郑忠国 著

高等统计学

北京大学出版社

书　　　名：高等统计学
著作责任者：郑忠国　著
责 任 编 辑：王明舟
标 准 书 号：ISBN 7-301-03880-0/O·422
出　版　者：北京大学出版社
地　　　址：北京市海淀区中关村北京大学校内　100871
网　　　址：http://cbs.pku.edu.cn/cbs.htm
电　　　话：出版部 62752015　发行部 62754140　编辑部 62753160
电 子 信 箱：zpup@pup.pku.edu.cn
印　刷　者：北京大学印刷厂
发　行　者：北京大学出版社
经　销　者：新华书店
850×1168　32 开本　8.875 印张　200 千字
1998 年 9 月第一版　1998 年 9 月第一次印刷
定　　　价：15.00 元

前　言

自 80 年代以来，我国统计学进入蓬勃发展时期．在此形势之下，北京大学为概率统计专业硕士生开设了硕士学位课程，高等统计学就是学位课程中三门必修课之一．在这门课程的讲稿的基础上，逐渐形成今天这个稿子．本书的目的是

(i) 向学生介绍数理统计学中最基本的概念和知识．

(ii) 对于一些重要的统计结论给以严格的证明，例如，关于统计假设检验 Neyman-Pearson 引理以及某些统计方法的优良性的证明等．

(iii) 适当介绍一些近代统计学的进展．

由于这门课是硕士学位的基础课，对于那些以后将进入统计研究领域的学生来说，本课程无疑是一门入门课，对于那些毕业以后不再进行统计学研究，而转入其它研究领域或从事实际工作的学生，也可从中得益，懂得数据分析的要领．本教材力图达到这个课程的宗旨．

由于统计学的飞速发展和研究门类的广泛，本课程的内容不可能涉及所有的统计学分支，也不可能包含统计学各研究方向所需的所有的基础知识．例如由于学时的限制，像回归分析这样重要的内容，本课程就没有涉及．从笔者的观点看来，这样重要的方向，似应开设单独的课程进行讲授．

全书共分五章．第一章介绍充分统计量．这一部分似应讲得慢一些．由于学生程度的不一致，根据情况，任课教师还得补充一些统计和数学的知识．第二至第五章分别讲授检验、估计、大样本性质和 Beyes 分析等统计的基本知识．全书的材料略多于一个 18×4 学时的课程所需的份量．讲课教师可根据进度或同学接受程

度对内容作一些增删. 每章后面, 配有一定数量的习题, 协助同学加深理解课堂讲授的内容. 本书倘有疏漏之处, 诚望读者批评指正.

郑忠国

目　　录

第一章　充分统计量 ……………………………… (1)
§1　充分统计量的定义与判别法 ……………………… (1)
§2　极小充分统计量 ……………………………… (18)
§3　完全性 ……………………………………… (25)
§4　指数族分布中统计量的完全性，极小充分性 ………… (28)
§5　统计判决问题和充分统计量的优良性 ……………… (36)
练习题 ………………………………………… (42)
第二章　假设检验 ……………………………… (47)
§1　一般概念 …………………………………… (47)
§2　简单假设检验，N-P 引理 ……………………… (49)
§3　关于单调似然比族的检验问题 ………………… (55)
§4　最不利的分布 ………………………………… (63)
§5　一致最优无偏检验 …………………………… (69)
§6　带讨厌参数指数族分布的参数
　　的 UMPU 检验问题 ………………………… (82)
§7　不变检验 …………………………………… (96)
练习题 ………………………………………… (101)
第三章　估计 …………………………………… (105)
§1　引言 ………………………………………… (105)
§2　无偏估计 …………………………………… (106)
§3　信息不等式 ………………………………… (116)
§4　同变估计 (位置参数) ………………………… (124)
§5　同变估计 (一般情况) ………………………… (134)
§6　风险无偏性 ………………………………… (152)

