

Markov Processes

Markov Processes

Characterization and Convergence

STEWART N. ETHIER
THOMAS G. KURTZ



A JOHN WILEY & SONS, INC., PUBLICATION

Copyright © 1986, 2005 by John Wiley & Sons, Inc. All rights reserved.

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.

Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the U.S. at (800) 762-2974, outside the U.S. at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic format. For information about Wiley products, visit our web site at www.wiley.com.

Library of Congress Cataloging-in-Publication is available.

ISBN-13 978-0-471-76986-6

ISBN-10 0-471-76986-X

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

PREFACE

The original aim of this book was a discussion of weak approximation results for Markov processes. The scope has widened with the recognition that each technique for verifying weak convergence is closely tied to a method of characterizing the limiting process. The result is a book with perhaps more pages devoted to characterization than to convergence.

The Introduction illustrates the three main techniques for proving convergence theorems applied to a single problem. The first technique is based on operator semigroup convergence theorems. Convergence of generators (in an appropriate sense) implies convergence of the corresponding semigroups, which in turn implies convergence of the Markov processes. Trotter's original work in this area was motivated in part by diffusion approximations. The second technique, which is more probabilistic in nature, is based on the martingale characterization of Markov processes as developed by Stroock and Varadhan. Here again one must verify convergence of generators, but weak compactness arguments and the martingale characterization of the limit are used to complete the proof. The third technique depends on the representation of the processes as solutions of stochastic equations, and is more in the spirit of classical analysis. If the equations "converge," then (one hopes) the solutions converge.

Although the book is intended primarily as a reference, problems are included in the hope that it will also be useful as a text in a graduate course on stochastic processes. Such a course might include basic material on stochastic processes and martingales (Chapter 2, Sections 1-6), an introduction to weak convergence (Chapter 3, Sections 1-9, omitting some of the more technical results and proofs), a development of Markov processes and martingale problems (Chapter 4, Sections 1-4 and 8), and the martingale central limit theorem (Chapter 7, Section 1). A selection of applications to particular processes could complete the course.

As an aid to the instructor of such a course, we include a flowchart for all proofs in the book. Thus, if one's goal is to cover a particular section, the chart indicates which of the earlier results can be skipped with impunity. (It also reveals that the course outline suggested above is not entirely self-contained.)

Results contained in standard probability texts such as Billingsley (1979) or Breiman (1968) are assumed and used without reference, as are results from measure theory and elementary functional analysis. Our standard reference here is Rudin (1974). Beyond this, our intent has been to make the book self-contained (an exception being Chapter 8). At points where this has not seemed feasible, we have included complete references, frequently discussing the needed material in appendixes.

Many people contributed toward the completion of this project. Cristina Costantini, Eimear Goggin, S. J. Sheu, and Richard Stockbridge read large portions of the manuscript and helped to eliminate a number of errors. Carolyn Birr, Dee Frana, Diane Reppert, and Marci Kurtz typed the manuscript. The National Science Foundation and the University of Wisconsin, through a Romnes Fellowship, provided support for much of the research in the book.

We are particularly grateful to our editor, Beatrice Shube, for her patience and constant encouragement. Finally, we must acknowledge our teachers, colleagues, and friends at Wisconsin and Michigan State, who have provided the stimulating environment in which ideas germinate and flourish. They contributed to this work in many uncredited ways. We hope they approve of the result.

STEWART N. ETHIER
THOMAS G. KURTZ

*Salt Lake City, Utah
Madison, Wisconsin
August 1985*

CONTENTS

Introduction	1
1 Operator Semigroups	6
1 Definitions and Basic Properties,	6
2 The Hille–Yosida Theorem,	10
3 Cores,	16
4 Multivalued Operators,	20
5 Semigroups on Function Spaces,	22
6 Approximation Theorems,	28
7 Perturbation Theorems,	37
8 Problems,	42
9 Notes,	47
2 Stochastic Processes and Martingales	49
1 Stochastic Processes,	49
2 Martingales,	55
3 Local Martingales,	64
4 The Projection Theorem,	71
5 The Doob–Meyer Decomposition,	74
6 Square Integrable Martingales,	78
7 Semigroups of Conditioned Shifts,	80
8 Martingales Indexed by Directed Sets,	84
9 Problems,	89
10 Notes,	93

3 Convergence of Probability Measures	95
1 The Prohorov Metric, 96	
2 Prohorov's Theorem, 103	
3 Weak Convergence, 107	
4 Separating and Convergence Determining Sets, 111	
5 The Space $D_E[0, \infty)$, 116	
6 The Compact Sets of $D_E[0, \infty)$, 122	
7 Convergence in Distribution in $D_E[0, \infty)$, 127	
8 Criteria for Relative Compactness in $D_E[0, \infty)$, 132	
9 Further Criteria for Relative Compactness in $D_E[0, \infty)$, 141	
10 Convergence to a Process in $C_E[0, \infty)$, 147	
11 Problems, 150	
12 Notes, 154	
4 Generators and Markov Processes	155
1 Markov Processes and Transition Functions, 156	
2 Markov Jump Processes and Feller Processes, 162	
3 The Martingale Problem: Generalities and Sample Path Properties, 173	
4 The Martingale Problem: Uniqueness, the Markov Property, and Duality, 182	
5 The Martingale Problem: Existence, 196	
6 The Martingale Problem: Localization, 216	
7 The Martingale Problem: Generalizations, 221	
8 Convergence Theorems, 225	
9 Stationary Distributions, 238	
10 Perturbation Results, 253	
11 Problems, 261	
12 Notes, 273	
5 Stochastic Integral Equations	275
1 Brownian Motion, 275	
2 Stochastic Integrals, 279	
3 Stochastic Integral Equations, 290	
4 Problems, 302	
5 Notes, 305	
6 Random Time Changes	306
1 One-Parameter Random Time Changes, 306	
2 Multiparameter Random Time Changes, 311	
3 Convergence, 321	

4	Markov Processes in \mathbb{Z}^d,	329
5	Diffusion Processes,	328
6	Problems,	332
7	Notes,	335
7	Invariance Principles and Diffusion Approximations	337
1	The Martingale Central Limit Theorem,	338
2	Measures of Mixing,	345
3	Central Limit Theorems for Stationary Sequences,	350
4	Diffusion Approximations,	354
5	Strong Approximation Theorems,	356
6	Problems,	360
7	Notes,	364
8	Examples of Generators	365
1	Nondegenerate Diffusions,	366
2	Degenerate Diffusions,	371
3	Other Processes,	376
4	Problems,	382
5	Notes,	385
9	Branching Processes	386
1	Galton–Watson Processes,	386
2	Two-Type Markov Branching Processes,	392
3	Branching Processes in Random Environments,	396
4	Branching Markov Processes,	400
5	Problems,	407
6	Notes,	409
10	Genetic Models	410
1	The Wright–Fisher Model,	411
2	Applications of the Diffusion Approximation,	415
3	Genotypic-Frequency Models,	426
4	Infinitely-Many-Allele Models,	435
5	Problems,	448
6	Notes,	451
11	Density Dependent Population Processes	452
1	Examples,	452
2	Law of Large Numbers and Central Limit Theorem,	455

x CONTENTS

3	Diffusion Approximations,	459
4	Hitting Distributions,	464
5	Problems,	466
6	Notes,	467
12	Random Evolutions	468
1	Introduction,	468
2	Driving Process in a Compact State Space,	472
3	Driving Process in a Noncompact State Space,	479
4	Non-Markovian Driving Process,	483
5	Problems,	491
6	Notes,	491
Appendixes		492
1	Convergence of Expectations,	492
2	Uniform Integrability,	493
3	Bounded Pointwise Convergence,	495
4	Monotone Class Theorems,	496
5	Gronwall's Inequality,	498
6	The Whitney Extension Theorem,	499
7	Approximation by Polynomials,	500
8	Bimeasures and Transition Functions,	502
9	Tulcea's Theorem,	504
10	Measurable Selections and Measurability of Inverses,	506
11	Analytic Sets,	506
References		508
Index		521
Flowchart		529

INTRODUCTION

The development of any stochastic model involves the identification of properties and parameters that, one hopes, uniquely characterize a stochastic process. Questions concerning continuous dependence on parameters and robustness under perturbation arise naturally out of any such characterization. In fact the model may well be derived by some sort of limiting or approximation argument. The interplay between characterization and approximation or convergence problems for Markov processes is the central theme of this book. Operator semigroups, martingale problems, and stochastic equations provide approaches to the characterization of Markov processes, and to each of these approaches correspond methods for proving convergence results.

The processes of interest to us here always have values in a complete, separable metric space E , and almost always have sample paths in $D_E[0, \infty)$, the space of right continuous E -valued functions on $[0, \infty)$ having left limits. We give $D_E[0, \infty)$ the Skorohod topology (Chapter 3), under which it also becomes a complete, separable metric space. The type of convergence we are usually concerned with is convergence in distribution; that is, for a sequence of processes $\{X_n\}$ we are interested in conditions under which $\lim_{n \rightarrow \infty} E[f(X_n)] = E[f(X)]$ for every $f \in \bar{C}(D_E[0, \infty))$. (For a metric space S , $\bar{C}(S)$ denotes the space of bounded continuous functions on S . Convergence in distribution is denoted by $X_n \Rightarrow X$.) As an introduction to the methods presented in this book we consider a simple but (we hope) illuminating example.

For each $n \geq 1$, define

$$(1) \quad \lambda_n(x) = 1 + 3x\left(x - \frac{1}{n}\right), \quad \mu_n(x) = 3x + x\left(x - \frac{1}{n}\right)\left(x - \frac{2}{n}\right).$$

2 INTRODUCTION

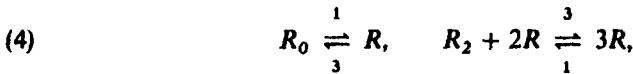
and let Y_n be a birth-and-death process in \mathbb{Z}_+ with transition probabilities satisfying

$$(2) \quad P\{Y_n(t+h) = j+1 \mid Y_n(t) = j\} = n\lambda_n\left(\frac{j}{n}\right)h + o(h)$$

and

$$(3) \quad P\{Y_n(t+h) = j-1 \mid Y_n(t) = j\} = n\mu_n\left(\frac{j}{n}\right)h + o(h)$$

as $h \rightarrow 0+$. In this process, known as the **Schlögl model**, $Y_n(t)$ represents the number of molecules at time t of a substance R in a volume n undergoing the chemical reactions



with the indicated rates. (See Chapter 11, Section 1.)

We rescale and renormalize letting

$$(5) \quad X_n(t) = n^{1/4}(n^{-1}Y_n(n^{1/2}t) - 1), \quad t \geq 0.$$

The problem is to show that X_n converges in distribution to a Markov process X to be characterized below.

The first method we consider is based on a semigroup characterization of X . Let $E_n = \{n^{1/4}(n^{-1}y - 1) : y \in \mathbb{Z}_+\}$, and note that

$$(6) \quad T_n(t)f(x) \equiv E[f(X_n(t)) \mid X_n(0) = x]$$

defines a semigroup $\{T_n(t)\}$ on $B(E_n)$ with generator of the form

$$(7) \quad G_n f(x) = n^{3/2}\lambda_n(1 + n^{-1/4}x)\{f(x + n^{-3/4}) - f(x)\} \\ + n^{3/2}\mu_n(1 + n^{-1/4}x)\{f(x - n^{-3/4}) - f(x)\}.$$

(See Chapter 1.) Letting $\lambda(x) \equiv 1 + 3x^2$, $\mu(x) \equiv 3x + x^3$, and

$$(8) \quad Gf(x) = 4f''(x) - x^3f'(x),$$

a Taylor expansion shows that

$$(9) \quad G_n f(x) = Gf(x) + n^{3/2}\{\lambda_n(1 + n^{-1/4}x) - \lambda(1 + n^{-1/4}x)\}\{f(x + n^{-3/4}) - f(x)\} \\ + n^{3/2}\{\mu_n(1 + n^{-1/4}x) - \mu(1 + n^{-1/4}x)\}\{f(x - n^{-3/4}) - f(x)\} \\ + \lambda(1 + n^{-1/4}x) \int_0^1 (1-u)\{f''(x + un^{-3/4}) - f''(x)\} du \\ + \mu(1 + n^{-1/4}x) \int_0^1 (1-u)\{f''(x - un^{-3/4}) - f''(x)\} du \\ + \{(\lambda + \mu)(1 + n^{-1/4}x) - (\lambda + \mu)(1)\}f''(x),$$

for all $f \in C^2(\mathbb{R})$ with $f' \in C_c(\mathbb{R})$ and all $x \in E_n$. Consequently, for such f ,

$$(10) \quad \lim_{n \rightarrow \infty} \sup_{x \in E_n} |G_n f(x) - Gf(x)| = 0.$$

Now by Theorem 1.1 of Chapter 8,

$$(11) \quad A \equiv \{(f, Gf) : f \in C[-\infty, \infty] \cap C^2(\mathbb{R}), Gf \in C[-\infty, \infty]\}$$

is the generator of a Feller semigroup $\{T(t)\}$ on $C[-\infty, \infty]$. By Theorem 2.7 of Chapter 4 and Theorem 1.1 of Chapter 8, there exists a diffusion process X corresponding to $\{T(t)\}$, that is, a strong Markov process X with continuous sample paths such that

$$(12) \quad E[f(X(t)) | \mathcal{F}_s^X] = T(t-s)f(X(s))$$

for all $f \in C[-\infty, \infty]$ and $t \geq s \geq 0$. ($\mathcal{F}_s^X = \sigma(X(u) : u \leq s)$.)

To prove that $X_n \Rightarrow X$ (assuming convergence of initial distributions), it suffices by Corollary 8.7 of Chapter 4 to show that (10) holds for all f in a core D for the generator A , that is, for all f in a subspace D of $\mathcal{D}(A)$ such that A is the closure of the restriction of A to D . We claim that

$$(13) \quad D \equiv \{f + g : f, g \in C^2(\mathbb{R}), f' \in C_c(\mathbb{R}), (x^2g)' \in C_c(\mathbb{R})\}$$

is a core, and that (10) holds for all $f \in D$. To see that D is a core, first check that

$$(14) \quad \mathcal{D}(A) = \{f \in C[-\infty, \infty] \cap C^2(\mathbb{R}) : f'' \in \hat{C}(\mathbb{R}), x^3f' \in C[-\infty, \infty]\}.$$

Then let $h \in C_c^2(\mathbb{R})$ satisfy $\chi_{[l-1, l]} \leq h \leq \chi_{[l-2, 2]}$ and put $h_m(x) = h(x/m)$. Given $f \in \mathcal{D}(A)$, choose $g \in D$ with $(x^2g)' \in C_c(\mathbb{R})$ and $x^3(f-g)' \in \hat{C}(\mathbb{R})$ and define

$$(15) \quad f_m(x) = f(0) - g(0) + \int_0^x (f-g)'(y)h_m(y) dy.$$

Then $f_m + g \in D$ for each m , $f_m + g \rightarrow f$, and $G(f_m + g) \rightarrow Gf$.

The second method is based on the characterization of X as the solution of a martingale problem. Observe that

$$(16) \quad f(X_n(t)) - \int_0^t G_n f(X_n(s)) ds$$

is an $\{\mathcal{F}_s^{X_n}\}$ -martingale for each $f \in B(E_n)$ with compact support. Consequently, if some subsequence $\{X_n\}$ converges in distribution to X , then, by the continuous mapping theorem (Corollary 1.9 of Chapter 3) and Problem 7 of Chapter 7,

$$(17) \quad f(X(t)) - \int_0^t Gf(X(s)) ds$$

is an $\{\mathcal{F}_t^X\}$ -martingale for each $f \in C_c^2(\mathbb{R})$, or in other words, X is a solution of the martingale problem for $\{(f, Gf) : f \in C_c^2(\mathbb{R})\}$. But by Theorem 2.3 of Chapter 8, this property characterizes the distribution on $D_{\mathbb{R}}[0, \infty)$ of X . Therefore, Corollary 8.16 of Chapter 4 gives $X_n \Rightarrow X$ (assuming convergence of initial distributions), provided we can show that

$$(18) \quad \lim_{\alpha \rightarrow \infty} \overline{\lim}_{n \rightarrow \infty} P \left\{ \sup_{0 \leq t \leq T} |X_n(t)| \geq \alpha \right\} = 0, \quad T > 0.$$

Let $\varphi(x) \equiv e^x + e^{-x}$, and check that there exist constants $C_{n,a} \geq 0$ such that $G_n \varphi \leq C_{n,a} \varphi$ on $[-\alpha, \alpha]$ for each $n \geq 1$ and $a > 0$, and $\overline{\lim}_{a \rightarrow \infty} \lim_{n \rightarrow \infty} C_{n,a} < \infty$. Letting $\tau_{n,a} = \inf\{t \geq 0 : |X_n(t)| \geq \alpha\}$, we have

$$\begin{aligned} (19) \quad & e^{-C_{n,a}T} \inf_{|y| \geq a} \varphi(y) P \left\{ \sup_{0 \leq t \leq T} |X_n(t)| \geq \alpha \right\} \\ & \leq E[\exp \{-C_{n,a}(\tau_{n,a} \wedge T)\} \varphi(X_n(\tau_{n,a} \wedge T))] \\ & \leq E[\varphi(X_n(0))] \end{aligned}$$

by Lemma 3.2 of Chapter 4 and the optional sampling theorem. An additional (mild) assumption on the initial distributions therefore guarantees (18).

Actually we can avoid having to verify (18) by observing that the uniform convergence of $G_n f$ to Gf for $f \in C_c^2(\mathbb{R})$ and the uniqueness for the limiting martingale problem imply (again by Corollary 8.16 of Chapter 4) that $X_n \Rightarrow X$ in $D_{\mathbb{R}^A}[0, \infty)$ where \mathbb{R}^A denotes the one-point compactification of \mathbb{R} . Convergence in $D_{\mathbb{R}}[0, \infty)$ then follows from the fact that X_n and X have sample paths in $D_{\mathbb{R}}[0, \infty)$.

Both of the approaches considered so far have involved characterizations in terms of generators. We now consider methods based on stochastic equations. First, by Theorems 3.7 and 3.10 of Chapter 5, we can characterize X as the unique solution of the stochastic integral equation

$$(20) \quad X(t) = X(0) + 2\sqrt{2}W(t) - \int_0^t X(s)^3 ds,$$

where W is a standard, one-dimensional, Brownian motion. (In the present example, the term $2\sqrt{2}W(t)$ corresponds to the stochastic integral term.) A convergence theory can be developed using this characterization of X , but we do not do so here. The interested reader is referred to Kushner (1974).

The final approach we discuss is based on a characterization of X involving random time changes. We observe first that Y_n satisfies

$$(21) \quad Y_n(t) = Y_n(0) + N_+ \left(n \int_0^t \lambda_n(n^{-1} Y_n(s)) ds \right) - N_- \left(n \int_0^t \mu_n(n^{-1} Y_n(s)) ds \right),$$

where N_+ and N_- are independent, standard (parameter 1), Poisson processes. Consequently, X_n satisfies

$$(22) \quad X_n(t) = X_n(0) + n^{-3/4} \tilde{N}_+ \left(n^{3/2} \int_0^t \lambda_n(1 + n^{-1/4} X_n(s)) ds \right) \\ - n^{-3/4} \tilde{N}_- \left(n^{3/2} \int_0^t \mu_n(1 + n^{-1/4} X_n(s)) ds \right) \\ + n^{3/4} \int_0^t (\lambda_n - \mu_n)(1 + n^{-1/4} X_n(s)) ds,$$

where $\tilde{N}_+(u) = N_+(u) - u$ and $\tilde{N}_-(u) = N_-(u) - u$ are independent, centered, standard, Poisson processes. Now it is easy to see that

$$(23) \quad (n^{-3/4} \tilde{N}_+(n^{3/2} \cdot), n^{-3/4} \tilde{N}_-(n^{3/2} \cdot)) \Rightarrow (W_+, W_-),$$

where W_+ and W_- are independent, standard, one-dimensional Brownian motions. Consequently, if some subsequence $\{X_{n_k}\}$ converges in distribution to X , one might expect that

$$(24) \quad X(t) = X(0) + W_+(4t) + W_-(4t) - \int_0^t X(s)^3 ds.$$

(In this simple example, (20) and (24) are equivalent, but they will not be so in general.) Clearly, (24) characterizes X , and using the estimate (18) we conclude $X_n \Rightarrow X$ (assuming convergence of initial distributions) from Theorem 5.4 of Chapter 6.

For a further discussion of the Schlögl model and related models see Schlögl (1972) and Malek-Mansour et al. (1981). The martingale proof of convergence is from Costantini and Nappo (1982), and the time change proof is from Kurtz (1981c).

Chapters 4–7 contain the main characterization and convergence results (with the emphasis in Chapters 5 and 7 on diffusion processes). Chapters 1–3 contain preliminary material on operator semigroups, martingales, and weak convergence, and Chapters 8–12 are concerned with applications.

1

OPERATOR SEMIGROUPS

Operator semigroups provide a primary tool in the study of Markov processes. In this chapter we develop the basic background for their study and the existence and approximation results that are used later as the basis for existence and approximation theorems for Markov processes. Section 1 gives the basic definitions, and Section 2 the Hille–Yosida theorem, which characterizes the operators that are generators of semigroups. Section 3 concerns the problem of verifying the hypotheses of this theorem, and Sections 4 and 5 are devoted to generalizations of the concept of the generator. Sections 6 and 7 present the approximation and perturbation results.

Throughout the chapter, L denotes a real Banach space with norm $\|\cdot\|$.

1. DEFINITIONS AND BASIC PROPERTIES

A one-parameter family $\{T(t): t \geq 0\}$ of bounded linear operators on a Banach space L is called a *semigroup* if $T(0) = I$ and $T(s + t) = T(s)T(t)$ for all $s, t \geq 0$. A semigroup $\{T(t)\}$ on L is said to be *strongly continuous* if $\lim_{t \rightarrow 0} T(t)f = f$ for every $f \in L$; it is said to be a *contraction semigroup* if $\|T(t)\| \leq 1$ for all $t \geq 0$.

Given a bounded linear operator B on L , define

$$(1.1) \quad e^{tB} = \sum_{k=0}^{\infty} \frac{1}{k!} t^k B^k, \quad t \geq 0.$$

A simple calculation gives $e^{(s+t)B} = e^{sB}e^{tB}$ for all $s, t \geq 0$, and hence $\{e^{tB}\}$ is a semigroup, which can easily be seen to be strongly continuous. Furthermore we have

$$(1.2) \quad \|e^{tB}\| \leq \sum_{k=0}^{\infty} \frac{1}{k!} t^k \|B^k\| \leq \sum_{k=0}^{\infty} \frac{1}{k!} t^k \|B\|^k = e^{t\|B\|}, \quad t \geq 0.$$

An inequality of this type holds in general for strongly continuous semigroups.

1.1 Proposition Let $\{T(t)\}$ be a strongly continuous semigroup on L . Then there exist constants $M \geq 1$ and $\omega \geq 0$ such that

$$(1.3) \quad \|T(t)\| \leq M e^{\omega t}, \quad t \geq 0.$$

Proof. Note first that there exist constants $M \geq 1$ and $t_0 > 0$ such that $\|T(t)\| \leq M$ for $0 \leq t \leq t_0$. For if not, we could find a sequence $\{t_n\}$ of positive numbers tending to zero such that $\|T(t_n)\| \rightarrow \infty$, but then the uniform boundedness principle would imply that $\sup_n \|T(t_n)f\| = \infty$ for some $f \in L$, contradicting the assumption of strong continuity. Now let $\omega = t_0^{-1} \log M$. Given $t \geq 0$, write $t = kt_0 + s$, where k is a nonnegative integer and $0 \leq s < t_0$; then

$$(1.4) \quad \|T(t)\| = \|T(s)T(t_0)^k\| \leq M M^k \leq M M^{t/t_0} = M e^{\omega t}. \quad \square$$

1.2 Corollary Let $\{T(t)\}$ be a strongly continuous semigroup on L . Then, for each $f \in L$, $t \mapsto T(t)f$ is a continuous function from $[0, \infty)$ into L .

Proof. Let $f \in L$. By Proposition 1.1, if $t \geq 0$ and $h \geq 0$, then

$$(1.5) \quad \begin{aligned} \|T(t+h)f - T(t)f\| &= \|T(t)[T(h)f - f]\| \\ &\leq M e^{\omega t} \|T(h)f - f\|, \end{aligned}$$

and if $0 \leq h \leq t$, then

$$(1.6) \quad \begin{aligned} \|T(t-h)f - T(t)f\| &= \|T(t-h)[T(h)f - f]\| \\ &\leq M e^{\omega t} \|T(h)f - f\|. \end{aligned} \quad \square$$

1.3 Remark Let $\{T(t)\}$ be a strongly continuous semigroup on L such that (1.3) holds, and put $S(t) = e^{-\omega t}T(t)$ for each $t \geq 0$. Then $\{S(t)\}$ is a strongly continuous semigroup on L such that

$$(1.7) \quad \|S(t)\| \leq M, \quad t \geq 0.$$

8 OPERATOR SEMIGROUPS

In particular, if $M = 1$, then $\{S(t)\}$ is a strongly continuous contraction semigroup on L .

Let $\{S(t)\}$ be a strongly continuous semigroup on L such that (1.7) holds, and define the norm $\|\cdot\|$ on L by

$$(1.8) \quad \|\cdot\| = \sup_{t \geq 0} \|S(t)f\|.$$

Then $\|f\| \leq \|\cdot\| \leq M\|f\|$ for each $f \in L$, so the new norm is equivalent to the original norm; also, with respect to $\|\cdot\|$, $\{S(t)\}$ is a strongly continuous contraction semigroup on L .

Most of the results in the subsequent sections of this chapter are stated in terms of strongly continuous contraction semigroups. Using these reductions, however, many of them can be reformulated in terms of noncontraction semigroups. \square

A (possibly unbounded) *linear operator* A on L is a linear mapping whose domain $\mathcal{D}(A)$ is a subspace of L and whose range $\mathcal{R}(A)$ lies in L . The *graph* of A is given by

$$(1.9) \quad \mathcal{G}(A) = \{(f, Af) : f \in \mathcal{D}(A)\} \subset L \times L.$$

Note that $L \times L$ is itself a Banach space with componentwise addition and scalar multiplication and norm $\|(f, g)\| = \|f\| + \|g\|$. A is said to be *closed* if $\mathcal{G}(A)$ is a closed subspace of $L \times L$.

The (*infinitesimal*) *generator* of a semigroup $\{T(t)\}$ on L is the linear operator A defined by

$$(1.10) \quad Af = \lim_{t \rightarrow 0} \frac{1}{t} \{T(t)f - f\}.$$

The domain $\mathcal{D}(A)$ of A is the subspace of all $f \in L$ for which this limit exists.

Before indicating some of the properties of generators, we briefly discuss the calculus of Banach space-valued functions.

Let Δ be a closed interval in $(-\infty, \infty)$, and denote by $C_L(\Delta)$ the space of continuous functions $u: \Delta \rightarrow L$. Let $C_L^1(\Delta)$ be the space of continuously differentiable functions $u: \Delta \rightarrow L$.

If Δ is the finite interval $[a, b]$, $u: \Delta \rightarrow L$ is said to be (*Riemann*) *integrable* over Δ if $\lim_{\delta \rightarrow 0} \sum_{k=1}^n u(s_k)(t_k - t_{k-1})$ exists, where $a = t_0 \leq s_1 \leq t_1 \leq \dots \leq t_{n-1} \leq s_n \leq t_n = b$ and $\delta = \max(t_k - t_{k-1})$; the limit is denoted by $\int_a^b u(t) dt$ or $\int_a^b u(t) dt$. If $\Delta = [a, \infty)$, $u: \Delta \rightarrow L$ is said to be *integrable* over Δ if $u|_{[a, b]}$ is integrable over $[a, b]$ for each $b \geq a$ and $\lim_{b \rightarrow \infty} \int_a^b u(t) dt$ exists; again, the limit is denoted by $\int_a^\infty u(t) dt$ or $\int_a^\infty u(t) dt$.

We leave the proof of the following lemma to the reader (Problem 3).

1.4 Lemma (a) If $u \in C_L(\Delta)$ and $\int_{\Delta} \|u(t)\| dt < \infty$, then u is integrable over Δ and

$$(1.11) \quad \left\| \int_{\Delta} u(t) dt \right\| \leq \int_{\Delta} \|u(t)\| dt.$$

In particular, if Δ is the finite interval $[a, b]$, then every function in $C_L(\Delta)$ is integrable over Δ .

(b) Let B be a closed linear operator on L . Suppose that $u \in C_L(\Delta)$, $u(t) \in \mathcal{D}(B)$ for all $t \in \Delta$, $Bu \in C_L(\Delta)$, and both u and Bu are integrable over Δ . Then $\int_{\Delta} u(t) dt \in \mathcal{D}(B)$ and

$$(1.12) \quad B \int_{\Delta} u(t) dt = \int_{\Delta} Bu(t) dt.$$

(c) If $u \in C_L^1[a, b]$, then

$$(1.13) \quad \int_a^b \frac{d}{dt} u(t) dt = u(b) - u(a).$$

1.5 Proposition Let $\{T(t)\}$ be a strongly continuous semigroup on L with generator A .

(a) If $f \in L$ and $t \geq 0$, then $\int_0^t T(s)f ds \in \mathcal{D}(A)$ and

$$(1.14) \quad T(t)f - f = A \int_0^t T(s)f ds.$$

(b) If $f \in \mathcal{D}(A)$ and $t \geq 0$, then $T(t)f \in \mathcal{D}(A)$ and

$$(1.15) \quad \frac{d}{dt} T(t)f = AT(t)f = T(t)Af.$$

(c) If $f \in \mathcal{D}(A)$ and $t \geq 0$, then

$$(1.16) \quad T(t)f - f = \int_0^t AT(s)f ds = \int_0^t T(s)Af ds.$$

Proof. (a) Observe that

$$\begin{aligned} (1.17) \quad \frac{1}{h} [T(h) - I] \int_0^t T(s)f ds &= \frac{1}{h} \int_0^t [T(s+h)f - T(s)f] ds \\ &= \frac{1}{h} \left\{ \int_h^{t+h} T(s)f ds - \int_0^t T(s)f ds \right\} \\ &= \frac{1}{h} \int_t^{t+h} T(s)f ds - \frac{1}{h} \int_0^h T(s)f ds \end{aligned}$$

for all $h > 0$, and as $h \rightarrow 0$ the right side of (1.17) converges to $T(t)f - f$.

(b) Since

$$(1.18) \quad \frac{1}{h} [T(t+h)f - T(t)f] = A_h T(t)f = T(t)A_h f$$

for all $h > 0$, where $A_h = h^{-1}[T(h) - I]$, it follows that $T(t)f \in \mathcal{D}(A)$ and $(d/dt)^+ T(t)f = AT(t)f = T(t)Af$. Thus, it suffices to check that $(d/dt)^- T(t)f = T(t)Af$ (assuming $t > 0$). But this follows from the identity

$$(1.19) \quad \begin{aligned} \frac{1}{-h} [T(t-h)f - T(t)f] &= T(t)Af \\ &= T(t-h)[A_h - A]f + [T(t-h) - T(t)]Af, \end{aligned}$$

valid for $0 < h \leq t$.

(c) This is a consequence of (b) and Lemma 1.4(c). \square

1.6 Corollary If A is the generator of a strongly continuous semigroup $\{T(t)\}$ on L , then $\mathcal{D}(A)$ is dense in L and A is closed.

Proof. Since $\lim_{t \rightarrow 0^+} t^{-1} \int_0^t T(s)f ds = f$ for every $f \in L$, Proposition 1.5(a) implies that $\mathcal{D}(A)$ is dense in L . To show that A is closed, let $\{f_n\} \subset \mathcal{D}(A)$ satisfy $f_n \rightarrow f$ and $Af_n \rightarrow g$. Then $T(t)f_n - f_n = \int_0^t T(s)Af_n ds$ for each $t > 0$, so, letting $n \rightarrow \infty$, we find that $T(t)f - f = \int_0^t T(s)g ds$. Dividing by t and letting $t \rightarrow 0$, we conclude that $f \in \mathcal{D}(A)$ and $Af = g$. \square

2. THE HILLE-YOSIDA THEOREM

Let A be a closed linear operator on L . If, for some real λ , $\lambda - A$ ($\equiv \lambda I - A$) is one-to-one, $\mathcal{D}(\lambda - A) = L$, and $(\lambda - A)^{-1}$ is a bounded linear operator on L , then λ is said to belong to the *resolvent set* $\rho(A)$ of A , and $R_\lambda = (\lambda - A)^{-1}$ is called the *resolvent* (at λ) of A .

2.1 Proposition Let $\{T(t)\}$ be a strongly continuous contraction semigroup on L with generator A . Then $(0, \infty) \subset \rho(A)$ and

$$(2.1) \quad (\lambda - A)^{-1}g = \int_0^\infty e^{-\lambda t} T(t)g dt$$

for all $g \in L$ and $\lambda > 0$.

Proof. Let $\lambda > 0$ be arbitrary. Define U_λ on L by $U_\lambda g = \int_0^\infty e^{-\lambda t} T(t)g dt$. Since

$$(2.2) \quad \|U_\lambda g\| \leq \int_0^\infty e^{-\lambda t} \|T(t)g\| dt \leq \lambda^{-1} \|g\|$$

for each $g \in L$, U_λ is a bounded linear operator on L . Now given $g \in L$,

$$(2.3) \quad \begin{aligned} \frac{1}{h} [T(h) - I] U_\lambda g &= \frac{1}{h} \int_0^\infty e^{-\lambda t} [T(t+h)g - T(t)g] dt \\ &= \frac{e^{\lambda h} - 1}{h} \int_0^\infty e^{-\lambda t} T(t)g dt - \frac{e^{\lambda h}}{h} \int_0^h e^{-\lambda t} T(t)g dt \end{aligned}$$

for every $h > 0$, so, letting $h \rightarrow 0$, we find that $U_\lambda g \in \mathcal{D}(A)$ and $AU_\lambda g = \lambda U_\lambda g - g$, that is,

$$(2.4) \quad (\lambda - A)U_\lambda g = g, \quad g \in L.$$

In addition, if $g \in \mathcal{D}(A)$, then (using Lemma 1.4(b))

$$(2.5) \quad \begin{aligned} U_\lambda Ag &= \int_0^\infty e^{-\lambda t} T(t)Ag dt = \int_0^\infty A(e^{-\lambda t} T(t)g) dt \\ &= A \int_0^\infty e^{-\lambda t} T(t)g dt = AU_\lambda g, \end{aligned}$$

so

$$(2.6) \quad U_\lambda(\lambda - A)g = g, \quad g \in \mathcal{D}(A).$$

By (2.6), $\lambda - A$ is one-to-one, and by (2.4), $\mathcal{R}(\lambda - A) = L$. Also, $(\lambda - A)^{-1} = U_\lambda$ by (2.4) and (2.6), so $\lambda \in \rho(A)$. Since $\lambda > 0$ was arbitrary, the proof is complete. \square

Let A be a closed linear operator on L . Since $(\lambda - A)(\mu - A) = (\mu - A)(\lambda - A)$ for all $\lambda, \mu \in \rho(A)$, we have $(\mu - A)^{-1}(\lambda - A)^{-1} = (\lambda - A)^{-1}(\mu - A)^{-1}$, and a simple calculation gives the *resolvent identity*

$$(2.7) \quad R_\lambda R_\mu = R_\mu R_\lambda = (\lambda - \mu)^{-1}(R_\mu - R_\lambda), \quad \lambda, \mu \in \rho(A).$$

If $\lambda \in \rho(A)$ and $|\lambda - \mu| < \|R_\lambda\|^{-1}$, then

$$(2.8) \quad \sum_{n=0}^{\infty} (\lambda - \mu)^n R_\lambda^{n+1}$$

defines a bounded linear operator that is in fact $(\mu - A)^{-1}$. In particular, this implies that $\rho(A)$ is open in \mathbb{R} .

A linear operator A on L is said to be *dissipative* if $\|\lambda f - Af\| \geq \lambda \|f\|$ for every $f \in \mathcal{D}(A)$ and $\lambda > 0$.

2.2 Lemma Let A be a dissipative linear operator on L and let $\lambda > 0$. Then A is closed if and only if $\mathcal{R}(\lambda - A)$ is closed.

Proof. Suppose A is closed. If $\{f_n\} \subset \mathcal{D}(A)$ and $(\lambda - A)f_n \rightarrow h$, then the dissipativity of A implies that $\{f_n\}$ is Cauchy. Thus, there exists $f \in L$ such that

$f_n \rightarrow f$, and hence $Af_n \rightarrow \lambda f - h$. Since A is closed, $f \in \mathcal{D}(A)$ and $h = (\lambda - A)f$. It follows that $\mathcal{R}(\lambda - A)$ is closed.

Suppose $\mathcal{R}(\lambda - A)$ is closed. If $\{f_n\} \subset \mathcal{D}(A)$, $f_n \rightarrow f$, and $Af_n \rightarrow g$, then $(\lambda - A)f_n \rightarrow \lambda f - g$, which equals $(\lambda - A)f_0$ for some $f_0 \in \mathcal{D}(A)$. By the dissipativity of A , $f_n \rightarrow f_0$, and hence $f = f_0 \in \mathcal{D}(A)$ and $Af = g$. Thus, A is closed. \square

2.3 Lemma Let A be a dissipative closed linear operator on L , and put $\rho^+(A) = \rho(A) \cap (0, \infty)$. If $\rho^+(A)$ is nonempty, then $\rho^+(A) = (0, \infty)$.

Proof. It suffices to show that $\rho^+(A)$ is both open and closed in $(0, \infty)$. Since $\rho(A)$ is necessarily open in \mathbb{R} , $\rho^+(A)$ is open in $(0, \infty)$. Suppose that $\{\lambda_n\} \subset \rho^+(A)$ and $\lambda_n \rightarrow \lambda > 0$. Given $g \in L$, let $g_n = (\lambda - A)(\lambda_n - A)^{-1}g$ for each n , and note that, because A is dissipative,

$$(2.9) \quad \lim_{n \rightarrow \infty} \|g_n - g\| = \lim_{n \rightarrow \infty} \|(\lambda - \lambda_n)(\lambda_n - A)^{-1}g\| \leq \lim_{n \rightarrow \infty} \frac{|\lambda - \lambda_n|}{\lambda_n} \|g\| = 0.$$

Hence $\mathcal{R}(\lambda - A)$ is dense in L , but because A is closed and dissipative, $\mathcal{R}(\lambda - A)$ is closed by Lemma 2.2, and therefore $\mathcal{R}(\lambda - A) = L$. Using the dissipativity of A once again, we conclude that $\lambda - A$ is one-to-one and $\|(\lambda - A)^{-1}\| \leq \lambda^{-1}$. It follows that $\lambda \in \rho^+(A)$, so $\rho^+(A)$ is closed in $(0, \infty)$, as required. \square

2.4 Lemma Let A be a dissipative closed linear operator on L , and suppose that $\mathcal{D}(A)$ is dense in L and $(0, \infty) \subset \rho(A)$. Then the Yosida approximation A_λ of A , defined for each $\lambda > 0$ by $A_\lambda = \lambda A(\lambda - A)^{-1}$, has the following properties:

- (a) For each $\lambda > 0$, A_λ is a bounded linear operator on L and $\{e^{tA_\lambda}\}$ is a strongly continuous contraction semigroup on L .
- (b) $A_\lambda A_\mu = A_\mu A_\lambda$ for all $\lambda, \mu > 0$.
- (c) $\lim_{\lambda \rightarrow \infty} A_\lambda f = Af$ for every $f \in \mathcal{D}(A)$.

Proof. For each $\lambda > 0$, let $R_\lambda = (\lambda - A)^{-1}$ and note that $\|R_\lambda\| \leq \lambda^{-1}$. Since $(\lambda - A)R_\lambda = I$ on L and $R_\lambda(\lambda - A) = I$ on $\mathcal{D}(A)$, it follows that

$$(2.10) \quad A_\lambda = \lambda^2 R_\lambda - \lambda I \quad \text{on } L, \quad \lambda > 0,$$

and

$$(2.11) \quad A_\lambda = \lambda R_\lambda A \quad \text{on } \mathcal{D}(A), \quad \lambda > 0.$$

By (2.10), we find that, for each $\lambda > 0$, A_λ is bounded and

$$(2.12) \quad \|e^{tA_\lambda}\| = e^{-t\lambda} \|e^{t\lambda^2 R_\lambda}\| \leq e^{-t\lambda} e^{t\lambda^2 \|R_\lambda\|} \leq 1$$

for all $t \geq 0$, proving (a). Conclusion (b) is a consequence of (2.10) and (2.7). As for (c), we claim first that

$$(2.13) \quad \lim_{\lambda \rightarrow \infty} \lambda R_\lambda f = f, \quad f \in L.$$

Noting that $\|\lambda R_\lambda f - f\| = \|R_\lambda Af\| \leq \lambda^{-1} \|Af\| \rightarrow 0$ as $\lambda \rightarrow \infty$ for each $f \in \mathcal{D}(A)$, (2.13) follows from the facts that $\mathcal{D}(A)$ is dense in L and $\|\lambda R_\lambda - I\| \leq 2$ for all $\lambda > 0$. Finally, (c) is a consequence of (2.11) and (2.13). \square

2.5 Lemma If B and C are bounded linear operators on L such that $BC = CB$ and $\|e^{tB}\| \leq 1$ and $\|e^{tC}\| \leq 1$ for all $t \geq 0$, then

$$(2.14) \quad \|e^{tB}f - e^{tC}f\| \leq t \|Bf - Cf\|$$

for every $f \in L$ and $t \geq 0$.

Proof. The result follows from the identity

$$(2.15) \quad \begin{aligned} e^{tB}f - e^{tC}f &= \int_0^t \frac{d}{ds} [e^{sB}e^{(t-s)C}]f ds = \int_0^t e^{sB}(B - C)e^{(t-s)C}f ds \\ &= \int_0^t e^{sB}e^{(t-s)C}(B - C)f ds. \end{aligned}$$

(Note that the last equality uses the commutivity of B and C). \square

We are now ready to prove the Hille-Yosida theorem.

2.6 Theorem A linear operator A on L is the generator of a strongly continuous contraction semigroup on L if and only if:

- (a) $\mathcal{D}(A)$ is dense in L .
- (b) A is dissipative.
- (c) $\mathcal{R}(\lambda - A) = L$ for some $\lambda > 0$.

Proof. The necessity of the conditions (a)-(c) follows from Corollary 1.6 and Proposition 2.1. We therefore turn to the proof of sufficiency.

By (b), (c), and Lemma 2.2, A is closed and $\rho(A) \cap (0, \infty)$ is nonempty, so by Lemma 2.3, $(0, \infty) \subset \rho(A)$. Using the notation of Lemma 2.4, we define for each $\lambda > 0$ the strongly continuous contraction semigroup $\{T_\lambda(t)\}$ on L by $T_\lambda(t) = e^{tA_\lambda}$. By Lemmas 2.4(b) and 2.5,

$$(2.16) \quad \|T_\lambda(t)f - T_\mu(t)f\| \leq t \|A_\lambda f - A_\mu f\|$$

for all $f \in L$, $t \geq 0$, and $\lambda, \mu > 0$. Thus, by Lemma 2.4(c), $\lim_{\lambda \rightarrow \infty} T_\lambda(t)f$ exists for all $t \geq 0$, uniformly on bounded intervals, for all $f \in \mathcal{D}(A)$, hence for every $f \in \overline{\mathcal{D}(A)} = L$. Denoting the limit by $T(t)f$ and using the identity

$$(2.17) \quad \begin{aligned} T(s+t)f - T(s)T(t)f &= [T(s+t) - T_\lambda(s+t)]f \\ &\quad + T_\lambda(s)[T_\lambda(t) - T(t)]f + [T_\lambda(s) - T(s)]T(t)f, \end{aligned}$$

we conclude that $\{T(t)\}$ is a strongly continuous contraction semigroup on L .

It remains only to show that A is the generator of $\{T(t)\}$. By Proposition 1.5(c),

$$(2.18) \quad T_\lambda(t)f - f = \int_0^t T_\lambda(s)A_\lambda f \, ds$$

for all $f \in L$, $t \geq 0$, and $\lambda > 0$. For each $f \in \mathcal{D}(A)$ and $t \geq 0$, the identity

$$(2.19) \quad T_\lambda(s)A_\lambda f - T(s)Af = T_\lambda(s)(A_\lambda f - Af) + [T_\lambda(s) - T(s)]Af,$$

together with Lemma 2.4(c), implies that $T_\lambda(s)A_\lambda f \rightarrow T(s)Af$ as $\lambda \rightarrow \infty$, uniformly in $0 \leq s \leq t$. Consequently, (2.18) yields

$$(2.20) \quad T(t)f - f = \int_0^t T(s)Af \, ds$$

for all $f \in \mathcal{D}(A)$ and $t \geq 0$. From this we find that the generator B of $\{T(t)\}$ is an extension of A . But, for each $\lambda > 0$, $\lambda - B$ is one-to-one by the necessity of (b), and $\mathcal{R}(\lambda - A) = L$ since $\lambda \in \rho(A)$. We conclude that $B = A$, completing the proof. \square

The above proof and Proposition 2.9 below yield the following result as a by-product.

2.7 Proposition Let $\{T(t)\}$ be a strongly continuous contraction semigroup on L with generator A , and let A_λ be the Yosida approximation of A (defined in Lemma 2.4). Then

$$(2.21) \quad \|e^{tA_\lambda}f - T(t)f\| \leq t \|A_\lambda f - Af\|, \quad f \in \mathcal{D}(A), t \geq 0, \lambda > 0,$$

so, for each $f \in L$, $\lim_{\lambda \rightarrow \infty} e^{tA_\lambda}f = T(t)f$ for all $t \geq 0$, uniformly on bounded intervals.

2.8 Corollary Let $\{T(t)\}$ be a strongly continuous contraction semigroup on L with generator A . For $M \subset L$, let

$$(2.22) \quad \Lambda_M = \{\lambda > 0: \lambda(\lambda - A)^{-1}: M \rightarrow M\}.$$

If either (a) M is a closed convex subset of L and Λ_M is unbounded, or (b) M is a closed subspace of L and Λ_M is nonempty, then

$$(2.23) \quad T(t): M \rightarrow M, \quad t \geq 0.$$

Proof. If $\lambda, \mu > 0$ and $|1 - \mu/\lambda| < 1$, then (cf. (2.8))

$$(2.24) \quad \mu(\mu - A)^{-1} = \sum_{n=0}^{\infty} \frac{\mu}{\lambda} \left(1 - \frac{\mu}{\lambda}\right)^n [\lambda(\lambda - A)^{-1}]^{n+1}.$$

Consequently, if M is a closed convex subset of L , then $\lambda \in \Lambda_M$ implies $(0, \lambda] \subset \Lambda_M$, and if M is a closed subspace of L , then $\lambda \in \Lambda_M$ implies $(0, 2\lambda) \subset \Lambda_M$. Therefore, under either (a) or (b), we have $\Lambda_M = (0, \infty)$. Finally, by (2.10),

$$(2.25) \quad \begin{aligned} \exp\{tA_\lambda\} &= \exp\{-t\lambda\} \exp\{t\lambda[\lambda(\lambda - A)^{-1}]\} \\ &= e^{-t\lambda} \sum_{n=0}^{\infty} \frac{(t\lambda)^n}{n!} [\lambda(\lambda - A)^{-1}]^n \end{aligned}$$

for all $t \geq 0$ and $\lambda > 0$, so the conclusion follows from Proposition 2.7. \square

2.9 Proposition Let $\{T(t)\}$ and $\{S(t)\}$ be strongly continuous contraction semigroups on L with generators A and B , respectively. If $A = B$, then $T(t) = S(t)$ for all $t \geq 0$.

Proof. This result is a consequence of the next proposition. \square

2.10 Proposition Let A be a dissipative linear operator on L . Suppose that $u: [0, \infty) \rightarrow L$ is continuous, $u(t) \in \mathcal{D}(A)$ for all $t > 0$, $Au: (0, \infty) \rightarrow L$ is continuous, and

$$(2.26) \quad u(t) = u(\varepsilon) + \int_{\varepsilon}^t Au(s) ds,$$

for all $t > \varepsilon > 0$. Then $\|u(t)\| \leq \|u(0)\|$ for all $t \geq 0$.

Proof. Let $0 < \varepsilon = t_0 < t_1 < \cdots < t_n = t$. Then

(2.27)

$$\begin{aligned} \|u(t)\| &= \|u(\varepsilon)\| + \sum_{i=1}^n [\|u(t_i)\| - \|u(t_{i-1})\|] \\ &= \|u(\varepsilon)\| + \sum_{i=1}^n [\|u(t_i)\| - \|u(t_i) - (t_i - t_{i-1})Au(t_i)\|] \\ &\quad + \sum_{i=1}^n [\|u(t_i) - (t_i - t_{i-1})Au(t_i)\| - \|u(t_i) - (u(t_i) - u(t_{i-1}))\|] \\ &\leq \|u(\varepsilon)\| + \sum_{i=1}^n \left[\|u(t_i) - (t_i - t_{i-1})Au(t_i)\| - \left\| u(t_i) - \int_{t_{i-1}}^t Au(s) ds \right\| \right] \\ &\leq \|u(\varepsilon)\| + \sum_{i=1}^n \int_{t_{i-1}}^t \|Au(t_i) - Au(s)\| ds, \end{aligned}$$

where the first inequality is due to the dissipativity of A . The result follows from the continuity of Au and u by first letting $\max(t_i - t_{i-1}) \rightarrow 0$ and then letting $\varepsilon \rightarrow 0$. \square

In many applications, an alternative form of the Hille–Yosida theorem is more useful. To state it, we need two definitions and a lemma.

A linear operator A on L is said to be *closable* if it has a closed linear extension. If A is closable, then the *closure* \bar{A} of A is the minimal closed linear extension of A ; more specifically, it is the closed linear operator B whose graph is the closure (in $L \times L$) of the graph of A .

2.11 Lemma Let A be a dissipative linear operator on L with $\mathcal{D}(A)$ dense in L . Then A is closable and $\mathcal{R}(\lambda - A) = \mathcal{R}(\lambda - \bar{A})$ for every $\lambda > 0$.

Proof. For the first assertion, it suffices to show that if $\{f_n\} \subset \mathcal{D}(A)$, $f_n \rightarrow 0$, and $Af_n \rightarrow g \in L$, then $g = 0$. Choose $\{g_m\} \subset \mathcal{D}(A)$ such that $g_m \rightarrow g$. By the dissipativity of A ,

$$\begin{aligned} (2.28) \quad \|(\lambda - A)g_m - \lambda g\| &= \lim_{n \rightarrow \infty} \|(\lambda - A)(g_m + \lambda f_n)\| \\ &\geq \lim_{n \rightarrow \infty} \lambda \|g_m + \lambda f_n\| = \lambda \|g_m\| \end{aligned}$$

for every $\lambda > 0$ and each m . Dividing by λ and letting $\lambda \rightarrow \infty$, we find that $\|g_m - g\| \geq \|g_m\|$ for each m . Letting $m \rightarrow \infty$, we conclude that $g = 0$.

