Nonlinear
Stochastic
Differential
Equations

Nonlinear Stochastic Differential Equations

◉ 周少波　编著

非线性随机微分方程

華中科技大學出版社
http://www.hustp.com

No ear
tochastic
Diff ial
quations

◉ 周少波 编著

非线性随机微分方程

華中科技大學出版社
http://www.hustp.com
中国 · 武汉

内 容 简 介

本书主要介绍了 Itô 型非线性随机微分方程，包括时滞随机微分方程和中立型随机微分方程的基本理论，深入讨论了非线性随机微分方程的稳定性、稳定化及其数值方法的收敛性及稳定性等。此外，本书还综述了近年来国内外非线性随机微分方程最新研究成果。

本书可作为高等院校数学系本科生、研究生的教材，高校教师的参考书，也可作为研究工作涉及随机微分方程的科技工作者的阅读资料。

图书在版编目(CIP)数据

非线性随机微分方程＝Nonlinear Stochastic differential equations：英文/周少波编著. —武汉：华中科技大学出版社，2014.7

ISBN 978-7-5680-0223-3

Ⅰ.①非…　Ⅱ.①周…　Ⅲ.①非线性-随机微分方程-教材-英文　Ⅳ.①O211.63

中国版本图书馆 CIP 数据核字(2014)第 155192 号

非线性随机微分方程
Nonlinear Stochastic Differential Equations
周少波　编著

策划编辑：吴　晗
责任编辑：姚同梅
封面设计：刘　卉
责任校对：封力煊
责任监印：张正林
出版发行：华中科技大学出版社(中国·武汉)
武昌喻家山　邮编：430074　电话：(027)81321915
录　　排：武汉市洪山区佳年华文印部
印　　刷：华中理工大学印刷厂
开　　本：787mm×1092mm　1/16
印　　张：20.5　插页：2
字　　数：536 千字
版　　次：2014 年 8 月第 1 版第 1 次印刷
定　　价：59.80 元

Nonlinear Stochastic Differential Equations

Preface

Nonlinear stochastic system has come to play an important role in many branches of science and industry. More and more researches have involved nonlinear stochastic differential equations. Recently, many research efforts have been devoted to deal with nonlinear stochastic differential systems. Many papers on them were published in different journals, which is not convenient for readers to understand the theory systematically. This book is therefore written. The main aim of this book is to explore systematically all various of nonlinear stochastic differential systems. Some important features of this text are as follows:

The text will be the first systematic presentation of the basic principles of various types of nonlinear stochastic systems, including stochastic differential equations, stochastic functional differential equations, stochastic equations of neutral type. It will emphasize the current research trends in the field of nonlinear systems at an advanced level, in which the local Lipschitz and one-sided polynomial growth conditions will replace the classical uniform Lipschitz and linear growth conditions.

This text emphasizes the analysis of stability which is vital in the automatic control of stochastic systems. The Lyapunov method can be adopted to study all various of stable properties of stochastic systems. Especially, this text demonstrates that the Khasminskii-type criteria on stability is very effective for highly nonlinear stochastic systems with delay.

The text explains systematically the use of the Razumikhin technique in the study of exponential stability for stochastic functional differential equations and functional equations of neutral-type with finite or infinite delays as well as stability of the discrete Euler-Maruyama approximate solution.

The text will be the first systematic presentation of the basic theory of nonlinear stochastic functional differential equations with infinite delays. It discusses the existence and exponential stability of stochastic functional differential equations on special and general measure spaces.

The text demonstrates systematically the stabilization of nonlinear deterministic system. It indicates a nonlinear Brownian noise feedback to suppress the potential explosion of the system, and a linear Brownian noise feedback to stabilize exponentially this system.

This text discusses new developments of the Euler-Maruyama approximation schemes under the local and one-sided Lipschitz conditions as well as one-sided polynomial growth condition. This text studies linear stability of the Euler-Maruyama approximate schemes and nonlinear stability of the backward Euler-Maruyama approximate schemes. The advantage of the backward Euler-Maruyama approximate schemes is that the approximate solution converges to the accurate solution under the local Lipschitz and one-sided polynomial growth conditions.

This text is mainly based on the papers of Professor Fuke Wu and Professor Xuerong Mao as well as some recent research papers, for example, Wu and Hu (2009a), Wu and Hu (2010a,b), Wu, Hu and Mao(2011), Wu and Hu(2011e), Mao and Szpruch(2012), Mao and Szpruch(2013), Zhou and Xie(2014), Zhou(2014a), Zhou(2014b). It hence discusses many hot topics including the discrete Razumikhin-type theorem, numerical convergence and stability, population dynamics and stochastic stabilization.

The text is suitable for advanced undergraduate students, graduate students and teachers in colleges and universities as well as research workers involving stochastic dynamic systems.

I have to thank Professor Shigeng Hu for his constant support and kind assistance and encouragement. I have to thank Professor Fuke Wu, who has provided me a great deal of material during the writing process, for his support and help. I also wish to thank Ph. D. Yangzi Hu for her support and encouragement. Moreover, I should thank my family for their constant support and understanding.

