随机逼近及自适应算法

西安交通大学数学研究生教学丛书

聂赞坎　徐宗本　编著

科学出版社
www.sciencep.com

内 容 简 介

本书系统地介绍了随机逼近理论，着重介绍了几类自适应算法以及随机逼近理论对它们的应用．作为一本研究生教材，本书难度适中，注重基础知识的讲述，深入浅出，易于自学．

读者对象：高等院校数学系高年级学生，信息科学、计算机科学、运筹学、应用数学等专业的研究生．

图书在版编目(CIP)数据

随机逼近及自适应算法/聂赞坎，徐宗本编著．—北京：科学出版社，2003．1

(西安交通大学数学研究生教学丛书)

ISBN 7-03-010534-6

Ⅰ．随… Ⅱ．①聂… ②徐… Ⅲ．①随机逼近-研究生-教材②自适应控制-算法理论-研究生-教材 Ⅳ．①O174．41 ②TP273

中国版本图书馆 CIP 数据核字(2002)第 047935 号

责任编辑：林 鹏 杨 波/责任校对：宣 慧

责任印制：安春生/封面设计：高海英 王 浩

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

新蕾印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2003年1月第 一 版 开本：B5 (720×1000)

2003年1月第一次印刷 印张：9 1/2

印数：1—3 000 字数：176 000

定价：19.00 元

(如有印装质量问题，我社负责调换〈新欣〉)

前　言

随机逼近算法可以理解为是利用观测值估计未知函数的极值或未知方程解的自适应问题求解技术. 它起源于20世纪50年代初Ribbins和Monro所提出的求未知函数零点的一个递推算法(称之为RM算法)以及此后Kiefer和Wolfowitz等人有关未知函数极值问题的研究. 随机逼近理论(即随机逼近算法的理论基础与分析)近几十年来得到了飞速发展,已成为数理统计与人工智能的交叉分支,并已广泛地应用于系统辩识、自适应控制、模式识别、自适应滤波、神经元网络等领域. 有关随机逼近理论的研究,近几年出现了Albert Benveniste、Lennart Ljung、陈翰馥等国内外著名学者的专著,它们对随机逼近理论的方方面面作了系统、深入的阐述. 这些著作虽各有深入讨论的侧重面,但由于篇幅过大,论题过多,对初学者来说,不易掌握. 本书的编写目的在于向读者(特别是数学系研究生)提供一本难度适中、内容基本、易于自学的教学用书. 因此,本书的前一部分(第一章~第四章)着重介绍随机逼近基础理论中的核心问题(随机逼近算法的收敛性)和基本的收敛性结果,以及这些基本结果在一些具体自适应性算法中的应用. 本书后一部分集中介绍近几年随高性能计算机发展而出现的几类自适应算法(遗传算法、模拟退火算法、主成分分析神经网络算法)以及随机逼近理论对它们的应用.

遗传算法是模拟自然界生物进化过程与机制求解极值问题的一类自适应算法. 它的产生归功于美国Michigan大学的Holland在20世纪60年代末70年代初的开创性工作. 他不仅设计了遗传算法的模拟与操作原理,而且更重要的是运用统计决策理论对遗传算法的搜索机理进行了理论分析,从而为遗传算法的发展奠定了基础. 近十几年来,遗传算法无论是在应用上,还是在基础理论上都取得了长足发展,已成为信息科学、计算机科学、运筹学和应用数学等诸多学科所共同关注的热点研究领域. 但相对卓有成效的、广泛的实际应用而言,遗传算法的数学基础理论研究还相对滞后,还不能说很完善或很深入. 近期研究主要是围绕如何提高算法效率和建立算法理论基础方面. 本书第五章和第六章专门讨论遗传算法的随机过程分析及其收敛性分析,其中大部分内容是西安交通大学理学院信息与系统科学研究所师生们近期的科研成果.

模拟退火算法源于对固体退火过程的模拟,是解全局优化问题(特别是组合优化问题)的一种通用自适应方法. 该类方法以随机方法为基础,结合一系列相关迭代算法,既克服了迭代算法的一些本质缺陷(如收敛的局部性,通常只收敛到问题的局部极小),而又保持了它们的突出优点,如快速局部收敛和

所获得的解不依赖于初始状态等.本书第七章专门讨论马尔可夫链模型的模拟退火算法,介绍齐次算法和非齐次算法的一些渐近收敛性结果.

主成分分析是一种经典的统计技术,它用于分析多变量统计观察的方差结构,它与估计理论中的最小方差技术,时间序列分析中的 Karhunen Loeve 变换和数值分析中的奇异值分解密切相关.这些方法在信号处理、图像编码、信息压缩等领域有十分基本而重要的应用.近年来,人们利用神经网络技术求解主成分显示出极大的优越性,许多用于求解主成分的神经网络算法问世.这些主成分分析神经网络算法的突出优点是它们的自组织性和自适应性,而且易于硬件实现.本书最后一章介绍与此相关的一些简单、通用算法,并且作为随机逼近理论的应用,研究这些算法的收敛性质.