练习题 (155)
第四章 估计的大样本性质 (158)
§1 相合性 (158)
§2 渐近正态性 (173)
§3 估计序列的大样本比较 (178)
§4 渐近有效性 (184)
§5 局部渐近正态性 (190)
§6 样本中位数 (202)
§7 L 估计 (205)
§8 M 估计和 R 估计 (216)
练习题 (223)
第五章 整体性质 (229)
§1 Bayes 估计 (229)
§2 极小极大估计 (236)
§3 容许性 (245)
§4 关于正则族中参数估计的容许性 (251)
§5 正态均值的估计问题 (265)
练习题 (274)

第一章　充分统计量

§1 充分统计量的定义与判别法

在开始学习统计学之前，首先必须回答这样的问题，什么是统计学？按目前统计学界流行的说法是：统计学是收集、处理和分析数据的科学．由于数据来自自然和社会的各个方面，“应用”是统计学的一个十分重要的特征，在统计学中，离开应用的理论是没有价值的．

人们在研究客观世界时，往往用某些量去刻画事物．这些刻画事物的量往往是非常重要的．它们能提供有关事物的重要信息．这些量称为参数．由于种种原因，这些量不能直接观察到．获得这些量的有关信息，正是统计学的任务．

现在举一些例子来说明之．人们可用心脏年龄这个量来刻画心脏衰老程度．这个量与人的实际年龄不完全一样．如果一个人的心脏年龄已经到了老年阶段，即使这个人的年龄不算老，他的心脏也会出现老年人的特征．他犯心脏病的可能性较大．在某种意义下，心脏年龄比人的自然年龄更重要．但是心脏年龄不像人的自然年龄有出生记录作为根据．心脏年龄是不可观察的．因此，心脏年龄作为刻画人的心脏健康程度的一个参数，是人体的一个很重要的指标．利用统计学方法获得心脏年龄的有关信息就是统计学的一个任务．又例如，有一批产品，其批量为 N．在 N 件产品中有 b 件是次品．通常 b 是不大可能观察到的，除非对整批产品进行逐个检查．它反映了这一批产品的质量的优劣．因此，b 是这一批产品的很重要的一个参数．

为了了解参数，我们需要寻找与参数有关的量．通常我们借助可观察的量 X 来提取有关的信息，其中 X 是一个数量、向量或其它的量．在一般情况下，我们不可能由 X 直接解出参数的值．X 通常是一个随机变量，其分布与我们所关心的参数有关．有时候， X 的分布也与我们不关心的参数 (我们称这些参数为讨厌参数) 有关．总之一句话， X 的分布与参数有关．通常我们用 θ 表示参数，用 Θ 表示参数空间 (θ 的可能范围)．这样 X 的分布可用 P_θ 表示，其中 $\theta\in\Theta$．记 $\mathcal{P}=\{P_\theta:\theta\in\Theta\}$ 为 X 的可能分布之集合．我们称之为 X 的分布族．我们的目的是通过观察值 X 来了解有关 θ 的信息．下面我们将用一个例子来说明之．

例 1.1 设某一批产品，其批量为 $N=100$, 其中 b 个为废品，b 的可能值为 $0,1,\cdots,100$. 但 b 是未知的．现在从中随机地选取 $n=3$ 个产品进行检验．设经过检验后得知废品的个数为 $x=3$. 这时 $(n,x)=(3,3)$ 就是数据．它表示 3 件产品全部不合格．如果没有统计学的帮助，从 (3,3) 这一组数据中得到的关于 b 的信息量是很少的．我们只可能说，这一批产品质量不好，废品个数较多．现在我们将 $x=3$ 看成一次试验的结果，即从 (N,b) 中抽取的一组样本 (n,x)．在这一个问题中， N,n 均为已知， b 为未知 (参数), x 是试验结果．由于抽样的随机性， x 可以看成一个代表随机试验的随机变量 X 的试验结果．随机变量的最主要特点是：(i) 试验之前，X 的值是不确定的， (ii) 试验结果出来以后，X 的值就完全确定， (iii) X 的取值遵从确定的统计规律 (分布)．本例中随机样品中废品个数刚好是某随机变量 X 的取值，现求 X 的分布．为简单起见，假定 $b\geq n, N-b\geq n$, 此时，废品个数可取 0 个到 n 个．现在讨论事件 $X=x$ 的概率．为了讨论方便，我们将 N 个产品编上号，写成 $1,2,\cdots,N$. 记 $1,2,\cdots,b$ 为废品， $b+1,\cdots,N$ 为正

