国防特色院士文库

随机无穷维动力系统

SUIJI WUQIONGWEI DONGLI XITONG

郭柏灵　蒲学科　著

北京航空航天大学出版社
北京理工大学出版社　哈尔滨工业大学出版社
哈尔滨工程大学出版社　西北工业大学出版社

国防特色院士文库

随机无穷维动力系统

郭柏灵　蒲学科　著

北京航空航天大学出版社

北京理工大学出版社　哈尔滨工业大学出版社
哈尔滨工程大学出版社　西北工业大学出版社

内容简介

本书共分 10 章，主要内容涉及几类重要的随机偏微分方程及其随机动力系统。前 3 章着重介绍概率论以及随机过程中的一些预备知识，包括 Itô 随机积分理论；从第 4 章开始，主要讨论由布朗运动以及 Lévy 过程驱动的随机非线性偏微分方程。本书详细介绍了这些随机偏微分方程的解的存在性理论及其长时间行为，如随机整体吸引子及其 Hausdorff 维数估计等理论，涵盖了这些方程的一些前沿结果以及作者研究的最新成果。

本书可供大学数学专业、应用数学专业和计算数学专业的高年级学生、研究生、教师以及相关的科技工作者阅读参考。

图书在版编目(CIP)数据

随机无穷维动力系统/郭柏灵，蒲学科著. —北京：北京航空航天大学出版社，2009.11

ISBN 978-7-81124-909-5

Ⅰ.随… Ⅱ.①郭…②蒲… Ⅲ.无限维一动力系统(数学) Ⅳ.O19

中国版本图书馆 CIP 数据核字（2009）第 161479 号

随机无穷维动力系统

郭柏灵　蒲学科　著

责任编辑　刘晓明

*

北京航空航天大学出版社出版发行

北京市海淀区学院路 37 号(100191)　发行部电话 010-82317024　传真 010-82328026

http://www.buaapress.com.cn　E-mail:bhpress@263.net

涿州市新华印刷有限公司印装　各地书店经销

*

开本:787×1 092　1/16　印张:18.25　字数:467 千字

2009 年 11 月第 1 版　2009 年 11 月第 1 次印刷　印数:3 000 册

ISBN 978-7-81124-909-5　定价:78.00 元

前　言

近10年来，随机非线性偏微分方程及其动力系统问题大量出现于物理、力学、金融、生物等相关领域，例如大气海洋中的环流问题、非线性波在随机介质中的传播问题、风险资产、股票等价格的波动规律等均有相应的随机偏微分方程描述。早在20世纪70年代，Bensoussan，Temam[1]，Pardoux[2,3]等不少数学家就对随机非线性偏微分方程进行了研究。随机无穷维动力系统的研究晚了一些，1994—1997年，数学家Crauel，Flandoli以及Debussche[4,5]等建立了有关随机无穷维动力系统理论的基本框架并研究了其在某些随机非线性发展方程中的应用，例如，建立了随机整体吸引子的存在性及其Hausdorff维数估计以及随机不变测度理论等。特别是最近10多年来，随机非线性偏微分方程及其动力系统以及它的数值计算的研究得到了蓬勃的发展，不少数学家，如G. Da Prato，J. Zabczyk，H. Crauel，F. Flandoli，A. de Bouard以及A. Debussche等均得出了一系列很有价值的研究结果，其中Prato和Zabczyk还出版了一些很好的专著[6,7]。

作者及其合作者从2005年起开始收集、学习有关随机过程(其中包括Lévy过程及分数阶Wiener过程)和随机非线性偏微分方程及其动力系统的著作和文献，并在讨论班上作了报告，同时和国外学者也进行了学术交流和讨论；结合我国大气、海洋问题以及非线性波在随机介质中的传播问题等进行了初步的研究，得出了一些和实际物理问题有关的理论结果。

写这本书的目的，一方面是总结我们这几年学习的心得体会和一些研究结果；另一方面也是更重要的是介绍目前国际上的某些前沿进展和结果，并盼能引起我国广大偏微分方程以及数值计算研究工作者的重视和关注。作者试图以简洁的方式和通俗易懂的语言介绍有关这方面的最基本的内容，希望能使读者节省一些时间掌握这些内容，并在此基础上开展一些研究工作。必须指出，由于这一方向是有关概率论和偏微分方程的交叉学科，因此是具有一定难度的，但我们认为这是可以克服的。

作者衷心感谢陈木法院士，特别是他的优秀博士生(已工作)王健对书稿进行了认真的审阅，并提出了许多宝贵意见。由于作者水平有限，书中错误在所难免，敬请读者指正。

郭柏灵

2009年3月

目　　录

第 1 章　概率论和随机过程的一些预备知识 …… 1

1.1　概率论的预备知识 …… 1

1.1.1　概率空间 …… 1

1.1.2　随机变量及其概率分布 …… 4

1.1.3　随机变量的数字特征 …… 6

1.2　随机过程的预备知识 …… 10

1.2.1　Markov 过程 …… 13

1.2.2　遍历论的基本知识 …… 18

1.3　鞅 …… 21

1.4　Wiener 过程和布朗运动 …… 29

1.5　Poisson 过程 …… 36

1.6　Lévy 过程 …… 40

1.6.1　特征函数和无穷可分性 …… 40

1.6.2　Lévy 过程概述 …… 42

1.6.3　Lévy－Itô 分解 …… 43

1.7　分数阶布朗运动 …… 46

第 2 章　随机积分及 Itô 公式 …… 48

2.1　随机积分 …… 48

2.1.1　Itô 积分 …… 49

2.1.2　一般情形的随机积分 …… 55

2.2　Itô 公式 …… 58

2.3　无穷维情形 …… 63

2.3.1　Q－Wiener 过程及其随机积分 …… 63

2.3.2　随机积分的性质及 Itô 公式 …… 70

2.4　核算子以及 Hilbert-Schmidt 算子 …… 74

第 3 章　广义 O－U 过程与随机微分方程 …… 77

3.1　广义 O－U 过程 …… 77

3.2　线性随机微分方程 …… 82

3.3　非线性随机微分方程 …… 89

第 4 章　随机吸引子 …… 94

4.1　确定的非自治系统 …… 94

4.2 随机动力系统 …… 96
4.3 在随机发展方程中的应用 …… 99
4.3.1 具有可加噪声的 Navier - Stokes 方程 …… 100
4.3.2 白噪声驱动的 Burgers 方程 …… 104
4.3.3 随机非线性波动方程 …… 107
4.4 Ginzburg-Landau 方程及其随机动力系统 …… 112
4.4.1 随机吸引子的存在性 …… 114
4.4.2 随机吸引子的 Hausdorff 维数 …… 117
4.4.3 随机广义 Ginzburg - Landau 方程的一些结果 …… 121

