□ 现代统计学丛书

高等统计学概论

赵林城 王占锋 编著

高等教育出版社

□ 现代统计学丛书

高等
统计学概论

赵林城 王占锋 编著

GAODENG TONGJIXUE GAILUN

高等教育出版社·北京

图书在版编目（C I P）数据

高等统计学概论 / 赵林城，王占锋编著 . -- 北京：高等教育出版社，2016. 3

ISBN 978-7-04-044895-5

Ⅰ. ①高… Ⅱ. ①赵… ②王… Ⅲ. ①统计学－教材 Ⅳ. ① C8

中国版本图书馆 CIP 数据核字（2016）第 030449 号

策划编辑 王丽萍　　责任编辑 李华英　　封面设计 王凌波　　版式设计 于 婕
责任校对 张小镝　　责任印制 尤 静

出版发行	高等教育出版社	咨询电话	400-810-0598
社　　址	北京市西城区德外大街4号	网　　址	http://www.hep.edu.cn
邮政编码	100120		http://www.hep.com.cn
印　　刷	北京佳信达欣艺术印刷有限公司	网上订购	http://www.landraco.com
开　　本	787mm× 1092mm　1/16		http://www.landraco.com.cn
印　　张	18	版　　次	2016年3月第1版
字　　数	370 千字	印　　次	2016年3月第1次印刷
购书热线	010-58581118	定　　价	69.00 元

物 料 号　44895-00

应用统计学丛书

前言

编写一本教科书, 断断续续, 前后经历了近八年的时间, 这在我还是第一次。退休前做科研, 完成一篇论文, 最短用几天时间, 最长的也不过一年。和陈希孺老师合作写过两本书, 也没费多长时间, 因为需要我写的是我们研究工作的一部分。教科书不一样, 总得写点与别人不一样的东西, 还得把难讲的东西写得简明易懂, 要做到这一点也颇费心思; 过去自己以为清楚了的东西, 等到下笔, 才发现仍有需要推敲的地方; 为了弄清一些问题的提法或者有关研究的来龙去脉, 还需翻阅若干参考文献。尤其不容易的是自己居然仍有这样的兴趣和耐心。

本书是为数理统计和相关专业的研究生或从事相关工作的人士撰写的。俗话说众口难调, 不敢奢望本书能适合所有人的需求, 但在作者心中, 还是希望本书能对朋友们在求知和求学方面有所裨益。本书着力于讨论数理统计的基本概念、理论、思想和方法, 在材料处理上有详有略, 有所侧重。随着统计科学和计量经济研究的发展, 为了处理一些复杂的统计模型, 越来越多的统计文献使用了过程收敛和经验过程的工具和方法。为了适应这种发展趋势, 使研究生尽早涉足统计研究的前沿, 本书在统计渐近理论方面加大篇幅讨论, 并着重介绍了经验过程理论在渐近统计中的应用。关于极值分布的渐近理论, 我们也尽可能引用了文献中最新的结果。

本书内容大致安排如下。作为统计理论体系的基础, 本书首先介绍了条件期望、条件概率和条件分布的理论, 以及统计决策理论的框架和有关概念。除了经典的点估计、假设检验、区间估计和 Bayes 统计外, 我们在第五章用较大篇幅介绍了统计渐近理论: 首先介绍了极限理论和渐近统计的基础知识, 以及 Cramér 关于似然方程解的经典理论; 由于 M-估计在统计应用中的重要地位, 以及经验过程理论在统计渐近理论中的主导作用和对未来的预期, 我们用了较大篇幅介绍如何借助于经验过程的理论和方法, 得出 M-估计的渐近分布, 进而建立 Le Cam 关于极大似然估计渐近分布的一种精妙方法; 而有关的过程收敛和经验过程的预备知识可以在附录中找到。之所以要做这样的努力, 不在于介绍多少具体的知识, 而是希望通过展现不同的视

角和处理方法, 使读者能在无形中提高一种内在的能力。

本书选取了一定数量的例题, 有些是作为理论或定理的应用, 有些则是理论的延伸。一部分例题采用了独特的解法, 希望能起一些示范作用。为了提高学生的解题能力, 每章最后都附有适量的习题, 除了供读者自我训练之外, 有些则是正文的补充。较难的习题前加了 * 号, 一些习题则加上了提示。为了提高自身解决问题的能力, 希望有志于这一学科的读者除了做好基本的习题之外, 也要能解一部分难题, 通过努力, 使自己的解题能力再上一个台阶。作者切身的体会是, 在上学时自己解出一些难题, 不仅会提高自己的信心和兴趣, 而且也是一种让人久久难忘的经历。无论将来从事科学研究、教学还是实际应用工作, 不断打好基础也许会成为一生的功课。