品. 这样, 每一个可能的试验结果为集合 $\{1,2,\cdots,N\}$ 的一个子集合 $\{i_1,i_2,\cdots,i_n\}$. 不同的子集合的个数是 $\binom{N}{n}$(组合数). 显然, 这是一个等概模型, 即每一个可能的结果 $\{i_1,\cdots,i_n\}$ 的概率为 $1/\binom{N}{n}$. 所谓对 $X=x$ 有利的试验结果是指 $\{i_1,\cdots,i_n\}$ 中恰有 $x\leq b$ (废品). 这样的试验结果共有 $\binom{b}{x}\binom{N-b}{n-x}$ 个. 按等概模型之定义

$$P_b\{X=x\}=\frac{\binom{b}{x}\binom{N-b}{n-x}}{\binom{N}{n}},\quad 0\leq x\leq n.$$

对于 $n>b$ 和 $n>N-b$ 的情况, 上式也成立, 只不过 x 的取值范围受到一定限制. 这样我们得到由下式给出的 X 的分布

$$P_b\{X=x\}=\frac{\binom{b}{x}\binom{N-b}{n-x}}{\binom{N}{n}},\quad \max\{n-(N-b),0\}\leq x\leq\min\{b,n\}.$$

在这里, 废品个数 b 是作为参数形式出现的. 在此我们强调指出, 概率论的推理之基础是认为废品个数是已知的, 由此可推导出 X 之分布. 但从统计学的角度看来, b 是未知的常数, 我们的任务是确定 b 的值或获得有关 b 的信息. 进一步研究发现, b 与废品个数 X 的分布律之间有一一对应关系. 即对于不同的 b 的值, X 的分布律也不同, 反之亦然. 因此, 统计问题归结为由 X 的观察值获得有关 X 的分布律的知识. 当 X 的分布律的信息获得以后, 关于废品个数 b 的信息也相应获得, 统计学的任务就完成了.

一般情况下, 设 X 为取值于某可测空间 $(\mathcal{X},\mathcal{B}_{\mathcal{X}})$ 的随机变量, 其中 $\mathcal{X}$ 为 X 可能取值之集合, $\mathcal{B}_{\mathcal{X}}$ 为 $\mathcal{X}$ 中某些集合形成的 σ 域. $(\mathcal{X},\mathcal{B}_{\mathcal{X}})$ 上的由 X 产生的概率测度 P 称为 X 的分布. 在统计问题中, X 的分布是未知的. 通常假定 X 的分布属于某分布族 $\mathcal{P}=\{P_\theta:\theta\in\Theta\}$, 其中 θ 称为参数, Θ 称为参数空间. 我们称 $(\mathcal{X},\mathcal{B}_{\mathcal{X}},\mathcal{P})$ 为统计模型, X 称为观察值随机变量. 必要时, 我们写成 $X\sim P_{\theta_0}$, 表示 X 的真分布为 P_{θ_0}, θ_0 为参数真值. 由于真

值往往不为人们所知，通常我们写成 $X \sim P_\theta$, 其中 θ 为 Θ 中的一员. 有时候，也写成 $X \sim P \in \mathcal{P}$.

在概率论中，通常需要建立概率空间 $(\mathcal{X}, \mathcal{B}_{\mathcal{X}}, P)$. 但在统计学中研究的对象却是统计模型 $(\mathcal{X}, \mathcal{B}_{\mathcal{X}}, \mathcal{P})$, 这里 $\mathcal{P}$ 是一个概率分布族. 关于统计模型，我们还需要引进统计学的一些术语.

