国际名校经典教材

[英] 戴维·威廉姆斯（剑桥大学统计实验室） 著
郑坚坚 译

概率和鞅

Probability with Martingales

学技术大学出版社

国际名校经典教材

[英] 戴维·威廉姆斯（剑桥大学统计实验室） 著
郑坚坚 译

概率和鞅

Probability with Martingales

中国科学技术大学出版社

安徽省版权局著作权合同登记号：第 12171794 号

图书在版编目(CIP)数据

概率和鞅/(英)戴维·威廉姆斯(David Williams)著;郑坚坚译. —合肥:中国科学技术大学出版社,2018.3
书名原文: Probability with Martingales
ISBN 978-7-312-04368-0

Ⅰ.概…　Ⅱ.①戴…　②郑…　Ⅲ.概率论　Ⅳ.①O21　②O211

中国版本图书馆 CIP 数据核字(2017)第 294420 号

出版　中国科学技术大学出版社
　　　安徽省合肥市金寨路 96 号,230026
　　　http://press.ustc.edu.cn
　　　https://zgkxjsdxcbs.tmall.com
印刷　安徽国文彩印有限公司
发行　中国科学技术大学出版社
经销　全国新华书店
开本　710 mm×1000 mm　1/16
印张　18.25
字数　388 千
版次　2018 年 3 月第 1 版
印次　2018 年 3 月第 1 次印刷
定价　48.00 元

序

请阅读!

本书最重要的章节是 E 章: 练习题 (Exercises). 这是我留给你们做的堪称有趣的事情. 你们可以现在就开始做其中的"EG"练习题, 但需要看一下本序言后文中的"更多关于习题".

本书 (它基本上是为剑桥大学的三年级学生所开设的课程的讲义) 是一本关于概率的严格理论的尽可能生动 (就我的能力而言) 的入门书. 由于其相当大的篇幅被用来论述鞅, 故它也应该是充满活力的: 请看第 10 章的那些练习题! 当然, 在涉足测度论基础时, 开始总会有深奥与滞重的感觉. 然而必须强调: 测度论的那些就其自身而言最枯燥的内容一旦应用于概率论, 便会变得非常活泼而具有生气. 这不仅仅是因为它得到了应用, 同时也由于其自身的内涵得到了极大的丰富.

你无法回避测度论: 概率论中的一个事件是一个可测集, 一个随机变量是样本空间上的可测函数, 一个随机变量的期望是其关于概率测度的积分, 等等. 当然, 也可以将测度论的一些主要结果作为不证自明的结论放在正文里, 而在附录中给出它们的详细证明; 这也恰好是我在本书中所做的.

测度论就其本身的结构而言是建立在测度的加法法则基础之上的. 概率论则以乘法法则对它加以补充, 后者是刻画独立性的. 事情已经得到改善. 但真正使内容得到充实并充满活力的是我们已能处理许多 σ-代数, 而不是像在测度论中那样, 只涉及一个 σ-代数.

在准备这本书的过程中, 我曾考虑过所有当时我认为多少有一些新意的论题, 但是对于其中的大部分最终我不得不忍痛割爱.

若想对本书中的许多议题有更深入的了解, 可以参考 Billingsley (1979)、Chow 和 Teicher (1978)、Chung (1968)、Kingman 和 Taylor (1966)、Laha 和 Rohatgi (1979) 以及 Neveu (1965) 等人的著作. 至于测度论, 我当初学习它是通过阅读 Dunford 和 Schwartz (1958) 以及 Halmos (1959) 等著作. 在读

完本书之后，你应该去阅读在现今仍属精彩的 Breiman（1968）的书，而为了在离散鞅论方面获得精辟的指点，你应该去看 Hall 和 Heyde（1980）的书.

当然，直觉要比测度论知识重要得多，你应该利用每一个机会来磨砺你的直觉. 没有比 Aldous（1989）更好的“磨刀石”了，尽管它是一本非常难读的书. 为了开阔概率论的视野并学习如何去思考其相关问题，我强力推荐 Karlin 和 Taylor（1981）、Grimmett 和 Stirzaker（1982）、Hall（1988）等的著作，还有近期 Grimmett（1989）关于过滤的极好的书.

更多关于习题. 在编写 E 章（它刚好是由我平时布置给剑桥大学学生的家庭作业题所构成）的时候我考虑了这样一点: 像其他任何一本数学书一样，本书也暗含着一大批另外的练习题，而其中的很多题都要比 E 章中的题容易. 当然，我更欣赏的还是你们在阅读有关结论之后能自创练习题，并且尝试亲自去证明它们，而不是先去找已有的证明. 关于习题还需加以说明的一点是希望你们能够原谅我的地方，诸如我在第 4 章的练习题中使用了数学期望 E，而 E 的严格定义却要到第 6 章才出现之类.

致谢. 首先我要感谢那些坚持听我的课（它是本书的基础）的学生们，他们出众的才华与素养促使我努力改进本课，使之能与他们的水平相称. 感谢那些在我之前就讲授了本课程的教师们，尤其是 David Kendall. 感谢 David Tranah 和剑桥大学出版社的其他员工，他们帮助我将该课程变成了这本书. 我还要感谢 Ben Garlin、James Norris 和 Chris Rogers，若没有他们，本书会包含更多的错误与含混之处.（本书残存的错误则应由我本人负责.）Helen Rutherford 和我打出了部分文稿，但本书的绝大部分文稿是由 Sarah Shea-Simonds 以娴熟的、堪与 Horowitz（钢琴家——译者注）相比的优美指法打出来的. 感谢 Helen，尤其要感谢 Sarah. 还要特别感谢我的妻子 Sheila 的全力相助.