全书共分八章.第一章给出本书必需的概率论方面的预备知识.第二章介绍研究随机逼近算法的三种常用方法.第三章讨论形式非常一般的具有局部有界矩随机逼近算法的几乎必然收敛性.第四章将第三章的收敛性结果应用于一些具体的适应性算法,获得这些算法的几乎必然收敛性.第五章介绍遗传算法的基本概念、过程分析和遗传算法的 Markov 链模型.第六章对抽象的(算子型)遗传算法在各种意义下的收敛性进行了讨论,并对两类特殊类型遗传算法的收敛性进行了细致分析.第七章建立模拟退火算法的马尔可夫链模型,介绍齐次算法和非齐次算法的一些渐近收敛性结果.第八章介绍一些简单、通用的主成分分析神经网络算法,并利用随机逼近理论研究这些算法的收敛性.

本书的主要内容曾经在西安交通大学理学院应用数学系硕士生的课程上讲授过,有不少内容是直接由文献改写的.限于作者自身水平,难免有遗漏与不妥之处,恳请读者批评与指正.

编　著

2002 年 8 月

于西安交通大学

目　　录

第一章　预备知识

本章给出阅读本书所必需的若干预备知识,包括概率论、离散参数鞅和马尔可夫链等领域中与本书有关的一些事实与概念.

§1.1　概率论的若干基本概念

1.1.1　随机变量及其分布

设$(\Omega,\mathscr{F},P)$为一概率空间,X 是在 Ω 上定义的实值函数,记 $X^{-1}(B)=\{\omega:X(\omega)\in B\}$,$B\subset R$,$X^{-1}(B)$称为集合 B 的原像.

定义 1.1.1　定义在可测空间$(\Omega,\mathscr{F})$上的实函数 X,如果对数直线上任意 Borel 集 B,$X^{-1}(B)\in\mathscr{F}$,则称 X 为 $\mathscr{F}$ 可测,概率空间上的可测函数也称为随机变量.

对可测函数 X,集系$\{X^{-1}(B):B\in\mathscr{B}(R)\}$构成一个 σ 域,称为由 X 生成的 σ 域,记为 $\sigma(X)$,其中 $\mathscr{B}(R)$是 R 上的 Borel 域. 对于随机变量 X,定义函数

$$F(x)=P\{\omega:X(\omega)\leqslant x\},\quad x\in R,$$

$F(x)$称为 X 的分布函数.

1.1.2　随机变量列的收敛性

设$\{X,X_n\}$为一列随机变量.

(1)如果对于任意正数 ε,有 $\lim\limits_{n\to\infty}P(|X_n-X|>\varepsilon)=0$,则称$\{X_n\}$依概率收敛到 X,记作 $X_n\xrightarrow{P}X$, $n\to\infty$.

(2)如果 $P(\lim\limits_{n\to\infty}X_n=X)=1$,则称随机变量列 X_n 概率 1 收敛到 X,或称随机变量列 X_n 几乎处处收敛到 X,记作 $X_n\xrightarrow{a.s.}X$, $n\to\infty$.

(3)如果 $\lim\limits_{n\to\infty}E(|X_n-X|^2)=0$,则称随机变量列 X_n 均方收敛到 X,记作 $X_n\xrightarrow{L^2}X$, $n\to\infty$.

上面各种收敛性有如下关系

$$X_n\xrightarrow{a.s.}X,n\to\infty\Rightarrow X_n\xrightarrow{P}X,n\to\infty\Leftarrow X_n\xrightarrow{L^2}X,n\to\infty.$$

$$X_n \xrightarrow{P} X, n \to \infty \Rightarrow \text{存在子列 } X_{n_k} \xrightarrow{a.s.} X, k \to \infty.$$

1.1.3 随机变量的期望和条件期望

定义 1.1.2 设随机变量 $X \geqslant 0, a.s.$，记

$$A_{ni} = \{\omega : i2^{-n} < X(\omega) \leqslant (i+1)2^{-n}\}, \quad n, i = 1, 2, \cdots$$

X 的数学期望 EX 定义为

$$EX = \int_\Omega X \mathrm{d}P = \lim_{n \to \infty} \Big[\sum_{i=0}^{n2^n - 1} i2^{-n} P(A_{ni}) + nP(X > n) \Big].$$

对不是非负的随机变量 X，可表为正部 $X^+ = \max(X, 0)$ 与负部 $X^- = \max(-X, 0)$ 的差 $X = X^+ - X^-$，如果 $EX^+ < \infty$ 或 $EX^- < \infty$，则定义 X 的数学期望 $EX = EX^+ - EX^-$；当 $EX^+ < \infty, EX^- < \infty$ 时，称 X 是可积的；当 EX 有定义时，它还可表为 Lebesgue-Stieltjes 积分

$$EX = \int_{-\infty}^{\infty} x \mathrm{d}F(x).$$

由初等概率论知，在事件 $A(P(A) > 0)$ 发生的条件下，事件 B 发生的条件概率定义为

$$P(B \mid A) = P(AB)/P(A), \tag{1.1.1}$$

如果 $P(A) = 0$，就规定 $P(B|A) = 0$.