I also would like to thank the Teaching Material Foundation of HUST, the National Natural Science Foundation of China (Grant No. 11301198 and 11422110) as well as National Excellent Young Foundation of China (61473125) for their financial supports.

General Notation

positive: >0

non-positive: $\leqslant 0$

negative: <0

non-negative: $\geqslant 0$

a. s. : almost surely or $\mathbb{P}$-almost surely or with probability one

$a(t)\equiv b(t)$: $a(t)$ and $b(t)$ are identically equal, i. e. $a(t)=b(t)$ for all t

$\varnothing$: the empty set

I_A: the indicator function of a set A, i. e. $I_A(x)=1$ if $x\in A$, otherwise 0

A^c: the complement of A in Ω, i. e. $A^c=\Omega-A$

$A\subset B$: A is a subset of B, i. e. $A\cap B^c=\varnothing$

$A\subset B$ a. s. : $\mathbb{P}(A\cap B^c)=0$

$\sigma(C)$: the σ-algebra generated by C

$a\vee b$: the maximum of a and b

$a\wedge b$: the minimum of a and b

$f:A\to B$: the mapping f from A to B

$\mathbb{S}:=\{1,2,\cdots,N\}$, the finite state space of a Markov chain

$\mathbb{R}=\mathbb{R}^1$: the real line

$\mathbb{R}_+$: $[0,\infty)$, the set of all non-negative real numbers

$\mathbb{R}^d$: the d-dimensional Euclidean space

$\mathbb{R}_+^d$: $\{(x_1,x_2,\cdots,x_d)\in\mathbb{R}^d: x_i>0, 1\leqslant i\leqslant d\}$

F: σ-algebra

F_t: σ-algebra filtration

$F_\infty=\sigma(\bigcup_{t\geqslant 0}F_t)$

$\mathbb{B}^d$: the Borel-σ-algebra on $\mathbb{R}^d$

$\mathbb{B}:=\mathbb{B}^1$

$\mathbb{R}^{d\times m}$: the space of real $d\times m$-matrices

$\mathbb{B}^{d\times m}$: the Borel-σ-algebra on $\mathbb{R}^{d\times m}$

$\mathbb{Z}^{d\times d}:=\{(a_{ij})_{d\times d}\in\mathbb{R}^{d\times d}: a_{ij}<0, i\neq j\}$

$|\boldsymbol{x}|$: the Euclidean norm of a vector $\boldsymbol{x}$

$\mathbb{S}_h:=\{\boldsymbol{x}\in\mathbb{R}^d: |\boldsymbol{x}|<h\}$

$\bar{\mathbb{S}}_h:=\{\boldsymbol{x}\in\mathbb{R}^n: |\boldsymbol{x}|\leqslant h\}$

$\boldsymbol{A}^{\mathrm{T}}$: the transpose of a vector or matrix $\boldsymbol{A}$

$\boldsymbol{A}=\boldsymbol{A}^{\mathrm{T}}$: $\boldsymbol{A}$ is a symmetric matrix

$A=A^T>0$: A is a symmetric positive-definite matrix

$A=A^T\geqslant 0$: A is a symmetric non-negative-definite matrix

$A\geqslant 0$: each element of A is non-negative

$A>0$: $A\geqslant 0$ and at least one element of A is positive

$A\gg 0$: all elements of A are positive

$A_1\geqslant A_2$: if and only if $A_1-A_2\geqslant 0$

$A_1>A_2$: if and only if $A_1-A_2>0$

$A_1\gg A_2$: if and only if $A_1-A_2\gg 0$

$\langle x,y\rangle$: the inner product of vectors x and y, namely $\langle x,y\rangle=x^T y$

$\mathrm{tr}(A)$: the trace of a square matrix $A=(a_{ij})_{d\times d}$ i. e. $\mathrm{tr}(A)=\sum_{1\leqslant i\leqslant n} a_{ii}$

$\lambda_{\min}(A)$: the smallest eigenvalue of a symmetric matrix A

$\lambda_{\max}(A)$: the largest eigenvalue of a symmetric matrix A

$\lambda_{\max}^{+}(A):=\sup_{x\in\mathbb{R}_+^d,|x|=1} x^T Ax$ for a symmetric matrix A

$\lambda(A)$: the spectrum of A

$\rho(A)$: the spectrum radius of A

$|A|$: $|A|=\sqrt{\mathrm{tr}(A^T A)}$ i. e. the trace norm of A

$\|A\|$: $\|A\|=\sup\{|Ax|:|x|=1\}=\sqrt{\lambda_{\max}(A^T A)}$ i. e. the operator norm of A

δ_{ij}: Dirac's delta function, that is $\delta_{ij}=1$ if $i=j$, otherwise 0

V_x: $V_x=(V_{x_1},V_{x_2},\cdots,V_{x_d})=\left(\frac{\partial V}{\partial x_1},\frac{\partial V}{\partial x_2},\cdots,\frac{\partial V}{\partial x_d}\right)$