但是我的最大感谢（也是你们的，若你们从本书中有所获益的话）应该献给三个其姓名以大写字母出现在本书索引中的人 J. L. 杜布、A. N. 柯尔莫哥洛夫和 P. 莱维: 如果没有他们的工作，则当杜布在其著作（1953）中作精彩总结时将没有多少内容可写.

戴维 · 威廉姆斯

1990 年 10 月于剑桥大学统计实验室

一个术语问题
——随机变量: 是函数还是等价类?

按本书的水平, 若将一个随机变量视为样本空间上一个可测函数的等价类, 则有关的理论会变得更“漂亮”, 两个函数属于相同的等价类当且仅当它们几乎处处相等. 那样的话, 条件期望映射

$$X \longmapsto E(X|\mathcal{G})$$

将成为一个真正定义明确的从 $L^p(\Omega,\mathcal{F},P)$ 到 $L^p(\Omega,\mathcal{G},P)(p \geqslant 1)$ 的压缩映射; 同时我们也不必总是提及 (随机变量) 不同的版本 (等价类的不同代表元), 且可以避免那些没完没了的“几乎必然”之类的限制条件.

然而我选择了“不漂亮”的方式: 首先, 我更喜欢处理函数, 例如我更喜欢

$$4+5=2 \bmod 7,$$

而不是

$$[4]_7+[5]_7=[2]_7.$$

但是还有一个实质性的原因. 我希望本书能吸引读者将来深入了解更加有趣, 也更加重要的理论, 其中我们的随机过程的参数集合是不可数的 (例如时间参数集合可能是 $[0,\infty)$). 在那里, 等价类的构想是不适用的: 引进商空间的所谓“聪明”会丧失那种即便是建立基本的、基于连续变化的结果也必须具有的精细与灵敏度, 还有其他一些东西, 除非你想表演那种扭曲的“柔身术”, 然而其结果却很难做到完美. 即便这些柔身术允许你建立某些结果, 但是要证明它们却仍然要使用真正的函数. 所以, 其现实性究竟何在?!

符号说明

► 表示重要的事, ►► 非常重要的事, ►►► 诸如鞅收敛定理.

我用 “:=” 来表示 “定义为” . 这一帕斯卡 (Pascal) 符号用起来特别方便, 因为它同时也具有反向的意义.

我使用分析论者 (相对于范畴论者) 的约定:

►
$$\mathbf{N} := \{1,2,3,\cdots\} \subseteq \{0,1,2,\cdots\} =: \mathbf{Z}^+.$$

另外众所周知: $\mathbf{R}^+ := [0,\infty)$.

对于一个包含于某全集 S 中的集合 B, I_B 表示 B 的示性函数, 即 $I_B : S \to \{0,1\}$ 且

$$I_B(s) := \begin{cases} 1, & \text{若 } s \in B, \\ 0, & \text{否则}. \end{cases}$$

对于 $a,b \in \mathbf{R}$,

$$a \wedge b := \min(a,b), \quad a \vee b := \max(a,b).$$

CF: 特征函数; DF: 分布函数; pdf: 概率密度函数.

σ- 代数: $\sigma(\mathcal{C})$(1.1); $\sigma(Y_\gamma : \gamma \in \mathcal{C})$(3.8,3.13). π- 系 (1.6); d- 系 (A1.2).

---

a.e.: 几乎处处, 几乎每点 (1.5).

a.s.: 几乎必然 (2.4).

$b\Sigma$: 有界 Σ- 可测函数空间 (3.1).

$\mathcal{B}(S)$: S 上的博雷尔 σ- 代数, $\mathcal{B} := \mathcal{B}(\mathbf{R})$(1.2).

$C \bullet X$: 离散随机积分 (10.6).

$\mathrm{d}\lambda/\mathrm{d}\mu$: 拉东 – 尼可丁导数 (5.14).

$\mathrm{d}Q/\mathrm{d}P$: 似然比 (14.13).

$E(X):X$ 的期望 $E(X):=\int_\Omega X(\omega)P(\mathrm{d}\omega)$(6.3).

$E(X;F):\int_F X\mathrm{d}P$(6.3).

$E(X/\mathcal{G})$: 条件期望 (9.3).

$(E_n,\mathrm{ev}):\liminf E_n$(2.8).

$(E_n,\mathrm{i.o.}):\limsup E_n$(2.6).

f_X:X 的概率密度函数 (pdf)(6.12).

$f_{X,Y}$: 联合 pdf(8.3).

$f_{X|Y}$: 条件 pdf(9.6).

$F_X:X$ 的分布函数 (3.9).

$\liminf$: 关于集合的 (2.8).

$\limsup$: 关于集合的 (2.6).

$x=\uparrow\lim x_n:x_n\uparrow x$, 即满足: $x_n\leqslant x_{n+1}(\forall n)$ 且 $x_n\to x$.

log: 自然对数 (以 e 为底的).

$\mathcal{L}_X,\Lambda_X:X$ 的分布（3.9）.

$\mathcal{L}^p,L^p$: 勒贝格空间（6.7,6.13）.

Leb: 勒贝格测度（1.8）.

$m\Sigma$: Σ- 可测函数空间（3.1）.

M^T: 在时间 T 停止的过程 M(10.9).

$\langle M\rangle$: 尖括号过程 (12.12).

$\mu(f):f$ 的关于 μ 的积分（5.0,5.2）.

$\mu(f;A):\int_A f\mathrm{d}\mu$（5.0,5.2）.

$\varphi_X:X$ 的特征函数（第 16 章）.

φ: 标准正态分布 $N(0,1)$ 的 pdf.

$\Phi:N(0,1)$ 的分布函数.

X^T: 在时间 T 停止的 X(10.9).

目　　录

B 部分 鞅 论

C 部分　特征函数

附 录