数理统计

——基本概念及专题

李泽慧　王嘉澜　林　亨　等译

陈希孺　　　　　　　　　校订

兰州大学出版社

数 理 统 计

——基本概念及专题

P·J·比克尔　K·A·道克苏　著

李泽慧　王嘉澜　林　亨　等译

陈希孺　　　　　　　　　　校订

兰 州 大 学 出 版 社

数理统计——基本概念及专题

李泽慧 王嘉澜 林 亨 等译
陈希孺 校订

兰州大学出版社出版
（兰州大学校内）

甘肃省静宁印刷厂印刷 甘肃省新华书店发行
开本：850×1168毫米 1/32 印张：20.125

1991年8月第1版 1991年8月第1次印刷
字数：499千字 印数：1—2000册

ISBN7-311-00372-5/O·57 定价：16.5元

译者的话

本书是美国当代著名统计学家P·J·Bickel等所著的一本教材。在同水平的书中具有深度和广度的满意结合，实为一本优秀著作。主要内容有：概率论中的一些课题，统计模型，估计方法及最优化理论，置信区间与假设检验的方法及优化理论，线性模型，离散数据分析，非参数模型和决策理论等。

本书有如下特色：1. 取材面广而不偏，统计中有实用意义的分支大多有论述；2. 数学程度适中，适用于具有大学工科二年级数学水平的读者；统计概念论述严谨，能做到深入浅出；例题习题多，特别是习题的选择，不仅量多而且多结合课文的论述，很便于初学者和自学者。本书适合作为数理统计系本科生基础课教材或主要参考书，适当删节后也可作为综合性院校或师范院校数学系的教材，对数理统计研究生来说也是巩固知识的良好参考书。对主要兴趣在应用的读者来说，本书这种水平的素养也是必要的。

参加本书初稿翻译的有兰州大学数学系李泽慧、王嘉澜、林亨、彭英伟、陈怀明，还有杜延军、张书献、岳天祥，并由李泽慧、王嘉澜整理修改了全部译稿，林亨、彭英伟也参加了部分译稿的修改工作。中国科学院研究生院教授陈希孺对书稿作了校订。

译　者

1990年5月

序

《*Mathematical Statistics*》是美国当代著名统计学家*P. J. Bickel*等所著的一种教本．*Bickel*是数理统计学的奠基者之一*J Neyman*的学生，工作面广且有深度，是这个领域国际上公认的权威之一，这本著作有以下几个特色：1．取材面广但不偏，统计中有实用意义的分支大多有论述。2．数学程度适中，可适用于具有大学工科二年级数学水平的学生。3．对统计概念论述严谨，能做到深入浅出。4．例题、习题多特别是习题的选择，不仅量多，且多结合课文的论述，很便于初学者和自学者。本书适合于作为数理统计系、科本科生基础课教本或主要参考书，一般综合大学或师院数学系也可作为教本（材料适当删节）。对统计研究生来说也是一本巩固基础知识的良好参考书，对主要兴趣在于应用的读者来说，本书这种水平的理论素养也是必要的。

因此，我认为此书是一本优秀著作，且同类书籍在中文中尚阙如，现翻译成中文出版，一定会受到统计方面的学生和广大读者的欢迎。

陈希孺

88年11月15日

前 言

本书为具有良好数学基础的学生提供了一本数理统计的入门书。所谓良好的数学基础，是指线性代数、矩阵理论及高等微积分（不含测度论）。作为一本统计入门书，本书需要一些概率论的知识。我们希望读者具有P・Hoel，S・Port和C・Stone的《概率论入门》的水平。事实上，附录已经列出本书需要的概率论知识。然而，附录只是简要介绍，证明少，没有例子和习题。

我们认为一本统计入门书至少要包含下列内容：

(1) 描述表明理论与实际的关系的数理统计的基本概念。

(2) 对重要的"基本"结论给出详细的证明。如*Neyman-Pearson*引理，Lehmann—Scheffé定理、信息不等式和Gauss—Markoff定理。

(3) 对于比较高深的结果进行启发讨论。如最大似然估计的大样本理论，决策理论中贝叶斯解和容许解的结构。同时还应明确指出，在讨论中的漏洞可以进一步补充、修改的程度，以及可以找到补充、修改的文献。

(4) 指出统计思想和结果怎样应用于各种重要的子领城。如高斯线性模型、多项式模型及非参数模型。

虽然已有几本书在这几方面做得不错，但我们觉得没有一本书在这个水平上具有深度和广度的满意结合。*Rao*的《线性统计推断及其应用》（第二版）包含本书的大部分内容，而且比本书还多，但由于采用测度论处理而显得抽象。在这一水平上的书的另一极端是仅用大量的例子来说明，而没有进行必要的详细讨论。如Hogg和Craig的《数理统计入门》（第三版），它确实包含了本书处理的大部分论题，但缺乏必要的讨论。例如，我们认为统计方法的存在和计算及大样本问题的详细讨论是必需的。