V_{xx}: $V_{xx}=(V_{x_i x_j})_{d\times d}=\left(\frac{\partial^2 V}{\partial x_i\partial x_j}\right)_{d\times d}$

$\|\xi\|_{L^p}:=(\mathbb{E}|\xi|^p)^{1/p}$

$L^p(\Omega;\mathbb{R}^d)$: the family of $\mathbb{R}^d$-valued random variables ξ with $\mathbb{E}|\xi|^p<\infty$

$L^p_{F_t}(\Omega;\mathbb{R}^d)$: the family of $\mathbb{R}^d$-valued F_t-measurable random variables ξ with $\mathbb{E}|\xi|^p<\infty$

$C([-\tau,0];\mathbb{R}^d)$: the space of continuous $\mathbb{R}^d$-valued functions φ defined on $[-\tau,0]$ with a norm $\|\varphi\|=\sup_{-\tau\leqslant\theta\leqslant 0}|\varphi(\theta)|$

$L^p_F([-\tau,0];\mathbb{R}^d)$: the family of $C([-\tau,0];\mathbb{R}^d)$-valued random variables φ such that $\mathbb{E}\|\varphi\|^p<\infty$

$L^p_{F_t}([-\tau,0];\mathbb{R}^d)$: the family of F_t-measurable $C([-\tau,0];\mathbb{R}^d)$-valued random variables φ such that $\mathbb{E}\|\varphi\|^p<\infty$

$C^b_{F_t}([-\tau,0];\mathbb{R}^d)$: the family of F_t-measurable bounded $C([-\tau,0];\mathbb{R}^d)$-valued random variables

$L^p([a,b];\mathbb{R}^d)$: the family of Borel measurable functions $h:[a,b]\to\mathbb{R}^d$ such that $\int_a^b|h(t)|^p\mathrm{d}t<\infty$

$\mathcal{L}^p([a,b];\mathbb{R}^d)$: the family of $\mathbb{R}^d$ -valued F_t-adapted processes $\{f(t)\}_{a\leqslant t\leqslant b}$ such that $\int_a^b |f(t)|^p dt < \infty$ a. s.

$M^p([a, b];\mathbb{R}^d)$: the family of processes $\{f(t)\}_{a\leqslant t\leqslant b}$ in $\mathcal{L}^p([a, b];\mathbb{R}^d)$ such that $\mathbb{E}\int_a^b |f(t)|^p dt < \infty$

$\mathcal{L}^p(\mathbb{R}_+;\mathbb{R}^d)$: the family of processes $\{f(t)\}_{t\geqslant 0}$ such that for every $T>0$, $\{f(t)\}_{0\leqslant t\leqslant T} \in L^p([0, T];\mathbb{R}^d)$

$M^p(\mathbb{R}_+;\mathbb{R}^d)$: the family of processes $\{f(t)\}_{t\geqslant 0}$ such that for every $T>0$, $\{f(t)\}_{0\leqslant t\leqslant T} \in M^p([0, T];\mathbb{R}^d)$

$W([-\tau, 0];\mathbb{R}_+)$: the family of all Borel measurable bounded non-negative functions $\zeta(s)$ defined on $-\tau\leqslant s\leqslant 0$ satisfying $\int_c^{c+\tau} \zeta(c-s)ds = 1$ for any constant c

$W((-\infty,0];\mathbb{R}_+)$: the family of all Borel measurable bounded non-negative functions $\zeta(s)$ defined on $-\infty<s\leqslant 0$ such that $\int_{-\infty}^0 \zeta(s)ds = 1$

$GB((-\infty,0];\mathbb{R}^d)$: the family of bounded continuous $\mathbb{R}^d$-valued stochastic process $\boldsymbol{\xi} = \{\boldsymbol{\xi}(s), -\infty<s\leqslant 0\}$ with the norm $\|\boldsymbol{\varphi}\| = \sup_{\theta\leqslant 0}|\boldsymbol{\varphi}(\theta)|$

K: the family of all continuous increasing functions $\boldsymbol{\kappa}: \mathbb{R}_+ \to \mathbb{R}_+$ such that $\boldsymbol{\kappa}(0)=0$ while $\boldsymbol{\kappa}(u)>0$ for $u>0$

K_∞: the family of all functions $\boldsymbol{\kappa}\in K$ with property that $\boldsymbol{\kappa}(\infty)=\infty$

$K_\vee$: the family of all convex functions $\boldsymbol{\kappa}\in K$

$K_\wedge$: the family of all concave functions $\boldsymbol{\kappa}\in K$

$L^1(\mathbb{R}_+;\mathbb{R}^d)$: the family of functions $\gamma: \mathbb{R}_+ \to \mathbb{R}_+$ such that $\int_0^\infty \gamma(t)dt < \infty$

$\Phi(\mathbb{R}_+;\mathbb{R}_+)$: the family of functions $\varphi: \mathbb{R}_+ \to \mathbb{R}_+$ such that $\int_0^\infty \varphi(t)dt = \infty$