固定 $A(P(A) > 0)$，$P(\cdot|A)$ 是 $\mathscr{F}$ 上的概率测度，(1.1.1)式可写成

$$\int I_B(\omega) P(\mathrm{d}\omega \mid A) = \frac{1}{P(A)} \int_A I_B(\omega) P(\mathrm{d}\omega),$$

其中 I_B 表示集 B 的示性函数. 因此随机变量 X 关于 A 的条件数学期望可定义为

$$E(X \mid A) = \int_\Omega X(\omega) P(\mathrm{d}\omega \mid A) = \frac{1}{P(A)} \int_A X(\omega) P(\mathrm{d}\omega). \tag{1.1.2}$$

如果 $P(A) = 0$，则规定 $E(X|A) = 0$. 特别当 $X(\omega) = I_B(\omega)$ 时，(1.1.2)就化为(1.1.1)，故我们只须考虑条件期望.

将 $E(X|A)$ 推广如下：考虑 Ω 的某一可测分割 $A_1, A_2, \cdots$，即对一切 $i \geqslant 1$，$A_i \in \mathscr{F}$，$A_iA_j = \varnothing$，$i \neq j$，$\sum_{i=1}^{\infty} A_i = \Omega$. 为了从整体上考查 $E(X|A_i)$，$i = 1, 2, \cdots$，定义随机变量 $Y(\omega)$ 如下

$$Y(\omega) = \sum_{i=1}^{\infty} E(X \mid A_i) I_{A_i}(\omega), \tag{1.1.3}$$

于是当 $E(|X|) < \infty$，$P(A_i) > 0$ 时，由(1.1.2)，(1.1.3)知

$$\int_{A_i} X(\omega)P(\mathrm{d}\omega) = P(A_i)E(X \mid A_i) = \int_{A_i} Y(\omega)P(\mathrm{d}\omega), \quad (1.1.4)$$

如果 $P(A_i)=0$,上式两边都等于零,故(1.1.4)仍然成立. 易见,当取$\{A, A^c\}$,$P(A)>0$,为 Ω 的可测分割时,在(1.1.4)中以 A 代替 A_i 即得(1.1.2). 令 $\mathscr{G}=\sigma(A_i: i=1,2,\cdots)$是含一切$\{A_i: i=1,2,\cdots\}$的最小 σ 域. 由于 $\mathscr{G}$ 中元素(除空集外)都是某些 A_i 的併集,故由(1.1.4)知,对一切 $B\in\mathscr{G}$,有

$$\int_B X(\omega)P(\mathrm{d}\omega) = \int_B Y(\omega)P(\mathrm{d}\omega), \quad (1.1.5)$$

记 $Y=E(X|\mathscr{G})$,则 $E(X|\mathscr{G})$具有性质:

(1)$E(X|\mathscr{G})$是 $\mathscr{G}$ 可测. (1.1.6)

(2)$\forall B\in\mathscr{G}, \int_B X(\omega)P(\mathrm{d}\omega) = \int_B E(X \mid \mathscr{G})(\omega)P(\mathrm{d}\omega).$ (1.1.7)

根据上面对特殊的 σ 域 $\mathscr{G}$ 讨论的启示,我们将条件期望推广到一般的 σ 域.

定义 1.1.3 给定随机变量 X 及 $\mathscr{F}$ 的子 σ 域 $\mathscr{G}$,$E(|X|)<\infty$,称随机变量 $E(X|\mathscr{G})$为 X 关于 $\mathscr{G}$ 的条件数学期望,如果它满足(1.1.6)和(1.1.7).

为使上述定义有意义,必须保证满足(1)和(2)的随机变量存在. 为此考查集函数

$$\varphi(B) = \int_B X\mathrm{d}P, \qquad \forall B \in \mathscr{G},$$

易知,$\varphi(\cdot)$是 $\mathscr{G}$ 上广义测度,而且如果 $P(B)=0$,则有 $\varphi(B)=0$. 即 φ 在 $\mathscr{G}$ 上关于测度 P 绝对连续. 根据 Radon-Nikodym 定理,在关于测度 P 几乎处处相等的意义下,满足(1)和(2)的随机变量 $E(X|\mathscr{G})$惟一存在. 我们说 X 关于 $\mathscr{G}$ 的条件数学期望时,指的是上述等价类中的一个代表.

例 1.1.4 若随机变量 X 关于 $\mathscr{G}$ 独立,则有

$$E(X \mid \mathscr{G}) = EX. \quad (1.1.8)$$

证明 因为 EX 关于 $\mathscr{G}$ 可测,且 $\forall B\in\mathscr{G}$,

$$\int_B X\mathrm{d}P = E(XI_B) = E(X)P(B) = \int_B EX\mathrm{d}P.$$

例 1.1.5 若 $\mathscr{G}=\mathscr{F}$或 X 为 $\mathscr{G}$ 可测,则有

$$E(X \mid \mathscr{G}) = X, \quad a.s. \quad (1.1.9)$$

特别$\forall B\in\mathscr{G}$有

$$P(B \mid \mathscr{G}) = I_B, \quad a.s. \quad (1.1.10)$$

例1.1.6 假设X与Y同时为连续型随机变量,$E(|Y|)<\infty$,而且(X,Y)有联合概率密度 $f(x,y)$,$x,y\in R$,如果 $\mathscr{G}=\sigma(X)$,即 $\mathscr{G}$ 是包含一切$\{(X\in B), B\in\mathscr{B}(X)\}$的最小 σ 域,则 $E(Y|\mathscr{G})$与概率论中定义的 Y 关于X 的条件期望一致. 事实上,Y 在$X=x$ 条件下的条件概率密度定义为

