现代数学基础丛书

概率论基础和随机过程

王寿仁 著

科学出版社

现代数学基础丛书

概率论基础和随机过程

王寿仁　著

科学出版社

1997

内 容 简 介

本书在测度论的基础上，叙述了随机过程的理论． 本书理论性较强，叙述亦较严谨，主要内容包括一般理论(过程的可分性、可测性和样本连续性)；独立增量过程；鞅论；Brown 运动和随机微分方程．

本书可供大学高年级学生和有关研究工作者参考．

现代数学基础丛书

概率论基础和随机过程

王寿仁　著

责任编辑　苏芳霞

科学出版社出版

北京东黄城根北街16号

邮政编码：100717

北京朝阳科地亚印刷厂印刷

新华书店北京发行所发行　各地新华书店经售

*

1986年6月第一版　开本：850×1168　1/32

1997年6月第二次印刷　印张：9 7/8

印数：7501—10520　字数：259,000

ISBN7-03-005996-4/O・931

定　价：20.00元

前　　言

我于 1963—1964 年为中国科学技术大学概率论专门化的五年级学生讲授概率论，又于 1979 年为中国科学技术大学研究生院讲授概率论，这本书就是以这两次讲稿的内容为基础写成的.

概率论的内容非常丰富，要想在一个课程里把概率论的主要分支全都系统地加以介绍，那是不可能的. 因此就需要有所选择，在选择时，难免要受到讲授者的偏爱而有所取舍. 我在设计这个课程时，除了讲授概率论和随机过程的基础部分，如条件概率和随机分析外，着重介绍对今后的工作带普遍性而有影响的内容，所以在随机过程方面，我选了三个过程作为主要的内容，这就是独立增量过程、鞅和 Brown 运动. 目前关于这三个过程，国内也缺乏系统介绍的书，所以本书恰是一个补充.

独立增量过程和 Brown 运动是随机过程里的经典分支，独立增量过程的研究已较完善，但高维 Brown 运动的研究还在发展. 许多随机分析的基本概念和问题，都起源于对独立增量过程的研究，而且这两个具体过程的深刻结果对马氏过程和鞅的研究提供了背景材料，从而它们具有启发作用. 鞅论在六十年代得到了划时代的发展，直到现在仍是概率论里很活跃的一个分支，通过对它的研究，已经发展起来所谓一般过程论. 同时它为其它分支，例如滤波、随机控制、随机微分方程及点过程等也提供了新的研究工具. 所以我认为这三种随机过程是从事概率论及其应用的学者必不可少的基本知识.

由于水平有限，缺点和错误在所难免，希读者指正.

目　　录

第一章 基本概念

§1. 引　言

概率论是研究随机现象的规律的数学理论，由于在它的发展中总是结合着具体的随机现象来讨论问题，所以这一数学分支里的术语都带有形象性，直到现在我们仍然沿用这些形象性的术语，因为这些术语的直观含义对概率论的发展还有启示作用，而且不断地引出新概念、新方法及研究方向．概率论在整个数学领域里占有一定的地位，与其它分支有着密切的联系，对整个数学的发展也起着影响．现在欲掌握概率论的本质，必须采用测度论的观点．本节目的在于给出用测度论观点的直观背景．

所谓随机现象就是当我们做实验或观察这种现象时，其结果是许多可能结果中的一个，而不能在实验或观察前完全预言．例如掷一个骰子，我们不能说掷下去准得什么点．用数学研究客观规律时，首先要把客观事物理想化，在概率论里，第一个理想化是关于实验或观察的可能结果应由哪些东西构成的问题．我们把实验或观察的结果叫做事件．事件分为简单事件和复合事件，以掷一个骰子为例，“掷得结果为⚀”是一个简单事件，“掷得为⚁”，…“掷得为⚅”都是简单事件．又如“掷得为双”，这就是或者“为⚁”或者“为⚃”或者“为⚅”．又如“掷得结果少于四点”，这就是或者“为⚀”或者“为⚁”或者“为⚂”．由此可见“掷得为双”．“掷得结果少于四点”都可以分解为简单事件，这样的事件叫做复合事件．简单事件代表可以想象的结果，不可能再分解，所有这些简单事件构成理想化实验的基本的可能结果．我们把一个实验的所有简单事件的全体叫做基本空间，记之为 Ω，每个简单事件叫做点．于是牵涉到实验的一切事件都可用点来表达，复合事件可以分解为点．