$\Psi(\mathbb{R}_+;\mathbb{R}_+)$: the family of all continuous functions $\psi: \mathbb{R}_+ \to \mathbb{R}_+$ such that for any $\delta>0$ and any increasing sequence $\{t_k\}_{k\geqslant 1}$, $\sum_{k=1}^\infty \int_{t_k}^{t_k+\delta} \psi(t)dt = \infty$

$\mathrm{sign}(x)$: the sign function, namely, $\mathrm{sign}(x)=1$ if $x\geqslant 0$, otherwise -1

□: The proof of theorem or lemma is complete

Contents

1 Stochastic Integral

The well-known Lotka-Volterra models proposed by Lotka(1925) and Volterra(1926) play an important role in ecological population theories. When two and more species live in proximity and share the same basic requirements, they usually compete for common resources including food, habitat or territory. A competitive Lotka-Volterra system with d species is described by

$$\frac{\mathrm{d}x_i(t)}{\mathrm{d}t} = x_i(t)\Big[b_i - \sum_{j=1}^{d} a_{ij}x_j(t)\Big], \quad i = 1,2,\cdots,d, \tag{1.1}$$

where $x_i(t)$ represents the population size of species i at time t, the constant b_i is the growth rate of species i, and a_{ij} represents the effect of interspecific (if $i \neq j$) or intraspecific($i=j$) interaction. The quotient b_i/a_{ii} is the carrying capacity of the i-th species in absence of other species. Eq. (1. 1) can also be written as

$$\frac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t} = \mathrm{diag}(x_1(t),\ x_2(t),\ \cdots,\ x_d(t))[\boldsymbol{b} - \boldsymbol{A}x(t)], \tag{1.2}$$

where $\boldsymbol{x} = (x_1,\ x_2,\cdots,\ x_d)^{\mathrm{T}}$ is a d-dimensional state vector, $\boldsymbol{b} = (b_1,\ b_2,\ \cdots,\ b_d)^{\mathrm{T}}$, $\boldsymbol{A} = (a_{ij})_{d\times d}$ is a $d\times d$ matrix, known as the community matrix, and $\boldsymbol{z}^{\mathrm{T}}$ denotes the transpose of $\boldsymbol{z}$.

Note that the parameter b_i represents the intrinsic growth rate of species i. In practice, it might happen that b_i is not completely known but subject to some random environment effects, so we usually estimate it by an average value plus an error which follows a normal distribution. Let b_i denote the average growth rate, then the intrinsic growth rate becomes

$$b_i \to b_i + \sigma_i\text{“noise”}.$$

So Eq. (1. 2) becomes

$$\frac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t} = \mathrm{diag}(x_1(t),\ x_2(t),\ \cdots,\ x_d(t))[\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{\sigma}\text{“noise”}], \tag{1.3}$$

where $\boldsymbol{\sigma} = (\sigma_1,\ \sigma_2,\ \cdots,\ \sigma_d)^{\mathrm{T}}$. The integral form is

$$x_i(t) = x_i(0) + \int_0^t x_i(s)\Big[b_i - \sum_{j=1}^{d} a_{ij}x_j(s)\Big]\mathrm{d}s + \int_0^t \sigma_i x_i(s)\text{“noise”}\mathrm{d}s. \tag{1.4}$$

The questions are: What is the mathematical interpretation for the “noise” term and what is the integration $\int_0^t \sigma_i x_i(s)\text{“noise”}\mathrm{d}s$?

A reasonable mathematical interpretation for the “noise” term is the so-called white noise $\dot{w}(t)$, which is formally regarded as the derivative of a Brownian motion $w(t)$, i. e. $\dot{w}(t) = \mathrm{d}w(t)/\mathrm{d}t$. So the term “noise”$\mathrm{d}t$ can be expressed as $\dot{w}(t)\mathrm{d}t = \mathrm{d}w(t)$, and

$$\int_0^t \sigma_i x_i(s)\text{“noise”}ds = \int_0^t \sigma_i x_i(s)dw(s). \quad (1.5)$$

The question is: What is the integration $\int_0^t \sigma_i x_i(s)dw(s)$? If the Brownian motion $w(t)$ was differentiable with its derivative $\dot{w}(t) = dw(t)/dt$, then the integral would have no problem at all as it could be done as the classical Lebesgue integral $\int_0^t \sigma_i x_i(s)\dot{w}(s)ds$. Unfortunately, we shall see that the Brownian motion $w(t)$ is nowhere differentiable, hence the integral is different to the ordinary integration. On the other hand, if $w(t)$ is a process of finite variation, the integral $\int_0^t \sigma_i x_i(s)dw(s)$ could be regarded as the Lebesgue-Stieltjes one. However, we shall see that almost every sample path of the Browian motion has infinite variation in any finite time interval later. Hence the integral cannot be defined in the ordinary way. The integral was first defined by K. Itô in 1949 and is now known as the Itô stochastic integral.

