SCHAUM'S ouTlines

全美经典学习指导系列

统计学原理

（下册）

——推断性统计学

〔美〕S. 伯恩斯坦　R. 伯恩斯坦 著

史道济 译

- 456道有完全解答的习题
- 221道附答案的练习题
- 涵盖本课程的所有基础，是任一教材的补充
- 获取高分的最佳助手

科学出版社

McGraw Hill 麦格劳-希尔教育出版集团

全美经典学习指导系列

统 计 学 原 理

（下 册）

——推断性统计学

〔美〕S. 伯恩斯坦
R. 伯恩斯坦 著

史道济 译

科 学 出 版 社

麦格劳-希尔教育出版集团

2002

内 容 简 介

本书论述工程技术、自然科学和生命科学中常用的统计学原理和方法.全书分为两册.下册主要是推断统计学的方法,包括抽样分布、估计理论、假设检验、回归分析和非参数方法等内容.

本书每一章都有相同的形式.第一部分以大纲的形式论述所有的新概念和新方法.第二部分是各种各样的习题解答,包括许多理论的应用.第三部分是补充的习题,只有答案.这部分内容是检验读者对本书内容的理解程度.

本书内容丰富,叙述严谨.全书有相互参照系统,方便读者阅读.本书适合于高等院校理工科学生、教师和有关的工程技术人员阅读.

Schaum's Outlines

Stephen Bernstein and Ruth Bernstein: Elements of Statistics II: Inferential Statistics

ISBN:0-07-005023-6

图字:01-2001-2118 号

图书在版编目(CIP)数据

统计学原理(下册)——推断性统计学/S.伯恩斯坦,R.伯恩斯坦著;史道济译.—北京:科学出版社,2002

(全美经典学习指导系列)

ISBN 7-03-009772-6

Ⅰ.统… Ⅱ.①伯…②伯…③史… Ⅲ.统计学 Ⅳ.C8

中国版本图书馆 CIP 数据核字(2001)第 063366 号

科学出版社 出版

北京东黄城根北街16号

邮政编码:100717

http://www.sciencep.com

西源印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2002年1月第 一 版 开本:A4 (890×1240)

2002年1月第一次印刷 印张:22 1/4

印数:1—5 000 字数:637 000

定价: 30.00 元

(如有印装质量问题,我社负责调换〈环伟〉)

目　　录

第十一章　离散型概率分布

11.1　离散型概率分布和概率质量函数

在上册第十章我们已经讲过离散型概率分布的一般性质.当时(见上册,10.3节)指出,离散型概率分布就是在离散随机变量定义的样本空间中指定事件发生概率的概率函数.当离散随机变量在定义域中取任意值($X=x$)时,概率函数对这个值就有指定的概率[$P(X=x)=f(x)$].离散型概率分布的概率函数称为概率质量函数,因为概率是散布在随机变量的各离散值上.

当所有的离散型概率分布都用概率质量函数的唯一确定的公式定义时,它们可以表示为以下4种形式:函数本身,由函数计算得到的概率分布列,概率表或概率图(见上册,表10.1和图10-1).虽然,对数学家而言,术语"离散型概率分布"和"概率质量函数"是同义词,但是,我们将区别离散型概率分布的定义函数和用该函数计算出的概率值的理论分布.

本章将讨论4个最重要的离散型概率分布:**二项分布,多项分布,超几何分布**和 **Poisson 分布**的具体特征和应用.先从二项分布开始.

11.2　Bernoulli 试验和多重 Bernoulli 试验

要理解二项分布,必须首先理解两个基本概念:**Bernoulli 试验**和**多重 Bernoulli 试验**;它们以瑞士数学家 James Bernoulli(也称 Jakob Bernoulli)(1645－1705)的名字命名,Bernoulli 首先研究了二者的性质.Bernoulli 对概率论的其它贡献是提出了 Bernoulli 定理(见上册,8.2节).Bernoulli 试验和多重 Bernoulli 试验有以下性质:

(1) 一次 Bernoulli 试验有且仅有两种可能结果,称为"成功"和"失败".两个结果是随机决定且互斥的.

(2) 多重 Bernoulli 试验是同一 Bernoulli 试验重复 n 次的序列.