在本书写作过程中, 部分内容已经给研究生讲了两次, 每次 80 学时, 实践看来这些学时仍然不够, 讲授时可以在内容上有所取舍。初学时有些较难的内容可以略去, 较难的证明也可以不讲。关于 M-估计渐近理论的讲述, 可以限制在经验过程有关方法和结果的应用层面, 而不涉及经验过程的理论研究本身。有关的专门理论可以鼓励学生自己钻研, 或者另外组织讨论班仔细研读。作者过去就曾这样做过。

本书在正文或习题中多处引用了陈希孺 (1981) (1999), Lehmann (1983) (1986), 茆诗松, 王静龙, 濮晓龙 (1998), van der Vaart (1998), Reiss (1989), Ferguson (1996) 以及 Shao (2005) 等著作中的有关材料, 有些已在引用处注明, 但遗漏之处在所难免, 除在此做统一说明之外, 还望读者和有关方面谅解。

第四章区间估计是本书另一作者王占锋撰写的。其中信仰推断部分引进了杨振海等人采用的垂直密度表示的内容, 并在内容和篇幅上做了微型化处理, 以便让读者用尽量少的时间了解这一方法的概貌。本书其余各章和附录部分为本人撰写。为使风格一致, 统筹工作也由本人完成。

本书在写作过程中, 得到了缪柏其、韦来生、吴耀华、杨亚宁、陈敏诸位教授以及张伟平、胡治水副教授等人的鼓励和实际帮助。陈子沐同学帮助作者解决了绘图以及录入中遇到的一些疑难问题。在此向他们一并致谢。

由于作者水平所限, 本书错误和疏漏之处在所难免, 还望读者不吝指教。

赵林城

2015 年 5 月

目录

第一章

基本概念

1.1 条件期望和条件概率

在数理统计的理论体系中, 条件期望和条件概率的理论占有重要的地位. 本节将以适合于统计需要的形式, 介绍这一理论的一些基本事实. 这些内容都是读者应当掌握的.

为了理论发展的需要, 下面先介绍一些重要的预备定理.

1.1.1 与可测变换有关的两个定理

在以下的陈述中, 空间 $\mathcal{X}$ 是具有某种性质的元素 (称为点) 构成的一个非空集合, $\mathcal{B}$ 为 $\mathcal{X}$ 的某些子集构成的一个 σ-域, 此时称 $(\mathcal{X},\mathcal{B})$ 为一**可测空间**, $\mathcal{B}$ 中的集合称为可测集. 最为常见的可测空间是 $(\mathbf{R}^k,\mathcal{B}_k)$, 其中 $\mathbf{R}^k$ 是一 k 维欧氏空间, $\mathcal{B}_k$ 是由 $\mathbf{R}^k$ 中的开集系生成的 σ-域, 即包含所有开集的最小 σ-域, 通常把这个 σ-域称为 Borel-域, 其中的集合则称为 Borel 集.

考虑两个非空集合 $\mathcal{X}$ 和 $\mathcal{T}$. 集合 $\mathcal{X}$ 到集合 $\mathcal{T}$ 内的一个变换 (或函数) T 可记为 $T:\mathcal{X}\mapsto\mathcal{T}$. 定义于 $\mathcal{X}$ 上取值于 $\mathbf{R}^1=(-\infty,\infty)$① 的函数称为 $\mathcal{X}$ 上的实值函数. 对于 $\mathcal{T}$ 的任一子集 G 和它的任一子集类 $\mathcal{G}$, 分别称 $T^{-1}(G):=\{x\in\mathcal{X}:T(x)\in G\}:=\{T\in G\}$ 和 $T^{-1}(\mathcal{G}):=\{T^{-1}(G):G\in\mathcal{G}\}$ 为集合 G 和集类 $\mathcal{G}$ 的原像.

由于 T^{-1} 保持集合的运算和包含关系, 所以一个域的原像为一个域, σ-域的原像为一个 σ-域.

设 $T:\mathcal{X}\mapsto\mathcal{T}$ 为可测空间 $(\mathcal{X},\mathcal{B})$ 到可测空间 $(\mathcal{T},\mathcal{G})$ 内的一个变换, 如果对任

① 实数集 $\mathbf{R}^1$ 也可简记为 $\mathbf{R}$, 而广义直线 $\bar{\mathbf{R}}^1$ 即可简记为 $\bar{\mathbf{R}}$.

