高等学校教材

概率论及数理统计

下 册

(第二版)

中山大学数学系
梁之舜 邓集贤
杨维权 司徒荣 邓永录
编 著

高等教育出版社

本书是中山大学数学力学系编《概率论及数理统计》(1980)的修订第二版，现改为编者个人署名．第二版与第一版相比，有不少小的修改．总的结构与第一版相同，仍分上、下册出版．

为了突出本课程教学的主要内容，编者在下册的修订中删去了第一版中个别的节、段以及个别定理的证明，还有些改为小字排印或标上了星号，以作为选学内容．总的篇幅略有减少．

本书可作为综合大学和师范院校数学专业、数理统计专业、应用数学专业的教材，也可作其它有关专业的参考书．

学习本书只要求具有初等微积分基础知识，因此本书具有适应面广、便于自学的特点．

高等学校教材

概率论及数理统计

下　册

（第二版）

中山大学数学系

梁之舜　邓集贤
杨维权　司徒荣　邓永录　编著

*

高等教育出版社出版

新华书店北京发行所发行

河北省香河县印刷厂印装

*

开本 850×1168 1/32　印张 11.75 字数 280 000

1980年7月第1版　1988年10月第2版　1988年10月第1次印刷

印数 0001— 2 820

ISBN7 04-000569-7/O·207

定价 3.15 元

目　录

第六章 抽 样 分 布

前述各章，是概率论的引论，讲述了概率论的基本内容，为数理统计学建立了重要的数学基础. 从本章起的接连四章，是数理统计学的初步，主要讲述估计与检验等原理，回归分析与方差分析等统计方法.

数理统计学是运用概率论的基本知识，对要研究的随机现象进行多次观察或试验，研究如何合理地获得数据资料，建立有效的数学方法，根据所获得的数据资料，对所关心的问题作出估计与检验. 数理统计学的重要分支有统计推断、多元统计分析、试验设计等，其具体方法甚多，应用相当广泛，已成为各学科从事科学研究及生产、经济等部门进行有效工作的必不可少的数学工具.

在这一章中，我们从数理统计学的基本概念讲起，讨论抽样分布及其重要的定理. 这些抽样分布及其几个重要的定理，在前述各章中尚未提到，而在后述三章中却经常要用到它们.

§6.1 基 本 概 念

一、总体、个体、简单随机子样

总体、个体、子样是数理统计学中三个最基本的术语. 我们把对某一个问题的研究对象的全体称为总体(或母体)，组成总体的每个基本单元称为个体，从总体中随机抽取的 n 个个体称为容量为 n 的子样. 例如，把某月的整批产品视为一总体，则每个产品为个体；又如把某批灯泡视为一总体，则每个灯泡为个体；某地在某季度内每天的日平均气温的全体视为一总体，则其中某天的日平

均气温为个体．在数理统计学中，我们是对总体成员的一个或者若干个数量表征进行研究．如日平均气温的度数用ξ表示，灯泡的使用寿命(小时)用η表示，这样对总体的研究就归结为讨论随机变数ξ或η的分布函数及其主要数字特征(如数学期望$E(\xi)$、方差$D(\xi)$等)的研究．又如某月的整批产品为一总体，若只要检查产品的质量，可用数量表征ξ来反映．产品为一等品，令ξ取值为1；二等品，ξ取值为2；不合格品，ξ取值为3．我们研究这批产品的质量规律，就归结为讨论随机变数ξ的分布函数及其主要数字特征了．

今后我们常用“总体ξ服从什么分布”这样的术语，它是指总体的某个具体数量表征ξ服从什么分布规律．我们说对总体进行n次独立的重复试验或n次独立的观察，就是从总体中随机地抽取容量为n的子样，对应的数量表征以$\xi_1, \cdots, \xi_n$表示，不难理解，ξ及$\xi_1, \cdots, \xi_n$都是随机变数，$(\xi_1, \cdots, \xi_n)$是n维随机向量．例如我们说整批灯泡的寿命分布如何，是指灯泡的寿命ξ服从什么分布．从整批灯泡中随机抽取容量为n的子样，是指随机抽取n个灯泡，观察这n个灯泡的寿命，用$\xi_1, \cdots, \xi_n$表示，其中ξ_i表示随机抽取到的第i个灯泡的寿命，显然ξ_i也是一随机变数$(i=1, \cdots, n)$．为区别总体与子样的概念，我们今后常用诸如“总体ξ服从什么分布”、“从总体ξ中抽取子样$\xi_1, \cdots, \xi_n$”等这类术语．