$$f(y|x)=f(x,y)/p(x),\qquad p(x)>0,\tag{1.1.11}$$

其中 $p(x)=\int_{-\infty}^{\infty}f(x,y)\mathrm{d}y$，当 $p(x)=0$ 时，规定(1.1.11)左方值为零. 于是 Y 在 $X=x$ 条件下的条件期望为

$$E(Y|X=x)=\int_{-\infty}^{\infty}yf(y|x)\mathrm{d}y=g(x),\tag{1.1.12}$$

由(1.1.11)和(1.1.12)知右方为 x 的可测函数，因此得到可测函数

$$g(x)=E(Y|X=x),\qquad x\in R,$$

可见，Y 关于 X 的条件期望 $E(Y|X)=g(X)$ 为 $\sigma(X)$ 可测；往证 $\forall B\in\sigma(X)$，有

$$\int_B g(X)\mathrm{d}P=\int_B Y\mathrm{d}P,\tag{1.1.13}$$

假设 $h(x)$ 为任意有界可测函数，$A=\{x:p(x)>0\}$，则有

$$\begin{aligned}E(h(X)Y)&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}h(x)yf(x,y)\mathrm{d}x\mathrm{d}y\\&=\int_A\left[\int_{-\infty}^{\infty}y\frac{f(x,y)}{p(x)}\mathrm{d}y\right]p(x)h(x)\mathrm{d}x\\&=\int_A g(x)p(x)h(x)\mathrm{d}p\\&=\int_\Omega g(X)h(X)\mathrm{d}p.\end{aligned}$$

如果在上式中令 $h(x)=I_\Gamma(x)$，$\Gamma\in\mathscr{B}(R)$ 则有

$$\int_{(X\in\Gamma)}Y\mathrm{d}p=\int_{(X\in\Gamma)}g(X)\mathrm{d}p,$$

即(1.1.13)成立，可见 $g(X)=E(Y|\mathscr{G})$，$a.s.$.

1.1.4 条件期望的基本性质

以下提到的随机变量 X,X_n 均假设是可积的. 不等式，等式及极限关系都是几乎处处成立，不另再声明.

性质 1.1.7 对于任意实数 c_i，$i=1,2$，有

$$E\Big(\sum_{i=1}^{2}c_iX_i|\mathscr{G}\Big)=\sum_{i=1}^{2}c_iE(X_i|\mathscr{G}).$$

证明 由定义知 $\sum_{i=1}^{2}c_iE(X_i|\mathscr{G})$ 为 $\mathscr{G}$ 可测，而且 $\forall B\in\mathscr{G}$，有

$$\begin{aligned}\int_B\Big[\sum_{i=1}^{2}c_iE(X_i|\mathscr{G})\Big]\mathrm{d}p&=\sum_{i=1}^{2}c_i\int_B E(X_i|\mathscr{G})\mathrm{d}p\\&=\sum_{i=1}^{2}c_i\int_B X_i\mathrm{d}p=\int_B\Big(\sum_{i=1}^{2}c_iX_i\Big)\mathrm{d}p.\end{aligned}$$

性质 1.1.8

(1)如果 $X\geqslant 0$,则有 $E(X|\mathscr{G})\geqslant 0$.

(2)如果 $X_1\geqslant X_2$,则有 $E(X_1|\mathscr{G})\geqslant E(X_2|\mathscr{G})$.

证明 设 $X\geqslant 0$,令

$$B=\{\omega:E(X|\mathscr{G})(\omega)<0\},$$
$$B_m=\{\omega:E(X|\mathscr{G})(\omega)<-1/m\},\quad m=1,2,\cdots,$$

易见 $B=\sum\limits_{m=1}^{\infty}B_m$,而且

$$-\frac{1}{m}p(B_m)\geqslant\int_{B_m}E(X|\mathscr{G})\mathrm{d}p=\int_{B_m}X\mathrm{d}P\geqslant 0,$$

所以 $p(B_m)=0,p(B)=0$,(1)成立;由(1)及性质 1.1.7 推得(2)成立.

性质 1.1.9 $|E(X|\mathscr{G})|\leqslant E(|X||\mathscr{G})$.

证明 由性质 1.1.7 和 1.1.8 直接推得.