(3) 在每次试验中,成功的概率是 p,失败的概率是 $q=1-p$.

(4) 各次 Bernoulli 试验之间**相互独立**(见上册,9.4节):每次试验结果不受其它各次试验结果的影响.

(5) 对每次 Bernoulli 试验,成功的概率 p 均相同(为**常数**)(这表明失败的概率 q 也是常数).

例 11.1　一只碗中放有20粒除颜色外别无差异的石子:10粒为红色,10粒为绿色.以下哪个是多重 Bernoulli 试验:(a) 闭上眼睛相继取出10粒石子,每次记录石子是否为红色,然后将石子放回,(b)重复(a)的试验,但每次取出的石子不放回?

解 ☞　(a) 是多重 Bernoulli 试验.这里,对每一次试验:成功＝红色,失败＝绿色,$p=10/20=\frac{1}{2}$,$q=1-\frac{1}{2}=\frac{1}{2}$,并且各次试验相互独立.

(b) 不是多重 Bernoulli 试验.因为取出的石子没有重新放回,p 在各次试验中不是保持不变的,且各次试验相互不独立(一次试验的结果对随后的试验结果有影响).

在例11.1(b)描述的试验中,假设碗中放有150粒红色石子和150粒绿色石子,而不是红色和绿色各10粒,再次进行试验,即闭上眼睛相继取出10粒石子,每次记录石子是否为红色,并且石子取出后不放回.试验似乎同例11.1(b)是一样的,但是,对统计学家而言,它们差异很

大,因为总体容量改变了[①].对于 300 粒石子,样本容量 n 对总体容量 N 的比例是 10/300 = 0.033,而对于 20 粒石子,比例是 10/20 = 0.50. 一般,统计学家认为,在抽样条件下,如果样本容量对总体容量的比例不大于 0.05(即,样本不大于总体的 5%,或 $n \leqslant 0.05N$),对不太保守的统计学家而言不超过 0.10($n \leqslant 0.10N$),就可以假设 p"基本上"保持不变,各次试验"基本上"相互独立;如果多重 Bernoulli 试验要求的所有其它性质均能满足,就可以认为抽样是多重 Bernoulli 试验.

我们在这里处理的问题以及之前和今后还将多次处理的问题是:纯粹的抽象统计模型和待解决的实际问题之间的冲突.统计方法建立在理论的,理想化环境的数学模型的基础之上,但这些模型在现实世界中极少存在.因此,一个统计模型的严格假设和要求往往不能完全满足.但是,在许多实际情形,就像这里,如果假设和要求可以"基本上"满足,则给定的统计方法能够给出合理的准确结果.

多重 Bernoulli 试验的假设仅在以下两种情形可完全满足:(1) **从有限或无限总体中有放回随机抽样**(见上册,3.16 节),(2) **从无限总体中无放回抽样**.在第一种情形,总体可以较小也可以相对于样本是较大的,由于每次将抽样个体放回,p 为常数且各次试验相互独立.在第二种情形,总体必须足够大,使得不放回抽样个体不会对实际问题造成影响;这样,p 仍为常数且各次试验相互独立.

11.3 二项随机变量,二项试验和二项概率函数

如果随机变量 X 用作记录 n 重 Bernoulli 试验中出现的成功次数($X = x$),则 X 称为**二项随机变量**,n 重 Bernoulli 试验序列称为一个**二项**(binomial)**试验**(bi, 两个;nomial, 项或结果).这样,如果一个概率函数被用作给出由二项随机变量定义的样本空间的每一样本点的概率值,则该函数称为**二项概率质量函数**或**二项概率函数**或**二项概率分布**或**二项分布**.

例 11.2 由上面和 11.2 节的定义说明:掷一枚硬币 3 次,出现正面的次数的离散型概率分布是一个二项分布.

解 ☞ 二项分布必须有以下基本要素:Bernoulli 试验,多重 Bernoulli 试验,二项随机变量,二项试验以及对二项变量定义的样本空间中每一样本点指定概率值的概率函数.