定义 1.1 设 $(\mathcal{X}, \mathcal{B}_{\mathcal{X}}, \mathcal{P})$ 为一统计模型，可测集 A 称为 $\mathcal{P}$ 零测集是指 $P(A) = 0, \forall P \in \mathcal{P}$, 记作 $\mathcal{P}(A) = 0$. 通常说 "a.e.$\mathcal{P}$" 是指"除了一个 $\mathcal{P}$ 零测集之外". 例如 $f(x) = 0$ a.e. $\mathcal{P}$ 是指例外集合 $\{x, f(x) \neq 0\}$ 是一个 $\mathcal{P}$ 零测集.

定义 1.2 设 $\mathcal{P}, \mathcal{Q}$ 为 $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$ 上的两个 σ 有限测度族.

(i) 若对任何 $A \in \mathcal{B}_{\mathcal{X}}$, $\mathcal{Q}(A) = 0 \Rightarrow \mathcal{P}(A) = 0$, 则称 $\mathcal{P}$ 相对于 $\mathcal{Q}$ 绝对连续. 记作 $\mathcal{P} \ll \mathcal{Q}$.

(ii) 若 $\mathcal{P} \ll \mathcal{Q}$, $\mathcal{Q} \ll \mathcal{P}$, 则称 $\mathcal{P}$ 与 $\mathcal{Q}$ 等价，记作 $\mathcal{P} \equiv \mathcal{Q}$.

(iii) 若存在 $\mathcal{Q} = \{\mu\}$ 为单点集，即只含一个测度的测度族，其中 μ 是一个 $\mathcal{X}$ 上的 σ 有限测度，使得 $\mathcal{P} \ll \mathcal{Q}$, 则称 $\mathcal{P}$ 为被控的. 今后，我们直接用记号 $\mathcal{P} \ll \mu$ 表示 $\mathcal{P}$ 被 μ 所控制.

在统计学中，最常见的分布族是被控的分布族. 下面的引理给出一个分布族 $\mathcal{P}$ 为被控的充要条件.

引理 1.1 统计模型 $(\mathcal{X}, \mathcal{B}_{\mathcal{X}}, \mathcal{P})$ 中 $\mathcal{P}$ 被控的充要条件是存在一个可数子族 $\mathcal{P}_0 \subset \mathcal{P}$, 使得 $\mathcal{P}_0 \equiv \mathcal{P}$.

证明 充分性. 设 $\mathcal{P}_0 = \{P_i, i \geq 1\}$ 是满足充分条件的那个可数子族. 令

$$P_0 = \sum_{i=1}^{\infty} \frac{1}{2^i} P_i.$$

由上述表达式可知，对任何随机事件 $A \in \mathcal{B}_{\mathcal{X}}$, $P_0(A) = 0$ 成立的充要条件为 $P_i(A) = 0, i = 1, 2, \cdots$. 由测度族等价之定义知 $P_0 \equiv \mathcal{P}_0$. 由充分性条件知 $\mathcal{P}_0 \equiv \mathcal{P}$. 再由关系 "$\equiv$" 的传递性知 $P_0 \equiv \mathcal{P}$. 故

分布族 $\mathcal{P}$ 是被控的.

必要性. 设 $\mathcal{P}$ 为被控的, 即存在一个 σ 有限测度 μ, 使得 $\mathcal{P} \ll \mu$. 无妨假定 μ 是一个有限测度. 由 $\mathcal{P} \ll \mu$ 可知, 对任何 $P \in \mathcal{P}$, $P \ll \mu$. 再由关于测度绝对连续的定义可知, 存在可测函数 $f(x)$, 它是概率测度 P 相对于 μ 的 Radon-Nikodym 导数. 即对任何 $A \in \mathcal{B}_{\mathcal{X}}$, $f(x)$ 满足

$$\int_A f(x)d\mu = P(A).$$

以后, 我们通常用 $dP/d\mu$ 表示这个 Radon-Nikodym 导数. 记

$$K_P = \left\{x : \frac{dP}{d\mu} > 0\right\},$$

其中 $dP/d\mu$ 是概率测度 P 相对于测度 μ 的 Radon-Nikodym 导数. 由于 μ 是一个有限测度, 我们有

$$\sup_{P\in\mathcal{P}} \mu(K_P) < \infty.$$

为证明必要性, 我们必需构造出一个与 $\mathcal{P}$ 等价的可数子族 $\mathcal{P}_0$. 取 $P_1, P_2, P_3, \cdots, P_n, \cdots$, 使