第 5 章 随机非线性 Schrödinger 方程 …… 123

5.1 L^2 理论 …… 123
5.1.1 逼近方程 …… 126
5.1.2 定理的证明 …… 131
5.2 H^1 理论 …… 136
5.2.1 可加噪声情形 …… 138
5.2.2 乘积噪声情形 …… 145

第 6 章 随机 KdV 方程 …… 152

6.1 准备工作 …… 152
6.2 可加噪声情形 …… 155
6.2.1 线性方程 …… 156
6.2.2 非线性方程 …… 165
6.3 乘积噪声情形 …… 168
6.4 随机 KdV 方程的吸引子 …… 172
6.4.1 解的存在性 …… 173
6.4.2 弱紧集的存在性及主要结果 …… 175
6.5 随机 KdV - BO 方程 …… 181
6.5.1 随机 KdV - BO 方程解的存在性 …… 181
6.5.2 弱阻尼随机 KdV - BO 方程解的长时间行为 …… 194

第 7 章 Lévy 过程驱动的随机偏微分方程 …… 203

7.1 Poisson 白噪声驱动的随机抛物方程 …… 203
7.1.1 主要结论 …… 205
7.1.2 定理的证明 …… 206
7.2 Lévy 噪声驱动的随机抛物方程 …… 213
7.2.1 估 计 …… 215
7.2.2 存在性的证明 …… 221

第 8 章　大气海洋模型及其随机动力系统 ………… 223

8.1　模型的提出 ………… 223
8.2　解的存在唯一性 ………… 224
8.2.1　局部存在性 ………… 225
8.2.2　整体存在性 ………… 227
8.3　随机吸引子的存在性 ………… 229
8.3.1　问题(P_2)的解的存在唯一性以及正则性 ………… 229
8.3.2　在 $L^2(D)$中的耗散性质 ………… 232

第 9 章　随机 Landau－Lifshitz 方程 ………… 234

9.1　问题的提出与随机积分 ………… 234
9.1.1　方程的提出 ………… 234
9.1.2　Strotonovich 积分 ………… 235
9.2　SLL 方程的整体弱解 ………… 236
9.3　光滑解的整体存在性 ………… 239
9.3.1　$\varepsilon>0$ 时的局部解 ………… 239
9.3.2　先验估计与整体解 ………… 242
9.4　方程($SLL_\varepsilon-1$)和($SLL_\varepsilon-2$)的等价性 ………… 248

第 10 章　随机微分方程在金融中的应用 ………… 249

10.1　一些基本概念及其模型 ………… 249
10.2　Girsanov 定理 ………… 252
10.3　期权定价模型 ………… 255
10.3.1　欧式期权 ………… 255
10.3.2　美式期权 ………… 263
10.3.3　亚洲期权 ………… 267
10.4　一类倒向随机微分方程 ………… 268

参考文献 ………… 271

第1章　概率论和随机过程的一些预备知识

这一章是为以后各章做准备的，主要介绍概率论和随机过程中的一些预备知识，特别是Wiener过程、Poisson过程以及Lévy过程的一些基本性质。由于这章中的不少内容可参阅其他文献，因此有关内容只写出结论，不予证明；对于鲜见于其他文献而又比较重要的定理，则予以简单的证明。

1.1　概率论的预备知识

1.1.1　概率空间

在我们所处的物质世界和社会环境中，每时每刻都会遇到许多不确定性与随机性。在随机现象的研究中，需要做大量的观测和试验。所谓随机试验指的是具有某种特性的试验，通常要求试验可以在相同的条件下重复，试验的可能结果不止一个，且能事先明确地确定所有可能结果的范围，但不能准确预言哪个结果会出现。为方便起见，随机试验简称为试验，通常用E表示。随机试验E的每个可能结果称为一个基本事件或样本点，常用ω表示；而E的全体样本点构成的集合称为基本事件空间，通常用Ω表示。事件则定义为样本点的某个集合，通常用大写字母$A,B,C,\cdots$表示。当且仅当它所包含的一个样本点出现时，称某事件发生。

以大家熟悉的“掷骰子”为例。在投掷一个骰子的试验中，虽然无法预测结果如何，但无非有“出现1点”，……，“出现6点”这6个可能结果之一，从而$\Omega=\{1,2,3,4,5,6\}$，由6个元素组成，反映掷骰子试验的6个基本结果。在这个试验中，“掷出素数点”则是一个事件，它由2,3,5这3个基本事件构成。

成功地将概率论实现公理化的是前苏联数学大师柯尔莫哥洛夫(Kolmogrov)，时间是在1933年。他不仅实现了概率论的公理化，而且提出的公理为数很少且极为简单，他就是在这样的基础上建立起了概率论的宏伟大厦。

在柯氏公理体系中，与事件相对应的概念是考虑由Ω的子集构成的一个集类$\mathscr{F}$。$\mathscr{F}$不必包含Ω的一切可能的子集，但必须满足一定的条件。$\mathscr{F}$的每个成员就称为“事件”。事件有概率，其大小随事件而异，即概率实际上就是事件的函数。与此对应，在柯氏的公理化体系中，引进了一个定义在$\mathscr{F}$上的函数P。对$\mathscr{F}$中的元素A，$P(A)$对应于事件A的概率。