Let $\lambda > 0$. The inclusion $\mathcal{R}(\lambda - A) \supset \mathcal{R}(\lambda - \bar{A})$ is obvious, so to prove equality, we need only show that $\mathcal{R}(\lambda - \bar{A})$ is closed. But this is an immediate consequence of Lemma 2.2. \square

2.12 Theorem A linear operator A on L is closable and its closure \bar{A} is the generator of a strongly continuous contraction semigroup on L if and only if:

- (a) $\mathcal{D}(A)$ is dense in L .
- (b) A is dissipative.
- (c) $\mathcal{R}(\lambda - A)$ is dense in L for some $\lambda > 0$.

Proof. By Lemma 2.11, A satisfies (a)–(c) above if and only if A is closable and \bar{A} satisfies (a)–(c) of Theorem 2.6. \square

3. CORES

In this section we introduce a concept that is of considerable importance in Sections 6 and 7.

Let A be a closed linear operator on L . A subspace D of $\mathcal{D}(A)$ is said to be a *core* for A if the closure of the restriction of A to D is equal to A (i.e., if $\overline{A|_D} = A$).

3.1 Proposition Let A be the generator of a strongly continuous contraction semigroup on L . Then a subspace D of $\mathcal{D}(A)$ is a core for A if and only if D is dense in L and $\mathcal{R}(\lambda - A|_D)$ is dense in L for some $\lambda > 0$.

3.2 Remark A subspace of L is dense in L if and only if it is weakly dense (Rudin (1973), Theorem 3.12). \square

Proof. The sufficiency follows from Theorem 2.12 and from the observation that, if A and B generate strongly continuous contraction semigroups on L and if A is an extension of B , then $A = B$. The necessity depends on Lemma 2.11. \square

3.3 Proposition Let A be the generator of a strongly continuous contraction semigroup $\{T(t)\}$ on L . Let D_0 and D be dense subspaces of L with $D_0 \subset D \subset \mathcal{D}(A)$. (Usually, $D_0 = D$.) If $T(t): D_0 \rightarrow D$ for all $t \geq 0$, then D is a core for A .

Proof. Given $f \in D_0$ and $\lambda > 0$,

$$(3.1) \quad f_n \equiv \frac{1}{n} \sum_{k=0}^{n^2} e^{-\lambda k/n} T\left(\frac{k}{n}\right) f \in D$$

for $n = 1, 2, \dots$. By the strong continuity of $\{T(t)\}$ and Proposition 2.1,

$$\begin{aligned} (3.2) \quad \lim_{n \rightarrow \infty} (\lambda - A)f_n &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n^2} e^{-\lambda k/n} T\left(\frac{k}{n}\right)(\lambda - A)f \\ &= \int_0^\infty e^{-\lambda t} T(t)(\lambda - A)f dt \\ &= (\lambda - A)^{-1}(\lambda - A)f = f. \end{aligned}$$

so $\overline{\mathcal{R}(\lambda - A|_D)} \supset D_0$. This suffices by Proposition 3.1 since D_0 is dense in L . \square

Given a dissipative linear operator A with $\mathcal{D}(A)$ dense in L , one often wants to show that \tilde{A} generates a strongly continuous contraction semigroup on L . By Theorem 2.12, a necessary and sufficient condition is that $\mathcal{R}(\lambda - A)$ be dense in L for some $\lambda > 0$. We can view this problem as one of characterizing a core (namely, $\mathcal{D}(A)$) for the generator of a strongly continuous contraction semigroup, except that, unlike the situation in Propositions 3.1 and 3.3, the generator is not provided in advance. Thus, the remainder of this section is primarily concerned with verifying the range condition (condition (c)) of Theorem 2.12.

Observe that the following result generalizes Proposition 3.3.

3.4 Proposition Let A be a dissipative linear operator on L , and D_0 a subspace of $\mathcal{D}(A)$ that is dense in L . Suppose that, for each $f \in D_0$, there exists a continuous function $u_f: [0, \infty) \rightarrow L$ such that $u_f(0) = f$, $u_f(t) \in \mathcal{D}(A)$ for all $t > 0$, $Au_f: (0, \infty) \rightarrow L$ is continuous, and

$$(3.3) \quad u_f(t) - u_f(\varepsilon) = \int_{\varepsilon}^t Au_f(s) ds$$

for all $t > \varepsilon > 0$. Then A is closable, the closure of A generates a strongly continuous contraction semigroup $\{T(t)\}$ on L , and $T(t)f = u_f(t)$ for all $f \in D_0$ and $t \geq 0$.

Proof. By Lemma 2.11, A is closable. Fix $f \in D_0$ and denote u_f by u . Let $t_0 > \varepsilon > 0$, and note that $\int_{\varepsilon}^{t_0} e^{-t} u(t) dt \in \mathcal{D}(\bar{A})$ and

$$(3.4) \quad \bar{A} \int_{\varepsilon}^{t_0} e^{-t} u(t) dt = \int_{\varepsilon}^{t_0} e^{-t} Au(t) dt.$$

Consequently,

$$\begin{aligned} (3.5) \quad \int_{\varepsilon}^{t_0} e^{-t} u(t) dt &= (e^{-\varepsilon} - e^{-t_0})u(\varepsilon) + \int_{\varepsilon}^{t_0} e^{-t} \int_{\varepsilon}^t Au(s) ds dt \\ &= (e^{-\varepsilon} - e^{-t_0})u(\varepsilon) + \int_{\varepsilon}^{t_0} (e^{-s} - e^{-t_0})Au(s) ds \\ &= \bar{A} \int_{\varepsilon}^{t_0} e^{-t} u(t) dt + e^{-\varepsilon}u(\varepsilon) - e^{-t_0}u(t_0). \end{aligned}$$

Since $\|u(t)\| \leq \|f\|$ for all $t \geq 0$ by Proposition 2.10, we can let $\varepsilon \rightarrow 0$ and $t_0 \rightarrow \infty$ in (3.5) to obtain $\int_0^{\infty} e^{-t} u(t) dt \in \mathcal{D}(\bar{A})$ and

$$(3.6) \quad (1 - \bar{A}) \int_0^{\infty} e^{-t} u(t) dt = f.$$

We conclude that $\mathcal{D}(1 - \bar{A}) \supset D_0$, which by Theorem 2.6 proves that \bar{A} generates a strongly continuous contraction semigroup $\{T(t)\}$ on L . Now for each $f \in D_0$,

$$(3.7) \quad T(t)f - T(\varepsilon)f = \int_{\varepsilon}^t \bar{A}T(s)f ds$$

for all $t > \varepsilon > 0$. Subtracting (3.3) from this and applying Proposition 2.10 once again, we obtain the second conclusion of the proposition. \square

The next result shows that a sufficient condition for \bar{A} to generate is that A be triangulizable. Of course, this is a very restrictive assumption, but it is occasionally satisfied.

3.5 Proposition Let A be a dissipative linear operator on L , and suppose that L_1, L_2, L_3, \dots is a sequence of finite-dimensional subspaces of $\mathcal{D}(A)$ such that $\bigcup_{n=1}^{\infty} L_n$ is dense in L . If $A: L_n \rightarrow L_n$ for $n = 1, 2, \dots$, then A is closable and the closure of A generates a strongly continuous contraction semigroup on L .

Proof. For $n = 1, 2, \dots$, $(\lambda - A)(L_n) = L_n$ for all λ not belonging to the set of eigenvalues of $A|_{L_n}$, hence for all but at most finitely many $\lambda > 0$. Consequently, $(\lambda - A)\left(\bigcup_{n=1}^{\infty} L_n\right) = \bigcup_{n=1}^{\infty} L_n$ for all but at most countably many $\lambda > 0$ and in particular for some $\lambda > 0$. Thus, the conditions of Theorem 2.12 are satisfied. \square

We turn next to a generalization of Proposition 3.3 in a different direction. The idea is to try to approximate A sufficiently well by a sequence of generators for which the conditions of Proposition 3.3 are satisfied. Before stating the result we record the following simple but frequently useful lemma.

3.6 Lemma Let A_1, A_2, \dots and A be linear operators on L , D_0 a subspace of L , and $\lambda > 0$. Suppose that, for each $g \in D_0$, there exists $f_n \in \mathcal{D}(A_n) \cap \mathcal{D}(A)$ for $n = 1, 2, \dots$ such that $g_n \equiv (\lambda - A_n)f_n \rightarrow g$ as $n \rightarrow \infty$ and

$$(3.8) \quad \lim_{n \rightarrow \infty} \|(\lambda - A)f_n\| = 0.$$

Then $\overline{\mathcal{R}(\lambda - A)} \supset D_0$.

Proof. Given $g \in D_0$, choose $\{f_n\}$ and $\{g_n\}$ as in the statement of the lemma, and observe that $\lim_{n \rightarrow \infty} \|(\lambda - A)f_n - g_n\| = 0$ by (3.8). It follows that $\lim_{n \rightarrow \infty} \|(\lambda - A)f_n - g\| = 0$, giving the desired result. \square

3.7 Proposition Let A be a linear operator on L and D_0 and D_1 dense subspaces of L satisfying $D_0 \subset \mathcal{D}(A) \subset D_1 \subset L$. Let $\|\cdot\|$ be a norm on D_1 . For $n = 1, 2, \dots$, suppose that A_n generates a strongly continuous contraction semigroup $\{T_n(t)\}$ on L and $\mathcal{D}(A) \subset \mathcal{D}(A_n)$. Suppose further that there exist $\omega \geq 0$ and a sequence $\{\varepsilon_n\} \subset (0, \infty)$ tending to zero such that, for $n = 1, 2, \dots$,

$$(3.9) \quad \|(\lambda - A)f\| \leq \varepsilon_n \|f\|, \quad f \in \mathcal{D}(A),$$

$$(3.10) \quad T_n(t): D_1 \rightarrow D_1, \quad \|T_n(t)|_{D_1}\| \leq e^{\omega t}, \quad t \geq 0,$$

and

$$(3.11) \quad T_n(t): D_0 \rightarrow \mathcal{D}(A), \quad t \geq 0.$$

Then A is closable and the closure of A generates a strongly continuous contraction semigroup on L .

Proof. Observe first that $\mathcal{D}(A)$ is dense in L and, by (3.9) and the dissipativity of each A_n , A is dissipative. It therefore suffices to verify condition (c) of Theorem 2.12.

Fix $\lambda > \omega$. Given $g \in D_0$, let

$$(3.12) \quad f_{m,n} = \frac{1}{m} \sum_{k=0}^{m^2} e^{-\lambda k/m} T_n \left(\frac{k}{m} \right) g \in \mathcal{D}(A)$$

for each $m, n \geq 1$ (cf. (3.1)). Then, for $n = 1, 2, \dots$, $(\lambda - A_n)f_{m,n} \rightarrow \int_0^\infty e^{-\lambda t} T_n(t)(\lambda - A_n)g dt = g$ as $m \rightarrow \infty$, so there exists a sequence $\{m_n\}$ of positive integers such that $(\lambda - A_n)f_{m_n,n} \rightarrow g$ as $n \rightarrow \infty$. Moreover,

$$(3.13) \quad \begin{aligned} \|(\lambda - A_n)f_{m_n,n}\| &\leq \varepsilon_n \|f_{m_n,n}\| \\ &\leq \varepsilon_n m_n^{-1} \sum_{k=0}^{m_n^2} e^{-\lambda k/m_n} e^{\omega k/m_n} \|g\| \\ &\rightarrow 0 \quad \text{as } n \rightarrow \infty \end{aligned}$$

by (3.9) and (3.10), so Lemma 3.6 gives the desired conclusion. \square

3.8 Corollary Let A be a linear operator on L with $\mathcal{D}(A)$ dense in L , and let $\|\cdot\|$ be a norm on $\mathcal{D}(A)$ with respect to which $\mathcal{D}(A)$ is a Banach space. For $n = 1, 2, \dots$, let T_n be a linear $\|\cdot\|$ -contraction on L such that $T_n: \mathcal{D}(A) \rightarrow \mathcal{D}(A)$, and define $A_n = n(T_n - I)$. Suppose there exist $\omega \geq 0$ and a sequence $\{\varepsilon_n\} \subset (0, \infty)$ tending to zero such that, for $n = 1, 2, \dots$, (3.9) holds and

$$(3.14) \quad \|T_n|_{\mathcal{D}(A)}\| \leq 1 + \frac{\omega}{n}.$$

Then A is closable and the closure of A generates a strongly continuous contraction semigroup on L .

Proof. We apply Proposition 3.7 with $D_0 = D_1 = \mathcal{D}(A)$. For $n = 1, 2, \dots$, $\exp\{tA_n\}: \mathcal{D}(A) \rightarrow \mathcal{D}(A)$ and

$$(3.15) \quad \|\exp\{tA_n\}|_{\mathcal{D}(A)}\| \leq \exp\{-nt\} \exp\{nt\|T_n|_{\mathcal{D}(A)}\|\} \leq \exp\{\omega t\}$$

for all $t \geq 0$, so the hypotheses of the proposition are satisfied. \square

4. MULTIVALUED OPERATORS

Recall that if A is a linear operator on L , then the graph $\mathcal{G}(A)$ of A is a subspace of $L \times L$ such that $(0, g) \in \mathcal{G}(A)$ implies $g = 0$. More generally, we regard an arbitrary subset A of $L \times L$ as a *multivalued operator* on L with domain $\mathcal{D}(A) = \{f: (f, g) \in A \text{ for some } g\}$ and range $\mathcal{R}(A) = \{g: (f, g) \in A \text{ for some } f\}$. $A \subset L \times L$ is said to be *linear* if A is a subspace of $L \times L$. If A is linear, then A is said to be *single-valued* if $(0, g) \in A$ implies $g = 0$; in this case,

A is a graph of a linear operator on L , also denoted by A , so we write $Ag = g$ if $(f, g) \in A$. If $A \subset L \times L$ is linear, then A is said to be *dissipative* if $\|\lambda f - g\| \geq \lambda \|f\|$ for all $(f, g) \in A$ and $\lambda > 0$; the closure \bar{A} of A is of course just the closure in $L \times L$ of the subspace A . Finally, we define $\lambda - A = \{(f, \lambda f - g) : (f, g) \in A\}$ for each $\lambda > 0$.

Observe that a (single-valued) linear operator A is closable if and only if the closure of A (in the above sense) is single-valued. Consequently, the term “closable” is no longer needed.

We begin by noting that the generator of a strongly continuous contraction semigroup is a maximal dissipative (multivalued) linear operator.

4.1 Proposition Let A be the generator of a strongly continuous contraction semigroup on L . Let $B \subset L \times L$ be linear and dissipative, and suppose that $A \subset B$. Then $A = B$.

Proof. Let $(f, g) \in B$ and $\lambda > 0$. Then $(f, \lambda f - g) \in \lambda - B$. Since $\lambda \in \rho(A)$, there exists $h \in \mathcal{D}(A)$ such that $\lambda h - Ah = \lambda f - g$. Hence $(h, \lambda f - g) \in \lambda - A \subset \lambda - B$. By linearity, $(f - h, 0) \in \lambda - B$, so by dissipativity, $f = h$. Hence $g = Ah$, so $(f, g) \in A$. \square

We turn next to an extension of Lemma 2.11.

4.2 Lemma Let $A \subset L \times L$ be linear and dissipative. Then

$$(4.1) \quad A_0 = \{(f, g) \in \bar{A} : g \in \overline{\mathcal{D}(A)}\}$$

is single-valued and $\overline{\mathcal{R}(\lambda - A)} = \mathcal{R}(\lambda - \bar{A})$ for every $\lambda > 0$.

Proof. Given $(0, g) \in A_0$, we must show that $g = 0$. By the definition of A_0 , there exists a sequence $\{(g_n, h_n)\} \subset A$ such that $g_n \rightarrow g$. For each n , $(g_n, h_n + \lambda g) \in \bar{A}$ by the linearity of \bar{A} , so $\|\lambda g_n - h_n - \lambda g\| \geq \lambda \|g_n\|$ for every $\lambda > 0$ by the dissipativity of \bar{A} . Dividing by λ and letting $\lambda \rightarrow \infty$, we find that $\|g_n - g\| \geq \|g_n\|$ for each n . Letting $n \rightarrow \infty$, we conclude that $g = 0$.

The proof of the second assertion is similar to that of the second assertion of Lemma 2.11. \square

The main result of this section is the following version of the Hille–Yosida theorem.

4.3 Theorem Let $A \subset L \times L$ be linear and dissipative, and define A_0 by (4.1). Then A_0 is the generator of a strongly continuous contraction semigroup on $\mathcal{D}(A)$ if and only if $\mathcal{R}(\lambda - A) \supset \mathcal{D}(A)$ for some $\lambda > 0$.

Proof. A_0 is single-valued by Lemma 4.2 and is clearly dissipative, so by the Hille–Yosida theorem (Theorem 2.6), A_0 generates a strongly continuous contraction semigroup on $\mathcal{D}(A)$ if and only if $\mathcal{D}(A_0)$ is dense in $\mathcal{D}(A)$ and $\mathcal{R}(\lambda - A_0) = \mathcal{D}(A)$ for some $\lambda > 0$. The latter condition is clearly equivalent to

$\mathcal{R}(\lambda - \tilde{A}) \supseteq \overline{\mathcal{D}(A)}$ for some $\lambda > 0$, which by Lemma 4.2 is equivalent to $\mathcal{R}(\lambda - A) \supseteq \mathcal{D}(A)$ for some $\lambda > 0$. Thus, to complete the proof, it suffices to show that $\mathcal{D}(A_0)$ is dense in $\overline{\mathcal{D}(A)}$ assuming that $\mathcal{R}(\lambda - A_0) = \overline{\mathcal{D}(A)}$ for some $\lambda > 0$.

By Lemma 2.3, $\mathcal{R}(\lambda - A_0) = \overline{\mathcal{D}(A)}$ for every $\lambda > 0$, so $\mathcal{R}(\lambda - \tilde{A}) = \mathcal{R}(\lambda - \tilde{A}) \supseteq \overline{\mathcal{D}(A)}$ for every $\lambda > 0$. By the dissipativity of \tilde{A} , we may regard $(\lambda - \tilde{A})^{-1}$ as a (single-valued) bounded linear operator on $\mathcal{R}(\lambda - \tilde{A})$ of norm at most λ^{-1} for each $\lambda > 0$. Given $(f, g) \in \tilde{A}$ and $\lambda > 0$, if $g \in \mathcal{R}(\lambda - \tilde{A})$ and $f \in \mathcal{D}(\tilde{A}) \subset \overline{\mathcal{D}(A)} \subset \mathcal{R}(\lambda - \tilde{A})$, so $g \in \mathcal{R}(\lambda - \tilde{A})$, and therefore $\|\lambda(\lambda - \tilde{A})^{-1}f - f\| = \|(\lambda - \tilde{A})^{-1}g\| \leq \lambda^{-1}\|g\|$. Since $\mathcal{D}(\tilde{A})$ is dense in $\overline{\mathcal{D}(A)}$, it follows that

$$(4.2) \quad \lim_{\lambda \rightarrow \infty} \lambda(\lambda - \tilde{A})^{-1}f = f, \quad f \in \overline{\mathcal{D}(A)}.$$

(Note that this does *not* follow from (2.13).) But clearly, $(\lambda - \tilde{A})^{-1}: \mathcal{R}(\lambda - A_0) \rightarrow \mathcal{D}(A_0)$, that is, $(\lambda - \tilde{A})^{-1}: \overline{\mathcal{D}(A)} \rightarrow \mathcal{D}(A_0)$, for all $\lambda > 0$. In view of (4.2), this completes the proof. \square

Multivalued operators arise naturally in several ways. For example, the following concept is crucial in Sections 6 and 7.

For $n = 1, 2, \dots$, let L_n , in addition to L , be a Banach space with norm also denoted by $\|\cdot\|$, and let $\pi_n: L \rightarrow L_n$ be a bounded linear transformation. Assume that $\sup_n \|\pi_n\| < \infty$. If $A_n \subset L_n \times L_n$ is linear for each $n \geq 1$, the *extended limit* of the sequence $\{A_n\}$ is defined by

$$(4.3) \quad \text{ex-lim}_{n \rightarrow \infty} A_n = \{(f, g) \in L \times L: \text{there exists } (f_n, g_n) \in A_n \text{ for each } n \geq 1 \text{ such that } \|f_n - \pi_n f\| \rightarrow 0 \text{ and } \|g_n - \pi_n g\| \rightarrow 0\}.$$

We leave it to the reader to show that $\text{ex-lim}_{n \rightarrow \infty} A_n$ is necessarily closed in $L \times L$ (Problem 11).

To see that $\text{ex-lim}_{n \rightarrow \infty} A_n$ need not be single-valued even if each A_n is, let $L_n = L$, $\pi_n = I$, and $A_n = B + nC$ for each $n \geq 1$, where B and C are bounded linear operators on L . If f belongs to $\mathcal{N}(C)$, the null space of C , and $h \in L$, then $A_n(f + (1/n)h) \rightarrow Bf + Ch$, so

$$(4.4) \quad \{(f, Bf + Ch): f \in \mathcal{N}(C), h \in L\} \subset \text{ex-lim}_{n \rightarrow \infty} A_n.$$

Another situation in which multivalued operators arise is described in the next section.

5. SEMIGROUPS ON FUNCTION SPACES

In this section we want to extend the notion of the generator of a semigroup, but to do so we need to be able to integrate functions $u: [0, \infty) \rightarrow L$ that are

not continuous and to which the Riemann integral of Section 1 does not apply. For our purposes, the most efficient way to get around this difficulty is to restrict the class of Banach spaces L under consideration. We therefore assume in this section that L is a "function space" that arises in the following way.

Let (M, \mathcal{M}) be a measurable space, let Γ be a collection of positive measures on \mathcal{M} , and let \mathcal{L} be the vector space of \mathcal{M} -measurable functions f such that

$$(5.1) \quad \|f\| \equiv \sup_{\mu \in \Gamma} \int |f| d\mu < \infty.$$

Note that $\|\cdot\|$ is a seminorm on \mathcal{L} but need not be a norm. Let $\mathcal{N} = \{f \in \mathcal{L} : \|f\| = 0\}$ and let L be the quotient space \mathcal{L}/\mathcal{N} , that is, L is the space of equivalence classes of functions in \mathcal{L} , where $f \sim g$ if $\|f - g\| = 0$. As is typically the case in discussions of L^p -spaces, we do not distinguish between a function in \mathcal{L} and its equivalence class in L unless necessary.

L is a Banach space, the completeness following as for L^p -spaces. In fact, if v is a σ -finite measure on \mathcal{M} , $1 \leq q \leq \infty$, $p^{-1} + q^{-1} = 1$, and

$$(5.2) \quad \Gamma = \left\{ \mu \geq 0 : \mu \ll v, \left\| \frac{d\mu}{dv} \right\|_q \leq 1 \right\},$$

where $\|\cdot\|_q$ is the norm on $L^q(v)$, then $L = L^p(v)$. Of course, if Γ is the set of probability measures on \mathcal{M} , then $L = B(M, \mathcal{M})$, the space of bounded \mathcal{M} -measurable functions on M with the sup norm.

Let (S, \mathcal{S}, v) be a σ -finite measure space, let $f: S \times M \rightarrow \mathbb{R}$ be $\mathcal{S} \times \mathcal{M}$ -measurable, and let $g: S \rightarrow [0, \infty)$ be \mathcal{S} -measurable. If $\|f(s, \cdot)\| \leq g(s)$ for all $s \in S$ and $\int g(s)v(ds) < \infty$, then

$$(5.3) \quad \begin{aligned} & \sup_{\mu \in \Gamma} \int \left| \int f(s, x)v(dx) \right| \mu(ds) \\ & \leq \sup_{\mu \in \Gamma} \iint |f(s, x)| \mu(dx)v(ds) \\ & \leq \int g(s)v(ds) < \infty, \end{aligned}$$

and we can define $\int f(s, \cdot)v(ds) \in L$ to be the equivalence class of functions in \mathcal{L} equivalent to h , where

$$(5.4) \quad h(x) = \begin{cases} \int f(s, x)v(ds), & \int |f(s, x)|v(ds) < \infty, \\ 0, & \text{otherwise.} \end{cases}$$

With the above in mind, we say that $u: S \rightarrow L$ is *measurable* if there exists an $\mathcal{S} \times \mathcal{M}$ -measurable function v such that $v(s, \cdot) \in u(s)$ for each $s \in S$. We define a semigroup $\{T(t)\}$ on L to be *measurable* if $T(\cdot)f$ is measurable as a function on $([0, \infty), \mathcal{B}[0, \infty))$ for each $f \in L$. We define the *full generator* \hat{A} of a measurable contraction semigroup $\{T(t)\}$ on L by

$$(5.5) \quad \hat{A} = \left\{ (f, g) \in L \times L : T(t)f - f = \int_0^t T(s)g \, ds, \quad t \geq 0 \right\}.$$

We note that \hat{A} is not, in general, single-valued. For example, if $L = B(\mathbb{R})$ with the sup norm and $T(t)f(x) \equiv f(x+t)$, then $(0, g) \in \hat{A}$ for each $g \in B(\mathbb{R})$ that is zero almost everywhere with respect to Lebesgue measure.

5.1 Proposition Let L be as above, and let $\{T(t)\}$ be a measurable contraction semigroup on L . Then the full generator \hat{A} of $\{T(t)\}$ is linear and dissipative and satisfies

$$(5.6) \quad (\lambda - \hat{A})^{-1}h = \int_0^\infty e^{-\lambda t} T(t)h \, dt$$

for all $h \in \mathcal{R}(\lambda - \hat{A})$ and $\lambda > 0$. If

$$(5.7) \quad T(s) \int_0^\infty e^{-\lambda t} T(t)h \, dt = \int_0^\infty e^{-\lambda t} T(s+t)h \, dt$$

for all $h \in L$, $\lambda > 0$, and $s \geq 0$, then $\mathcal{R}(\lambda - \hat{A}) = L$ for every $\lambda > 0$.

Proof. Let $(f, g) \in \hat{A}$, $\lambda > 0$, and $h = \lambda f - g$. Then

$$\begin{aligned} (5.8) \quad \int_0^\infty e^{-\lambda t} T(t)h \, dt &= \lambda \int_0^\infty e^{-\lambda t} T(t)f \, dt - \int_0^\infty e^{-\lambda t} T(t)g \, dt \\ &= \lambda \int_0^\infty e^{-\lambda t} T(t)f \, dt - \lambda \int_0^\infty e^{-\lambda t} \int_0^t T(s)g \, ds \, dt \\ &= f. \end{aligned}$$

Consequently, $\|f\| \leq \lambda^{-1} \|h\|$, proving dissipativity, and (5.6) holds.

Assuming (5.7), let $h \in L$ and $\lambda > 0$, and define $f = \int_0^\infty e^{-\lambda t} T(t)h \, dt$ and $g = \lambda f - h$. Then

$$\begin{aligned} (5.9) \quad \int_0^t T(s)g \, ds &= \lambda \int_0^t \int_0^\infty e^{-\lambda u} T(s+u)h \, du \, ds - \int_0^t T(s)h \, ds \\ &= \lambda \int_0^t e^{\lambda s} \int_s^\infty e^{-\lambda u} T(u)h \, du \, ds - \int_0^t T(s)h \, ds \\ &= e^{\lambda t} \int_t^\infty e^{-\lambda u} T(u)h \, du - \int_0^\infty e^{-\lambda u} T(u)h \, du \\ &\quad + \int_0^t T(s)h \, ds - \int_0^t T(s)h \, ds \\ &= T(t)f - f \end{aligned}$$

for all $t \geq 0$, so $(f, g) \in \hat{A}$ and $h = \lambda f - g \in \mathcal{R}(\lambda - \hat{A})$. \square

The following proposition, which is analogous to Proposition 1.5(a), gives a useful description of some elements of \hat{A} .

5.2 Proposition Let L and $\{T(t)\}$ be as in Theorem 5.1, let $h \in L$ and $u \geq 0$, and suppose that

$$(5.10) \quad T(t) \int_0^u T(s)h \, ds = \int_0^u T(t+s)h \, ds$$

for all $t \geq 0$. Then

$$(5.11) \quad \left(\int_0^u T(s)h \, ds, \quad T(u)h - h \right) \in \hat{A}.$$

Proof. Put $f = \int_0^u T(s)h \, ds$. Then

$$\begin{aligned} (5.12) \quad T(t)f - f &= \int_0^u T(t+s)h \, ds - \int_0^u T(s)h \, ds \\ &= \int_u^{t+u} T(s)h \, ds - \int_0^t T(s)h \, ds \\ &= \int_0^t T(s)(T(u)h - h) \, ds \end{aligned}$$

for all $t \geq 0$. □

In the present context, given a dissipative closed linear operator $A \subset L \times L$, it may be possible to find measurable functions $u: [0, \infty) \rightarrow L$ and $v: [0, \infty) \rightarrow L$ such that $(u(t), v(t)) \in A$ for every $t > 0$ and

$$(5.13) \quad u(t) = u(0) + \int_0^t v(s) \, ds, \quad t \geq 0.$$

One would expect u to be continuous, and since A is closed and linear, it is reasonable to expect that

$$(5.14) \quad \left(\int_0^t u(s) \, ds, \quad \int_0^t v(s) \, ds \right) = \left(\int_0^t u(s) \, ds, u(t) - u(0) \right) \in A$$

for all $t > 0$. With these considerations in mind, we have the following multi-valued extension of Proposition 2.10. Note that this result is in fact valid for arbitrary L .

5.3 Proposition Let $A \subset L \times L$ be a dissipative closed linear operator. Suppose $u: [0, \infty) \rightarrow L$ is continuous and $(\int_0^t u(s) ds, u(t) - u(0)) \in A$ for each $t > 0$. Then

$$(5.15) \quad \|u(t)\| \leq \|u(0)\|$$

for all $t \geq 0$. Given $\lambda > 0$, define

$$(5.16) \quad f = \int_0^\infty e^{-\lambda t} u(t) dt, \quad g = \lambda \int_0^\infty e^{-\lambda t} (u(t) - u(0)) dt.$$

Then $(f, g) \in A$ and $\lambda f - g = u(0)$.

Proof. Fix $t \geq 0$, and for each $\varepsilon > 0$, put $u_\varepsilon(t) = \varepsilon^{-1} \int_t^{t+\varepsilon} u(s) ds$. Then

$$(5.17) \quad u_\varepsilon(t) = u_\varepsilon(0) + \int_0^t \varepsilon^{-1} (u(s+\varepsilon) - u(s)) ds.$$

Since $(u_\varepsilon(t), \varepsilon^{-1}(u(t+\varepsilon) - u(t))) \in A$, it follows as in Proposition 2.10 that $\|u_\varepsilon(t)\| \leq \|u_\varepsilon(0)\|$. Letting $\varepsilon \rightarrow 0$, we obtain (5.15).

Integrating by parts,

$$(5.18) \quad f = \int_0^\infty e^{-\lambda t} u(t) dt = \lambda \int_0^\infty e^{-\lambda t} \int_0^t u(s) ds dt,$$

so $(f, g) \in A$ by the continuity of u and the fact that A is closed and linear. The equation $\lambda f - g = u(0)$ follows immediately from the definition of f and g . \square

Heuristically, if $\{S(t)\}$ has generator B and $\{T(t)\}$ has generator $A + B$, then (cf. Lemma 6.2)

$$(5.19) \quad T(t)f = S(t)f + \int_0^t S(t-s)A T(s)f ds$$

for all $t \geq 0$. Consequently, a weak form of the equation $u_t = (A + B)u$ is

$$(5.20) \quad u(t) = S(t)u(0) + \int_0^t S(t-s)Au(s) ds.$$

We extend Proposition 5.3 to this setting.

5.4 Proposition Let L be as in Proposition 5.1, let $A \subset L \times L$ be a dissipative closed linear operator, and let $\{S(t)\}$ be a strongly continuous, measurable, contraction semigroup on L . Suppose $u: [0, \infty) \rightarrow L$ is continuous, $v: [0, \infty) \rightarrow L$ is bounded and measurable, and

$$(5.21) \quad u(t) = S(t)u(0) + \int_0^t S(t-s)v(s) ds$$

for all $t \geq 0$. If

$$(5.22) \quad \left(\int_0^t u(s) ds, \int_0^t v(s) ds \right) \in A$$

for every $t > 0$, and

$$(5.23) \quad \int_0^t S(q+r)v(s) ds = S(q) \int_0^t S(r)v(s) ds$$

for all $q, r, t \geq 0$, then (5.15) holds for all $t \geq 0$.

5.5 Remark The above result holds in an arbitrary Banach space under the assumption that v is strongly measurable, that is, v can be uniformly approximated by measurable simple functions. \square

Proof. Assume first that $u: [0, \infty) \rightarrow L$ is continuously differentiable, $v: [0, \infty) \rightarrow L$ is continuous, and $(u(t), v(t)) \in A$ for all $t \geq 0$. Let $0 = t_0 < t_1 < \dots < t_n = t$. Then, as in the proof of Proposition 2.10,

$$\begin{aligned}
 (5.24) \quad \|u(t)\| &= \|u(0)\| + \sum_{i=1}^n [\|u(t_i)\| - \|u(t_{i-1})\|] \\
 &= \|u(0)\| + \sum_{i=1}^n \left[\|u(t_i)\| - \left\| u(t_i) - (S(t_i - t_{i-1}) - I)u(t_{i-1}) \right. \right. \\
 &\quad \left. \left. - \int_{t_{i-1}}^{t_i} S(t_i - s)v(s) ds \right\| \right] \\
 &\leq \|u(0)\| + \sum_{i=1}^n [\|u(t_i)\| - \|u(t_i) - (S(t_i - t_{i-1}) - I)u(t_{i-1}) - (t_i - t_{i-1})v(t_i)\|] \\
 &\quad + \sum_{i=1}^n \left\| (S(t_i - t_{i-1}) - I)(u(t_i) - u(t_{i-1})) - \int_{t_{i-1}}^{t_i} (S(t_i - s)v(s) - v(t_i)) ds \right\| \\
 &\leq \|u(0)\| + \sum_{i=1}^n [\|u(t_i)\| - \|2u(t_i) - (t_i - t_{i-1})v(t_i)\| + \|S(t_i - t_{i-1})u(t_i)\|] \\
 &\quad + \sum_{i=1}^n \left\| \int_{t_{i-1}}^{t_i} [(S(t_i - t_{i-1}) - I)u'(s) - S(t_i - s)v(s) + v(t_i)] ds \right\| \\
 &\leq \|u(0)\| + \int_0^t \|(S(s'' - s') - I)u'(s) - S(s'' - s)v(s) + v(s'')\| ds,
 \end{aligned}$$

where $s' = t_{i-1}$ and $s'' = t_i$ for $t_{i-1} \leq s < t_i$. Since the integrand on the right is bounded and tends to zero as $\max(t_i - t_{i-1}) \rightarrow 0$, we obtain (5.15) in this case.

In the general case, fix $t \geq 0$, and for each $\varepsilon > 0$, put

$$(5.25) \quad u_\varepsilon(t) = \varepsilon^{-1} \int_t^{t+\varepsilon} u(s) ds, \quad v_\varepsilon(t) = \varepsilon^{-1} \int_t^{t+\varepsilon} v(s) ds.$$

Then

$$\begin{aligned} (5.26) \quad u_\varepsilon(t) &= \varepsilon^{-1} \int_0^t u(t+s) ds \\ &= \varepsilon^{-1} \int_0^t S(t+s)u(0) ds + \varepsilon^{-1} \int_0^t \int_0^{t+s} S(t+s-r)v(r) dr ds \\ &= \varepsilon^{-1} S(t) \int_0^t S(s)u(0) ds + \varepsilon^{-1} \int_0^t \int_0^s S(t+s-r)v(r) dr ds \\ &\quad + \varepsilon^{-1} \int_0^t \int_0^s S(t-r)v(r+s) dr ds \\ &= S(t) \left[\varepsilon^{-1} \int_0^t S(s)u(0) ds + \varepsilon^{-1} \int_0^t \int_0^s S(s-r)v(r) dr ds \right] \\ &\quad + \int_0^t S(t-r)v_\varepsilon(r) dr. \end{aligned}$$

By the special case already treated,

$$(5.27) \quad \|u_\varepsilon(t)\| \leq \left\| \varepsilon^{-1} \int_0^t S(s)u(0) ds + \varepsilon^{-1} \int_0^t \int_0^s S(s-r)v(r) dr ds \right\|,$$

and letting $\varepsilon \rightarrow 0$, we obtain (5.15) in general. \square

6. APPROXIMATION THEOREMS

In this section, we adopt the following conventions. For $n = 1, 2, \dots, L_n$, in addition to L , is a Banach space (with norm also denoted by $\|\cdot\|$) and $\pi_n: L \rightarrow L_n$ is a bounded linear transformation. We assume that $\sup_n \|\pi_n\| < \infty$. We write $f_n \rightarrow f$ if $f_n \in L_n$ for each $n \geq 1$, $f \in L$, and $\lim_{n \rightarrow \infty} \|f_n - \pi_n f\| = 0$.

6.1 Theorem For $n = 1, 2, \dots$, let $\{T_n(t)\}$ and $\{T(t)\}$ be strongly continuous contraction semigroups on L_n and L with generators A_n and A . Let D be a core for A . Then the following are equivalent:

- (a) For each $f \in L$, $T_n(t)\pi_n f \rightarrow T(t)f$ for all $t \geq 0$, uniformly on bounded intervals.

- (b) For each $f \in L$, $T_n(t)\pi_n f \rightarrow T(t)f$ for all $t \geq 0$.
(c) For each $f \in D$, there exists $f_n \in \mathcal{D}(A_n)$ for each $n \geq 1$ such that $f_n \rightarrow f$ and $A_n f_n \rightarrow Af$ (i.e., $\{(f, Af): f \in D\} \subset \text{ex-lim}_{n \rightarrow \infty} A_n$).

The proof of this result depends on the following two lemmas, the first of which generalizes Lemma 2.5.

6.2 Lemma Fix a positive integer n . Let $\{S_n(t)\}$ and $\{S(t)\}$ be strongly continuous contraction semigroups on L_n and L with generators B_n and B . Let $f \in \mathcal{D}(B)$ and assume that $\pi_n S(s)f \in \mathcal{D}(B_n)$ for all $s \geq 0$ and that $B_n \pi_n S(\cdot)f: [0, \infty) \rightarrow L_n$ is continuous. Then, for each $t \geq 0$,

$$(6.1) \quad S_n(t)\pi_n f - \pi_n S(t)f = \int_0^t S_n(t-s)(B_n \pi_n - \pi_n B)S(s)f ds,$$

and therefore

$$(6.2) \quad \|S_n(t)\pi_n f - \pi_n S(t)f\| \leq \int_0^t \|(B_n \pi_n - \pi_n B)S(s)f\| ds.$$

Proof. It suffices to note that the integrand in (6.1) is $-(d/ds)S_n(t-s)\pi_n S(s)f$ for $0 \leq s \leq t$. \square

6.3 Lemma Suppose that the hypotheses of Theorem 6.1 are satisfied together with condition (c) of that theorem. For $n = 1, 2, \dots$ and $\lambda > 0$, let A_n^λ and A^λ be the Yosida approximations of A_n and A (cf. Lemma 2.4). Then $A_n^\lambda \pi_n f \rightarrow A^\lambda f$ for every $f \in L$ and $\lambda > 0$.

Proof. Fix $\lambda > 0$. Let $f \in D$ and $g = (\lambda - A)f$. By assumption, there exists $f_n \in \mathcal{D}(A_n)$ for each $n \geq 1$ such that $f_n \rightarrow f$ and $A_n f_n \rightarrow Af$, and therefore $(\lambda - A_n)f_n \rightarrow g$. Now observe that

$$\begin{aligned} (6.3) \quad & \|A_n^\lambda \pi_n g - \pi_n A^\lambda g\| \\ &= \|[\lambda^2(\lambda - A_n)^{-1} - \lambda I]\pi_n g - \pi_n [\lambda^2(\lambda - A)^{-1} - \lambda I]g\| \\ &= \lambda^2 \|(\lambda - A_n)^{-1}\pi_n g - \pi_n (\lambda - A)^{-1}g\| \\ &\leq \lambda^2 \|(\lambda - A_n)^{-1}\pi_n g - f_n\| + \lambda^2 \|f_n - \pi_n (\lambda - A)^{-1}g\| \\ &\leq \lambda \|\pi_n g - (\lambda - A_n)f_n\| + \lambda^2 \|f_n - \pi_n g\| \end{aligned}$$

for every $n \geq 1$. Consequently, $\|A_n^\lambda \pi_n g - \pi_n A^\lambda g\| \rightarrow 0$ for all $g \in \mathcal{R}(\lambda - A|_D)$. But $\mathcal{R}(\lambda - A|_D)$ is dense in L and the linear transformations $A_n^\lambda \pi_n - \pi_n A^\lambda$, $n = 1, 2, \dots$, are uniformly bounded, so the conclusion of the lemma follows. \square

Proof of Theorem 6.1. (a \Rightarrow b) Immediate.

(b \Rightarrow c) Let $\lambda > 0$, $f \in \mathcal{D}(A)$, and $g = (\lambda - A)f$, so that $f = \int_0^\infty e^{-\lambda t} T(t)g dt$. For each $n \geq 1$, put $f_n = \int_0^\infty e^{-\lambda t} T_n(t)\pi_n g dt \in \mathcal{D}(A_n)$. By (b) and the dominated convergence theorem, $f_n \rightarrow f$, so since $(\lambda - A_n)f_n = \pi_n g \rightarrow g = (\lambda - A)f$, we also have $A_n f_n \rightarrow Af$.

(c \Rightarrow a) For $n = 1, 2, \dots$ and $\lambda > 0$, let $\{T_n^\lambda(t)\}$ and $\{T^\lambda(t)\}$ be the strongly continuous contraction semigroups on L_n and L generated by the Yosida approximations A_n^λ and A^λ . Given $f \in D$, choose $\{f_n\}$ as in (c). Then

$$(6.4) \quad \begin{aligned} T_n(t)\pi_n f - \pi_n T(t)f &= T_n(t)(\pi_n f - f_n) + [T_n(t)f_n - T_n^\lambda(t)f_n] \\ &\quad + T_n^\lambda(t)(f_n - \pi_n f) + [T_n^\lambda(t)\pi_n f - \pi_n T^\lambda(t)f] \\ &\quad + \pi_n[T^\lambda(t)f - T(t)f] \end{aligned}$$

for every $n \geq 1$ and $t \geq 0$. Fix $t_0 \geq 0$. By Proposition 2.7 and Lemma 6.3,

$$(6.5) \quad \begin{aligned} \overline{\lim}_{n \rightarrow \infty} \sup_{0 \leq t \leq t_0} \|T_n(t)f_n - T_n^\lambda(t)f_n\| &\leq \overline{\lim}_{n \rightarrow \infty} t_0 \|A_n f_n - A_n^\lambda f_n\| \\ &\leq \overline{\lim}_{n \rightarrow \infty} t_0 \{ \|A_n f_n - \pi_n Af\| + \|\pi_n(Af - A^\lambda f)\| \\ &\quad + \|A_n^\lambda f - A_n^\lambda \pi_n f\| + \|A_n^\lambda(\pi_n f - f_n)\| \} \\ &\leq Kt_0 \|Af - A^\lambda f\|, \end{aligned}$$

where $K = \sup_n \|\pi_n\|$. Using Lemmas 6.2, 6.3, and the dominated convergence theorem, we obtain

$$(6.6) \quad \begin{aligned} \overline{\lim}_{n \rightarrow \infty} \sup_{0 \leq t \leq t_0} \|T_n^\lambda(t)\pi_n f - \pi_n T^\lambda(t)f\| \\ &\leq \overline{\lim}_{n \rightarrow \infty} \int_0^{t_0} \|(A_n^\lambda \pi_n - \pi_n A^\lambda)T^\lambda(s)f\| ds = 0. \end{aligned}$$

Applying (6.5), (6.6), and Proposition 2.7 to (6.4), we find that

$$(6.7) \quad \overline{\lim}_{n \rightarrow \infty} \sup_{0 \leq t \leq t_0} \|T_n(t)\pi_n f - \pi_n T(t)f\| \leq 2Kt_0 \|Af - A^\lambda f\|.$$

Since λ was arbitrary, Lemma 2.4(c) shows that the left side of (6.7) is zero. But this is valid for all $f \in D$, and since D is dense in L , it holds for all $f \in L$. \square

There is a discrete-parameter analogue of Theorem 6.1, the proof of which depends on the following lemma.

6.4 Lemma Let B be a linear contraction on L . Then

$$(6.8) \quad \|B^n f - e^{n(B-\lambda)} f\| \leq \sqrt{n} \|Bf - f\|$$

for all $f \in L$ and $n = 0, 1, \dots$

Proof. Fix $f \in L$ and $n \geq 0$. For $k = 0, 1, \dots$,

$$(6.9) \quad \begin{aligned} \|B^n f - B^k f\| &\leq \|B^{|k-n|} f - f\| \\ &= \left\| \sum_{j=0}^{|k-n|-1} B^j (Bf - f) \right\| \\ &\leq |k-n| \|Bf - f\|. \end{aligned}$$

Therefore

$$(6.10) \quad \begin{aligned} \|B^n f - e^{n(B-I)} f\| &= \left\| B^n f - e^{-n} \sum_{k=0}^{\infty} \frac{n^k}{k!} B^k f \right\| \\ &= e^{-n} \left\| \sum_{k=0}^{\infty} (B^n f - B^k f) \frac{n^k}{k!} \right\| \\ &\leq e^{-n} \sum_{k=0}^{\infty} |k-n| \frac{n^k}{k!} \|Bf - f\| \\ &\leq \left\{ e^{-n} \sum_{k=0}^{\infty} (k-n)^2 \frac{n^k}{k!} \right\}^{1/2} \|Bf - f\| \\ &= \sqrt{n} \|Bf - f\|. \end{aligned}$$

(Note that the last equality follows from the fact that a Poisson random variable with parameter n has mean n and variance n .) \square

6.5 Theorem For $n = 1, 2, \dots$, let T_n be a linear contraction on L_n , let ε_n be a positive number, and put $A_n = \varepsilon_n^{-1}(T_n - I)$. Assume that $\lim_{n \rightarrow \infty} \varepsilon_n = 0$. Let $\{T(t)\}$ be a strongly continuous contraction semigroup on L with generator A , and let D be a core for A . Then the following are equivalent:

- (a) For each $f \in L$, $T_n^{(t/\varepsilon_n)} \pi_n f \rightarrow T(t)f$ for all $t \geq 0$, uniformly on bounded intervals.
- (b) For each $f \in L$, $T_n^{(t/\varepsilon_n)} \pi_n f \rightarrow T(t)f$ for all $t \geq 0$.
- (c) For each $f \in D$, there exists $f_n \in L_n$ for each $n \geq 1$ such that $f_n \rightarrow f$ and $A_n f_n \rightarrow Af$ (i.e., $\{(f, Af) : f \in D\} \subset \text{ex-lim}_{n \rightarrow \infty} A_n$).

Proof. (a \Rightarrow b) Immediate.

(b \Rightarrow c) Let $\lambda > 0$, $f \in \mathcal{D}(A)$, and $g = (\lambda - A)f$, so that $f = \int_0^\infty e^{-\lambda t} T(t)g dt$. For each $n \geq 1$, put

$$(6.11) \quad f_n = \varepsilon_n \sum_{k=0}^{\infty} e^{-\lambda k \varepsilon_n} T_n^k \pi_n g.$$

By (b) and the dominated convergence theorem, $f_n \rightarrow f$, and a simple calculation shows that

$$(6.12) \quad (\lambda - A_n)f_n = \pi_n g + \lambda \varepsilon_n \pi_n g$$

$$+ (\lambda \varepsilon_n e^{-\lambda t_n} - 1 + e^{-\lambda t_n}) \sum_{k=0}^{\infty} e^{-\lambda k t_n} T_n^{k+1} \pi_n g$$

for every $n \geq 1$, so $(\lambda - A_n)f_n \rightarrow g = (\lambda - A)f$. It follows that $A_n f_n \rightarrow Af$.

(c \Rightarrow a) Given $f \in D$, choose $\{f_n\}$ as in (c). Then

$$(6.13) \quad T_n^{t/\varepsilon_n} \pi_n f - \pi_n T(t)f$$

$$\begin{aligned} &= T_n^{t/\varepsilon_n}(\pi_n f - f_n) + T_n^{t/\varepsilon_n}f_n - \exp\left\{\varepsilon_n\left[\frac{t}{\varepsilon_n}\right]A_n\right\}f_n \\ &\quad + \exp\left\{\varepsilon_n\left[\frac{t}{\varepsilon_n}\right]A_n\right\}(f_n - \pi_n f) + \exp\left\{\varepsilon_n\left[\frac{t}{\varepsilon_n}\right]A_n\right\}\pi_n f - \pi_n T(t)f \end{aligned}$$

for every $n \geq 1$ and $t \geq 0$. Fix $t_0 \geq 0$. By Lemma 6.4,

$$\begin{aligned} (6.14) \quad &\overline{\lim}_{n \rightarrow \infty} \sup_{0 \leq t \leq t_0} \left\| T_n^{t/\varepsilon_n}f_n - \exp\left\{\varepsilon_n\left[\frac{t}{\varepsilon_n}\right]A_n\right\}f_n \right\| \\ &\leq \overline{\lim}_{n \rightarrow \infty} \left(\left[\frac{t_0}{\varepsilon_n} \right]^{1/2} \varepsilon_n \|A_n f_n\| \right) = 0, \end{aligned}$$

and by Theorem 6.1,

$$(6.15) \quad \lim_{n \rightarrow \infty} \sup_{0 \leq t \leq t_0} \left\| \exp\left\{\varepsilon_n\left[\frac{t}{\varepsilon_n}\right]A_n\right\}\pi_n f - \pi_n T(t)f \right\| = 0.$$

Consequently,

$$(6.16) \quad \lim_{n \rightarrow \infty} \sup_{0 \leq t \leq t_0} \|T_n^{t/\varepsilon_n} \pi_n f - \pi_n T(t)f\| = 0.$$

But this is valid for all $f \in D$, and since D is dense in L , it holds for all $f \in L$. \square

6.6 Corollary Let $\{V(t): t \geq 0\}$ be a family of linear contractions on L with $V(0) = I$, and let $\{T(t)\}$ be a strongly continuous contraction semigroup on L with generator A . Let D be a core for A . If $\lim_{\varepsilon \rightarrow 0} \varepsilon^{-1}[V(\varepsilon)f - f] = Af$ for every $f \in D$, then, for each $f \in L$, $V(t/n)^n f \rightarrow T(t)f$ for all $t \geq 0$, uniformly on bounded intervals.

Proof. It suffices to show that if $\{t_n\}$ is a sequence of positive numbers such that $t_n \rightarrow t \geq 0$, then $V(t_n/n)^n f \rightarrow T(t)f$ for every $f \in L$. But this is an immediate consequence of Theorem 6.5 with $T_n = V(t_n/n)$ and $\varepsilon_n = t_n/n$ for each $n \geq 1$. \square

6.7 Corollary Let $\{T(t)\}$, $\{S(t)\}$, and $\{U(t)\}$ be strongly continuous contraction semigroups on L with generators A , B , and C , respectively. Let D be a core for A , and assume that $D \subset \mathcal{D}(B) \cap \mathcal{D}(C)$ and that $A = B + C$ on D . Then, for each $f \in L$,

$$(6.17) \quad \lim_{n \rightarrow \infty} \left[S\left(\frac{t}{n}\right) U\left(\frac{t}{n}\right) \right]^n f = T(t)f$$

for all $t \geq 0$, uniformly on bounded intervals. Alternatively, if $\{\varepsilon_n\}$ is a sequence of positive numbers tending to zero, then, for each $f \in L$,

$$(6.18) \quad \lim_{n \rightarrow \infty} [S(\varepsilon_n)U(\varepsilon_n)]^{t/\varepsilon_n} f = T(t)f$$

for all $t \geq 0$, uniformly on bounded intervals.