例如“掷得结果小于四点”这一事件（用A表示），则$A = \{⚀, ⚁, ⚂\}$是由三个点构成的．在数学里我们称A为基本空间Ω里的一个子集合，又如令事件B表示“掷得为单”，则$B = \{⚀, ⚂, ⚄\}$，B也是Ω里的子集合．由此可见事件无非就是基本空间Ω里的点构成的集合．为了在理论上合理起见，我们需要加上空集这个概念，它就是不包含任何点的集合，我们记之为$\varnothing$，在实验中$\varnothing$相应于“不可能事件”，总之我们一旦谈论概率时，首先必须结合着所给定的基本空间来讨论，我们从基本空间和它的点的概念出发，这些是原始的无定义的概念，正如欧氏几何的公理化中点和直线是无定义的概念一样．事件就是由点构成的集合，我们说事件A由某些点构成，就是说这些点表示在实验里能使事件A发生的那些结果．今后用$A, B, \cdots$等大写拉丁字母表示事件．“事件A, B为不相容或互斥”、“事件A是事件B的反事件”、“事件A, B同时发生”、“事件A, B至少一个发生”以及事件A发生就引起事件B发生，这些直观说法分别相应于集合的如下关系及运算：“集A, B无公共点或不相交”、“A, B互为余集”、“A, B的交”、“A, B的并”以及“A是B的子集”．交及并的运算对可数个集合也有重要的意义，并且有直观的背景．例如我们做这样的实验，一次接一次地重复掷一个骰子，这时基本空间Ω的点（也可以叫元素）是由⚀⚁⚂⚃⚄⚅这六个符号组成的无穷序列．在这个实验里可以谈论这样的事件A_1：“直到头一次出现⚀”我们令A_n表示事件“直到第n次才出现⚀”，A_n包含点$(X_1, \cdots, X_{n-1}, ⚀, Y_{n+1}, Y_{n+2}, \cdots)$其中每个$X_1, \cdots, X_{n-1}$都不是⚀，每个$Y_{n+1}, Y_{n+2}, \cdots$可以是那六个符号里的任何一个，$A_n$是由无穷个点构成的集，而$A = \sum_{n=1}^{\infty} A_n$，这表示事件$A$是事件$(A_i, i = 1, 2, \cdots)$的可数并．要求事件经过可数个集合运算所得到的集合仍为事件，反映在集合论里，就是要求Ω里的某些集构成的族对可列交及并都为封闭，这就是说Ω里所有事件的总体构成σ域，通常我们把事件σ域记为$\mathscr{F}$．

事件的概率是描述这个事件出现的可能性的大小，对每个事件 A 要赋予一个数值 $P(A)$ 以表示它出现的可能性. 事件的概率就是 σ 域 $\mathscr{F}$ 上的集函数 $P(\cdot)$，此集函数反映事件的概率的客观规律. 例如在掷一次骰子的试验里，每个符号出现的概率应为 1/6，而集 {⚀⚁⚂} 的概率为1/2等等. 与概率的直观常识相对照，容易看出 $\mathscr{F}$ 上的集函数 $P(\cdot)$ 是一个测度，且有就范性：$P(\Omega)=1$.

直观看，随机变量就是随着机会的不同而取不同值的变量，即是依赖机会的一个函数，也就是随着随机试验的结果而变的函数，它的值域为 $\boldsymbol{R}=(-\infty,\infty)$，定义域为 Ω. 这样说还不完全，因为谈到随机变量 $X(\omega)$ 时，还要知道这个随机变量取某个或某些值的概率，例如需要知道 $\{X<a\}$ 的概率是多少，其中 $a\in\boldsymbol{R}$. 回答这个问题首先需知 $\{X<a\}$ 是一个事件，这样才能谈到它的概率，这一要求反映在实函数论里就是要求随机变量 X 是可测空间 $(\Omega,\mathscr{F})$ 到 $(\boldsymbol{R},\mathscr{B}(\boldsymbol{R}))$ 的可测映象，此处 $\mathscr{B}(\boldsymbol{R})$ 是 $\boldsymbol{R}$ 中的 Borel 域.

