac{\xi}{\sqrt{\eta/n}}$ 所服从的分布为**自由度为 n 的 t 分布**，记为 $t(n)$.

(3) 若 $\xi\sim\chi^2(m)$，$\eta\sim\chi^2(n)$，且 ξ 与 η 独立，则称 $\frac{\xi/m}{\eta/n}$ 所服从的分布为**第一自由度为 m、第二自由度为 n 的 F 分布**，记为 $F(m,n)$.

(4) 由于 $\chi^2(n)$、$t(n)$ 和 $F(m,n)$ 的分布密度函数的不定积分没有有限形式，有关概率的计算只能查表，我们在此不写出它们的分布密度函数. 但需指出，$t(n)$ 分布与 $N(0,1)$ 分布很接近，$\chi^2(n)$ 和 $F(m,n)$ 分布的随机变量非负，其密度函数在其变元为负时取零值. 这三种分布的密度函数曲线如图 1.1 ～ 1.3 所示.

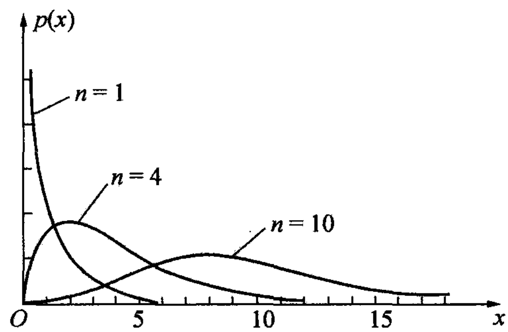

图 1.1 $\chi^2(n)$ 分布密度函数曲线示意图

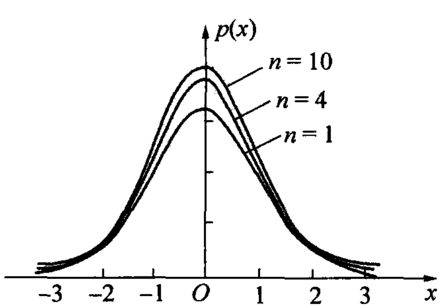

图 1.2 $t(n)$ 分布密度函数曲线示意图

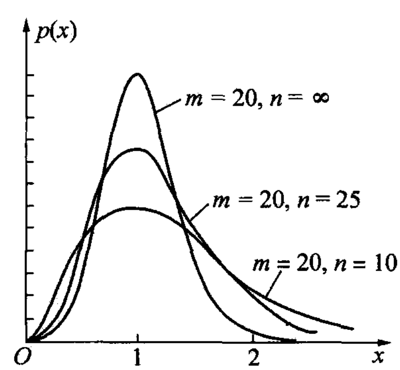

图 1.3 $F(m,n)$ 分布密度函数曲线示意图