

概率统计

(修订版)

张尧庭 编

工程数学

中央广播电视台大学出版社

工 程 数 学
概 率 统 计
(修 订 版)
张尧庭 主编

中央广播电视台大学出版社

工程数学
概率统计
(修订版)
张尧庭 主编

*
中央广播电视台出版社出版
新华书店北京发行所发行
二二〇七工厂印装

*
开本 787×1092 1/32 印张 10 千字 208
1984年9月第1版 1987年5月第1次印刷
印数 1—38000
书号 13300·51 定价 1.65 元

第二版序言

这本教材经修改后出第二版了，在此我们说明一下这一版与原书相比作了哪些改动，以及改动的原因。原书使用以后，收到了各种各样的意见，主要集中在两点：1) 体系和别的书不同，无法看别的参考书；2) 书本身写得太简略，有的地方很难自学、看懂，例题及说明太少。

根据这些意见，我们作了如下的变动：

1. 体系方面进一步把概率和数理统计结合起来。这两者的结合对于教与学都是有利的，这个方向不能变，应该结合得越紧密越好。所以第二版在这方面的变动较大，原书的八章现改为十章，将难点分散，适当增加了随机事件、排列组合的内容，使初学者容易看其他的参考书。把条件分布、贝叶斯公式和贝叶斯方法集中在后面，又能结合，又避免了联合分布这一段难点过多。

2. 增加了不少例子，为了说明各种方法的应用和计算步骤。每一小节后面基本上都附有思考练习题，以便学生及时抓住本节的主要内容，弄清有关的概念。在内容上从两方面补充了一些材料：一是要用到的微积分中的一些知识，如 Γ 函数， β 函数，我们均在附录中列出，以便查阅，不需去找其他参考书。再如排列、组合，我们也另写一节，以便复习提高。二是一些涉及理论方面的讨论或证明，或作为附录，或作为正文，都详细写明，正文前打一个*号，表示这部份内容是不讲的，有兴趣的读者可以自己去看。需要注意的是：附录的证明是不要求的，但附录中的结果是正文中常常要用的，为了

节省篇幅，我们不在正文中将这些结果再复述一遍，目的也是希望大家去看看、翻翻附录的内容。

希望这样改动后，大家学习起来可以更少一些困难。欢迎大家提出批评、修改意见。参加教材修改的还有陆永余和张璋同志。借此机会，我们诚恳地向所有提出批评、建议的读者表示诚挚的感谢。

编 者

1986年8月

初 版 前 言

一本 50 学时能教完的书要介绍概率论、数理统计的全貌是困难的，必需有所取舍，才能重点突出，使学的人有所收获。本书的重点是数理统计，介绍一些概率论的概念是使得对一些统计的理论有更好的了解，因此舍去了一般教材中关于古典概型的比较详细的讨论。在数理统计中也是重点选取最有用的回归分析作为一条主线，目的是使学的人至少能掌握一种比较有用的方法。为此，书中对一些例子的计算步骤及结论作了较详细的讨论，希望通过这些实例的讨论，读者对于应用中遇到的问题会进行分析。本书所引用的数字的实例，基本上都是选自国内的实际应用的问题，并注明了数据的来源，但计算及讨论大都是不一样的，个别的例是完全引用别的书上的，我也在注中说明了。为了使读者能对数理统计应用的广泛性有一点感性的认识，本书的例子和习题有意选用了各个不同行业的材料，并不只限于工业的应用。

很明显，这是一本带有探索性的教材，是想探索一种新的编写体系与方法，使读者在学习时少遇到一些困难，多感到一些兴趣。是否能达到预期的效果，只能由读者来评定了。欢迎提出各种改进、批评的意见，使得如有再版、改写的可能时，来作进一步的修改。

编 者

1984 年 9 月

1984.9.6.