我们在对总体进行研究中，当总体相当大时（指个体相当多时)或当对个体进行试验具有破坏性(如炮弹能否引发爆炸）或费时耗资大时，只可能从总体中抽取部分个体来作研究，即抽取容量为n的子样来作研究．我们要从子样的观察或试验结果的特性来对总体的特性作出估计与推断，一方面自然要研究应该怎样从总体中抽取子样，使得子样在尽可能大的程度上反映总体的特性；同时必须建立一整套的方法，使能根据所选取的子样的性质，来对总

体的特性进行估计与推断．因此，我们在抽取子样来对总体作出估计与推断时，从总体中抽取子样必须是随机的，即每一个体都有同等概率被抽取(当总体中的个体是有限个时，要用有返回抽取方式)．其具体要求为两个方面：独立性，是指$\xi_1,\cdots,\xi_n$相互独立；代表性，是指$\xi_1,\cdots,\xi_n$中每一个都与总体ξ有相同分布．

定义 6.1.1 (简单随机子样) 设$\xi_1,\cdots,\xi_n$为来自总体ξ的容量为n的子样，如果$\xi_1,\cdots,\xi_n$相互独立且每一个都是与总体ξ有相同分布的随机变数，则称$\xi_1,\cdots,\xi_n$为总体ξ的容量为n的简单随机子样，简称为简单子样或子样①．

在这一章中，我们所讲的子样$\xi_1,\cdots,\xi_n$，无特别声明的话，都是指简单子样，即$\xi_1,\cdots,\xi_n$相互独立且每一个都与总体ξ有相同的分布．引入简单随机子样，是基于要从子样所获得的概率统计特性来对总体的概率统计特性作出估计与推断．对于简单随机子样，我们可以应用概率论中对独立随机变数的情形所建立的许多重要的定理，这些重要的结论为数理统计学提供了必要的基础．

二、统计量

子样$\xi_1,\cdots,\xi_n$也可用n维随机向量$(\xi_1,\cdots,\xi_n)$表示．记x_i为ξ_i的一次观察值，并称$(x_1,\cdots,x_n)$为子样的一次观察值．子样$(\xi_1,\cdots,\xi_n)$的所有可能取值的全体称为子样空间，记作$\mathscr{X}$，它是n维空间．子样的一次观察值$(x_1,\cdots,x_n)$就是子样空间$\mathscr{X}$中的一个点，即$(x_1,\cdots,x_n)\in\mathscr{X}$．

定义 6.1.2(统计量) 设$\xi_1,\cdots,\xi_n$为总体ξ的子样，T为子样空间$\mathscr{X}$中点$(x_1,\cdots,x_n)$的实值函数，作子样的函数$T=T(\xi_1,$

---

① 是指若ξ为定义在概率空间$(\Omega,\mathscr{F},P)$上的随机变数，则$\xi_1,\cdots,\xi_n$为概率空间(Ω,ξ,P)上n个相互独立的并同ξ具有相同分布的随机变数．

$\cdots,\xi_n)$，T 的取值记为 $t=T(x_1,\cdots,x_n)$. 若 $T(\xi_1,\cdots,\xi_n)$ 也为一随机变数①，且不带未知参数，则称 T 或 $T(\xi_1,\cdots,\xi_n)$ 为统计量.

我们在用子样 $\xi_1,\cdots,\xi_n$ 获得的信息来对总体 ξ 作出估计与推断时，是按不同的统计问题的要求而规定子样的各种函数. 在这本教材中，所涉及到的子样的各种函数 $T(\xi_1,\cdots,\xi_n)$，$T(x_1,\cdots,x_n)$ 一般都是多维随机变量的连续函数，因而 $T(\xi_1,\cdots,\xi_n)$ 都是随机变数.