Proof. The first result follows easily from Corollary 6.6 with $V(t) = S(t)U(t)$ for all $t \geq 0$. The second follows directly from Theorem 6.5. \square

6.8 Corollary Let $\{T(t)\}$ be a strongly continuous contraction semigroup on L with generator A . Then, for each $f \in L$, $(I - (t/n)A)^{-n}f \rightarrow T(t)f$ for all $t \geq 0$, uniformly on bounded intervals. Alternatively, if $\{\varepsilon_n\}$ is a sequence of positive numbers tending to zero, then, for each $f \in L$, $(I - \varepsilon_n A)^{-t/\varepsilon_n}f \rightarrow T(t)f$ for all $t \geq 0$, uniformly on bounded intervals.

Proof. The first result is a consequence of Corollary 6.6. Simply take $V(t) = (I - tA)^{-1}$ for each $t \geq 0$, and note that if $\varepsilon > 0$ and $\lambda = \varepsilon^{-1}$, then

$$(6.19) \quad \varepsilon^{-1}\{V(\varepsilon)f - f\} = \lambda^2(\lambda - A)^{-1}f - \lambda f = A_\lambda f,$$

where A_λ is the Yosida approximation of A (cf. Lemma 2.4). The second result follows from (6.19) and Theorem 6.5. \square

We would now like to generalize Theorem 6.1 in two ways. First, we would like to be able to use some extension \hat{A}_n of the generator A_n in verifying the conditions for convergence. That is, given $(f, g) \in A$, it may be possible to find $(f_n, g_n) \in \hat{A}_n$ for each $n \geq 1$ such that $f_n \rightarrow f$ and $g_n \rightarrow g$ when it is not possible (or at least more difficult) to find $(f_n, g_n) \in A_n$ for each $n \geq 1$. Second, we would like to consider notions of convergence other than norm convergence. For example, convergence of bounded sequences of functions pointwise or uniformly on compact sets may be more appropriate than uniform convergence for some applications. An analogous generalization of Theorem 6.5 is also given.

Let LIM denote a notion of convergence of certain sequences $f_n \in L_n$, $n = 1, 2, \dots$, to elements $f \in L$ satisfying the following conditions:

$$(6.20) \quad \text{LIM } f_n = f \text{ and } \text{LIM } g_n = g \text{ imply}$$

$$\text{LIM } (\alpha f_n + \beta g_n) = \alpha f + \beta g \text{ for all } \alpha, \beta \in \mathbb{R}.$$

$$(6.21) \quad \text{LIM } f_n^{(k)} = f^{(k)} \text{ for each } k \geq 1 \text{ and}$$

$$\limsup_{k \rightarrow \infty} \sup_{n \geq 1} \|f_n^{(k)} - f\| \vee \|f^{(k)} - f\| = 0 \text{ imply } \text{LIM } f_n = f.$$

$$(6.22) \quad \text{There exists } K > 0 \text{ such that for each } f \in L, \text{ there is a sequence } f_n \in L_n \text{ with } \|f_n\| \leq K \|f\|, \quad n = 1, 2, \dots, \text{ satisfying } \text{LIM } f_n = f.$$

If $A_n \subset L_n \times L_n$ is linear for each $n \geq 1$, then, by analogy with (4.3), we define

$$(6.23) \quad \text{ex-LIM } A_n = \{(f, g) \in L \times L : \text{there exists } (f_n, g_n) \in A_n$$

$$\text{for each } n \geq 1 \text{ such that } \text{LIM } f_n = f \text{ and } \text{LIM } g_n = g\}.$$

6.9 Theorem For $n = 1, 2, \dots$, let $A_n \subset L_n \times L_n$ and $A \subset L \times L$ be linear and dissipative with $\mathcal{R}(\lambda - A_n) = L_n$ and $\mathcal{R}(\lambda - A) = L$ for some (hence all) $\lambda > 0$, and let $\{T_n(t)\}$ and $\{T(t)\}$ be the corresponding strongly continuous contraction semigroups on $\overline{\mathcal{D}(A_n)}$ and $\overline{\mathcal{D}(A)}$. Let LIM satisfy (6.20)–(6.22) together with

$$(6.24) \quad \text{LIM } f_n = 0 \text{ implies } \text{LIM } (\lambda - A_n)^{-1} f_n = 0 \text{ for all } \lambda > 0.$$

(a) If $A \subset \text{ex-LIM } A_n$, then, for each $(f, g) \in A$, there exists $(f_n, g_n) \in A_n$ for each $n \geq 1$ such that $\sup_n \|f_n\| < \infty$, $\sup_n \|g_n\| < \infty$, $\text{LIM } f_n = f$, $\text{LIM } g_n = g$, and $\text{LIM } T_n(t)f_n = T(t)f$ for all $t \geq 0$.

(b) If in addition $\{T_n(t)\}$ extends to a contraction semigroup (also denoted by $\{T_n(t)\}$) on L_n for each $n \geq 1$, and if

$$(6.25) \quad \text{LIM } f_n = 0 \text{ implies } \text{LIM } T_n(t)f_n = 0 \text{ for all } t \geq 0,$$

then, for each $f \in \overline{\mathcal{D}(A)}$, $\text{LIM } f_n = f$ implies $\text{LIM } T_n(t)f_n = T(t)f$ for all $t \geq 0$.

6.10 Remark Under the hypotheses of the theorem, $\text{ex-LIM } A_n$ is closed in $L \times L$ (Problem 16). Consequently, the conclusion of (a) is valid for all $(f, g) \in \bar{A}$. \square

Proof. By renorming L_n , $n = 1, 2, \dots$, if necessary, we can assume $K = 1$ in (6.22).

Let \mathcal{L} denote the Banach space $(\prod_{n \geq 1} L_n) \times L$ with norm given by $\|(\{f_n\}, f)\| = \sup_{n \geq 1} \|f_n\| \vee \|f\|$, and let

$$(6.26) \quad \mathcal{L}_0 = \{(\{f_n\}, f) \in \mathcal{L} : \text{LIM } f_n = f\}.$$

Conditions (6.20) and (6.21) imply that \mathcal{L}_0 is a closed subspace of \mathcal{L} , and Condition (6.22) (with $K = 1$) implies that, for each $f \in L$, there is an element $(\{f_n\}, f) \in \mathcal{L}_0$ with $\|(\{f_n\}, f)\| = \|f\|$.

Let

$$(6.27) \quad \mathcal{A} = \{[(\{f_n\}, f), (\{g_n\}, g)] \in \mathcal{L} \times \mathcal{L} : (f_n, g_n) \in A_n \text{ for each}$$

$$n \geq 1 \quad \text{and} \quad (f, g) \in A\}.$$

Then \mathcal{A} is linear and dissipative, and $\mathcal{R}(\lambda - \mathcal{A}) = \mathcal{L}$ for all $\lambda > 0$. The corresponding strongly continuous semigroup $\{\mathcal{T}(t)\}$ on $\overline{\mathcal{D}(\mathcal{A})}$ is given by

$$(6.28) \quad \mathcal{T}(t)(\{f_n\}, f) = (\{T_n(t)f_n\}, T(t)f).$$

We would like to show that

$$(6.29) \quad \mathcal{T}(t) : \mathcal{L}_0 \cap \overline{\mathcal{D}(\mathcal{A})} \rightarrow \mathcal{L}_0 \cap \overline{\mathcal{D}(\mathcal{A})}, \quad t \geq 0.$$

To do so, we need the following observation. If $(f, g) \in A$, $\lambda > 0$, $h = \lambda f - g$, $(\{h_n\}, h) \in \mathcal{L}_0$, and

$$(6.30) \quad (f_n, g_n) = ((\lambda - A_n)^{-1}h_n, \lambda f_n - h_n)$$

for each $n \geq 1$, then

$$(6.31) \quad [(\{f_n\}, f), (\{g_n\}, g)] \in (\mathcal{L}_0 \times \mathcal{L}_0) \cap \mathcal{A}.$$

To prove this, since $A \subset \text{ex-LIM } A_n$, choose $(f_n, \hat{g}_n) \in A_n$ for each $n \geq 1$ such that $\text{LIM } f_n = f$ and $\text{LIM } \hat{g}_n = g$. Then $\text{LIM } (h_n - (\lambda f_n - \hat{g}_n)) = 0$, so by (6.24), $\text{LIM } (\lambda - A_n)^{-1}h_n - f_n = 0$. It follows that $\text{LIM } f_n = \text{LIM } (\lambda - A_n)^{-1}h_n = \text{LIM } \hat{f}_n = f$ and $\text{LIM } g_n = \text{LIM } (\lambda f_n - h_n) = \lambda f - h = g$. Also, $\sup_n \|f_n\| \leq \lambda^{-1} \sup_n \|h_n\| < \infty$ and $\sup_n \|g_n\| \leq 2 \sup_n \|h_n\| < \infty$. Consequently, $[(\{f_n\}, f), (\{g_n\}, g)]$ belongs to $\mathcal{L}_0 \times \mathcal{L}_0$, and it clearly also belongs to \mathcal{A} .

Given $(\{h_n\}, h) \in \mathcal{L}_0$ and $\lambda > 0$, there exists $(f, g) \in A$ such that $\lambda f - g = h$. Define $(f_n, g_n) \in A_n$ for each $n \geq 1$ by (6.30). Then $(\lambda - \mathcal{A})^{-1}(\{h_n\}, h) = (\{f_n\}, f) \in \mathcal{L}_0$ by (6.31), so

$$(6.32) \quad (\lambda - \mathcal{A})^{-1} : \mathcal{L}_0 \rightarrow \mathcal{L}_0, \quad \lambda > 0.$$

By Corollary 2.8, this proves (6.29).

To prove (a), let $(f, g) \in A$, $\lambda > 0$, and $h = \lambda f - g$. By (6.22), there exists $(\{h_n\}, h) \in \mathcal{L}_0$ with $\|(\{h_n\}, h)\| = \|h\|$. Define $(f_n, g_n) \in A_n$ for each $n \geq 1$ by (6.30). By (6.31), (6.29), and (6.28), $(\{T_n(t)f_n\}, T(t)f) \in \mathcal{L}_0$ for all $t \geq 0$, so the conclusion of (a) is satisfied.

As for (b), observe that, by (a) together with (6.25), $\text{LIM } f_n = f \in \mathcal{D}(A)$ implies $\text{LIM } T_n(t)f_n = T(t)f$ for all $t \geq 0$. Let $f \in \overline{\mathcal{D}(A)}$ and choose $\{f^{(k)}\} \subset \mathcal{D}(A)$ such that $\|f^{(k)} - f\| \leq 2^{-k}$ for each $k \geq 1$. Put $f^{(0)} = 0$, and by (6.22), choose

36 OPERATOR SEMIGROUPS

$(\{u_n^{(k)}\}, f^{(k)} - f^{(k-1)}) \in \mathcal{L}_0$ such that $\|(\{u_n^{(k)}\}, f^{(k)} - f^{(k-1)})\| = \|f^{(k)} - f^{(k-1)}\|$ for each $k \geq 1$. Fix $t \geq 0$. Then

$$(6.33) \quad \text{LIM } \sum_1^k u_n^{(l)} = f^{(k)}, \quad \text{LIM } T_n(t) \sum_1^k u_n^{(l)} = T(t)f^{(k)}$$

for each $k \geq 1$. Since

$$(6.34) \quad \left\| \sum_{k+1}^{\infty} u_n^{(l)} \right\| \leq 3 \cdot 2^{-k}, \quad \|f^{(k)} - f\| \leq 2^{-k},$$

and

$$(6.35) \quad \left\| T_n(t) \sum_{k+1}^{\infty} u_n^{(l)} \right\| \leq 3 \cdot 2^{-k}, \quad \|T(t)f^{(k)} - T(t)f\| \leq 2^{-k},$$

for each $n \geq 1$ and $k \geq 1$, (6.21) implies that

$$(6.36) \quad \text{LIM } \sum_1^{\infty} u_n^{(l)} = f, \quad \text{LIM } T_n(t) \sum_1^{\infty} u_n^{(l)} = T(t)f,$$

so the conclusion of (b) follows from (6.25). \square

6.11 Theorem For $n = 1, 2, \dots$, let T_n be a linear contraction on L_n , let $\varepsilon_n > 0$, and put $A_n = \varepsilon_n^{-1}(T_n - I)$. Assume that $\lim_{n \rightarrow \infty} \varepsilon_n = 0$. Let $A \subset L \times L$ be linear and dissipative with $\mathcal{R}(\lambda - A) = L$ for some (hence all) $\lambda > 0$, and let $\{T(t)\}$ be the corresponding strongly continuous contraction semigroup on $\mathcal{D}(A)$. Let LIM satisfy (6.20)–(6.22), (6.24), and

$$(6.37) \quad \lim \|f_n\| = 0 \text{ implies LIM } f_n = 0.$$

(a) If $A \subset \text{ex-LIM } A_n$, then, for each $(f, g) \in A$, there exists $f_n \in L_n$ for each $n \geq 1$ such that $\sup_n \|f_n\| < \infty$, $\sup_n \|A_n f_n\| < \infty$, $\text{LIM } f_n = f$, $\text{LIM } A_n f_n = g$, and $\text{LIM } T_n^{(t/\varepsilon_n)} f_n = T(t)f$ for all $t \geq 0$.

(b) If in addition

$$(6.38) \quad \text{LIM } f_n = 0 \text{ implies LIM } T_n^{(t/\varepsilon_n)} f_n = 0 \text{ for all } t \geq 0,$$

then for each $f \in \overline{\mathcal{D}(A)}$, $\text{LIM } f_n = f$ implies $\text{LIM } T_n^{(t/\varepsilon_n)} f_n = T(t)f$ for all $t \geq 0$.

Proof. Let $(f, g) \in A$. By Theorem 6.9, there exists $f_n \in L_n$ for each $n \geq 1$ such that $\sup_n \|f_n\| < \infty$, $\sup_n \|A_n f_n\| < \infty$, $\text{LIM } f_n = f$, $\text{LIM } A_n f_n = g$, and $\text{LIM } e^{tA_n} f_n = T(t)f$ for all $t \geq 0$. Since

$$(6.39) \quad \begin{aligned} \lim_{n \rightarrow \infty} & \left\| \exp \left\{ \varepsilon_n \left[\frac{t}{\varepsilon_n} \right] A_n \right\} f_n - \exp \{tA_n\} f_n \right\| \\ & \leq \lim_{n \rightarrow \infty} \left| \varepsilon_n \left[\frac{t}{\varepsilon_n} \right] - t \right| \|A_n f_n\| = 0 \end{aligned}$$

for all $t \geq 0$, we deduce from (6.37) that

$$(6.40) \quad \text{LIM } \exp \left\{ \varepsilon_n \left[\frac{t}{\varepsilon_n} \right] A_n \right\} f_n = T(t)f, \quad t \geq 0.$$

The conclusion of (a) therefore follows from (6.14) and (6.37).

The proof of (b) is completely analogous to that of Theorem 6.9 (b). \square

7. PERTURBATION THEOREMS

One of the main results of this section concerns the approximation of semigroups with generators of the form $A + B$, where A and B themselves generate semigroups. (By definition, $\mathcal{D}(A + B) = \mathcal{D}(A) \cap \mathcal{D}(B)$.) First, however, we give some sufficient conditions for $A + B$ to generate a semigroup.

7.1 Theorem Let A be a linear operator on L such that \bar{A} is single-valued and generates a strongly continuous contraction semigroup on L . Let B be a dissipative linear operator on L such that $\mathcal{D}(B) \supset \mathcal{D}(A)$. (In particular, \bar{B} is single-valued by Lemma 4.2.) If

$$(7.1) \quad \|Bf\| \leq \alpha \|Af\| + \beta \|f\|, \quad f \in \mathcal{D}(A),$$

where $0 \leq \alpha < 1$ and $\beta \geq 0$, then $\overline{A + B}$ is single-valued and generates a strongly continuous contraction semigroup on L . Moreover, $\overline{A + B} = \bar{A} + \bar{B}$.

Proof. Let $\gamma \geq 0$ be arbitrary. Clearly, $\mathcal{D}(A + \gamma B) = \mathcal{D}(A)$ is dense in L . In addition, $A + \gamma B$ is dissipative. To see this, let A_μ be the Yosida approximation of \bar{A} for each $\mu > 0$, so that $A_\mu = \mu[\mu(\mu - \bar{A})^{-1} - I]$. If $f \in \mathcal{D}(A)$ and $\lambda > 0$, then

$$\begin{aligned} (7.2) \quad \|\lambda f - (A + \gamma B)f\| &= \lim_{\mu \rightarrow \infty} \|\lambda f - (A_\mu + \gamma B)f\| \\ &= \lim_{\mu \rightarrow \infty} \|(\lambda + \mu)f - \gamma Bf - \mu^2(\mu - \bar{A})^{-1}f\| \\ &\geq \overline{\lim}_{\mu \rightarrow \infty} \{ \|(\lambda + \mu)f - \gamma Bf\| - \|\mu^2(\mu - \bar{A})^{-1}f\| \} \\ &\geq \lim_{\mu \rightarrow \infty} \{ (\lambda + \mu)\|f\| - \mu\|f\| \} \\ &= \lambda\|f\| \end{aligned}$$

by Lemma 2.4(c) and the dissipativity of γB .

If $f \in \mathcal{D}(\bar{A})$, then there exists $\{f_n\} \subset \mathcal{D}(A)$ such that $f_n \rightarrow f$ and $Af_n \rightarrow \bar{A}f$. By (7.1), $\{Bf_n\}$ is Cauchy, so $f \in \mathcal{D}(\bar{B})$ and $Bf_n \rightarrow \bar{B}f$. Hence $\mathcal{D}(\bar{A}) \subset \mathcal{D}(\bar{B})$ and (7.1) extends to

$$(7.3) \quad \|\bar{B}f\| \leq \alpha \|\bar{A}f\| + \beta \|f\|, \quad f \in \mathcal{D}(\bar{A}).$$

In addition, if $f \in \mathcal{D}(\bar{A})$ and if $\{f_n\}$ is as above, then

$$(7.4) \quad (\bar{A} + \gamma \bar{B})f = \lim_{n \rightarrow \infty} Af_n + \gamma \lim_{n \rightarrow \infty} Bf_n = \overline{(A + \gamma B)f_n} = \overline{(A + \gamma B)f},$$

implying that $\overline{A + \gamma B}$ is a dissipative extension of $\bar{A} + \gamma \bar{B}$.

Let

$$(7.5) \quad \Gamma = \{\gamma \geq 0 : \mathcal{R}(\lambda - \bar{A} - \gamma \bar{B}) = L \text{ for some (hence all) } \lambda > 0\}.$$

To complete the proof, it suffices by Theorem 2.6 and Proposition 4.1 to show that $1 \in \Gamma$. Noting that $0 \in \Gamma$ by assumption, it is enough to show that

$$(7.6) \quad \gamma \in \Gamma \cap [0, 1] \text{ implies } \left[\gamma, \gamma + \frac{1 - \alpha\gamma}{2\alpha} \right] \subset \Gamma.$$

To prove (7.6), let $\gamma \in \Gamma \cap [0, 1]$, $0 \leq \varepsilon < (2\alpha)^{-1}(1 - \alpha\gamma)$, and $\lambda > 0$. If $g \in \mathcal{D}(\bar{A})$, then $f \equiv (\lambda - \bar{A} - \gamma \bar{B})^{-1}g$ satisfies

$$(7.7) \quad \|\bar{B}f\| \leq \alpha \|\bar{A}f\| + \beta \|f\| \leq \alpha \|(\bar{A} + \gamma \bar{B})f\| + \alpha\gamma \|\bar{B}f\| + \beta \|f\|$$

by (7.3), that is,

$$(7.8) \quad \|\bar{B}f\| \leq \alpha(1 - \alpha\gamma)^{-1} \|(\bar{A} + \gamma \bar{B})f\| + \beta(1 - \alpha\gamma)^{-1} \|f\|,$$

and consequently,

$$(7.9) \quad \|\bar{B}(\lambda - \bar{A} - \gamma \bar{B})^{-1}g\| \leq [2\alpha(1 - \alpha\gamma)^{-1} + \beta(1 - \alpha\gamma)^{-1}\lambda^{-1}] \|g\|.$$

Thus, for λ sufficiently large, $\|\varepsilon \bar{B}(\lambda - \bar{A} - \bar{B})^{-1}\| < 1$, which implies that $I - \varepsilon \bar{B}(\lambda - \bar{A} - \gamma \bar{B})^{-1}$ is invertible. We conclude that

$$\begin{aligned} (7.10) \quad \mathcal{R}(\lambda - \bar{A} - (\gamma + \varepsilon)\bar{B}) &\supset \mathcal{R}((\lambda - \bar{A} - (\gamma + \varepsilon)\bar{B})(\lambda - \bar{A} - \gamma \bar{B})^{-1}) \\ &= \mathcal{R}(I - \varepsilon \bar{B}(\lambda - \bar{A} - \gamma \bar{B})^{-1}) \\ &= L \end{aligned}$$

for such λ , so $\gamma + \varepsilon \in \Gamma$, implying (7.6) and completing the proof. \square

7.2 Corollary If A generates a strongly continuous contraction semigroup on L and B is a bounded linear operator on L , then $A + B$ generates a strongly continuous semigroup $\{T(t)\}$ on L such that

$$(7.11) \quad \|T(t)\| \leq e^{\|B\|t}, \quad t \geq 0.$$

Proof. Apply Theorem 7.1 with $B - \|B\|I$ in place of B . \square

Before turning to limit theorems, we state the following lemma, the proof of which is left to the reader (Problem 18). For an operator A , let $\mathcal{N}(A) \equiv \{f \in \mathcal{D}(A) : Af = 0\}$ denote the null space of A .

7.3 Lemma Let B generate a strongly continuous contraction semigroup $\{S(t)\}$ on L , and assume that

$$(7.12) \quad \lim_{\lambda \rightarrow 0^+} \lambda \int_0^\infty e^{-\lambda t} S(t)f dt \equiv Pf \text{ exists for all } f \in L.$$

Then the following conclusions hold:

- (a) P is a linear contraction on L and $P^2 = P$.
- (b) $S(t)P = PS(t) = P$ for all $t \geq 0$.
- (c) $\mathcal{R}(P) = \mathcal{N}(B)$.
- (d) $\mathcal{N}(P) = \overline{\mathcal{R}(B)}$.

7.4 Remark If in the lemma

$$(7.13) \quad B = \gamma^{-1}(Q - I),$$

where Q is a linear contraction on L and $\gamma > 0$, then a simple calculation shows that (7.12) is equivalent to

$$(7.14) \quad \lim_{\rho \rightarrow 1^-} (1 - \rho) \sum_{k=0}^{\infty} \rho^k Q^k f \equiv Pf \text{ exists for all } f \in L. \quad \square$$

7.5 Remark If in the lemma $\lim_{t \rightarrow \infty} S(t)f$ exists for every $f \in L$, then (7.12) holds and

$$(7.15) \quad Pf = \lim_{t \rightarrow \infty} S(t)f, \quad f \in L.$$

If B is as in Remark 7.4 and if $\lim_{k \rightarrow \infty} Q^k f$ exists for every $f \in L$, then (7.14) holds (in fact, so does (7.15)) and

$$(7.16) \quad Pf = \lim_{k \rightarrow \infty} Q^k f, \quad f \in L.$$

The proofs of these assertions are elementary. \square

For the following result, recall the notation introduced in the first paragraph of Section 6, as well as the notion of the extended limit of a sequence of operators (Section 4).

7.6 Theorem Let $A \subset L \times L$ be linear, and let B generate a strongly continuous contraction semigroup $\{S(t)\}$ on L satisfying (7.12). Let D be a subspace

of $\mathcal{D}(A)$ and D' a core for B . For $n = 1, 2, \dots$, let A_n be a linear operator on L_n and let $\alpha_n > 0$. Suppose that $\lim_{n \rightarrow \infty} \alpha_n = \infty$ and that

$$(7.17) \quad \{(f, g) \in A : f \in D\} \subset \text{ex-lim}_{n \rightarrow \infty} A_n,$$

$$(7.18) \quad \{(h, Bh) : h \in D'\} \subset \text{ex-lim}_{n \rightarrow \infty} \alpha_n^{-1} A_n.$$

Define $C = \{(f, Pg) : (f, g) \in A, f \in D\}$ and assume that $\{(f, g) \in C : g \in \bar{D}\}$ is single-valued and generates a strongly continuous contraction semigroup $\{T(t)\}$ on \bar{D} .

(a) If A_n is the generator of a strongly continuous contraction semigroup $\{T_n(t)\}$ on L_n for each $n \geq 1$, then, for each $f \in \bar{D}$, $T_n(t)\pi_n f \rightarrow T(t)f$ for all $t \geq 0$, uniformly on bounded intervals.

(b) If $A_n = \varepsilon_n^{-1}(T_n - I)$ for each $n \geq 1$, where T_n is a linear contraction on L_n and $\varepsilon_n > 0$, and if $\lim_{n \rightarrow \infty} \varepsilon_n = 0$, then, for each $f \in \bar{D}$, $T_n^{(t/\varepsilon_n)}\pi_n f \rightarrow T(t)f$ for all $t \geq 0$, uniformly on bounded intervals.

Proof. Theorems 6.1 and 6.5 are applicable, provided we can show that

$$(7.19) \quad \{(f, g) \in C : g \in \bar{D}\} \subset \left(\text{ex-lim}_{n \rightarrow \infty} A_n \right) \cap (\bar{D} \times \bar{D}).$$

Since $\text{ex-lim}_{n \rightarrow \infty} A_n$ is closed, it suffices to show that $C \subset \text{ex-lim}_{n \rightarrow \infty} A_n$. Given $(f, g) \in C$ with $f \in D$, choose $f_n \in \mathcal{D}(A_n)$ for each $n \geq 1$ such that $f_n \rightarrow f$ and $A_n f_n \rightarrow g$. Given $h \in D'$, choose $h_n \in \mathcal{D}(A_n)$ for each $n \geq 1$ such that $h_n \rightarrow h$ and $\alpha_n^{-1} A_n h_n \rightarrow Bh$. Then $f_n + \alpha_n^{-1} h_n \rightarrow f$ and $A_n(f_n + \alpha_n^{-1} h_n) \rightarrow g + Bh$. Consequently,

$$(7.20) \quad \{(f, g + Bh) : (f, g) \in C, f \in D, h \in D'\} \subset \text{ex-lim}_{n \rightarrow \infty} A_n.$$

But since $\text{ex-lim}_{n \rightarrow \infty} A_n$ is closed and since, by Lemma 7.3(d),

$$(7.21) \quad Pg - g \in \mathcal{N}(P) = \overline{\mathcal{R}(B)} = \overline{\mathcal{R}(B|_D)}$$

for all $g \in L$, we conclude that

$$(7.22) \quad \{(f, Pg) : (f, g) \in C, f \in D\} \subset \text{ex-lim}_{n \rightarrow \infty} A_n,$$

completing the proof. \square

We conclude this section with two corollaries. The first one extends the conclusions of Theorem 7.6, and the other describes an important special case of the theorem.

7.7 Corollary Assume the hypotheses of Theorem 7.6(a) and suppose that (7.15) holds. If $h \in \mathcal{N}(P)$ and if $\{t_n\} \subset [0, \infty)$ satisfies $\lim_{n \rightarrow \infty} t_n \alpha_n = \infty$,

then $T_n(t_n)\pi_n h \rightarrow 0$. Consequently, for each $f \in P^{-1}(\bar{D})$ and $\delta \in (0, 1)$, $T_n(t)\pi_n f \rightarrow T(t)Pf$, uniformly in $\delta \leq t \leq \delta^{-1}$.

Assume the hypotheses of Theorem 7.6(b), and suppose that either (i) $\lim_{n \rightarrow \infty} \alpha_n \varepsilon_n = 0$ and (7.15) holds, or (ii) $\lim_{n \rightarrow \infty} \alpha_n \varepsilon_n = \gamma > 0$ and (7.16) holds (where Q is as in (7.13)). If $h \in \mathcal{N}(P)$ and if $\{k_n\} \subset \{0, 1, \dots\}$ satisfies $\lim_{n \rightarrow \infty} k_n \alpha_n \varepsilon_n = \infty$, then $T_n^{k_n} \pi_n h \rightarrow 0$. Consequently, for each $f \in P^{-1}(\bar{D})$ and $\delta \in (0, 1)$, $T_n^{(t/t_n)} \pi_n f \rightarrow T(t)Pf$, uniformly in $\delta \leq t \leq \delta^{-1}$.

Proof. We give the proof assuming the hypotheses of Theorem 7.6(a), the other case being similar. Let $h \in \mathcal{N}(P)$, let $\{t_n\}$ be as above, and let $\varepsilon > 0$. Choose $s \geq 0$ such that $\|S(s)h\| \leq \varepsilon/2K$, where $K = \sup_{n \geq 1} \|\pi_n\|$, and let $s_n = s \wedge t_n \alpha_n$ for each $n \geq 1$. Then

$$(7.23) \quad \|T_n(t_n)\pi_n h\| \leq \left\| T_n \left(\frac{s_n}{\alpha_n} \right) \pi_n h - \pi_n S(s)h \right\| + \|\pi_n S(s)h\| \leq \varepsilon$$

for all n sufficiently large by (7.18) and Theorem 6.1. If $f \in L$, then $f - Pf \in \mathcal{N}(P)$, so $T_n(t_n)\pi_n(f - Pf) \rightarrow 0$ whenever $\{t_n\} \subset [0, \infty)$ satisfies $\lim_{n \rightarrow \infty} t_n = t \neq 0$. If $f \in P^{-1}(\bar{D})$, this, together with the conclusion of the theorem applied to Pf , completes the proof. \square

7.8 Corollary Let Π , A , and B be linear operators on L such that B generates a strongly continuous contraction semigroup $\{S(t)\}$ on L satisfying (7.12). Assume that $\mathcal{D}(\Pi) \cap \mathcal{D}(A) \cap \mathcal{D}(B)$ is a core for B . For each α sufficiently large, suppose that an extension of $\Pi + \alpha A + \alpha^2 B$ generates a strongly continuous contraction semigroup $\{T_\alpha(t)\}$ on L . Let D be a subspace of

$$(7.24) \quad \{f \in \mathcal{D}(\Pi) \cap \mathcal{D}(A) \cap \mathcal{N}(B):$$

there exists $h \in \mathcal{D}(\Pi) \cap \mathcal{D}(A) \cap \mathcal{D}(B)$ with $Bh = -Af\}$,

and define

$$(7.25) \quad C = \{(f, P\Pi f + PAh): f \in D, h \in \mathcal{D}(\Pi) \cap \mathcal{D}(A) \cap \mathcal{D}(B), Bh = -Af\}.$$

Then C is dissipative, and if $\{(f, g) \in \bar{C}: g \in \bar{D}\}$, which is therefore single-valued, generates a strongly continuous contraction semigroup $\{T(t)\}$ on \bar{D} , then, for each $f \in \bar{D}$, $\lim_{n \rightarrow \infty} T_n(t)f = T(t)f$ for all $t \geq 0$, uniformly on bounded intervals.

Proof. Let $\{\alpha_n\}$ be a sequence of (sufficiently large) positive numbers such that $\lim_{n \rightarrow \infty} \alpha_n = \infty$, and apply Theorem 7.6(a) with $L_n = L$, $\pi_n = I$, A replaced by

$$(7.26) \quad \{(f, \Pi f + Ah): f \in D, h \in \mathcal{D}(\Pi) \cap \mathcal{D}(A) \cap \mathcal{D}(B), Bh = -Af\},$$

A_n equal to the generator of $\{T_{\alpha_n}(t)\}$, α_n replaced by α_n^2 , and $D' = \mathcal{D}(\Pi) \cap \mathcal{D}(A) \cap \mathcal{D}(B)$. Since $A_n(f + \alpha_n^{-1}h) = \Pi f + Ah + \alpha_n^{-1}\Pi h$ whenever $f \in D$, $h \in \mathcal{D}(\Pi) \cap \mathcal{D}(A) \cap \mathcal{D}(B)$, $Bh = -Af$, and $n \geq 1$, and since $\lim_{n \rightarrow \infty}$

$\alpha_n^{-2} A_n h = Bh$ for all $h \in D'$, we find that (7.17) and (7.18) hold, so the theorem is applicable. The dissipativity of C follows from the dissipativity of $\text{ex-lim}_{n \rightarrow \infty} A_n$. \square

7.9 Remark (a) Observe that in Corollary 7.8 it is necessary that $PAf = 0$ for all $f \in D$ by Lemma 7.3(d).

(b) Let $f \in \mathcal{D}(A)$ satisfy $PAf = 0$. To actually solve the equation $Bh = -Af$ for h , suppose that

$$(7.27) \quad \int_0^\infty \| (S(t) - P)g \| dt < \infty, \quad g \in L.$$

Then $h \equiv \lim_{\lambda \rightarrow 0+} (\lambda - B)^{-1} Af = \int_0^\infty (S(t) - P)Af dt$ belongs to $\mathcal{D}(B)$ (since B is closed) and satisfies $Bh = -Af$. Of course, the requirement that h belong to $\mathcal{D}(\Pi) \cap \mathcal{D}(A)$ must also be satisfied.

(c) When applying Corollary 7.8, it is not necessary to determine C explicitly. Instead, suppose a linear operator C_0 on \bar{D} can be found such that C_0 generates a strongly continuous contraction semigroup on \bar{D} and $C_0 \subset \bar{C}$. Then $\{(f, g) \in \bar{C} : g \in \bar{D}\} = C_0$ by Proposition 4.1.

(d) See Problem 20 for a generalization and Problem 22 for a closely related result. \square

8. PROBLEMS

- Define $\{T(t)\}$ on $\hat{C}(\mathbb{R})$ by $T(t)f(x) = f(x + t)$. Show that $\{T(t)\}$ is a strongly continuous contraction semigroup on L , and determine its generator A . (In particular, this requires that $\mathcal{D}(A)$ be characterized.)
- Define $\{T(t)\}$ on $\hat{C}(\mathbb{R})$ by

$$(8.1) \quad T(t)f(x) = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} f(y) \exp \left\{ -\frac{(y-x)^2}{2t} \right\} dy$$

for each $t > 0$ and $T(0) = I$. Show that $\{T(t)\}$ is a strongly continuous contraction semigroup on L , and determine its generator A .

- Prove Lemma 1.4.
- Let $\{T(t)\}$ be a strongly continuous contraction semigroup on L with generator A , and let $f \in \mathcal{D}(A^2)$.
 - Prove that

$$(8.2) \quad T(t)f = f + tAf + \int_0^t (t-s)T(s)A^2f ds, \quad t \geq 0.$$

- (b) Show that $\|Af\|^2 \leq 4\|A^2f\|\|f\|$.
5. Let A generate a strongly continuous semigroup on L . Show that $\bigcap_{n=1}^{\infty} \mathcal{D}(A^n)$ is dense in L .
6. Show directly that the linear operator $A = \frac{1}{2}d^2/dx^2$ on L satisfies conditions (a)–(c) of Theorem 2.6 when $\mathcal{D}(A)$ and L are as follows:
- $\mathcal{D}(A) = \{f \in C^2[0, 1]: \alpha_i f''(i) - (-1)^i \beta_i f'(i) = 0, i = 0, 1\}$,
 $L = C[0, 1], \alpha_0, \beta_0, \alpha_1, \beta_1 \geq 0, \alpha_0 + \beta_0 > 0, \alpha_1 + \beta_1 > 0$.
 - $\mathcal{D}(A) = \{f \in \hat{C}^2[0, \infty): \alpha_0 f''(0) - \beta_0 f'(0) = 0\}$
 $L = \hat{C}[0, \infty), \alpha_0, \beta_0 \geq 0, \alpha_0 + \beta_0 > 0$.
 - $\mathcal{D}(A) = \hat{C}^2(\mathbb{R}), L = \hat{C}(\mathbb{R})$.
- Hint:* Look for solutions of $\lambda f - \frac{1}{2}f'' = g$ of the form $f(x) = \exp\{-\sqrt{2\lambda}x\}h(x)$.
7. Show that $C_c^\infty(\mathbb{R})$ is a core for the generators of the semigroups of Problems 1 and 2.
8. In this problem, every statement involving k, l , or n is assumed to hold for all $k, l, n \geq 1$.

Let $L_1 \subset L_2 \subset L_3 \subset \dots$ be a sequence of closed subspaces of L . Let U_k, M_k , and $M_k^{(n)}$ be bounded linear operators on L . Assume that U_k and $M_k^{(n)}$ map L_n into L_n , and that for some $\mu_k > 0$, $\|M_k^{(n)}\| \leq \mu_k$ and

$$(8.3) \quad \lim_{n \rightarrow \infty} \|M_k^{(n)} - M_k\| = 0.$$

Suppose that the restriction of $A_n \equiv \sum_{j=1}^n M_j^{(n)} U_j$ to L_n is dissipative and that there exist nonnegative constants α_{kl} ($= \alpha_{lk}$), β_{kl} , and γ such that

$$(8.4) \quad \|U_k U_l f - U_l U_k f\| \leq \alpha_{kl} (\|U_k f\| + \|U_l f\|), \quad f \in L,$$

$$(8.5) \quad \sum_{j=1}^{\infty} \mu_j \alpha_{jl} \leq \gamma,$$

$$(8.6) \quad \|U_k M_l^{(n)} - M_l^{(n)} U_k\| \leq \beta_{kl},$$

and

$$(8.7) \quad \sum_{j=1}^{\infty} \mu_j \beta_{jl} \leq \gamma \mu_l.$$

Define $A = \sum_{j=1}^{\infty} M_j U_j$ on

$$(8.8) \quad \mathcal{D}(A) = \left\{ f \in \bigcup_{m=1}^{\infty} L_m : \sum_{j=1}^{\infty} \mu_j \|U_j f\| < \infty \right\}.$$

If $\mathcal{D}(A)$ is dense in L , show that \bar{A} is single-valued and generates a strongly continuous contraction semigroup on L .

Hint: Fix $\lambda > 3\gamma$ and apply Lemma 3.6. Show first that for $g \in \mathcal{D}(A)$ and $f_n = (\lambda - A_n)^{-1}g$,

$$(8.9) \quad (\lambda - \gamma) \|U_k f_n\| \leq \|U_k g\| + \sum_{j=1}^n (\beta_{kj} + \mu_j \alpha_{kj}) \|U_j f_n\|.$$

Denoting by μ the positive measure on the set of positive integers that gives mass μ_k to k , observe that the formula

$$(8.10) \quad (BF)_k = \sum_{j=1}^{\infty} (\beta_{kj} + \mu_j \alpha_{kj}) F_j$$

defines a positive bounded linear operator on $L^1(\mu)$ of norm at most 2γ .

9. As an application of Corollary 3.8, prove the following result, which yields the conclusion of Theorem 7.1 under a different set of hypotheses.

Let A and B generate strongly continuous contraction semigroups $\{T(t)\}$ and $\{S(t)\}$ on L . Let D be a dense subspace of L and $\|\cdot\|$ a norm on D with respect to which D is a Banach space. Assume that $\|f\| \geq \|f\|$ for all $f \in D$. Suppose there exists $\mu \geq 0$ such that

$$(8.11) \quad D \subset \mathcal{D}(A^2); \quad \|A^2 f\| \leq \mu \|f\|, \quad f \in D;$$

$$(8.12) \quad D \subset \mathcal{D}(B^2); \quad \|B^2 f\| \leq \mu \|f\|, \quad f \in D;$$

$$(8.13) \quad T(t): D \rightarrow D, \quad S(t): D \rightarrow D, \quad t \geq 0;$$

$$(8.14) \quad \|T(t)\| \leq e^{\mu t}, \quad \|S(t)\| \leq e^{\mu t}, \quad t \geq 0.$$

Then the closure of the restriction of $A + B$ to D is single-valued and generates a strongly continuous contraction semigroup on L .

We remark that only one of the two conditions (8.11) and (8.12) is really needed. See Ethier (1976).

10. Define the bounded linear operator B on $L \equiv C([0, 1] \times [0, 1])$ by $Bf(x, y) = \int_0^1 f(x, z) dz$, and define $A \subset L \times L$ by

$$(8.15) \quad A = \{(f, \frac{1}{2}f_{xx} + h): f \in C^2([0, 1] \times [0, 1]) \cap \mathcal{D}(B),$$

$$f_x(0, y) = f_x(1, y) = 0 \quad \text{for all } y \in [0, 1],$$

$$h \in \mathcal{N}(B)\}.$$

Show that A satisfies the conditions of Theorem 4.3.

11. Show that $\text{ex-lim}_{n \rightarrow \infty} A_n$, defined by (4.3), is closed in $L \times L$.
12. Does the dissipativity of A_n for each $n \geq 1$ imply the dissipativity of $\text{ex-lim}_{n \rightarrow \infty} A_n$?
13. In Theorem 6.1 (and Theorem 6.5), show that (a)–(c) are equivalent to the following:

- (d) There exists $\lambda > 0$ such that $(\lambda - A_n)^{-1}\pi_n g \rightarrow (\lambda - A)^{-1}g$ for all $g \in L$.
14. Let L , $\{L_n\}$, and $\{\pi_n\}$ be as in Section 6. For each $n \geq 1$, let $\{T_n(t)\}$ be a contraction semigroup on L_n , or, for each $n \geq 1$, let $\{T_n(t)\}$ be defined in terms of a linear contraction T_n on L_n and a number $\varepsilon_n > 0$ by $T_n(t) = T_n^{(t/\varepsilon_n)}$ for all $t \geq 0$; in the latter case assume that $\lim_{n \rightarrow \infty} \varepsilon_n = 0$. Let $\{T(t)\}$ be a contraction semigroup on L , let $f, g \in L$, and suppose that $\lim_{t \rightarrow \infty} T(t)f = g$ and
- $$(8.16) \quad \lim_{n \rightarrow \infty} \sup_{0 \leq t \leq t_0} \|T_n(t)\pi_n f - \pi_n T(t)f\| = 0$$
- for every $t_0 > 0$. Show that
- $$(8.17) \quad \lim_{n \rightarrow \infty} \sup_{t \geq 0} \|T_n(t)\pi_n f - \pi_n T(t)f\| = 0$$
- if and only if
- $$(8.18) \quad \lim_{n \rightarrow \infty} \sup_{t \geq 0} \|T_n(t)\pi_n g - \pi_n T(t)g\| = 0.$$
15. Using the results of Problem 2 and Theorem 6.5, prove the central limit theorem. That is, if X_1, X_2, \dots are independent, identically distributed, real-valued random variables with mean 0 and variance 1, show that $n^{-1/2} \sum_{i=1}^n X_i$ converges in distribution to a standard normal random variable as $n \rightarrow \infty$. (Define $T_n f(x) = E[f(x + n^{-1/2}X_1)]$ and $\varepsilon_n = n^{-1}$.)
16. Under the hypotheses of Theorem 6.9, show that $\text{ex-LIM } A_n$ is closed in $L \times L$.
17. Show that (6.21) implies (6.37) under the following (very reasonable) additional assumption.
- $$(8.19) \quad \text{If } f_n \in L_n \text{ for each } n \geq 1 \text{ and if, for some } n_0 \geq 1, f_n = 0 \text{ for all } n \geq n_0, \text{ then } \text{LIM } f_n = 0.$$
18. Prove Lemma 7.3 and the remarks following it.
19. Under the assumptions of Corollary 6.7, prove (6.18) using Theorem 7.6.
Hint: For each $n \geq 1$, define the contraction operator T_n on $L \times L$ by

$$(8.20) \quad T_n \begin{pmatrix} f \\ g \end{pmatrix} = \begin{pmatrix} S(\varepsilon_n)g \\ U(\varepsilon_n)f \end{pmatrix}.$$

20. Corollary 7.8 has been called a second-order limit theorem. Prove the following k th-order limit theorem as an application of Theorem 7.6.
Let A_0, A_1, \dots, A_k be linear operators on L such that A_k generates a strongly continuous contraction semigroup $\{S(t)\}$ on L satisfying (7.12). Assume that $\mathcal{D} \equiv \bigcap_{j=0}^k \mathcal{D}(A_j)$ is a core for A_k . For each α sufficiently

large, suppose that an extension of $\sum_{j=0}^k \alpha^j A_j$ generates a strongly continuous contraction semigroup $\{T_\alpha(t)\}$ on L . Let D be a subspace of

$$(8.21) \quad \left\{ f_0 \in D : \text{there exist } f_1, f_2, \dots, f_{k-1} \in D \text{ with} \right.$$

$$\left. \sum_{j=0}^m A_{k-m+j} f_j = 0 \text{ for } m = 0, \dots, k-1 \right\},$$

and define

$$(8.22) \quad C = \left\{ \left(f_0, \sum_{j=0}^{k-1} P A_j f_j \right) : f_0 \in D, f_1, \dots, f_{k-1} \text{ as above} \right\}.$$

Then C is dissipative and if $\{(f, g) \in \bar{C} : g \in \bar{D}\}$, which is therefore single-valued, generates a strongly continuous contraction semigroup $\{T(t)\}$ on \bar{D} , then, for each $f \in \bar{D}$, $\lim_{t \rightarrow \infty} T_\alpha(t)f = T(t)f$ for all $t \geq 0$, uniformly on bounded intervals.

21. Prove the following generalization of Theorem 7.6.

Let M be a closed subspace of L , let $A \subset L \times L$ be linear, and let B_1 and B_2 generate strongly continuous contraction semigroups $\{S_1(t)\}$ and $\{S_2(t)\}$ on M and L , respectively, satisfying

$$(8.23) \quad \lim_{\lambda \rightarrow 0+} \lambda \int_0^\infty e^{-\lambda t} S_1(t) f dt \equiv P_1 f \text{ exists for all } f \in M,$$

$$(8.24) \quad \lim_{\lambda \rightarrow 0+} \lambda \int_0^\infty e^{-\lambda t} S_2(t) f dt \equiv P_2 f \text{ exists for all } f \in L.$$

Assume that $\mathcal{R}(P_2) \subset M$. Let D be a subspace of $\mathcal{D}(A)$, D_1 a core for B_1 , and D_2 a core for B_2 . For $n = 1, 2, \dots$, let A_n be a linear operator on L and let $\alpha_n, \beta_n > 0$. Suppose that $\lim_{n \rightarrow \infty} \alpha_n = \infty$, $\lim_{n \rightarrow \infty} \beta_n = \infty$, and

$$(8.25) \quad \{(f, g) \in A : f \in D\} \subset \text{ex-lim}_{n \rightarrow \infty} A_n,$$

$$(8.26) \quad \{(h, B_1 h) : h \in D_1\} \subset \text{ex-lim}_{n \rightarrow \infty} \alpha_n^{-1} A_n,$$

$$(8.27) \quad \{(k, B_2 k) : k \in D_2\} \subset \text{ex-lim}_{n \rightarrow \infty} \beta_n^{-1} A_n.$$

Define $C = \{(f, P_1 P_2 g) : (f, g) \in A, f \in D\}$ and assume that $\{(f, g) \in \bar{C} : g \in \bar{D}\}$ generates a strongly continuous contraction semigroup $\{T(t)\}$ on \bar{D} . Then conclusions (a) and (b) of Theorem 7.6 hold.

22. Prove the following modification of Corollary 7.8.

Let Π , A , and B be linear operators on L such that B generates a strongly continuous contraction semigroup $\{S(t)\}$ on L satisfying (7.12). Assume that $\mathcal{D}(\Pi) \cap \mathcal{D}(A) \cap \mathcal{D}(B)$ is a core for B . For each α sufficiently large, suppose that an extension of $\Pi + \alpha A + \alpha^2 B$ generates a strongly

continuous contraction semigroup $\{T_\alpha(t)\}$ on L . Let D be a subspace of $\mathcal{D}(\Pi) \cap \mathcal{D}(A) \cap \mathcal{N}(B)$ with $\mathcal{R}(P) \subset \bar{D}$, and define $C = \{(f, PAf) : f \in D\}$. Then C is dissipative. Suppose that \bar{C} generates a strongly continuous contraction semigroup $\{U(t)\}$ on \bar{D} , and that

$$(8.28) \quad \lim_{\lambda \rightarrow 0^+} \lambda \int_0^\infty e^{-\lambda t} U(t)f dt \equiv P_0 f \text{ exists for every } f \in \bar{D}.$$

Let D_0 be a subspace of $\{f \in D : \text{there exists } h \in \mathcal{D}(\Pi) \cap \mathcal{D}(A) \cap \mathcal{D}(B) \text{ with } Bh = -Af\}$, and define

$$(8.29) \quad C_0 = \{(f, P_0 P\Pi f + P_0 PAh) : f \in D_0,$$

$$h \in \mathcal{D}(\Pi) \cap \mathcal{D}(A) \cap \mathcal{D}(B), \quad Bh = -Af\}.$$

Then C_0 is dissipative, and if $\{(f, g) \in \bar{C}_0 : g \in \bar{D}_0\}$ generates a strongly continuous contraction semigroup $\{T(t)\}$ on \bar{D}_0 , then, for each $f \in \bar{D}_0$, $\lim_{n \rightarrow \infty} T_n(t)f = T(t)f$ for all $t \geq 0$, uniformly on bounded intervals.

23. Let A generate a strongly continuous semigroup $\{T(t)\}$ on L , let $B(t) : L \rightarrow L$, $t \geq 0$, be bounded linear operators such that $\{B(t)\}$ is strongly continuous in $t \geq 0$ (i.e., $t \mapsto B(t)f$ is continuous for each $f \in L$).
- (a) Show that for each $f \in L$ there exists a unique $u : [0, \infty) \rightarrow L$ satisfying

$$(8.30) \quad u(t) = T(t)f + \int_0^t T(t-s)B(s)u(s) ds.$$

- (b) Show that if $B(t)g$ is continuously differentiable in t for each $g \in L$, and $f \in \mathcal{D}(A)$, then the solution of (8.30) satisfies

$$(8.31) \quad \frac{\partial}{\partial t} u(t) = Au(t) + B(t)u(t).$$

9. NOTES

Among the best general references on operator semigroups are Hille and Phillips (1957), Dynkin (1965), Davies (1980), Yosida (1980), and Pazy (1983).

Theorem 2.6 is due to Hille (1948) and Yosida (1948).

To the best of our knowledge, Proposition 3.3 first appeared in a paper of Watanabe (1968).

Theorem 4.3 is the linear version of a theorem of Crandall and Liggett (1971). The concept of the extended limit is due to Sova (1967) and Kurtz (1969).

Sufficient conditions for the convergence of semigroups in terms of convergence of their generators were first obtained by Neveu (1958), Skorohod (1958), and Trotter (1958). The necessary and sufficient conditions of Theorems

6.1 and 6.5 were found by Sova (1967) and Kurtz (1969). The proof given here follows Goldstein (1976). Hasegawa (1964) and Kato (1966) found necessary and sufficient conditions of a different sort. Lemma 6.4 and Corollary 6.6 are due to Chernoff (1968). Corollary 6.7 is known as the Trotter (1959) product formula. Corollary 6.8 can be found in Hille (1948). Theorems 6.9 and 6.11 were proved by Kurtz (1970a).