目 录

第一章 数据的简单分析	(1)
§ 1 准备知识.....	(1)
§ 2 几个概念.....	(9)
§ 3 加权平均数.....	(19)
§ 4 直方图.....	(26)
*§ 5 平方和分解.....	(30)
习题一.....	(32)
第二章 事件与概率	(35)
§ 1 随机事件与概率.....	(35)
*§ 2 排列、组合.....	(38)
§ 3 加法公式.....	(41)
§ 4 乘法公式.....	(50)
习题二.....	(56)
第三章 随机变量与概率分布	(58)
§ 1 随机变量.....	(58)
§ 2 分布密度.....	(62)
§ 3 分布函数.....	(67)
§ 4 期望与方差.....	(74)
§ 5 随机变量的线性变换.....	(79)
习题三.....	(83)
附录.....	(84)

第四章 数理统计的基本概念	(86)
§ 1 随机变量的独立性	(86)
§ 2 样品、样本、总体、统计量	(94)
§ 3 一个基本公式	(98)
§ 4 切比谢夫不等式及大数定律	(108)
习题四	(113)
附录	(114)
第五章 统计推断	(124)
§ 1 参数估计	(124)
§ 2 区间估计	(133)
§ 3 假设检验	(138)
§ 4 正态总体的几个检验问题	(145)
习题五	(150)
第六章 回归分析	(154)
§ 1 $1 \rightarrow 1$ 回归	(154)
§ 2 检验及预测	(160)
§ 3 非线性回归	(169)
§ 4 多 $\rightarrow 1$ 回归	(174)
习题六	(180)
附录	(182)
第七章 方差分析与试验设计	(189)
§ 1 单因素方差分析	(189)
§ 2 主效应、交互作用效应	(194)
§ 3 多因素方差分析	(198)

§ 4 正交设计	(204)
§ 5 表头设计	(213)
习题七.....	(221)

第八章 条件分布、贝叶斯公式 (225)

§ 1 贝叶斯公式	(225)
§ 2 条件分布(离散随机变量)	(230)
§ 3 条件分布(连续随机变量)	(233)
§ 4 贝叶斯估计	(238)
习题八.....	(245)

第九章 可靠性统计分析 (248)

§ 1 可靠性问题	(248)
*§ 2 次序统计量	(255)
§ 3 参数估计(1).....	(259)
§ 4 参数估计(2).....	(266)
习题九.....	(271)

第十章 中心极限定理及其应用 (273)

*§ 1 中心极限定理的证明	(273)
§ 2 中心极限定理的几种用法	(278)
§ 3 抽样检查方案	(283)
§ 4 大样本理论和方法的介绍	(288)

习题答案 (291)

附表 1 正态分布数值表	(300)
附表 2 t 分布临界值表	(301)
附表 3 χ^2 分布临界值表	(302)

附表 4	<i>F</i> 分布临界值表 ($\alpha=0.05$)	(303)
附表 5	<i>F</i> 分布临界值表 ($\alpha=0.025$)	(305)
附表 6	<i>F</i> 分布临界值表 ($\alpha=0.01$)	(307)
介绍进一步阅读的参考书目		(309)

第一章 数据的简单分析

本章叙述数据分析的一些直观和简单的方法，对均值、方差、中位数、频数等概念给以说明，为今后各章提供一个直观的背景。§1是一些准备知识，熟悉它的读者可以跳过不看。*§5是本章加深的内容，可以在学习后面章节时再看。

§1 准备知识

一、求和号 Σ

Σ 是一个符号，表示若干个数连加。 n 个数 a_1, a_2, \dots, a_n 相加，可用 $\sum_{i=1}^n a_i$ 表示，即

$$\sum_{i=1}^n a_i \triangleq a_1 + a_2 + \dots + a_n$$

(\triangle 表示按定义相等，或用左端符号表示右端)

式中 i 称为流动足标(均为整数)， a_i 称为通项。很明显

$$\sum_{i=1}^n a_i = a_1 + a_2 + \dots + a_n = \sum_{a=1}^n a_a$$

可见流动足标用什么符号表示对求和没有影响，这一点以后常常要用到。其次，只有通项 a_i 有了明确的定义时，求

和才有意义。例如 $a_i = x^i$, 则有 $\sum_{i=1}^n a_i = x + x^2 + \cdots + x^n$ 。

下面给出求和号今后常用的几个运算性质。

性质 1.1 $\sum_{i=1}^n c = nc$ (特别有 $\sum_{i=1}^n 1 = n$)。

证 按定义, 通项 $a_i = c$, 因此

$$\sum_{i=1}^n c = \underbrace{c + c + \cdots + c}_{n \text{ 个}} = nc$$

性质 1.2 $\sum_{i=1}^n (a_i + b_i) = \sum_{i=1}^n a_i + \sum_{i=1}^n b_i$

证 $\begin{aligned} \sum_{i=1}^n (a_i + b_i) &= (a_1 + b_1) + (a_2 + b_2) + \cdots + (a_n + b_n) \\ &= (a_1 + a_2 + \cdots + a_n) + (b_1 + b_2 + \cdots + b_n) \\ &= \sum_{i=1}^n a_i + \sum_{i=1}^n b_i \end{aligned}$