性质 1.1.10(单调收敛定理) 假设 $0\leqslant X_n\uparrow X,n\to\infty$,则有

$$\lim_{n\to\infty}E(X_n|\mathscr{G})=E(X|\mathscr{G}).$$

证明 由 $0\leqslant X_n\leqslant X_{n+1}\leqslant X$ 得

$$0\leqslant E(X_1|\mathscr{G})\leqslant E(X_2|\mathscr{G})\leqslant\cdots\leqslant E(X|\mathscr{G}),$$

故对几乎所有 ω,$\lim\limits_{n\to\infty}E(X_n|\mathscr{G})(\omega)$存在,对极限不存在的 ω 定义极限值为零.如此规定后 $\lim\limits_{n\to\infty}E(X_n|\mathscr{G})(\omega)$为 $\mathscr{G}$ 可测.往证它等于 $E(X|\mathscr{G})$.任取 $B\in\mathscr{G}$,根据积分的单调收敛定理得

$$\begin{aligned}\int_B\lim_{n\to\infty}E(X_n|\mathscr{G})\mathrm{d}p&=\lim_{n\to\infty}\int_B E(X_n|\mathscr{G})\mathrm{d}p\\&=\lim_{n\to\infty}\int_B X_n\mathrm{d}p=\int_B X\mathrm{d}p,\end{aligned}$$

所以

$$\lim_{n\to\infty}E(X_n|\mathscr{G})=E(X|\mathscr{G}).$$

性质 1.1.11(Fatou 引理) 设随机变量列$\{X_n;n=1,\cdots\}$的每个 X_n 的期望存在.

(1)若存在随机变量 X,使 $EX>-\infty$,且 $\forall\, n\geqslant 1,X_n\geqslant X,a.s.$,则 $\underline{\lim}_{n\to\infty}X_n$ 的期望存在,且有

$$E[\underline{\lim}_{n\to\infty}X_n|\mathscr{G}]\leqslant\underline{\lim}_{n\to\infty}E[X_n|\mathscr{G}].$$

(2)若存在随机变量 X,使 $EX<\infty$,且 $\forall\, n\geqslant 1,X_n\leqslant X,a.s.$,则 $\overline{\lim}_{n\to\infty}X_n$ 的期望存在,且有

$$E[\overline{\lim_{n\to\infty}} X_n \mid \mathscr{G}] \geqslant \overline{\lim_{n\to\infty}} E[X_n \mid \mathscr{G}].$$

证明 因为

$$0 \leqslant \inf_{k\geqslant n} X_k - X \uparrow \underline{\lim_{n\to\infty}} X_n - X, \quad n \to \infty,$$

由单调收敛定理得

$$\begin{aligned} E[\underline{\lim_{n\to\infty}} X_n \mid \mathscr{G}] - E[X \mid \mathscr{G}] &= E[\underline{\lim_{n\to\infty}} X_n - X \mid \mathscr{G}] \\ &= \lim_{n\to\infty} E[(\inf_{k\geqslant n} X_k - X) \mid \mathscr{G}] \\ &= \underline{\lim_{n\to\infty}} E[X_n - X \mid \mathscr{G}] \\ &= \underline{\lim_{n\to\infty}} E[X_n \mid \mathscr{G}] - E[X \mid \mathscr{G}]. \end{aligned}$$

从而(1)得证,类似地可以证明(2).

性质 1.1.12 (控制收敛定理) 假设$|X_n| \leqslant Y$, Y 可积,而且$\lim\limits_{n\to\infty} X_n = X$,则有

$$\lim_{n\to\infty} E(X_n \mid \mathscr{G}) = E(X \mid \mathscr{G}).$$

证明 令

$$S_n = \sup_{k\geqslant n} X_k, \quad I_n = \inf_{k\geqslant n} X_k,$$

因为

$$|X_n| \leqslant Y, \quad S_n \downarrow X, \quad I_n \uparrow X, \quad n \to \infty,$$
$$0 \leqslant Y - S_n \downarrow Y - X, \quad 0 \leqslant Y + I_n \uparrow Y + X, \quad n \to \infty,$$

由单调收敛定理得

$$E(Y - S_n \mid \mathscr{G}) \downarrow E(Y - X \mid \mathscr{G}), \quad n \to \infty,$$
$$E(Y + I_n \mid \mathscr{G}) \uparrow E(Y + X \mid \mathscr{G}), \quad n \to \infty,$$

从而

$$\begin{aligned} E(X \mid \mathscr{G}) &= \lim_{n\to\infty} E(I_n \mid \mathscr{G}) \\ &\leqslant \lim_{n\to\infty} E(X_n \mid \mathscr{G}) \\ &\leqslant \lim_{n\to\infty} E(S_n \mid \mathscr{G}) \\ &= E(X \mid \mathscr{G}). \end{aligned}$$

性质 1.1.13 如果随机变量 X 是 $\mathscr{G}$ 可测的,而且 $E(|XY|) < \infty$,则有

$$E(XY \mid \mathscr{G}) = XE(Y \mid \mathscr{G}). \tag{1.1.14}$$

证明 令

$$L = \{X : E(|XY|) < \infty\},$$
$$\boldsymbol{L} = \{X : X \text{ 使(1.1.14)满足}\},$$

由性质 1.1.7 和 1.1.10 容易证明 L 是$\boldsymbol{L}$ 系,另外当 $X = I_A, A \in \mathscr{G}$时,$X$ 是 $\mathscr{G}$

可测的,而且对于任意 $B\in\mathscr{G}$,有

$$\int_B XY\mathrm{d}p=\int_{AB}Y\mathrm{d}p=\int_{AB}E(Y|\mathscr{G})\mathrm{d}p=\int_B XE(Y|\mathscr{G})\mathrm{d}p,$$

故 $I_A\in L$,根据 $\boldsymbol{L}$ 系方法(参见文献[2]463 页引理 4)知 L 包含一切属于 $\boldsymbol{L}$ 且关于 $\mathscr{G}$ 可测的随机变量 X.