例 6.1.1 设 $\xi_1,\cdots,\xi_n$ 为总体 ξ 的子样，其容量为 n. 记

$$\bar{\xi}=\frac{1}{n}\sum_{i=1}^{n}\xi_i,\qquad S^2=\frac{1}{n}\sum_{i=1}^{n}(\xi_i-\bar{\xi})^2,$$

则 $\bar{\xi}$ 及 S^2 都是统计量，称 $\bar{\xi}$ 及 S^2 分别为子样 $\xi_1,\cdots,\xi_n$ 的平均值及方差. 子样的观察值为 $x_1,\cdots,x_n$，$\bar{\xi}$ 及 S^2 的观察值分别记作

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i,\qquad s^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2,$$

今后，大写的 S^2 表示统计量，小写的 s^2 表示统计量 S^2 的观察值.

定义 6.1.3(顺序统计量) 设 $\xi_1,\cdots,\xi_n$ 为总体 ξ 的子样，今由子样建立 n 个函数:

$$\xi_k^*=\xi_k^*(\xi_1,\cdots,\xi_n)\qquad k=1,\cdots,n,$$

其中 ξ_k^* 为这样的统计量，它的观察值为 x_k^*，x_k^* 为子样 $\xi_1,\cdots,\xi_n$ 的观察值 $x_1,\cdots,x_n$ 中由小至大排列(即 $x_1^*\leqslant\cdots\leqslant x_k^*\leqslant\cdots\leqslant x_n^*$)后的第 k 个数值，则称 $\xi_1^*,\cdots,\xi_n^*$ 为顺序统计量.

易见，$\xi_1^*=\min\{\xi_1,\cdots,\xi_n\}$，$\xi_n^*=\max\{\xi_1,\cdots,\xi_n\}$. 称 ξ_1^* 为最小项统计量，ξ_n^* 为最大项统计量. 若 n 为奇数，则称 $\xi_{\frac{n+1}{2}}^*$ 为子样

① 具体要求就是 $T(\xi_1,\cdots,\xi_n)$ 是 n 元可测实函数，也就是在第二章中所述的 n 元随机变数，它满足对每一 $t\in R_1$，有 $\{T(\xi_1,\cdots,\xi_n)\leqslant t\}\in\mathscr{F}$.

的中值；若 n 为偶数，则称 $\xi^*_{\frac{n}{2}+1}$ 为子样的中值.

定义 6.1.4(极差) 设 $\xi_1, \cdots, \xi_n$ 为总体 ξ 的子样，则称统计量 $D_n^* \underset{def}{=} \xi_n^* - \xi_1^*$ 为子样的极差.

它是子样中最大值与最小值之差，反映了子样观察值的波动幅度. 它同方差一样是反映观察值离散程度的数量指标，而且计算方便.

例 6.1.2 设 $\xi_1, \cdots, \xi_5$ 为 ξ 的容量为 5 的子样，今对这个子样作了三次观察，其值如表 6.1.1 所示，试求 $\bar{\xi}$、S^2 及 D_n^* 的观察值.

表 6.1.1

ξ	ξ_1	ξ_2	ξ_3	ξ_4	ξ_5
1	3	1	10	5	6
2	2	6	7	2	8
3	8	3	9	10	5

表 6.1.2

ξ_1^*	ξ_2^*	ξ_3^*	ξ_4^*	ξ_5^*	$\bar{\xi}$	s^2	D_n^*
1	3	5	6	10	5	9.2	9
2	2	6	7	8	5	6.4	6
3	5	8	9	10	7	6.8	7

三、小样问题与大样问题

统计量是我们对总体 ξ 的分布函数或数字特征进行估计与推断最重要的基本概念，求出统计量 $T(\xi_1, \cdots, \xi_n)$ 的分布函数是数理统计学的基本问题之一. 统计量的分布，称为抽样分布.