一 $G \in \mathcal{G}$, $T^{-1}(G) \in \mathcal{B}$, 则称 T 是 $\mathcal{B}|\mathcal{G}$ 可测的, 这一条件等价于 $T^{-1}(\mathcal{G}) \subset \mathcal{B}$. 此时 $T^{-1}(\mathcal{G})$ 为 $\mathcal{B}$ 的一个子 σ-域. $(\mathcal{X}, \mathcal{B})$ 到空间 $(\mathbf{R}^k, \mathcal{B}_k)$ 内的可测函数, 常简称为 $\mathcal{B}$-可测函数. 在没有歧义时涉及的 σ-域则无需交代.

设 μ 为 $\mathcal{B}$ 上的一个测度, T 为 $(\mathcal{X}, \mathcal{B})$ 到 $(\mathcal{T}, \mathcal{G})$ 内的一个可测变换. 对于每个 $G \in \mathcal{G}$, 定义 $\mu^*(G) = \mu(T^{-1}(G))$. 不难证明 μ^* 为 $\mathcal{G}$ 上的一个测度, 称为 μ 在 T 之下的导出测度. 注意 μ 和 μ^* 是定义在不同的 σ-域上的测度. 在测度论和概率论中, 称 $(\mathcal{X}, \mathcal{B}, \mu)$ 为一**测度空间**, 而当 μ 为一概率测度时, 则称其为**概率空间**.

设 B 为 $\mathcal{X}$ 的任一子集, 其余集 $\mathcal{X} \setminus B$ 记为 B^c. 集合 B 的示性函数 I_B 定义为

$$I_B(x) = \begin{cases} 1, & \text{当 } x \in B, \\ 0, & \text{当 } x \in B^c. \end{cases}$$

集合 B 的示性函数有时也记为 $I(B)$.

随机元和样本空间是数理统计中的基本概念. 统计问题研究的是带有随机性的数据, 这些数据可以用一个随机元 X 来表示. 在数理统计中, 空间 $\mathcal{X}$ 上的一个**随机元** X 即为有关的概率空间 $(\Omega, \mathcal{A}, P)$ 到可测空间 $(\mathcal{X}, \mathcal{B})$ 内的一个可测映照. 作为随机元 X 取值的空间, $(\mathcal{X}, \mathcal{B})$ 称为**样本空间**. 一个重要的情形是上述定义中的空间 $\mathcal{X}$ 为一距离空间, $\mathcal{B}$ 为它的 Borel σ-域, 即由 $\mathcal{X}$ 中的开集系生成的 σ-域. 在许多常见的统计问题中, X 为一有限维随机向量, 在本书中, $\mathcal{X}$ 常常取为 $\mathbf{R}^k$ 或它的一个适当的 Borel 子集, 连同相应的 Borel-域 $\mathcal{B}_k$, 称这样的样本空间是**欧氏的**.

以后我们将会看到, 一个统计问题的表述, 一般不再提及有关的概率空间 $(\Omega, \mathcal{A}, P)$, 样本空间及由随机元导出的概率分布族即为统计研究的出发点.

需要指出的是, 仅有欧氏样本空间是远远不够的, 函数型数据和过程统计的研究, 就要涉及可距离化样本空间.

下面引出两个重要引理.

引理 1.1 *设 T 为可测空间 $(\mathcal{X}, \mathcal{B})$ 到可测空间 $(\mathcal{T}, \mathcal{G})$ 的可测变换, $\mathcal{B}_0 := T^{-1}(\mathcal{G})$ 为由 T 导出的 $\mathcal{B}$ 的子 σ-域, 则一个实值 $\mathcal{B}$-可测函数 f 为 $\mathcal{B}_0$-可测的充要条件为: 存在一个 $\mathcal{G}$-可测函数 g, 使 $f(x) = g(T(x))$ 对所有 $x \in \mathcal{X}$ 成立.*

证明 充分性显然. 事实上, 假定这样的函数 g 存在, 则对任一 Borel 集 C,

$$f^{-1}(C) = T^{-1}\big(g^{-1}(C)\big) \in T^{-1}(\mathcal{G}) = \mathcal{B}_0,$$

故 f 为 $\mathcal{B}_0$-可测.

为证必要性, 设 f 为 $\mathcal{B}_0$-可测, 则对固定的 n, 集合

$$B_{in} = \{x : i/2^n < f(x) \leqslant (i+1)/2^n\}, \quad i = 0, \pm 1, \pm 2, \cdots$$

为互斥的 $\mathcal{B}_0$ 集, 其并集为 $\mathcal{X}$, 因而存在 $G_{in} \in \mathcal{G}$, 使 $B_{in} = T^{-1}(G_{in})$. 不难证明, 存在 G_{in} 的子集 $G_{in}^* \in \mathcal{G}$, 使得对固定的 n, $G_{in}^*, i = 0, \pm 1, \pm 2, \cdots$ 两两不交, 且满足 $B_{in} = T^{-1}(G_{in}^*)$. 实际上, 只需取 $G_{in}^* = G_{in} \setminus \bigcup_{j \neq i} G_{jn}$ (请读者补出证明的细节).