现在让我们再以掷一次骰子为例. 令 $X(\omega)=$"ω 的点数"，所以 $X(\omega)$ 的可能值为 $1,2,\cdots,6$ 中的一个，$\{X=1\}$ 的概率为 1/6，$\{X\in(1,2,3)\}$ 或 $\{X\leqslant 3\}$ 的概率为 1/2. X 的每个可能值乘以相应的概率的总和：$1\cdot\frac{1}{6}+2\cdot\frac{1}{6}+\cdots+6\cdot\frac{1}{6}=\frac{1+2+\cdots+6}{6}$ 叫做随机变量 X 的均值，这个数在概率论里很重要. 如果骰子不是匀称的，即如

$$P(X=i)=p_i,\quad i=1,2,\cdots,6$$

其中 $p_i\geqslant 0$, $p_1,\cdots,p_6$ 不全相等，$\sum_{i=1}^{6}p_i=1$. 这时 X 的均值等于 $\sum_{k=1}^{6}kp_k$. 把均值概念推广于一般的随机变量上，就是通常的测度空间上的积分概念.

经过这样的初步讨论，可以看出概率与测度之间的对应关系，可以把对应关系列表如下:

概率空间	就范的测度空间 $(\Omega, \mathscr{F}, P)$
简单事件	Ω中的点ω
事件A	可测集A，即$A \in \mathscr{F}$
不可能事件	空集$\varnothing$
概率为1 (a.s.)	几乎处处 (a.e.)
随机变量	$(\Omega, \mathscr{F})$到$(\boldsymbol{R}, \mathscr{B}(\boldsymbol{R}))$上的可测映象
均值(或期望)	可测函数对P的积分

易知随机变量的 Borel 函数 (有穷值的)仍是随机变量. 若随机变量X使得$\int_\Omega X^+(\omega)P(d\omega)$及$\int_\Omega X^-(\omega)P(d\omega)$中至少一个为有穷，其中 $X^+(\omega)=\max(X(\omega), 0)\triangleq X(\omega)\vee 0$，$X^-(\omega)=\max(-X(\omega), 0)=-X(\omega)\vee 0=-(X(\omega)\wedge 0)$，则称随机变量$X$的积分存在. 当随机变量$X$的积分存在时，我们把$X$的积分叫做$X$的期望，并记之为

$$EX=\int_\Omega X(\omega)P(d\omega).$$

令X为随机变量，对$\boldsymbol{R}$中 Borel 集S，定义

$$P_X(S)=P\{X\in S\}, \quad S\in\mathscr{B}(\boldsymbol{R}).$$

这个定义于$\mathscr{B}(\boldsymbol{R})$上的集函数叫做随机变量$X$的分布. 因为$X$是有穷值的，故 $X^{-1}(\boldsymbol{R})=\Omega$，又因在$X$的映象下，Borel 集的和的逆象为其各加项的逆象的和，故有

$$P_X(\boldsymbol{R})=1,$$

$$P_X(\Sigma S_i)=\Sigma P_X(S_i), \quad S_i\in\mathscr{B}(\boldsymbol{R}), \quad (S_i) \text{ 不相交}.$$

由此可见P_X是$\mathscr{B}(\boldsymbol{R})$上的概率测度. 这样一来，随机变量$X$在它的值域空间$\boldsymbol{R}$上诱导出一个新的概率空间$(\boldsymbol{R}, \mathscr{B}(\boldsymbol{R}), P_X)$，这个概率空间叫做随机变量$X$的样本概率空间.

X的分布函数$F_X(x)$是由下列关系定出:

$$F_X(x)=P_X(-\infty, x]=P\{X\leqslant x\}, \quad x\in\boldsymbol{R}.$$

显见 $F_X(x)$ 为非降、右连续函数，且

$$F_X(-\infty) \equiv \lim_{x\to-\infty} F(x) = 0,$$

$$F_X(+\infty) \equiv \lim_{x\to\infty} F(x) = 1.$$

定理 1.1.1 设 $F(x)$ $(x\in \boldsymbol{R})$ 为非降右连续函数，且 $F(-\infty)=0$, $F(+\infty)=1$，于是它必是某一概率空间上某个随机变量的分布函数.