To define this integration, we shall briefly review the basic concepts of probability theory and stochastic processes. By the aid of these properties, we proceed to define the Brownian motions and the stochastic integral and establish the well-known Itô formula referring to Mao(1997), Klebaner(2008), Hu, Huang and Wu(2008).

1.1 Variation

If f is a function of real variable, its variation over the interval $[0, t]$ is defined by

$$V_f([0,t]) = \lim_{\sigma_n \to 0} \sum_{k=1}^{n} | f(t_k^n) - f(t_{k-1}^n) |,$$

where the limit is taken over partitions: $0=t_0^n<t_1^n<\cdots<t_n^n=t$ with $\sigma_n = \max_{1\leq k\leq n}(t_k^n - t_{k-1}^n)$.

If f is a function of real variable, its quadratic variation over the interval $[0, t]$ is defined by

$$[f,f](t) = [f,f]([0,t]) = \lim_{\sigma_n \to 0} \sum_{k=1}^{n} | f(t_k^n) - f(t_{k-1}^n) |^2,$$

where the limit is taken over partitions: $0=t_0<t_1^n<\cdots<t_n^n=t$ with $\sigma_n = \max_{1\leq k\leq n}(t_k^n - t_{k-1}^n)$.

If f and g are functions of real variable, their quadratic covariation over the interval $[0, t]$ is defined by

$$[f,g](t) = \lim_{\sigma_n \to 0} \sum_{k=1}^{n} (f(t_k^n) - f(t_{k-1}^n))(g(t_k^n) - g(t_{k-1}^n)),$$

where the limit is taken over partitions: $0=t_0<t_1^n<\cdots<t_n^n=t$, with $\sigma_n = \max_{1\leq k\leq n}(t_k^n - t_{k-1}^n)$.

Quadratic variation and covariation play important role in stochastic calculus, but it hardly ever meets in standard calculus due to the fact that smooth functions have zero quadratic variation. That is, if f is continuous and of finite variation then its quadratic variation is zero. If f is continuous and g is of finite variation, then their quadratic covariation is zero(see Klebaner(2008)).

1.2 Random Variable

1.2.1 Probability Space

The outcomes of a random experiment cannot be determined in advance. The set of all possible outcomes is called the *sample space* Ω with typical element $\omega \in \Omega$. A subset of a sample space is said to be *an event*. Not every subset of Ω is in general an observable or interesting event. So we only group these observable or interesting events together as a family F of subsets of Ω. A family F is said to be *σ-algebra* if

(i) $\varnothing \in F$, where $\varnothing$ denotes the empty set;

(ii) $A \in F \Rightarrow A^c \in F$, where $A^c = \Omega - A$ is the complement of A in Ω;

(iii) $\{A_i\}_{i \geqslant 1} \subset F \Rightarrow \bigcup_{i=1}^{\infty} A_i \in F$ (also $\bigcap_{i=1}^{\infty} A_i \in F$).

That is, a σ-algebra is a collection of subsets of Ω, which is closed with respect to countable unions and countable intersections of its members. The pair (Ω, F) is called a *measurable space*, and the elements of F are said to be *F-measurable* sets. If C is a family of subsets of Ω, there exists a smallest σ-algebra $\sigma(C)$ on Ω which contains C. This $\sigma(C)$ is called the *σ-algebra generated by* C. $B^d = \sigma(C)$ is said to be the *Borel* σ-algebra if $\Omega = \mathbb{R}^d$ and C is the family of all open sets in $\mathbb{R}^d$, and the elements of B^d are called the *Borel sets*.

A real-valued function $X: \Omega \to \mathbb{R}$ is said to be F-measurable if

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

The function X is also called a real-valued (F-measurable) *random variable*. A is an F-measurable set(i. e. $A \in F$) if and only if the indicator function I_A is *F-measurable*, where the *indicator function* I_A of a set $A \subset \Omega$ is defined by

$$I_A(\omega) = \begin{cases} 1 & \text{for} \quad \omega \in A, \\ 0 & \text{for} \quad \omega \overline{\in} A. \end{cases}$$

Theorem 1.1 (i) X is a random variable on (Ω, F) if and only if it is a simple function or a limit of simple functions.

(ii) If $X_n (n \geqslant 1)$ are random variables on (Ω, F) and $X(\omega) = \lim_{n \to \infty} X_n(\omega)$, then X is a random variable.

(iii) If X is a random variable on (Ω, F) and g is a B-measurable function, then $g(X)$ is a random variable.

Example 1.1 Define a sequence of simple functions

$$X_n(\omega) = \sum_{k=-n2^n}^{n2^n-1} \frac{k}{2^n} I_{\left[\frac{k}{2^n}, \frac{k+1}{2^n}\right)} X(\omega).$$

Clearly, X_n's are increasing and all the sets $\left\{\omega: \frac{k}{2^n} \leqslant X(\omega) < \frac{k+1}{2^n}\right\}$ are measurable by the definition of a random variable. Moreover, we shall see that X_n converge to a limit X, i. e. $X(\omega)=\lim_{n\to\infty} X_n(\omega)$ for all ω.