定义

$$f_n(x) = i/2^n, \quad \text{如果 } x \in B_{in}, \quad i = 0, \pm 1, \pm 2, \cdots,$$

$$g_n(t)=\begin{cases} i/2^n, & \text{如果 } t\in G_{in}^*, \quad i=0,\pm1,\pm2,\cdots, \\ 0, & \text{否则,}\end{cases}$$

那么 $f_n(x)=g_n(T(x)), x\in\mathcal{X}$. 因为函数 g_n 为 $\mathcal{G}$-可测, 故使得 $g_n(t)$ 收敛于有限极限的点 t 所成的集合 $V\in\mathcal{G}$. 令 $S=T(\mathcal{X})$ 为 T 的值域, 则对 $t\in S$, 存在 $x\in\mathcal{X}$, 使 $t=T(x)$, 故

$$\lim g_n(t)=\lim g_n[T(x)]=\lim f_n(x)=f(x)$$

对所有 $x\in\mathcal{X}$ 成立, 因而 $S\subset V$. 定义

$$g(t)=\begin{cases} \lim g_n(t), & \text{当 } t\in V, \\ 0, & \text{否则,}\end{cases}$$

则 g 就具有所要的性质. ■

参看 Doob (1953) 和 Lehmann (1986). 严加安 (1981) 和严士健, 刘秀芳 (1994) 提供了另一种证明方法. 上述定理称为 **Doob 复合函数可测性定理**, 它也可等价地表述如下.

设 T 为 $\mathcal{X}$ 到 $(\mathcal{T},\mathcal{G})$ 内的一个变换, $\mathcal{B}:=T^{-1}(\mathcal{G})$ 为 $\mathcal{X}$ 上由 T 导出的 σ-域, 则 $\mathcal{X}$ 上的一个实值函数 f 为 $\mathcal{B}$-可测的充要条件为: 存在一个 $\mathcal{G}$-可测函数 g, 使 $f(x)=g(T(x))$ 对所有 $x\in\mathcal{X}$ 成立.

上述函数 f 和 g 的积分之间的关系由下面的引理给出.

引理 1.2 (积分变换定理) *设 T 为 $(\mathcal{X},\mathcal{B})$ 到 $(\mathcal{T},\mathcal{G})$ 的可测变换, μ 为 $\mathcal{B}$ 上的一个测度, μ^* 为其导出测度, 则对任何 $\mathcal{G}$-可测的实值函数 $g(t)$ 和任一 $G\in\mathcal{G}$, 有*

$$\int_{T^{-1}(G)} g[T(x)]d\mu(x)=\int_G g(t)d\mu^*(t). \tag{1.1}$$

上式成立是指只要一边的积分存在, 另一边的积分也存在, 并且两者相等.

证明 不失一般性, 不妨假定 G 就是整个空间 $\mathcal{T}$. 若 $g=I_{G^*}$ 为集合 G^* 的示性函数, $G^*\in\mathcal{G}$, 则 (1.1) 的左右两边分别简化为 $\mu(T^{-1}(G^*))$ 和 $\mu^*(G^*)$, 由 μ^* 的定义, 引理显然成立. 因而 (1.1) 对简单函数, 对非负可测函数, 继而对积分存在的所有函数都成立. ■

1.1.2 其他有关的预备定理

首先介绍著名的 Radon-Nikodym 定理. 为此, 设 μ 和 ν 为可测空间 $(\mathcal{X},\mathcal{B})$ 上的两个测度. 若对任何 $B\in\mathcal{B}$,

$$\mu(B)=0\Rightarrow\nu(B)=0,$$

则称 ν 关于 μ 绝对连续, 或称测度 ν 受控于测度 μ, 记为 $\nu\ll\mu$. 如果 $\nu\ll\mu$, 且 $\mu\ll\nu$, 则称这两个测度等价. 这些定义容易推广到 μ 为 $(\mathcal{X},\mathcal{B})$ 上的有号测度的情形.

称 μ 为可测空间 $(\mathcal{X},\mathcal{B})$ 上的一个 **σ-有限测度**, 如果存在一列 μ-测度有限的可

测集 A_n, 使 $\mathcal{X}=\bigcup_{n=1}^{\infty}A_n$. 定义

$$\nu(B)=\sum_{n=1}^{\infty}\frac{\mu(A_nB)}{1+2^n\mu(A_n)},\quad B\in\mathcal{B},$$

则 ν 与 μ 等价, 且 ν 为有限测度.