性质 1.1.14

(1)全数学期望公式:$E(E(X|\mathscr{G}))=EX$.

(2)重条件期望公式:当 $\mathscr{G}_1\subset\mathscr{G}_2$ 时,有

$$E(E(X|\mathscr{G}_1)|\mathscr{G}_2)=E(X|\mathscr{G}_1)=E(E(X|\mathscr{G}_2)|\mathscr{G}_1).$$

证明 (1)在(1.1.7)中令 $B=\Omega$,即得到(1).

(2)因为 $\mathscr{G}_1\subset\mathscr{G}_2$,故 $E(X|\mathscr{G}_1)$ 为 $\mathscr{G}_2$ 可测,从而由例 1.1.5 的结论得第 1 个等式,为证明第 2 个等式,只需注意,如果 $B\in\mathscr{G}_1$,则 $B\in\mathscr{G}_2$,故有

$$\int_B X\mathrm{d}p=\int_B E(X|\mathscr{G}_2)\mathrm{d}p.$$

性质 1.1.15 (Holder 不等式) 设 $1<p,q<\infty,\dfrac{1}{p}+\dfrac{1}{q}=1$,则有

$$E[|XY||\mathscr{G}]\leqslant(E[|X|^p|\mathscr{G}])^{1/p}(E[|Y|^q|\mathscr{G}])^{1/q},$$

特别当 $p=q=2$ 时,有 Cauchy-Schwarz 不等式

$$E[|XY||\mathscr{G}]^2\leqslant E(X^2|\mathscr{G})E(Y^2|\mathscr{G}).$$

证明 当 $E[|X|^p|\mathscr{G}]\neq0,E[|Y|^q|\mathscr{G}]\neq0$ 时,在不等式

$$|ab|\leqslant\frac{1}{p}|a|^p+\frac{1}{q}|b|^q$$

中置

$$a=X/E[|X|^p|\mathscr{G}]\neq0,$$
$$b=Y/E[|Y|^q|\mathscr{G}]\neq0,$$

利用条件期望的线性性,单调性和性质 1.1.11,对所得不等式稍加整理,即得所需结果.

为了便于证明下面 Jensen 不等式,我们介绍关于凸函数的一个性质.称函数 $f(x),x\in R$ 为凸函数,如果对于任意 $x,y\in R,0\leqslant\alpha\leqslant1$ 有

$$f(\alpha x+(1-\alpha)y)\leqslant\alpha f(x)+(1-\alpha)f(y),\qquad(1.1.15)$$

对任意固定的 $x<y$,令 $z=\alpha x+(1-\alpha)y$,由(1.1.15)得

$$\frac{f(z)-f(x)}{z-x}\leqslant\frac{\alpha f(x)+(1-\alpha)f(y)-f(x)}{y-x}=\frac{f(y)-f(x)}{y-x}.$$

令 $\alpha\uparrow1$(这等价于 $z\downarrow x$)得 $f(x)$的右导数 $f'_+(x)$存在,且满足下面不等式

$$f'_+(x) \leqslant \frac{f(y)-f(x)}{y-x},$$

$$f'_+(x)(y-x) \leqslant f(y)-f(x), \tag{1.1.16}$$

显然当 $y<x$ 时,此不等式仍然成立.

性质 1.1.16 (Jensen 不等式) 设 $f(x), x\in R$ 为凸函数,随机变量 X 是可积的,$\mathscr{G}$ 是 $\mathscr{F}$ 的任意子 σ 域则有

$$f[E(X|\mathscr{G})] \leqslant E[f(X)|\mathscr{G}], \quad a.s.$$

证明 在不等式(1.1.16)中,令 $y=X, x=E(X|\mathscr{G})$ 得

$$f'_+(E(X|\mathscr{G}))(X-E(X|\mathscr{G})) \leqslant f(X)-f(E(X|\mathscr{G}))$$

两边对 $\mathscr{G}$ 求条件期望得

$$f[E(X)|\mathscr{G}] \leqslant E[f(X)|\mathscr{G}], \quad a.s.$$

§1.2 离散参数鞅

1.2.1 停时

定义 1.2.1 定义在可测空间$(\Omega,\mathscr{F})$上取值于$\{0,1,\cdots,+\infty\}$的可测函数 $\tau(\omega)$,称为关于 $\mathscr{F}$ 的子 σ 域流$\{\mathscr{F}_n: n=0,1,\cdots\}$是停时,如果对任意非负整数 n,$(\tau\leqslant n)\in\mathscr{F}_n$,这等价于$(\tau=n)\in\mathscr{F}_n$.

若 τ_1, τ_2 为停时,根据定义不难验证 $\tau_1+\tau_2$, $\tau_1\vee\tau_2\equiv\max(\tau_1,\tau_2)$, $\tau_1\wedge\tau_2\equiv\min(\tau_1,\tau_2)$都是停时.进一步若假设$\{\tau_n\}$为一列停时,容易证明 $\sup\limits_{n\geqslant1}\{\tau_n\}$, $\inf\limits_{n\geqslant1}\{\tau_n\}$, $\limsup\limits_{n\to\infty}\limits_{n\geqslant1}\{\tau_n\}$, $\liminf\limits_{n\to\infty}\limits_{n\geqslant1}\{\tau_n\}$都是停时.