This example is due to Lebesgue (see Klebaner(2008)), who gave it for non-negative functions, demonstrating that a measurable function X is a limit of a monotone sequence of simple function X_n, $X_{n+1} \geqslant X_n$.

More generally, let $\{\Omega', F'\}$ be another measurable space. A mapping $X: \Omega \to \Omega'$ is said to *be* $\{F, F'\}$ *-measurable* if

$$\{\omega: \boldsymbol{X}(\omega) \leqslant A'\} \in F, \quad A' \in F'.$$

The mapping X is called an *Ω'-valued* $\{F, F'\}$*-measurable* (or simply, *F'-measurable*) *random variable*.

An $\mathbb{R}^d$-valued function $\boldsymbol{X}(\omega)=(X_1(\omega), X_2(\omega), \cdots, X_d(\omega))^{\mathrm{T}}$ is said to be *F-measurable* if all the elements X_i are F-measurable. A $d \times m$-matrix-valued function $\boldsymbol{X}(\omega)=(X_{ij}(\omega))_{d\times m}$ is said to be *F-measurable* if all the elements X_{ij} are F-measurable. If the measurable space is $\{\mathbb{R}^d, B^d\}$, a B^d-measurable function is called a *Borel measurable function*. Let $\boldsymbol{X}: \Omega \to \mathbb{R}^d$ be any function. The σ-algebra $\sigma(\boldsymbol{X})$ generated by $\boldsymbol{X}$ is the smallest σ-algebra on Ω containing all the sets $\{\omega: \boldsymbol{X}(\omega) \in B\}$, $B \subset \mathbb{R}^d$ open. That is

$$\sigma(\boldsymbol{X})=\sigma(\{\omega: \boldsymbol{X}(\omega) \in B\}: B \subset \mathbb{R}^d \text{ open}).$$

Clearly, $\boldsymbol{X}$ will be $\sigma(\boldsymbol{X})$-measurable and $\sigma(\boldsymbol{X})$ is the smallest σ-algebra with this property. If $\boldsymbol{X}$ is F-measurable, then $\sigma(\boldsymbol{X}) \subset F$, i. e. $\boldsymbol{X}$ generates a sub-σ-algebra of F. If $\{\boldsymbol{X}_i: i \in I\}$ is a collection of $\mathbb{R}^d$-valued functions, define

$$\sigma\{\boldsymbol{X}_i: i \in I\}=\sigma\Big(\bigcup_{i\in I} \sigma(\boldsymbol{X}_i)\Big),$$

which is called the σ-algebra generated by $\{\boldsymbol{X}_i: i \in I\}$. It is the smallest σ-algebra with respect to which every $\boldsymbol{X}_i$ is measurable.

Lemma 1.1 (The Doob-Dynkin lemma) If $\boldsymbol{X}, \boldsymbol{Y}: \Omega \to \mathbb{R}^d$ are two given function, then $\boldsymbol{Y}$ is $\sigma(\boldsymbol{X})$ measurable if and only if there exists a Borel measurable function $\boldsymbol{g}: \mathbb{R}^d \to \mathbb{R}^d$, such that $\boldsymbol{Y}=\boldsymbol{g}(\boldsymbol{X})$.

A function $\mathbb{P}: F \to [0, 1]$ is said to be a *probability measure* $\mathbb{P}$ on a measurable space (Ω, F) if

(i) $\mathbb{P}(\Omega)=1$;

(ii) $A \in F \Rightarrow \mathbb{P}(A^c)=1-\mathbb{P}(A)$, where $A^c=\Omega/A$;

(iii) (Countable additivity) For any disjoint sequence $\{A_i\}_{i\geqslant 1} \subset F$(i. e. $A_i \cap A_j = \varnothing$ if $i \neq j$)

$$\mathbb{P}\Big(\bigcup_{i=1}^{\infty} A_i\Big)=\sum_{i=1}^{\infty} \mathbb{P}(\boldsymbol{A}_i).$$

The triple $(\Omega, F, \mathbb{P})$ is called *a probability space*.

If $(\Omega,F,\mathbb{P})$ is a probability space, define

$$\bar{F}=\{A\subset\Omega:\ \exists B,\ C\in F:B\subset A\subset C,\ \mathbb{P}(B)=\mathbb{P}(C)\}.$$

Then $\bar{F}$ is a σ-algebra and is called the *completion of* F. If $F=\bar{F}$, the probability space $(\Omega,F,\mathbb{P})$ is said to be *complete*. If not, we can extend $\mathbb{P}$ to $\bar{F}$ by defining $\mathbb{P}(A)=\mathbb{P}(B)=\mathbb{P}(C)$ for $A\in\bar{F}$, where B, $C\in F$ with the properties that $B\subset A\subset C$ and $\mathbb{P}(B)=\mathbb{P}(C)$. So$(\Omega,\bar{F},\mathbb{P})$ is a complete probability space, called the *completion* of $(\Omega,F,\mathbb{P})$.