定理 1.3 (Radon-Nikodym 定理) 设 μ 和 ν 为可测空间 $(\mathcal{X},\mathcal{B})$ 上的两个 σ-有限测度, 则 ν 关于 μ 绝对连续的充要条件为: 存在一个几乎处处有限的非负的 $\mathcal{B}$-可测函数 f, 使

$$\nu(B)=\int_B f(x)d\mu(x),\qquad 对任何\ \ B\in\mathcal{B}. \tag{1.2}$$

如果不计其在一个 μ-零测集上的取值, 则 f 由关系式 (1.2) 唯一确定.

更一般地, 上述定理也适用于 ν 为 $(\mathcal{X},\mathcal{B})$ 上的 σ-有限的有号测度的情形.

此时称 f 为 ν 关于 μ 的 Radon-Nikodym 导数, 记为

$$d\nu=fd\mu,\quad 或\quad f=d\nu/d\mu.$$

在 ν 为概率测度的情形, f 称为 ν 关于 μ 的概率密度. 有时为了指出上述 f 确定到不计它在一个 μ-零测集上的取值, 常记为

$$f=d\nu/d\mu,\quad \text{a.e. }\mu\quad 或\quad f=d\nu/d\mu,\quad \text{a.e. }(\mathcal{B},\mu).$$

此处我们约定, 如果一个命题 $Q(x)$ 对除去一个 μ-零测集的所有 x 成立, 则称 $Q(x)$ a.e. μ 成立; 如要进一步强调 μ 的定义域 $\mathcal{B}$, 则写成 $Q(x)$, a.e. $(\mathcal{B},\mu)$. 设 $\mathcal{P}$ 为一概率测度族, 若对于每个 $P\in\mathcal{P}$ 都有 $Q(x)$, a.e. P, 则称 $Q(x)$ a.e. $\mathcal{P}$ 成立.

在此我们顺便指出, 在证明上述定理中 f 的唯一性时用到一个有用的事实: 设 f 和 g 是两个 $\mathcal{B}$-可测且 μ-可积的函数, 使 $\int_B f(x)d\mu(x)=\int_B g(x)d\mu(x)$ 对任何 $B\in\mathcal{B}$ 成立, 则 $f=g$, a.e. μ. 事实上, 令 $B=\{f>g\}$, 则 $B\in\mathcal{B}$, 由假定,

$$\int_{\{f>g\}}[f(x)-g(x)]d\mu(x)=0,$$

因而 $\mu(f>g)=0$. 同样也有 $\mu(f<g)=0$.

需要介绍的下一个定理是关于重积分的 Fubini 定理. 设 B 和 C 为两个集合, 集合 $B\times C:=\{(x,y):x\in B,y\in C\}$ 称为 B 和 C 的直积 (或 Cartesian 积). 设 $(\mathcal{X},\mathcal{B})$ 和 $(\mathcal{Y},\mathcal{C})$ 为两个可测空间, $\mathcal{B}\times\mathcal{C}$ 是包含形如 $B\times C(B\in\mathcal{B},C\in\mathcal{C})$ 的所有集合的最小 σ-域. 设 μ 和 ν 分别为 $(\mathcal{X},\mathcal{B})$ 和 $(\mathcal{Y},\mathcal{C})$ 上的 σ-有限测度, 则存在 $(\mathcal{X}\times\mathcal{Y},\mathcal{B}\times\mathcal{C})$ 上唯一的测度 $\lambda=\mu\times\nu$, 称为 μ 和 ν 的乘积测度, 使得对所有的 $B\in\mathcal{B},C\in\mathcal{C}$,

$$\lambda(B\times C)=\mu(B)\nu(C).$$

在可测性方面, 以下事实值得一提: 设 $f(\cdot,\cdot)$ 为 $\mathcal{B}\times\mathcal{C}$-可测函数, 则对 $\forall x\in\mathcal{X}$, $f(x,\cdot)$ 为一 $\mathcal{C}$-可测函数, 而对 $\forall y\in\mathcal{Y}$, $f(\cdot,y)$ 为一 $\mathcal{B}$-可测函数.