本书的内容一学期讲不完。在我们教过的数学、统计、物理和

工程专业的研究生一学期课程中，我们讲授本书的核心章节第二章至第七章。这些章节包含估计和检验的模型化以及线性模型。另外，我们认为第十章的决策理论非常重要，还至少讲授前两节。最后，我们从第八章离散数据和第九章非参数模型中选一些讨论题目。

第一章是概率论的内容而不是统计的内容。遗憾的是这些内容的大部分在基础概率论教科书中没有，而本书的其余部分要用到它们。这一章有两种处理方法。一是讲授2—7章时穿插讲授这些内容，而不在开头单独讲授；二是在统计的先行课概率论的末尾讲授。

本书的特别之处是习题丰富。既有最平常的计算练习，也有用以强化学生概念的基本习题，还有超出本书内容的习题。这些习题一方面用于衡量对内容的掌握程度，另一方面是大量的思想和结果的体现，而这些思想和结果出于篇幅的原因，未被收集在正文中。

约定：

(I) 为了减少脚注的数量，我们在每章末尾习题节之前增加了注释节。注释按所属的节安排。每节可有多个注释，用阿拉伯数字标明。注释有题外话、补充和其它参考书。只有读者足够好奇时，才需读它们。

(II) 本书使用了许多符号约定和缩写。最常用的符号和缩写在书末给出，并指出了引入的地方。

(III) 概率术语的基本符号由附录给出。如随机变量、随机向量、密度、分布函数和矩的符号。

致谢部分（略）。

伯克莱　　Peter J·Bickel

1976　　Kjell A·Doksum

目　录

第一章　概率论中的一些课题

在这一章，我们要介绍概率论中的一些结论，这些结论在数理统计中是必要的，可是，在一些概率教科书中却没有十分详细的阐述。

在这本书里，我们不使用测度论。但总括地假设所有考虑到的集合和函数都是可测的。

1.1. 用随机变量或向量施行条件化

在研究随机变量或随机向量之间的联系时，条件这一概念是重要的。在这一节我们给出某些对预测理论、估计理论和回归理论有用的结果。

1.1.A. 离散情况

读者已经熟悉了事件B发生的条件下事件A发生的概率的概念。如果$\mathbf{X}$和$\mathbf{Y}$是不同维数的离散型随机向量，我们要研究$\mathbf{Y}$已经取得了特定值$\mathbf{y}$时，$\mathbf{X}$的条件概率的结构。

定义给定$\mathbf{Y}=\mathbf{y}$时，$\mathbf{X}$的**条件频率函数**$p(\cdot|\mathbf{y})$为

$$p(\mathbf{x}|\mathbf{y})=P[\mathbf{X}=\mathbf{x}|\mathbf{Y}=\mathbf{y}]=\frac{p(\mathbf{x},\mathbf{y})}{p_Y(\mathbf{y})} \tag{1.1.1}$$

这里p和p_Y是$(\mathbf{X},\mathbf{Y})$和$\mathbf{Y}$的频率函数。条件频率函数p仅仅对$p_Y(\mathbf{y})>0$的$\mathbf{y}$值有意义。由这个定义，$p(\cdot|\mathbf{y})$是一个频率分布函数。这是显然的，因为由$(A.8.11)$，

$$\Sigma_Y p(\mathbf{x}|\mathbf{y})=\frac{\Sigma_Y p(\mathbf{x},\mathbf{y})}{p_Y(\mathbf{y})}=\frac{p_Y(\mathbf{y})}{p_Y(\mathbf{y})}=1$$

$p(\cdot|\mathbf{y})$称为**给出**$\mathbf{Y}=\mathbf{y}$**时**，$\mathbf{X}$**的条件分布**。

例1.1.1. 设$\mathbf{X}=(X_1,\cdots,X_n)$，这里$X_i$表示具有成功概率$p$的事件在$n$重二项式试验中第$i$次试验结果的示性函数。设$Y=\sum_{i=1}^{n}X_i$表示成功的总次数，则$Y$服从二项分布$\mathscr{B}(n,p)$，

且有

$$p(\mathbf{x}\mid y)=\frac{P[\mathbf{X}=\mathbf{x},Y=y]}{\binom{n}{y}p^y(1-p)^{n-y}} \tag{1.1.2}$$

$$=\frac{p^y(1-p)^{n-y}}{\binom{n}{y}p^y(1-p)^{n-y}}=\frac{1}{\binom{n}{y}}$$

如果所有的x_i为0或1，且$\sum_{i=1}^{n}x_i=y$。

因此，如果说在n次二项式试验中观察到K次成功，那么这K次成功发生在n次试验中任何K次试验的可能性都相等。□

例1.1.2. 设X和Y有联合频率函数如下表，

表1.1.1

x \ y	0	10	20	$P_X(x)$
0	0.25	0.05	0.05	0.35
1	0.05	0.15	0.05	0.25
2	0.05	0.10	0.25	0.40
$P_Y(y)$	0.35	0.30	0.35	1