在这个问题中,Bernoulli 试验是掷一枚硬币,它只有两个可能结果:正面和反面,而且两个结果是随机决定且互斥的.可以将其中一个结果认定为成功结果,这种选择依赖于所研究的实际概率问题.这里,将观测到正面视为成功,目的是确定出现各种正面次数的概率.应该注意的是:将某一结果视为成功并不意味着该结果更能令人接受;它只表明,该结果是所要研究的.在任何情形,将正面视为成功从而反面视为失败之后,Bernoulli 试验就可以定义为:成功(正面)的概率是 $p = \frac{1}{2}$,同时,失败(反面)的概率是 $q = 1 - \frac{1}{2} = \frac{1}{2}$.

n 重 Bernoulli 试验序列是掷一枚硬币 3 次,每掷一次就是一次 Bernoulli 试验.在这个序列中,对每次试验均成立:(1) 只有同样的两个互斥且随机决定的可能结果:成功(正面)或失败(反面),(2) 对各次试验,成功的概率恒为常数 $p = \frac{1}{2}$,(3) 各次试验相互独立[假定它是理想化的游戏(见上册,8.1 节),硬币的任何结果都不影响随后的投掷结果].

离散随机变量 X 记录 $n = 3$ 重 Bernoulli 试验中出现成功的次数($X = x$),即出现正面的不同次数.X 是一个二项随机变量,这使得多重 Bernoulli 试验序列成为一个二项试验.最后,一个概率函数被用来(见上册,例 10.5)指定由二项随机变量定义的样本空间 $S = \{0, 1, 2, 3\}$ 的每一样本点的概率,从而,得到一个二项概率分布(见上册,表 10.2).

11.4 二项系数

上册第九章介绍了三个计数规则:乘法原理(见上册,9.12 节),排列(见上册,9.13 节)和组合(见上册,9.14 节).**二项系数**也是一个计数规则,它是二项概率函数(见下面 11.5 节)的

① 原文为样本容量改变了.——译者注

一个组成部分.如果 n 个个体只由不同的两类构成.一类有 x 个,另一类有 $n-x$ 个,则二项系数用于确定这 n 个个体构成的可能排列数.可以证明,这个排列数由 n 个不同个体中一次选取 x 个的组合数公式[上册,公式(9.20)]给出

$$\text{二项系数} = {}_nC_x = \binom{n}{x} = \frac{n!}{x!(n-x)!} \tag{11.1}$$

为理解上述含义,我们用 4 粒石子,红色(记为 R)和绿色(记为 G)各有 2 粒,试问:将这 $n=4$个个体**按次序**排成一列共有多少种排法? 由于我们考虑个体的次序,故这显然是排列问题.如果 4 个个体全不相同,则答案是(见上册,习题 9.30)

$${}_nP_n = n! = 4! = 24$$

如果 4 个个体全不相同,但不考虑次序,即仅仅关心 n 个个体中同时取出 4 个的可能组合数,则答案是(见上册,习题 9.41)

$${}_nC_n = 1$$

当 $n=4$ 且有$(x=2)$个为红色和$(n-x=2)$个为绿色时,两个答案都不正确;将 4 粒石子实际排成一列,可以看出,可能的排法是

$$RRGG \quad RGRG \quad RGGR \quad GGRR \quad GRGR \quad GRRG$$

这里一类有 $x=2$ 个,另一类有 $n-x=2$ 个,$(n=4)$个个体只有 6 种可能的排列.我们可用二项系数给出这个答案:

$$\binom{n}{x} = \frac{n!}{x!(n-x)!} = \frac{4!}{2!2!} = 6$$

例 11.3 一只碗中放有 20 粒除颜色外别无差异的石子,12 粒为红色,8 粒为绿色.闭上眼睛从碗中取出 6 粒石子,观测每粒石子的颜色,然后将其放回.试验中,有多少种取出 4 粒红色石子的方法?