$$\mu(K_{P_1}) \geq \sup_{P\in\mathcal{P}} \mu(K_P) - \frac{1}{2},$$

$$\mu(K_{P_2} \backslash K_{P_1}) \geq \sup_{P\in\mathcal{P}} \mu(K_P \backslash K_{P_1}) - \frac{1}{2^2},$$

$$\mu\left(K_{P_3} \Big\backslash \bigcup_{i=1}^{2} K_{P_i}\right) \geq \sup_{P\in\mathcal{P}} \mu\left(K_P \Big\backslash \bigcup_{i=1}^{2} K_{P_i}\right) - \frac{1}{2^3},$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\mu\left(K_{P_n} \Big\backslash \bigcup_{i=1}^{n-1} K_{P_i}\right) \geq \sup_{P\in\mathcal{P}} \mu\left(K_P \backslash \bigcup_{i=1}^{n-1} K_{P_i}\right) - \frac{1}{2^n},$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

如此可得到可数分布子族 $\mathcal{P}_0=\{P_i: i=1,2,\cdots\}$. 由于 μ 为有限测度，并且 $K_{P_n}\Big\backslash\bigcup_{i=1}^{n-1}K_{P_i}, n=1,2,\cdots$ 是样本空间 $\mathcal{X}$ 中互不相交的集合，可知

$$\mu\left(K_{P_n}\Big\backslash\bigcup_{i=1}^{n-1}K_{P_i}\right)\to 0,\quad n\to\infty.$$

再利用不等式

$$\mu\left(K_P\Big\backslash\bigcup_{i=1}^{\infty}K_{P_i}\right)\le\mu\left(K_P\Big\backslash\bigcup_{i=1}^{n-1}K_{P_i}\right)\le\mu\left(K_{P_n}\Big\backslash\bigcup_{i=1}^{n-1}K_{P_i}\right)+\frac{1}{2^n}$$

可知

$$\mu\left(K_P\Big\backslash\bigcup_{i=1}^{\infty}K_{P_i}\right)=0,\quad \forall P\in\mathcal{P}. \tag{1.1}$$

此外，由集合 K_P 之定义知

$$P(B)=0\Longleftrightarrow\mu\{K_P\cap B\}=0. \tag{1.2}$$

为证明 $\mathcal{P}\equiv\mathcal{P}_0$, 我们只需证明 $\mathcal{P}\ll\mathcal{P}_0$. 设 B 为满足条件 $\mathcal{P}_0(B)=0$, 即

$$P_i(B)=0,\quad i=1,2,\cdots$$

的可测集. 由 (1.2) 知

$$\mu(B\cap K_{P_i})=0,\quad i=1,2,\cdots\quad 或\quad \mu\left(B\cap\left(\bigcup_{i=1}^{\infty}K_{P_i}\right)\right)=0. \tag{1.3}$$

利用 (1.1) 和 (1.3) 可知

$$\begin{aligned}\mu(B\cap K_P)&\le\mu\left(B\cap\left(K_P\Big\backslash\bigcup_{i=1}^{\infty}K_{P_i}\right)\right)+\mu\left(B\cap\left(\bigcup_{i=1}^{\infty}K_{P_i}\right)\right)\\&\le\mu\left(K_P\Big\backslash\bigcup_{i=1}^{\infty}K_{P_i}\right)+\mu\left(B\cap\left(\bigcup_{i=1}^{\infty}K_{P_i}\right)\right)=0.\end{aligned}$$

再利用 (1.2) 知 $P(B)=0$. 在证明中，概率测度 P 可以是 $\mathcal{P}$ 中任意概率测度，这就证明了 $\mathcal{P} \ll \mathcal{P}_0$. ∎