定义 1.2.2 设 τ 是$\{\mathscr{F}_n: n=0,1,\cdots\}$停时,令

$$\mathscr{F}_\tau = \{A: A\in\mathscr{F}_\infty, A(\tau\leqslant n)\in\mathscr{F}_n, \forall n\geqslant 0\}, \tag{1.2.1}$$

称 $\mathscr{F}_\tau$ 为 τ 前 σ 域.

其中 $\mathscr{F}_\infty=\sigma(\bigcup_{n=1}^\infty\mathscr{F}_n)$;易知当 $\tau\equiv n$ 时,$\mathscr{F}_\tau=\mathscr{F}_n$.

性质 1.2.3 设 τ, τ_1, τ_2 是$\{\mathscr{F}_n: n=0,1,\cdots\}$停时,则有

(1) τ 为 $\mathscr{F}_\tau$ 可测.

(2) 当 $\tau_1\leqslant\tau_2$ 时,$\mathscr{F}_{\tau_1}\subset\mathscr{F}_{\tau_2}$.

证明 (1)对任意非负整数 m, n,令 $A=(\tau\leqslant m)$,则

$$A(\tau\leqslant n) = (\tau\leqslant n\wedge m)\in\mathscr{F}_{n\wedge m}\subset\mathscr{F}_n \Rightarrow A\in\mathscr{F}_\tau.$$

故 τ 为 $\mathscr{F}_\tau$ 可测.

(2) $\forall A\in\mathscr{F}_{\tau_1}$,由定义知,

$$\forall n\geqslant 0, \quad A(\tau_1\leqslant n)\in\mathscr{F}_n,$$

$$A(\tau_2\leqslant n) = A(\tau_1\leqslant n)(\tau_2\leqslant n)\in\mathscr{F}_n,$$

故 $A\in\mathscr{F}_{\tau_2}$, $\mathscr{F}_{\tau_1}\subset\mathscr{F}_{\tau_2}$.

性质 1.2.4 设 $\tau_1\geqslant\tau_2\geqslant\cdots\geqslant\tau_n\geqslant\cdots$是一列$\{\mathscr{F}_n:n=0,1,\cdots\}$停时,则 $\tau=\inf\limits_{n\geqslant1}\{\tau_n\}$是停时,而且 $\mathscr{F}_\tau=\bigcap\limits_{n\geqslant1}\mathscr{F}_{\tau_n}$.

证明 由性质 1.2.3 知,对任意正整数 n, $\mathscr{F}_\tau\subset\mathscr{F}_{\tau_n}\Rightarrow\mathscr{F}_\tau\subset\bigcap\limits_{n\geqslant1}\mathscr{F}_{\tau_n}$,反之任取 $A\in\bigcap\limits_{n\geqslant1}\mathscr{F}_{\tau_n}$,有

$$A(\tau\leqslant n)=\bigcup_{k\geqslant1}[A(\tau_k\leqslant n)]\in\mathscr{F}_n\Rightarrow A\in\mathscr{F}_\tau.$$

1.2.2 鞅

定义 1.2.5 称过程 $X=\{X_n:n=0,1,\cdots\}$为$\{\mathscr{F}_n:n=0,1,\cdots\}$鞅(相应地上鞅,下鞅),如果

(1) X 是$\{\mathscr{F}_n:n=0,1,\cdots\}$适应的,即对每个 n, X_n 是 $\mathscr{F}_n$ 可测.

(2) $E(|X_n|)<\infty$, $n=0,1,\cdots$.

(3) 对任意非负整数 $m<n$, $E(X_n|\mathscr{F}_m)=X_m$(相应地$\leqslant X_m$, $\geqslant X_m$).

例 1.2.6 假设$\{Y_n:n=0,1,\cdots\}$是独立的随机变量列, $E(|Y_n|^2)<\infty$, $E(Y_n)=0$, g_k 是k 元 Borel 可测函数, $k=1,2,\cdots$令

$$b_k=g_k(Y_0,Y_1,\cdots,Y_{k-1}),$$

$$X_n=X_0+\sum_{k=1}^n b_kY_k,$$

$\mathscr{F}_n=\sigma(Y_0,Y_1,\cdots,Y_n)$, $n=1,2,\cdots$.其中 X_0 为常量,假定 $E(|b_k|^2)<\infty$, $k=1,2,\cdots$,则$\{X_n,\mathscr{F}_n,n=0,1,\cdots\}$是鞅.

事实上,显然对任意 $n=1,2,\cdots$, $E(|X_n|)<\infty$;又因为

$$X_{n+1}=X_n+b_{n+1}Y_{n+1},$$

$$\begin{aligned}E(X_{n+1}|\mathscr{F}_n)&=E(X_n|\mathscr{F}_n)+E(b_{n+1}Y_{n+1}|\mathscr{F}_n)\\&=X_n+b_{n+1}E(Y_{n+1}|\mathscr{F}_n)\\&=X_n+b_{n+1}EY_{n+1}=X_n.\end{aligned}$$

定理 1.2.7 假设$\{X_n,\mathscr{F}_n,n=0,1,\cdots\}$和$\{Y_n,\mathscr{F}_n,n=0,1,\cdots\}$是鞅(或下鞅),则

(1) $\{X_n+Y_n,\mathscr{F}_n,n=0,1,\cdots\}$是鞅(或下鞅).