性质 1.3 $\sum_{i=1}^n ca_i = c \sum_{i=1}^n a_i$

证 $\begin{aligned} \sum_{i=1}^n ca_i &= ca_1 + ca_2 + \cdots + ca_n \\ &= c(a_1 + a_2 + \cdots + a_n) = c \sum_{i=1}^n a_i \end{aligned}$

例 1.1(等差数列求和)

设 $a_i = a + id$ ($i = 1, 2, \dots, n$), 求 $\sum_{i=0}^n a_i$ 的值。

设 $a_i = a + id$, $i = 0, 1, 2, \dots, n$,

解: $\sum_{i=0}^n i = 0 + 1 + 2 + \dots + n$

$$= \frac{1}{2}(1 + 2 + \dots + n + 1 + 2 + \dots + n)$$

$$= \frac{1}{2}[(1+n) + (2+n-1) + \dots + (n+1)]$$

$$= \frac{1}{2}n(n+1)$$

即

$$\sum_{i=0}^n i = \frac{1}{2}n(n+1) \quad (1.1)$$

$$\therefore \sum_{i=0}^n (a + id) \xrightarrow{\text{性质 1.2}} \sum_{i=0}^n a + \sum_{i=0}^n id$$

$$\xrightarrow{\text{性质 1.3}} \sum_{i=0}^n a + d \sum_{i=0}^n i$$

$$\xrightarrow{\text{性质 1.1}} (n+1)a + d \sum_{i=0}^n i$$

$$\xrightarrow{(1.1)} (n+1)a + d \cdot \frac{1}{2}n(n+1)$$

$$= (n+1) \cdot \frac{1}{2}[a + (a + nd)]$$

$$= \frac{1}{2} \cdot \text{项数} \cdot (\text{首项} + \text{末项})$$

例 1.2(等比数列求和)

设 $a_i = x^i$, $x \neq 1$, $i = 0, 1, 2, \dots, n$, 求 $\sum_{i=0}^n a_i$ 的值。

解：记 $s_n \triangleq \sum_{i=0}^n a_i$, $n = 0, 1, 2, \dots$, 于是

$$s_0 = a_0 = x^0 = 1$$

$$s_1 = a_0 + a_1 = x^0 + x = 1 + x = \frac{1 - x^2}{1 - x} \quad (x \neq 1)$$

$$s_2 = a_0 + a_1 + a_2 = x^0 + x + x^2 = 1 + x + x^2 = \frac{1 - x^3}{1 - x} \quad (x \neq 1)$$

容易看出

$$\begin{aligned} & (1 - x)(1 + x + x^2 + \dots + x^n) \\ &= 1 + x + x^2 + \dots + x^n - x - x^2 - \dots - x^n - x^{n+1} \\ &= 1 - x^{n+1} \end{aligned}$$

因此有

$$1 + x + \dots + x^n = \frac{1 - x^{n+1}}{1 - x} \quad (x \neq 1)$$

即 $\sum_{i=0}^n a_i = \frac{1 - x^{n+1}}{1 - x} \quad (x \neq 1)$

由此可知 $\sum_{i=0}^n ax^i = \frac{a(1 - x^{n+1})}{1 - x} \quad (x \neq 1)$

二、双重求和号

若有 k 组数据，用 x_{ij} 表示，足标 i 表示第 i 组，足标 j 表示在该组内第 j 个数据。若第 i 组内共有 n_i 个数据，则 $j = 1, 2, \dots, n_i$ 有意义。此时 $i = 1, 2, \dots, k$ 。将全部数据依行分组，写成如下的形式：

$x_{11}, x_{12}, \dots, x_{1n_1}$

$x_{21}, x_{22}, \dots, x_{2n_2}$

.....