在统计学中还有一个非常重要的概念，即统计量的概念. 通常用 $T(X)$ 表示之. 统计量实际上是观察值的函数. 在此必需指出的是，统计量 $T(X)$ 必须与参数无直接关系. 当观察值 X 给出以后，我们能够从函数关系 $T(X)$ 计算出具体的数值. 因此统计量不能与未知的参数直接有关. 例如设 $X=(x_1,\cdots,x_n)$ 是一组数据，$T(X)=\overline{X}=\sum X_i/n$ 是统计量. 归根结底，统计量 $T(X)$ 的值可通过 X 的值，经过计算得到. 广义上说，统计量也是试验的结果，它可用具体的数据表达出来. 但是，$\overline{X}-EX_1$ 这个量，看起来很像统计量，但它不是统计量. 其原因是，当给定 X 的值以后，无法计算出 $\overline{X}-EX_1$ 的值. 统计量的一个作用是压缩数据. 这一点是非常明确和直观的，我们不在此赘述. 一个好的统计量还必须具有保留信息的作用. 现在让我们讨论统计量与参数的关系. 前面已经提到过，作为函数，$T(X)$ 与参数是没有关系的. 但是，$T(X)$ 的分布是有可能与参数有关的. 设 X 的分布的参数是 θ. 若 $T(X)$ 的分布与参数 θ 有关，这时侯，T 多少包含一些关于参数 θ 的信息. 若 $T(X)$ 的分布与参数 θ 无关，则 $T(X)$ 不包含参数 θ 的任何信息. 一般说来，统计量越简单，它对数据的压缩就越厉害，但它所保留的信息就越少. 我们的目的是找到适当的统计量 T, 使它尽可能地保留 θ 的信息和最大限度地压缩数据. 设 $S(X)$ 和 $T(X)$ 为两个统计量. 若 $S(X)$ 在 $T(X)=t$ 之条件分布与参数无关，其中 t 是任意数，则说明 $S(X)$ 在 T 之值已知的条件下不再包含参数 θ 的信息. 若对任何统计量 $S(X)$, $S(X)$ 在 T 之值已知的条件下不再包含参数 θ 的信息，此时，T 包含了全部有关参数 θ 的信息，并称 T 为充分统计量. 实际上，由条件概

率之理论知，关于充分统计量的定义可以写成：若对于任何事件 $A \in \mathcal{B}_{\mathcal{X}}$, 在 $T(X)=t$ 之下事件 A 的条件概率与参数无关，其中 t 是任意数，则称 T 是充分统计量.

为了叙述充分统计量的严格定义，首先叙述条件概率之严格定义. 现举一个例子.

例 1.2 设 $\mathcal{X}=(0,1]$, $\mathcal{B}_{\mathcal{X}}$ 为 (0,1] 上 Borel 域， $P(A)=A$ 的 Lebesgue 测度. 又设事件 $B_1=(0,1/2]$. 由古典条件概率之定义，事件 A 在 B_1 之下的条件概率为

$$P(A|B_1)=\frac{P(A\cap B_1)}{P(B_1)}.$$

这个古典条件概率之定义有缺陷. 当 $P(B_1)$ 为 0 时， $P(A|B_1)$ 就没有定义. 现在我们从另一角度来解释上述条件概率之定义. 将条件概率的公式写成如下的形式

$$P(A|B)=\begin{cases}\dfrac{P(A\cap B_1)}{P(B_1)}, & \text{当} B=B_1,\\[2ex] \dfrac{P(A\cap B_1^c)}{P(B_1^c)}, & \text{当} B=B_1^c.\end{cases}$$

其中 B_1^c 表示事件 B_1 的余事件， $B_1^c=\mathcal{X}\backslash B_1$. 这个公式综合了条件 B_1 和 B_1^c. 上面的公式又可以改写成下面的随机变量的形式，

$$f_A(x)=\begin{cases}\dfrac{P(A\cap B_1)}{P(B_1)}, & \text{当 } x\in B_1,\\[2ex] \dfrac{P(A\cap B_1^c)}{P(B_1^c)}, & \text{当 } x\in B_1^c.\end{cases}\tag{1.4}$$

当 $x\in B_1$ 时，事件 B_1 发生， $f_A(x)$ 表示事件 A 相对于条件 B_1 的条件概率. 当 $x\in B_1^c$ 时， $f_A(x)$ 表示事件 A 相对于条件 B_1^c 的条件概率.