例如，假设Y是从某一个确定的总体中随机地抽查到的一个人每天吸烟的支数（精确到整10）。X是这个人通常的健康级别，$X=0$对应良好，$X=2$对应差，$X=1$对应既不是良好又不是差。我们发现，对于$y=20$有

x	0	1	2
$p(\mathbf{x}\|20)$	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{5}{7}$

该表指出了重度吸烟者与差健康状况之间的联系，因为$p(2|20)$几乎是$p_X(2)$的两倍。□

在两种特殊情况下，计算给定$\mathbf{Y}=\mathbf{y}$时$\mathbf{X}$的条件分布是容易的。

(I) 如果$\mathbf{X}$和$\mathbf{Y}$是独立的，即$p(\mathbf{x}|\mathbf{y})=p_X(\mathbf{x})$，这时条件分布重合于边际分布。

(II) 如果$\mathbf{X}$是$\mathbf{Y}$的一个函数$h(\mathbf{Y})$，则$\mathbf{X}$的条件分布是退化的，以概率1有$\mathbf{X}=h(\mathbf{y})$。

以上两个结论由定义(1.1.1)立即可得。

由定义(1.1.1)及(A.4.5)可得两个重要的公式. 以$q(\mathbf{y}|\mathbf{x})$记$\mathbf{X}=\mathbf{x}$时$\mathbf{Y}$的条件频率函数，则

(1.1.3) $$p(\mathbf{x},\mathbf{y})=p(\mathbf{x}|\mathbf{y})p_Y(\mathbf{y})$$

(1.1.4) $$p(\mathbf{x}|\mathbf{y})=\frac{q(\mathbf{y}|\mathbf{x})p_X(\mathbf{x})}{\sum_{z}q(\mathbf{y}|\mathbf{z})p_X(\mathbf{z})}$$ （巴叶斯公式）

只要上式右边的分母是正的。

方程(1.1.3)可以用来构造模型. 例如，假设在一个生产过程中，大量N件产品中的次品数Y服从二项式分布$\mathscr{B}(N,\theta)$. 又假设在N件产品中不还原地抽取n件作为样本，X是样本中的次品数. 我们知道，给定$Y=y$时X服从超几何分布$\mathscr{H}(y,N,n)$. 现在我们可以利用(1.1.3)写出X与Y的联合分布。

(1.1.5) $$P[X=x,Y=y]=\binom{N}{y}\theta^{y}(1-\theta)^{N-y}\frac{\binom{y}{x}\binom{N-y}{n-x}}{\binom{N}{n}}$$

这里的组合系数$\binom{a}{b}$中a,b是满足$b\leqslant a$的整数，否则$\binom{a}{b}$成为零。

我们还能使用这个模型说明(1.1.4)．由于通常仅仅能观察X，我们想知道给定$X=x$时Y的条件分布是什么．由(1.1.4)知这是，

$$P[Y=y\mid X=x]=\binom{N}{y}\theta^{y}(1-\theta)^{N-y}\binom{y}{x}\binom{N-y}{n-x}\Big/c(x).$$

这里$c(x)=\Sigma_y\binom{N}{y}\theta^{y}(1-\theta)^{N-y}\binom{y}{x}\binom{N-y}{n-x}$．这个公式简化为二项式概率（见习题1.1.11），

$$(1.1.6)\qquad P[Y=y\mid X=x]=\binom{N-n}{y-x}\theta^{y-x}(1-\theta)^{N-n-(y-x)}$$

1.1.B．离散型随机变量的条件数学期望

假设X是一个随机变量，且$E(|X|)<\infty$．定义给出$\mathbf{Y}=\mathbf{y}$时X的条件期望，记为$E(X\mid\mathbf{Y}=\mathbf{y})$，由下式给出

$$(1.1.7)\qquad E(X\mid\mathbf{Y}=\mathbf{y})=\sum_x xp(x\mid\mathbf{y})$$

注意由(1.1.1)，如果$p_Y(\mathbf{y})>0$，

$$(1.1.8)\qquad \sum_x|x|p(x\mid\mathbf{y})\leqslant\sum_x|x|\frac{p_X(x)}{p_Y(\mathbf{y})}=\frac{E(|X|)}{p_Y(\mathbf{y})}$$

例1.1.3．假设X和Y具有表1.1.1的联合频率函数．我们求得

$$E(X\mid Y=20)=0\times\frac{1}{7}+1\times\frac{2}{7}+2\times\frac{5}{7}=1.57$$

类似地，$E(X\mid Y=10)=\frac{7}{6}=1.17$，$E(X\mid Y=0)=\frac{3}{7}=0.43$．

注意在健康与抽烟的对应关系里，我们可以认为$E(X\mid Y=y)$是每天吸y支烟的人的平均健康级别。□

设$g(\mathbf{y})=E(X\mid\mathbf{Y}=\mathbf{y})$，随机变量$g(\mathbf{y})$就可记作$E(X\mid\mathbf{Y})$，且称为给出$\mathbf{Y}$时$\mathbf{X}$的条件期望*。

---

* 我们遵循称$E(X|\mathbf{Y})$为以概率1等于$g(\mathbf{Y})$的随机变量的惯例．