在实际中，往往要对事件进行各种运算，从而自然就有Ω可测子集的运算是否仍为可测的问题。为此引入σ代数的概念。

定义1.1.1　Ω为一样本空间，如果它满足

① $\Omega\in\mathscr{F}$；

② 如果$A\in\mathscr{F}$，则$A^c\xlongequal{\text{def}}\Omega-A\in\mathscr{F}$；

③ 如果$A_i\in\mathscr{F}$，则$\bigcup_{i=1}^{\infty}A_i\in\mathscr{F}$，

则称Ω的某些子集组成的集类$\mathscr{F}=\{A:A\subset\Omega\}$为$\Omega$中的一个$\sigma$-代数。

约定，如果$\mathscr{F}$由样本空间Ω中的一些可测子集组成且满足σ-代数的三条公理的集类，就称$\mathscr{F}$为事件域，并把$\mathscr{F}$中的元素称为事件。对于固定的样本空间Ω，可以构造很多σ-代数，然而并不是每一个σ-代数都是事件域。同时Ω中的事件域也不是只有一个。如$\mathscr{F}_1=\{\varnothing,\Omega\}$，$\mathscr{F}_2=\{\varnothing,A,A^c,\Omega\}$，$A\subset\Omega$，可以断定$\mathscr{F}_1$是事件域，然而却不能断定$\mathscr{F}_2$是否为事件域，因为并不知道$A$是否可测。

以后如不加说明，则总把$\mathscr{F}$看成事件域，并当且仅当元素$A\in\mathscr{F}$时，称A为事件，此时$(\Omega,\mathscr{F})$称为可测空间。下面给出概率的定义。

定义 1.1.2 设$(\Omega,\mathscr{F})$为一可测空间，P为定义于事件域$\mathscr{F}$上的实值集合函数，如果P满足

① 对每个$A\in\mathscr{F}$都有$P(A)\geqslant 0$；

② $P(\Omega)=1$；

③ $A_i\in\mathscr{F}(i=1,2,\cdots,\infty)$，且当$i\neq j$时，$A_iA_j=\varnothing$，则

$$P\left(\bigcup_{i=1}^{\infty}A_i\right)=\sum_{i=1}^{\infty}P(A_i)$$

则称P为概率测度，简称概率。

这样的三元组$(\Omega,\mathscr{F},P)$称为概率空间。以后总将$(\Omega,\mathscr{F},P)$中的Ω看成样本空间，$\mathscr{F}$为Ω中的事件域，而将P看成对应于$(\Omega,\mathscr{F})$的一个确定的概率。定义中的性质分别称为概率的非负性、规范性与完全可加性。

引入概率空间之后，就可以考察事件之间的关系以及事件的运算，并且可以讨论条件概率。如果A,B不能在同一次试验中都发生(但可以都不发生)，则称两事件A,B互斥。如果一些事件中任意两个都互斥，则称这些事件是两两互斥的。关于互斥事件的概率有如下定理。

定理 1.1.1 若干个互斥事件之和的概率，等于各事件的概率之和：

$$P(A_1+A_2+\cdots)=P(A_1)+P(A_2)+\cdots$$

这里事件的个数可以是有限的或者是无限的。

此定理称为概率的加法定理，其重要条件是各事件必须两两互斥。

所谓条件概率，就是在附加一定的条件之后的概率。从广义上讲，任何事件都是条件概率，因为人们是在一定的试验基础上去考虑事件的概率的，而试验即规定有条件。由于在概率论中，规定试验的那些基础条件被认为是已定不变的，如果不再加入其他条件或者假设，则算出的概率就叫做“无条件概率”，即通常所说的概率。当说到条件概率时，总是指另外附加的条件，通常是在“已知某事件发生了”的条件下的概率。

例如，仍考虑“掷骰子”的试验。这里，在投掷的过程中要求骰子必须是均匀的正方体，投掷高度也要达到一定的要求。这些条件是试验固有的，不作为附加条件。考虑事件A：“掷出素数点”，B：“掷出奇数点”，C：“掷出偶数点”，即

$$A=\{2,3,5\},\qquad B=\{1,3,5\},\qquad C=\{2,4,6\}$$

可以计算出A的(无条件)概率是$1/2$。现在若加上条件“已知B发生”，则可能情况只有3种：1,3,5，其中2种有利于A发生，故在此条件下，A的条件概率是$P(A\mid B)=2/3$。同样，在给定事件C发生的条件下，A的条件概率为$P(A\mid C)=1/3$。

定义 1.1.3 设$(\Omega,\mathscr{F},P)$为一概率空间，$A,B\in\mathscr{F}$为两事件，且$P(B)\neq 0$，则“在给定B发生的条件下”A的条件概率，记为$P(A\mid B)$，定义为

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

其中，$P(AB)$ 表示事件 A,B 的积事件，即 $P(AB) = P(A \cap B) = P(A,B$ 都发生)。

概率论中一个非常重要的概念是事件的独立性。考虑两事件 A,B，一般来说，事件 A 的无条件概率 $P(A)$ 与在给定 B 发生下的条件概率 $P(A \mid B)$ 是有差异的。这说明事件 A,B 之间是有一定的联系的。若 $P(A \mid B) > P(A)$，则说明 B 的发生使 A 发生的可能性增大了；反之，若 $P(A) = P(A \mid B)$，则 B 的发生与否对 A 发生的可能性毫无影响。这时在概率论上就称 A,B 两事件独立，即

$$P(AB) = P(A)P(B) \tag{1.1.2}$$

定义 1.1.4　事件 A,B 若满足式(1.1.2)，则称 A,B 独立。

由此出发，还可以考虑多个事件的独立性。

定义 1.1.5　设 $A_1,A_2,\cdots$ 为有限个或者无限个事件。如果其中任意有限个事件 A_{i_1}，$A_{i_2},\cdots,A_{i_m}$ 都成立，即

$$P(A_{i_1},A_{i_2},\cdots,A_{i_m}) = P(A_{i_1})P(A_{i_2})\cdots P(A_{i_m}) \tag{1.1.3}$$

则称事件 $A_1,A_2,\cdots$ 相互独立。

由独立性定义立即可以得到如下的概率乘法定理。

定理 1.1.2　若干个独立事件 $A_1,A_2,\cdots,A_n$ 的积的概率等于其各自概率的乘积，即

$$P(A_1A_2\cdots A_n) = P(A_1)P(A_2)\cdots P(A_n)$$

乘法定理的作用与加法定理一样，把复杂事件的概率计算归结为更简单的事件概率的计算。当然，这是要有条件的：相加是互斥，相乘是独立。

应该注意到独立事件的一部分也是相互独立的，即若 A,B,C,D 相互独立，则 A,C 或者 B,C,D 也相互独立。若事件 $A_1,A_2,\cdots$ 相互独立，那么将其任意部分改为对立事件时，所得的事件仍相互独立。