Theorem 7.1 was obtained by Kato (1966) assuming $\alpha < \frac{1}{2}$ and in general by Gustafsson (1966). Lemma 7.3 appears in Hille (1948). Theorem 7.6 is due to Ethier and Nagylaki (1980) and Corollary 7.7 to Kurtz (1977). Corollary 7.8 was proved by Kurtz (1973) and Kertz (1974); related results are given in Davies (1980).

Problem 4(b) is due to Kallman and Rota (1970), Problem 8 to Liggett (1972), Problem 9 to Kurtz (see Ethier (1976)), Problem 13 to Kato (1966), and Problem 14 to Norman (1977). Problem 20 is closely related to a theorem of Kertz (1978).

2

STOCHASTIC PROCESSES AND MARTINGALES

This chapter consists primarily of background material that is needed later. Section 1 defines various concepts in the theory of stochastic processes, in particular the notion of a stopping time. Section 2 gives a basic introduction to martingale theory including the optional sampling theorem, and local martingales are discussed in Section 3, in particular the existence of the quadratic variation or square bracket process. Section 4 contains additional technical material on processes and conditional expectations, including a Fubini theorem. The Doob–Meyer decomposition theorem for submartingales is given in Section 5, and some of the special properties of square integrable martingales are noted in Section 6. The semigroup of conditioned shifts on the space of progressive processes is discussed in Section 7. The optional sampling theorem for martingales indexed by a metric lattice is given in Section 8.

1. STOCHASTIC PROCESSES

A *stochastic process* X (or simply a process) with *index set* \mathcal{I} and *state space* (E, \mathcal{B}) (a measurable space) defined on a probability space (Ω, \mathcal{F}, P) is a function defined on $\mathcal{I} \times \Omega$ with values in E such that for each $t \in \mathcal{I}$, $X(t, \cdot): \Omega \rightarrow E$ is an E -valued random variable, that is, $\{\omega: X(t, \omega) \in \Gamma\} \in \mathcal{F}$ for every $\Gamma \in \mathcal{B}$. We assume throughout that E is a metric space with metric r .

and that \mathcal{B} is the Borel σ -algebra $\mathcal{B}(E)$. As is usually done, we write $X(t)$ and $X(t, \cdot)$ interchangeably.

In this chapter, with the exception of Section 8, we take $\mathcal{I} = [0, \infty)$. We are primarily interested in viewing X as a “random” function of time. Consequently, it is natural to put further restrictions on X . We say that X is *measurable* if $X: [0, \infty) \times \Omega \rightarrow E$ is $\mathcal{B}[0, \infty) \times \mathcal{F}$ -measurable. We say that X is (almost surely) *continuous* (*right continuous*, *left continuous*) if for (almost) every $\omega \in \Omega$, $X(\cdot, \omega)$ is continuous (right continuous, left continuous). Note that the statements “ X is measurable” and “ X is continuous” are not parallel in that “ X is measurable” is stronger than the statement that $X(\cdot, \omega)$ is measurable for each $\omega \in \Omega$. The function $X(\cdot, \omega)$ is called the *sample path* of the process at ω .

A collection $\{\mathcal{F}_t\} \equiv \{\mathcal{F}_t, t \in [0, \infty)\}$ of σ -algebras of sets in \mathcal{F} is a *filtration* if $\mathcal{F}_t \subset \mathcal{F}_{t+s}$ for $t, s \in [0, \infty)$. Intuitively \mathcal{F}_t corresponds to the information known to an observer at time t . In particular, for a process X we define $\{\mathcal{F}_t^X\}$ by $\mathcal{F}_t^X = \sigma(X(s): s \leq t)$; that is, \mathcal{F}_t^X is the information obtained by observing X up to time t .

We occasionally need additional structure on $\{\mathcal{F}_t\}$. We say $\{\mathcal{F}_t\}$ is *right continuous* if for each $t \geq 0$, $\mathcal{F}_t = \mathcal{F}_{t+} \equiv \bigcap_{\epsilon > 0} \mathcal{F}_{t+\epsilon}$. Note the filtration $\{\mathcal{F}_{t+}\}$ is always right continuous (Problem 7). We say $\{\mathcal{F}_t\}$ is *complete* if (Ω, \mathcal{F}, P) is complete and $\{A \in \mathcal{F}: P(A) = 0\} \subset \mathcal{F}_0$.

A process X is *adapted* to a filtration $\{\mathcal{F}_t\}$ (or simply $\{\mathcal{F}_t\}$ -adapted) if $X(t)$ is \mathcal{F}_t -measurable for each $t \geq 0$. Since \mathcal{F}_t is increasing in t , X is $\{\mathcal{F}_t\}$ -adapted if and only if $\mathcal{F}_t^X \subset \mathcal{F}_t$ for each $t \geq 0$.

A process X is $\{\mathcal{F}_t\}$ -*progressive* (or simply *progressive* if $\{\mathcal{F}_t\} = \{\mathcal{F}_t^X\}$) if for each $t \geq 0$ the restriction of X to $[0, t] \times \Omega$ is $\mathcal{B}[0, t] \times \mathcal{F}_t$ -measurable. Note that if X is $\{\mathcal{F}_t\}$ -progressive, then X is $\{\mathcal{F}_t\}$ -adapted and measurable, but the converse is not necessarily the case (see Section 4 however). However, every right (left) continuous $\{\mathcal{F}_t\}$ -adapted process is $\{\mathcal{F}_t\}$ -progressive (Problem 1).

There are a variety of notions of equivalence between two stochastic processes. For $0 \leq t_1 < t_2 < \dots < t_m$, let μ_{t_1, \dots, t_m} be the probability measure on $\mathcal{B}(E) \times \dots \times \mathcal{B}(E)$ induced by the mapping $(X(t_1), \dots, X(t_m)) \rightarrow E^m$, that is, $\mu_{t_1, \dots, t_m}(\Gamma) = P\{(X(t_1), \dots, X(t_m)) \in \Gamma\}$, $\Gamma \in \mathcal{B}(E) \times \dots \times \mathcal{B}(E)$. The probability measures $\{\mu_{t_1, \dots, t_m}: m \geq 1, 0 \leq t_1 < \dots < t_m\}$ are called the *finite-dimensional distributions* of X . If X and Y are stochastic processes with the same finite-dimensional distributions, then we say Y is a *version* of X (and X is a version of Y). Note that X and Y need not be defined on the same probability space. If X and Y are defined on the same probability space and for each $t \geq 0$, $P\{X(t) = Y(t)\} = 1$, then we say Y is a *modification* of X . (We are implicitly assuming that $(X(t), Y(t))$ is an $E \times E$ -valued random variable, which is always the case if E is separable.) If Y is a modification of X , then clearly Y is a version of X . Finally if there exists $N \in \mathcal{F}$ such that $P(N) = 0$ and $X(\cdot, \omega) = Y(\cdot, \omega)$ for all $\omega \notin N$, then we say X and Y are *indistinguishable*. If X and Y are indistinguishable, then clearly Y is a modification of X .

A random variable τ with values in $[0, \infty]$ is an $\{\mathcal{F}_t\}$ -stopping time if $\{\tau \leq t\} \in \mathcal{F}_t$, for every $t \geq 0$. (Note that we allow $\tau = \infty$.) If $\tau < \infty$ a.s., we say τ is finite a.s. If $\tau \leq T < \infty$ for some constant T , we say τ is bounded. In some sense a stopping time is a random time that is recognizable by an observer whose information at time t is \mathcal{F}_t .

If τ is an $\{\mathcal{F}_t\}$ -stopping time, then for $s < t$, $\{\tau \leq s\} \in \mathcal{F}_s \subset \mathcal{F}_t$, $\{\tau < t\} = \bigcup_n \{\tau \leq t - 1/n\} \in \mathcal{F}_t$, and $\{\tau = t\} = \{\tau \leq t\} - \{\tau < t\} \in \mathcal{F}_t$. If τ is discrete (i.e., if there exists a countable set $D \subset [0, \infty]$ such that $\{\tau \in D\} = \Omega$), then τ is an $\{\mathcal{F}_t\}$ -stopping time if and only if $\{\tau = t\} \in \mathcal{F}_t$, for each $t \in D \cap [0, \infty)$.

1.1 Lemma A $[0, \infty]$ -valued random variable τ is an $\{\mathcal{F}_{t+}\}$ -stopping time if and only if $\{\tau < t\} \in \mathcal{F}_t$, for every $t \geq 0$.

Proof. If $\{\tau < t\} \in \mathcal{F}_t$, for every $t \geq 0$, then $\{\tau < t + n^{-1}\} \in \mathcal{F}_{t+m-1}$ for $n \geq m$ and $\{\tau \leq t\} = \bigcap_n \{\tau < t + n^{-1}\} \in \bigcap_m \mathcal{F}_{t+m-1} = \mathcal{F}_{t+}$. The necessity was observed above. \square

1.2 Proposition Let τ_1, τ_2, \dots be $\{\mathcal{F}_t\}$ -stopping times and let $c \in [0, \infty)$. Then the following hold.

- (a) $\tau_1 + c$ and $\tau_1 \wedge c$ are $\{\mathcal{F}_t\}$ -stopping times.
- (b) $\sup_n \tau_n$ is an $\{\mathcal{F}_t\}$ -stopping time.
- (c) $\min_{k \leq n} \tau_k$ is an $\{\mathcal{F}_t\}$ -stopping time for each $n \geq 1$.
- (d) If $\{\mathcal{F}_t\}$ is right continuous, then $\inf_n \tau_n$, $\underline{\lim}_{n \rightarrow \infty} \tau_n$, and $\overline{\lim}_{n \rightarrow \infty} \tau_n$ are $\{\mathcal{F}_t\}$ -stopping times.

Proof. We prove (b) and (d) and leave (a) and (c) to the reader. Note that $\{\sup_n \tau_n \leq t\} = \bigcap_n \{\tau_n \leq t\} \in \mathcal{F}_t$, so (b) follows. Similarly $\{\inf_n \tau_n < t\} = \bigcup_n \{\tau_n < t\} \in \mathcal{F}_t$, so if $\{\mathcal{F}_t\}$ is right continuous, then $\inf_n \tau_n$ is a stopping time by Lemma 1.1. Since $\underline{\lim}_{n \rightarrow \infty} \tau_n = \sup_m \inf_{n \geq m} \tau_n$ and $\overline{\lim}_{n \rightarrow \infty} \tau_n = \inf_m \sup_{n \geq m} \tau_n$, (d) follows. \square

By Proposition 1.2(a) every stopping time τ can be approximated by a sequence of bounded stopping times, that is, $\lim_{n \rightarrow \infty} \tau \wedge n = \tau$. This fact is very useful in proving theorems about stopping times. A second equally useful approximation is the approximation of arbitrary stopping times by a nonincreasing sequence of discrete stopping times.

1.3 Proposition For $n = 1, 2, \dots$, let $0 = t_0^n < t_1^n < \dots$ and $\lim_{k \rightarrow \infty} t_k^n = \infty$, and suppose that $\lim_{n \rightarrow \infty} \sup_k (t_{k+1}^n - t_k^n) = 0$. Let τ be an $\{\mathcal{F}_{t+}\}$ -stopping time and define

$$(1.1) \quad \tau_n = \begin{cases} t_{k+1}^n & \text{if } t_k^n \leq \tau < t_{k+1}^n, \\ \infty & \text{if } \tau = \infty. \end{cases} \quad k \geq 0$$

52 STOCHASTIC PROCESSES AND MARTINGALES

Then τ_n is an $\{\mathcal{F}_t\}$ -stopping time and $\lim_{n \rightarrow \infty} \tau_n = \tau$. If in addition $\{t_k^n\} \subset \{t_k^{n+1}\}$, then $\tau_n \geq \tau_{n+1}$.

Proof. Let $\gamma_n(t) = \max \{t_k^n : t_k^n \leq t\}$. Then

$$(1.2) \quad \{\tau_n \leq t\} = \{\tau_n \leq \gamma_n(t)\} = \{\tau < \gamma_n(t)\} \in \mathcal{F}_{\gamma_n(t)} \subset \mathcal{F}_t.$$

The rest is clear. \square

Recall the intuitive description of \mathcal{F}_t as the information known to an observer at time t . For an $\{\mathcal{F}_t\}$ -stopping time τ , the σ -algebra \mathcal{F}_τ should have the same intuitive meaning. For technical reasons \mathcal{F}_τ is defined by

$$(1.3) \quad \mathcal{F}_\tau = \{A \in \mathcal{F} : A \cap \{\tau \leq t\} \in \mathcal{F}_t \text{ for all } t \geq 0\}.$$

Similarly, $\mathcal{F}_{\tau+}$ is defined by replacing \mathcal{F}_τ by $\mathcal{F}_{\tau+}$. See Problem 6 for some motivation as to why the definition is reasonable. Given an E -valued process X , define $X(\infty) \equiv x_0$ for some fixed $x_0 \in E$.

1.4 Proposition Let τ and σ be $\{\mathcal{F}_t\}$ -stopping times, let γ be a nonnegative \mathcal{F}_t -measurable random variable, and let X be an $\{\mathcal{F}_t\}$ -progressive E -valued process. Define X^τ and Y by $X^\tau(t) = X(\tau \wedge t)$ and $Y(t) = X(\tau + t)$, and define $\mathcal{G}_t = \mathcal{F}_{t \wedge \tau}$ and $\mathcal{H}_t = \mathcal{F}_{t + \tau}$, $t \geq 0$. (Recall that $\tau \wedge t$ and $\tau + t$ are stopping times.) Then the following hold:

- (a) \mathcal{F}_τ is a σ -algebra.
- (b) τ and $\tau \wedge \sigma$ are \mathcal{F}_τ -measurable.
- (c) If $\tau \leq \sigma$, then $\mathcal{F}_\tau \subset \mathcal{F}_\sigma$.
- (d) $X(\tau)$ is \mathcal{F}_τ -measurable.
- (e) $\{\mathcal{G}_t\}$ is a filtration and X^τ is both $\{\mathcal{G}_t\}$ -progressive and $\{\mathcal{F}_t\}$ -progressive.
- (f) $\{\mathcal{H}_t\}$ is a filtration and Y is $\{\mathcal{H}_t\}$ -progressive.
- (g) $\tau + \gamma$ is an $\{\mathcal{F}_t\}$ -stopping time.

Proof. (a) Clearly \emptyset and Ω are in \mathcal{F}_τ , since \mathcal{F}_τ is a σ -algebra and $\{\tau \leq t\} \in \mathcal{F}_t$. If $A \cap \{\tau \leq t\} \in \mathcal{F}_t$, then $A^c \cap \{\tau \leq t\} = \{\tau \leq t\} - A \cap \{\tau \leq t\} \in \mathcal{F}_t$, and hence $A \in \mathcal{F}_\tau$ implies $A^c \in \mathcal{F}_\tau$. Similarly $A_k \cap \{\tau \leq t\} \in \mathcal{F}_t$, $k = 1, 2, \dots$, implies $(\bigcup_k A_k) \cap \{\tau \leq t\} = \bigcup_k (A_k \cap \{\tau \leq t\}) \in \mathcal{F}_t$, and hence \mathcal{F}_τ is closed under countable unions.

(b) For each $c \geq 0$ and $t \geq 0$,

$$(1.4) \quad \{\tau \wedge \sigma \leq c\} \cap \{\tau \leq t\} = \{\tau \wedge \sigma \leq c \wedge t\} \cap \{\tau \leq t\} \\ = (\{\tau \leq c \wedge t\} \cup \{\sigma \leq c \wedge t\}) \cap \{\tau \leq t\} \in \mathcal{F}_t.$$

Hence $\{\tau \wedge \sigma \leq c\} \in \mathcal{F}_\tau$ and $\tau \wedge \sigma$ is \mathcal{F}_τ -measurable, as is τ (take $\sigma = \tau$).

(c) If $A \in \mathcal{F}_t$, then $A \cap \{\sigma \leq t\} = A \cap \{\tau \leq t\} \cap \{\sigma \leq t\} \in \mathcal{F}$, for all $t \geq 0$. Hence $A \in \mathcal{F}_\sigma$.

(d) Fix $t \geq 0$. By (b), $\tau \wedge t$ is \mathcal{F}_t -measurable. Consequently the mapping $\omega \rightarrow (\tau(\omega) \wedge t, \omega)$ is a measurable mapping of (Ω, \mathcal{F}) into $([0, t] \times \Omega, \mathcal{B}[0, t] \times \mathcal{F})$ and since X is $\{\mathcal{F}_t\}$ -progressive, $(s, \omega) \rightarrow X(s, \omega)$ is a measurable mapping of $([0, t] \times \Omega, \mathcal{B}[0, t] \times \mathcal{F})$ into $(E, \mathcal{B}(E))$. Since $X(\tau \wedge t)$ is the composition of these two mappings, it is \mathcal{F}_t -measurable. Finally, for $\Gamma \in \mathcal{B}(E)$, $\{X(\tau) \in \Gamma\} \cap \{\tau \leq t\} = \{X(\tau \wedge t) \in \Gamma\} \cap \{\tau \leq t\} \in \mathcal{F}_t$, and hence $\{X(\tau) \in \Gamma\} \in \mathcal{F}_t$.

(e) By (a) and (c), $\{\mathcal{G}_t\}$ is a filtration, and since $\mathcal{G}_t \subset \mathcal{F}_t$, by (c), X^t is $\{\mathcal{F}_t\}$ -progressive if it is $\{\mathcal{G}_t\}$ -progressive. To see that X^t is $\{\mathcal{G}_t\}$ -progressive, we begin by showing that if $s \leq t$ and $H \in \mathcal{B}[0, t] \times \mathcal{F}_s$, then

$$(1.5) \quad H \cap ([0, t] \times \{\tau \wedge t \geq s\}) \in \mathcal{B}[0, t] \times \mathcal{F}_{\tau \wedge t} = \mathcal{B}[0, t] \times \mathcal{G}_t.$$

To verify this, note that the collection $\mathcal{H}_{s,t}$ of $H \in \mathcal{B}[0, t] \times \mathcal{F}_s$, satisfying (1.5) is a σ -algebra. Since $A \in \mathcal{F}_s$ implies $A \cap \{\tau \wedge t \geq s\} \in \mathcal{F}_{\tau \wedge t}$, it follows that if $B \in \mathcal{B}[0, t]$ and $A \in \mathcal{F}_s$, then

$$(1.6) \quad (B \times A) \cap ([0, t] \times \{\tau \wedge t \geq s\}) \\ = B \times (A \cap \{\tau \wedge t \geq s\}) \in \mathcal{B}[0, t] \times \mathcal{G}_t,$$

so $B \times A \in \mathcal{H}_{s,t}$. But the collection of $B \times A$ of this form generates $\mathcal{B}[0, t] \times \mathcal{F}_s$.

Finally, for $\Gamma \in \mathcal{B}(E)$ and $t \geq 0$,

$$(1.7) \quad \{(s, \omega) \in [0, t] \times \Omega: X(\tau(\omega) \wedge s, \omega) \in \Gamma\} \\ = \{(s, \omega): X(\tau(\omega) \wedge s, \omega) \in \Gamma, \tau(\omega) \wedge t \leq s \leq t\} \\ \cup \{(s, \omega): X(s, \omega) \in \Gamma, s < \tau(\omega) \wedge t\} \\ = (\{(s, \omega): \tau(\omega) \wedge t \leq s \leq t\} \cap ([0, t] \times \{X(\tau \wedge t) \in \Gamma\})) \\ \cup \bigcup_n \bigcup_k \left(\left\{ (s, \omega): X(s, \omega) \in \Gamma, s < \frac{k}{n} \right\} \right. \\ \left. \cap \left\{ (s, \omega): \frac{k}{n} \leq \tau(\omega) \wedge t \right\} \right) \in \mathcal{B}[0, t] \times \mathcal{G}_t,$$

since

$$(1.8) \quad \{(s, \omega): \tau(\omega) \wedge t \leq s \leq t\} \\ = \bigcap_n \bigcup_k \left(\left[\frac{k}{n}, t \right] \times \left\{ \frac{k}{n} \leq \tau \wedge t < \frac{k+1}{n} \right\} \right) \in \mathcal{B}[0, t] \times \mathcal{G}_t,$$

and since the last set on the right in (1.7) is in $\mathcal{B}[0, t] \times \mathcal{G}_t$ by (1.5).

(f) Again $\{\mathcal{H}_t\}$ is a filtration by (a) and (c). Fix $t \geq 0$. By part (e) the mapping $(s, \omega) \rightarrow X((\tau(\omega) + t) \wedge s, \omega)$ from $([0, \infty] \times \Omega, \mathcal{B}[0, \infty] \times \mathcal{F}_{t+1})$

54 STOCHASTIC PROCESSES AND MARTINGALES

into $(E, \mathcal{B}(E))$ is measurable, as is the mapping $(u, \omega) \rightarrow (\tau(\omega) + u, \omega)$ from $([0, t] \times \Omega, \mathcal{B}[0, t] \times \mathcal{F}_{t+})$ into $([0, \infty] \times \Omega, \mathcal{B}[0, \infty] \times \mathcal{F}_{t+})$. The mapping $(u, \omega) \rightarrow X(\tau(\omega) + u, \omega)$ from $([0, t] \times \Omega, \mathcal{B}[0, t] \times \mathcal{F}_{t+})$ into $(E, \mathcal{B}(E))$ is a composition of the first two mappings so it too is measurable. Since $\mathcal{H}_t = \mathcal{F}_{t+}$, Y is $\{\mathcal{H}_t\}$ -progressive.

(g) Let $\gamma_n = [ny]/n$. Note that $\{\tau + \gamma_n \leq t\} \cap \{\gamma_n = k/n\} = \{\tau \leq t - k/n\} \cap \{\gamma_n = k/n\} \in \mathcal{F}_{t-k/n}$, since $\{\gamma_n = k/n\} \in \mathcal{F}_t$. Consequently, $\{\tau + \gamma_n \leq t\} \in \mathcal{F}_t$. Since $\tau + \gamma = \sup_n (\tau + \gamma_n)$, part (g) follows by Proposition 1.2(b). \square

Let X be an E -valued process and let $\Gamma \in \mathcal{B}(E)$. The *first entrance time into* Γ is defined by

$$(1.9) \quad \tau_e(\Gamma) = \inf \{t: X(t) \in \Gamma\}$$

(where $\inf \emptyset = \infty$), and for a $[0, \infty]$ -valued random variable σ , the *first entrance time into* Γ *after* σ is defined by

$$(1.10) \quad \tau_e(\Gamma, \sigma) = \inf \{t \geq \sigma: X(t) \in \Gamma\}.$$

For each $\omega \in \Omega$ and $0 \leq s \leq t$, let $F_X(s, t, \omega) \subset E$ be the closure of $\{X(u, \omega): s \leq u \leq t\}$. The *first contact time with* Γ is defined by

$$(1.11) \quad \tau_c(\Gamma) = \inf \{t: F_X(0, t) \cap \Gamma \neq \emptyset\}$$

and the *first contact time with* Γ *after* σ by

$$(1.12) \quad \tau_c(\Gamma, \sigma) = \inf \{t \geq \sigma: F_X(\sigma, t) \cap \Gamma \neq \emptyset\}.$$

The *first exit time from* Γ (*after* σ) is the first entrance time of Γ^c (*after* σ). Although intuitively the above times are “recognizable” to our observer, they are not in general stopping times (or even random variables). We do, however, have the following result, which is sufficient for our purposes.

1.5 Proposition Suppose that X is a right continuous, $\{\mathcal{F}_t\}$ -adapted, E -valued process and that σ is an $\{\mathcal{F}_t\}$ -stopping time.

(a) If Γ is closed and X has left limits at each $t > 0$ or if Γ is compact, then $\tau_e(\Gamma, \sigma)$ is an $\{\mathcal{F}_t\}$ -stopping time.

(b) If Γ is open, then $\tau_e(\Gamma, \sigma)$ is an $\{\mathcal{F}_{t+}\}$ -stopping time.

Proof. Using the right continuity of X , if Γ is open,

$$(1.13) \quad \{\tau_e(\Gamma, \sigma) < t\} = \bigcup_{s \in \Omega \cap (0, t)} \{X(s) \in \Gamma\} \cap \{\sigma < s\} \in \mathcal{F}_t,$$

implying part (b). For $n = 1, 2, \dots$ let $\Gamma_n = \{x: r(x, \Gamma) < 1/n\}$. Then, under the conditions of part (a), $\tau_e(\Gamma, \sigma) = \lim_{n \rightarrow \infty} \tau_e(\Gamma_n, \sigma)$, and

$$(1.14) \quad \{\tau_e(\Gamma, \sigma) \leq t\} = (\{\sigma \leq t\} \cap \{X(t) \in \Gamma\}) \cup \bigcap_n \{\tau_e(\Gamma_n, \sigma) < t\} \in \mathcal{F}_t. \quad \square$$

Under slightly more restrictive hypotheses on $\{\mathcal{F}_t\}$, a much more general result than Proposition 1.5 holds. We do not need this generality, so we simply state the result without proof.

1.6 Theorem Let $\{\mathcal{F}_t\}$ be complete and right continuous, and let X be an E -valued $\{\mathcal{F}_t\}$ -progressive process. Then for each $\Gamma \in \mathcal{B}(E)$, $\tau_\epsilon(\Gamma)$ is an $\{\mathcal{F}_t\}$ -stopping time.

Proof. See, for example, Elliott (1982), page 50. \square

2. MARTINGALES

A real-valued process X with $E[|X(t)|] < \infty$ for all $t \geq 0$ and adapted to a filtration $\{\mathcal{F}_t\}$ is an $\{\mathcal{F}_t\}$ -martingale if

$$(2.1) \quad E[X(t+s)|\mathcal{F}_t] = X(t), \quad t, s \geq 0,$$

is an $\{\mathcal{F}_t\}$ -submartingale if

$$(2.2) \quad E[X(t+s)|\mathcal{F}_t] \geq X(t), \quad t, s \geq 0,$$

and is an $\{\mathcal{F}_t\}$ -supermartingale if the inequality in (2.2) is reversed. Note that X is a supermartingale if $-X$ is a submartingale, and that X is a martingale if both X and $-X$ are submartingales. Consequently, results proved for submartingales immediately give analogous results for martingales and supermartingales. If $\{\mathcal{F}_t\} = \{\mathcal{F}_t^X\}$ we simply say X is a martingale (submartingale, supermartingale).

Jensen's inequality gives the following.

2.1 Proposition (a) Suppose X is an $\{\mathcal{F}_t\}$ -martingale, φ is convex, and $E[|\varphi(X(t))|] < \infty$ for all $t \geq 0$. Then $\varphi \circ X$ is an $\{\mathcal{F}_t\}$ -submartingale.

(b) Suppose X is an $\{\mathcal{F}_t\}$ -submartingale, φ is convex and nondecreasing, and $E[|\varphi(X(t))|] < \infty$ for all $t \geq 0$. Then $\varphi \circ X$ is an $\{\mathcal{F}_t\}$ -submartingale.

Proof. By Jensen's inequality, for $t, s \geq 0$,

$$(2.3) \quad E[\varphi(X(t+s))|\mathcal{F}_t] \geq \varphi(E[X(t+s)|\mathcal{F}_t]) \geq \varphi(X(t)).$$

Note that for part (a) the last inequality is in fact equality, and in part (b) the last inequality follows from the assumption that φ is nondecreasing. \square

2.2 Lemma Let τ_1 and τ_2 be $\{\mathcal{F}_t\}$ -stopping times assuming values in $\{t_1, t_2, \dots, t_m\} \subset [0, \infty)$. If X is an $\{\mathcal{F}_t\}$ -submartingale, then

$$(2.4) \quad E[X(\tau_2)|\mathcal{F}_{\tau_1}] \geq X(\tau_1 \wedge \tau_2).$$

56 STOCHASTIC PROCESSES AND MARTINGALES

Proof. Assume $t_1 < t_2 < \dots < t_m$. We must show that for every $A \in \mathcal{F}_{t_1}$

$$(2.5) \quad \int_A X(\tau_2) dP \geq \int_A X(\tau_2 \wedge t_1) dP.$$

Since $A = \bigcup_{i=1}^m (A \cap \{\tau_1 = t_i\})$, it is sufficient to show that

$$(2.6) \quad \int_{A \cap \{\tau_1 = t_i\}} X(\tau_2) dP \geq \int_{A \cap \{\tau_1 = t_i\}} X(\tau_2 \wedge t_i) dP = \int_{A \cap \{\tau_1 = t_i\}} X(\tau_2 \wedge t_i) dP,$$

but since $A \cap \{\tau_1 = t_i\} \in \mathcal{F}_{t_i}$, (2.5) holds if

$$(2.7) \quad E[X(\tau_2) | \mathcal{F}_{t_i}] \geq X(\tau_2 \wedge t_i).$$

Finally, observe that

$$\begin{aligned} (2.8) \quad & E[X(\tau_2 \wedge t_{k+1}) | \mathcal{F}_{t_k}] \\ &= E[X(t_{k+1}) \chi_{\{\tau_2 > t_k\}} + X(\tau_2 \wedge t_k) \chi_{\{\tau_2 \leq t_k\}} | \mathcal{F}_{t_k}] \\ &= E[X(t_{k+1}) | \mathcal{F}_{t_k}] \chi_{\{\tau_2 > t_k\}} + X(\tau_2 \wedge t_k) \chi_{\{\tau_2 \leq t_k\}} \\ &\geq X(t_k) \chi_{\{\tau_2 > t_k\}} + X(\tau_2 \wedge t_k) \chi_{\{\tau_2 \leq t_k\}} \\ &= X(\tau_2 \wedge t_k). \end{aligned}$$

Starting with $k = m - 1$ and observing that $\tau_2 = \tau_2 \wedge t_m$, (2.8) may be iterated to give (2.7). \square

The following is a simple application of Lemma 2.2. Let $x^+ = x \vee 0$.

2.3 Lemma Let X be a submartingale, $T > 0$, and $F \subset [0, T]$ be finite. Then for each $x > 0$,

$$(2.9) \quad P\left\{\max_{t \in F} X(t) \geq x\right\} \leq x^{-1} E[X^+(T)]$$

and

$$(2.10) \quad P\left\{\min_{t \in F} X(t) \leq -x\right\} \leq x^{-1} (E[X^+(T)] - E[X(0)]).$$

Proof. Let $\tau = \min\{t \in F: X(t) \geq x\}$ and set $\tau_1 = \tau \wedge T$ and $\tau_2 = T$ in (2.4). Then

$$(2.11) \quad E[X(T)] \geq E[X(\tau \wedge T)] = E[X(\tau) \chi_{\{\tau < \infty\}}] + E[X(T) \chi_{\{\tau = \infty\}}],$$

and hence

$$(2.12) \quad E[X(T) \chi_{\{\tau < \infty\}}] \geq E[X(\tau) \chi_{\{\tau < \infty\}}] \geq x P\{\tau < \infty\} = x P\left\{\max_{t \in F} X(t) \geq x\right\},$$

which implies (2.9). The proof of (2.10) is similar. \square

2.4 Corollary Let X be a submartingale and let $F \subset [0, \infty)$ be countable. Then for each $x > 0$ and $T > 0$,

$$(2.13) \quad P\left\{\sup_{t \in F \cap [0, T]} X(t) \geq x\right\} \leq x^{-1} E[X^+(T)]$$

and

$$(2.14) \quad P\left\{\inf_{t \in F \cap [0, T]} X(t) \leq -x\right\} \leq x^{-1}(E[X^+(T)] - E[X(0)]).$$

Proof. Let $F_1 \subset F_2 \subset \dots$ be finite and $F = \bigcup F_n$. Then, for $0 < y < x$,

$$(2.15) \quad P\left\{\sup_{t \in F \cap [0, T]} X(t) \geq x\right\} \leq \lim_{n \rightarrow \infty} P\left\{\max_{t \in F_n \cap [0, T]} X(t) \geq y\right\} \leq y^{-1} E[X^+(T)].$$

Letting $y \rightarrow x$ we obtain (2.13), and (2.14) follows similarly. \square

Let X be a real-valued process, and let $F \subset [0, \infty)$ be finite. For $a < b$ define $\tau_1 = \min\{t \in F : X(t) \leq a\}$, and for $k = 1, 2, \dots$ define $\sigma_k = \min\{t > \tau_k : t \in F, X(t) \geq b\}$ and $\tau_{k+1} = \min\{t > \sigma_k : t \in F, X(t) \leq a\}$. Define

$$(2.16) \quad U(a, b, F) = \max\{k : \sigma_k < \infty\}.$$

The quantity $U(a, b, F)$ is called the number of *upcrossings* of the interval (a, b) by X restricted to F .

2.5 Lemma Let X be a submartingale. If $T > 0$ and $F \subset [0, T]$ is finite, then

$$(2.17) \quad E[U(a, b, F)] \leq \frac{E[(X(T) - a)^+]}{b - a}.$$

Proof. Since $\sigma_k \wedge T \leq \tau_{k+1} \wedge T$, Lemma 2.2 implies

$$\begin{aligned} (2.18) \quad 0 &\leq E\left[\sum_{k=1}^{\infty} (X(\tau_{k+1} \wedge T) - X(\sigma_k \wedge T))\right] \\ &= E\left[\sum_{k=1}^{U(a, b, F)} (X(\tau_{k+1} \wedge T) - X(\sigma_k \wedge T))\right] \\ &= E\left[-\sum_{k=2}^{U(a, b, F)} (X(\sigma_k \wedge T) - X(\tau_k \wedge T)) \right. \\ &\quad \left. + X(\tau_{U(a, b, F)+1} \wedge T) - (X(\sigma_1 \wedge T) - a) - a\right] \\ &\leq E[-(b - a)U(a, b, F) + X(\tau_{U(a, b, F)+1} \wedge T) - a] \\ &\leq E[-(b - a)U(a, b, F) + (X(T) - a)^+], \end{aligned}$$

which gives (2.17). \square

58 STOCHASTIC PROCESSES AND MARTINGALES

2.6 Corollary Let X be a submartingale. Let $T > 0$, let F be a countable subset of $[0, T]$, and let $F_1 \subset F_2 \subset \dots$ be finite subsets with $F = \bigcup F_n$. Define $U(a, b, F) = \lim_{n \rightarrow \infty} U(a, b, F_n)$. Then $U(a, b, F)$ depends only on F (not the particular sequence $\{F_n\}$) and

$$(2.19) \quad E[U(a, b, F)] \leq \frac{E[(X(T) - a)^+]}{b - a}.$$

Proof. The existence of the limit defining $U(a, b, F)$ as well as the independence of $U(a, b, F)$ from the choice of $\{F_n\}$ follows from the fact that $G \subset H$ implies $U(a, b, G) \leq U(a, b, H)$. Consequently (2.19) follows from (2.17) and the monotone convergence theorem. \square

One implication of the upcrossing inequality (2.19) is that submartingales have modifications with “nice” sample paths. To see this we need the following lemmas.

2.7 Lemma Let (E, r) be a metric space and let $x: [0, \infty) \rightarrow E$. Suppose $x(t+) \equiv \lim_{s \rightarrow t, +} x(s)$ exists for all $t \geq 0$ and $x(t-) \equiv \lim_{s \rightarrow t, -} x(s)$ exists for all $t > 0$. Then there exists a countable set Γ such that for $t \in (0, \infty) - \Gamma$, $x(t-) = x(t) = x(t+)$.

Let $\Gamma_n = \{t: r(x(t-), x(t)) \vee r(x(t-), x(t+)) \vee r(x(t), x(t+)) > n^{-1}\}$. Then $\Gamma_n \cap [0, T]$ is finite for each $T > 0$.

Proof. Since we may take $\Gamma = \bigcup_n \Gamma_n$, it is enough to verify the last statement. If $\Gamma_n \cap [0, T]$ had a limit point t then either $x(t-)$ or $x(t+)$ would fail to exist. Consequently $\Gamma_n \cap [0, T]$ must be finite. \square

2.8 Lemma Let (E, r) be a metric space, let F be a dense subset of $[0, \infty)$, and let $x: F \rightarrow E$. If for each $t \geq 0$

$$(2.20) \quad y(t) \equiv \lim_{\substack{s \rightarrow t, + \\ s \in F}} x(s)$$

exists, then y is right continuous. If for each $t > 0$

$$(2.21) \quad y^-(t) \equiv \lim_{\substack{s \rightarrow t, - \\ s \in F}} x(s)$$

exists, then y^- is left continuous (on $(0, \infty)$). If for each $t > 0$ both (2.20) and (2.21) exist, then $y^-(t) = y(t-)$ for all $t > 0$.

Proof. Suppose (2.20) exists for all $t \geq 0$. Given $t_0 > 0$ and $\varepsilon > 0$, there exists a $\delta > 0$ such that $r(y(t_0), x(s)) \leq \varepsilon$ for all $s \in F \cap (t_0, t_0 + \delta)$, and hence

$$(2.22) \quad r(y(t_0), y(s)) = \lim_{\substack{u \rightarrow s+ \\ u \in F}} r(y(t_0), x(u)) \leq \varepsilon$$

for all $s \in (t_0, t_0 + \delta)$ and the right continuity of y follows. The proof of the other parts is similar. \square

Let F be a countable dense subset of $[0, \infty)$. For a submartingale X , Corollary 2.4 implies $P\{\sup_{t \in F \cap [0, T]} X(t) < \infty\} = 1$ and $P\{\inf_{t \in F \cap [0, T]} X(t) > -\infty\} = 1$ for each $T > 0$, and Corollary 2.6 gives $P\{U(a, b, F \cap [0, T]) < \infty\} = 1$ for all $a < b$ and $T > 0$. Let

$$(2.23) \quad \Omega_0 = \bigcap_{n=1}^{\infty} \left(\left\{ \sup_{t \in F \cap [0, n]} X(t) < \infty \right\} \cap \left\{ \inf_{t \in F \cap [0, n]} X(t) > -\infty \right\} \cap \bigcap_{\substack{a < b \\ a, b \in \mathbb{Q}}} \{U(a, b, F \cap [0, n]) < \infty\} \right).$$

Then $P(\Omega_0) = 1$. For $\omega \in \Omega_0$,

$$(2.24) \quad Y(t, \omega) = \lim_{\substack{s \rightarrow t+ \\ s \in F}} X(s, \omega)$$

exists for all $t \geq 0$, and

$$(2.25) \quad Y^-(t, \omega) = \lim_{\substack{s \rightarrow t- \\ s \in F}} X(s, \omega)$$

exists for all $t > 0$; furthermore, $Y(\cdot, \omega)$ is right continuous and has left limits with $Y(t^-, \omega) = Y^-(t, \omega)$ for all $t > 0$ (Problem 9). Define $Y(t, \omega) = 0$ for all $\omega \notin \Omega_0$ and $t \geq 0$.

2.9 Proposition Let X be a submartingale, and let Y be defined by (2.24). Then $\Gamma \equiv \{t: P\{Y(t) \neq Y(t^-)\} > 0\}$ is countable, $P\{X(t) = Y(t)\} = 1$ for $t \notin \Gamma$, and

$$(2.26) \quad \tilde{X}(t) = \begin{cases} Y(t), & t \in [0, \infty) - \Gamma, \\ X(t), & t \in \Gamma, \end{cases}$$

defines a modification of X almost all of whose sample paths have right and left limits at all $t \geq 0$ and are right continuous at all $t \notin \Gamma$.

Proof. For real-valued random variables η and ξ (defined on (Ω, \mathcal{F}, P)) define

$$(2.27) \quad \gamma(\eta, \xi) = \inf \{\varepsilon > 0: P\{|\eta - \xi| > \varepsilon\} < \varepsilon\}.$$

60 STOCHASTIC PROCESSES AND MARTINGALES

Then γ is a metric corresponding to convergence in probability (Problem 8). Since Y has right and left limits in this metric at all $t \geq 0$, Lemma 2.7 implies Γ is countable.

Let $\alpha \in \mathbb{R}$. Then $X \vee \alpha$ is a submartingale by Proposition 2.1 so for any $T > 0$,

$$(2.28) \quad \alpha \leq X(t) \vee \alpha \leq E[X(T) \vee \alpha | \mathcal{F}_t^X], \quad 0 \leq t \leq T,$$

and since $\{E[X(T) \vee \alpha | \mathcal{F}_t^X] : 0 \leq t \leq T\}$ is uniformly integrable (Problem 10), it follows that $\{X(t) \vee \alpha : 0 \leq t \leq T\}$ is uniformly integrable. Therefore

$$(2.29) \quad X(t) \vee \alpha \leq \lim_{\substack{s \rightarrow t+ \\ s \in \mathbb{Q}}} E[X(s) \vee \alpha | \mathcal{F}_s^X] = E[Y(t) \vee \alpha | \mathcal{F}_t^X], \quad t \geq 0.$$

Furthermore if $t \notin \Gamma$, then

$$(2.30) \quad E[E[Y(t) \vee \alpha | \mathcal{F}_t^X] - X(t) \vee \alpha] \leq \lim_{\substack{s \rightarrow t- \\ s \in \mathbb{Q}}} E[Y(t) \vee \alpha - X(s) \vee \alpha] = 0,$$

and hence, since $Y(t) = Y(t-)$ a.s. and $Y(t-)$ is \mathcal{F}_t^X -measurable,

$$(2.31) \quad X(t) \vee \alpha = E[Y(t) \vee \alpha | \mathcal{F}_t^X] = Y(t) \vee \alpha \quad \text{a.s.}$$

Since α is arbitrary, $P\{X(t) = Y(t)\} = 1$ for $t \notin \Gamma$.

To see that almost all sample paths of \tilde{X} have right and left limits at all $t \geq 0$ and are right continuous at all $t \notin \Gamma$, replace F in the construction of Y by $F \cup \Gamma$. Note that this replaces Ω_0 by $\tilde{\Omega}_0 \subset \Omega_0$, but that for $\omega \in \tilde{\Omega}_0$, $Y(\cdot, \omega)$ and $\tilde{X}(\cdot, \omega)$ do not change. Since for $\omega \in \tilde{\Omega}_0$

$$(2.32) \quad Y(t, \omega) = \lim_{\substack{s \rightarrow t+ \\ s \in F \cup \Gamma}} Y(s, \omega) = \lim_{\substack{s \rightarrow t+ \\ s \in F \cup \Gamma}} X(s, \omega), \quad t \geq 0,$$

it follows that

$$(2.33) \quad Y(t, \omega) = \lim_{s \rightarrow t+} \tilde{X}(s, \omega), \quad t \geq 0,$$

which gives both the existence of right limits and the right continuity of $\tilde{X}(\cdot, \omega)$ at $t \notin \Gamma$. The existence of left limits follows similarly. \square

2.10 Corollary Let Z be a random variable with $E[|Z|] < \infty$. Then for any filtration $\{\mathcal{F}_s\}$ and $t \geq 0$, $E[Z | \mathcal{F}_{s+}] \rightarrow E[Z | \mathcal{F}_{t+}]$ in L^1 as $s \rightarrow t+$.

Proof. Let $X(t) = E[Z | \mathcal{F}_t]$, $t \geq 0$. Then X is a martingale and by Proposition 2.9 we may assume X has right limits a.s. at each $t \geq 0$. Since $\{X(t)\}$ is uniformly integrable, $X(t+) \equiv \lim_{s \rightarrow t+} X(s)$ exists a.s. and in L^1 for all $t \geq 0$.

We need only check that $X(t+) = E[Z|\mathcal{F}_{t+}]$. Clearly $X(t+)$ is \mathcal{F}_{t+} -measurable and for $A \in \mathcal{F}_{t+}$,

$$(2.34) \quad \int_A X(t+) dP = \lim_{s \rightarrow t+} \int_A X(s) dP = \int_A Z dP,$$

hence $X(t+) = E[Z|\mathcal{F}_{t+}]$. \square

2.11 Corollary If $\{\mathcal{F}_t\}$ is a right continuous filtration and X is an $\{\mathcal{F}_t\}$ -martingale, then X has a right continuous modification.

2.12 Remark It follows from the construction of Y in (2.24) that almost all sample paths of a right continuous submartingale have left limits at all $t > 0$. \square

Proof. With reference to (2.24) and Corollary 2.10, for $t < T$,

$$(2.35) \quad \begin{aligned} Y(t) &= \lim_{\substack{s \rightarrow t+ \\ s \in F}} X(s) = \lim_{\substack{s \rightarrow t+ \\ s \in F}} E[X(T)|\mathcal{F}_s] \\ &= E[X(T)|\mathcal{F}_{t+}] = E[X(T)|\mathcal{F}_t] = X(t) \text{ a.s.,} \end{aligned}$$

so Y is the desired modification. \square

Essentially, Proposition 2.9 says that we may assume every submartingale has well-behaved sample paths, that is, if all that is prescribed about a submartingale is its finite-dimensional distributions, then we may as well assume that the sample paths have the properties given in the proposition. In fact, in virtually all cases of interest, $\Gamma = \emptyset$, so we can assume right continuity at all $t \geq 0$. We do just that in the remainder of this section. Extension of the results to the somewhat more general case is usually straightforward.

Our next result is the optional sampling theorem.

2.13 Theorem Let X be a right continuous $\{\mathcal{F}_t\}$ -submartingale, and let τ_1 and τ_2 be $\{\mathcal{F}_t\}$ -stopping times. Then for each $T > 0$,

$$(2.36) \quad E[X(\tau_2 \wedge T)|\mathcal{F}_{\tau_1}] \geq X(\tau_1 \wedge \tau_2 \wedge T).$$

If, in addition, τ_2 is finite a.s., $E[|X(\tau_2)|] < \infty$, and

$$(2.37) \quad \lim_{T \rightarrow \infty} E[|X(T)|\chi_{(\tau_2 > T)}] = 0,$$

then

$$(2.38) \quad E[X(\tau_2)|\mathcal{F}_{\tau_1}] \geq X(\tau_1 \wedge \tau_2).$$

2.14 Remark Note that if X is a martingale (X and $-X$ are submartingales), then equality holds in (2.36) and (2.38). Note also that any right continuous $\{\mathcal{F}_t\}$ -submartingale is an $\{\mathcal{F}_{t+}\}$ -submartingale, and hence corresponding inequalities hold for $\{\mathcal{F}_{t+}\}$ -stopping times. \square

62 STOCHASTIC PROCESSES AND MARTINGALES

Proof. For $i = 1, 2$, let $\tau_i^{(n)} = \infty$ if $\tau_i = \infty$ and let $\tau_i^{(n)} = (k+1)/2^n$ if $k/2^n \leq \tau_i < (k+1)/2^n$. Then by Proposition 1.3, $\tau_i^{(n)}$ is an $\{\mathcal{F}_t\}$ -stopping time, and by Lemma 2.2, for each $\alpha \in \mathbb{R}$ and $T > 0$,

$$(2.39) \quad E[X(\tau_2^{(n)} \wedge T) \vee \alpha | \mathcal{F}_{\tau_1^{(n)}}] \geq X(\tau_1^{(n)} \wedge \tau_2^{(n)} \wedge T) \vee \alpha.$$

Since $\mathcal{F}_{\tau_1} \subset \mathcal{F}_{\tau_1^{(n)}}$ by Proposition 1.4(c), (2.39) implies

$$(2.40) \quad E[X(\tau_2^{(n)} \wedge T) \vee \alpha | \mathcal{F}_{\tau_1}] \geq E[X(\tau_1^{(n)} \wedge \tau_2^{(n)} \wedge T) \vee \alpha | \mathcal{F}_{\tau_1}].$$

Since Lemma 2.2 implies

$$(2.41) \quad \alpha \leq X(\tau_2^{(n)} \wedge T) \vee \alpha \leq E[X(T) \vee \alpha | \mathcal{F}_{\tau_1^{(n)}}],$$

$\{X(\tau_2^{(n)} \wedge T) \vee \alpha\}$ is uniformly integrable as is $\{X(\tau_1^{(n)} \wedge \tau_2^{(n)} \wedge T) \vee \alpha\}$ (Problem 10). Letting $n \rightarrow \infty$, the right continuity of X and the uniform integrability of the sequences gives

$$(2.42) \quad \begin{aligned} E[X(\tau_2 \wedge T) \vee \alpha | \mathcal{F}_{\tau_1}] &\geq E[X(\tau_1 \wedge \tau_2 \wedge T) \vee \alpha | \mathcal{F}_{\tau_1}] \\ &= X(\tau_1 \wedge \tau_2 \wedge T) \vee \alpha. \end{aligned}$$

Letting $\alpha \rightarrow -\infty$ gives (2.36), and under the additional hypotheses, letting $T \rightarrow \infty$ gives (2.38). \square

The following is an application of the optional sampling theorem.

2.15 Proposition Let X be a right continuous nonnegative $\{\mathcal{F}_t\}$ -supermartingale, and let $\tau_c(0)$ be the first contact time with 0. Then $X(t) = 0$ for all $t \geq \tau_c(0)$ with probability one.

Proof. For $n = 1, 2, \dots$, let $\tau_n = \tau_c([0, n^{-1}])$, the first entrance time into $[0, n^{-1}]$. (By Proposition 1.5, τ_n is an $\{\mathcal{F}_{t+}\}$ -stopping time.) Then $\tau_c(0) = \lim_{n \rightarrow \infty} \tau_n$. If $\tau_n < \infty$, then $X(\tau_n) \leq n^{-1}$. Consequently, for every $t \geq 0$,

$$(2.43) \quad E[X(t) | \mathcal{F}_{\tau_n+}] \leq X(t \wedge \tau_n),$$

and hence

$$(2.44) \quad E[X(t) | \mathcal{F}_{\tau_n+}] \chi_{\{\tau_n \leq t\}} \leq n^{-1}.$$

Taking expectations and letting $n \rightarrow \infty$, we have

$$(2.45) \quad E[X(t) \chi_{\{\tau_c(0) \leq t\}}] = 0.$$

The proposition follows by the nonnegativity and right continuity. \square

Next we extend Lemma 2.3.

2.16 Proposition (a) Let X be a right continuous submartingale. Then for each $x > 0$ and $T > 0$,

$$(2.46) \quad P\left\{\sup_{t \leq T} X(t) \geq x\right\} \leq x^{-1} E[X^+(T)]$$

and

$$(2.47) \quad P\left\{\inf_{t \leq T} X(t) \leq -x\right\} \leq x^{-1}(E[X^+(T)] - E[X(0)]).$$

(b) Let X be a nonnegative right continuous submartingale. Then for $\alpha > 1$ and $T > 0$,

$$(2.48) \quad E\left[\sup_{t \leq T} X(t)^\alpha\right] \leq \left(\frac{\alpha}{\alpha-1}\right)^\alpha E[X(T)^\alpha].$$

Proof. Corollary 2.4 implies (2.46) and (2.47), but we need to extend (2.46) in order to obtain (2.48). Under the assumptions of part (b) let $x > 0$, and define $\tau = \inf\{t: X(t) > x\}$. Then τ is an $\{\mathcal{F}_{t+}\}$ -stopping time by Proposition 1.5(b), and the right continuity of X implies $X(\tau) \geq x$ if $\tau < \infty$. Consequently for $T > 0$,

$$(2.49) \quad \left\{\sup_{t \leq T} X(t) > x\right\} \subset \{\tau \leq T\} \subset \left\{\sup_{t \leq T} X(t) \geq x\right\},$$

and the three events have equal probability for all but countably many $x > 0$. By Theorem 2.13,

$$(2.50) \quad E[X(\tau \wedge T)] \leq E[X(T)],$$

and hence

$$(2.51) \quad xP\{\tau \leq T\} \leq E[X(\tau)\chi_{\{\tau \leq T\}}] \leq E[X(T)\chi_{\{\tau \leq T\}}].$$

Let φ be absolutely continuous on bounded intervals of $[0, \infty)$ with $\varphi' \geq 0$ and $\varphi(0) = 0$. Define $Z = \sup_{t \leq T} X(t)$. Then for $\beta > 0$,

$$\begin{aligned} (2.52) \quad E[\varphi(Z \wedge \beta)] &= \int_0^\beta \varphi'(x)P\{Z > x\} dx \\ &\leq \int_0^\beta \varphi'(x)x^{-1}E[X(T)\chi_{\{Z \geq x\}}] dx \\ &= E[X(T)\psi(Z \wedge \beta)] \end{aligned}$$

where $\psi(z) = \int_0^z \varphi'(x)x^{-1} dx$.