现在令 $\mathcal{B}=\{\varnothing, B_1, B_1^c, \mathcal{X}\}$. 在 $\mathcal{B}$ 上定义两个测度, $\phi_A(B)=P(A\cap B)$ 和 $P^{\mathcal{B}}(B)=P(B), B\in\mathcal{B}$. 显然, $\phi_A(B)\ll P^{\mathcal{B}}$. 利用 Radon-Nikodym 定理, 可求得 Radon-Nikodym 导数 $d\phi_A/dP^{\mathcal{B}}$. 事实上, $d\phi_A/dP^{\mathcal{B}}$ 刚好等于 (1.4) 中的 $f_A(x)$. 这样, $d\phi_A/dP^{\mathcal{B}}$ 就可作为条件概率的另一定义. 由于这个定义综合了两个条件之条件概率, 我们用 $P(A|\mathcal{B})$ 代替原来的记号 $P(A|B)$. $P(A|\mathcal{B})$ 具有下面的性质:

(i) $P(A|\mathcal{B})$ 是 $\mathcal{B}$ 可测函数;

(ii)
$$\int_B P(A|\mathcal{B})dP=P(A\cap B),\quad \forall B\in\mathcal{B}. \tag{1.5}$$

当 $B=B_1$ 或 B_1^c 时, 上式变成

$$P(A\cap B)=P(B)P(A|B),$$

它就是古典的条件概率公式.

由例 1.2 可以看出, 例中给出的两种条件概率之定义是相互等价的 (一种是古典定义, 一种依赖于测度论的定义). 但是后一种定义具有推广价值. 下面是关于条件概率的一般定义.

定义 1.3 设 $(\mathcal{X},\mathcal{B}_{\mathcal{X}},P)$ 是某概率空间, $\mathcal{B}$ 是 $\mathcal{B}_{\mathcal{X}}$ 的子 σ 域. 对于事件 $A\in\mathcal{B}_{\mathcal{X}}$, 定义集合函数 $\phi_A(B)=P(A\cap B)$, $B\in\mathcal{B}$. $\phi_A(B)$ 是定义在 $\mathcal{B}$ 上的 σ 可加有限测度, 并且在 $\mathcal{B}$ 上相对于 $P^{\mathcal{B}}$ (概率测度 P 在 $\mathcal{B}$ 上的限制) 是绝对连续的. 因此, 按照 Randon-Nikodym 定理, 存在一个 $\mathcal{B}$ 可测函数 $f_A(x)=d\phi_A/dP^{\mathcal{B}}$, 使

$$\int_B f_A(x)dP=\phi_A(B)=P(A\cap B),\quad B\in\mathcal{B}. \tag{1.6}$$

可测函数 $f_A(x)$ 称为事件 A 相对于子 σ 域 $\mathcal{B}$ 的条件概率.

现在设 T 为可测空间 $(\mathcal{X},\mathcal{B}_{\mathcal{X}})$ 到可测空间 $(\mathcal{T},\mathcal{B}_{\mathcal{T}})$ 上的一个可测变换. 记 $\mathcal{B}_T$ 为由 T 产生的 $\mathcal{B}_{\mathcal{X}}$ 的子 σ 域. 用测度论的记