(2) $\{X_n\vee Y_n,\mathscr{F}_n,n=0,1,\cdots\}$是下鞅.

证明 (1)X_n+Y_n 的适应性和可积性是显然的,只需验证定义 1.2.5 中的条件(3),对任意非负整数 $m<n$,

$$\begin{aligned}E(X_n+Y_n|\mathscr{F}_m)&=E(X_n|\mathscr{F}_m)+E(Y_n|\mathscr{F}_m)\\&=X_m+Y_m\quad(\text{相应地}\geqslant X_m+Y_m).\end{aligned}$$

(2)因为 $X_n \vee Y_n \geqslant X_n, X_n \vee Y_n \geqslant Y_n$,故有

$$E(X_n \vee Y_n | \mathscr{F}_m) \geqslant E(X_n | \mathscr{F}_m) \vee E(Y_n | \mathscr{F}_m) \geqslant X_m \vee Y_m.$$

由定理 1.2.7(2)知,如果$\{X_n, \mathscr{F}_n, n=0,1,\cdots\}$和$\{Y_n, \mathscr{F}_n, n=0,1,\cdots\}$是上鞅,则$\{X_n \wedge Y_n, \mathscr{F}_n, n=0,1,\cdots\}$也是上鞅.

定理 1.2.8 (1)假设$\{X_n, \mathscr{F}_n, n=0,1,\cdots\}$是鞅,$f$ 是定义在R_1 上的凸函数.如果,$\forall n \geqslant 0; E|f(X_n)| < \infty$,则$\{f(X_n), \mathscr{F}_n, n=0,1,\cdots\}$是下鞅.

(2)假设$\{X_n, \mathscr{F}_n, n=0,1,\cdots\}$是下鞅(上鞅),$f$ 是定义在R_1 上的非降凸函数.如果,$\forall n \geqslant 0, E|f(X_n)| < \infty$,则$\{f(X_n),{}_n, \mathscr{F}_n, n=0,1,\cdots\}$是下鞅(上鞅).

(3)假设$\{X_n, \mathscr{F}_n, n=0,1,\cdots\}$是上鞅,$f$ 是定义在R_1 上的非降凹函数.如果,$\forall n \geqslant 0, E|f(X_n)| < \infty$,则$\{f(X_n), \mathscr{F}_n, n=0,1,\cdots\}$是上鞅.

证明 我们只证明(2),其余的证明是类似的.对任意非负整数 $m<n$,因为函数 f 是非降的,而且 $E(X_n|\mathscr{F}_m) \geqslant X_m$,所以 $f(E(X_n|\mathscr{F}_m)) \geqslant f(X_m)$,由 Jensen 不等式有

$$E(f(X_n)|\mathscr{F}_m) \geqslant f(E(X_n|\mathscr{F}_m)) \geqslant f(X_m).$$

推论 1.2.9 假设$\{X_n, \mathscr{F}_n, n=0,1,\cdots\}$是鞅(或非负下鞅),$\lambda \geqslant 1$ 为一常数.如果对 $n \geqslant 0$, $|X_n|^\lambda$ 是可积的.则$\{|X_n|^\lambda, \mathscr{F}_n, n \geqslant 0\}$是下鞅.

推论 1.2.10 假设$\{X_n, \mathscr{F}_n, n=0,1,\cdots\}$是下鞅,则$\{X_n^+, \mathscr{F}_n, n=0,1,\cdots\}$也是下鞅.

1.2.3 离散鞅的基本不等式

设$\{X_n, \mathscr{F}_n, n=0,1,\cdots\}$是鞅(或下鞅),则对于 $n \geqslant m$ 有$E(X_n|\mathscr{F}_m) = X_m$(或$\geqslant X_m$),如果将 n, m 换成停时后,上面式子是否仍然成立?下面的定理指出,当停时有界时,回答是肯定的.

定理 1.2.11 (Doob 有界停时定理) 设$\{X_n, \mathscr{F}_n, n=0,1,\cdots\}$是下鞅,$\sigma \leqslant \tau$ 是一对有界停时,则有

(1) $E|X_\sigma| < \infty, E|X_\tau| < \infty$.

(2) $E(X_\tau|\mathscr{F}_\sigma) \geqslant X_\sigma$. (1.2.2)

如果$\{X_n, \mathscr{F}_n, n=0,1,\cdots\}$是鞅,则(1.2.2)中不等号改为等号.

证明 设 $\tau \leqslant N$,则有 $|X_\tau| \leqslant \sum_{i=0}^{N} |X_i| \in L^1$,同理$|X_\sigma| \in L^1$.设 $A \in \mathscr{F}_\sigma$,则有

$$A(\sigma = j)(\tau > j) \in \mathscr{F}_j. \tag{1.2.3}$$

如果 $\tau - \sigma \leqslant 1$,那么由下鞅性及(1.2.3),对 $A \in \mathscr{F}_\sigma$ 有