64 STOCHASTIC PROCESSES AND MARTINGALES

If $\varphi(x) \equiv x^\alpha$ for some $\alpha > 1$, then

$$(2.53) \quad \begin{aligned} E[(Z \wedge \beta)^\alpha] &\leq \frac{\alpha}{\alpha - 1} E[X(T)(Z \wedge \beta)^{\alpha-1}] \\ &\leq \frac{\alpha}{\alpha - 1} E[X(T)^\alpha]^{1/\alpha} E[(Z \wedge \beta)^\alpha]^{(\alpha-1)/\alpha}, \end{aligned}$$

and hence

$$(2.54) \quad E[(Z \wedge \beta)^\alpha]^{1/\alpha} \leq \frac{\alpha}{\alpha - 1} E[X(T)^\alpha]^{1/\alpha}.$$

Letting $\beta \rightarrow \infty$ gives (2.48). \square

2.17 Corollary Let X be a right continuous martingale. Then for $x > 0$ and $T > 0$,

$$(2.55) \quad P\left\{\sup_{t \leq T} |X(t)| \geq x\right\} \leq x^{-1} E[|X(T)|],$$

and for $\alpha > 1$ and $T > 0$,

$$(2.56) \quad E\left[\sup_{t \leq T} |X(t)|^\alpha\right] \leq \left(\frac{\alpha}{\alpha - 1}\right)^\alpha E[|X(T)|^\alpha].$$

Proof. Since $|X|$ is a submartingale by Proposition 2.1, (2.55) and (2.56) follow directly from (2.46) and (2.48). \square

3. LOCAL MARTINGALES

A real-valued process X is an $\{\mathcal{F}_t\}$ -local martingale if there exist $\{\mathcal{F}_t\}$ -stopping times $\tau_1 \leq \tau_2 \leq \dots$ with $\tau_n \rightarrow \infty$ a.s. such that $X^{\tau_n} \equiv X(\cdot \wedge \tau_n)$ is an $\{\mathcal{F}_t\}$ -martingale. Local submartingales and local supermartingales are defined similarly. Of course a martingale is a local martingale. In studying stochastic integrals (Chapter 5) and random time changes (Chapter 6), one is led naturally to local martingales that are not martingales.

3.1 Proposition If X is a right continuous $\{\mathcal{F}_t\}$ -local martingale and τ is an $\{\mathcal{F}_t\}$ -stopping time, then $X^\tau \equiv X(\cdot \wedge \tau)$ is an $\{\mathcal{F}_t\}$ -local martingale.

Proof. There exist $\{\mathcal{F}_t\}$ -stopping times $\tau_1 \leq \tau_2 \leq \dots$ such that $\tau_n \rightarrow \infty$ a.s. and X^{τ_n} is an $\{\mathcal{F}_t\}$ -martingale. But then $X^\tau(\cdot \wedge \tau_n) = X^{\tau_n}(\cdot \wedge \tau)$ is an $\{\mathcal{F}_t\}$ -martingale, and hence X^τ is an $\{\mathcal{F}_t\}$ -local martingale. \square

In the next result the stochastic integral is just a Stieltjes integral and consequently needs no special definition. As before, when we say a process V is continuous and locally of bounded variation, we mean that for all $\omega \in \Omega$, $V(\cdot, \omega)$ is continuous and of bounded variation on bounded intervals.

3.2 Proposition Suppose X is a right continuous $\{\mathcal{F}_t\}$ -local martingale, and V is real-valued, continuous, locally of bounded variation, and $\{\mathcal{F}_t\}$ -adapted. Then

$$(3.1) \quad M(t) \equiv \int_0^t V(s) dX(s) = V(t)X(t) - V(0)X(0) - \int_0^t X(s) dV(s)$$

is an $\{\mathcal{F}_t\}$ -local martingale.

Proof. The last equality in (3.1) is just integration by parts. There exist $\{\mathcal{F}_t\}$ -stopping times $\tau_1 \leq \tau_2 \leq \dots$ such that $\tau_n \rightarrow \infty$ a.s. and X^{τ_n} is an $\{\mathcal{F}_t\}$ -martingale. Without loss of generality we may assume $\tau_n \leq \tau_c((-\infty, -n] \cup [n, \infty))$, the first contact time of $(-\infty, -n] \cup [n, \infty)$ by X . (If not, replace τ_n by the minimum of the two stopping times.) Let R be the total variation process

$$(3.2) \quad R(t) = \sup \sum_{i=0}^{m-1} |V(s_{i+1}) - V(s_i)|,$$

where the supremum is over partitions of $[0, t]$, $0 = s_0 < s_1 < \dots < s_m = t$. For $n = 1, 2, \dots$ let $\gamma_n = \inf\{t: R(t) \geq n\}$. Since R is continuous, γ_n is the first contact time of $[n, \infty)$ and is an $\{\mathcal{F}_t\}$ -stopping time by Proposition 1.5. The continuity of R also implies $\gamma_n \rightarrow \infty$ a.s.

Let $\sigma_n = \gamma_n \wedge \tau_n$. Then $\sigma_n \rightarrow \infty$ a.s. and we claim $M(\cdot \wedge \sigma_n)$ is an $\{\mathcal{F}_t\}$ -martingale. To verify this we must show

$$(3.3) \quad E \left[\int_{t \wedge \sigma_n}^{(t+s) \wedge \sigma_n} V(u) dX(u) \middle| \mathcal{F}_t \right] = E \left[\int_t^{t+s} V^{\sigma_n}(u) dX^{\sigma_n}(u) \middle| \mathcal{F}_t \right] = 0,$$

for all $t, s \geq 0$.

Let $t = u_0 < u_1 < \dots < u_m = t + s$. Then

$$(3.4) \quad E \left[\sum_{k=0}^{m-1} V^{\sigma_n}(u_k) (X^{\sigma_n}(u_{k+1}) - X^{\sigma_n}(u_k)) \middle| \mathcal{F}_t \right] = 0,$$

since X^{σ_n} is an $\{\mathcal{F}_t\}$ -martingale and $V^{\sigma_n}(u_k)$ is \mathcal{F}_{u_k} -measurable. Letting $\max_k |u_{k+1} - u_k| \rightarrow 0$, the sum in (3.4) converges to the second integral in (3.3)

66 STOCHASTIC PROCESSES AND MARTINGALES

a.s. However, to obtain (3.3), we must show that the convergence is in L^1 . Observe

$$\begin{aligned}
 (3.5) \quad & \left| \sum_{k=0}^{m-1} V^{\sigma_n}(u_k)(X^{\sigma_n}(u_{k+1}) - X^{\sigma_n}(u_k)) \right| \\
 & = \left| V^{\sigma_n}(t+s)X^{\sigma_n}(t+s) - V^{\sigma_n}(0)X^{\sigma_n}(0) \right. \\
 & \quad \left. - \sum_{k=0}^{m-1} X^{\sigma_n}(u_{k+1})(V^{\sigma_n}(u_{k+1}) - V^{\sigma_n}(u_k)) \right| \\
 & \leq |V^{\sigma_n}(t+s)X^{\sigma_n}(t+s) - V^{\sigma_n}(0)X^{\sigma_n}(0)| \\
 & \quad + \sum_{k=0}^{m-1} |X^{\sigma_n}(u_{k+1})| |V^{\sigma_n}(u_{k+1}) - V^{\sigma_n}(u_k)| \\
 & \leq |V^{\sigma_n}(t+s)X^{\sigma_n}(t+s) - V^{\sigma_n}(0)X^{\sigma_n}(0)| + (n \vee |X^{\sigma_n}(t+s)|)R(\sigma_n).
 \end{aligned}$$

The right side is in L^1 , so the desired convergence follows by the dominated convergence theorem. \square

3.3 Corollary Let X and Y be real-valued, right continuous, $\{\mathcal{F}_t\}$ -adapted processes. Suppose that for each t , $\inf_{s \leq t} X(s) > 0$. Then

$$(3.6) \quad M_1(t) \equiv X(t) - \int_0^t Y(s) \, ds$$

is an $\{\mathcal{F}_t\}$ -local martingale if and only if

$$(3.7) \quad M_2(t) \equiv X(t) \exp \left\{ - \int_0^t \frac{Y(s)}{X(s)} \, ds \right\}$$

is an $\{\mathcal{F}_t\}$ -local martingale.

Proof. Suppose M_1 is an $\{\mathcal{F}_t\}$ -local martingale. Then by Proposition 3.2,

$$\begin{aligned}
 (3.8) \quad & \int_0^t \exp \left\{ - \int_0^s \frac{Y(u)}{X(u)} \, du \right\} dM_1(s) \\
 & = \int_0^t \exp \left\{ - \int_0^s \frac{Y(u)}{X(u)} \, du \right\} dX(s) \\
 & \quad - \int_0^t \exp \left\{ - \int_0^s \frac{Y(u)}{X(u)} \, du \right\} Y(s) \, ds \\
 & = X(t) \exp \left\{ - \int_0^t \frac{Y(u)}{X(u)} \, du \right\} - X(0)
 \end{aligned}$$

is an $\{\mathcal{F}_t\}$ -local martingale. Conversely, if M_2 is an $\{\mathcal{F}_t\}$ -local martingale, then

$$(3.9) \quad \begin{aligned} & \int_0^t \exp \left\{ \int_0^s \frac{Y(u)}{X(u)} du \right\} dM_2(s) \\ &= X(t) - X(0) - \int_0^t Y(s) ds \end{aligned}$$

is an $\{\mathcal{F}_t\}$ -local martingale. \square

We close this section with a result concerning the quadratic variation of the sample paths of a local martingale. By an "increasing" process, we mean a process whose sample paths are nondecreasing.

3.4 Proposition Let X be a right continuous $\{\mathcal{F}_t\}$ -local martingale. Then there exists a right continuous increasing process, denoted by $[X]$, such that for each $t \geq 0$ and each sequence of partitions $\{u_k^{(n)}\}$ of $[0, t]$ with $\max_k (u_{k+1}^{(n)} - u_k^{(n)}) \rightarrow 0$,

$$(3.10) \quad \sum_k (X(u_{k+1}^{(n)}) - X(u_k^{(n)}))^2 \xrightarrow{P} [X](t)$$

as $n \rightarrow \infty$. If, in addition, X is a martingale and $E[X(t)^2] < \infty$ for all $t \geq 0$, then the convergence in (3.10) is in L^1 .

Proof. Convergence in probability is metrizable (Problem 8); consequently we want to show that $\{\sum_k (X(u_{k+1}^{(n)}) - X(u_k^{(n)}))^2\}$ is a Cauchy sequence. If this were not the case, then there would exist $\varepsilon > 0$ and $\{n_i\}$ and $\{m_i\}$ such that $n_i \rightarrow \infty$, $m_i \rightarrow \infty$, and

$$(3.11) \quad P \left\{ \left| \sum_k (X(u_{k+1}^{(n)}) - X(u_k^{(n)}))^2 - \sum_k (X(u_{k+1}^{(m)}) - X(u_k^{(m)}))^2 \right| \geq \varepsilon \right\} \geq \varepsilon$$

for all i .

Since any pair of partitions of $[0, t]$ has a common refinement, that is, there exists $\{v_k\}$ such that $\{u_k^{(n)}\} \subset \{v_k\}$ and $\{u_k^{(m)}\} \subset \{v_k\}$, the following lemma contradicts (3.11) and hence proves that the left side of (3.10) converges in probability.

3.5 Lemma Let X be a right continuous $\{\mathcal{F}_t\}$ -local martingale. Fix $T > 0$. For $n = 1, 2, \dots$, let $\{u_k^{(n)}\}$ and $\{v_k^{(n)}\}$ be partitions of $[0, T]$ with $\{v_k^{(n)}\} \subset \{u_k^{(n)}\}$ and $\max_k (v_{k+1}^{(n)} - v_k^{(n)}) \rightarrow 0$. Then

$$(3.12) \quad \sum_k (X(v_{k+1}^{(n)}) - X(v_k^{(n)}))^2 - \sum_k (X(u_{k+1}^{(n)}) - X(u_k^{(n)}))^2 \xrightarrow{P} 0.$$

68 STOCHASTIC PROCESSES AND MARTINGALES

Proof. Without loss of generality we can assume X is a martingale (otherwise consider the stopped process X^*), and $X(0) = 0$. Fix $M > 0$, and let $\tau = \inf\{s: |X(s)| \geq M \text{ or } |X(s-)| \geq M\}$. Note that $P\{\tau \leq t\} \leq E[|X(t)|]/M$ by Corollary 2.17.

Let $\{u_k\}$ and $\{v_k\}$ be partitions of $[0, t]$, and suppose $\{v_k\} \subset \{u_k\}$. Let $w_k = \max\{v_i: v_i \leq u_k\}$, and define $\eta_k \equiv X^*(v_k) - X^*(u_{k-1})$ and

$$(3.13) \quad Z = \sum (X^*(v_i) - X^*(v_{i-1}))^2 - \sum (X^*(u_k) - X^*(u_{k-1}))^2 \\ = \sum \zeta_k,$$

where $\zeta_k = 2(X^*(u_k) - X^*(u_{k-1}))(X^*(u_{k-1}) - X^*(w_{k-1}))$. Note that either $\zeta_k = 0$ or $|\zeta_k| \leq 4M|X^*(u_k) - X^*(u_{k-1})|$ and that $E[\zeta_{k+1} | \mathcal{F}_{u_k}] = 0$. Consequently,

$$(3.14) \quad Z_m = \sum_{k=1}^m \zeta_k$$

is a discrete-parameter martingale.

Let

$$(3.15) \quad \varphi(x) = \begin{cases} x^2, & |x| \leq 4M^2, \\ 8M^2|x| - 16M^4, & |x| > 4M^2. \end{cases}$$

Let α be an $\{\mathcal{F}_t\}$ -stopping time with values in $\{u_k\}$ and let β be \mathcal{F}_α -measurable with values in $\{u_k\}$ and $\beta \geq \alpha$. Let k_α and k_β satisfy $\alpha = u_{k_\alpha}$ and $\beta = u_{k_\beta}$ and let $K = \max\{k: u_k < \tau\}$. Then

$$(3.16) \quad E[\varphi(X^*(\beta)) | \mathcal{F}_\alpha] - \varphi(X^*(\alpha)) \\ = \sum_{k=k_\alpha+1}^{k_\beta} E[\varphi(\eta_k + X^*(u_{k-1})) - \varphi(X^*(u_{k-1})) | \mathcal{F}_\alpha] \\ = \sum_{k=k_\alpha+1}^{k_\beta} E[\varphi(\eta_k + X^*(u_{k-1})) - \varphi(X^*(u_{k-1})) - \eta_k \varphi'(X^*(u_{k-1})) | \mathcal{F}_\alpha] \\ \geq E\left[\sum_{k=k_\alpha+1}^{k_\beta \wedge K} \eta_k^2 \middle| \mathcal{F}_\alpha\right],$$

where the last inequality follows from the convexity of φ and the fact that for $k \leq K$, $|X^*(u_k)| \leq M$.

Using the fact that $\{Z_m\}$ is a discrete-parameter martingale,

$$(3.17) \quad E[\varphi(Z)] = \sum E[\varphi(\xi_k + Z_{k-1}) - \varphi(Z_{k-1}) - \xi_k \varphi'(Z_{k-1})] \\ = \sum E\left[\int_0^{\xi_k} \int_0^y \varphi''(z + Z_{k-1}) dz dy\right] \\ \leq \sum 2E\left[\int_0^{\xi_k} \int_0^{y/2} \varphi''(z) dz dy\right]$$

$$\begin{aligned}
&= \sum 4E\left[\varphi\left(\frac{\xi_k}{2}\right)\right] \\
&= 4E\left[\varphi\left(\frac{\xi_{K+1}}{2}\right)\right] \\
&\quad + 4E\left[\sum_{k=1}^K \eta_k^2 (X'(u_{k-1}) - X'(w_{k-1}))^2\right].
\end{aligned}$$

Fix $\varepsilon > 0$. Let $\alpha_i = \min\{u_k : |X'(u_k) - X'(v_i)| \geq \varepsilon\} \cup \{v_{i+1}\}$ and $\beta_i = v_{i+1}$. Note that if $v_i = w_{k-1} \leq u_{k-1} < v_{i+1}$ and $\alpha_i > v_{i+1}$, then $(X'(u_{k-1}) - X'(w_{k-1}))^2 \leq \varepsilon^2$. Consequently, by (3.16) and (3.17),

$$\begin{aligned}
(3.18) \quad E[\varphi(Z)] &\leq 4E\left[\varphi\left(\frac{\xi_{K+1}}{2}\right)\right] + 4\varepsilon^2 E\left[\sum_{k=1}^K \eta_k^2\right] \\
&\quad + \sum_i E[\chi_{\{\alpha_i < v_{i+1}\}} 16M^2 (\varphi(X'(\beta_i)) - \varphi(X'(\alpha_i)))].
\end{aligned}$$

Fix $N \geq 1$, and let $L = \min\{l : \sum_{i=0}^l \chi_{\{\alpha_i < v_{i+1}\}} = N\}$. Let $\gamma = \alpha_L$ if $L < \infty$ and $\gamma = T$ otherwise. Then γ is an $\{\mathcal{F}_t\}$ -stopping time, and hence by (3.18) and the convexity of φ ,

$$\begin{aligned}
(3.19) \quad E[\varphi(Z)] &\leq 4E\left[\varphi\left(\frac{\xi_{K+1}}{2}\right)\right] + 4\varepsilon^2 E[\varphi(X'(T))] \\
&\quad + 16M^2 E[\varphi(X'(T)) - \varphi(X'(\gamma))] \\
&\quad + 16M^2 E\left[\sum_{i \leq L} \chi_{\{\alpha_i < v_{i+1}\}} (\varphi(X'(\beta_i)) - \varphi(X'(\alpha_i)))\right].
\end{aligned}$$

Given $\varepsilon, \varepsilon' > 0$, let $D = \{s \in [0, T] : |X(s) - X(s-)| > \varepsilon/2\}$. Then there exists a positive random variable δ such that $s \in D$ and $s \leq t \leq s + \delta$ imply $|X(t) - X(s)| \leq \varepsilon'$, and $0 \leq s < t \leq T$, $t - s \leq \delta$, and $|X(t) - X(s)| \geq \varepsilon$ imply $(s, t] \cap D \neq \emptyset$. Let $|D|$ denote the cardinality of D . On $\{\max(v_{i+1} - v_i) \leq \delta\}$,

$$\begin{aligned}
(3.20) \quad &\sum_{i \leq L} \chi_{\{\alpha_i < v_{i+1}\}} (\varphi(X'(\beta_i)) - \varphi(X'(\alpha_i))) \\
&\leq (|D| \wedge N) 8M^2 \varepsilon'.
\end{aligned}$$

Let $S(T)$ be the collection of $\{\mathcal{F}_t\}$ -stopping times α with $\alpha \leq T$. Since

$$(3.21) \quad \varphi(X'(\alpha)) \leq E[\varphi(X'(T)) | \mathcal{F}_\alpha]$$

for all $\alpha \in S(T)$, $\{\varphi(X'(\alpha)) : \alpha \in S(T)\}$ is uniformly integrable. Consequently, the right side of (3.19) can be made arbitrarily small by taking ε small, N large (so that $P\{N < |D|\}$ is small), ε' small, and $\max(v_{i+1} - v_i)$ small. Note that if $N > |D|$ and $\max(v_{i+1} - v_i) < \delta$, then $\gamma = T$.

70 STOCHASTIC PROCESSES AND MARTINGALES

Thus, if $Z^{(n)}$ is defined for $\{v_k^{(n)}\}$ and $\{u_k^{(n)}\}$ as in (3.13), the estimate in (3.19) implies

$$(3.22) \quad \lim_{n \rightarrow \infty} E[\varphi(Z^{(n)})] = 0,$$

which, since M is arbitrary, implies (3.12). \square

Proof of Proposition 3.4—continued. Assume X is a martingale and $E[X(T)^2] < \infty$. Let $\{u_k\}$ be a partition of $[0, T]$, and let X' be as in the proof of Lemma 3.5. Then

$$(3.23) \quad E[|\sum(X(u_{k+1}) - X(u_k))^2 - \sum(X'(u_{k+1}) - X'(u_k))^2|] \\ \leq E[(X(T) - X(T \wedge \tau))^2] + E[|(X(u_{K+1}) - X(\tau))(X(\tau) - X(u_K))| \chi_{\{t < \tau\}}],$$

where $K = \max\{k: u_k < \tau\}$. Since for $M \geq 1$, φ defined by (3.15) satisfies $|\varphi(x)| \leq \varepsilon + \varphi(x)/\varepsilon$ for every $\varepsilon > 0$, the estimates in the proof of Lemma 3.5 imply $\{\sum(X'(u_{k+1}^{(n)}) - X'(u_k^{(n)}))^2\}$ is a Cauchy sequence in L^1 (note that we need $E[X(T)^2] < \infty$ in order that this sequence be in L^1). Consequently, since the right side of (3.23) can be made small by taking M large and $\max(u_{k+1} - u_k)$ small, it follows that $\{\sum(X(u_{k+1}^{(n)}) - X(u_k^{(n)}))^2\}$ is a Cauchy sequence in L^1 .

Convergence of the left side of (3.10) determines $[X](t)$ a.s. for each $t \geq 0$. We must show that $[X]$ has a right continuous modification. Since $\{Z_m\}$ given by (3.14) is a discrete-parameter martingale, Proposition 2.16 gives

$$(3.24) \quad P\left\{\sup_m |Z_m| > \varepsilon\right\} \leq \frac{E[\varphi(Z)]}{\varphi(\varepsilon)}.$$

Consequently, for $l \geq n \rightarrow \infty$ and $T > 0$,

$$(3.25) \quad \sup_{k \leq 2^n T} \left| \sum_{i=1}^k \left(X\left(\frac{i}{2^n}\right) - X\left(\frac{i-1}{2^n}\right) \right)^2 \right. \\ \left. - \sum_{j=1}^{k2^{l-n}} \left(X\left(\frac{j}{2^l}\right) - X\left(\frac{j-1}{2^l}\right) \right)^2 \right| \xrightarrow{P} 0,$$

and it follows that we can define $[X]$ on the dyadic rationals so that it is nondecreasing and satisfies

$$(3.26) \quad \sup_{k \leq 2^n T} \left| \sum_{i=1}^k \left(X\left(\frac{i}{2^n}\right) - X\left(\frac{i-1}{2^n}\right) \right)^2 - [X]\left(\frac{k}{2^n}\right) \right| \xrightarrow{P} 0.$$

The right continuity of $[X]$ on the dyadic rationals follows from the right continuity of X . For arbitrary $t \geq 0$, define

$$(3.27) \quad [X](t) = \lim_{n \rightarrow \infty} [X]\left(\frac{\lfloor 2^n t \rfloor + 1}{2^n}\right).$$

Clearly this definition makes $[X]$ right continuous. We must verify that (3.10) is satisfied.

Let $\{u_k^n\} = \{i/2^n : 0 \leq i \leq [2^n t]\} \cup \{t\}$. Then

$$\begin{aligned}
 (3.28) \quad & \left| \sum_k (X(u_{k+1}^n) - X(u_k^n))^2 - [X](t) \right| \\
 & \leq \left| [X]\left(\frac{[2^n t] + 1}{2^n}\right) - [X](t) \right| \\
 & \quad + \left| \left(X(t) - X\left(\frac{[2^n t]}{2^n}\right) \right)^2 - \left(X\left(\frac{[2^n t] + 1}{2^n}\right) - X\left(\frac{[2^n t]}{2^n}\right) \right)^2 \right| \\
 & \quad + \left| \sum_{i=1}^{[2^n t]+1} \left(X\left(\frac{i}{2^n}\right) - X\left(\frac{i-1}{2^n}\right) \right)^2 - [X]\left(\frac{[2^n t] + 1}{2^n}\right) \right| \\
 & \xrightarrow{P} 0,
 \end{aligned}$$

and (3.10) follows. \square

3.6 Proposition Let X be a continuous $\{\mathcal{F}_t\}$ -local martingale. Then $[X]$ can be taken to be continuous.

Proof. Let $[X]$ be as in Proposition 3.4. Almost sure continuity of $[X]$ restricted to the dyadic rationals follows from (3.26) and the continuity of X . Since $[X]$ is nondecreasing, it must therefore be almost surely continuous. \square

4. THE PROJECTION THEOREM

Recall that an E -valued process X is $\{\mathcal{F}_t\}$ -progressive if the restriction of X to $[0, t] \times \Omega$ is $\mathcal{B}[0, t] \times \mathcal{F}_t$ -measurable for each $t \geq 0$, that is, if

$$(4.1) \quad \{(s, \omega) : X(s, \omega) \in \Gamma\} \cap ([0, t] \times \Omega) \in \mathcal{B}[0, t] \times \mathcal{F}_t$$

for each $t \geq 0$ and $\Gamma \in \mathcal{B}(E)$. Alternatively, we define the σ -algebra of $\{\mathcal{F}_t\}$ -progressive sets \mathcal{W} by

$$\begin{aligned}
 (4.2) \quad \mathcal{W} = \{A \in \mathcal{B}[0, \infty) \times \mathcal{F} : A \cap ([0, t] \times \Omega) \in \mathcal{B}[0, t] \times \mathcal{F}_t, \\
 \text{for all } t \geq 0\}.
 \end{aligned}$$

(The proof that \mathcal{W} is a σ -algebra is the same as for \mathcal{F}_t in Proposition 1.4.) Then (4.1) is just the requirement that X is a \mathcal{W} -measurable function on $[0, \infty) \times \Omega$.

The σ -algebra of $\{\mathcal{F}_t\}$ -optional sets \mathcal{O} is the σ -algebra of subsets of $[0, \infty) \times \Omega$ generated by the real-valued, right continuous $\{\mathcal{F}_t\}$ -adapted processes. An E -valued process X is $\{\mathcal{F}_t\}$ -optional if it is an \mathcal{O} -measurable function on $[0, \infty) \times \Omega$. Since every right continuous $\{\mathcal{F}_t\}$ -adapted process is $\{\mathcal{F}_t\}$ -progressive, $\mathcal{O} \subset \mathcal{W}$, and every $\{\mathcal{F}_t\}$ -optional process is $\{\mathcal{F}_t\}$ -progressive.

72 STOCHASTIC PROCESSES AND MARTINGALES

Throughout the remainder of this section we fix $\{\mathcal{F}_t\}$ and simply say adapted, optional, progressive, and so on to mean $\{\mathcal{F}_t\}$ -adapted, and so on. In addition we assume that $\{\mathcal{F}_t\}$ is complete.

4.1 Lemma Every martingale has an optional modification.

Proof. By Proposition 2.9, every martingale has a modification X whose sample paths have right and left limits at every $t \in [0, \infty)$ and are right continuous at every t except possibly for t in a countable, deterministic set Γ . We show that X is optional. Since we are assuming $\{\mathcal{F}_t\}$ is complete, X is adapted. First define

$$(4.3) \quad Y_n(t) = X\left(\frac{k-1}{n}\right), \quad \frac{k}{n} \leq t < \frac{k+1}{n},$$

(set $X(-1/n) = X(0)$), and note that $\lim_{n \rightarrow \infty} Y_n(t) \equiv Y(t) = X(t-)$. Since Y_n is adapted and right continuous, Y is optional.

Fix $\varepsilon > 0$. Define $\tau_0 = 0$ and, for $n = 0, 1, 2, \dots$,

$$(4.4) \quad \tau_{n+1} = \inf \{s > \tau_n : |X(s) - X(s-)| > \varepsilon \text{ or } |X(s+) - X(s-)| > \varepsilon \text{ or } |X(s+) - X(s)| > \varepsilon\}.$$

Since $X(s+) = X(s)$ except for $s \in \Gamma$,

$$(4.5) \quad \{\tau_n < t\} = \bigcup_i \bigcap_m \bigcup_{(s_i, t_i)} \bigcap_{i=1}^n \left\{ |X(t_i) - X(s_i)| > \varepsilon + \frac{1}{i} \right\} \in \mathcal{F}_t,$$

where $\{s_i, t_i\}$ ranges over all sets of the form $0 \leq s_1 < t_1 < s_2 < t_2 < \dots < s_n < t_n < t$, $|t_i - s_i| < 1/m$, and $t_i, s_i \in \Gamma \cup \mathbb{Q}$. Define

$$(4.6) \quad \begin{aligned} Z_m^\varepsilon(t) &= \sum_{n=1}^{\infty} \chi_{[\tau_n, \tau_n + 1/m]}(t) \chi_{\{|X(\tau_n) - X(\tau_n-)| > \varepsilon\}} (X(\tau_n) - X(\tau_n-)) \\ &= \sum_{n=1}^{\infty} \chi_{\{\tau_n < t\}} \chi_{[\tau_n, \tau_n + 1/m]}(t) \chi_{\{|X(\tau_n) - X(\tau_n-)| > \varepsilon\}} (X(\tau_n) - X(\tau_n-)) \\ &\quad + \chi_{\{|X(t) - X(t-)| > \varepsilon\}} (X(t) - X(t-)). \end{aligned}$$

Since X has right and left limits at each $t \in [0, \infty)$, $\lim_{n \rightarrow \infty} \tau_n = \infty$, and hence Z_m^ε is right continuous and has left limits. By (4.5), $\{\tau_n < s\} \in \mathcal{F}_s$ for $s \leq t$, and an examination of the right side of (4.6) shows that $Z_m^\varepsilon(t)$ is \mathcal{F}_t -measurable. Therefore Z_m^ε is optional. Finally observe that $|Y(t) + \lim_{m \rightarrow \infty} Z_m^\varepsilon(t) - X(t)| \leq \varepsilon$, and since ε is arbitrary, X is optional. \square

4.2 Theorem Let X be a nonnegative real-valued, measurable process. Then there exists a $[0, \infty]$ -valued optional process Y such that

$$(4.7) \quad E[X(\tau) | \mathcal{F}_\tau] = Y(\tau)$$

for all stopping times τ with $P\{\tau < \infty\} = 1$. (Note that we allow both sides of (4.7) to be infinite.)

4.3 Remark Y is called the *optional projection* of X . This theorem implies a partial converse to the observation that an optional process is progressive. Every real-valued, progressive process has an optional modification. The optional process Y is unique in the sense that, if Y_1 and Y_2 are optional processes satisfying (4.7), then Y_1 and Y_2 are indistinguishable. (See Dellacherie and Meyer (1982), page 103.) \square

Proof. Let $A \in \mathcal{F}$ and $B \in \mathcal{B}[0, \infty)$, and let Z be an optional process satisfying $E[\chi_A | \mathcal{F}_t] = Z(t)$. Z exists, since $E[\chi_A | \mathcal{F}_t]$ is a martingale. The optional sampling theorem implies $E[\chi_B | \mathcal{F}_\tau] = Z(\tau)$. Consequently, $\chi_B(\tau)Z(\tau)$ is optional, and

$$(4.8) \quad E[\chi_B(\tau)\chi_A | \mathcal{F}_\tau] = \chi_B(\tau)Z(\tau).$$

Therefore the collection M of bounded nonnegative measurable processes X for which there exists an optional Y satisfying (4.7) contains processes of the form $\chi_B\chi_A$, $B \in \mathcal{B}[0, \infty)$, $A \in \mathcal{F}$. Since M is closed under nondecreasing limits, and $X_1, X_2 \in M$, $X_1 \geq X_2$ implies $X_1 - X_2 \in M$, the Dynkin class theorem implies M contains all indicators of sets in $\mathcal{B}[0, \infty) \times \mathcal{F}$, and hence all bounded nonnegative measurable processes. The general case is proved by approximating X by $X \wedge n$, $n = 1, 2, \dots$. \square

4.4 Corollary Let X be a nonnegative real-valued, measurable process. Then there exists $Y: [0, \infty) \times [0, \infty) \times \Omega \rightarrow [0, \infty]$, measurable with respect to $\mathcal{B}[0, \infty) \times \mathcal{C}$, such that

$$(4.9) \quad E[X(\tau + s) | \mathcal{F}_\tau] = Y(s, \tau)$$

for all a.s. finite stopping times τ and all $s \geq 0$.

Proof. Replace $\chi_B(t)$ by $\chi_B(t + s)$ in the proof above. \square

4.5 Corollary Let $X: E \times [0, \infty) \times \Omega \rightarrow [0, \infty)$ be $\mathcal{B}(E) \times \mathcal{B}[0, \infty) \times \mathcal{F}$ -measurable. Then there exists $Y: E \times [0, \infty) \times \Omega \rightarrow [0, \infty]$, measurable with respect to $\mathcal{B}(E) \times \mathcal{C}$, such that

$$(4.10) \quad E[X(x, \tau) | \mathcal{F}_\tau] = Y(x, \tau)$$

for all a.s. finite stopping times τ and all $x \in E$.

Proof. Replace $\chi_B(t)$ by $\chi_B(x, t)$, $B \in \mathcal{B}(E) \times \mathcal{B}[0, \infty)$, in the proof of Theorem 4.2. \square

The argument used in the proof of Theorem 4.2 also gives us a Fubini theorem for conditional expectations.

74 STOCHASTIC PROCESSES AND MARTINGALES

4.6 Proposition Let $X: E \times \Omega \rightarrow \mathbb{R}$ be $\mathcal{B}(E) \times \mathcal{F}$ -measurable, and let μ be a σ -finite measure on $\mathcal{B}(E)$. Suppose $\int E[|X(x)|]\mu(dx) < \infty$. Then for every σ -algebra $\mathcal{D} \subset \mathcal{F}$, there exists $Y: E \times \Omega \rightarrow \mathbb{R}$ such that Y is $\mathcal{B}(E) \times \mathcal{D}$ -measurable, $Y(x) = E[X(x)|\mathcal{D}]$ for all $x \in E$, $\int |Y(x)|\mu(dx) < \infty$ a.s., and

$$(4.11) \quad \int Y(x)\mu(dx) = E\left[\int X(x)\mu(dx) \middle| \mathcal{D}\right].$$

4.7 Remark With this proposition in mind, we do not hesitate to write

$$(4.12) \quad \int E[X(x)|\mathcal{D}]\mu(dx) = E\left[\int X(x)\mu(dx) \middle| \mathcal{D}\right]. \quad \square$$

Proof. First assume μ is finite, verify the result for $X = \chi_B \chi_A$, $B \in \mathcal{B}(E)$, $A \in \mathcal{F}$, and then apply the Dynkin class theorem. The σ -finite case follows by writing μ as a sum of finite measures. \square

5. THE DOOB-MEYER DECOMPOSITION

Let S denote the collection of all $\{\mathcal{F}_t\}$ -stopping times. A right continuous $\{\mathcal{F}_t\}$ -submartingale is of *class DL* if for each $T > 0$, $\{X(\tau \wedge T): \tau \in S\}$ is uniformly integrable. If X is an $\{\mathcal{F}_t\}$ -martingale or if X is bounded below, then X is of class *DL* (Problem 10).

A process A is *increasing* if $A(\cdot, \omega)$ is nondecreasing for all $\omega \in \Omega$. Every right continuous nondecreasing function a on $[0, \infty)$ with $a(0) = 0$ determines a Borel measure μ_a on $[0, \infty)$ by $\mu_a[0, t] = a(t)$. We define

$$(5.1) \quad \int_0^t f(s) da(s) = \int_{[0, t]} f(s) \mu_a(ds)$$

when the integral on the right exists. Note that this is not a Stieltjes integral if f and a have common discontinuities.

5.1 Theorem Let $\{\mathcal{F}_t\}$ be complete and right continuous, and let X be a right continuous $\{\mathcal{F}_t\}$ -submartingale of class *DL*. Then there exists a unique (up to indistinguishability) right continuous $\{\mathcal{F}_t\}$ -adapted increasing process A with $A(0) = 0$ and the following properties:

- (a) $M = X - A$ is an $\{\mathcal{F}_t\}$ -martingale.

(b) For every nonnegative right continuous $\{\mathcal{F}_t\}$ -martingale Y and every $t \geq 0$ and $\tau \in S$,

$$(5.2) \quad E\left[\int_0^{t \wedge \tau} Y(s-) dA(s)\right] = E\left[\int_0^{t \wedge \tau} Y(s) dA(s)\right] \\ = E[Y(t \wedge \tau)A(t \wedge \tau)].$$

5.2 Remark (a) We allow the possibility that all three terms in (5.2) are infinite. If (5.2) holds for all bounded nonnegative right continuous $\{\mathcal{F}_t\}$ -martingales Y , then it holds for all nonnegative $\{\mathcal{F}_t\}$ -martingales, since on every bounded interval $[0, T]$ a nonnegative martingale Y is the limit of an increasing sequence $\{Y_n\}$ of bounded nonnegative martingales (e.g., take Y_n to be a right continuous modification of $Y_n^0(t) = E[Y(T) \wedge n | \mathcal{F}_t]$).

(b) If A is continuous, then the first equality in (5.2) is immediate. The second equality always holds, since (assuming Y is bounded) by the right continuity of Y

(5.3)

$$\begin{aligned} E\left[\int_0^{t \wedge \tau} Y(s) dA(s)\right] &= \lim_{n \rightarrow \infty} E\left[\int_0^{t \wedge \tau} Y\left(\frac{[ns] + 1}{n}\right) dA(s)\right] \\ &= \lim_{n \rightarrow \infty} E\left[\sum_k \chi_{[t \wedge \tau \geq (k-1)/n]} Y\left(\frac{k}{n}\right) \left(A\left(\frac{k}{n} - \right) - A\left(\frac{k-1}{n} - \right)\right)\right] \\ &= \lim_{n \rightarrow \infty} E\left[\sum_k \chi_{[t \wedge \tau \geq (k-1)/n]} Y(t \wedge \tau + n^{-1}) \left(A\left(\frac{k}{n} - \right) - A\left(\frac{k-1}{n} - \right)\right)\right] \\ &= \lim_{n \rightarrow \infty} E\left[Y(t \wedge \tau + n^{-1}) A\left(\frac{[n(t \wedge \tau)] + 1}{n} - \right)\right] \\ &= E[Y(t \wedge \tau)A(t \wedge \tau)]. \end{aligned}$$

The third equality in (5.3) follows from the fact that Y is a martingale.

(c) Property (b) is usually replaced by the requirement that A be *predictable*, but we do not need this concept elsewhere and hence do not introduce it here. See Dellacherie and Meyer (1982), page 194, or Elliott (1982), Theorem 8.15. \square

Proof. For each $\varepsilon > 0$, let X_ε be the optional projection of $\varepsilon^{-1} \int_0^\varepsilon X(\cdot + s) ds$,

$$(5.4) \quad X_\varepsilon(t) = E\left[\varepsilon^{-1} \int_0^\varepsilon X(t + s) ds \mid \mathcal{F}_t\right].$$

Then X_ε is a submartingale and

$$(5.5) \quad \lim_{\varepsilon \rightarrow 0} E[|X_\varepsilon(t) - X(t)|] = 0, \quad t \geq 0.$$

76 STOCHASTIC PROCESSES AND MARTINGALES

Let Y_ε be the optional projection of $\varepsilon^{-1}(X(\cdot + \varepsilon) - X(\cdot))$, and define

$$(5.6) \quad A_\varepsilon(t) = \int_0^t Y_\varepsilon(s) \, ds.$$

Since X is a submartingale,

$$(5.7) \quad Y_\varepsilon(t) = \varepsilon^{-1} E[X(t + \varepsilon) - X(t) | \mathcal{F}_t] \geq 0,$$

and hence A_ε is an increasing process. Furthermore

$$(5.8) \quad M_\varepsilon = X_\varepsilon - A_\varepsilon$$

is a martingale, since for $t, u \geq 0$ and $B \in \mathcal{F}_t$,

$$(5.9) \quad \begin{aligned} & \int_B (M_\varepsilon(t+u) - M_\varepsilon(t)) \, dP \\ &= \int_B \left(\varepsilon^{-1} \int_0^t X(t+u+s) \, ds - \varepsilon^{-1} \int_0^t X(t+s) \, ds \right. \\ &\quad \left. - \int_t^{t+u} \varepsilon^{-1} (X(s+\varepsilon) - X(s)) \, ds \right) \, dP = 0. \end{aligned}$$

We next observe that $\{A_\varepsilon(t): 0 < \varepsilon \leq 1\}$ is uniformly integrable for each $t \geq 0$. To see this, let $\tau_\lambda^\varepsilon = \inf\{s: A_\varepsilon(s) \geq \lambda\}$. Then

$$\begin{aligned} (5.10) \quad E[A_\varepsilon(t) - \lambda \wedge A_\varepsilon(t)] &= E[A_\varepsilon(t) - A_\varepsilon(\tau_\lambda^\varepsilon \wedge t)] \\ &= E[X_\varepsilon(t) - X_\varepsilon(\tau_\lambda^\varepsilon \wedge t)] \\ &= E[\chi_{\{\tau_\lambda^\varepsilon < t\}} (X_\varepsilon(t) - X_\varepsilon(\tau_\lambda^\varepsilon \wedge t))] \\ &= \varepsilon^{-1} \int_0^t E[\chi_{\{\tau_\lambda^\varepsilon < s\}} (X(t+s) - X(\tau_\lambda^\varepsilon \wedge t+s))] \, ds. \end{aligned}$$

Since $P\{\tau_\lambda^\varepsilon < t\} \leq \lambda^{-1} E[A_\varepsilon(t)] \leq \lambda^{-1} E[X(t + \varepsilon) - X(0)]$, the uniform integrability of $\{X(\tau \wedge (t+1)): \tau \in S\}$ implies the right side of (5.10) goes to zero as $\lambda \rightarrow \infty$ uniformly in $0 < \varepsilon \leq 1$. Consequently $\{A_\varepsilon(t): 0 < \varepsilon \leq 1\}$ is uniformly integrable (Appendix 2). For each $t \geq 0$, this uniform integrability implies the existence of a sequence $\{\varepsilon_n\}$ with $\varepsilon_n \rightarrow 0$, and a random variable $A(t)$ on (Ω, \mathcal{F}) such that

$$(5.11) \quad \lim_{n \rightarrow \infty} E[A_{\varepsilon_n}(t)\chi_B] = E[A(t)\chi_B],$$

for every $B \in \mathcal{F}$ (Appendix 2). By a diagonalization argument we may assume the same sequence $\{\varepsilon_n\}$ works for all $t \in \mathbb{Q} \cap [0, \infty)$.

Let $0 \leq s < t$, $s, t \in \mathbb{Q}$, and $B = \{A(t) < A(s)\}$. Then

$$(5.12) \quad E[(A(t) - A(s))\chi_B] = \lim_{n \rightarrow \infty} E[(A_{\varepsilon_n}(t) - A_{\varepsilon_n}(s))\chi_B] \geq 0,$$

so $A(s) \leq A(t)$ a.s. For $s, t \geq 0$, $s, t \in \mathbb{Q}$, and $B \in \mathcal{F}_t$,

$$(5.13) \quad E[(X(t+s) - A(t+s))\chi_B] = \lim_{n \rightarrow \infty} E[(M_{\epsilon_n}(t+s) - M_{\epsilon_n}(t))\chi_B] = 0,$$

and defining $M(t) = X(t) - A(t)$ for $t \in \mathbb{Q} \cap [0, \infty)$, we have $E[M(t+s)|\mathcal{F}_s] = M(t)$ for all $s, t \in \mathbb{Q} \cap [0, \infty)$. By the right continuity of $\{\mathcal{F}_t\}$ and Corollary 2.11, M extends to a right continuous $\{\mathcal{F}_t\}$ -martingale, and it follows that A has a right continuous increasing modification.

To see that (5.2) holds, let Y be a bounded right continuous $\{\mathcal{F}_t\}$ -martingale. Then for $t \geq 0$,

$$\begin{aligned} (5.14) \quad E[Y(t)A(t)] &= \lim_{n \rightarrow \infty} E[Y(t)A_{\epsilon_n}(t)] \\ &= \lim_{n \rightarrow \infty} E\left[\int_0^t Y(s-) dA_{\epsilon_n}(s)\right] \\ &= \lim_{n \rightarrow \infty} E\left[\int_0^t Y(s-) \epsilon_n^{-1} E[X(s+\epsilon_n) - X(s)|\mathcal{F}_s] ds\right] \\ &= \lim_{n \rightarrow \infty} E\left[\int_0^t Y(s-) \epsilon_n^{-1} (A(s+\epsilon_n) - A(s)) ds\right] \\ &= \lim_{n \rightarrow \infty} E\left[\int_0^{t+\epsilon_n} \int_0^t Y(s-) \epsilon_n^{-1} \chi_{(s, s+\epsilon_n)}(u) ds dA(u)\right] \\ &= E\left[\int_0^t Y(u-) dA(u)\right], \end{aligned}$$

and the same argument works with t replaced by $t \wedge \tau$.

Finally, to obtain the uniqueness, suppose A_1 and A_2 are processes with the desired properties. Then $A_1 - A_2$ is a martingale, and by Problem 15, if Y is a bounded, right continuous martingale,

$$\begin{aligned} (5.15) \quad E[Y(t)A_1(t)] &= E\left[\int_0^t Y(s-) dA_1(s)\right] \\ &= E\left[\int_0^t Y(s-) dA_2(s)\right] \\ &= E[Y(t)A_2(t)]. \end{aligned}$$

Let $B = \{A_1(t) > A_2(t)\}$ and $Y(s) = E[\chi_B|\mathcal{F}_s]$ (by Corollary 2.11, Y can be taken to be right continuous). Then (5.15) implies

$$(5.16) \quad E[(A_1(t) - A_2(t))\chi_B] = 0.$$

Similarly take $B = \{A_2(t) > A_1(t)\}$ and it follows that $A_1(t) = A_2(t)$ a.s. for each $t \geq 0$. The fact that A_1 and A_2 are indistinguishable follows from the right continuity. \square

5.3 Corollary If, in addition to the assumptions of Theorem 5.1, X is continuous, then A can be taken to be continuous.

Proof. Let A be as in Theorem 5.1. Let $\alpha > 0$ and $\tau = \inf\{t: X(t) - A(t) \notin [-\alpha, \alpha]\}$, and define $Y = A(\cdot \wedge \tau) - X(\cdot \wedge \tau) + \alpha$. Since X is continuous, $Y \geq 0$, and hence by (5.2),

$$(5.17) \quad E\left[\int_0^{t \wedge \tau} Y(s-) dA(s)\right] = E\left[\int_0^{t \wedge \tau} Y(s) dA(s)\right], \quad t \geq 0.$$

For $0 \leq s \leq \tau$, $Y(s-) \leq \alpha$, and hence (5.17) is finite, and

$$(5.18) \quad \begin{aligned} E\left[\int_0^{t \wedge \tau} (Y(s) - Y(s-)) dA(s)\right] \\ = E\left[\int_0^{t \wedge \tau} (A(s) - A(s-)) dA(s)\right] = 0, \quad t \geq 0. \end{aligned}$$

Since α is arbitrary, it follows that A is almost surely continuous. \square

5.4 Corollary Let X be a right continuous, $\{\mathcal{F}_t\}$ -local submartingale. Then there exists a right continuous, $\{\mathcal{F}_t\}$ -adapted, increasing process A satisfying Property (b) of Theorem 5.1 such that $M \equiv X - A$ is an $\{\mathcal{F}_t\}$ -local martingale.

Proof. Let $\tau_1 \leq \tau_2 \leq \dots$ be stopping times such that $\tau_n \rightarrow \infty$ and X^{τ_n} is a submartingale, and let $\gamma_n = \inf\{t: X(t) \leq -n\}$. Then $X^{\tau_n \wedge \gamma_n}$ is a submartingale of class DL , since for any $\{\mathcal{F}_t\}$ -stopping time τ ,

$$(5.19) \quad X^{\tau_n \wedge \gamma_n}(T) \wedge (-n) \leq X^{\tau_n \wedge \gamma_n}(T \wedge \tau) \leq E[X^{\tau_n \wedge \gamma_n}(T) | \mathcal{F}_\tau].$$

Let A_n be the increasing process for $X^{\tau_n \wedge \gamma_n}$ given by Theorem 5.1. Then $A = \lim_{n \rightarrow \infty} A_n$. \square

6. SQUARE INTEGRABLE MARTINGALES

Fix a filtration $\{\mathcal{F}_t\}$, and assume $\{\mathcal{F}_t\}$ is complete and right continuous. In this section all martingales, local martingales, and so on are $\{\mathcal{F}_t\}$ -martingales, $\{\mathcal{F}_t\}$ -local martingales, and so on.

A martingale M is *square integrable* if $E[|M(t)|^2] < \infty$ for all $t \geq 0$. A right continuous process M is a *local square integrable* martingale if there exist stopping times $\tau_1 \leq \tau_2 \leq \dots$ such that $\tau_n \rightarrow \infty$ a.s. and for each $n \geq 1$, $M^{\tau_n} \equiv M(\cdot \wedge \tau_n)$ is a square integrable martingale. Let \mathcal{M} denote the collection of right continuous, square integrable martingales, and let \mathcal{M}_{loc} denote the collection of right continuous local square integrable martingales. We also need to

define \mathcal{M}_c , the collection of continuous square integrable martingales, and $\mathcal{M}_{c, loc}$, the collection of continuous local martingales. (Note that a continuous local martingale is necessarily a local square integrable martingale.)

Each of these collections is a linear space. Let τ be a stopping time. If $M \in \mathcal{M}(\mathcal{M}_{loc}, \mathcal{M}_c, \mathcal{M}_{c, loc})$, then clearly $M^\tau \equiv M(\cdot \wedge \tau) \in \mathcal{M}(\mathcal{M}_{loc}, \mathcal{M}_c, \mathcal{M}_{c, loc})$.

6.1 Proposition If $M \in \mathcal{M}(\mathcal{M}_{loc})$, then $M^2 - [M]$ is a martingale (local martingale).

Proof. Let $M \in \mathcal{M}$. Since for $t, s \geq 0$ and $t = u_0 < u_1 < \dots < u_m = t + s$,

$$(6.1) \quad E[M^2(t+s) - M^2(t) | \mathcal{F}_t] = E[(M(t+s) - M(t))^2 | \mathcal{F}_t] \\ = E\left[\sum_{k=0}^{m-1} (M(u_{k+1}) - M(u_k))^2 \middle| \mathcal{F}_t\right],$$

the result follows by Proposition 3.4. The extension to local martingales is immediate. \square

If $M \in \mathcal{M}(\mathcal{M}_{loc})$, then M^2 satisfies the conditions of Theorem 5.1 (Corollary 5.4). Let $\langle M \rangle$ be the increasing process given by the theorem (corollary) with $X = M^2$. Then $M^2 - \langle M \rangle$ is a martingale (local martingale). If $M \in \mathcal{M}_c(\mathcal{M}_{c, loc})$, then by Proposition 3.6, $[M]$ is continuous, and Proposition 6.1 implies $[M]$ has the properties required for A in Theorem 5.1 (Corollary 5.4). Consequently, by uniqueness, $[M] = \langle M \rangle$ (up to indistinguishability).