设总体 ξ 的分布函数表达式已知，对于任一自然数 n，如能求

出给定统计量$T(\xi_1,\cdots,\xi_n)$的分布函数，这分布称为统计量T的精确分布．求出统计量T的精确分布，这对于数理统计学中的所谓小样问题（即在子样容量n比较小的情况下所讨论的各种统计问题）的研究是很重要的．

但一般说来，要确定一个统计量的精确分布其难度比较大．只对一些重要的特殊情形，如总体ξ服从正态分布时，已求出t统计量、χ^2统计量、F统计量等的精确分布．它们在参数的估计及检验中起很重要的作用．

若统计量$T(\xi_1,\cdots,\xi_n)$的精确分布求不出来，或其表达式非常复杂而难于应用，但如能求出它在$n\to\infty$时的极限分布，那么这个统计量的极限分布对于数理统计学中的所谓大样问题（即在子样容量n比较大的情况下讨论的各种统计问题）的研究很有用．但要注意，在应用极限分布时，要求子样的容量n比较大．如第八章的§8.3所讨论的非参数性假设检验问题，是用检验统计量的极限分布，因而子样容量n应取得比较大才行．

§6.2　子样的数字特征及其分布

*一、经验分布与格列汶科定理

在实际工作中遇到各种各样的随机变数，怎样确定它的分布函数$F(x)$？在概率论中，我们介绍了常用的几种分布函数以及它们的一些性质，在那里我们假定它们都是事先给定了的．现在，我们介绍用独立重复试验的方法，利用子样建立一定的概率模型，用由此所获得的概率统计特性来对总体ξ的分布函数$F(x)$等作出估计与推断．

定义6.2.1（经验分布函数）　从总体ξ中抽取容量为n的子样$\xi_1,\cdots,\xi_n$，当顺序统计量$\xi_1^*,\xi_2^*,\cdots,\xi_n^*$的值给定时，对任何实数$x$，我们定义函数$F_n^*(x)$：

$$F_n^*(x)=\begin{cases}0, x\leqslant x_1^*\\ \dfrac{k}{n}, x_k^*<x\leqslant x_{k+1}^*, k=1,\cdots,n-1,\\ 1, x>x_n^*\end{cases} \tag{6.2.1}$$

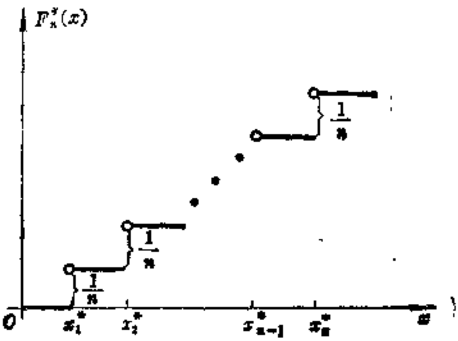

图 6.2.1

称 $F_n^*(x)$ 为总体 ξ 的经验分布函数①.

易见,对于每一组观察值 $\xi_i^*=x_i^*, i=1,\cdots,n$, $F_n^*(x)$ 单调、非降、左连续且在 $x=x_k^*$ 点有间断点, 在每个间断点上跳跃量都是 $\frac{1}{n}$. 显然, $0\leqslant F_n^*(x)\leqslant 1$, 并具有分布函数的其它性质.

由定义 6.2.1 知, 对于 x 的每一数值而言, 经验分布函数 $F_n^*(x)$ 为子样 $\xi_1,\cdots,\xi_n$ 的函数, 它是一统计量, 即为一随机变数, 其可能取值为 $0,\frac{1}{n},\cdots,\frac{n-1}{n}, 1$. 事件"$F_n^*(x)=\frac{k}{n}$"发生的概率, 由于 $\xi_1,\cdots,\xi_n$ 相互独立且有相同的分布函数 $F(x)$, 因而它等价于

① 对于每一 ω, 即子样的一次观察值, 由(6.2.1)定义的 $F_n^*(x)$ 是实数 x 的函数, $x\in R_1$. 而对于每一 x 值, $F_n^*(x)$ 依赖于子样的观察值, 即为 ω 的函数, 因而 $F_n^*(x)$ 为一统计量. 这里, 如果 $x_k<x_{k+1}=x_{k+2}=\cdots$, 而 $x_k<x=x_{k+1}=x_{k+2}=\cdots$, 则应在(6.2.1)中把 $\frac{k}{n}$ 改为 $\frac{N_k(x)}{n}$, 其中 $N_k(x)$ 是子样中小于 x_{k+1} 的观察值的个数.

n 次独立重复试验的贝努里概型中事件“$\xi<x$”发生 k 次而其余 $n-k$ 次不发生的概率，即有：

$$P\left\{F_n^*(x)=\frac{k}{n}\right\}=C_n^k\{F(x)\}^k\{1-F(x)\}^{n-k} \qquad (6.2.2)$$

其中 $F(x)=P(\xi<x)$，它是总体 ξ 的分布函数.