定理 1.4 (Fubini 定理) 设 μ 和 ν 分别为 $(\mathcal{X},\mathcal{B})$ 和 $(\mathcal{Y},\mathcal{C})$ 上的 σ-有限测度, $\lambda=\mu\times\nu$. 设 $f(x,y)$ 关于 λ 可积, 则

(i) 按测度 ν 对几乎所有的 y, $f(\cdot,y)$ 关于 μ 可积;

(ii) 函数 $\int_{\mathcal{X}} f(x,y)d\mu(x)$ 关于 ν 可积, 且

$$\int_{\mathcal{X}\times\mathcal{Y}} f(x,y)d\lambda(x,y) = \int_{\mathcal{Y}}\left[\int_{\mathcal{X}} f(x,y)d\mu(x)\right]d\nu(y).$$

如果积分区域是全空间, 今后常常将其略去不写.

1.1.3 条件期望的定义和性质

设 $(\mathcal{X},\mathcal{B},P)$ 为一概率空间, T 为 $(\mathcal{X},\mathcal{B})$ 到 $(\mathcal{T},\mathcal{G})$ 的可测变换, $\mathcal{B}_0 := T^{-1}(\mathcal{G})$ 为由 T 导出的 $\mathcal{B}$ 的子 σ-域. 在统计中, P 为随机元 X (也可视为一个样本) 服从的概率分布, T 为一个统计量. 设 f 是一个 $(\mathcal{B},P)$-可积 (即 $\mathcal{B}$-可测且 P-可积) 的实值函数, 则 $\nu(B) := \int_B f dP$ 对所有 $B\in\mathcal{B}$ 有定义, 因而也对所有 $B\in\mathcal{B}_0$ 有定义. 局限于 σ-域 $\mathcal{B}_0$ 上, $\nu \ll P$, 由 Radon-Nikodym 定理, 存在 $(\mathcal{B}_0,P)$-可积函数 f_0, 使

$$\int_{B_0} f dP = \int_{B_0} f_0 dP, \quad 对所有\ \ B_0\in\mathcal{B}_0, \tag{1.3}$$

此处, f_0 由上式 a. e. $(\mathcal{B}_0,P)$ 唯一确定. 定义 T 的概率分布 P^* 如下: $P^*(G) = P(T^{-1}(G)), G\in\mathcal{G}$. 由引理 1.1, 存在一个 $\mathcal{G}$-可测函数 g, 使 $f_0(x) = g(T(x))$ 对所有 $x\in\mathcal{X}$ 成立. 由引理 1.2, 函数 g 可直接借助于 f 由

$$\int_{T^{-1}(G)} f(x)dP(x) = \int_G g(t)dP^*(t), \quad 对所有\ G\in\mathcal{G} \tag{1.4}$$

定义. 如果不计其在一个 P^*-零测集上的取值, 则 g 是唯一确定的.

定义 1.1 设 f_0 为满足 (1.3) 的任一 $\mathcal{B}_0$-可测函数, 则称 $f_0(X)$ 为在给定子 σ-域 $\mathcal{B}_0$ 或统计量 $T(X)$ 的条件下 $f(X)$ 的**条件期望**, 记为

$$f_0(X) = \mathrm{E}[f(X)|\mathcal{B}_0] = \mathrm{E}[f(X)|T(X)],$$

或

$$f_0(x) = \mathrm{E}[f(X)|T(x)] := \mathrm{E}[f(X)|T(X)=T(x)].$$

或者等价地, 称 $g(T)$ 为在给定 T 的条件下 $f(X)$ 的**条件期望**, 其中 g 为满足 (1.4) 的任一 $\mathcal{G}$-可测函数, 记为

$$g(T) = \mathrm{E}[f(X)|T], \text{ 或者 } g(t) = \mathrm{E}[f(X)|t] := \mathrm{E}[f(X)|T=t].$$

需要特别指出的是, 条件期望的上述定义只是存在性的, 而不是构造性的, 没有提供任何计算方法. 为求解条件期望, 一般是先借助直观 "找" 出一个解, 再用上述定义加以验证. 对初学者而言, 这就是困难之所在, 只有多做练习才能逐步把握它.

例 1.1 设 $X_1,\cdots,X_n$ 是独立同分布 (i.i.d.) 的随机变量, 具有相同概率分布 P 且其分布函数连续. 令

$$T(x_1,\cdots,x_n) = (x_{(1)},\cdots,x_{(n)}),$$

其中 $x_{(1)} \leqslant x_{(2)} \leqslant \cdots \leqslant x_{(n)}$ 指出这些 x_i 之间的次序. 称 $(X_{(1)},\cdots,X_{(n)})$ 为 $X_1,\cdots,X_n$ 的次序统计量, 或者称其为从分布 P 中抽取的容量为 n 的次序样本.