A collection of sets $\{A_i:i\in I\}\subset F$ is said to be *independent* if

$$\mathbb{P}(A_{i_1}\cap A_{i_2}\cap\cdots\cap A_{i_k})=\mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2})\cdots\mathbb{P}(A_{i_k})$$

for all possible choices of finite indices $i_1,i_2,\cdots,i_k\in I$, where I be an index set.

Two sub-σ-algebras F_1 and F_2 of F are said to be *independent* if

$$\mathbb{P}(A_1\cap A_2)=\mathbb{P}(A_1)\mathbb{P}(A_2)$$

for all $A_1\in F_1$, $A_2\in F_2$.

A collection of sub-σ-algebras $\{F_i:i\in I\}$ is said to be *independent* if for every possible choice of indices i_1, $i_2,\cdots$, $i_k\in I$,

$$\mathbb{P}(A_{i_1}\cap A_{i_2}\cap\cdots\cap A_{i_k})=\mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2})\cdots\mathbb{P}(A_{i_k})$$

for all $A_{i_1}\in F_{i_1}$, $A_{i_2}\in F_{i_2},\cdots$, $A_{i_k}\in F_{i_k}$. A family of random variables $\{X_i:i\in I\}$ is said to be *independent* if the σ-algebras $\sigma(X_i)$, $i\in I$ generated by them are independent.

For example, two random variables $\boldsymbol{X}:\Omega\to\mathbb{R}^d$ and $\boldsymbol{Y}:\Omega\to\mathbb{R}^m$ are independent if and only if

$$\mathbb{P}\{\omega:\boldsymbol{X}(\omega)\in A,\ \boldsymbol{Y}(\omega)\in B\}=\mathbb{P}\{\omega:\boldsymbol{X}(\omega)\in A\}\mathbb{P}\{\omega:\boldsymbol{Y}(\omega)\in B\}$$

holds for all $A\in B^d$, $B\in B^m$.

1.2.2 Expectation and Moment

Let $(\Omega,F,\mathbb{P})$ be a probability space. X is real-valued random variable, then

$$\mathbb{E}X=\int_\Omega X(\omega)\,\mathrm{d}\,\mathbb{P}(\omega)$$

is said to be the *expectation of* X (with respect to $\mathbb{P}$) if X is integrable with respect to the probability measure $\mathbb{P}$. For an $\mathbb{R}^d$-valued random variable $\boldsymbol{X}=(X_1,\ X_2,\cdots,\ X_d)^{\mathrm{T}}\in L^1(\Omega;\mathbb{R}^d)$, define $\mathbb{E}\boldsymbol{X}=(\mathbb{E}X_1,\ \mathbb{E}X_2,\cdots,\mathbb{E}X_d)^{\mathrm{T}}\in\mathbb{R}^d$. For a $d\times m$-matrix-valued random variable $\boldsymbol{X}=(X_{ij})_{d\times m}\in L^1(\Omega;\mathbb{R}^{d\times m})$, define $\mathbb{E}\boldsymbol{X}=(\mathbb{E}X_{ij})_{d\times m}\in\mathbb{R}^{d\times m}$. Denote $L^p(\Omega;\mathbb{R}^d)$ by the family of $\mathbb{R}^d$-valued random variable $\boldsymbol{X}$ with $\mathbb{E}\,|\boldsymbol{X}|^p<\infty$. If $1\leqslant p<\infty$, then $L^p(\Omega;\mathbb{R}^d)$ is a Banach space with the norm $(\mathbb{E}\,|\boldsymbol{X}|^p)^{1/p}$. $L^2(\Omega;\mathbb{R}^d)$ is a Hilbert space with its inner product $\langle\boldsymbol{X},\ \boldsymbol{Y}\rangle=\mathbb{E}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y})$. $\mathbb{E}\,|\boldsymbol{X}|^p(p>0)$ is said to be the *p-th moment* of $\boldsymbol{X}$ for $\boldsymbol{X}\in L^p(\Omega;\mathbb{R}^d)$. $\mathrm{V}(\boldsymbol{X})=\mathbb{E}(\boldsymbol{X}-\mathbb{E}\boldsymbol{X})^2$ is said to be the *variance of* $\boldsymbol{X}$ for $\boldsymbol{X}\in L^2(\Omega;\mathbb{R}^d)$. $\mathrm{Cov}(X,\ Y)=\mathbb{E}[(X-\mathbb{E}X)(Y-\mathbb{E}Y)]$ is said to be the *covariance* of X and Y for X, $Y\in L^2(\Omega;\mathbb{R})$.