证 考虑可测空间 $(\boldsymbol{R}, \mathscr{B}(\boldsymbol{R}))$，令

$$P(a,b] = F(b) - F(a), \quad 其中\ a, b\in \boldsymbol{R},\ a\leqslant b.$$

由测度拓张定理知 P 可以拓张成为 $\mathscr{B}(\boldsymbol{R})$ 上的概率测度.这样便得基本概率空间 $(\boldsymbol{R}, \mathscr{B}(\boldsymbol{R}), P)$. 定义随机变量 $X(x)=x$, $x\in \boldsymbol{R}$，于是 $F(x)$ 就是 $X(x)$ 的分布函数. 证毕

注 不只有一个这样的概率空间，例如还可以这样取基本概率空间: $\Omega=[0,1]$, $\mathscr{F}=\mathscr{B}[0,1]$($[0,1]$ 里的 Borel 集的全体)，P 为 $\mathscr{F}$ 上的 Lebesgue 测度 λ，取 F 的逆 F^{-1}:

$$F^{-1}(\omega) = \inf\{s: F(s) > \omega\}, \quad \omega\in[0,1].$$

易知 $F^{-1}(\omega)$, $\omega\in[0,1]$ 是单调增的. 每个 ω 为 $F^{-1}(\omega)$ 右连续点，否则存在 h，使得下式成立

$F^{-1}(\omega) < h < F^{-1}(\omega+) \leqslant F^{-1}(\omega+\varepsilon)$，对任意 $\varepsilon>0$. 这就是说 $F(h)>\omega$，且 $F(h)\leqslant \omega+\varepsilon$，对任意 $\varepsilon>0$，这是不可能的，故 $F^{-1}(\omega)$ 为右连续增函数. 对每个 s 往证

$$F(s) = \inf\{\omega\in[0,1],\ F^{-1}(\omega) > s\}. \tag{1}$$

事实上，对每个 $s\in \boldsymbol{R}$,

$$F^{-1}(F(s)) \geqslant s,$$

从而对任意 $\varepsilon>0$, $F^{-1}(F(s+\varepsilon)) \geqslant s+\varepsilon > s$，由此可知

$$F(s+\varepsilon) \geqslant \inf\{\omega: F^{-1}(\omega) > s\}.$$

但 F 为右连续，由上式得

$$F(s) \geqslant \inf\{\omega: F^{-1}(\omega) > s\}. \tag{2}$$

设 ω 使得 $F^{-1}(\omega)>s$，由 F^{-1} 的定义知有 $F(s)\leqslant\omega$，从而

$$F(s) \leqslant \inf\{\omega: F^{-1}(\omega) > s\}. \tag{3}$$

故由 (2) 及 (3) 得证 (1).

由 (1) 知

$$\lambda\{\omega: F^{-1}(\omega) \leqslant y\} = \inf\{\omega: F^{-1}(\omega) > y\} = F(y).$$

由此可知 $\{[0, 1], \mathscr{B}[0, 1], \lambda\}$ 上的随机变量 $F^{-1}(\omega)$ 的分布函数为 F.

定理 1.1.2 令 $g(\cdot)$ 为 $\boldsymbol{R}$ 到 $\boldsymbol{R}$ 的 Borel 函数, X 为基本概率空间 $(\Omega, \mathscr{F}, P)$ 上的随机变量 (今后提到随机变量就是指在给定的基本概率空间 $(\Omega, \mathscr{F}, P)$ 上的随机变量), 于是随机变量 $g(X)$ 的分布由随机变量 X 的分布 P_X 所决定, 若等式一方存在, 则有

$$Eg(X) = \int_{\boldsymbol{R}} g(x)P_X(dx) = \int_{\boldsymbol{R}} g(x)dF_X(x). \tag{4}$$

证 由前知 $g(X)$ 是随机变量. 由定义有

$$\{g(X) \in S\} = \{X \in g^{-1}(S)\}, \quad S \in \mathscr{B}(\boldsymbol{R}),$$