接下来介绍全概率公式和贝叶斯公式。设 $B_1,B_2,\cdots$ 为有限或无限个事件，它们两两互斥且在每次试验中至少发生一个，即

$$B_iB_j = \varnothing(\text{不可能事件}), \qquad i \neq j$$

$$B_1 + B_2 + \cdots = \Omega(\text{必然事件})$$

考虑任意事件 A。因为 Ω 是必然事件，从而 $A = A\Omega = AB_1 + AB_2 + \cdots$。因为 $B_1,B_2,\cdots$ 两两互斥，从而 $AB_1,AB_2,\cdots$ 也两两互斥。依据加法定理可知

$$P(A) = P(AB_1) + P(AB_2) + \cdots \tag{1.1.4}$$

再由条件概率的定义可知 $P(AB_i) = P(B_i)P(A \mid B_i)$，从而

$$P(A) = P(B_1)P(A \mid B_1) + P(B_2)P(A \mid B_2) + \cdots \tag{1.1.5}$$

此式就称为“全概率公式”。从式(1.1.4)和式(1.1.5)可知：全部概率 $P(A)$ 被分解为许多部分之和。还可以这样理解，把 B_i 看作导致事件 A 的一种可能途径。

对不同途径，A 发生的概率即条件概率 $P(A \mid B_i)$ 各不相同，而采取哪种途径则是随机的。直观上理解，在这种机制下，A 的综合概率 $P(A)$ 应该在最小的 $P(A \mid B_i)$ 和最大的 $P(A \mid B_i)$ 之间，而又因为各种途径被使用的机会 $P(B_i)$ 不同，从而概率 $P(A)$ 应该是各 $P(A \mid B_i)$ 的加权平均，而权正好是对应的 $P(B_i)$。

在全概率公式的假设之下，有

$$P(B \mid A) = P(AB_i)/P(A) = P(B_i)P(A \mid B_i)/\sum_j P(B_j)P(A \mid B_j) \tag{1.1.6}$$

这个公式就叫做贝叶斯公式。从形式上，这个公式只不过是条件概率与全概率公式的简单推论。它之所以著名，原因在于其具有现实以至哲理意义。先看 $P(B_i)$，它们是在没有进一步的信息（不知事件 A 是否发生）的情况下，人们对诸事件 B_i 发生可能性大小的共识。现在有了新的信息（知道 A 发生），人们对 B_i 发生可能性大小则有了新的估计。如果把事件 A 看成是结果，诸事件 $B_1,B_2,\cdots$ 看成是导致 A 的可能的原因，则可以形象地把全概率公式看作是“由原因推结果”；而贝叶斯公式则正好相反，其作用在于“由结果推原因”。事实上，这一思想现在已经发展成为一整套统计推断方法，称为“贝叶斯统计”。

1.1.2 随机变量及其概率分布

随机变量，顾名思义就是其值随机会而定的变量。严格地讲，给定概率空间$(\Omega,\mathscr{F},P)$，随机变量定义为Ω上的可测映射$X:\Omega\to R^d$。当 $d\geqslant 2$ 时，X 也通常称为随机向量，而 d 就是随机向量的维数。随机变量可以分为离散型和连续型两种，视随机变量的取值而定。随机变量的研究是概率论的中心内容，这是因为在一个随机试验中，人们所关心的往往是与所研究的特定问题有关的某些量，而这些量就是随机变量。

接下来考虑随机变量的分布。

定义 1.1.6 设 X 为一随机变量，则函数

$$F(x)=P(X\leqslant x),\qquad -\infty<x<\infty \tag{1.1.7}$$

称为 X 的分布函数，其中 $P(X\leqslant x)$ 表示事件$\{\omega:X(\omega)\leqslant x\}$ 的概率。

这里并没有要求随机变量是离散的或是连续的。显然，分布函数具有如下性质：$F(x)$ 是单调非降的。当 $x\to-\infty$ 时，$F(x)\to 0$；而当 $x\to\infty$ 时，$F(x)\to 1$。

首先考虑离散型随机变量 X，其全部可能的取值为$\{a_1,a_2,\cdots\}$，则

$$p_i=P(X=a_i),\qquad i=1,2,\cdots$$

称为X的概率函数。一个重要的例子是 Poisson 分布。若随机变量X的取值为$\{0,1,2,\cdots\}$，且概率函数为

$$P(X=i)=\mathrm{e}^{-\lambda}\lambda^i/i!$$

则称 X 服从 Poisson 分布，常记为 $X\sim P(\lambda)$。此处，$\lambda>0$，为一常数。

对于连续型随机变量的分布，不能用离散型变量的方法描述。原因在于连续型随机变量的取值充满一个区间，不能一一列出。刻画连续型随机变量的一个方法是使用分布函数以及概率密度函数。

定义 1.1.7 设连续型随机变量 X 具有概率分布函数 $F(x)$，则 $F(x)$ 的导数 $f(x)=F'(x)$ 就称为 X 的概率密度函数。

密度函数 $f(x)$ 具有如下性质：

① $f(x)\geqslant 0$；

② $\int_{-\infty}^{\infty}f(x)\mathrm{d}x=1$；

③ 对任意的 $a<b$，有

$$P(a\leqslant X\leqslant b)=F(b)-F(a)=\int_a^b f(x)\mathrm{d}x$$

一个重要的连续型分布的例子是正态分布。其概率密度函数为

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-(x-\mu)^2/2\sigma^2},\qquad -\infty<x<\infty$$

通常将这样的随机变量记为 $X\sim N(\mu,\sigma^2)$。

自然地可以将上述结论推广到随机向量的情形。考虑 d 维随机向量 $\boldsymbol{X}=(X_1,X_2,\cdots,X_n)$，其分量 $X_1,X_2,\cdots,X_n$ 都是一维随机变量。为了避免重复，下面仅考虑连续型随机向量的情形，此时 $\boldsymbol{X}$ 的取值充满 R^n 的某个区域。对 R^n 的某个集合 A，引入记号：$\boldsymbol{X}\in A$ 表示事件 $\{\omega:\boldsymbol{X}(\omega)\in A\}$。