解 在这个二项试验中,$n=6$ 且 $x=4$.由二项系数得

$$\binom{n}{x} = \frac{n!}{x!(n-x)!} = \frac{6!}{4!2!} = 15$$

11.5 二项概率函数

用上册例 10.5 中的概率函数计算二项分布-掷一枚硬币 3 次出现正面的次数-的概率值,是基于对 k 个独立事件的特殊乘法原理的推广[上册,公式(9.9)]和 k 个互斥事件的特殊加法原理的推广(上册,8.6 节,性质 4).现在,我们给出一个公式,称为**二项概率函数**,它将这些乘法和加法合成一个计算公式

$$f(x) = \binom{n}{x} p^x q^{n-x}, \ x = 0, 1, 2, \cdots, n \tag{11.2}$$

这说明,在一个二项试验的 n 重 Bernoulli 试验中,x 次成功的概率$[f(x)]$等于二项系数[公式(11.1)]和成功概率 p 自乘x 次(p^x)(见上册,1.9 节)以及失败概率 q 自乘$n-x$ 次q^{n-x}的乘积.

例 11.4 如例 11.3,一只碗中放有 12 粒红色石子和 8 粒绿色石子,仍取出 6 粒,不同的是一粒一粒地

取,每次观测颜色并将石子放回.试问:在 6 次选取中,取出 4 粒红色的概率是多少?

解 ☞ 从例 11.3 知,二项系数 $\binom{n}{x}$ 是 15;即取得一个容量为 6 的样本使之含有 4 粒红色石子的方法有 15 种.因此,为求得取出 4 粒红色的概率[$P(X=4)=f(4)$],只需将取出 4 粒红色石子和 2 粒绿色石子的概率乘以这个系数.对每次试验,取得一粒红色的概率(p)是 $\frac{12}{20}=0.6$,取得一粒绿色的概率(q)是 $1-p=0.4$. 4 次试验每次均取得一粒红色的概率是(0.6)(0.6)(0.6)(0.6),或 0.6^4;2 次试验每次均取得一粒绿色的概率是(0.4)(0.4),或 0.4^2.由公式(11.2),取得一个容量为 6 且含有 4 粒红色石子的样本的概率是

$$f(x)=\binom{n}{x}p^xq^{n-x}$$

$$f(4)=(15)(0.6)^4(0.4)^2=0.311$$

11.6 二项概率分布的均值,方差和标准差

任意离散型概率分布都有均值,方差和标准差.由上册公式(10.10),离散随机变量的概率分布的期望(或均值)的一般公式为

$$E(X)=\mu=\sum_x xf(x)$$

为得到**二项概率分布的均值**的公式,只须简单的以公式(11.2)替换 $f(x)$:

$$E(X)=\mu=\sum_x x\left[\binom{n}{x}p^xq^{n-x}\right] \tag{11.3}$$

在数学上可以证明,上式可简化为

$$E(X)=\mu=np \tag{11.4}$$

由上册公式(10.22),离散型概率分布的方差的一般计算公式为

$$\sigma^2=\sum_x x^2f(x)-\mu^2$$

用公式(11.2)替换 $f(x)$,公式(11.4)替换 μ,得到**二项概率分布的方差**的公式:

$$\sigma^2=\sum_x x^2\left[\binom{n}{x}p^xq^{n-x}\right]-(np)^2 \tag{11.5}$$

在数学上可以证明,上式可简化为

$$\sigma^2=npq \tag{11.6}$$

离散型概率分布的标准差是

$$\sigma=\sqrt{\sigma^2}$$

从而**二项概率分布的标准差**是

$$\sigma=\sqrt{npq} \tag{11.7}$$

例 11.5 求二项概率分布(掷一枚硬币 3 次,出现正面的次数)的均值,方差和标准差.

解 ☞ 对于这个试验,p = 掷一次硬币出现正面的概率 $=\frac{1}{2}$,q = 掷一次硬币出现反面的概率 $=\frac{1}{2}$,$n=3$ 将 n 和 p 的值代入公式(11.4),

$$E(X) = \mu = np = 3\left(\frac{1}{2}\right) = 1.5$$

将 n, p 和 q 的值代入(11.6),

$$\sigma^2 = npq = (3)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = 0.75$$

取方差的平方根可以计算标准差,

$$\sigma = \sqrt{npq} = \sqrt{0.75} = 0.87$$

11.7　二项式展开和二项式定理

二项概率函数和二项分布的名称来源于它们与**二项式展开**的关系.形如$(a+b)$的**二项代数表达式**(见上册,1.11 节)自乘 n 次后产生的**若干项和**,称为二项式展开.下面是二项式展开的 2 个例子:

$$(a+b)^2 = a^2 + 2ab + b^2 \quad (a+b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$$