但 $g^{-1}(S)$, 是 Borel 集, 故

$$P_{g(X)}(S) = P_X(g^{-1}(S)), \quad S \in \mathscr{B}(\boldsymbol{R}).$$

由此知定理的第一部分成立. 欲证第二部分, 只需对 $g \geqslant 0$ 的特殊情形加以证明, 又由单调收敛定理知, 只需对 g 为非负简单函数的情形加以证明即可. 又由积分的可加性知, 只需对 $g(x) = I_S(x)$ 的情形加以证明即可, 此处 $I_S(x)$ 为 S 的示性函数,

$$I_S(x) = \begin{cases} 1, & x \in S, \\ 0, & x \notin S. \end{cases}$$

这时

$$g(X(\omega)) = I_S(X(\omega)) = I_{\{X \in S\}}(\omega),$$

于是 (4) 的右方等于

$$\int_{\boldsymbol{R}} g(x)P_X(dx) = \int_{\boldsymbol{R}} I_S(x)P_X(dx) = P_X(S),$$

而 (4) 的左方等于

$$Eg(X) = \int_{\Omega} I_{\{X \in S\}}(\omega)P(d\omega) = P\{X \in S\}.$$

由分布 P_X 的定义知上二式相等, 故 (4) 得证. 证毕

由定理 1.1.1 及 1.1.2 可知, 我们在基本概率空间 $(\Omega, \mathscr{F}, P)$

上考虑随机变量 X 的分布问题时，可以代之考虑样本空间 $(\boldsymbol{R}, \mathscr{B}(\boldsymbol{R}), P_X)$ 上定义的随机变量 $\tilde{X}(x)=x$ 的分布问题. 这样一来，当 $A \in \mathscr{B}(\boldsymbol{R})$ 时

$$P\{X \in A\} = P_X(A) = P_X(x \in A) = P_X\{\tilde{X} \in A\}.$$

如果记 $\sigma(X) \triangleq X^{-1}(\mathscr{B}(\boldsymbol{R})) = \{X^{-1}(B),\ B \in \mathscr{B}(\boldsymbol{R})\}$ (这就是使 X 为可测的最小子 σ 域，也叫做由 X 所产生的子 σ 域)，那么 $X(\omega)$ 是 $(\Omega, \sigma(X), P)$ 到 $(\boldsymbol{R}, \mathscr{B}(\boldsymbol{R}), P_X)$ 的可测映象，使得 $(\Omega, \sigma(X), P)$ 上的随机变量 X 变到 $(\boldsymbol{R}, \mathscr{B}(\boldsymbol{R}), P_X)$ 上的随机变量 $\tilde{X}$，且对任意 Borel 函数 g，在此映象下，随机变量 $g(X)$ 变为 $(\boldsymbol{R}, \mathscr{B}(\boldsymbol{R}), P_X)$ 上的随机变量 $g(X)$，且有

$$\int_\Omega g(X(\omega))P(d\omega) = \int_{\boldsymbol{R}} g(x)P_X(dx),$$

假若上式一方存在.

§2. 矩及常用不等式

随机变量的矩在概率论的研究中起着重要的作用. 在简单的 Марков 及 Чебышев 不等式里，在特征函数的展开里都出现矩. 依 r 阶矩收敛也是概率论里的重要概念. 研究独立随机变量和时，特别是采用截尾法研究独立随机变量和时，就是以矩作为基础的.

定义 1.2.1 令 X 为随机变量，对任意 $r > 0$，X^r 的期望 EX^r 叫做 X 的 r 阶矩；如果 EX^r 存在，则 $|X|^r$ 的期望 $E|X|^r$ (永远存在) 叫做 X 的 r 阶绝对矩.

定义 1.2.2 若 $E|X|^r < \infty$，则说 X 属于 $L_r(r > 0)$，此处 L_r 是由所有的具有有穷 r 阶绝对矩的随机变量组成的空间.

关于 L_r 空间的一般性质，例如它是线性完备距离空间等等性质，可在一般的实变函数论书中找到，这里不再重复.

下面我们只给出一些常用的不等式和简单性质:

(1) 由 $|X|^{r'} \leqslant 1 + |X|^r$，$0 < r' \leqslant r$ 立刻推知若 $E|X|^r < \infty$，则 $E|X|^{r'} < \infty$，对一切 $0 < r' \leqslant r$，因而 $E|X|^{r'}$ 存在

且有穷.