我们从第五章所讲的大数定律知道，在一定的条件下，事件发生的频率依概率收敛于这个事件发生的概率. 人们自然要问，总体 ξ 的经验分布函数 $F_n^*(x)$，即“$\xi<x$”事件发生的频率，当 n 足够大时，是否也渐近于事件“$\xi<x$”发生的概率，即总体 ξ 的分布函数 $F(x)$ 呢？格列汶科于 1933 年作了肯定的回答.

格列汶科定理 设总体 ξ 的分布函数为 $F(x)$，经验分布函数为 $F_n^*(x)$，对于任何实数 x，记

$$D_n=\sup_{-\infty<x<+\infty}|F_n^*(x)-F(x)|, \qquad (6.2.3)$$

则有

$$P\{\lim_{n\to\infty}D_n=0\}=1. \qquad (6.2.4)$$

我们知道 $F_n^*(x)$ 为一统计量，因而 D_n 也为一统计量. D_n 这个统计量用来衡量 $F_n^*(x)$ 同 $F(x)$ 之间在所有的 x 值上最大的差异程度. 格列汶科定理证明了统计量 D_n 以概率为 1 地收敛于零. 通俗地说，就是当 n 足够大时，对于所有的 x 值，$F_n^*(x)$ 同 $F(x)$ 之差的绝对值都很小这个事件发生的概率等于 1. 但是格列汶科定理还未阐明统计量 D_n 服从什么分布或以什么分布为其极限分布. 我们将在第八章的 §8.3 中叙述两个重要定理，即柯尔莫哥洛夫定理和斯米尔诺夫定理. 这两个定理找到了统计量 D_n 的极限分布表达式，这些定理在分布函数的假设检验中有重要的作用.

二、子样的数字特征

1. 子样矩

设 $F_n^*(x)$ 为总体 ξ 的经验分布函数，记

$$A_r=\int_{-\infty}^{\infty} x^r dF_n^*(x)=\frac{1}{n}\sum_{i=1}^{n}\xi_i^r, \tag{6.2.5}$$

$$B_r=\int_{-\infty}^{\infty}(x-\bar{\xi})^r dF_n^*(x)=\frac{1}{n}\sum_{i=1}(\xi_i-\bar{\xi})^r. \tag{6.2.6}$$

称 A_r 及 B_r 分别为子样的 r 阶原点矩及 r 阶中心矩，其中 r 为正整数. 显然 $A_1=\bar{\xi}$，$B_2=S^2$. 由于 A_r 及 B_r 都是子样的连续函数，因而都是统计量，也即都是随机变数.

2. 协方差与相关系数

设 $(\xi_1,\eta_1),\cdots,(\xi_n,\eta_n)$ 为二维总体 (ξ,η) 的子样，记

$$\bar{\xi}=\frac{1}{n}\sum_{i=1}^{n}\xi_i \qquad S_1^2=\frac{1}{n}\sum_{i=1}^{n}(\xi_i-\bar{\xi})^2$$

$$\bar{\eta}=\frac{1}{n}\sum_{i=1}^{n}\eta_i \qquad S_2^2=\frac{1}{n}\sum_{i=1}^{n}(\eta_i-\bar{\eta})^2 \tag{6.2.7}$$

$$S_{12}=\frac{1}{n}\sum_{i=1}^{n}(\xi_i-\bar{\xi})(\eta_i-\bar{\eta})$$

称统计量 S_{12} 为子样的协方差，称统计量

$$R=\frac{S_{12}}{S_1\cdot S_2} \tag{6.2.8}$$

为子样的相关系数.

三、子样数字特征的分布

我们在这一段叙述子样的平均值、极值与极差的分布，子样相关系数 R 的分布将在§8.3 中叙述，子样方差 S^2 的分布在下一节中叙述.