定义 1.1.8　*若 $f(x_1,x_2,\cdots,x_n)$ 为 R^n 上的非负函数，使得对 R^n 中的任意集合 A，有*

$$P(X\in A)=\int_A f(x_1,x_2,\cdots,x_n)\mathrm{d}x_1\cdots\mathrm{d}x_n \tag{1.1.8}$$

则称 f 是 X 的密度函数。

与一维情形类似，可以用概率分布去描述多维随机向量的概率分布，其定义为

$$F(x_1,x_2,\cdots,x_n)=P(X_1\leqslant x_1,X_2\leqslant x_2,\cdots,X_n\leqslant x_n)$$

然而在多维情形，分布函数极少应用。

对于随机向量 $\boldsymbol{X}=(X_1,X_2,\cdots,X_n)$，它的每个分量 X_i 都是一维随机变量，它们都有各自的分布 $F_i,i=1,2,\cdots,n$，它们都是一维分布，称为随机向量 $\boldsymbol{X}$ 或者其分布 F 的“边缘分布”。边缘分布完全由原分布 F 确定。考虑连续型随机向量 $\boldsymbol{X}=(X_1,X_2)$，其概率密度函数为 $f(x_1,x_2)$。由于事件$(X_1\leqslant x_1)=(X_1\leqslant x_1,X_2<\infty)$，从而

$$F_1(x_1)=P(X_1\leqslant x_1)=\int_{-\infty}^{x_1}\mathrm{d}t_1\int_{-\infty}^{\infty}f(t_1,t_2)\mathrm{d}t_2$$

以 x_2 替代 t_2，还可以得到 X_1 的概率密度函数：

$$f_1(x_1)\xlongequal{\text{def}}\frac{\mathrm{d}F_1(x_1)}{\mathrm{d}x_1}=\int_{-\infty}^{\infty}f(x_1,x_2)\mathrm{d}x_2$$

对 X_2 可作同样的处理。在高维情形，对 $X=(X_1,X_2,\cdots,X_n)$ 可以类似地得到

$$f(x_1)\xlongequal{\text{def}}\frac{\mathrm{d}F_1(x_1)}{\mathrm{d}x_1}=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}f(x_1,x_2,\cdots,x_n)\mathrm{d}x_2\cdots\mathrm{d}x_n$$

应当指出，“边缘”分布就是通常的分布，并无任何特殊含义。它只不过强调了：这个分布是由 X_i 作为随机向量 $\boldsymbol{X}=(X_1,X_2,\cdots,X_n)$ 之一的分量，且是从后者的分布之中派生出的分布而已。

前面引入了条件概率，并由此讨论了事件的独立性。下面接着讨论条件概率分布以及随机变量的独立性。所谓一个随机变量或向量 $\boldsymbol{X}$ 的条件概率分布，就是在给定某种条件下的 $\boldsymbol{X}$ 的概率分布。一般具有这样的形式：给定两个随机变量 X,Y，在给定了 Y 取某些值的条件下，去求 X 的条件分布。

仍以连续型随机变量为例。设二维随机向量 $\boldsymbol{X}=(X_1,X_2)$ 有概率密度函数 $f(x_1,x_2)$，考虑在限定 $a\leqslant X_2\leqslant b$ 的条件下，X_1 的条件分布。由于

$$P(X_1\leqslant x_1\mid a\leqslant X_1\leqslant b)=P(X_1\leqslant x_1,a\leqslant X_1\leqslant b)/P(a\leqslant X_1\leqslant b)$$

利用 X_2 的边缘分布的密度函数 f_2，可知

$$P(X_1\leqslant x_1\mid a\leqslant X_1\leqslant b)=\int_{-\infty}^{x_1}\mathrm{d}t_1\int_a^b f(t_1,t_2)\mathrm{d}t_2\Big/\int_a^b f_2(t_2)\mathrm{d}t_2$$

这就是 X_1 的条件分布函数。对此关于 x_1 求导可得条件密度函数

$$f_1(X_1\mid a\leqslant X_1\leqslant b)=\int_a^b f(x_1,t_2)\mathrm{d}t_2\Big/\int_a^b f_2(t_2)\mathrm{d}t_2$$

有意思的是考虑极限情形 $a=b$。通过极限步骤可以说明

$$\begin{aligned} f_1(x_1 \mid x_2) = f_1(x_1 \mid X_2 = x_2) = \\ \lim_{h\to 0} f_1(x_1 \mid x_2 \leqslant X_2 \leqslant x_2 + h) = \\ \lim_{h\to 0}\int_{x_1}^{x_2+h} f(x_1,t_2)\mathrm{d}t_2 \Big/ \lim_{h\to 0}\int_{x_1}^{x_2+h} f_2(t_2)\mathrm{d}t_2 = \\ f(x_1,x_2)/f_2(x_2) \end{aligned}$$

这就是在给定 $X_2=x_2$ 的条件下 X_1 的条件概率密度函数。在上式中，当然需要假设 $f_2(x_2)>0$，以保证此式有意义。此式还可改写为

$$f(x_1,x_2) = f_2(x_2)f_1(x_1 \mid x_2) \tag{1.1.9}$$

这对应于条件概率公式 $P(AB)=P(A)P(B)$。对于高维情形 $X=(X_1,X_2,\cdots,X_n)$，其概率密度函数为 $f(x_1,x_2,\cdots,x_n)$，则

$$f(x_1,x_2,\cdots,x_n) = g(x_1,x_2,\cdots,x_k)h(x_{k+1},x_{k+2},\cdots,x_n \mid x_1,x_2,\cdots,x_k)$$

其中，g 是$(X_1,X_2\cdots,X_k)$ 的概率密度，而 h 则是在给定 $X_1=x_1,x_2,\cdots,x_k$ 的条件下，X_{k+1}，$X_{k+2}\cdots,X_n$ 的条件概率密度。此式也可以视为条件概率密度 h 的定义。将式(1.1.9) 关于 x_2 积分可得

$$f_1(x_1) = \int_{-\infty}^{\infty} f(x_1,x_2)\mathrm{d}x_2 = \int_{-\infty}^{\infty} f_1(x_1 \mid x_2)f_2(x_2)\mathrm{d}x_2 \tag{1.1.10}$$