(2) 由

$$|a+b|^r \leqslant c_r|a|^r + c_r|b|^r,\ r>0,$$

此处

$$c_r = \begin{cases} 1, & r \leqslant 1, \\ 2^{r-1}, & r \geqslant 1, \end{cases}$$

立刻得到 c_r 不等式

$$E|X+Y|^r \leqslant c_r E|X|^r + c_r|Y|^r,\ r>0,$$

其中 X, Y 都是随机变量.

(3) 在下列不等式中

$$|ab| \leqslant \frac{1}{r}|a|^r + \frac{1}{s}|b|^s,\ \ r>1,\ \frac{1}{r}+\frac{1}{s}=1,$$

取 $a = X/E^{\frac{1}{r}}|X|^r$, $b = Y/E^{\frac{1}{s}}|Y|^s$ 得到 Hölder 不等式

$$E|XY| \leqslant E^{\frac{1}{r}}|X|^r \cdot E^{\frac{1}{s}}|Y|^s,\ r>1,\ \frac{1}{r}+\frac{1}{s}=1.$$

Hölder 不等式的特款为 Schwarz 不等式

$$E|XY| \leqslant \sqrt{E|X|^2 \cdot E|Y|^2}.$$

(4) Minkowski 不等式

$$E^{\frac{1}{r}}|X+Y|^r \leqslant E^{\frac{1}{r}}|X|^r + E^{\frac{1}{r}}|Y|^r,\ r\geqslant 1.$$

事实上,当 $r=1$ 时,此不等式显然成立. 当 $r>1$ 时,由 Hölder 不等式知

$$\begin{aligned} E|X+Y|^r &\leqslant E|X||X+Y|^{r-1} + E|Y||X+Y|^{r-1} \\ &\leqslant E^{\frac{1}{r}}|X|^r E^{\frac{1}{s}}|X+Y|^{(r-1)s} + E^{\frac{1}{r}}|Y|^r E^{\frac{1}{s}}|X \\ &\qquad + Y|^{(r-1)s} \\ &\leqslant (E^{\frac{1}{r}}|X|^r + E^{\frac{1}{r}}|Y|^r)E^{\frac{1}{s}}|X+Y|^r. \end{aligned}$$

把此不等式两边除以 $E^{\frac{1}{s}}|X+Y|^r$ 即得 Minkowski 不等式.

(5) $\log E|X|^r$, $r\geqslant 0$ 是 r 的凸函数

欲证此,取 $r'<r$, 在 Schwarz 不等式里把 X 代以 $|X|^{(r-r')/2}$, Y

代以 $|X|^{(r+r')/2}$ 得

$$E^2|X|^r \leqslant E|X|^{r-r'} \cdot E|X|^{r+r'},$$

取对数，得

$$\log E|X|^r \leqslant \frac{1}{2}(\log E|X|^{r-r'} + \log E|X|^{r+r'}).$$

由此知 $\log E|X|^r$ 是 r 的凸函数.

(6) $E^{\frac{1}{r}}|X|^r$ 是 r 的非降函数

欲证此，在 (r, y) 平面上考虑 $y = \log E|X|^r$ 的图象，此图象经过 $(0, 0)$ 点. 由于 $\log E|X|^r$ 是 r 的凸函数，故当 $r \geqslant 0$ 时，由 $(0, 0)$ 到 $(r, \log E|X|^r)$ 的直线的斜率 $\frac{1}{r}\log E|X|^r$ 是 r 的非降函数. 所以 $E^{\frac{1}{r}}|X|^r$ 也是 r 的非降函数.

(7) 若 f 为 $\boldsymbol{R}^n$ 里凸区域 D 上的连续凸函数，$X_1, \cdots, X_n$ 为可积随机变量，且 $(X_1, \cdots, X_n) \in D$, a.s. 则

$$E[f(X_1, \cdots, X_n)] \leqslant f(EX_1, \cdots, EX_n).$$

往证此不等式. 由 $(X_1, \cdots, X_n) \in D$, a.s. 不难看出 $(EX_1, \cdots, EX_n) \in D$. 取 D 里任意一点 $(x_1^0, \cdots, x_n^0)$，于是经过 $\boldsymbol{R}^{n+1}$ 里点 $(x_1^0, \cdots, x_n^0, f(x_1^0, \cdots, x_n^0))$ 的超平面必在曲面 f 之上，设 $(\lambda_1(x^0), \cdots, \lambda_n(x^0))$ 为这个超平面的方向余弦，其中 $x^0 = (x_1^0, \cdots, x_n^0)$，则对 D 里任意点 $(x_1, \cdots, x_n)$ 有

$$f(x_1, \cdots, x_n) \leqslant f(x_1^0, \cdots, x_n^0) + \sum_1^n \lambda_i(x^0)(x_i - x_i^0).$$