不失一般性, 我们可以只限于考虑 $x_{(1)} < x_{(2)} < \cdots < x_{(n)}$ 的那些点, 因为有两个坐标相等的概率为 0. 因而 $\mathcal{X}$ 可取为不同坐标的所有 n 数组的集合, $\mathcal{T}$ 为所有有序 n 数组的集合, $\mathcal{B}$ 和 $\mathcal{G}$ 分别为 $\mathcal{X}$ 和 $\mathcal{T}$ 中 Borel 子集的类. 在 T^{-1} 之下, 由单个向量 $a = (a_1, \cdots, a_n)$ 组成的单点集变换成由 $n!$ 个点组成的集合, 它可通过将 a 的坐标进行所有可能的置换而得到. 因此, $\mathcal{B}_0$ 是所有对称的 Borel 子集 B_0 组成的集类, 对称的含义为, 如果 B_0 包含点 $x = (x_1, \cdots, x_n)$, 则它也包含所有形如 $(x_{i_1}, \cdots, x_{i_n})$ 的点, 此处 $(i_1, \cdots, i_n)$ 是 $(1, \cdots, n)$ 的任一置换. 对于任何可积函数 f, 令

$$f_0(x) = \frac{1}{n!} \sum f(x_{i_1}, \cdots, x_{i_n}),$$

此处, $(i_1, \cdots, i_n)$ 跑遍 $(1, \cdots, n)$ 的所有置换. 函数 f_0 是 $\mathcal{B}_0$-可测的, 因为它关于 n 个自变量是对称的. 由于

$$\int_{B_0} f(x_1, \cdots, x_n) dP(x_1) \cdots dP(x_n) = \int_{B_0} f(x_{i_1}, \cdots, x_{i_n}) dP(x_1) \cdots dP(x_n),$$

f_0 满足 (1.3), 因而 $f_0(x)$ 即为 $f(X)$ 在给定 $T(x)$ 时的条件期望.

由上述定义容易验证, 对于期望的每一条通常的性质, 都有条件期望的一条性质与之相应.

引理 1.5 设 T 是 $(\mathcal{X}, \mathcal{B})$ 到 $(\mathcal{T}, \mathcal{G})$ 的可测变换, $f, g, h, \cdots$ 均为 $(\mathcal{B}, P)$-可积的函数, $a, b, \cdots$ 为实数, 则有

(i) $\mathrm{E}[af(X) + bg(X)|T] = a\mathrm{E}[f(X)|T] + b\mathrm{E}[g(X)|T]$ a.s.;

(ii) 若 f 非负, 则 $\mathrm{E}[f(X)|T] \geqslant 0$ a.s.;

更一般地, 若 $f \leqslant g$, 则 $\mathrm{E}[f(X)|T] \leqslant \mathrm{E}[g(X)|T]$ a.s.;

(iii) $\mathrm{E}[h(T)f(X)|T] = h(T)\mathrm{E}[f(X)|T]$ a.s.;

(iv) $\mathrm{E}[\mathrm{E}(f(X)|T)] = \mathrm{E}[f(X)]$ (条件期望的平滑性);

(v) 若又设 S 是 $(\mathcal{T}, \mathcal{G})$ 到 $(\mathcal{S}, \mathcal{H})$ 的可测变换, 则 $\mathrm{E}[\mathrm{E}(f(X)|T)|S] = \mathrm{E}[f(X)|S]$ a.s.;

(vi) 设 $0 \leqslant f_n \uparrow f$, 则 $\mathrm{E}[f_n(X)|T] \uparrow \mathrm{E}[f(X)|T]$ a.s.;

(vii) 设所有 f_n 非负, 则 $\mathrm{E}[\liminf_{n\to\infty} f_n(X)|T] \leqslant \liminf_{n\to\infty} \mathrm{E}[f_n(X)|T]$ a.s.;

(viii) 设 $|f_n| \leqslant h$, $f_n \to f$ a.s., 则 $\mathrm{E}[f_n(X)|T] \to \mathrm{E}[f(X)|T]$ a.s. 成立.

事实上, 性质 (i)—(v) 可直接由条件期望定义得出, 性质 (vi)—(viii) 可由积分的有关定理得出. 以 (vi) 为例. 记 $g_n(t) = \mathrm{E}[f_n(X)|t]$. 由性质 (ii), 序列 g_n 为 a.e. P^* 上升且有一 $\mathcal{G}$-可测的极限函数 g. 依条件期望的定义, 对 $\forall G \in \mathcal{G}$, $\int_G g_n(t) dP^*(t) = \int_{T^{-1}(G)} f_n(x) dP(x)$. 因 $g_n \uparrow g$ 而 $f_n \uparrow f$, 由单调收敛定理, 对 $\forall G \in \mathcal{G}$, $\int_G g(t) dP^*(t) = \int_{T^{-1}(G)} f(x) dP(x)$, 即 $g(T) = \mathrm{E}[f(X)|T]$, 因而 $\mathrm{E}[f_n(X)|T] \uparrow \mathrm{E}[f(X)|T]$ a.s..