此式当然可以认为是全概率公式在概率密度情况下的表现形式。

接下来讨论随机变量的独立性。沿用上面的记号，如果 $f_1(x_1 \mid x_2)$ 不依赖于 x_2，而只是 x_1 的函数，则表明 X_1 的分布情况与 X_2 的取值完全无关，这时就称 X_1 和 X_2 这两个随机变量在概率意义上独立。此独立的概念与事件的独立概念完全类似。

定义 1.1.9 设 n 维随机变量 $X=(X_1,X_2,\cdots,X_n)$ 的联合概率密度函数为 $f(x_1,x_2,\cdots,x_n)$，而 X_i 的边缘密度函数为 $f_i(x_i)$，$i=1,2,\cdots,n$。如果

$$f(x_1,x_2,\cdots,x_n) = f_1(x_1)f_2(x_2)\cdots f_n(x_n)$$

就称随机变量 $X_1,X_2,\cdots,X_n$ 相互独立或者简称独立。

变量的独立性概念还可从下面的角度去考虑。按照前面的分析，如果 $X_1,X_2,\cdots,X_n$ 相互独立，则变量的概率不受其他变量的影响，从而事件

$$A_1 = (a_1 \leqslant X_1 \leqslant b_1),\cdots,A_n = (a_n \leqslant X_n \leqslant b_n)$$

是相互独立的。也可以将这个要求作为变量 $X_1,X_2,\cdots,X_n$ 独立的定义，它和上面的定义是等价的。

1.1.3 随机变量的数字特征

前面讨论了随机变量的概率分布，这种分布是随机变量的概率性质最完整的刻画。而随机变量的数字特征则是由随机变量的分布决定的常数，它也刻画了随机变量的某一方面的性质。

最简单也是最重要的数字特征莫过于数学期望和方差。首先考虑离散型随机变量 X，它的所有可能取值为有限个 $a_1,a_2,\cdots,a_m$，其相应的概率分布为 $P(X=a_i)=p_i$，$i=1,2,\cdots,m$，则其数学期望定义为 $\mathrm{E}(X)=\mathrm{E}X=\sum_{i=1}^{m} a_i p_i$。这导致如下一般情形的定义。

定义 1.1.10　设 X 为离散型随机变量，取无穷个值 $a_1,a_2,\cdots$，相应的概率分布为 $P(X=a_i)=p_i, i=1,2,\cdots$，如果

$$\sum_{i=1}^{\infty}|a_i|p_i<\infty \tag{1.1.11}$$

则称

$$\mathrm{E}X=\sum_{i=1}^{\infty}a_ip_i$$

为随机变量 X 的数学期望；相应地对连续型随机变量 X，设其概率密度函数为 $f(x)$，如果

$$\int_{-\infty}^{\infty}|x|f(x)\mathrm{d}x<\infty$$

则称

$$\mathrm{E}(X)=\int_{-\infty}^{\infty}xf(x)\mathrm{d}x \tag{1.1.12}$$

为 X 的数学期望。

条件式(1.1.11) 以及式(1.1.12) 是为了保证级数收敛。

与条件分布的定义相似，随机变量 Y 的条件数学期望，就是它在给定某个附加条件之下的数学期望。以只有两个变量 X 和 Y 的情形为例，要求在给定 $X=x$ 的情况下 Y 的期望，记为 $\mathrm{E}(Y\mid X=x)$。在不至于混淆的情况下，也简记为 $\mathrm{E}(Y\mid x)$。设已知 (X,Y) 的联合密度，在给定 $X=x$ 条件下，Y 的条件概率密度函数为 $f(y\mid x)$。由定义可知

$$\mathrm{E}(Y\mid x)=\int_{-\infty}^{\infty}yf(y\mid x)\mathrm{d}y$$

如果说条件分布是变量 X 和 Y 之间的相依关系在概率上的完全刻画，那么条件期望就在一个很重要的方面刻画了二者的关系。它反映了随着 X 的取值 x 的变化，Y 的平均值的变化情况，从而 $\mathrm{E}(Y\mid X)$ 为一随机变量，它随 X 的变化而变化。在统计学中，常把条件期望 $\mathrm{E}(Y\mid x)$ 作为 x 的函数，称为 Y 对 X 的“回归函数”。

从条件数学期望，可以得出求通常的无条件数学期望的一个重要公式。回忆全概率公式 $P(A)=\sum_i P(B_i)P(A\mid B_i)$，它可以理解为通过 A 的条件概率 $P(A\mid B_i)$ 来求得无条件概率 $P(A)$ 的一个表达式。更确切地，$P(A)$ 就是条件概率 $P(A\mid B_i)$ 的加权平均，权即是事件 B_i 的概率 $P(B_i)$。以此类推，变量 Y 的无条件期望，应等于其条件期望 $\mathrm{E}(Y\mid x)$ 对 x 的加权平均，其权与 X 在 x 点的概率密度 $f_1(x)$ 成比例，即

$$\mathrm{E}(Y)=\int_{-\infty}^{\infty}\mathrm{E}(Y\mid x)f_1(x)\mathrm{d}x$$

这个式子的证明是容易的。上式右端恰好是随机变量 $\mathrm{E}(Y\mid X)$ 关于 X 的期望，从而得到

$$\mathrm{E}(Y)=\mathrm{E}(\mathrm{E}(Y\mid X))$$

这个公式的意义如下：一个变量 Y 的期望，等于其条件期望的期望。

顺带提一下，在上面条件期望的定义中，重要的并不是 Y 的值，而是在 Y 的信息下能得到关于 X 的什么信息，从而可以直接考虑 Y 生成的 σ- 代数，同时也为了在方程中应用。下面讨论关于 $\mathscr{F}$ 的子 σ- 代数条件期望。给出如下定义。

定义 1.1.11　令 $(\Omega,\mathscr{F},P)$ 为一概率空间，$\mathscr{G}\subset\mathscr{F}$。如果 $X:\Omega\to R^n$ 为可积的随机变量，则定义 $\mathrm{E}(X\mid\mathscr{G})$ 为满足下述条件的随机变量：

① $\mathrm{E}(X \mid \mathscr{G})$ 是 $\mathscr{G}$- 可测的；

② $\int_A X \mathrm{d}P = \int_A \mathrm{E}(X \mid \mathscr{G}) \mathrm{d}P, \forall A \in \mathscr{G}$。