For $M, N \in \mathcal{M}_{loc}$ we define

$$(6.2) \quad [M, N] = \frac{1}{2}([M+N, M+N] - [M, M] - [N, N])$$

and

$$(6.3) \quad \langle M, N \rangle = \frac{1}{2}(\langle M+N, M+N \rangle - \langle M, M \rangle - \langle N, N \rangle).$$

Of course, $[M, N]$ is the cross variation of M and N , that is (cf. (3.10)),

$$(6.4) \quad [M, N](t) = \lim_{n \rightarrow \infty} \sum_k (M(u_{k+1}^{(n)}) - M(u_k^{(n)}))(N(u_{k+1}^{(n)}) - N(u_k^{(n)}))$$

in probability. Note that $[M, M] = [M]$ and $\langle M, M \rangle = \langle M \rangle$. The following proposition indicates the interest in these quantities.

6.2 Proposition If $M, N \in \mathcal{M}(\mathcal{M}_{loc})$, then $MN - [M, N]$ and $MN - \langle M, N \rangle$ are martingales (local martingales).

Proof. Observe that

$$(6.5) \quad MN - [M, N] = \frac{1}{2}((M + N)^2 - [M + N, M + N] - (M^2 - [M]) - (N^2 - [N])),$$

and similarly for $MN - \langle M, N \rangle$. \square

If $\langle M, N \rangle = 0$, then M and N are said to be *orthogonal*. Note that $\langle M, N \rangle = 0$ implies MN and $[M, N]$ are martingales (local martingales).

7. SEMIGROUPS OF CONDITIONED SHIFTS

Let $\{\mathcal{F}_t\}$ be a complete filtration. Again all martingales, stopping times, and so on are $\{\mathcal{F}_t\}$ -martingales, $\{\mathcal{F}_t\}$ -stopping times, and so on. Let \mathcal{L} be the space of progressive (i.e., $\{\mathcal{F}_t\}$ -progressive) processes Y such that $\sup_t E[|Y(t)|] < \infty$. Defining

$$(7.1) \quad \|Y\| = \sup_t E[|Y(t)|]$$

and $\mathcal{N} = \{Y \in \mathcal{L}: \|Y\| = 0\}$, then \mathcal{L}/\mathcal{N} (the quotient space) is a Banach space with norm $\|\cdot\|$ satisfying the conditions of Chapter 1, Section 5, that is, (7.1) is of the form (5.1) of Chapter 1 ($\Gamma = \{\delta_t \times P: t \in [0, \infty)\}$). Since there is little chance of confusion, we do not distinguish between \mathcal{L} and \mathcal{L}/\mathcal{N} .

We define a semigroup of operators $\{\mathcal{T}(s)\}$ on \mathcal{L} by

$$(7.2) \quad \mathcal{T}(s)Y(t) = E[Y(t+s) | \mathcal{F}_t].$$

By Corollary 4.4, we can assume $(s, t, \omega) \rightarrow \mathcal{T}(s)Y(t, \omega)$ is $\mathcal{B}[0, \infty) \times \mathcal{C}$ -measurable. The semigroup property follows by

$$(7.3) \quad \begin{aligned} \mathcal{T}(u)\mathcal{T}(s)Y(t) &= E[E[Y(t+u+s) | \mathcal{F}_{t+u}] | \mathcal{F}_t] \\ &= E[Y(t+u+s) | \mathcal{F}_t] \\ &= \mathcal{T}(u+s)Y(t). \end{aligned}$$

Since

$$(7.4) \quad \sup_t E[|\mathcal{T}(s)Y(t)|] \leq \sup_t E[|Y(t)|],$$

$\{\mathcal{T}(s)\}$ is a measurable contraction semigroup on \mathcal{L} .

Integrals of the form $W = \int_a^b f(u)\mathcal{T}(u)Z du$ are well defined for Borel measurable f with $\int_a^b |f(u)| du < \infty$ and $Z \in \mathcal{L}$ by (5.4) of Chapter 1,

$$(7.5) \quad W(t) = E\left[\int_a^b f(u)Z(t+u) du \middle| \mathcal{F}_t\right],$$

and

$$(7.6) \quad \begin{aligned} \left\| \int_a^b f(u) \mathcal{T}(u) Z \, du \right\| &\leq \int_a^b |f(u)| \| \mathcal{T}(u) Z \| \, du \\ &\leq \| Z \| \int_a^b |f(u)| \, du. \end{aligned}$$

Define

$$(7.7) \quad \hat{\mathcal{A}} = \left\{ (Y, Z) \in \mathcal{L} \times \mathcal{L} : Y(t) - \int_0^t Z(s) \, ds \text{ is a martingale} \right\}.$$

Since $(Y, Z) \in \hat{\mathcal{A}}$ if and only if

$$(7.8) \quad \mathcal{T}(s)Y = Y + \int_0^s \mathcal{T}(u)Z \, du, \quad s \geq 0,$$

$\hat{\mathcal{A}}$ is the full generator for $\{\mathcal{T}(s)\}$ as defined in Chapter 1, Section 5. Note that the “harmonic functions”, that is, the solutions of $\hat{\mathcal{A}}Y = 0$, are the martingales in \mathcal{L} .

7.1 Theorem The operator $\hat{\mathcal{A}}$ defined in (7.7) is a dissipative linear operator with $\mathcal{R}(\lambda - \hat{\mathcal{A}}) = \mathcal{L}$ for all $\lambda > 0$ and resolvent

$$(7.9) \quad (\lambda - \hat{\mathcal{A}})^{-1}W = \int_0^\infty e^{-\lambda s} \mathcal{T}(s)W \, ds.$$

The largest closed subspace \mathcal{L}_0 of \mathcal{L} on which $\{\mathcal{T}(s)\}$ is strongly continuous is the closure of $\mathcal{D}(\hat{\mathcal{A}})$, and for each $Y \in \mathcal{L}_0$ and $s \geq 0$,

$$(7.10) \quad \mathcal{T}(s)Y = \lim_{n \rightarrow \infty} \left(I - \frac{s}{n} \hat{\mathcal{A}} \right)^{-n} Y.$$

Proof. (Cf. the proof of Proposition 5.1 of Chapter 1.) Suppose $(Y, Z) \in \hat{\mathcal{A}}$. Then

$$\begin{aligned} (7.11) \quad & \int_0^\infty e^{-\lambda s} \mathcal{T}(s)(\lambda Y - Z)(t) \, ds \\ &= \int_0^\infty e^{-\lambda s} E[\lambda Y(t+s) - Z(t+s) | \mathcal{F}_t] \, ds \\ &= \int_0^\infty e^{-\lambda s} E \left[\lambda Y(t) + \lambda \int_0^s Z(t+u) \, du - Z(t+s) \middle| \mathcal{F}_t \right] \, ds \\ &= Y(t) + E \left[\int_0^\infty e^{-\lambda s} \lambda \int_0^s Z(t+u) \, du \, ds \middle| \mathcal{F}_t \right] \\ &\quad - E \left[\int_0^\infty e^{-\lambda s} Z(t+s) \, ds \middle| \mathcal{F}_t \right] = Y(t). \end{aligned}$$

82 STOCHASTIC PROCESSES AND MARTINGALES

The last equality follows by interchanging the order of integration in the second term. This identity implies (7.9), which since $\mathcal{F}(s)$ is a contraction, implies $\hat{\mathcal{A}}$ is dissipative.

To see that $\mathcal{R}(\lambda - \hat{\mathcal{A}}) = \mathcal{L}$, let $W \in \mathcal{L}$, $Y = \int_0^\infty e^{-\lambda s} \mathcal{F}(s)W ds$, and $Z = \lambda Y - W$. An interchange in the order of integration gives the identity

$$(7.12) \quad \begin{aligned} \mathcal{F}(r)Y &= \int_0^\infty \lambda e^{-\lambda s} \int_0^s \mathcal{F}(r+u)W du ds \\ &= \int_0^\infty \lambda e^{-\lambda s} \int_r^{r+s} \mathcal{F}(u)W du ds, \end{aligned}$$

and we have

$$(7.13) \quad \begin{aligned} \int_0^r \mathcal{F}(u)Z du &= \int_0^\infty \lambda e^{-\lambda s} \int_0^r \mathcal{F}(s+u)W du ds - \int_0^r \mathcal{F}(u)W du \\ &= \int_0^\infty \lambda e^{-\lambda s} \int_s^{r+s} \mathcal{F}(u)W du ds - \int_0^r \mathcal{F}(u)W du. \end{aligned}$$

Subtracting (7.13) from (7.12) gives

$$(7.14) \quad \begin{aligned} \mathcal{F}(r)Y - \int_0^r \mathcal{F}(u)Z du &= \int_r^\infty \lambda e^{-\lambda s} \int_r^s \mathcal{F}(u)W du ds \\ &\quad - \int_0^r \lambda e^{-\lambda s} \int_s^r \mathcal{F}(u)W du ds + \int_0^r \mathcal{F}(u)W du \\ &= \int_r^\infty e^{-\lambda u} \mathcal{F}(u)W du - \int_0^r (1 - e^{-\lambda u}) \mathcal{F}(u)W du \\ &\quad + \int_0^r \mathcal{F}(u)W du = Y, \end{aligned}$$

which verifies (7.8) and implies $(Y, Z) \in \hat{\mathcal{A}}$.

If $W \in \mathcal{L}_0$, then $\lambda \int_0^\infty e^{-\lambda s} \mathcal{F}(s)W ds \in \mathcal{D}(\hat{\mathcal{A}})$ and $\lim_{\lambda \rightarrow \infty} \lambda \int_0^\infty e^{-\lambda s} \mathcal{F}(s)W ds = W$ (the limit being the strong limit in the Banach space \mathcal{L}). If $(Y, Z) \in \hat{\mathcal{A}}$, then

$$(7.15) \quad \begin{aligned} \| \mathcal{F}(s)Y - Y \| &= \sup_t E \left[\left| \int_0^s E[Z(t+u) | \mathcal{F}_t] du \right| \right] \leq s \sup_t E[|Z(t)|] \\ &= s \| Z \|, \end{aligned}$$

and hence $\mathcal{D}(\hat{\mathcal{A}}) \subset \mathcal{L}_0$. Therefore \mathcal{L}_0 is the closure of $\mathcal{D}(\hat{\mathcal{A}})$. Corollary 6.8 of Chapter 1 gives (7.10). \square

The following lemma may be useful in showing that a process is in $\mathcal{D}(\hat{\mathcal{A}})$.

7.2 Lemma Let $Y, Z_1, Z_2 \in \mathcal{L}$ and suppose that Y is right continuous and that $Z_1(t) \leq Z_2(t)$ a.s. for all t . If $Y(t) - \int_0^t Z_1(s) ds$ is a submartingale, and $Y(t) - \int_0^t Z_2(s) ds$ is a supermartingale, then there exists $Z \in \mathcal{L}$ satisfying $Z_1(t) \leq Z(t) \leq Z_2(t)$ a.s. for all $t \geq 0$, such that $Y(t) - \int_0^t Z(s) ds$ is a martingale.

7.3 Remark The assumption that Y is right continuous is for convenience only. There always exists a modification of Y that has right limits (since $Y(t) - \int_0^t Z_2(s) ds$ is a supermartingale). The lemma as stated then implies the existence of Z (adapted to $\{\mathcal{F}_{t+}\}$) for which $Y(t+) - \int_0^t Z(s) ds$ is an $\{\mathcal{F}_{t+}\}$ -martingale. Since $E[Y(t+) | \mathcal{F}_t] = Y(t)$ and $E[\int_t^{t+u} Z(s) ds | \mathcal{F}_t] = E[\int_t^{t+u} E[Z(s) | \mathcal{F}_s] ds | \mathcal{F}_t]$, $Y(t) - \int_0^t E[Z(s) | \mathcal{F}_s] ds$ is an $\{\mathcal{F}_t\}$ -martingale. \square

Proof. Without loss of generality we may assume $Z_1 = 0$. Then Y is a submartingale, and since $Y \vee 0$ and $(\int_0^t Z_2(s) ds - Y(t)) \vee 0$ are submartingales of class DL , Y and $\int_0^t Z_2(s) ds - Y(t)$ are also (note that $|Y(t)| \leq Y(t) \vee 0 + (\int_0^t Z_2(s) ds - Y(t)) \vee 0$). Consequently, by Theorem 5.1, there exist right continuous increasing processes A_1 and A_2 , with Property (b) of Theorem 5.1 such that $Y - A_1$ and $Y(t) - \int_0^t Z_2(s) ds + A_2(t)$ are martingales. Since $Y + A_2$ is a submartingale of class DL , and $Y + A_2 - (A_1 + A_2)$ and $Y(t) + A_2(t) - \int_0^t Z_2(s) ds$ are martingales, the uniqueness in Theorem 5.1 implies that with probability one,

$$(7.16) \quad A_1(t) + A_2(t) = \int_0^t Z_2(s) ds, \quad t \geq 0.$$

Since A_2 is increasing,

$$(7.17) \quad A_1(t+u) - A_1(t) \leq \int_t^{t+u} Z_2(s) ds, \quad t, u \geq 0,$$

so A_1 is absolutely continuous with derivative Z , where $0 \leq Z \leq Z_2$. \square

7.4 Corollary If $Y \in \mathcal{L}$, Y is right continuous, and there exists a constant M such that

$$(7.18) \quad |E[Y(t+s) - Y(t) | \mathcal{F}_t]| \leq Ms, \quad t, s \geq 0,$$

then there exists $Z \in \mathcal{L}$ with $|Z| \leq M$ a.s. such that $Y(t) - \int_0^t Z(s) ds$ is a martingale.

Proof. Take $Z_1(t) = -M$ and $Z_2(t) = M$ in Lemma 7.2. \square

7.5 Proposition Let $Y \in \mathcal{L}$ and let ξ be the optional projection of $\int_0^b Y(\cdot + s) ds$ and η the optional projection of $Y(\cdot + b) - Y$, that is, $\xi(t) = E[\int_0^b Y(t+s) ds | \mathcal{F}_t]$ and $\eta(t) = E[Y(t+b) - Y(t) | \mathcal{F}_t]$. Then $(\xi, \eta) \in \hat{\mathcal{A}}$.

Proof. This is just Proposition 5.2 of Chapter 1. \square

7.6 Proposition Let $\xi \in \mathcal{L}$. If $\{s^{-1}E[\xi(t+s) - \xi(t)|\mathcal{F}_t]: s > 0, t \geq 0\}$ is uniformly integrable and

$$(7.19) \quad s^{-1}E[\xi(t+s) - \xi(t)|\mathcal{F}_t] \xrightarrow{P} \eta(t) \quad \text{as } s \rightarrow 0+, \text{ a.e. } t,$$

then $(\xi, \eta) \in \hat{\mathcal{A}}$.

Proof. Let $\xi_\varepsilon(t) = \varepsilon^{-1}E[\int_0^\varepsilon \xi(t+s) ds | \mathcal{F}_t]$ and $\eta_\varepsilon(t) = \varepsilon^{-1}E[\xi(t+\varepsilon) - \xi(t) | \mathcal{F}_t]$. Then $(\xi_\varepsilon, \eta_\varepsilon) \in \mathcal{A}$ and as $\varepsilon \rightarrow 0$, $\xi_\varepsilon(t) \rightarrow \xi(t)$ and

$$(7.20) \quad \int_0^t \eta_\varepsilon(s) ds \rightarrow \int_0^t \eta(s) ds$$

in L^1 for each $t \geq 0$. \square

We close this section with some observations about the relationship between the semigroup of conditioned shifts and the semigroup associated with a Markov process.

For an adapted process X with values in a metric space (E, r) , let \mathcal{M} be the subspace of \mathcal{L} of processes of the form $\{f(X(t), t)\}$, where $f \in B(E \times [0, \infty))$, and let \mathcal{M}_0 be the subspace of processes of the form $\{f(X(t))\}$, $f \in B(E)$. Then X is a Markov process if and only if $\mathcal{T}(s): \mathcal{M} \rightarrow \mathcal{M}$ for all $s \geq 0$, and it is natural to call X temporally homogeneous if $\mathcal{T}(s): \mathcal{M}_0 \rightarrow \mathcal{M}_0$ for all $s \geq 0$.

Suppose X is a Markov process corresponding to a transition function $P(s, x, \Gamma)$, define the semigroup $\{T(t)\}$ on $B(E)$ by

$$(7.21) \quad T(s)f(x) = \int f(y)P(s, x, dy),$$

and let $\hat{\mathcal{A}}$ denote its full generator. Then for $Y \equiv f \circ X \in \mathcal{M}_0$,

$$(7.22) \quad \mathcal{T}(s)Y = T(s)f \circ X, \quad s \geq 0,$$

and for $(f, h) \in \hat{\mathcal{A}}, (f \circ X, h \circ X) \in \hat{\mathcal{A}}$.

8. MARTINGALES INDEXED BY DIRECTED SETS

In Chapter 6, we need a generalization of the optional sampling theorem to martingales indexed by directed sets. We give this generalization here because of its close relationship to the other material in this chapter.

A set \mathcal{I} is *partially ordered* if some pairs $(u, v) \in \mathcal{I} \times \mathcal{I}$ are ordered by a relation denoted $u \leq v$ (or $v \geq u$) that has the following properties:

$$(8.1) \quad \text{For all } u \in \mathcal{I}, \quad u \leq u.$$

$$(8.2) \quad \text{If } u \leq v \text{ and } v \leq u, \quad \text{then } u = v.$$

$$(8.3) \quad \text{If } u \leq v \text{ and } v \leq w, \quad \text{then } u \leq w.$$

A partially ordered set \mathcal{I} (together with a metric ρ on \mathcal{I}) is a *metric lattice* if (\mathcal{I}, ρ) is a metric space, if for $u, v \in \mathcal{I}$ there exist unique elements $u \wedge v \in \mathcal{I}$ and $u \vee v \in \mathcal{I}$ such that

$$(8.4) \quad \{w \in \mathcal{I}: w \leq u\} \cap \{w \in \mathcal{I}: w \leq v\} = \{w \in \mathcal{I}: w \leq u \wedge v\}$$

and

$$(8.5) \quad \{w \in \mathcal{I}: w \geq u\} \cap \{w \in \mathcal{I}: w \geq v\} = \{w \in \mathcal{I}: w \geq u \vee v\},$$

and if $(u, v) \rightarrow u \wedge v$ and $(u, v) \rightarrow u \vee v$ are continuous mappings of $\mathcal{I} \times \mathcal{I}$ onto \mathcal{I} . We write $\min\{u_1, \dots, u_m\}$ for $u_1 \wedge \dots \wedge u_m$, and $\max\{u_1, \dots, u_m\}$ for $u_1 \vee \dots \vee u_m$. We assume throughout this section that \mathcal{I} is a metric lattice.

For $u, v \in \mathcal{I}$ with $u \leq v$, the set $[u, v] \equiv \{w \in \mathcal{I}: u \leq w \leq v\}$ is called an *interval*. Note that $[u, v]$ is a closed subset of \mathcal{I} . A subset $F \subset \mathcal{I}$ is *separable from above* if there exists a sequence $\{\alpha_n\} \subset F$ such that $w = \lim_{n \rightarrow \infty} \min\{\alpha_i: w \leq \alpha_i, i \leq n\}$ for all $w \in F$. We call the sequence $\{\alpha_n\}$ a *separating sequence*. Note that F can be separable without being separable from above. Define $\mathcal{I}_u = \{v \in \mathcal{I}: v \leq u\}$.

Let (Ω, \mathcal{F}, P) be a probability space. As in the case $\mathcal{I} = [0, \infty)$, a collection $\{\mathcal{F}_u\} = \{\mathcal{F}_u, u \in \mathcal{I}\}$ of sub- σ -algebras of \mathcal{F} is a *filtration* if $u \leq v$ implies $\mathcal{F}_u \subset \mathcal{F}_v$, and an \mathcal{I} -valued random variable τ is a *stopping time* if $\{\tau \leq u\} \in \mathcal{F}_u$ for all $u \in \mathcal{I}$. For a stopping time τ ,

$$(8.6) \quad \mathcal{F}_\tau = \{A \in \mathcal{F}: A \cap \{\tau \leq u\} \in \mathcal{F}_u \text{ for all } u \in \mathcal{I}\}.$$

A filtration $\{\mathcal{F}_u\}$ is *complete* if (Ω, \mathcal{F}, P) is complete and $\mathcal{F}_u \supset \{A \in \mathcal{F}: P(A) = 0\}$ for all $u \in \mathcal{I}$.

Let $\Gamma_u^n = \{v: \inf_{w \leq u} \rho(v, w) < n^{-1}\}$. We say that $\{\mathcal{F}_u\}$ is *right continuous* if

$$(8.7) \quad \mathcal{F}_u = \bigcap_n \bigvee_{v \in \Gamma_u^n} \mathcal{F}_v.$$

See Problem 20 for an alternative definition of right continuity.

8.1 Proposition Let τ_1, τ_2, \dots be $\{\mathcal{F}_u\}$ -stopping times. Then the following hold:

(a) $\max_{k \leq n} \tau_k$ is an $\{\mathcal{F}_u\}$ -stopping time.

(b) Suppose $\{\mathcal{F}_u\}$ is right continuous and complete. If τ is an \mathcal{I} -valued random variable and $\tau = \lim_{n \rightarrow \infty} \tau_n$ a.s., then τ is an $\{\mathcal{F}_u\}$ -stopping time.

86 STOCHASTIC PROCESSES AND MARTINGALES

Proof. (a) As in the case $\mathcal{I} = [0, \infty)$, $\{\max_{k \leq n} \tau_k \leq u\} = \bigcap_{k \leq n} \{\tau_k \leq u\} \in \mathcal{F}_u$.

(b) By the right continuity,

$$(8.8) \quad B = \bigcap_n \bigcup_m \bigcap_{k \geq m} \{\tau_k \in \Gamma_u^m\} \in \mathcal{F}_u,$$

and hence

$$(8.9)$$

$$\{\tau \leq u\} = \left(\{\tau \leq u\} \cap \left\{ \tau \neq \lim_{n \rightarrow \infty} \tau_n \right\} \right) \cup \left(B \cap \left\{ \tau = \lim_{n \rightarrow \infty} \tau_n \right\} \right) \in \mathcal{F}_u. \quad \square$$

8.2 Proposition Suppose τ is an $\{\mathcal{F}_u\}$ -stopping time and $a \in \mathcal{I}$. Define

$$(8.10) \quad \tau^a = \begin{cases} \tau & \text{on } \{\tau \leq a\} \\ a & \text{otherwise.} \end{cases}$$

Then τ^a is an $\{\mathcal{F}_u\}$ -stopping time.

8.3 Remark Note that τ^a is not in general equal to $\tau \wedge a$, which need not be a stopping time. \square

Proof. If $u = a$, $\{\tau^a \leq u\} = \Omega \in \mathcal{F}_u$. If $u \leq a$, but $u \neq a$, then $\{\tau^a \leq u\} = \{\tau \leq u\} \in \mathcal{F}_u$. In general, $\{\tau^a \leq u\} = \{\tau^a \leq u \wedge a\} \in \mathcal{F}_{u \wedge a} \subset \mathcal{F}_u$. \square

8.4 Proposition Suppose τ is an $\{\mathcal{F}_u\}$ -stopping time, $a \in \mathcal{I}$ with $\tau \leq a$, and \mathcal{I}_a is separable from above. Let $\{\alpha_i\}$ be a separating sequence for \mathcal{I}_a with $\alpha_1 = a$, and define

$$(8.11) \quad \tau_n = \min \{\alpha_i : \tau \leq \alpha_i, i \leq n\}, \quad n \geq 1.$$

Then τ_n is a stopping time for each $n \geq 1$, $a = \tau_1 \geq \tau_2 \geq \dots$, and $\lim_{n \rightarrow \infty} \tau_n = \tau$.

Proof. Let F_n be the finite collection of possible values of τ_n . For $u \in F_n$,

$$(8.12) \quad \{\tau_n = u\} = \{\tau \leq u\} \cap \bigcap_{\substack{v \in F_n \cap \mathcal{I}_a \\ v \neq u}} \{\tau \leq v\}^c \in \mathcal{F}_u,$$

and in general

$$(8.13) \quad \{\tau_n \leq u\} = \bigcup_{v \in F_n \cap \mathcal{I}_a} \{\tau_n = v\} \in \mathcal{F}_u.$$

The rest follows from the definition of a separating sequence. \square

Let X be an E -valued process indexed by \mathcal{I} . Then X is $\{\mathcal{F}_u\}$ -adapted if $X(u)$ is \mathcal{F}_u -measurable for each $u \in \mathcal{I}$, and X is $\{\mathcal{F}_u\}$ -progressive if for each $u \in \mathcal{I}$, the restriction of X to $\mathcal{I}_u \times \Omega$ is $\mathcal{B}(\mathcal{I}_u) \times \mathcal{F}_u$ -measurable. As in the

case $\mathcal{I} = [0, \infty)$, if \mathcal{F}_u is separable from above for each $u \in \mathcal{I}$, and X is right continuous (i.e., $\lim_{v \rightarrow u} X(u \vee v, \omega) = X(u, \omega)$ for all $u \in \mathcal{I}$ and $\omega \in \Omega$) and $\{\mathcal{F}_u\}$ -adapted, then X is $\{\mathcal{F}_u\}$ -progressive.

8.5 Proposition Let τ and σ be $\{\mathcal{F}_u\}$ -stopping times with $\tau \leq \sigma$, and let X be $\{\mathcal{F}_u\}$ -progressive. Then the following hold:

- (a) \mathcal{F}_τ is a σ -algebra.
- (b) $\mathcal{F}_\tau \subset \mathcal{F}_\sigma$.
- (c) If \mathcal{F}_u is separable from above for each $u \in \mathcal{I}$, then τ and $X(\tau)$ are \mathcal{F}_τ -measurable.

Proof. The proofs for parts (a) and (b) are the same as for the corresponding results in Proposition 1.4. Fix $a \in \mathcal{I}$. We first want to show that τ^a is \mathcal{F}_a -measurable. Let $\{\alpha_n\}$ be a separating sequence for \mathcal{I}_a with $\alpha_1 = a$, and define $\tau_n^a = \min\{\alpha_i : \tau^a \leq \alpha_i, i \leq n\}$. Then (8.12) implies τ_n^a is \mathcal{F}_a -measurable. Since $\lim_{n \rightarrow \infty} \tau_n^a = \tau^a$, τ^a is \mathcal{F}_a -measurable, and $X(\tau^a)$ is \mathcal{F}_a -measurable by the argument in the proof of Proposition 1.4(d). Finally, $\{\tau \in \Gamma\} \cap \{\tau \leq a\} = \{\tau^a \in \Gamma\} \cap \{\tau \leq a\} \in \mathcal{F}_a$ for all $a \in \mathcal{I}$ and $\Gamma \in \mathcal{B}(\mathcal{I})$, so $\{\tau \in \Gamma\} \in \mathcal{F}_\tau$, and τ is \mathcal{F}_τ -measurable. The same argument implies that $X(\tau)$ is \mathcal{F}_τ -measurable. \square

8.6 Proposition Suppose $\{\mathcal{F}_u\}$ is a right continuous filtration indexed by \mathcal{I} . For each $t \geq 0$, let $\tau(t)$ be an $\{\mathcal{F}_u\}$ -stopping time such that $s \leq t$ implies $\tau(s) \leq \tau(t)$ and $\tau(t)$ is a right continuous function of t . Let $\mathcal{H}_t = \mathcal{F}_{\tau(t)}$, and let η be an $\{\mathcal{H}_t\}$ -stopping time. Then $\tau(\eta)$ is an $\{\mathcal{F}_u\}$ -stopping time.

Proof. First assume η is discrete. Then

$$(8.14) \quad \{\tau(\eta) \leq u\} = \bigcup_i (\{\eta = t_i\} \cap \{\tau(t_i) \leq u\}).$$

Since $\{\eta = t_i\} \in \mathcal{F}_{\tau(t_i)}$, $\{\eta = t_i\} \cap \{\tau(t_i) \leq u\} \in \mathcal{F}_u$. For general η approximate η by a decreasing sequence of discrete stopping times (cf. Proposition 1.3) and apply Proposition 8.1(b). \square

A real-valued process X indexed by \mathcal{I} is an $\{\mathcal{F}_u\}$ -martingale if $E[|X(u)|] < \infty$ for all $u \in \mathcal{I}$, X is $\{\mathcal{F}_u\}$ -adapted, and

$$(8.15) \quad E[X(v) | \mathcal{F}_u] = X(u)$$

for all $u, v \in \mathcal{I}$ with $u \leq v$.

8.7 Theorem Let \mathcal{I} be a metric lattice and suppose each interval in \mathcal{I} is separable from above. Let X be a right continuous $\{\mathcal{F}_u\}$ -martingale, and let τ_1

88 STOCHASTIC PROCESSES AND MARTINGALES

and τ_2 be $\{\mathcal{F}_u\}$ -stopping times with $\tau_1 \leq \tau_2$. Suppose there exist $\{u_m\}, \{v_m\} \subset \mathcal{S}$ such that

$$(8.16) \quad \lim_{m \rightarrow \infty} P\{\tau_m \leq \tau_1 \leq \tau_2 \leq v_m\} = 1$$

and

$$(8.17) \quad \lim_{m \rightarrow \infty} E[|X(v_m)| \chi_{\{\tau_2 \leq v_m\}}] = 0,$$

and that $E[|X(\tau_2)|] < \infty$. Then

$$(8.18) \quad E[X(\tau_2) | \mathcal{F}_{\tau_1}] = X(\tau_1).$$

Proof. Fix $m \geq 1$ and for $i = 1, 2$ define

$$(8.19) \quad \tau_{im} = \begin{cases} \tau_i \vee u_m & \text{on } \{\tau_i \leq v_m\} \\ v_m & \text{otherwise.} \end{cases}$$

Let $\{\alpha_n\}$ be separating for $[u_m, v_m]$ with $\alpha_1 = v_m$, and define $\tau_{im}^n = \min\{\alpha_k : \tau_{im} \leq \alpha_k, k \leq n\}$. Note τ_{im}^n assumes finitely many values, and $\tau_{1m}^n \leq \tau_{2m}^n$. Fix n , and let Γ be the set of values assumed by τ_{1m}^n and τ_{2m}^n . For $\alpha \in \Gamma$ with $\alpha \neq v_m$, $\{\tau_{1m}^n = \alpha\} = \{\tau_1 \leq \alpha\} \cap \{\tau_{1m}^n = \alpha\}$, and hence $A \cap \{\tau_{1m}^n = \alpha\} = A \cap \{\tau_1 \leq \alpha\} \cap \{\tau_{1m}^n = \alpha\} \in \mathcal{F}_\alpha$. Consequently for $\alpha \in \Gamma$ with $\alpha \neq v_m$,

$$\begin{aligned} (8.20) \quad & \int_A \chi_{\{\tau_{1m}^n = \alpha\}} X(\tau_{2m}^n) dP \\ &= \int_A \chi_{\{\tau_{1m}^n = \alpha\}} \sum_{\beta \in \Gamma} E[X(v_m) | \mathcal{F}_\beta] \chi_{\{\tau_{1m}^n = \beta\}} dP \\ &= \sum_{\beta \in \Gamma} \int_A \chi_{\{\tau_{1m}^n = \alpha\} \cap \{\tau_{1m}^n = \beta\}} X(v_m) dP \\ &= \int_A \chi_{\{\tau_{1m}^n = \alpha\}} X(v_m) dP \\ &= \int_A \chi_{\{\tau_{1m}^n = \alpha\}} X(\alpha) dP. \end{aligned}$$

Since $\tau_{1m}^n = v_m$ implies $\tau_{2m}^n = v_m$, (8.20) is immediate for $\alpha = v_m$, and summing over $\alpha \in \Gamma$, (8.20) implies

$$(8.21) \quad \int_A X(\tau_{2m}^n) dP = \int_A X(\tau_{1m}^n) dP.$$

Letting $n \rightarrow \infty$ and then $m \rightarrow \infty$ gives

$$(8.22) \quad \int_A X(\tau_2) dP = \int_A X(\tau_1) dP,$$

which implies (8.18). \square

9. PROBLEMS

1. Show that if X is right (left) continuous and $\{\mathcal{F}_t\}$ -adapted, then X is $\{\mathcal{F}_t\}$ -progressive.

Hint: Fix $t > 0$ and approximate X on $[0, t] \times \Omega$ by X_n given by $X_n(s, \omega) = X(t \wedge (([ns] + 1)/n), \omega)$.

2. (a) Suppose X is E -valued and $\{\mathcal{F}_t\}$ -progressive, and $f \in B(E)$. Show that $f \circ X$ is $\{\mathcal{F}_t\}$ -progressive and $Y(t) \equiv \int_0^t f(X(s)) ds$ is $\{\mathcal{F}_t\}$ -adapted.
 (b) Suppose X is E -valued, measurable, and $\{\mathcal{F}_t\}$ -adapted, and $f \in B(E)$. Show that $f \circ X$ is $\{\mathcal{F}_t\}$ -adapted and that $Y(t) \equiv \int_0^t f(X(s)) ds$ has an $\{\mathcal{F}_t\}$ -adapted modification.

3. Let Y be a version of X . Suppose X is right continuous. Show that there is a modification of Y that is right continuous.

4. Let (E, r) be a complete, separable metric space.

- (a) Let ξ_1, ξ_2, \dots be E -valued random variables defined on (Ω, \mathcal{F}, P) . Let $A = \{\omega : \lim \xi_n(\omega) \text{ exists}\}$. Show that $A \in \mathcal{F}$, and that for $x \in E$,

$$(9.1) \quad \xi(\omega) = \begin{cases} \lim \xi_n(\omega), & \omega \in A, \\ x, & \omega \in A^c, \end{cases}$$

is a random variable (i.e., is \mathcal{F} -measurable).

- (b) Let X be an E -valued process that is right continuous in probability, that is, for each $\varepsilon > 0$ and $t \geq 0$,

$$(9.2) \quad \lim_{s \rightarrow t^+} P\{r(X(t), X(s)) > \varepsilon\} = 0.$$

Show that X has a modification that is progressive.

Hint: Show that for each n there exists a countable collection of disjoint intervals $[t_a^n, s_a^n]$ such that $[0, \infty) = \bigcup [t_a^n, s_a^n]$ and

$$(9.3) \quad P\{r(X(t_a^n), X(s)) > 2^{-n}\} < 2^{-n}, \quad t_a^n \leq s < s_a^n.$$

5. Suppose X is a modification of Y , and X and Y are right continuous. Show that X and Y are indistinguishable.

6. Let X be a stochastic process, and let τ be a discrete $\{\mathcal{F}_t^X\}$ -stopping time. Show that

$$(9.4) \quad \mathcal{F}_\tau^X = \sigma(X(t \wedge \tau) : t \geq 0).$$

7. Let $\{\mathcal{F}_t\}$ be a filtration. Show that $\{\mathcal{F}_{t+}\}$ is right continuous.

8. Let (Ω, \mathcal{F}, P) be a probability space. Let S be the collection of equivalence classes of real-valued random variables where two random vari-

90 STOCHASTIC PROCESSES AND MARTINGALES

ables are equivalent if they are almost surely equal. Let γ be defined by (2.27). Show that γ is a metric on S corresponding to convergence in probability.

9. (a) Let $\{x_n\}$ satisfy $\sup_n x_n < \infty$ and $\inf_n x_n > -\infty$, and assume that for each $a < b$ either $\{n: x_n \geq b\}$ or $\{n: x_n \leq a\}$ is finite. Show that $\lim_{n \rightarrow \infty} x_n$ exists.
 (b) Verify the existence of the limits in (2.24) and (2.25) and show that $Y^-(t, \omega) = Y(t-, \omega)$ for all $t > 0$ and $\omega \in \Omega_0$.
10. (a) Suppose X is a real-valued integrable random variable on (Ω, \mathcal{F}, P) . Let Γ be the collection of sub- σ -algebras of \mathcal{F} . Show that $\{E[X|\mathcal{D}]: \mathcal{D} \in \Gamma\}$ is uniformly integrable.
 (b) Let X be a right continuous $\{\mathcal{F}_t\}$ -martingale and let S be the collection of $\{\mathcal{F}_t\}$ -stopping times. Show that for each $T > 0$, $\{X(T \wedge \tau): \tau \in S\}$ is uniformly integrable.
 (c) Let X be a right continuous, nonnegative $\{\mathcal{F}_t\}$ -submartingale. Show that for each $T > 0$, $\{X(T \wedge \tau): \tau \in S\}$ is uniformly integrable.
11. (a) Let X be a right continuous $\{\mathcal{F}_t\}$ -submartingale and τ a finite $\{\mathcal{F}_t\}$ -stopping time. Suppose that for each $t > 0$, $E[\sup_{s \geq t} |X(\tau + s) - X(\tau)|] < \infty$. Show that $Y(t) \equiv X(\tau + t) - X(\tau)$ is an $\{\mathcal{F}_{t+s}\}$ -submartingale.
 (b) Let X be a right continuous $\{\mathcal{F}_t\}$ -submartingale and τ_1 and τ_2 be finite $\{\mathcal{F}_t\}$ -stopping times. Suppose $\tau_1 \leq \tau_2$ and $E[\sup_s |X((\tau_1 + s) \wedge \tau_2) - X(\tau_1)|] < \infty$. Show that $E[X(\tau_2) - X(\tau_1)|\mathcal{F}_{\tau_1}] \geq 0$.
12. Let X be a submartingale. Show that $\sup_t E[|X(t)|] < \infty$ if and only if $\sup_t E[X^+(t)] < \infty$.
13. Let η and ξ be independent random variables with $P\{\eta = 1\} = P\{\eta = -1\} = \frac{1}{2}$ and $E[|\xi|] = \infty$. Define

$$(9.5) \quad X(t) = \begin{cases} 0, & 0 \leq t < 1, \\ \eta \xi, & t \geq 1, \end{cases}$$

and

$$(9.6) \quad \mathcal{F}_t = \begin{cases} \sigma(\xi), & 0 \leq t < 1, \\ \sigma(\xi, \eta), & t \geq 1. \end{cases}$$

Show that X is an $\{\mathcal{F}_t\}$ -local martingale, but that X is not an $\{\mathcal{F}_t^X\}$ -local martingale.

14. Let E be a separable Banach space with norm $\|\cdot\|$, and let X be an E -valued random variable defined on (Ω, \mathcal{F}, P) .

- (a) Show that for every $\varepsilon > 0$ there exist $\{x_i\} \subset E$ and $\{B_i^\varepsilon\} \subset \mathcal{F}$ with $B_i^\varepsilon \cap B_j^\varepsilon = \emptyset$ for $i \neq j$, such that

$$(9.7) \quad X_\varepsilon = \sum x_i \chi_{B_i^\varepsilon}$$

satisfies $\|X - X_\varepsilon\| \leq \varepsilon$.

- (b) Suppose $E[\|X\|] < \infty$. Define

$$(9.8) \quad E[X_\varepsilon | \mathcal{D}] = \sum x_i P(B_i^\varepsilon | \mathcal{D}),$$

and show that

$$(9.9) \quad \|E[X_\varepsilon | \mathcal{D}] - E[X_\varepsilon | \mathcal{D}]\| \\ \leq E[\|X_{\varepsilon_1} - X_{\varepsilon_2}\| | \mathcal{D}] \leq \varepsilon_1 + \varepsilon_2$$

so that one can define

$$(9.10) \quad E[X | \mathcal{D}] = \lim_{\varepsilon \rightarrow 0} E[X_\varepsilon | \mathcal{D}].$$

- (c) Extend Theorem 4.2 and Corollary 4.5 to bounded, measurable, E -valued processes.
15. Let A_1 and A_2 be right continuous increasing processes with $A_1(0) = A_2(0) = 0$ and $E[A_i(t)] < \infty$, $i = 1, 2$, $t > 0$. Suppose that $A_1 - A_2$ is an $\{\mathcal{F}_t\}$ -martingale and that Y is bounded, right continuous, has left limits at each $t > 0$, and is $\{\mathcal{F}_t\}$ -adapted. Show that

$$(9.11) \quad \int_0^t Y(s-) d(A_1(s) - A_2(s)) \\ = \int_0^t Y(s-) dA_1(s) - \int_0^t Y(s-) dA_2(s)$$

is an $\{\mathcal{F}_t\}$ -martingale. (The integrals are defined as in (5.1).)

Hint: Let

$$(9.12) \quad Y_\varepsilon(t) = \varepsilon^{-1} \int_{t-\varepsilon}^t Y(s) ds$$

and apply Proposition 3.2 to

$$(9.13) \quad \int_0^t Y_\varepsilon(s) d(A_1(s) - A_2(s)).$$

16. Let Y be a unit Poisson process, and define $M(t) = Y(t) - t$. Show that M is a martingale and compute $[M]$ and $\langle M \rangle$.

92 STOCHASTIC PROCESSES AND MARTINGALES

17. Let W be standard Brownian motion. Use the law of large numbers to compute $[W]$. (Recall $[W] = \langle W \rangle$ since W is continuous.)
18. Let \mathcal{L}^2 be the space of real-valued $\{\mathcal{F}_t\}$ -progressive processes X satisfying $\int_0^\infty E[|X(t)|^2] dt < \infty$. Note that \mathcal{L}^2 is a Hilbert space with inner product

$$(9.14) \quad (X, Y) = \int_0^\infty E[X(t)Y(t)] dt$$

and norm $\|X\| = \sqrt{(X, X)}$. Let A be a bounded linear operator on \mathcal{L}^2 . Then A^* , the adjoint of A , is the unique bounded linear operator satisfying $(AX, Y) = (X, A^*Y)$ for all $X, Y \in \mathcal{L}^2$. Fix $s \geq 0$ and let $U(s)$ be the bounded linear operator defined by

$$(9.15) \quad U(s)X(t) = \begin{cases} X(t-s), & t \geq s, \\ 0, & t < s. \end{cases}$$

What is $U^*(s)$? (Remember that $U^*(s)X$ must be $\{\mathcal{F}_t\}$ -progressive.)

19. Let M_1, M_2, \dots, M_m be independent martingales. Let $\mathcal{I} = [0, \infty)^m$ and define

$$(9.16) \quad M(u) = \prod_{i=1}^m M_i(u_i), \quad u \in \mathcal{I}.$$

- (a) Show that M is a martingale indexed by \mathcal{I} .
- (b) Let $\mathcal{F}_u = \sigma(M(v): v \leq u)$, and let $\tau(t)$, $t \geq 0$, be $\{\mathcal{F}_u\}$ -stopping times satisfying $\tau(s) \leq \tau(t)$ for $s \leq t$. Suppose that for each t there exists $c_t \in \mathcal{I}$ such that $\tau(t) \leq c_t$ a.s. Let $X_i(t) = M_i(\tau_i(t))$. Show that X_1, \dots, X_m are orthogonal $\{\mathcal{F}_{\tau_i(t)}\}$ -martingales. More generally, show that for any $I \subset \{1, \dots, m\}$, $\prod_{i \in I} X_i$ is an $\{\mathcal{F}_{\tau_i(t)}\}$ -martingale.
20. Let \mathcal{I} be a metric lattice. Show that a filtration $\{\mathcal{F}_u, u \in \mathcal{I}\}$ is right continuous if and only if for every u , $\{u_n\} \subset \mathcal{I}$ with $u \leq u_n$, $n = 1, 2, \dots$, and $u = \lim_{n \rightarrow \infty} u_n$, we have $\mathcal{F}_u = \bigcap \mathcal{F}_{u_n}$.
21. (a) Suppose M is a local martingale and $\sup_{s \leq t} |M(s)| \in L^1$ for each $t > 0$. Show that M is a martingale.
(b) Suppose M is a positive local martingale. Show that M is a super-martingale.
22. Let X and Y be measurable and $\{\mathcal{G}_t\}$ -adapted. Suppose $E[|X(t)| \int_0^t |Y(s)| ds] < \infty$ and $E[\int_0^t |X(s)Y(s)| ds] < \infty$ for every $t \geq 0$, and that X is a $\{\mathcal{G}_t\}$ -martingale. Show that $X(t) \int_0^t Y(s) ds - \int_0^t X(s)Y(s) ds$ is a martingale. (Cf. Proposition 3.2 but note we are not assuming X is right continuous.)

23. Let X be a real-valued $\{\mathcal{G}_t\}$ -adapted process, with $E[|X(t)|] < \infty$ for every $t \geq 0$. Show that X is a $\{\mathcal{G}_t\}$ -martingale if and only if $E[X(\tau)] = E[X(0)]$ for every $\{\mathcal{G}_t\}$ -stopping time τ assuming only finitely many values.
24. Let M_1, \dots, M_n be right continuous $\{\mathcal{G}_t\}$ -martingales, and suppose that, for each $I \subset \{1, \dots, n\}$, $\prod_{i \in I} M_i$ is also a $\{\mathcal{G}_t\}$ -martingale. Let τ_1, \dots, τ_n be $\{\mathcal{G}_t\}$ -stopping times, and suppose $E[\prod_{i=1}^n \sup_{t \leq \tau_i} |M_i(t)|] < \infty$. Show that $M(t) \equiv \prod_{i=1}^n M_i(t \wedge \tau_i)$ is a $\{\mathcal{G}_t\}$ -martingale.
- Hint:* Use Problem 23 and induction on n .
25. Let X be a real-valued stochastic process, and $\{\mathcal{F}_t\}$ a filtration. (X is not necessarily $\{\mathcal{F}_t\}$ -adapted.) Suppose $E[X(t)|\mathcal{F}_t] \geq 0$ for each t . Show that $E[X(\tau)|\mathcal{F}_\tau] \geq 0$ for each finite, discrete $\{\mathcal{F}_t\}$ -stopping time τ .
26. Let (M, \mathcal{M}, μ) be a probability space, and let $\mathcal{M}_1 \subset \mathcal{M}_2 \subset \dots$ be an increasing sequence of discrete σ -algebras, that is, for $n = 1, 2, \dots$, $\mathcal{M}_n = \sigma(A_i^n, i = 1, 2, \dots)$ where the A_i^n are disjoint, and $M = \bigcup_i A_i^n$. Let $X \in L^1(\mu)$, and define

$$(9.17) \quad X_n = \sum_i \mu(A_i^n)^{-1} \int_{A_i^n} X \, d\mu \chi_{A_i^n}$$

- (a) Show that $\{X_n\}$ is an $\{\mathcal{M}_n\}$ -martingale.
 (b) Suppose $\mathcal{M} = \bigvee_n \mathcal{M}_n$. Show that $\lim_{n \rightarrow \infty} X_n = X$ μ -a.s. and in $L^1(\mu)$.

27. Let $\{X(t): t \in \mathcal{I}\}$ be a stochastic process. Show that

$$(9.18) \quad \sigma(X(s): s \in \mathcal{I}) = \bigcup_{I \subset \mathcal{I}} \sigma(X(s): s \in I)$$

where the union is over all countable subsets of \mathcal{I} .

28. Let τ and σ be $\{\mathcal{F}_t\}$ -stopping times. Show that $\mathcal{F}_{\tau \wedge \sigma} = \mathcal{F}_\tau \cap \mathcal{F}_\sigma$ and $\mathcal{F}_{\tau \vee \sigma} = \mathcal{F}_\tau \vee \mathcal{F}_\sigma$.
29. Let X be a right continuous, E -valued process adapted to a filtration $\{\mathcal{F}_t\}$. Let $f \in C(E)$ and $g, h \in B(E)$, and suppose that

$$(9.19) \quad M_f(t) \equiv f(X(t)) - f(X(0)) - \int_0^t g(X(s)) \, ds$$

and

$$(9.20) \quad f(X(t))^2 - f(X(0))^2 - \int_0^t h(X(s)) \, ds$$

94 STOCHASTIC PROCESSES AND MARTINGALES

are $\{\mathcal{F}_t\}$ -martingales. Show that

$$(9.21) \quad M_f(t)^2 - \int_0^t (h(X(s)) - 2f(X(s))g(X(s))) \, ds$$

is an $\{\mathcal{F}_t\}$ -martingale.

10. NOTES

Most of the material in Section 1 is from Doob (1953) and Dynkin (1961), and has been developed and refined by the Strasbourg school. For the most recent presentation see Dellacherie and Meyer (1978, 1982). Section 2 is almost entirely from Doob (1953).

The notion of a local martingale is due to Itô and Watanabe (1965). Proposition 3.2 is of course a special case of much more general results on stochastic integrals. See Dellacherie and Meyer (1982) and Chapter 5. Proposition 3.4 is due to Doleans-Dade (1969). The projection theorem is due to Meyer (1968). Theorem 5.1 is also due to Meyer. See Meyer (1966), page 122.

The semigroup of conditioned shifts appeared first in work of Rishel (1970). His approach is illustrated by Problem 18. The presentation in Section 7 is essentially that of Kurtz (1975). Chow (1960) gave one version of an optional sampling theorem for martingales indexed by directed sets. Section 8 follows Kurtz (1980b).

Problem 4(b) is essentially Theorem II.2.6 of Doob (1953). See Dellacherie and Meyer (1978), page 99, for a more refined version.

3 CONVERGENCE OF PROBABILITY MEASURES

In this chapter we study convergence of sequences of probability measures defined on the Borel subsets of a metric space (S, d) and in particular of $D_S[0, \infty)$, the space of right continuous functions from $[0, \infty)$ into a metric space (E, r) having left limits. Our starting point in Section 1 is the Prohorov metric ρ on $\mathcal{P}(S)$, the set of Borel probability measures on S , and in Section 2 we give Prohorov's characterization of the compact subsets of $\mathcal{P}(S)$. In Section 3 we define weak convergence of a sequence in $\mathcal{P}(S)$ and consider its relationship to convergence in the Prohorov metric (they are equivalent if S is separable). Section 4 concerns the concepts of separating and convergence determining classes of bounded continuous functions on S .

Sections 5 and 6 are devoted to a study of the space $D_E[0, \infty)$ with the Skorohod topology and Section 7 to weak convergence of sequences in $\mathcal{P}(D_E[0, \infty))$. In Section 8 we give necessary and sufficient conditions in terms of conditional expectations of $r^q(X_s(t+u), X_s(t)) \wedge 1$ (conditioning on $\mathcal{F}_t^{X_s}$) for a family of processes $\{X_s\}$ to be relatively compact (that is, for the family of distributions on $D_E[0, \infty)$ to be relatively compact). Criteria for relative compactness that are particularly useful in the study of Markov processes are given in Section 9. Finally, Section 10 contains necessary and sufficient conditions for a limiting process to have sample paths in $C_E[0, \infty)$.