特别，取 $(x_1^0, \cdots x_n^0) = (EX_1, \cdots, EX_n)$, $(x_1, \cdots, x_n) = (X_1 \cdots, X_n)$ 代入上式，右方变为 $f(EX_1, \cdots, EX_n) + \sum_1^n \lambda_i(EX)(X_i - EX_i)$，这是可积随机变量，其积分为 $f(EX_1, \cdots, EX_n)$，从而左方 $f(X_1, \cdots, X_n)$ 的积分存在，故得

$$Ef(X_1, \cdots, X_n) \leqslant f(EX_1, \cdots, EX_n).$$

特别，对任意可积正随机变量 X, Y 有

$$E(X^\alpha Y^{1-\alpha}) \leqslant E^\alpha X E^{1-\alpha} Y, \quad 0 < \alpha < 1,$$

$$E(X^{\frac{1}{p}}+Y^{\frac{1}{p}})^p \leqslant (E^{\frac{1}{p}}X+E^{\frac{1}{p}}Y)^p;\quad 1\leqslant p<\infty.$$

由这两个不等式容易推出 Hölder 及 Minkowski 不等式。

如果令 $\|X\|_p \triangleq E^{\frac{1}{p}}|X|^p$，$1\leqslant p<\infty$，

$$\|X\|_\infty \triangleq \sup\{x: P(|X|>x)>0\},$$

则 $0\leqslant \|X\|_1 \leqslant \|X\|_p \leqslant \|X\|_q \leqslant \|X\|_\infty \leqslant \infty, (1<p<q<\infty)$,

$$P(|X|\neq 0)=0 \Longleftrightarrow \|X\|_p=0, (1\leqslant p\leqslant\infty).$$

此外，对任意 $c<\|X\|_\infty$ 及 $1\leqslant p<\infty$ 有

$$\|X\|_\infty \geqslant E^{\frac{1}{p}}|X|^p \geqslant E^{\frac{1}{p}}(|X|^p I_{(|X|\geqslant c)})$$

$$\geqslant c[P(|X|\geqslant c)]^{\frac{1}{p}}$$

令 $p\to\infty$，再令 $c\uparrow\|X\|_\infty$，由上不等式得

$$\lim_{p\to\infty}\|X\|_p=\|X\|_\infty.$$

(8) 矩不等式

令 g 为 $\boldsymbol{R}$ 上非负 Borel 函数，X 为任意随机变量．若 g 在 $[0,\infty)$ 上非降，且为偶函数，则对每个 $a\geqslant 0$ 有

$$\frac{Eg(X)-g(a)}{\|g(X)\|_\infty} \leqslant P\{|X|\geqslant a\} \leqslant \frac{Eg(X)}{g(a)}.$$

若 g 只在 $\boldsymbol{R}$ 上非降，则上不等式里中间那一项可以换为 $P\{X\geqslant a\}$，$a\in\boldsymbol{R}$.

欲证此不等式．当 g 为偶函数且在 $[0,\infty)$ 上非降时，我们有

$$Eg(X)=\int_{|X|\geqslant a} g(X(\omega))P(d\omega)+\int_{|X|<a} g(X(\omega))P(d\omega),$$

但

$$g(a)P\{|X|\geqslant a\} \leqslant \int_{|X|\geqslant a} g(X(\omega))P(d\omega)$$

$$\leqslant \|g(X)\|_\infty P\{|X|\geqslant a\},$$

$$0\leqslant \int_{|X|<a} g(X(\omega))P(d\omega) \leqslant g(a),$$

故

$$g(a)P\{|X|\geqslant a\} \leqslant Eg(X) \leqslant \|g(X)\|_\infty P\{|X|\geqslant a\}+g(a).$$