条件期望有如下性质：

命题 1.1.1 ① 如果 X 是 $\mathscr{G}$- 可测的，则 $\mathrm{E}(X \mid \mathscr{G}) = X$ a.s.。

② 如果 a,b 是常数，那么 $\mathrm{E}(aX + bY \mid \mathscr{G}) = a\mathrm{E}(X \mid \mathscr{G}) + b\mathrm{E}(Y \mid \mathscr{G})$ a.s.。

③ 如果 X 是 $\mathscr{G}$- 可测的，且 XY 是可积的，则 $\mathrm{E}(XY \mid \mathscr{G}) = X\mathrm{E}(Y \mid \mathscr{G})$ a.s.。

④ 如果 X 独立于 $\mathscr{G}$，则 $\mathrm{E}(X \mid \mathscr{G}) = \mathrm{E}(X)$ a.s.。

⑤ 如果 $\mathscr{E} \subset \mathscr{G}$，则

$$\mathrm{E}(X \mid \mathscr{E}) = \mathrm{E}(\mathrm{E}(X \mid \mathscr{G}) \mid \mathscr{E}) = \mathrm{E}(\mathrm{E}(X \mid \mathscr{E}) \mid \mathscr{G}) \text{a.s.}。$$

⑥ 如果 $X \leqslant Y$ a.s.，则 $\mathrm{E}(X \mid \mathscr{G}) \leqslant \mathrm{E}(Y \mid \mathscr{G})$ a.s.。

这里不证明此命题，读者可以参阅参考文献[8]；同时，不加证明地引入下面的 Jensen 不等式。关于一般情形的 Jensen 不等式，可以参阅相关的实分析的教材。

引理 1.1.1 设 $\Phi: R \to R$ 是凸的，$\mathrm{E}(\mid \Phi(X) \mid) < \infty$，那么

$$\Phi(\mathrm{E}(X \mid \mathscr{G})) \leqslant \mathrm{E}(\Phi(X) \mid \mathscr{G})$$

为了刻画随机变量的分散程度，可以引入方差的概念。

定义 1.1.12 设 X 为随机变量，分布为 F，则

$$\sigma_X^2 = \mathrm{var}(X) = \mathrm{E}(X - \mathrm{E}X)^2$$

称为 X 或者 F 的方差，而 σ_X 称为 X 或 F 的标准差。

容易说明如下关系：$\mathrm{var}(X) = \mathrm{E}(X^2) - (\mathrm{E}X)^2$。作为推广，可以考虑 X 的矩(moment)的概念。令 c 为常数，k 为正整数，则称 $\mathrm{E}[(X - c)^k]$ 为随机变量 X 关于 c 的 k 阶矩。特别地，当 $c = 0$ 时称为原点矩，而当 $c = \mathrm{E}X$ 时则称为中心矩。

下面介绍协方差和相关系数的概念。仍以二维随机向量 (X,Y) 为例。由于 X,Y 本身为一维随机变量，可以分别定义其期望和方差，记为

$$\mathrm{E}X = m_1, \quad \mathrm{E}Y = m_2; \qquad \mathrm{var}(X) = \sigma_1^2, \quad \mathrm{var}(Y) = \sigma_2^2$$

人们感兴趣的是反映分量之间关系的量，其中重要的是协方差(covariance)和相关系数(correlation)。

定义 1.1.13 称 $\mathrm{E}[(X-m_1)(Y-m_2)]$ 为 X,Y 的协方差，记为 $\mathrm{cov}(X,Y)$；而称 $\mathrm{cov}(X,Y)/(\sigma_1\sigma_2)$ 为 X,Y 的相关系数，并记为 $\rho(X,Y)$。

不难由 Schwarz 不等式得出 $\mathrm{cov}(X,Y) \leqslant \sigma_1\sigma_2$，从而 $-1 \leqslant \rho(X,Y) \leqslant 1$。若 X,Y 独立，则有 $\mathrm{cov}(X,Y) = 0$，从而 $\rho(X,Y) = 0$。然而反之却不成立，相关系数 $\rho(X,Y)$ 为 0，却不能推出 X,Y 独立。原因在于相关系数实际上只是“线性相关系数”，它并不刻画 X,Y 之间一般的关系程度。当然也有特例。当 X,Y 是二维正态时，由 $\rho(X,Y) = 0$ 能推出 X,Y 独立，此时独立和相关是一回事。

最后介绍大数定理和中心极限定理。在概率论中存在这样的情况，如果 $X_1, X_2, \cdots, X_n$ 是一些随机变量，则 $X_1 + X_2 + \cdots + X_n$ 的分布算起来一般比较复杂。因而人们自然会问：是否可以利用极限的方法来进行近似计算呢？事实表明，这一方法不仅是可能的，在许多时候更是非常方便的。在很一般的情况下，和的极限分布就是正态分布。这使得正态分布的地位在概率论中得到大大提升。习惯上把和的分布收敛于正态分布的那一类定理叫做“中心极限

定理”。还有一类情况，即所谓的“大数定理”，它是由概率的统计定义“频率收敛于概率”而来的。考虑 n 次独立试验，每次均观察 A 是否发生。记

$$X_i = \begin{cases} 1, & \text{事件 } A \text{ 在第 } i \text{ 次试验中发生} \\ 0, & \text{事件 } A \text{ 在第 } i \text{ 次试验中不发生} \end{cases} \quad (i = 1,2,\cdots,n) \tag{1.1.13}$$

则 n 次试验中 A 一共出现了 $\sum_i X_i$ 次，即频率为 $p_n = \sum_i^n X_i/n = \overline{X}_{(n)}$。若令 $P(A) = p$，则“频率收敛于概率”是说，在某种意义下，当 n 很大时，p_n 接近于 p。这就是一般情况下的大数定理。“大数”的意思即是涉及大量数目的观测值 X_i，它表明定理描述的现象只有在大量次数的试验和观测之下才能成立。

定理 1.1.3 设 $X_1, X_2, \cdots, X_n, \cdots$ 是独立同分布的随机变量，假设它们公共均值为 μ，方差存在并记为 σ^2，则对任意给定的 $\varepsilon > 0$，有

$$\lim_{n\to\infty} P(|\overline{X}_{(n)} - \mu| \geqslant \varepsilon) = 0 \tag{1.1.14}$$