号，$\mathcal{B}_T=\sigma\{T\in B\}$，即 $\mathcal{B}_T$ 是形如 $\{T\in B\}$ 集合产生的 σ 域. A 相对于 $\mathcal{B}_T$ 的条件概率 $f_A(x)$ 就定义为事件 A 相对于统计量 T 的条件概率. 现在设 $f_A(x)$ 是事件 A 相对于统计量 T 的条件概率. 按定义，$f_A(x)$ 是一个 $\mathcal{B}_T$ 可测函数. 利用复合函数定理可知 $f_A(x)$ 可写成 $T(x)$ 函数的形式. 为符号简便，我们将这个函数写成 $f_A(T(x))$. 这样在叙述关于条件 T 的条件概率时，我们可以这样说：对于 $A\in\mathcal{B}_{\mathcal{X}}$，存在一个 $\mathcal{B}_T$ 可测函数 $f_A(T(x))$，满足

$$\int_{T\in B} f_A(T(x))d\,P=P(A\cap\{T\in B\}),\quad \forall B\in\mathcal{B}_T.$$

这个 $f_A(T(x))$ 就定义为事件 A 在条件 T 之下的条件概率，记作 $P(A|T(x))$. 显然，条件概率是 a.e.P 确定的. 若将上式右端的积分作积分变换，把它转化成 T 的象空间中的积分，我们可得公式

$$\int_B f_A(t)d\,P^T=P(A\cap\{T\in B\}),\quad \forall B\in\mathcal{B}_T.$$

此处，P^T 是概率测度 P 通过变换 T 在 T 的象空间中形成的测度. 通常 P^T 称为 T 之分布. $f_A(t)$ 为定义在 T 之象空间上的可测函数. $f_A(t)$ 也称为事件 A 的条件概率，记作

$$f_A(t)=P(A|T(X)=t).$$

例 1.3 设 (X,Y) 取值于 $(0,1]^2$, (X,Y) 的密度为

$$f(x,y)=\begin{cases} x+y, & x\in(0,1],\ y\in(0,1], \\ 0, & \text{其它}. \end{cases}$$

现需计算事件 $X\le c$ 在 $X+Y=t$ 之下条件概率

$$P\{X\le c|X+Y=t\}.$$

由于固定的 t ($t\in(0,2]$), $f(x,y)$ 在 $x+y=t$ 上取常数值，此时我们可以认为在 $X+Y=t$ 时，点 (X,Y) 在线段

$$l_t = \{(x,y) : x+y=t, x\in(0,1], y\in(0,1]\}$$

上均匀地分布. 利用这个想法和几何直观，经计算得

$$P\{X\le c|X+Y=t\} = \begin{cases} 1, & \min\{1,t\}\le c, \\ \dfrac{c}{t}, & c<t\le 1, \\ \dfrac{c-(t-1)}{1-(t-1)}, & 1<t\le c+1\le 2, \\ 0, & \text{其它}. \end{cases} \tag{1.7}$$

为严格证明上式，我们必须验证下列公式

$$\int_{u\le t} P\{X\le c|X+Y=u\} f_{X+Y}(u)du = P\{X\le c, X+Y\le t\}. \tag{1.8}$$

其中 $f_{X+Y}(u)$ 为 $X+Y$ 的分布密度. 经计算， $f_{X+Y}(u)$ 由下式给出

$$f_{X+Y}(u) = \begin{cases} u^2, & 0<u\le 1, \\ 1-(u-1)^2, & 1<u\le 2, \\ 0, & \text{其它}. \end{cases} \tag{1.9}$$

通过计算，可以证明 (1.8) 式两边是相等的. 它们的值等于

$$\begin{cases} \dfrac{1}{3}t^3, & 0<t\le c\le 1 \text{ 或 } 0<t\le 1,\ c>1, \\ 1/3+(t-1)-\dfrac{1}{3}(t-1)^3, & c>1,\ 1<t\le 2, \\ 1, & c>1,\ t\ge 2, \\ \dfrac{1}{2}ct^2-\dfrac{1}{6}c^3, & 0<c<t\le 1, \\ \dfrac{1}{2}(c+c^2)-(c-t+1)^2\left(\dfrac{1}{6}+\dfrac{1}{6}c+\dfrac{1}{3}t\right), & 1<t\le 1+c,\ 0<c<1, \\ \dfrac{1}{2}(c+c^2), & c+1<t,\ 0<c<1, \\ 0, & \text{其它}. \end{cases}$$