X and Y are said to be *uncorrelated* if $\mathrm{Cov}(X,\ Y)=0$ for X, $Y\in L^p(\Omega;\mathbb{R})$. The symmetric non-negative definite $d\times d$ matrix

$$\mathrm{Cov}(\boldsymbol{X},\ \boldsymbol{Y})=\mathbb{E}[(\boldsymbol{X}-\mathbb{E}\,\boldsymbol{X})(\boldsymbol{Y}-\mathbb{E}\,\boldsymbol{Y})^{\mathrm{T}}],\quad \boldsymbol{X},\ \boldsymbol{Y}\in L^2(\Omega;\mathbb{R}^d)$$

is called their *covariance matrix*. For any $X_i \in L^1(\Omega;\mathbb{R})(i=1,2,\cdots,d)$, if $X_1, X_2, \cdots, X_d$ are independent, then

$$\mathbb{E}\left(\prod_{i=1}^{d} X_i\right) = \prod_{i=1}^{d} \mathbb{E} X_i.$$

If $X, Y \in L^2(\Omega;\mathbb{R})$ are uncorrelated, then

$$V(X+Y) = V(X) + V(Y).$$

If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent, they are uncorrelated, the converse is not true. If $(\boldsymbol{X}, \boldsymbol{Y})$ has a joint normal distribution, then $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent if and only if they are uncorrelated.

Let $\boldsymbol{X}$ be an $\mathbb{R}^d$-valued random variable. μ_X is said to be a *probability measure induced* by $\boldsymbol{X}$ on the Borel measurable space $(\mathbb{R}^d, B^d)$ in the sense of

$$\mu_X(B) = P\{\omega : \boldsymbol{X}(\omega) \in B\}$$

for $B \in B^d$, and μ_X is said to be the *distribution* of $\boldsymbol{X}$. The expectation of $\boldsymbol{X}$ can now be expressed as

$$\mathbb{E}\boldsymbol{X} = \int_{\mathbb{R}^d} \boldsymbol{x}\,\mathrm{d}\mu_X(\boldsymbol{x}).$$

More generally, if $\boldsymbol{g} : \mathbb{R}^d \to \mathbb{R}^m$ is Borel measurable, then

$$\mathbb{E}\, g(\boldsymbol{X}) = \int_{\mathbb{R}^d} g(\boldsymbol{x})\,\mathrm{d}\mu_X(\boldsymbol{x}).$$

The integral of the right is Lebesgue-Stieltjes integral with respect to $\mu_X(\cdot)$.

The following moment inequalities will play an important role in this book.

(1) **Hölder inequality**

$$|\mathbb{E}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y})| \leqslant (\mathbb{E}|\boldsymbol{X}|^p)^{\frac{1}{p}}(\mathbb{E}|\boldsymbol{X}|^q)^{\frac{1}{q}},\ p>1,\ \frac{1}{p}+\frac{1}{q}=1,\ \boldsymbol{X}\in L^p(\Omega;\mathbb{R}^d),\ \boldsymbol{Y}\in L^q(\Omega;\mathbb{R}^d).$$

In particular, $|\mathbb{E}\boldsymbol{X}| \leqslant \mathbb{E}|\boldsymbol{X}|$ for $\boldsymbol{X} \in L^1(\Omega;\mathbb{R}^d)$.

(2) **Lyapunov's inequality**

$$(\mathbb{E}|\boldsymbol{X}|^r)^{\frac{1}{r}} \leqslant (\mathbb{E}|\boldsymbol{X}|^p)^{\frac{1}{p}},\ 0<r<p<\infty,\ \boldsymbol{X}\in L^p(\Omega;\mathbb{R}^d).$$

(3) **Minkovski's inequality**

$$(\mathbb{E}|\boldsymbol{X}+\boldsymbol{Y}|^p|)^{\frac{1}{p}} \leqslant (\mathbb{E}|\boldsymbol{X}|^p)^{\frac{1}{p}} + (\mathbb{E}|\boldsymbol{X}|^p)^{\frac{1}{p}},\ p>1,\ \boldsymbol{X}, \boldsymbol{Y}\in L^p(\Omega;\mathbb{R}^d).$$

(4) **Chebyshev's inequality**

$$\mathrm{P}\{\omega : |\boldsymbol{X}(\omega)| \geqslant \varepsilon\} \leqslant \varepsilon^{-p}\mathbb{E}|\boldsymbol{X}|^p,\ \varepsilon>0,\ p>0,\ \boldsymbol{X}\in L^p(\Omega;\mathbb{R}^d).$$

(5) C_p **inequality**

$$\mathbb{E}\left|\sum_{i=1}^{d} \boldsymbol{X}_i\right|^p \leqslant C_p \sum_{i=1}^{d} \mathbb{E}|\boldsymbol{X}_i|^p$$

for $\boldsymbol{X} \in L^p(\Omega;\mathbb{R}^d)$, $C_p = 1$ for $0<p\leqslant 1$ or $C_p = d^{p-1}$ for $p>1$.

(6) **Jensen's inequality**

$$g(\mathbb{E}\boldsymbol{X}) \leqslant \mathbb{E}(g(\boldsymbol{X})),$$

where $\boldsymbol{X} \in L^1(\Omega;\mathbb{R}^d)$, $g : \mathbb{R}^d \to \mathbb{R}$ is a convex function.

The following convergence concepts are very useful in this book.