1. 子样平均值的分布

设 $\varphi(t)$ 为总体 ξ 的特征函数，$\xi_1,\cdots,\xi_n$ 为总体 ξ 的子样，则子样平均值 $\bar{\xi}$ 的特征函数为：

$$\varphi_1(t)=[\varphi_0(t)]^n=\left[\varphi\left(\frac{t}{n}\right)\right]^n, \qquad (6.2.9)$$

其中 $\varphi_0(t)=\varphi\left(\frac{t}{n}\right)$，它是$\frac{1}{n}\xi$的特征函数.

例 6.2.1 设总体 ξ 服从正态 $N(a, \sigma)$，求子样平均值 $\bar{\xi}$ 的分布.

解 因 ξ 的特征函数为

$$\varphi(t)=\exp\left[-\frac{\sigma^2}{2}t^2+jat\right],$$

则 $\bar{\xi}$ 的特征函数为

$$\varphi_1(t)=\left\{\exp\left[-\frac{\sigma^2}{2}\left(\frac{t}{n}\right)^2+ja\frac{t}{n}\right]\right\}^n$$
$$=\exp\left[-\frac{t^2}{2}\left(\frac{\sigma}{\sqrt{n}}\right)^2+jat\right],$$

即 $\bar{\xi}$ 服从正态 $N\left(a, \frac{\sigma}{\sqrt{n}}\right)$.

例 6.2.2 设总体 ξ 服从具有参数 λ 的泊松分布，求子样平均值 $\bar{\xi}$ 的分布.

解 因 ξ 的特征函数为

$$\varphi(t)=\exp[\lambda(e^{jt}-1)],$$

则 $\bar{\xi}$ 的特征函数为

$$\varphi_1(t)=\exp[n\lambda(e^{j\frac{t}{n}}-1)],$$

即 $\bar{\xi}$ 的概率分布为

$$P\left(\bar{\xi}=\frac{k}{n}\right)=e^{-n\lambda}\frac{(n\lambda)^k}{k!}, k=0,1,2,\cdots.$$

右端是参数为 $n\lambda$ 的泊松分布.

例 6.2.3 设总体 ξ 服从具有参数为 λ 的指数分布，求子样平均值 $\bar{\xi}$ 的分布.

ξ的密度函数为

$$f(x)=\begin{cases}\lambda e^{-\lambda x}, & x>0,\lambda>0,\\ 0\quad, & x\leqslant 0.\end{cases}$$

解 ξ的特征函数为

$$\varphi(t)=\frac{\lambda}{\lambda-jt}=\left(1-j\frac{t}{\lambda}\right)^{-1}\qquad \lambda>0,$$

则$\bar{\xi}$的特征函数为

$$\varphi_{\bar{\xi}}(t)=\left(1-j\frac{t}{n\lambda}\right)^{-n}.$$

可见子样平均$\bar{\xi}$服从具有参数$\alpha=n-1,\beta=\frac{1}{n\lambda}$的$\Gamma$-分布，记作$\Gamma\left(n-1,\frac{1}{n\lambda}\right)$.

应用第五章的定理5.4.3,再举一例.

例6.2.4 设总体ξ服从参数为p的二项分布，$0<p<1$，N为正整数,概率函数为:

$$P\{\xi=k\}=C_N^k p^k(1-p)^{N-k}\qquad k=0,1,\cdots,N,$$

求子样平均值$\bar{\xi}$的极限分布.

解 因$E(\xi)=Np\qquad D(\xi)=Np(1-p)\leqslant\frac{N}{4}$

则统计量

$$\eta_n=\frac{\sum_{i=1}^{n}\xi_i-E\left(\sum_{i=1}^{n}\xi_i\right)}{\sqrt{D\left(\sum_{i=1}^{n}\xi_i\right)}}=\frac{n\bar{\xi}-nNp}{\sqrt{nNp(1-p)}}$$

$$=\frac{\bar{\xi}-Np}{\sqrt{Np(1-p)/n}}$$

有极限分布$N(0,1)$,称$\bar{\xi}$有渐近正态分布$N\left(Np,\sqrt{\frac{Np(1-p)}{n}}\right)$

*2. 极值的分布