在 (iii) 中对 $h(T)f(X)$ 取条件期望时, 因 T 是给定的, $h(T)$ 可作为一个常数提出来. 性质 (iv) 和 (v) 显示的是条件期望的平滑性质, 表明一次平均可以分两步进行, 在某些情况下这是一种方便的方法.

1.1.4 条件概率的定义和性质

以 I_B 记集合 B 的示性函数, 则 $P(X\in B)=\mathrm{E}[I_B(X)]$, 因而有如下定义.

定义 1.2 设 $B\in\mathcal{B}$, 则可以把在 T 给定的条件下 B 的**条件概率**定义为

$$P(B|T)=\mathrm{E}[I_B(X)|T]. \tag{1.5}$$

依据 (1.4), $P(B|t)$ 是一 $\mathcal{G}$-可测函数, 且满足方程

$$P[B\cap T^{-1}(G)]=\int_G P(B|t)dP^*(t),\quad \text{对所有}\quad G\in\mathcal{G}. \tag{1.6}$$

如果不计它在一个 P^*-零测集上的取值, 则它是唯一确定的.

记 $\mathcal{B}_0=T^{-1}(\mathcal{G})$, 则可等价地定义 $P\big(B|T(x)\big)$ 为一 $\mathcal{B}_0$-可测函数, 且满足方程

$$P[B\cap B_0]=\int_{B_0}P\big(B|T(x)\big)dP(x),\quad \text{对所有}\quad B_0\in\mathcal{B}_0. \tag{1.7}$$

关于条件概率的性质, 我们有如下引理.

引理 1.6 设 T 是 $(\mathcal{X},\mathcal{B})$ 到 $(\mathcal{T},\mathcal{G})$ 的可测变换, $B_1,B_2,\cdots$ 为 $\mathcal{B}$ 中的集合, 则以下性质 a.e. $(\mathcal{G},P^*)$ 成立.

(i) 若 $B_1\subset B_2$, 则 $P(B_1|t)\leqslant P(B_2|t)$;

(ii) 若 $B_1,B_2,\cdots$ 互斥, 则有

$$P\Big(\bigcup_{i=1}^{\infty}B_i|t\Big)=\sum_{i=1}^{\infty}P(B_i|t);$$

若 $B_n\uparrow B$ 或 $B_n\downarrow B$, 则 $P(B_n|t)\ \to\ P(B|t)$;

(iii) $P(\mathcal{X}|t)=1,\quad P(\emptyset|t)=0.$

1.1.5 条件概率分布

依定义 1.2, 条件概率 $P(B|t)$ 应看作 B 固定时 t 的一个 $\mathcal{G}$-可测函数. 而在初等情形, 我们是将 t 视为固定的, 而将 $P(B|t)$ 看作 B 变化时在 $\mathcal{B}$ 上定义的一个集合函数. 引理 1.6 暗示了对固定的 t, 将 $P(B|t)$ 解释成 $\mathcal{B}$ 上的概率测度或概率分布的可能性. 但是, 为使上述性质成立, 对于每一组 $\{B_i\}$, 都要去除 t 的一个 P^*-零测集, 而这个例外集会随着 $\{B_i\}$ 的变化而变化, 它们的并可能已不再是一个 P^*-零测集. 幸而在 $(\mathcal{X},\mathcal{B})$ 为欧氏样本空间这个重要情形, 我们可以利用条件概率函数在那些例外集上的非唯一性, 适当选取条件概率, 使得引理 1.6 三条性质中 "a.e. $(\mathcal{G},P^*)$" 的尾巴可以去掉. 这就引出了下述定义.

定义 1.3 设 P 为样本空间 $(\mathcal{X},\mathcal{B})$ 上的一个概率分布, T 为取值于 $(\mathcal{T},\mathcal{G})$ 的一个统计量. 函数 $P(B,t),B\in\mathcal{B},t\in\mathcal{T}$ 称为一个**条件概率分布**, 如果它满足以下两个条件:

(i) 对每个固定的 $t\in\mathcal{T}$, $P(\cdot,t)$ 是 $\mathcal{B}$ 上的概率测度;

(ii) 对每个固定的 $B\in\mathcal{B}$, $P(B,t)$ 为在 $T=t$ 给定的条件下 B 的条件概率.