二项式定理是如下展开$(a+b)^n$ 的一般公式,这里 a 和 b 是实数,n 和 x 是正整数

$$(a+b)^n = \sum_{x=0}^{n} \binom{n}{x} a^{n-x} b^x \tag{11.8}$$

它是(1)二项系数[公式(11.1)],(2)a 自乘$(n-x)$次(a^{n-x}),和(3)b 自乘 x 次(b^x)的乘积从 $x=0$ 到 $x=n$ 的和.例如,二项表达式$(a+b)^2$ 的二项式展开是

$$\begin{aligned}(a+b)^2 &= \sum_{x=0}^{2} \binom{2}{x} a^{2-x} b^x \\ &= \binom{2}{0} a^{2-0} b^0 + \binom{2}{1} a^{2-1} b^1 + \binom{2}{2} a^{2-2} b^2 \\ &= \frac{2!}{0!2!} a^2 + \frac{2!}{1!1!} ab + \frac{2!}{2!0!} b^2 \\ &= a^2 + 2ab + b^2\end{aligned}$$

如果设 $a=q$ 且 $b=p$,这里 q 和 p 分别是一次 Bernoulli 试验的失败和成功的概率,$n=$多重 Bernoulli 试验的次数,$x=$成功次数,则

$$\begin{aligned}(q+p)^n &= \sum_{x=0}^{n} \binom{n}{x} q^{n-x} p^x \\ &= \binom{n}{0} q^{n-0} p^0 + \binom{n}{1} q^{n-1} p^1 + \cdots + \binom{n}{n-1} q^{n-(n-1)} p^{n-1} + \binom{n}{n} q^{n-n} p^n \\ &= q^n + nq^{n-1}p + \cdots + nqp^{n-1} + p^n \end{aligned} \tag{11.9}$$

(注:这里颠倒 p 和 q 的次序,是为了与二项式定理中两个参数的次序保持一致,其中 a^{n-x} 先于 b^x). 从上面展开式可以看到,二项概率函数和二项式展开$(q+p)^n$ 之间有关系:对每一个整数 x, x 次成功的概率对应二项式展开的一项.

例 11.6　用二项式定理确定二项变量(掷一粒骰子 5 次得到 5 点的次数)的概率分布.

表 11.1

5 点的次数 x	概率 $f(x)=\binom{n}{x}p^xq^{n-x}$
0	$f(0)=\binom{5}{0}(1/6)^0(5/6)^{5-0}=0.40187$
1	$f(1)=\binom{5}{1}(1/6)^1(5/6)^{5-1}=0.40188$
2	$f(2)=\binom{5}{2}(1/6)^2(5/6)^{5-2}=0.16076$
3	$f(3)=\binom{5}{3}(1/6)^3(5/6)^{5-3}=0.03215$
4	$f(4)=\binom{5}{4}(1/6)^4(5/6)^{5-4}=0.00322$
5	$f(5)=\binom{5}{5}(1/6)^5(5/6)^{5-5}=0.00013$
$\sum$	1.00001

解 ☞　由公式(11.9),当 $n=5$ 时

$$
\begin{aligned}
(q+p)^5 &= \sum_{x=0}^{5}\binom{5}{x}q^{5-x}p^x \\
&= \binom{5}{0}q^{5-0}p^0+\binom{5}{1}q^{5-1}p^1+\binom{5}{2}q^{5-2}p^2 \\
&\quad +\binom{5}{3}q^{5-3}p^3+\binom{5}{4}q^{5-4}p^4+\binom{5}{5}q^{5-5}p^5 \\
&= \frac{5!}{0!5!}q^5+\frac{5!}{1!4!}q^4p+\frac{5!}{2!3!}q^3p^2+\frac{5!}{3!2!}q^2p^3+\frac{5!}{4!1!}qp^4+\frac{5!}{5!0!}p^5 \\
&= q^5+5q^4p+10q^3p^2+10q^2p^3+5qp^4+p^5
\end{aligned}
$$

如果将这个展开式与表 11.1 中变量的概率分布相比较,你会发现,任一 x 次成功的概率对应着展开式的一项.