此定理表明了当 n 很大时，“$\overline{X}_{(n)}$ 接近于 μ”的确切含义：它的意义是概率上的，不同于微积分意义下一数列 μ_n 收敛于 μ 的情形。在概率论中，将这样的收敛叫做“$\overline{X}_{(n)}$ 依概率收敛于 μ”。

这里不证明这个定理，但给出切比雪夫(Chebyshev)不等式，它在定理的证明中是有用的。

切比雪夫不等式如下：

若 $\text{var}(Y)$ 存在，则

$$P(|Y - \text{E}Y| \geqslant \varepsilon) \leqslant \text{var}(Y)/\varepsilon^2$$

定理 1.1.3 的一个重要特例是前面提到的“频率收敛到概率”：

$$\lim_{n\to\infty} P(|p_n - p| \geqslant \varepsilon) = 0 \tag{1.1.15}$$

此定理就是最早的一个大数定理，是伯努利(Bernoulli)在 1713 年证明的，常称为伯努利大数定理。在上一定理中要求 $X_1, X_2, \cdots$ 的方差存在，但是在这些随机变量服从相同分布的场合，并不需要这一要求，有如下的辛钦(Khintchine)定理。

定理 1.1.4 设随机变量 $X_1, X_2, \cdots$ 相互独立，服从相同的分布，且具有数学期望 $\text{E}X_k = \mu, k = 1,2,\cdots$，则对于任意正数 ε，有

$$\lim_{n\to\infty} P(|\overline{X}_{(n)} - \mu| < \varepsilon) = 1$$

伯努利大数定理是辛钦定理的特殊情况，辛钦定理在应用中是很重要的。

中心极限定理的意义已经阐述了，这里给出它们的具体形式。为此令

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-t^2/2} \mathrm{d}t$$

为标准正态分布 $N(0,1)$ 的分布函数。

定理 1.1.5 设 $X_1, X_2, \cdots, X_n, \cdots$ 为独立同分布的随机变量，$\text{E}X_i = \mu, \text{var}(X_i) = \sigma^2$，$0 < \sigma^2 < \infty$。则对任意实数 x，有

$$\lim_{n\to\infty} P\left[\frac{1}{\sqrt{n}\sigma}(X_1 + \cdots + X_n - n\mu) \leqslant x\right] = \Phi(x) \tag{1.1.16}$$

定理的意思是，均值为 μ、方差为 σ^2 的独立同分布的随机变量 $X_1, X_2, \cdots, X_n$ 之和 $\sum_{k=1}^{n} X_k$

的标准化变量，当 n 充分大时，近似地服从标准正态分布，即

$$\frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma} \sim N(0,1) \quad （近似地）$$

此定理通常称为林德伯格-列维定理。考虑如下的特例：令 X_i 为式(1.13) 所定义，$X_1 + X_2 + \cdots + X_n$ 就是某事件 A 在 n 次独立试验中发生的次数。

定理 1.1.6 设 $X_1, X_2, \cdots, X_n, \cdots$ 独立同分布，X_i 满足分布

$$P(X_i = 1) = p, \qquad P(X_i = 0) = 1 - p, \qquad 0 < p < 1$$

则对任意实数 x，有

$$\lim_{n\to\infty} P\left[\frac{1}{\sqrt{np(1-p)}}(X_1 + X_2 + \cdots + X_n - np) \leqslant x\right] = \Phi(x)$$

此定理称为棣莫弗-拉普拉斯定理，是历史上最早的中心极限定理。

作为本节的结束，下面列举几种重要的收敛性以及它们之间的关系(实际上前面已经提到了依概率收敛)。这些在概率论以及随机过程的研究中会经常遇到。

定义 1.1.14 设 $\{X_n; n \geqslant 1\}$ 为随机变量序列，若存在随机变量 X，使得对任意的 $\varepsilon > 0$，有

$$\lim_{n\to\infty} P(|X_n - X| \geqslant \varepsilon) = 0$$

则称随机变量序列 $\{X_n; n \geqslant 1\}$ 依概率收敛到 X，记为 $X_n \xrightarrow{P} X$。

如果事件 $\{\omega: \lim_{n\to\infty}(X_n(\omega) - X(\omega)) = 0\}$ 的概率为 1，即

$$P\left[\lim_{n\to\infty}(X_n - X) = 0\right] = 1$$

则称随机变量序列 $\{X_n; n \geqslant 1\}$ 几乎必然(a. s.) 收敛到 X，也称为随机变量序列以概率 1 收敛于 X，记为 $X_n \to X$ a. s. 。

设 $X, X_n (n \geqslant 1)$ 都有有限的 p- 阶矩。如果

$$\lim_{n\to\infty} \mathrm{E}(|X_n - X|^p) = 0$$

则称为随机变量序列 $\{X_n; n \geqslant 1\}$ 以 L^p 范数收敛到 X，记为 $X_n \xrightarrow{L^p} X$。特别地，如果 $p = 2$，则称 $\{X_n; n \geqslant 1\}$ 均方收敛到 X。

上述三种收敛性的关系是：均方收敛和几乎必然收敛都蕴含依概率收敛，反之不成立；另外，均方收敛和几乎必然收敛则互不包含。相反，如果 X_n 以概率收敛到 X，则由 Riesz 定理可知，存在子列 X_{n_k} 几乎处处收敛到 X。

1.2 随机过程的预备知识

下面简单介绍随机过程的一些基本概念。所谓随机过程，简单地讲就是一族定义在同一概率空间 $(\Omega, \mathscr{F}, P)$ 上的随机变量 $X = \{X(t, \omega)\}_{t\geqslant 0}$。有时也将 $X(t, \omega)$ 记为 $X_t(\omega)$ 或者 X_t。

定义 1.2.1 设 $(\Omega, \mathscr{F}, P)$ 为概率空间，T 是给定的参数集，若对每一个 $t \in T$，有一个随机变量 $X(t, \omega)$ 与之对应，则称随机变量族 $\{X(t, \omega)\}_{t\in T}$ 是 $(\Omega, \mathscr{F}, P)$ 上的随机过程，通常简记为 $\{X(t)\}_{t\in T}$。$X(t)$ 的所有可能状态构成的集合称为状态空间或者相空间 $\mathscr{S}$，T 称为参数集，通常表示时间。