1. THE PROHOROV METRIC

Throughout Sections 1–4, (S, d) is a metric space (d denoting the metric), $\mathcal{B}(S)$ is the σ -algebra of Borel subsets of S , and $\mathcal{P}(S)$ is the family of Borel probability measures on S . We topologize $\mathcal{P}(S)$ with the Prohorov metric

$$(1.1) \quad \rho(P, Q) = \inf \{\varepsilon > 0: P(F) \leq Q(F^\varepsilon) + \varepsilon \text{ for all } F \in \mathcal{C}\},$$

where \mathcal{C} is the collection of closed subsets of S and

$$(1.2) \quad F^\varepsilon = \left\{x \in S: \inf_{y \in F} d(x, y) < \varepsilon\right\}.$$

To see that ρ is a metric, we need the following lemma.

1.1 Lemma Let $P, Q \in \mathcal{P}(S)$ and $\alpha, \beta > 0$. If

$$(1.3) \quad P(F) \leq Q(F^\alpha) + \beta$$

for all $F \in \mathcal{C}$, then

$$(1.4) \quad Q(F) \leq P(F^\alpha) + \beta$$

for all $F \in \mathcal{C}$.

Proof. Given $F_1 \in \mathcal{C}$, let $F_2 = S - F_1^\alpha$, and note that $F_2 \in \mathcal{C}$ and $F_1 \subset S - F_2^\alpha$. Consequently, by (1.3) with $F = F_2$,

$$(1.5) \quad P(F_1^\alpha) = 1 - P(F_2) \geq 1 - Q(F_2^\alpha) - \beta \geq Q(F_1) - \beta,$$

implying (1.4) with $F = F_1$. \square

It follows immediately from Lemma 1.1 that $\rho(P, Q) = \rho(Q, P)$ for all $P, Q \in \mathcal{P}(S)$. Also, if $\rho(P, Q) = 0$, then $P(F) = Q(F)$ for all $F \in \mathcal{C}$ and hence for all $F \in \mathcal{B}(S)$; therefore, $\rho(P, Q) = 0$ if and only if $P = Q$. Finally, if $P, Q, R \in \mathcal{P}(S)$, $\rho(P, Q) < \delta$, and $\rho(Q, R) < \varepsilon$, then

$$(1.6) \quad \begin{aligned} P(F) &\leq Q(F^\delta) + \delta \leq Q(\bar{F}^\delta) + \delta \\ &\leq R((\bar{F}^\delta)^\varepsilon) + \delta + \varepsilon \leq R(F^{\delta+\varepsilon}) + \delta + \varepsilon \end{aligned}$$

for all $F \in \mathcal{C}$, so $\rho(P, R) \leq \delta + \varepsilon$, proving the triangle inequality.

The following theorem provides a probabilistic interpretation of the Prohorov metric when S is separable.

1.2 Theorem Let (S, d) be separable, and let $P, Q \in \mathcal{P}(S)$. Define $\mathcal{M}(P, Q)$ to be the set of all $\mu \in \mathcal{P}(S \times S)$ with marginals P and Q (i.e., $\mu(A \times S) = P(A)$ and $\mu(S \times A) = Q(A)$ for all $A \in \mathcal{B}(S)$). Then

$$(1.7) \quad \rho(P, Q) = \inf_{\mu \in \mathcal{M}(P, Q)} \inf \{\varepsilon > 0: \mu\{(x, y): d(x, y) \geq \varepsilon\} \leq \varepsilon\}.$$

Proof. If for some $\varepsilon > 0$ and $\mu \in \mathcal{M}(P, Q)$ we have

$$(1.8) \quad \mu\{(x, y): d(x, y) \geq \varepsilon\} \leq \varepsilon,$$

then

$$\begin{aligned} (1.9) \quad P(F) &= \mu(F \times S) \\ &\leq \mu((F \times S) \cap \{(x, y): d(x, y) < \varepsilon\}) + \varepsilon \\ &\leq \mu(S \times F^\varepsilon) + \varepsilon = Q(F^\varepsilon) + \varepsilon \end{aligned}$$

for all $F \in \mathcal{C}$, so $\rho(P, Q)$ is less than or equal to the right side of (1.7).

The reverse inequality is an immediate consequence of the following lemma. \square

1.3 Lemma Let S be separable. Let $P, Q \in \mathcal{P}(S)$, $\rho(P, Q) < \varepsilon$, and $\delta > 0$. Suppose that $E_1, \dots, E_N \in \mathcal{B}(S)$ are disjoint with diameters less than δ and that $P(E_0) \leq \delta$, where $E_0 = S - \bigcup_{i=1}^N E_i$. Then there exist constants $c_1, \dots, c_N \in [0, 1]$ and independent random variables X, Y_0, \dots, Y_N (S -valued) and ξ ($[0, 1]$ -valued) on some probability space (Ω, \mathcal{F}, v) such that X has distribution P , ξ is uniformly distributed on $[0, 1]$,

$$(1.10) \quad Y = \begin{cases} Y_i & \text{on } \{X \in E_i, \xi \geq c_i\}, \\ Y_0 & \text{on } \{X \in E_0\} \cup \bigcup_{i=1}^N \{X \in E_i, \xi < c_i\} \end{cases} \quad i = 1, \dots, N,$$

has distribution Q ,

$$(1.11) \quad \{d(X, Y) \geq \delta + \varepsilon\} \subset \{X \in E_0\} \cup \left\{ \xi < \max \left[\frac{\varepsilon}{P(E_i)} : P(E_i) > 0 \right] \right\},$$

and

$$(1.12) \quad v\{d(X, Y) \geq \delta + \varepsilon\} \leq \delta + \varepsilon.$$

The proof of this lemma depends on another lemma.

1.4 Lemma Let μ be a finite positive Borel measure on S , and let $p_i \geq 0$ and $A_i \in \mathcal{B}(S)$ for $i = 1, \dots, n$. Suppose that

$$(1.13) \quad \sum_{i \in I} p_i \leq \mu\left(\bigcup_{i \in I} A_i\right) \quad \text{for all } I \subset \{1, \dots, n\}.$$

Then there exist positive Borel measures $\lambda_1, \dots, \lambda_n$ on S such that $\lambda_i(A_i) = \lambda_i(S) = p_i$ for $i = 1, \dots, n$ and $\sum_{i=1}^n \lambda_i(A) \leq \mu(A)$ for all $A \in \mathcal{B}(S)$.

Proof. Note first that it involves no loss of generality to assume that each $p_i > 0$.

We proceed by induction on n . For $n = 1$, define λ_1 on $\mathcal{B}(S)$ by $\lambda_1(A) =$

$p_1\mu(A \cap A_1)/\mu(A_1)$. Then $\lambda_1(A_1) = \lambda_1(S) = p_1$, and since $p_1 \leq \mu(A_1)$ by (1.13), we have $\lambda_1(A) \leq \mu(A \cap A_1) \leq \mu(A)$ for all $A \in \mathcal{B}(S)$. Suppose now that the lemma holds with n replaced by m for $m = 1, \dots, n-1$ and that μ , p_i , and A_i ($1 \leq i \leq n$) satisfy (1.13). Define η on $\mathcal{B}(S)$ by $\eta(A) = \mu(A \cap A_n)/\mu(A_n)$, and let ε_0 be the largest ε such that

$$(1.14) \quad \sum_{i \in I} p_i \leq (\mu - \varepsilon\eta) \left(\bigcup_{i \in I} A_i \right) \text{ for all } I \subset \{1, \dots, n-1\}.$$

CASE 1. $\varepsilon_0 \geq p_n$. Let $\lambda_n = p_n\eta$ and put $\mu' = \mu - \lambda_n$. Since $p_n \leq \mu(A_n)$ by (1.13), μ' is a positive Borel measure on S , so by (1.14) (with $\varepsilon = p_n$) and the induction hypothesis, there exist positive Borel measures $\lambda_1, \dots, \lambda_{n-1}$ on S such that $\lambda_i(A_i) = \lambda_i(S) = p_i$ for $i = 1, \dots, n-1$ and $\sum_{i=1}^{n-1} \lambda_i(A) \leq \mu'(A)$ for all $A \in \mathcal{B}(S)$. Also, $\lambda_n(A_n) = \lambda_n(S) = p_n$, so $\lambda_1, \dots, \lambda_n$ have the required properties.

CASE 2. $\varepsilon_0 < p_n$. Put $\mu' = \mu - \varepsilon_0\eta$, and note that μ' is a positive Borel measure on S . By the definition of ε_0 , there exists $I_0 \subset \{1, \dots, n-1\}$ (nonempty) such that

$$(1.15) \quad \sum_{i \in I} p_i \leq \mu' \left(\bigcup_{i \in I} A_i \right) \text{ for all } I \subset I_0$$

with equality holding for $I = I_0$. By the induction hypothesis, there exist positive Borel measures λ_i on S , $i \in I_0$, such that $\lambda_i(A_i) = \lambda_i(S) = p_i$ for each $i \in I_0$ and $\sum_{i \in I_0} \lambda_i(A) \leq \mu'(A)$ for all $A \in \mathcal{B}(S)$. Let $p'_i = p_i$ for $i = 1, \dots, n-1$ and $p'_n = p_n - \varepsilon_0$. Put $B_0 = \bigcup_{i \in I_0} A_i$, define μ'' on $\mathcal{B}(S)$ by $\mu''(A) = \mu'(A) - \mu'(A \cap B_0)$, and let $I_1 = \{1, \dots, n\} - I_0$. Then, for all $I \subset I_1$,

$$\begin{aligned} (1.16) \quad \sum_{i \in I} p'_i + \mu'(B_0) &= \sum_{i \in I \cup I_0} p'_i \\ &\leq \mu' \left(\bigcup_{i \in I \cup I_0} A_i \right) \\ &= \mu' \left(\bigcup_{i \in I} A_i \right) + \mu'(B_0) - \mu' \left(\bigcup_{i \in I} A_i \cap B_0 \right) \\ &= \mu'' \left(\bigcup_{i \in I} A_i \right) + \mu'(B_0). \end{aligned}$$

Here, equality in the first line holds because equality in (1.15) holds for $I = I_0$, while the inequality in the second line follows from (1.14) if $n \notin I$ and from (1.13) if $n \in I$; more specifically, if $n \in I$, then

$$(1.17) \quad \sum_{i \in I \cup I_0} p'_i \leq \mu \left(\bigcup_{i \in I \cup I_0} A_i \right) - \varepsilon_0 = \mu' \left(\bigcup_{i \in I \cup I_0} A_i \right).$$

By (1.16),

$$(1.18) \quad \sum_{i \in I} p'_i \leq \mu''\left(\bigcup_{i \in I} A_i\right) \text{ for all } I \subset I_1,$$

so by the induction hypothesis, there exist positive Borel measures λ'_i on S , $i \in I_1$, such that $\lambda'_i(A_i) = \lambda'_i(S) = p'_i$ for each $i \in I_1$ and $\sum_{i \in I_1} \lambda'_i(A) \leq \mu''(A)$ for all $A \in \mathcal{B}(S)$. Finally, let $\lambda_i = \lambda'_i$ for $i \in I_1 - \{n\}$ and $\lambda_n = \lambda'_n + \varepsilon_0 n$. Then $\lambda_i(A_i) = \lambda_i(S) = p_i$ for each $i \in I_1$, hence for $i = 1, \dots, n$, and

$$\begin{aligned} (1.19) \quad \sum_{i=1}^n \lambda_i(A) &= \sum_{i \in I_0} \lambda_i(A) + \sum_{i \in I_1} \lambda_i(A) \\ &= \sum_{i \in I_0} \lambda_i(A \cap B_0) + \sum_{i \in I_1} \lambda'_i(A) + \varepsilon_0 \eta(A) \\ &\leq \mu'(A \cap B_0) + \mu''(A) + \varepsilon_0 \eta(A) \\ &= \mu'(A) + \varepsilon_0 \eta(A) \\ &= \mu(A) \end{aligned}$$

for all $A \in \mathcal{B}(S)$, so $\lambda_1, \dots, \lambda_n$ again have the required properties. \square

1.5 Corollary Let μ be a finite positive Borel measure on S , and let $p_i \geq 0$ and $A_i \in \mathcal{B}(S)$ for $i = 1, \dots, n$. Let $\varepsilon > 0$, and suppose that

$$(1.20) \quad \sum_{i \in I} p_i \leq \mu\left(\bigcup_{i \in I} A_i\right) + \varepsilon \text{ for all } I \subset \{1, \dots, n\}.$$

Then there exist positive Borel measures $\lambda_1, \dots, \lambda_n$ on S such that $\lambda_i(A_i) = \lambda_i(S) \leq p_i$ for $i = 1, \dots, n$, $\sum_{i=1}^n \lambda_i(S) \geq \sum_{i=1}^n p_i - \varepsilon$, and $\sum_{i=1}^n \lambda_i(A) \leq \mu(A)$ for all $A \in \mathcal{B}(S)$.

Proof. Let $S' = S \cup \{\Delta\}$, where Δ is an isolated point not belonging to S . Extend μ to a Borel measure on S' by defining $\mu(\{\Delta\}) = \varepsilon$. Letting $A'_i = A_i \cup \{\Delta\}$ for $i = 1, \dots, n$, we have

$$(1.21) \quad \sum_{i \in I} p_i \leq \mu\left(\bigcup_{i \in I} A'_i\right) \text{ for all } I \subset \{1, \dots, n\}.$$

By Lemma 1.4, there exist positive Borel measures $\lambda'_1, \dots, \lambda'_n$ on S' such that $\lambda'_i(A'_i) = \lambda'_i(S') = p_i$ for $i = 1, \dots, n$ and $\sum_{i=1}^n \lambda'_i(A) \leq \mu(A)$ for all $A \in \mathcal{B}(S')$. Let λ_i be the restriction of λ'_i to $\mathcal{B}(S)$ for $i = 1, \dots, n$. Then $\lambda_i(A_i) = \lambda'_i(A_i) \leq \lambda'_i(A'_i) = p_i$ and $\lambda_i(S - A_i) = \lambda'_i(S' - A'_i) = 0$ for $i = 1, \dots, n$. Also,

$$(1.22) \quad \sum_{i=1}^n \lambda_i(S) = \sum_{i=1}^n [p_i - \lambda'_i(\{\Delta\})] \geq \sum_{i=1}^n p_i - \mu(\{\Delta\}) = \sum_{i=1}^n p_i - \varepsilon$$

and $\sum_{i=1}^n \lambda_i(A) = \sum_{i=1}^n \lambda'_i(A) \leq \mu(A)$ for all $A \in \mathcal{B}(S)$. \square

Proof of Lemma 1.3 Let $P, Q, \varepsilon, \delta$, and E_0, \dots, E_N be as in the statement of the lemma. Let $p_i = P(E_i)$ and $A_i = E_i^c$ for $i = 1, \dots, N$. Then

$$(1.23) \quad \sum_{i \in I} p_i \leq P\left(\overline{\bigcup_{i \in I} E_i}\right) \leq Q\left(\bigcup_{i \in I} A_i\right) + \varepsilon \quad \text{for all } I \subset \{1, \dots, N\},$$

so by Corollary 1.5, there exist positive Borel measures $\lambda_1, \dots, \lambda_N$ on S such that $\lambda_i(A_i) = \lambda_i(S) \leq p_i$ for $i = 1, \dots, N$,

$$(1.24) \quad \sum_{i=1}^N \lambda_i(S) \geq \sum_{i=1}^N p_i - \varepsilon,$$

and $\sum_{i=1}^N \lambda_i(A) \leq Q(A)$ for all $A \in \mathcal{B}(S)$. Define $c_1, \dots, c_N \in [0, 1]$ by $c_i = (p_i - \lambda_i(S))/p_i$, where $0/0 = 0$, and note that $(1 - c_i)P(E_i) = \lambda_i(S)$ for $i = 1, \dots, N$ and $P(E_0) + \sum_{i=1}^N c_i P(E_i) = 1 - \sum_{i=1}^N \lambda_i(S)$. Consequently, there exist $Q_0, \dots, Q_N \in \mathcal{P}(S)$ such that

$$(1.25) \quad Q_i(B)(1 - c_i)P(E_i) = \lambda_i(B), \quad i = 1, \dots, N,$$

and

$$(1.26) \quad Q_0(B)\left(P(E_0) + \sum_{i=1}^N c_i P(E_i)\right) = Q(B) - \sum_{i=1}^N \lambda_i(B)$$

for all $B \in \mathcal{B}(S)$.

Let X, Y_0, \dots, Y_N , and ξ be independent random variables on some probability space $(\Omega, \mathcal{F}, \nu)$ with X, Y_0, \dots, Y_N having distributions P, Q_0, \dots, Q_N and ξ uniformly distributed on $[0, 1]$. We can assume that Y_1, \dots, Y_N take values in A_1, \dots, A_N , respectively. Defining Y by (1.10), we have by (1.25) and (1.26),

$$\begin{aligned} (1.27) \quad \nu\{Y \in B\} &= \sum_{i=1}^N Q_i(B)(1 - c_i)P(E_i) \\ &\quad + Q_0(B)\left(P(E_0) + \sum_{i=1}^N c_i P(E_i)\right) \\ &= Q(B) \end{aligned}$$

for all $B \in \mathcal{B}(S)$. Noting that $\{X \in E_i, \xi \geq c_i\} \subset \{X \in E_i, Y \in A_i\} \subset \{d(X, Y) < \delta + \varepsilon\}$ for $i = 1, \dots, N$, we have

$$\begin{aligned} (1.28) \quad \{d(X, Y) \geq \delta + \varepsilon\} &\subset \{X \in E_0\} \cup \bigcup_{i=1}^N \{X \in E_i, \xi < c_i\} \\ &\subset \{X \in E_0\} \cup \left\{ \xi < \bigvee_{i=1}^N c_i \right\} \\ &\subset \{X \in E_0\} \cup \left\{ \xi < \max \left[\frac{\varepsilon}{P(E_i)} : P(E_i) > 0 \right] \right\}. \end{aligned}$$

where the third containment follows from $p_i - \lambda_i(S) \leq \varepsilon$ for $i = 1, \dots, N$ (see (1.24)). Finally, by the first containment in (1.28) and by (1.24),

$$(1.29) \quad \begin{aligned} v\{d(X, Y) \geq \delta + \varepsilon\} &\leq P(E_0) + \sum_{i=1}^N c_i P(E_i) \\ &= P(E_0) + \sum_{i=1}^N (p_i - \lambda_i(S)) \\ &\leq \delta + \varepsilon. \end{aligned}$$

□

1.6 Corollary Let (S, d) be separable. Suppose that X_n , $n = 1, 2, \dots$, and X are S -valued random variables defined on the same probability space with distributions P_n , $n = 1, 2, \dots$, and P , respectively. If $d(X_n, X) \rightarrow 0$ in probability as $n \rightarrow \infty$, then $\lim_{n \rightarrow \infty} \rho(P_n, P) = 0$.

Proof. For $n = 1, 2, \dots$, let μ_n be the joint distribution of X_n and X . Then $\lim_{n \rightarrow \infty} \mu_n\{(x, y) : d(x, y) \geq \varepsilon\} = 0$ for every $\varepsilon > 0$, so the result follows from Theorem 1.2. □

The next result shows that the metric space $(\mathcal{P}(S), \rho)$ is complete and separable whenever (S, d) is. We note that while separability is a topological property, completeness is a property of the metric.

1.7 Theorem If S is separable, then $\mathcal{P}(S)$ is separable. If in addition (S, d) is complete, then $(\mathcal{P}(S), \rho)$ is complete.

Proof. Let $\{x_n\}$ be a countable dense subset of S , and let δ_x denote the element of $\mathcal{P}(S)$ with unit mass at $x \in S$. We leave it to the reader to show that the probability measures of the form $\sum_{i=1}^N a_i \delta_{x_i}$ with N finite, a_i rational, and $\sum_{i=1}^N a_i = 1$, comprise a dense subset of $\mathcal{P}(S)$ (Problem 3).

To prove completeness it is enough to consider sequences $\{P_n\} \subset \mathcal{P}(S)$ with $\rho(P_{n-1}, P_n) < 2^{-n}$ for each $n \geq 2$. For $n = 2, 3, \dots$, choose $E_1^{(n)}, \dots, E_{N_n}^{(n)} \in \mathcal{B}(S)$ disjoint with diameters less than 2^{-n} and with $P_{n-1}(E_0^{(n)}) \leq 2^{-n}$, where $E_0^{(n)} = S - \bigcup_{i=1}^{N_n} E_i^{(n)}$. By Lemma 1.3, there exists a probability space (Ω, \mathcal{F}, v) on which are defined S -valued random variables $Y_0^{(n)}, \dots, Y_{N_n}^{(n)}$, $n = 2, 3, \dots, [0, 1]$ -valued random variables $\xi^{(n)}$, $n = 2, 3, \dots$, and an S -valued random variable X_1 with distribution P_1 , all of which are independent, such that if the constants $c_1^{(n)}, \dots, c_{N_n}^{(n)} \in [0, 1]$, $n = 2, 3, \dots$, are appropriately chosen, then the random variable

$$(1.30) \quad X_n = \begin{cases} Y_i^{(n)} & \text{on } \{X_{n-1} \in E_i^{(n)}, \xi^{(n)} \geq c_i^{(n)}\}, \quad i = 1, \dots, N_n, \\ Y_0^{(n)} & \text{on } \{X_{n-1} \in E_0^{(n)}\} \cup \bigcup_{i=1}^{N_n} \{X_{n-1} \in E_i^{(n)}, \xi^{(n)} < c_i^{(n)}\} \end{cases}$$

has distribution P_n and

$$(1.31) \quad v\{d(X_{n-1}, X_n) \geq 2^{-n+1}\} \leq 2^{-n+1},$$

successively for $n = 2, 3, \dots$. By the Borel-Cantelli lemma,

$$(1.32) \quad v\left\{\sum_{n=2}^{\infty} d(X_{n-1}, X_n) < \infty\right\} = 1,$$

so by the completeness of (S, d) , $\lim_{n \rightarrow \infty} X_n \equiv X$ exists a.s. Letting P be the distribution of X , Corollary 1.6 implies that $\lim_{n \rightarrow \infty} \rho(P_n, P) = 0$. \square

As a further application of Lemma 1.3, we derive the so-called Skorohod representation.

1.8 Theorem Let (S, d) be separable. Suppose P_n , $n = 1, 2, \dots$, and P in $\mathcal{P}(S)$ satisfy $\lim_{n \rightarrow \infty} \rho(P_n, P) = 0$. Then there exists a probability space (Ω, \mathcal{F}, v) on which are defined S -valued random variables X_n , $n = 1, 2, \dots$, and X with distributions P_n , $n = 1, 2, \dots$, and P , respectively, such that $\lim_{n \rightarrow \infty} X_n = X$ a.s.

Proof. For $k = 1, 2, \dots$, choose $E_1^{(k)}, \dots, E_{N_k}^{(k)} \in \mathcal{B}(S)$ disjoint with diameters less than 2^{-k} and with $P(E_0^{(k)}) \leq 2^{-k}$, where $E_0^{(k)} = S - \bigcup_{i=1}^{N_k} E_i^{(k)}$, and assume (without loss of generality) that $\varepsilon_k \equiv \min_{1 \leq i \leq N_k} P(E_i^{(k)}) > 0$. Define the sequence $\{k_n\}$ by

$$(1.33) \quad k_n = \max\{1\} \cup \left\{k \geq 1 : \rho(P_n, P) < \frac{\varepsilon_k}{k}\right\},$$

and apply Lemma 1.3 with $Q = P_n$, $\varepsilon = \varepsilon_{k_n}/k_n$ if $k_n > 1$ and $\varepsilon = \rho(P_n, P) + 1/n$ if $k_n = 1$, $\delta = 2^{-k_n}$, $E_i = E_i^{(k_n)}$, and $N = N_{k_n}$ for $n = 1, 2, \dots$. We conclude that there exists a probability space (Ω, \mathcal{F}, v) on which are defined S -valued random variables $Y_0^{(n)}, \dots, Y_{N_n}^{(n)}$, $n = 1, 2, \dots$, a random variable ξ uniformly distributed on $[0, 1]$, and an S -valued random variable X with distribution P , all of which are independent, such that if the constants $c_1^{(n)}, \dots, c_{N_n}^{(n)} \in [0, 1]$, $n = 1, 2, \dots$, are appropriately chosen, then the random variable

$$(1.34) \quad X_n = \begin{cases} Y_i^{(n)} & \text{on } \{X \in E_i^{(k_n)}, \xi \geq c_i^{(n)}\}, \quad i = 1, \dots, N_{k_n}, \\ Y_0^{(n)} & \text{on } \{X \in E_0^{(k_n)}\} \cup \bigcup_{i=1}^{N_k} \{X \in E_i^{(k_n)}, \xi < c_i^{(n)}\} \end{cases}$$

has distribution P_n and

$$(1.35) \quad \begin{aligned} & \left\{d(X_n, X) \geq 2^{-k_n} + \frac{\varepsilon_{k_n}}{k_n}\right\} \\ & \subset \{X \in E_0^{(k_n)}\} \cup \left\{\xi < \frac{1}{k_n}\right\} \quad \text{if } k_n > 1 \end{aligned}$$

for $n = 1, 2, \dots$. If $K_n \equiv \min_{m \geq n} k_m > 1$, then

$$(1.36) \quad v\left(\bigcup_{m=n}^{\infty} \left\{d(X_m, X) \geq 2^{-k_m} + \frac{\varepsilon_{k_m}}{k_m}\right\}\right)$$

$$\leq \sum_{k=K_n}^{\infty} v\{X \in E_0^{(k)}\} + v\left\{\xi < \frac{1}{K_n}\right\}$$

$$\leq 2^{-K_n+1} + \frac{1}{K_n},$$

and since $\lim_{n \rightarrow \infty} K_n = \infty$, we have $\lim_{n \rightarrow \infty} X_n = X$ a.s. \square

We conclude this section by proving the continuous mapping theorem.

1.9 Corollary Let (S, d) and (S', d') be separable metric spaces, and let $h: S \rightarrow S'$ be Borel measurable. Suppose that P_n , $n = 1, 2, \dots$, and P in $\mathcal{P}(S)$ satisfy $\lim_{n \rightarrow \infty} \rho(P_n, P) = 0$, and define Q_n , $n = 1, 2, \dots$, and Q in $\mathcal{P}(S')$ by

$$(1.37) \quad Q_n = P_n h^{-1}, \quad Q = Ph^{-1}.$$

(By definition, $Ph^{-1}(B) = P\{s \in S : h(s) \in B\}$.) Let C_h be the set of points of S at which h is continuous. If $P(C_h) = 1$, then $\lim_{n \rightarrow \infty} \rho'(Q_n, Q) = 0$, where ρ' is the Prohorov metric on $\mathcal{P}(S')$.

Proof. By Theorem 1.8, there exists a probability space (Ω, \mathcal{F}, v) on which are defined S -valued random variables X_n , $n = 1, 2, \dots$, and X with distributions P_n , $n = 1, 2, \dots$, and P , respectively, such that $\lim_{n \rightarrow \infty} X_n = X$ a.s. Since $v\{X \in C_h\} = 1$, we have $\lim_{n \rightarrow \infty} h(X_n) = h(X)$ a.s., and by Corollary 1.6, this implies that $\lim_{n \rightarrow \infty} \rho'(Q_n, Q) = 0$. \square

2. PROHOROV'S THEOREM

We are primarily interested in the convergence of sequences of Borel probability measures on the metric space (S, d) . A common approach for verifying the convergence of a sequence $\{x_n\}$ in a metric space is to first show that $\{x_n\}$ is contained in some compact set and then to show that every convergent subsequence of $\{x_n\}$ must converge to the same element x . This then implies that $\lim_{n \rightarrow \infty} x_n = x$. We use this argument repeatedly in what follows, and, consequently, a characterization of the compact subsets of $\mathcal{P}(S)$ is crucial. This characterization is given by the theorem of Prohorov that relates compactness to the notion of tightness.

A probability measure $P \in \mathcal{P}(S)$ is said to be *tight* if for each $\varepsilon > 0$ there exists a compact set $K \subset S$ such that $P(K) \geq 1 - \varepsilon$. A family of probability measures $\mathcal{M} \subset \mathcal{P}(S)$ is *tight* if for each $\varepsilon > 0$ there exists a compact set $K \subset S$

such that

$$(2.1) \quad \inf_{P \in \mathcal{M}} P(K) \geq 1 - \varepsilon.$$

2.1 Lemma If (S, d) is complete and separable, then each $P \in \mathcal{P}(S)$ is tight.

Proof. Let $\{x_k\}$ be dense in S , and let $P \in \mathcal{P}(S)$. Given $\varepsilon > 0$, choose positive integers N_1, N_2, \dots such that

$$(2.2) \quad P\left(\bigcup_{k=1}^{N_n} B(x_k, \frac{1}{n})\right) \geq 1 - \frac{\varepsilon}{2^n}$$

for $n = 1, 2, \dots$. Let K be the closure of $\bigcap_{n \geq 1} \bigcup_{k=1}^{N_n} B(x_k, 1/n)$. Then K is totally bounded and hence compact, and

$$(2.3) \quad P(K) \geq 1 - \sum_{n=1}^{\infty} \frac{\varepsilon}{2^n} = 1 - \varepsilon. \quad \square$$

2.2 Theorem Let (S, d) be complete and separable, and let $\mathcal{M} \subset \mathcal{P}(S)$. Then the following are equivalent:

(a) \mathcal{M} is tight.

(b) For each $\varepsilon > 0$, there exists a compact set $K \subset S$ such that

$$(2.4) \quad \inf_{P \in \mathcal{M}} P(K^\varepsilon) \geq 1 - \varepsilon,$$

where K^ε is as in (1.2).

(c) \mathcal{M} is relatively compact.

Proof. (a \Rightarrow b) Immediate.

(b \Rightarrow c) Since the closure of \mathcal{M} is complete by Theorem 1.7, it is sufficient to show that \mathcal{M} is totally bounded. That is, given $\delta > 0$, we must construct a finite set $\mathcal{N} \subset \mathcal{P}(S)$ such that $\mathcal{M} \subset \{Q : \rho(P, Q) < \delta \text{ for some } P \in \mathcal{N}\}$.

Let $0 < \varepsilon < \delta/2$ and choose a compact set $K \subset S$ such that (2.4) holds. By the compactness of K , there exists a finite set $\{x_1, \dots, x_n\} \subset K$ such that $K^\varepsilon \subset \bigcup_{i=1}^n B_i$, where $B_i = B(x_i, 2\varepsilon)$. Fix $x_0 \in S$ and $m \geq n/\varepsilon$, and let \mathcal{N} be the collection of probability measures of the form

$$(2.5) \quad P = \sum_{i=0}^m \left(\frac{k_i}{m} \right) \delta_{x_i},$$

where $0 \leq k_i \leq m$ and $\sum_{i=0}^m k_i = m$.

Given $Q \in \mathcal{M}$, let $k_i = [mQ(E_i)]$ for $i = 1, \dots, n$, where $E_i = B_i$

$- \bigcup_{j=1}^{l-1} B_j$, and let $k_0 = m - \sum_{i=1}^n k_i$. Then, defining P by (2.5), we have

$$\begin{aligned} (2.6) \quad Q(F) &\leq Q\left(\bigcup_{F \cap E_i \neq \emptyset} E_i\right) + \varepsilon \\ &\leq \sum_{F \cap E_i \neq \emptyset} \frac{[mQ(E_i)]}{m} + \frac{n}{m} + \varepsilon \\ &\leq P(F^{2^n}) + 2\varepsilon \end{aligned}$$

for all closed sets $F \subset S$, so $\rho(P, Q) \leq 2\varepsilon < \delta$.

(c \Rightarrow a) Let $\varepsilon > 0$. Since \mathcal{M} is totally bounded, there exists for $n = 1, 2, \dots$ a finite subset $\mathcal{N}_n \subset \mathcal{M}$ such that $\mathcal{M} \subset \{Q : \rho(P, Q) < \varepsilon/2^{n+1} \text{ for some } P \in \mathcal{N}_n\}$. Lemma 2.1 implies that for $n = 1, 2, \dots$ we can choose a compact set $K_n \subset S$ such that $P(K_n) \geq 1 - \varepsilon/2^{n+1}$ for all $P \in \mathcal{N}_n$. Given $Q \in \mathcal{M}$, it follows that for $n = 1, 2, \dots$ there exists $P_n \in \mathcal{N}_n$ such that

$$(2.7) \quad Q(K_n^{1/2^{n+1}}) \geq P_n(K_n) - \varepsilon/2^{n+1} \geq 1 - \varepsilon/2^n.$$

Letting K be the closure of $\bigcap_{n \geq 1} K_n^{1/2^{n+1}}$, we conclude that K is compact and

$$(2.8) \quad Q(K) \geq 1 - \sum_{n=1}^{\infty} \frac{\varepsilon}{2^n} = 1 - \varepsilon. \quad \square$$

2.3 Corollary Let (S, d) be arbitrary, and let $\mathcal{M} \subset \mathcal{P}(S)$. If \mathcal{M} is tight, then \mathcal{M} is relatively compact.

Proof. For each $m \geq 1$ there exists a compact set $K_m \subset S$ such that

$$(2.9) \quad \inf_{P \in \mathcal{M}} P(K_m) \geq 1 - \frac{1}{m},$$

and we can assume that $K_1 \subset K_2 \subset \dots$. For every $P \in \mathcal{M}$ and $m \geq 1$, define $P^{(m)} \in \mathcal{P}(S)$ by $P^{(m)}(A) = P(A \cap K_m)/P(K_m)$, and note that $P^{(m)}$ may be regarded as belonging to $\mathcal{P}(K_m)$. Since compact metric spaces are complete and separable, $\mathcal{M}^{(m)} = \{P^{(m)} : P \in \mathcal{M}\}$ is relatively compact in $\mathcal{P}(K_m)$ for each $m \geq 1$ by Theorem 2.2.

We also have

$$(2.10) \quad P(A) \leq P(A \cap K_m) + \frac{1}{m} \leq P^{(m)}(A) + \frac{1}{m},$$

$$(2.11) \quad P^{(m')}(A) \leq \frac{(P(A \cap K_m) + 1/m)}{P(K_m)} \leq P^{(m)}(A) + \frac{2}{m}, \quad m' > m,$$

$$(2.12) \quad P(A) \geq P(K_m)P^{(m)}(A) \geq \left(1 - \frac{1}{m}\right)P^{(m)}(A),$$

$$(2.13) \quad P^{(m')}(A) \geq \frac{P(K_m)P^{(m)}(A)}{P(K_m)} \geq \left(1 - \frac{1}{m}\right)P^{(m)}(A), \quad m' > m,$$

for all $P \in \mathcal{M}$, $A \in \mathcal{B}(S)$, and $m \geq 1$. By (2.10),

$$(2.14) \quad \rho(P, P^{(m)}) \leq \frac{1}{m}$$

for all $P \in \mathcal{M}$ and $m \geq 1$. Given $A_1, A_2, \dots \in \mathcal{B}(S)$ disjoint, (2.13) and (2.11) imply that

$$\begin{aligned} (2.15) \quad \sum_i |P^{(m')}(A_i) - P^{(m)}(A_i)| \\ &\leq \sum_i \left(P^{(m')}(A_i) - \left(1 - \frac{1}{m}\right)P^{(m)}(A_i) + \frac{1}{m}P^{(m)}(A_i) \right) \\ &\leq P^{(m')}\left(\bigcup_i A_i\right) - P^{(m)}\left(\bigcup_i A_i\right) + \frac{2}{m} \\ &\leq \frac{2}{m} + \frac{2}{m} = \frac{4}{m} \end{aligned}$$

for every $P \in \mathcal{M}$ and $m' > m \geq 1$.

Let $\{P_n\} \subset \mathcal{M}$. By the relative compactness of $\mathcal{M}^{(m)}$ in $\mathcal{P}(K_m)$, there exists (through a diagonalization argument) a subsequence $\{P_{n_k}\} \subset \{P_n\}$ and $Q^{(m)} \in \mathcal{P}(K_m)$ such that

$$(2.16) \quad \lim_{k \rightarrow \infty} \rho(P_{n_k}^{(m)}, Q^{(m)}) = 0$$

for every $m \geq 1$. It follows that

$$(2.17) \quad Q^{(m)}(F) = \lim_{\epsilon \rightarrow 0} \overline{\lim}_{k \rightarrow \infty} P_{n_k}^{(m)}(F^\epsilon)$$

for all closed sets $F \subset S$ and $m \geq 1$, and therefore the inequalities (2.11) and (2.13) are preserved for the $Q^{(m)}$ for all closed sets $A \subset S$, hence for all $A \in \mathcal{B}(S)$ (using the regularity of the $Q^{(m)}$). Consequently, we have (2.15) for the $Q^{(m)}$, so

$$(2.18) \quad Q(A) = \lim_{m \rightarrow \infty} Q^{(m)}(A)$$

exists for every $A \in \mathcal{B}(S)$ and defines a probability measure $Q \in \mathcal{P}(S)$ (the countable additivity following from (2.15)). But

$$(2.19) \quad \begin{aligned} \rho(P_{nk}, Q) &\leq \rho(P_{nk}, P_{nk}^{(m)}) + \rho(P_{nk}^{(m)}, Q^{(m)}) + \rho(Q^{(m)}, Q) \\ &\leq \frac{1}{m} + \rho(P_{nk}^{(m)}, Q^{(m)}) + \frac{2}{m}, \end{aligned}$$

for each k and $m \geq 1$, implying that $\lim_{k \rightarrow \infty} \rho(P_{nk}, Q) = 0$. \square

We conclude this section with a result concerning countably infinite product spaces.

2.4 Proposition Let (S_k, d_k) , $k = 1, 2, \dots$, be metric spaces, and define the metric space (S, d) by letting $S = \prod_{k=1}^{\infty} S_k$ and $d(x, y) = \sum_{k=1}^{\infty} 2^{-k} (d_k(x_k, y_k) \wedge 1)$ for all $x, y \in S$. Let $\{P_{\alpha}\} \subset \mathcal{P}(S)$ (where α ranges over some index set), and for $k = 1, 2, \dots$ and each α , define $P_{\alpha}^k \in \mathcal{P}(S_k)$ to be the k th marginal distribution of P_{α} (i.e., $P_{\alpha}^k = P_{\alpha} \pi_k^{-1}$, where the projection $\pi_k: S \rightarrow S_k$ is given by $\pi_k(x) = x_k$). Then $\{P_{\alpha}\}$ is tight if and only if $\{P_{\alpha}^k\}$ is tight for $k = 1, 2, \dots$.

Proof. Suppose that $\{P_{\alpha}^k\}$ is tight for $k = 1, 2, \dots$, and let $\varepsilon > 0$. For $k = 1, 2, \dots$, choose a compact set $K_k \subset S_k$ such that $\inf_{\alpha} P_{\alpha}^k(K_k) \geq 1 - \varepsilon/2^k$. Then $K \equiv \prod_{k=1}^{\infty} K_k = \bigcap_{k=1}^{\infty} \pi_k^{-1}(K_k)$ is compact in S , and

$$(2.20) \quad P_{\alpha}(K) \geq 1 - \sum_{k=1}^{\infty} (1 - P_{\alpha}^k(K_k)) \geq 1 - \varepsilon$$

for all α . Consequently, $\{P_{\alpha}\}$ is tight.

The converse follows by observing that for each compact set $K \subset S$, $\pi_k(K)$ is compact in S_k and

$$(2.21) \quad \inf_{\alpha} P_{\alpha}^k(\pi_k(K)) \geq \inf_{\alpha} P_{\alpha}(K)$$

for $k = 1, 2, \dots$. \square

3. WEAK CONVERGENCE

Let $\bar{C}(S)$ be the space of real-valued bounded continuous functions on the metric space (S, d) with norm $\|f\| = \sup_{x \in S} |f(x)|$. A sequence $\{P_n\} \subset \mathcal{P}(S)$ is said to converge weakly to $P \in \mathcal{P}(S)$ if

$$(3.1) \quad \lim_{n \rightarrow \infty} \int f dP_n = \int f dP, \quad f \in \bar{C}(S).$$

The distribution of an S -valued random variable X , denoted by PX^{-1} , is the element of $\mathcal{P}(S)$ given by $PX^{-1}(B) = P\{X \in B\}$. A sequence $\{X_n\}$ of S -valued

random variables is said to *converge in distribution* to the S -valued random variable X if $\{PX_n^{-1}\}$ converges weakly to PX^{-1} , or equivalently, if

$$(3.2) \quad \lim_{n \rightarrow \infty} E[f(X_n)] = E[f(X)], \quad f \in \bar{C}(S).$$

Weak convergence is denoted by $P_n \Rightarrow P$ and convergence in distribution by $X_n \Rightarrow X$. When it is useful to emphasize which metric space is involved, we write " $P_n \Rightarrow P$ on S " or " $X_n \Rightarrow X$ in S ".

If S' is a second metric space and $f: S \rightarrow S'$ is continuous, we note that then $X_n \Rightarrow X$ in S implies $f(X_n) \Rightarrow f(X)$ in S' since $g \in \bar{C}(S')$ implies $g \circ f \in \bar{C}(S)$. For example, if $S = C[0, 1]$ and $S' = \mathbb{R}$, then $f(x) \equiv \sup_{0 \leq t \leq 1} x(t)$ is continuous, so $X_n \Rightarrow X$ in $C[0, 1]$ implies $\sup_{0 \leq t \leq 1} X_n(t) \Rightarrow \sup_{0 \leq t \leq 1} X(t)$ in \mathbb{R} . Recall that, if $S = \mathbb{R}$, then (3.2) is equivalent to

$$(3.3) \quad \lim_{n \rightarrow \infty} P\{X_n \leq x\} = P\{X \leq x\}$$

for all x at which the right side of (3.3) is continuous.

We now show that weak convergence is equivalent to convergence in the Prohorov metric. The *boundary* of a subset $A \subset S$ is given by $\partial A = \bar{A} \cap \overline{A^c}$ (\bar{A} and A^c denote the closure and complement of A , respectively). A is said to be a *P -continuity set* if $A \in \mathcal{B}(S)$ and $P(\partial A) = 0$.

3.1 Theorem Let (S, d) be arbitrary, and let $\{P_n\} \subset \mathcal{P}(S)$ and $P \in \mathcal{P}(S)$. Of the following conditions, (b) through (f) are equivalent and are implied by (a). If S is separable, then all six conditions are equivalent:

- (a) $\lim_{n \rightarrow \infty} \rho(P_n, P) = 0$.
- (b) $P_n \Rightarrow P$.
- (c) $\lim_{n \rightarrow \infty} \int f dP_n = \int f dP$ for all uniformly continuous $f \in \bar{C}(S)$.
- (d) $\overline{\lim}_{n \rightarrow \infty} P_n(F) \leq P(F)$ for all closed sets $F \subset S$.
- (e) $\underline{\lim}_{n \rightarrow \infty} P_n(G) \geq P(G)$ for all open sets $G \subset S$.
- (f) $\lim_{n \rightarrow \infty} P_n(A) = P(A)$ for all P -continuity sets $A \subset S$.

Proof. (a \Rightarrow b) For each n , let $\varepsilon_n = \rho(P_n, P) + 1/n$. Given $f \in \bar{C}(S)$ with $f \geq 0$,

$$(3.4) \quad \int f dP_n = \int_0^{\|f\|} P_n\{f \geq t\} dt \leq \int_0^{\|f\|} P(\{f \geq t\}^{\varepsilon_n}) dt + \varepsilon_n \|f\|$$

for every n , so

$$(3.5) \quad \begin{aligned} \overline{\lim}_{n \rightarrow \infty} \int f dP_n &\leq \lim_{n \rightarrow \infty} \int_0^{\|f\|} P(\{f \geq t\}^{\varepsilon_n}) dt \\ &= \int_0^{\|f\|} P\{f \geq t\} dt = \int f dP. \end{aligned}$$

Consequently,

$$(3.6) \quad \begin{aligned} \overline{\lim}_{n \rightarrow \infty} \int (\|f\| + f) dP_n &\leq \int (\|f\| + f) dP, \\ \overline{\lim}_{n \rightarrow \infty} \int (\|f\| - f) dP_n &\leq \int (\|f\| - f) dP \end{aligned}$$

for all $f \in \bar{C}(S)$, and this implies (3.1).

(b \Rightarrow c) Immediate.

(c \Rightarrow d) Let $F \subset S$ be closed. For each $\varepsilon > 0$, define $f_\varepsilon \in \bar{C}(S)$ by

$$(3.7) \quad f_\varepsilon(x) = \left(1 - \frac{d(x, F)}{\varepsilon}\right) \vee 0,$$

where $d(x, F) = \inf_{y \in F} d(x, y)$. Then f is uniformly continuous, so

$$(3.8) \quad \overline{\lim}_{n \rightarrow \infty} P_n(F) \leq \lim_{n \rightarrow \infty} \int f_\varepsilon dP_n = \int f_\varepsilon dP,$$

for each $\varepsilon > 0$, and therefore

$$(3.9) \quad \overline{\lim}_{n \rightarrow \infty} P_n(F) \leq \lim_{\varepsilon \rightarrow 0} \int f_\varepsilon dP = P(F).$$

(d \Rightarrow e) For every open set $G \subset S$,

$$(3.10) \quad \lim_{n \rightarrow \infty} P_n(G) = 1 - \overline{\lim}_{n \rightarrow \infty} P_n(G^c) \geq 1 - P(G^c) = P(G).$$

(e \Rightarrow f) Let A be a P -continuity set in S , and let A° denote its interior ($A^\circ = A - \partial A$). Then

$$(3.11) \quad \overline{\lim}_{n \rightarrow \infty} P_n(A) \leq \overline{\lim}_{n \rightarrow \infty} P_n(\bar{A}) = 1 - \overline{\lim}_{n \rightarrow \infty} P_n(\bar{A}^c) \leq 1 - P(\bar{A}^c) = P(A)$$

and

$$(3.12) \quad \underline{\lim}_{n \rightarrow \infty} P_n(A) \geq \underline{\lim}_{n \rightarrow \infty} P_n(A^\circ) \geq P(A^\circ) = P(A).$$

(f \Rightarrow b) Let $f \in \bar{C}(S)$ with $f \geq 0$. Then $\partial\{f \geq t\} \subset \{f = t\}$, so $\{f \geq t\}$ is a P -continuity set for all but at most countably many $t \geq 0$. Therefore,

$$(3.13) \quad \begin{aligned} \lim_{n \rightarrow \infty} \int f dP_n &= \lim_{n \rightarrow \infty} \int_0^{\|f\|} P_n\{f \geq t\} dt \\ &= \int_0^{\|f\|} P\{f \geq t\} dt = \int f dP \end{aligned}$$

for all nonnegative $f \in \bar{C}(S)$, which clearly implies (3.1).

($\epsilon \Rightarrow a$, assuming separability) Let $\epsilon > 0$ be arbitrary, and let $E_1, E_2, \dots \in \mathcal{B}(S)$ be a partition of S with diameter(E_i) $< \epsilon/2$ for $i = 1, 2, \dots$. Let N be the smallest positive integer n such that $P(\bigcup_{i=1}^n E_i) > 1 - \epsilon/2$, and let \mathcal{G} be the (finite) collection of open sets of the form $(\bigcup_{i \in I} E_i)^{\epsilon/2}$, where $I \subset \{1, \dots, N\}$. Since \mathcal{G} is finite, there exists n_0 such that $P(G) \leq P_n(G) + \epsilon/2$ for all $G \in \mathcal{G}$ and $n \geq n_0$. Given $F \in \mathcal{G}$, let

$$(3.14) \quad F_0 = \bigcup \{E_i : 1 \leq i \leq N, E_i \cap F \neq \emptyset\}.$$

Then $F_0^{\epsilon/2} \in \mathcal{G}$ and

$$(3.15) \quad \begin{aligned} P(F) &\leq P(F_0^{\epsilon/2}) + \epsilon/2 \\ &\leq P_n(F_0^{\epsilon/2}) + \epsilon \leq P_n(F^\epsilon) + \epsilon \end{aligned}$$

for all $n \geq n_0$. Hence $P_n \Rightarrow P$ for each $n \geq n_0$. \square

3.2 Corollary Let P_n , $n = 1, 2, \dots$, and P belong to $\mathcal{P}(S)$, and let $S' \in \mathcal{B}(S)$. For $n = 1, 2, \dots$, suppose that $P_n(S') = P(S') = 1$, and let P'_n and P' be the restrictions of P_n and P to $\mathcal{B}(S')$ (of course, S' has the relative topology). Then $P_n \Rightarrow P$ on S if and only if $P'_n \Rightarrow P'$ on S' .

Proof. If G' is open in S' , then $G' = G \cap S'$ for some open set $G \subset S$. Therefore, if $P_n \Rightarrow P$ on S ,

$$(3.16) \quad \lim_{n \rightarrow \infty} P'_n(G') = \lim_{n \rightarrow \infty} P_n(G) \geq P(G) = P'(G'),$$

so $P'_n \Rightarrow P'$ on S' by Theorem 3.1. The converse is proved similarly. \square

3.3 Corollary Let (S, d) be arbitrary, and let (X_n, Y_n) , $n = 1, 2, \dots$, and X be $(S \times S)$ - and S -valued random variables. If $X_n \Rightarrow X$ and $d(X_n, Y_n) \rightarrow 0$ in probability, then $Y_n \Rightarrow X$.

3.4 Remark If S is separable, then $\mathcal{B}(S \times S) = \mathcal{B}(S) \times \mathcal{B}(S)$, and hence (X, Y) is an $(S \times S)$ -valued random variable whenever X and Y are S -valued random variables defined on the same probability space. This observation has already been used implicitly in Section 1, and we use it again (without mention) in later sections of this chapter. \square

Proof. If $f \in C(S)$ is uniformly continuous, then

$$(3.17) \quad \lim_{n \rightarrow \infty} E[f(X_n) - f(Y_n)] = 0.$$

Consequently,

$$(3.18) \quad \lim_{n \rightarrow \infty} E[f(Y_n)] = \lim_{n \rightarrow \infty} E[f(X_n)] = E[f(X)],$$

and Theorem 3.1 is again applicable. \square

4. SEPARATING AND CONVERGENCE DETERMINING SETS

Let (S, d) be a metric space. A sequence $\{f_n\} \subset B(S)$ is said to converge *boundedly and pointwise* to $f \in B(S)$ if $\sup_n \|f_n\| < \infty$ (where $\|\cdot\|$ denotes the sup norm) and $\lim_{n \rightarrow \infty} f_n(x) = f(x)$ for every $x \in S$; we denote this by

$$(4.1) \quad \text{bp-lim}_{n \rightarrow \infty} f_n = f.$$

A set $M \subset B(S)$ is called *bp-closed* if whenever $\{f_n\} \subset M$, $f \in B(S)$, and (4.1) holds, we have $f \in M$. The *bp-closure* of $M \subset B(S)$ is the smallest bp-closed subset of $B(S)$ that contains M . Finally, if the bp-closure of $M \subset B(S)$ is equal to $B(S)$, we say that M is *bp-dense* in $B(S)$. We remark that if M is bp-dense in $B(S)$ and $f \in B(S)$, there need not exist a sequence $\{f_n\} \subset M$ such that (4.1) holds.

4.1 Lemma If $M \subset B(S)$ is a subspace, then the bp-closure of M is also a subspace.

Proof. Let H be the bp-closure of M . For each $f \in H$, define

$$(4.2) \quad H_f = \{g \in H : af + bg \in H \text{ for all } a, b \in \mathbb{R}\},$$

and note that H_f is bp-closed because H is. If $f \in M$, then $H_f \supset M$, so $H_f = H$. If $f \in H$, then $f \in H_g$ for every $g \in M$, hence $g \in H_f$ for every $g \in M$, and therefore $H_f = H$. \square

4.2 Proposition Let (S, d) be arbitrary. Then $\bar{C}(S)$ is bp-dense in $B(S)$. If S is separable, then there exists a sequence $\{f_n\}$ of nonnegative functions in $\bar{C}(S)$ such that $\text{span } \{f_n\}$ is bp-dense in $B(S)$.