$x_{k1}, x_{k2}, \dots, x_{kn_k}$

各组数据之和分别为 $\sum_{j=1}^{n_1} x_{1j}, \sum_{j=1}^{n_2} x_{2j}, \dots, \sum_{j=1}^{n_k} x_{kj}$ 。对这 k 个数

再求和，得全部数据之和为 $\sum_{i=1}^k \sum_{j=1}^{n_i} x_{ij}$ 。这就是双重足标求和，要注意的是 j 的流动范围是 1 到 n_i ，也即与 i 有关，因此

不能记作 $\sum_{j=1}^{n_1} \sum_{i=1}^k x_{ij}$ 。不难理解，如果足标变多时，求和号的先后顺序一般是不能随便交换的。如果 $n_i = l$ 与 i 无关，即

每组数据均有 l 个，则 $\sum_{i=1}^k \sum_{j=1}^l x_{ij}$ 与 $\sum_{j=1}^l \sum_{i=1}^k x_{ij}$ 的结果一样。

这一事实可以这样来理解，将 x_{11}, \dots, x_{kl} 这 kl 个数排成一个长方形，横的称行，竖的称列，就是

$$\begin{array}{ccccccc}
 x_{11} & x_{12} & \cdots & x_{1l} & \xrightarrow{\text{加}} & \sum_{j=1}^l x_{1j} \\
 x_{21} & x_{22} & \cdots & x_{2l} & \xrightarrow{\text{加}} & \sum_{j=1}^l x_{2j} \\
 \vdots & \vdots & \cdots & \vdots & & \vdots \\
 x_{k1} & x_{k2} & \cdots & x_{kl} & \xrightarrow{\text{加}} & \sum_{j=1}^l x_{kj} \\
 \text{加} \downarrow & \downarrow & & \downarrow & & \downarrow \text{相加}
 \end{array}$$

$$\sum_{i=1}^k x_{i1} \quad \sum_{i=1}^k x_{i2} \quad \cdots \quad \sum_{i=1}^k x_{il} \xrightarrow{\text{相加}} \sum_{i=1}^k \sum_{j=1}^l x_{ij}$$

先依行求和，将行和相加；或先依列求和，将列和相加，这两种方法都是将全部 x_{ij} 求和，结果自然应该一样，用公式表示就是

$$\sum_{i=1}^k \sum_{j=1}^l x_{ij} = \sum_{j=1}^l \sum_{i=1}^k x_{ij}$$

特别地，若 $a_{ij} = b_i c_j$ ，即第 (i, j) 个数据是 b_i 与 c_j 的乘积，此时有：

$$\text{性质 1.4} \quad (\sum_{i=1}^k b_i) \cdot (\sum_{j=1}^l c_j) = \sum_{i=1}^k \sum_{j=1}^l b_i \cdot c_j$$

$$\begin{aligned} \text{证} \quad \text{左边} &= (b_1 + b_2 + \cdots + b_k) \cdot (c_1 + c_2 + \cdots + c_l) \\ &= b_1(c_1 + \cdots + c_l) + b_2(c_1 + \cdots + c_l) + \cdots + \\ &\quad b_k(c_1 + \cdots + c_l) \\ &= \sum_{j=1}^l b_1 c_j + \sum_{j=1}^l b_2 c_j + \cdots + \sum_{j=1}^l b_k c_j \\ &= \sum_{i=1}^k \sum_{j=1}^l b_i c_j \end{aligned}$$

$$\text{特别有 } (\sum_{i=1}^k b_i)^2 = \sum_{i=1}^k \sum_{j=1}^k b_i \cdot b_j$$

三、几个应用的例

在数据处理时，经常要将公式变形，以便进行计算或显示其统计意义。下面举一些例子来说明求和号在这些变形中的作用。

以下用 \bar{x} 表示 n 个数 x_1, x_2, \dots, x_n 的平均数, 即

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \dots + x_n) = \frac{1}{n} \sum_{i=1}^n x_i$$

容易证明: 平均数有性质(见思考题)

$$\sum_{i=1}^n (x_i - \bar{x}) = 0 \quad (1.2)$$

例 1.3 求证: 对任给的常数 c, d , 有

$$\begin{aligned} & \sum_{i=1}^n (x_i - c)(y_i - d) \\ &= \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) + n(\bar{x} - c)(\bar{y} - d) \end{aligned}$$

其中 $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, $\bar{y} = \sum_{i=1}^n y_i$.

$$\begin{aligned} \text{证: 左端} &= \sum_{i=1}^n [(x_i - \bar{x}) + (\bar{x} - c)][(y_i - \bar{y}) + (\bar{y} - d)] \\ &= \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) + \sum_{i=1}^n (x_i - \bar{x})(\bar{y} - d) \\ &\quad + \sum_{i=1}^n (\bar{x} - c)(y_i - \bar{y}) + \sum_{i=1}^n (\bar{x} - c)(\bar{y} - d) \end{aligned}$$

性质 1.1 $\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) + \sum_{i=1}^n (x_i - \bar{x})(\bar{y} - d)$