由于一次 Bernoulli 试验只有成功和失败两个可能结果,并且它们之间互斥,

$$q+p=1$$

因此

$$(q+p)^n=\sum_{x=0}^{n}\binom{n}{x}q^{n-x}p^x=1$$

表 11.1 表明二项概率函数和二项式定理之间的这种关系:即对于离散型概率分布,有(见上册,图 10-3,性质 7)

$$\sum_x f(x)=\sum_x\binom{n}{x}p^xq^{n-x}=1$$

11.8　Pascal 三角形和二项系数

Pascal 三角形[①]是$(a+b)^n$ 的展开式中二项系数的一个排列. $n=5$ 的 Pascal 三角形如图 11-1 所示. 可以看出,每一行的第一个数字和最后一个数字都是 1,并且一行中间的任意数字

① 我国南宋时期数学家杨辉在他所著的《详解九章算术》中记载着有关二项系数的研究,Pascal 三角形与此类似,在我国也称杨辉三角形.——译者注

都能通过对该数字上方处于它左右两边的两个数字求和计算得到.例如,第 $n=5$ 行的两个 5 就是它们上方处于左右两边的数字 1 和 4 的和.在 $n=5$ 行,$(a+b)^5$[或$(q+p)^5$]的系数是:1,5,10,10,5,1.显然,这些与我们在例 11.6 得到的$(q+p)^5$ 的系数相同.

n											
0						1					
1					1		1				
2				1		2		1			
3			1		3		3		1		
4		1		4		6		4		1	
5	1		5		10		10		5		1

图 11-1

例 11.7　用 Pascal 三角形求$(q+p)^6$ 的展开式的二项系数.

解 ☞　用上面的计算方法,系数是: 1,6,15,20,15,6,1.

11.9　二项分布族

二项概率函数只有一种形式[公式(11.2)]

$$f(x)=\binom{n}{x}p^x q^{n-x}$$

但是它定义了无数多个具体的二项分布,对常数 n 和 p 的不同组合即为一个.指定 q 是没有必要的,因为 $q=1-p$.由于指定 n 和 p 即完全决定了所考虑的是哪个具体的二项分布,故 n 和 p 称为**二项分布的参数**.注意:在这里,术语“参数”的使用不同于在上册 3.4 节中介绍的描述性总体参数-由总体计算的描述性测量值.尽管理论方程的参数是确定常数,但是测量总体的参数仍然是该总体的一个描述性测量值.

如果一个理论方程被赋予不同参数时具有不同的函数形式,则该方程称为一个**方程族**.因此,我们可以说,二项概率函数定义了一个非常庞大的**二项概率分布族**.图 11-2 给出其中的 6 个,用图说明 n 和 p 的变化对分布的影响.其中,在图 11-2(a)中,p 从 0.25 到 0.50 到 0.75 变化(从上到下),n 恒为常数 5;在图 11-2(b)中,p 从 0.25 到 0.50 到 0.75 变化(从上到下),n 恒为常数 10.图 11-2(a)的分布值是用$(q+p)^5$ 的展开式(见例 11.6)计算得到,而图 11-2(b)的分布值由下面计算得到

$$(q+p)^{10}=q^{10}+10q^9p+45q^8p^2+120q^7p^3+210q^6p^4+252q^5p^5$$
$$+210q^4p^6+120q^3p^7+45q^2p^8+10qp^9+p^{10}$$

从图 11-2 可以看出二项分布的一些特性.第一,$p=0.50$ 时,分布是对称的;当 $p<0.50$(正偏,见上册,习题 5.6)和 $p>0.50$(负偏,见上册,习题 5.5)时,分布不对称.第二,如果 p 不变且 n 增大,则 μ 和 σ 都增大.第三,如果 n 不变且 p 增大,则 μ 也增大;而 σ 先增大,直到 $p=0.50$,然后对称的减小.

11.10　二项累积概率表

在解决实际问题中,经常需要知道一个二项变量(即 n 次试验成功的次数)小于或等于某个整数的概率,诸如掷一枚硬币 7 次至多出现 2 次正面的概率.计算这样的概率,要用到离散