Proof. Let H be the bp-closure of $\bar{C}(S)$. H is closed under monotone convergence of uniformly bounded sequences, H is a subspace of $B(S)$ by Lemma 4.1, and $\chi_G \in H$ for every open set $G \subset S$. By the Dynkin class theorem for functions (Theorem 4.3 of the Appendixes), $H = B(S)$.

If S is separable, let $\{x_i\}$ be dense in S . For every open set $G \subset S$ that is a finite intersection of $B(x_i, 1/k)$, $i, k \geq 1$, choose a sequence $\{f_n^G\}$ of nonnegative functions in $\bar{C}(S)$ such that $\text{bp-lim}_{n \rightarrow \infty} f_n^G = \chi_G$. The Dynkin class theorem for functions now applies to $\text{span } \{f_n^G : n, G \text{ as above}\}$. \square

For future reference, we extend two of the definitions given at the beginning of this section. A set $M \subset B(S) \times B(S)$ is called *bp-closed* if whenever $\{(f_n, g_n)\} \subset M$, $(f, g) \in B(S) \times B(S)$, $\text{bp-lim}_{n \rightarrow \infty} f_n = f$, and $\text{bp-lim}_{n \rightarrow \infty} g_n = g$, we have $(f, g) \in M$. The *bp-closure* of $M \subset B(S) \times B(S)$ is the smallest bp-closed subset of $B(S) \times B(S)$ that contains M .

A set $M \subset \bar{C}(S)$ is called *separating* if whenever $P, Q \in \mathcal{P}(S)$ and

$$(4.3) \quad \int f dP = \int f dQ, \quad f \in M,$$

we have $P = Q$. Also, M is called *convergence determining* if whenever $\{P_n\} \subset \mathcal{P}(S)$, $P \in \mathcal{P}(S)$, and

$$(4.4) \quad \lim_{n \rightarrow \infty} \int f dP_n = \int f dP, \quad f \in M,$$

we have $P_n \Rightarrow P$.

Given $P, Q \in \mathcal{P}(S)$, the set of all $f \in B(S)$ such that $\int f dP = \int f dQ$ is bp-closed. Consequently, Proposition 4.2 implies that $\bar{C}(S)$ is itself separating. It follows that if $M \subset \bar{C}(S)$ is convergence determining, then M is separating. The converse is false in general, as Problem 8 indicates. However, if S is compact, then $\mathcal{P}(S)$ is compact by Theorem 2.2, and the following lemma implies that the two concepts are equivalent.

4.3 Lemma Let $\{P_n\} \subset \mathcal{P}(S)$ be relatively compact, let $P \in \mathcal{P}(S)$, and let $M \subset \bar{C}(S)$ be separating. If (4.4) holds, then $P_n \Rightarrow P$.

Proof. If Q is the weak limit of a convergent subsequence of $\{P_n\}$, then (4.4) implies (4.3), so $Q = P$. It follows that $P_n \Rightarrow P$. \square

4.4 Proposition Let (S, d) be separable. The space of functions $f \in \bar{C}(S)$ that are uniformly continuous and have bounded support is convergence determining. If S is also locally compact, then $C_c(S)$, the space of $f \in \bar{C}(S)$ with compact support, is convergence determining.

Proof. Let $\{x_i\}$ be dense in S , and define $f_{ij} \in \bar{C}(S)$ for $i, j = 1, 2, \dots$ by

$$(4.5) \quad f_{ij}(x) = 2(1 - jd(x, x_i)) \vee 0.$$

Given an open set $G \subset S$, define $g_m \in M$ for $m = 1, 2, \dots$ by $g_m(x) = (\sum f_{ij}(x)) \wedge 1$, where the sum extends over those $i, j \leq m$ such that $B(x_i, 1/j) \subset G$ (and $B(x_i, 1/j)$ is compact if S is locally compact). If (4.4) holds, then

$$(4.6) \quad \lim_{n \rightarrow \infty} P_n(G) \geq \lim_{n \rightarrow \infty} \int g_m dP_n = \int g_m dP$$

for $m = 1, 2, \dots$, so by letting $m \rightarrow \infty$, we conclude that condition (e) of Theorem 3.1 holds. \square

Recall that a collection of functions $M \subset \bar{C}(S)$ is said to *separate points* if for every $x, y \in S$ with $x \neq y$ there exists $h \in M$ such that $h(x) \neq h(y)$. In

addition, M is said to *strongly separate points* if for every $x \in S$ and $\delta > 0$ there exists a finite set $\{h_1, \dots, h_k\} \subset M$ such that

$$(4.7) \quad \inf_{y: d(y, x) \geq \delta} \max_{1 \leq i \leq k} |h_i(y) - h_i(x)| > 0.$$

Clearly, if M strongly separates points, then M separates points.

4.5 Theorem Let (S, d) be complete and separable, and let $M \subset \bar{C}(S)$ be an algebra.

- (a) If M separates points, then M is separating.
- (b) If M strongly separates points, then M is convergence determining.

Proof. (a) Let $P, Q \in \mathcal{P}(S)$, and suppose that (4.3) holds. Then $\int h \, dP = \int h \, dQ$ for all h in the algebra $H = \{f + a: f \in M, a \in \mathbb{R}\}$, hence for all h in the closure (with respect to $\|\cdot\|$) of H . Let $g \in \bar{C}(S)$ and $\varepsilon > 0$ be arbitrary. By Lemma 2.1, there exists a compact set $K \subset S$ such that $P(K) \geq 1 - \varepsilon$ and $Q(K) \geq 1 - \varepsilon$. By the Stone-Weierstrass theorem, there exists a sequence $\{g_n\} \subset H$ such that $\sup_{x \in K} |g_n(x) - g(x)| \rightarrow 0$ as $n \rightarrow \infty$. Now observe that

$$\begin{aligned} (4.8) \quad & \left| \int_S g e^{-t g^2} \, dP - \int_S g e^{-t g^2} \, dQ \right| \\ & \leq \left| \int_S g e^{-t g^2} \, dP - \int_K g e^{-t g^2} \, dP \right| \\ & \quad + \left| \int_K g e^{-t g^2} \, dP - \int_K g_n e^{-t g_n^2} \, dP \right| \\ & \quad + \left| \int_K g_n e^{-t g_n^2} \, dP - \int_S g_n e^{-t g_n^2} \, dP \right| \\ & \quad + \left| \int_S g_n e^{-t g_n^2} \, dP - \int_S g_n e^{-t g_n^2} \, dQ \right| \\ & \quad + \left| \int_S g_n e^{-t g_n^2} \, dQ - \int_K g_n e^{-t g_n^2} \, dQ \right| \\ & \quad + \left| \int_K g_n e^{-t g_n^2} \, dQ - \int_S g e^{-t g^2} \, dQ \right| \\ & \quad + \left| \int_S g e^{-t g^2} \, dQ - \int_S g e^{-t g^2} \, dP \right| \end{aligned}$$

for each n , and the fourth term on the right is zero since $g_n e^{-t g_n^2}$ belongs to the closure of H . The second and sixth terms tend to zero as $n \rightarrow \infty$, so the left side of (4.8) is bounded by $4\gamma\sqrt{\varepsilon}$, where $\gamma = \sup_{t \geq 0} t e^{-t^2}$. Letting $\varepsilon \rightarrow 0$,

it follows that $\int g \, dP = \int g \, dQ$. Since $g \in \bar{C}(S)$ was arbitrary and $\bar{C}(S)$ is separating, we conclude that $P = Q$.

(b) Let $\{P_n\} \subset \mathcal{P}(S)$ and $P \in \mathcal{P}(S)$, and suppose that (4.4) holds. By Lemma 4.3 and part (a), it suffices to show that $\{P_n\}$ is relatively compact.

Let $f_1, \dots, f_k \in M$. Then

$$(4.9) \quad \lim_{n \rightarrow \infty} \int g \circ (f_1, \dots, f_k) \, dP_n = \int g \circ (f_1, \dots, f_k) \, dP$$

for all polynomials g in k variables by (4.4) and the assumption that M is an algebra. Since f_1, \dots, f_k are bounded, (4.9) holds for all $g \in \bar{C}(\mathbb{R}^k)$. We conclude that

$$(4.10) \quad P_n(f_1, \dots, f_k)^{-1} \Rightarrow P(f_1, \dots, f_k)^{-1}, \quad f_1, \dots, f_k \in M.$$

Let $K \subset S$ be compact, and let $\delta > 0$. For each $x \in S$, choose $\{h_i^x, \dots, h_{k(x)}^x\} \subset M$ satisfying

$$(4.11) \quad \varepsilon(x) \equiv \inf_{y: d(y, x) \geq \delta} \max_{1 \leq i \leq k(x)} |h_i^x(y) - h_i^x(x)| > 0,$$

and let $G_x = \{y \in S: \max_{1 \leq i \leq k(x)} |h_i^x(y) - h_i^x(x)| < \varepsilon(x)\}$. Then $K \subset \bigcup_{x \in K} G_x \subset K^\delta$, so, since K is compact, there exist $x_1, \dots, x_m \in K$ such that $K \subset \bigcup_{i=1}^m G_{x_i} \subset K^\delta$. Define $g_1, \dots, g_m \in \bar{C}(S)$ by

$$(4.12) \quad g_i(x) = \max_{1 \leq l \leq k(x_i)} |h_l^{x_i}(x) - h_l^{x_i}(x_i)|,$$

and observe that (4.10) implies that

$$(4.13) \quad P_n(g_1, \dots, g_m)^{-1} \Rightarrow P(g_1, \dots, g_m)^{-1}.$$

It follows that

$$\begin{aligned} (4.14) \quad \lim_{n \rightarrow \infty} P_n(K^\delta) &\geq \lim_{n \rightarrow \infty} P_n\left(\bigcup_{l=1}^m G_{x_l}\right) \\ &= \lim_{n \rightarrow \infty} P_n\left\{x \in S: \min_{1 \leq l \leq m} [g_l(x) - \varepsilon(x_l)] < 0\right\} \\ &\geq P\left\{x \in S: \min_{1 \leq l \leq m} [g_l(x) - \varepsilon(x_l)] < 0\right\} \\ &= P\left(\bigcup_{l=1}^m G_{x_l}\right) \\ &\geq P(K), \end{aligned}$$

where the middle inequality depends on (4.13) and Theorem 3.1. Applying Lemma 2.1 to P and to finitely many terms in the sequence $\{P_n\}$, we conclude that there exists a compact set $K \subset S$ such that $\inf_n P_n(K^\delta) \geq 1 - \delta$. By Theorem 2.2, $\{P_n\}$ is relatively compact. \square

We turn next to a result concerning countably infinite product spaces. Let (S_k, d_k) , $k = 1, 2, \dots$, be metric spaces, and define $S = \prod_{k=1}^{\infty} S_k$ and $d(x, y) = \sum_{k=1}^{\infty} 2^{-k}(d_k(x_k, y_k) \wedge 1)$ for all $x, y \in S$. Then (S, d) is separable if the S_k are separable and complete if the (S_k, d_k) are complete. If the S_k are separable, then $\mathcal{B}(S) = \prod_{k=1}^{\infty} \mathcal{B}(S_k)$.

4.6 Proposition Let (S_k, d_k) , $k = 1, 2, \dots$, and (S, d) be as above. Suppose $M_k \subset \bar{C}(S_k)$, and let

$$(4.15) \quad M = \{f(x) = \prod_{k=1}^n f_k(x_k) : n \geq 1, f_k \in M_k \cup \{1\} \text{ for } k = 1, \dots, n\}.$$

- (a) If the S_k are separable and the M_k are separating, then M is separating.
- (b) If the (S_k, d_k) are complete and separable and the M_k are convergence determining, then M is convergence determining.

Proof. (a) Suppose that $P, Q \in \mathcal{P}(S)$ and

$$(4.16) \quad \int f_1(x_1) \cdots f_n(x_n) P(dx) = \int f_1(x_1) \cdots f_n(x_n) Q(dx)$$

whenever $n \geq 1$ and $f_k \in M_k \cup \{1\}$ for $k = 1, \dots, n$. Given $n \geq 2$ and $f_k \in M_k \cup \{1\}$ for $k = 2, \dots, n$, put

$$(4.17) \quad \mu(dx) = f_2(x_2) \cdots f_n(x_n) P(dx), \quad v(dx) = f_2(x_2) \cdots f_n(x_n) Q(dx),$$

and let μ^1 and v^1 be the first marginals of μ and v on $\mathcal{B}(S_1)$. Since M_1 is separating (with respect to Borel probability measures), it is separating with respect to finite signed Borel measures as well. Therefore $\mu^1 = v^1$ and hence

$$(4.18) \quad \int \chi_{A_1}(x_1) f_2(x_2) \cdots f_n(x_n) P(dx) = \int \chi_{A_1}(x_1) f_2(x_2) \cdots f_n(x_n) Q(dx)$$

whenever $A_1 \in \mathcal{B}(S_1)$, $n \geq 2$, and $f_k \in M_k \cup \{1\}$ for $k = 2, \dots, n$. Proceeding inductively, we conclude that

$$(4.19) \quad \int \chi_{A_1}(x_1) \cdots \chi_{A_n}(x_n) P(dx) = \int \chi_{A_1}(x_1) \cdots \chi_{A_n}(x_n) Q(dx)$$

whenever $n \geq 1$ and $A_k \in \mathcal{B}(S_k)$ for $k = 1, \dots, n$. It follows that $P = Q$ on $\prod_{k=1}^{\infty} \mathcal{B}(S_k) = \mathcal{B}(S)$ and thus that M is separating.

(b) Let $\{P_n\} \subset \mathcal{P}(S)$ and $P \in \mathcal{P}(S)$, and suppose that (4.4) holds. Then, for $k = 1, 2, \dots$, $\lim_{n \rightarrow \infty} \int f dP_n^k = \int f dP^k$ for all $f \in M_k$, where P_n^k and P^k denote the k th marginals of P_n and P , and hence $P_n^k \Rightarrow P^k$. In particular, this implies that $\{P_n^k\}$ is relatively compact for $k = 1, 2, \dots$, and hence, by

Theorem 2.2 and Proposition 2.4, $\{P_n\}$ is relatively compact. By Lemma 4.3, $P_n \Rightarrow P$, so M is convergence determining. \square

We conclude this section by generalizing the concept of separating set. A set $M \subset B(S)$ is called *separating* if whenever $P, Q \in \mathcal{P}(S)$ and (4.3) holds, we have $P = Q$. More generally, if $\mathcal{M} \subset \mathcal{P}(S)$, a set $M \subset M(S)$ (the space of real-valued Borel functions on S) is called *separating on \mathcal{M}* if

$$(4.20) \quad \int |f| dP < \infty, \quad f \in M, P \in \mathcal{M},$$

and if whenever $P, Q \in \mathcal{M}$ and (4.3) holds, we have $P = Q$. For example, the set of monomials on \mathbb{R} (i.e., $1, x, x^2, x^3, \dots$) is separating on

$$(4.21) \quad \mathcal{M} = \left\{ P \in \mathcal{P}(\mathbb{R}): \overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \left(\int_{-\infty}^{\infty} |x|^n P(dx) \right)^{1/n} < \infty \right\}$$

(Feller (1971), p. 514).

5. THE SPACE $D_E[0, \infty)$

Throughout the remaining sections of this chapter, (E, r) denotes a metric space, and q denotes the metric $r \wedge 1$.

Most stochastic processes arising in applications have the property that they have right and left limits at each time point for almost every sample path. It has become conventional to assume that sample paths are actually right continuous when this can be done (as it usually can) without altering the finite-dimensional distributions. Consequently, the space $D_E[0, \infty)$ of right continuous functions $x: [0, \infty) \rightarrow E$ with left limits (i.e., for each $t \geq 0$, $\lim_{s \rightarrow t+} x(s) = x(t)$ and $\lim_{s \rightarrow t-} x(s) = x(t-)$ exists; by convention, $\lim_{s \rightarrow 0-} x(s) = x(0-) = x(0)$) is of considerable importance.

We begin by observing that functions in $D_E[0, \infty)$ are better behaved than might initially be suspected.

5.1 Lemma If $x \in D_E[0, \infty)$, then x has at most countably many points of discontinuity.

Proof. For $n = 1, 2, \dots$, let $A_n = \{t > 0: r(x(t), x(t-)) > 1/n\}$, and observe that A_n has no limit points in $[0, \infty)$ since $\lim_{s \rightarrow t+} x(s)$ and $\lim_{s \rightarrow t-} x(s)$ exist for all $t \geq 0$. Consequently, each A_n is countable. But the set of all discontinuities of x is $\bigcup_{n=1}^{\infty} A_n$, and hence it too is countable. \square

The results on convergence of probability measures in Sections 1–4 are best suited for complete separable metric spaces. With this in mind we now define a metric on $D_E[0, \infty)$ under which it is a separable metric space if E is separable,

and is complete if (E, r) is complete. Let Λ' be the collection of (strictly) increasing functions λ mapping $[0, \infty)$ onto $[0, \infty)$ (in particular, $\lambda(0) = 0$, $\lim_{t \rightarrow \infty} \lambda(t) = \infty$, and λ is continuous). Let Λ be the set of Lipschitz continuous functions $\lambda \in \Lambda'$ such that

$$(5.1) \quad \begin{aligned} \gamma(\lambda) &\equiv \text{ess sup}_{t \geq 0} |\log \lambda'(t)| \\ &= \sup_{s > t \geq 0} \left| \log \frac{\lambda(s) - \lambda(t)}{s - t} \right| < \infty. \end{aligned}$$

For $x, y \in D_E[0, \infty)$, define

$$(5.2) \quad d(x, y) = \inf_{\lambda \in \Lambda} \left[\gamma(\lambda) \vee \int_0^\infty e^{-u} d(x, y, \lambda, u) du \right],$$

where

$$(5.3) \quad d(x, y, \lambda, u) = \sup_{t \geq 0} q(x(t \wedge u), y(\lambda(t) \wedge u)).$$

It follows that, given $\{x_n\}, \{y_n\} \subset D_E[0, \infty)$, $\lim_{n \rightarrow \infty} d(x_n, y_n) = 0$ if and only if there exists $\{\lambda_n\} \subset \Lambda$ such that $\lim_{n \rightarrow \infty} \gamma(\lambda_n) = 0$ and

$$(5.4) \quad \lim_{n \rightarrow \infty} m\{u \in [0, u_0] : d(x_n, y_n, \lambda_n, u) \geq \varepsilon\} = 0$$

for every $\varepsilon > 0$ and $u_0 > 0$, where m is Lebesgue measure; moreover, since

$$(5.5) \quad \text{ess sup}_{t \geq 0} |\lambda'(t) - 1| \leq 1 - e^{-\gamma(\lambda)} \leq \gamma(\lambda)$$

for every $\lambda \in \Lambda$,

$$(5.6) \quad \lim_{n \rightarrow \infty} \gamma(\lambda_n) = 0$$

implies that

$$(5.7) \quad \lim_{n \rightarrow \infty} \sup_{0 \leq t \leq T} |\lambda_n(t) - t| = 0$$

for all $T > 0$.

Let $x, y \in D_E[0, \infty)$, and observe that

$$(5.8) \quad \sup_{t \geq 0} q(x(t \wedge u), y(\lambda(t) \wedge u)) = \sup_{t \geq 0} q(x(\lambda^{-1}(t) \wedge u), y(t \wedge u))$$

for all $\lambda \in \Lambda$ and $u \geq 0$, and therefore $d(x, y, \lambda, u) = d(y, x, \lambda^{-1}, u)$. Together with the fact that $\gamma(\lambda) = \gamma(\lambda^{-1})$ for every $\lambda \in \Lambda$, this implies that $d(x, y) = d(y, x)$. If $d(x, y) = 0$, then, by (5.4) and (5.7), $x(t) = y(t)$ for every

continuity point t of y , and hence $x = y$ by Lemma 5.1 and the right continuity of x and y . Thus, to show that d is a metric, we need only verify the triangle inequality. Let $x, y, z \in D_E[0, \infty)$, $\lambda_1, \lambda_2 \in \Lambda$, and $u \geq 0$. Then

$$\begin{aligned}
 (5.9) \quad & \sup_{t \geq 0} q(x(t \wedge u), z(\lambda_2(\lambda_1(t)) \wedge u)) \\
 & \leq \sup_{t \geq 0} q(x(t \wedge u), y(\lambda_1(t) \wedge u)) \\
 & \quad + \sup_{t \geq 0} q(y(\lambda_1(t) \wedge u), z(\lambda_2(\lambda_1(t)) \wedge u)) \\
 & = \sup_{t \geq 0} q(x(t \wedge u), y(\lambda_1(t) \wedge u)) \\
 & \quad + \sup_{t \geq 0} q(y(t \wedge u), z(\lambda_2(t) \wedge u)),
 \end{aligned}$$

that is, $d(x, z, \lambda_2 \circ \lambda_1, u) \leq d(x, y, \lambda_1, u) + d(y, z, \lambda_2, u)$. But since $\lambda_2 \circ \lambda_1 \in \Lambda$ and

$$(5.10) \quad \gamma(\lambda_2 \circ \lambda_1) \leq \gamma(\lambda_1) + \gamma(\lambda_2),$$

we obtain $d(x, z) \leq d(x, y) + d(y, z)$.

The topology induced on $D_E[0, \infty)$ by the metric d is called the *Skorohod topology*.

5.2 Proposition Let $\{x_n\} \subset D_E[0, \infty)$ and $x \in D_E[0, \infty)$. Then $\lim_{n \rightarrow \infty} d(x_n, x) = 0$ if and only if there exists $\{\lambda_n\} \subset \Lambda$ such that (5.6) holds and

$$(5.11) \quad \lim_{n \rightarrow \infty} d(x_n, x, \lambda_n, u) = 0 \quad \text{for all continuity points } u \text{ of } x.$$

In particular, $\lim_{n \rightarrow \infty} d(x_n, x) = 0$ implies that $\lim_{n \rightarrow \infty} x_n(u) = \lim_{n \rightarrow \infty} x_n(u^-) = x(u)$ for all continuity points u of x .

Proof. The sufficiency follows from Lemma 5.1. Conversely, suppose that $\lim_{n \rightarrow \infty} d(x_n, x) = 0$, and let u be a continuity point of x . Recalling (5.4), there exist $\{\lambda_n\} \subset \Lambda$ and $\{u_n\} \subset (u, \infty)$ such that (5.6) holds and

$$(5.12) \quad \lim_{n \rightarrow \infty} \sup_{t \geq 0} q(x_n(t \wedge u_n), x(\lambda_n(t) \wedge u_n)) = 0.$$

Now

$$\begin{aligned}
 (5.13) \quad & \sup_{t \geq 0} q(x_n(t \wedge u), x(\lambda_n(t) \wedge u)) \\
 & \leq \sup_{t \geq 0} q(x_n(t \wedge u), x(\lambda_n(t \wedge u) \wedge u_n)) \\
 & \quad + \sup_{t \geq 0} q(x(\lambda_n(t \wedge u) \wedge u_n), x(\lambda_n(t) \wedge u))
 \end{aligned}$$

$$\begin{aligned} &\leq \sup_{0 \leq t \leq u} q(x_n(t \wedge u_n), x(\lambda_n(t) \wedge u_n)) \\ &\quad + \sup_{u \leq s \leq \lambda_n(u) \vee u} q(x(s), x(u)) \\ &\quad \vee \sup_{\lambda_n(u) \wedge u \leq s \leq u} q(x(\lambda_n(u) \wedge u_n), x(s)) \end{aligned}$$

for each n , where the second half of the second inequality follows by considering separately the cases $t \leq u$ and $t > u$. Thus, $\lim_{n \rightarrow \infty} d(x_n, x, \lambda_n, u) = 0$ by (5.12), (5.7), and the continuity of x at u . \square

5.3 Proposition Let $\{x_n\} \subset D_E[0, \infty)$ and $x \in D_E[0, \infty)$. Then the following are equivalent:

- (a) $\lim_{n \rightarrow \infty} d(x_n, x) = 0$.
- (b) There exists $\{\lambda_n\} \subset \Lambda$ such that (5.6) holds and

$$(5.14) \quad \lim_{n \rightarrow \infty} \sup_{0 \leq t \leq T} r(x_n(t), x(\lambda_n(t))) = 0$$

for all $T > 0$.

- (c) For each $T > 0$, there exists $\{\lambda_n\} \subset \Lambda'$ (possibly depending on T) such that (5.7) and (5.14) hold.

5.4 Remark In conditions (b) and (c) of Proposition 5.3, (5.14) can be replaced by

$$(5.14') \quad \lim_{n \rightarrow \infty} \sup_{0 \leq t \leq T} r(x_n(\lambda_n(t)), x(t)) = 0.$$

Denoting the resulting conditions by (b') and (c'), this is easily established by checking that (b) is equivalent to (b') and (c) is equivalent to (c'). \square

Proof. (a \Rightarrow b) Assuming (a) holds, there exist $\{\lambda_n\} \subset \Lambda$ and $\{u_n\} \subset (0, \infty)$ such that (5.6) holds, $u_n \rightarrow \infty$, and $d(x_n, x, \lambda_n, u_n) \rightarrow 0$; in particular,

$$(5.15) \quad \lim_{n \rightarrow \infty} \sup_{t \geq 0} r(x_n(t \wedge u_n), x(\lambda_n(t) \wedge u_n)) = 0.$$

Given $T > 0$, note that $u_n \geq T \vee \lambda_n(T)$ for all n sufficiently large, so (5.15) implies (5.14).

(b \Rightarrow a) Let $\{\lambda_n\} \subset \Lambda$ satisfy the conditions of (b). Then

$$(5.16) \quad \lim_{n \rightarrow \infty} \sup_{t \geq 0} q(x_n(t \wedge u), x(\lambda_n(t) \wedge u)) = 0$$

for every continuity point u of x by (5.13) with $u_n > \lambda_n(u) \vee u$ for each n . Hence (a) holds.

(b \Rightarrow c) Immediate.

(c \Rightarrow b) Let N be a positive integer, and choose $\{\lambda_n^N\} \subset \Lambda'$ satisfying the conditions of (c) with $T = N$, such that for each n , $\lambda_n^N(t) = \lambda_n^N(N) + t - N$ for all $t > N$. Define $\tau_0^N = 0$ and, for $k = 1, 2, \dots$,

$$(5.17) \quad \tau_k^N = \inf \left\{ t > \tau_{k-1}^N : r(x(t), x(\tau_{k-1}^N)) > \frac{1}{N} \right\}$$

if $\tau_{k-1}^N < \infty$, $\tau_k^N = \infty$ if $\tau_{k-1}^N = \infty$. Observe that the sequence $\{\tau_k^N\}$ is (strictly) increasing (as long as its terms remain finite) by the right continuity of x and has no finite limit points since x has left limits. For each n , let $u_{k,n}^N = (\lambda_n^N)^{-1}(\tau_k^N)$ for $k = 0, 1, \dots$, where $(\lambda_n^N)^{-1}(\infty) = \infty$, and define $\mu_n^N \in \Lambda$ by

$$(5.18) \quad \begin{aligned} \mu_n^N(t) &= \tau_k^N + (t - u_{k,n}^N)(u_{k+1,n}^N - u_{k,n}^N)^{-1}(\tau_{k+1}^N - \tau_k^N), \\ t &\in [u_{k,n}^N, u_{k+1,n}^N) \cap [0, N], \quad k = 0, 1, \dots, \\ \mu_n^N(t) &= \mu_n^N(N) + t - N, \quad t > N, \end{aligned}$$

where, by convention, $\infty^{-1}\infty = 1$. With this convention,

$$(5.19) \quad \gamma(\mu_n^N) = \max_{N \geq u_{k,n}^N < N} |\log(u_{k+1,n}^N - u_{k,n}^N)^{-1}(\tau_{k+1}^N - \tau_k^N)|$$

and

$$\begin{aligned} (5.20) \quad &\sup_{0 \leq t \leq N} r(x_n(t), x(\mu_n^N(t))) \\ &\leq \sup_{0 \leq t \leq N} r(x_n(t), x(\lambda_n^N(t))) + \sup_{0 \leq t \leq N} r(x(\lambda_n^N(t)), x(\mu_n^N(t))) \\ &\leq \sup_{0 \leq t \leq N} r(x_n(t), x(\lambda_n^N(t))) + \frac{2}{N} \end{aligned}$$

for all n . Since $\lim_{n \rightarrow \infty} u_{k,n}^N = \tau_k^N$ for $k = 0, 1, \dots$, (5.18) implies that $\lim_{n \rightarrow \infty} \gamma(\mu_n^N) = 0$, which, together with (5.20) and (5.14) with $T = N$, implies that we can select $1 < n_1 < n_2 < \dots$ such that $\gamma(\mu_{n_i}^N) \leq 1/N$ and $\sup_{0 \leq t \leq N} r(x_n(t), x(\mu_n^N(t))) \leq 3/N$ for all $n \geq n_N$. For $1 \leq n < n_1$, let $\hat{\lambda}_n \in \Lambda$ be arbitrary. For $n_N \leq n < n_{N+1}$, where $N \geq 1$, let $\hat{\lambda}_n = \mu_n^N$. Then $\{\hat{\lambda}_n\} \subset \Lambda$ satisfies the conditions of (b). \square

5.5 Corollary

For $x, y \in D_E[0, \infty)$, define

$$(5.21) \quad d'(x, y) = \inf_{\lambda \in \Lambda'} \int_0^\infty e^{-u} \sup_{t \geq 0} [q(x(t \wedge u), y(\lambda(t) \wedge u)) \\ V(|\lambda(t) \wedge u - t \wedge u| \wedge 1)] du.$$

The d' is a metric on $D_E[0, \infty)$ that is equivalent to d . (However, the metric space $(D_E[0, \infty), d')$ is not complete.)

Proof. The proof that d' is a metric is essentially the same as that for d . The equivalence of d and d' follows from the equivalence of (a) and (c) in Proposition 5.3. We leave the verification of the parenthetical remark to the reader. \square

5.6 Theorem If E is separable, then $D_E[0, \infty)$ is separable. If (E, r) is complete, then $(D_E[0, \infty), d)$ is complete.

Proof. Let $\{\alpha_i\}$ be a countable dense subset of E and let Γ be the collection of functions of the form

$$(5.22) \quad x(t) = \begin{cases} \alpha_{i_k}, & t_{k-1} \leq t < t_k, \quad k = 1, \dots, n, \\ \alpha_{i_n}, & t \geq t_n, \end{cases}$$

where $0 = t_0 < t_1 < \dots < t_n$ are rationals, i_1, \dots, i_n are positive integers, and $n \geq 1$. We leave to the reader the proof that Γ is dense in $D_E[0, \infty)$ (Problem 14).

To prove completeness, it is enough to show that every Cauchy sequence has a convergent subsequence. If $\{x_n\} \subset D_E[0, \infty)$ is Cauchy, then there exist $1 \leq N_1 < N_2 < \dots$ such that $m, n \geq N_k$ implies

$$(5.23) \quad d(x_n, x_m) \leq 2^{-k-1}e^{-k}.$$

For $k = 1, 2, \dots$, let $y_k = x_{N_k}$ and, by (5.23), select $u_k > k$ and $\lambda_k \in \Lambda$ such that

$$(5.24) \quad \gamma(\lambda_k) \vee d(y_k, y_{k+1}, \lambda_k, u_k) \leq 2^{-k};$$

then, recalling (5.5),

$$(5.25) \quad \mu_k \equiv \lim_{n \rightarrow \infty} \lambda_{k+n} \circ \dots \circ \lambda_{k+1} \circ \lambda_k$$

exists uniformly on bounded intervals, is Lipschitz continuous, satisfies

$$(5.26) \quad \gamma(\mu_k) \leq \sum_{l=k}^{\infty} \gamma(\lambda_l) \leq 2^{-k+1},$$

and hence belongs to Λ . Since

$$\begin{aligned} (5.27) \quad & \sup_{t \geq 0} q(y_k(\mu_k^{-1}(t) \wedge u_k), y_{k+1}(\mu_{k+1}^{-1}(t) \wedge u_k)) \\ &= \sup_{t \geq 0} q(y_k(\mu_k^{-1}(t) \wedge u_k), y_{k+1}(\lambda_k(\mu_k^{-1}(t)) \wedge u_k)) \\ &= \sup_{t \geq 0} q(y_k(t \wedge u_k), y_{k+1}(\lambda_k(t) \wedge u_k)) \\ &\leq 2^{-k} \end{aligned}$$

for $k = 1, 2, \dots$ by (5.24), it follows from the completeness of (E, r) that $z_k \equiv y_k \circ \mu_k^{-1}$ converges uniformly on bounded intervals to a function $y: [0, \infty) \rightarrow E$.

But each $z_k \in D_E[0, \infty)$, so y must also belong to $D_E[0, \infty)$. Since $\lim_{k \rightarrow \infty} y(\mu_k^{-1}) = 0$ and

$$(5.28) \quad \lim_{k \rightarrow \infty} \sup_{0 \leq t \leq T} r(y_k(\mu_k^{-1}(t)), y(t)) = 0$$

for all $T > 0$, we conclude that $\lim_{k \rightarrow \infty} d(y_k, y) = 0$ by Proposition 5.3 (see Remark 5.4). \square

6. THE COMPACT SETS OF $D_E[0, \infty)$

Again let (E, r) denote a metric space. In order to apply Prohorov's theorem to $\mathcal{P}(D_E[0, \infty))$, we must have a characterization of the compact subsets of $D_E[0, \infty)$. With this in mind we first give conditions under which a collection of step functions is compact. Given a step function $x \in D_E[0, \infty)$, define $s_0(x) = 0$ and, for $k = 1, 2, \dots$,

$$(6.1) \quad s_k(x) = \inf \{t > s_{k-1}(x) : x(t) \neq x(t-)\}$$

if $s_{k-1}(x) < \infty$, $s_k(x) = \infty$ if $s_{k-1}(x) = \infty$.

6.1 Lemma Let $\Gamma \subset E$ be compact, let $\delta > 0$, and define $A(\Gamma, \delta)$ to be the set of step functions $x \in D_E[0, \infty)$ such that $x(t) \in \Gamma$ for all $t \geq 0$ and $s_k(x) - s_{k-1}(x) > \delta$ for each $k \geq 1$ for which $s_{k-1}(x) < \infty$. Then the closure of $A(\Gamma, \delta)$ is compact.

Proof. It is enough to show that every sequence in $A(\Gamma, \delta)$ has a convergent subsequence. Given $\{x_n\} \subset A(\Gamma, \delta)$, there exists by a diagonalization argument a subsequence $\{y_m\}$ of $\{x_n\}$ such that, for $k = 0, 1, \dots$, either (a) $s_k(y_m) < \infty$ for each m , $\lim_{m \rightarrow \infty} s_k(y_m) \equiv t_k$ exists (possibly ∞), and $\lim_{m \rightarrow \infty} y_m(s_k(y_m)) \equiv \alpha_k$ exists, or (b) $s_k(y_m) = \infty$ for each m . Since $s_k(y_m) - s_{k-1}(y_m) > \delta$ for each $k \geq 1$ and m for which $s_{k-1}(y_m) < \infty$, it follows easily that $\{y_m\}$ converges to the function $y \in D_E[0, \infty)$ defined by $y(t) = \alpha_k$, $t_k \leq t < t_{k+1}$, $k = 0, 1, \dots$. \square

The conditions for compactness are stated in terms of the following modulus of continuity. For $x \in D_E[0, \infty)$, $\delta > 0$, and $T > 0$, define

$$(6.2) \quad w'(x, \delta, T) = \inf \max_{\{\iota_i\}} \sup_{\substack{i- \\ \iota_i - \iota_{i-1} \in [t_{i-1}, t_i]}} r(x(s), x(t)),$$

where $\{\iota_i\}$ ranges over all partitions of the form $0 = \iota_0 < \iota_1 < \cdots < \iota_{n-1} < T \leq \iota_n$ with $\min_{1 \leq i \leq n} (\iota_i - \iota_{i-1}) > \delta$ and $n \geq 1$. Note that $w'(x, \delta, T)$ is nondecreasing in δ and in T , and that

$$(6.3) \quad w'(x, \delta, T) \leq w'(y, \delta, T) + 2 \sup_{0 \leq s \leq T+\delta} r(x(s), y(s)).$$

6.2 Lemma (a) For each $x \in D_E[0, \infty)$ and $T > 0$, $w'(x, \delta, T)$ is right continuous in δ and

$$(6.4) \quad \lim_{\delta \rightarrow 0} w'(x, \delta, T) = 0.$$

(b) If $\{x_n\} \subset D_E[0, \infty)$, $x \in D_E[0, \infty)$, and $\lim_{n \rightarrow \infty} d(x_n, x) = 0$, then

$$(6.5) \quad \overline{\lim}_{n \rightarrow \infty} w'(x_n, \delta, T) \leq w'(x, \delta, T + \varepsilon)$$

for every $\delta > 0$, $T > 0$, and $\varepsilon > 0$.

(c) For each $\delta > 0$ and $T > 0$, $w'(x, \delta, T)$ is Borel measurable in x .

Proof. (a) The right continuity follows from the fact that any partition that is admissible in the definition of $w'(x, \delta, T)$ is admissible for some $\delta' > \delta$. To obtain (6.4), let $N \geq 1$ and define $\{\tau_k^N\}$ as in (5.17). If $0 < \delta < \min\{\tau_{k+1}^N - \tau_k^N : \tau_k^N < T\}$, then $w'(x, \delta, T) \leq 2/N$.

(b) Let $\{x_n\} \subset D_E[0, \infty)$, $x \in D_E[0, \infty)$, $\delta > 0$, and $T > 0$. If $\lim_{n \rightarrow \infty} d(x_n, x) = 0$, then by Proposition 5.3, there exists $\{\lambda_n\} \subset \Lambda'$ such that (5.7) and (5.14) hold with T replaced by $T + \delta$. For each n , let $y_n(t) = x(\lambda_n(t))$ for all $t \geq 0$ and $\delta_n = \sup_{0 \leq t \leq T} [\lambda_n(t + \delta) - \lambda_n(t)]$. Then, using (6.3) and part (a),

$$\begin{aligned} (6.6) \quad \overline{\lim}_{n \rightarrow \infty} w'(x_n, \delta, T) &= \overline{\lim}_{n \rightarrow \infty} w'(y_n, \delta, T) \\ &\leq \overline{\lim}_{n \rightarrow \infty} w'(x, \delta_n, \lambda_n(T)) \\ &\leq \overline{\lim}_{n \rightarrow \infty} w'(x, \delta_n \vee \delta, T + \varepsilon) \\ &= w'(x, \delta, T + \varepsilon) \end{aligned}$$

for all $\varepsilon > 0$.

(c) By part (b) and the monotonicity of $w'(x, \delta, T)$ in T , $w'(x, \delta, T + \varepsilon) = \lim_{\varepsilon \rightarrow 0^+} w'(x, \delta, T + \varepsilon)$ is upper semicontinuous, hence Borel measurable, in x . Therefore it suffices to observe that $w'(x, \delta, T) = \lim_{\varepsilon \rightarrow 0^+} w'(x, \delta, (T - \varepsilon) +)$ for every $x \in D_E[0, \infty)$. \square

6.3 Theorem Let (E, r) be complete. Then the closure of $A \subset D_E[0, \infty)$ is compact if and only if the following two conditions hold:

(a) For every rational $t \geq 0$, there exists a compact set $\Gamma_t \subset E$ such that $x(t) \in \Gamma_t$ for all $x \in A$.

(b) For each $T > 0$,

$$(6.7) \quad \lim_{\delta \rightarrow 0} \sup_{x \in A} w'(x, \delta, T) = 0.$$

6.4 Remark In Theorem 6.3 it is actually necessary that for each $T > 0$ there exist a compact set $\Gamma_T \subset E$ such that $x(t) \in \Gamma_T$ for $0 \leq t \leq T$ and all $x \in A$. See Problem 16. \square

Proof. Suppose that A satisfies (a) and (b), and let $l \geq 1$. Choose $\delta_l \in (0, 1)$ such that

$$(6.8) \quad \sup_{x \in A} w'(x, \delta_l, l) \leq \frac{1}{l}$$

and $m_l \geq 2$ such that $1/m_l < \delta_l$. Define $\Gamma^{(l)} = \bigcup_{i=0}^{(l+1)m_l} \Gamma_{i/m_l}$ and, using the notation of Lemma 6.1, let $A_l = A(\Gamma^{(l)}, \delta_l)$.

Given $x \in A$, there is a partition $0 = t_0 < t_1 < \cdots < t_{n-1} < l \leq t_n < l+1 < t_{n+1} = \infty$ with $\min_{1 \leq i \leq n} (t_i - t_{i-1}) > \delta_l$ such that

$$(6.9) \quad \max_{1 \leq i \leq n} \sup_{s, t \in [t_{i-1}, t_i]} r(x(s), x(t)) \leq \frac{2}{l}.$$

Define $x' \in A_l$ by $x'(t) = x([m_l t] + 1)/m_l$ for $t_i \leq t < t_{i+1}$, $i = 0, 1, \dots, n$. Then $\sup_{0 \leq t \leq l} r(x'(t), x(t)) \leq 2/l$, so

$$(6.10) \quad \begin{aligned} d(x', x) &\leq \int_0^\infty e^{-u} \sup_{t \geq 0} [r(x'(t \wedge u), x(t \wedge u)) \wedge 1] du \\ &\leq 2/l + e^{-l} < 3/l. \end{aligned}$$

It follows that $A \subset A_l^{3/l}$. Now l was arbitrary, so since \bar{A}_l is compact for each $l \geq 1$ by Lemma 6.1, and since $A \subset \bigcap_{l \geq 1} A_l^{3/l}$, A is totally bounded and hence has compact closure.

Conversely, suppose that A has compact closure. We leave the proof of (a) to the reader (Problem 16). To see that (b) holds, suppose there exist $\eta > 0$, $T > 0$, and $\{x_n\} \subset A$ such that $w'(x_n, 1/n, T) \geq \eta$ for all n . Since A has compact closure, we may assume that $\lim_{n \rightarrow \infty} d(x_n, x) = 0$ for some $x \in D_E[0, \infty)$. But then Lemma 6.2(b) implies that

$$(6.11) \quad \eta \leq \overline{\lim}_{n \rightarrow \infty} w'(x_n, \delta, T) \leq w'(x, \delta, T + 1)$$

for all $\delta > 0$. Letting $\delta \rightarrow 0$, the right side of (6.11) tends to zero by Lemma 6.2(a), and this results in a contradiction. \square

We conclude this section with a further characterization of convergence of sequences in $D_E[0, \infty)$. (This result would have been included in Section 5 were it not for the fact that we need Lemma 6.2 in the proof.) We note that $(C_E[0, \infty), d_V)$ is a metric space, where

$$(6.12) \quad d_V(x, y) = \int_0^\infty e^{-u} \sup_{0 \leq t \leq u} [r(x(t), y(t)) \wedge 1] du.$$

Moreover, if $\{x_n\} \subset C_E[0, \infty)$ and $x \in C_E[0, \infty)$, then $\lim_{n \rightarrow \infty} d_U(x_n, x) = 0$ if and only if whenever $\{t_n\} \subset [0, \infty)$, $t \geq 0$, and $\lim_{n \rightarrow \infty} t_n = t$, we have $\lim_{n \rightarrow \infty} r(x_n(t_n), x(t)) = 0$. The following proposition gives an analogue of this result for $(D_E[0, \infty), d)$.

6.5 Proposition Let (E, r) be arbitrary, and let $\{x_n\} \subset D_E[0, \infty)$ and $x \in D_E[0, \infty)$. Then $\lim_{n \rightarrow \infty} d(x_n, x) = 0$ if and only if whenever $\{t_n\} \subset [0, \infty)$, $t \geq 0$, and $\lim_{n \rightarrow \infty} t_n = t$, the following conditions hold:

- (a) $\lim_{n \rightarrow \infty} r(x_n(t_n), x(t)) \wedge r(x_n(t_n), x(t-)) = 0$.
- (b) If $\lim_{n \rightarrow \infty} r(x_n(t_n), x(t)) = 0$, $s_n \geq t_n$ for each n , and $\lim_{n \rightarrow \infty} s_n = t$, then $\lim_{n \rightarrow \infty} r(x_n(s_n), x(t)) = 0$.
- (c) If $\lim_{n \rightarrow \infty} r(x_n(t_n), x(t-)) = 0$, $0 \leq s_n \leq t_n$ for each n , and $\lim_{n \rightarrow \infty} s_n = t$, then $\lim_{n \rightarrow \infty} r(x_n(s_n), x(t-)) = 0$.

Proof. Suppose $\lim_{n \rightarrow \infty} d(x_n, x) = 0$, and let $\{t_n\} \subset [0, \infty)$, $t \geq 0$, and $\lim_{n \rightarrow \infty} t_n = t$. Choose $T > 0$ such that $\{t_n\} \subset [0, T]$ and $t \leq T$. By Proposition 5.3, there exists $\{\lambda_n\} \subset \Lambda'$ such that (5.7) and (5.14) hold. Therefore, since

$$\begin{aligned} (6.13) \quad & r(x_n(t_n), x(t)) \wedge r(x_n(t_n), x(t-)) \\ & \leq \sup_{0 \leq u \leq T} r(x_n(u), x(\lambda_n(u))) \\ & \quad + r(x(\lambda_n(t_n)), x(t)) \wedge r(x(\lambda_n(t_n)), x(t-)) \end{aligned}$$

for each n , and since $\lim_{n \rightarrow \infty} \lambda_n(t_n) = t$, (a) holds. If, in addition, $t_n \leq s_n \leq T$ for each n and $\lim_{n \rightarrow \infty} s_n = t$, then

$$\begin{aligned} (6.14) \quad & r(x_n(s_n), x(t)) \leq \sup_{0 \leq u \leq T} r(x_n(u), x(\lambda_n(u))) \\ & \quad + r(x(\lambda_n(s_n)), x(t)) \end{aligned}$$

and

$$\begin{aligned} (6.15) \quad & r(x(\lambda_n(t_n)), x(t)) \leq \sup_{0 \leq u \leq T} r(x(\lambda_n(u)), x_n(u)) \\ & \quad + r(x_n(t_n), x(t)) \end{aligned}$$

for each n . If also $\lim_{n \rightarrow \infty} r(x_n(t_n), x(t)) = 0$, then $\lim_{n \rightarrow \infty} r(x(\lambda_n(t_n)), x(t)) = 0$ by (6.15), so since $\lambda_n(s_n) \geq \lambda_n(t_n)$ for each n and $\lim_{n \rightarrow \infty} \lambda_n(s_n) = t$, it follows that $\lim_{n \rightarrow \infty} r(x(\lambda_n(s_n)), x(t)) = 0$. Thus, (b) follows from (6.14), and the proof of (c) is similar.

We turn to the proof of the sufficiency of (a)–(c). Fix $T > 0$ and for each n define

$$(6.16) \quad \varepsilon_n = 2 \inf \{\varepsilon > 0: \Gamma(t, n, \varepsilon) \neq \emptyset \text{ for } 0 \leq t \leq T\},$$

where

$$(6.17) \quad \Gamma(t, n, \varepsilon) = \{s \in (t - \varepsilon, t + \varepsilon) \cap [0, \infty) : r(x_n(s), x(t)) < \varepsilon, \\ r(x_n(s-), x(t-)) < \varepsilon\}.$$

We claim that $\lim_{n \rightarrow \infty} \varepsilon_n = 0$. Suppose not. Then there exist $\varepsilon > 0$, a sequence $\{n_k\}$ of positive integers, and a sequence $\{t_k\} \subset [0, T]$ such that $\Gamma(t_k, n_k, \varepsilon) = \emptyset$ for all k . By choosing a subsequence if necessary, we can assume that $\lim_{k \rightarrow \infty} t_k \equiv t$ exists and that $t_k < t$ for all k , $t_k > t$ for all k , or $t_k = t$ for all k . In the first case, $\lim_{k \rightarrow \infty} x(t_k) = \lim_{k \rightarrow \infty} x(t_k-) = x(t-)$, and in the second case, $\lim_{k \rightarrow \infty} x(t_k) = \lim_{k \rightarrow \infty} x(t_k-) = x(t)$. Since (a) implies that $\lim_{n \rightarrow \infty} x_n(s) = \lim_{n \rightarrow \infty} x_n(s-) = x(s)$ for all continuity points s of x , there exist (by Lemma 5.1 and Proposition 5.2) sequences $\{a_n\}$ and $\{b_n\}$ of continuity points of x such that $a_n < t < b_n$ for each n and $a_n \rightarrow t$ and $b_n \rightarrow t$ sufficiently slowly that $\lim_{n \rightarrow \infty} x_n(a_n) = \lim_{n \rightarrow \infty} x_n(a_n-) = \lim_{n \rightarrow \infty} x(a_n) = x(t-)$ and $\lim_{n \rightarrow \infty} x_n(b_n) = \lim_{n \rightarrow \infty} x_n(b_n-) = \lim_{n \rightarrow \infty} x(b_n) = x(t)$. If $t_k < t$ (respectively, $t_k > t$) for all k , then a_{n_k} (respectively, b_{n_k}) belongs to $\Gamma(t_k, n_k, \varepsilon)$ for all k sufficiently large, a contradiction. It remains to consider the case $t_k = t$ for all k . If $x(t) = x(t-)$, then $t \in \Gamma(t, n_k, \varepsilon)$ for all k sufficiently large by condition (a). Therefore we suppose that $r(x(t), x(t-)) = \delta > 0$. By the choice of $\{a_n\}$ and $\{b_n\}$ and by condition (a), there exists $n_0 \geq 1$ such that for all $n \geq n_0$,

$$(6.18) \quad r(x_n(a_n), x(t-)) \vee r(x_n(b_n), x(t)) < \frac{\delta \wedge \varepsilon}{2},$$

$$(6.19) \quad \sup_{a_n \leq s \leq b_n} r(x_n(s), x(t)) \wedge r(x_n(s), x(t-)) < \frac{\delta \wedge \varepsilon}{2},$$

and $a_n, b_n \in (t - \varepsilon, t + \varepsilon)$. Let $n \geq n_0$ and define

$$(6.20) \quad s_n = \inf \left\{ s > a_n : r(x_n(s), x(t)) < \frac{\delta \wedge \varepsilon}{2} \right\}.$$

By (6.18), $a_n < s_n \leq b_n$, and therefore $s_n \in (t - \varepsilon, t + \varepsilon)$, $r(x_n(s_n), x(t)) \leq (\delta \wedge \varepsilon)/2$, and $r(x_n(s_n-), x(t)) \geq (\delta \wedge \varepsilon)/2$. The latter inequality, together with (6.19), implies that $r(x_n(s_n-), x(t-)) < (\delta \wedge \varepsilon)/2$. We conclude that $s_n \in \Gamma(t, n, \varepsilon)$ for all $n \geq n_0$, and this contradiction establishes the claim that $\lim_{n \rightarrow \infty} \varepsilon_n = 0$.

For each n , we construct $\lambda_n \in \Lambda'$ as follows. Choose a partition $0 = t_0^n < t_1^n < \dots < t_{m_n}^n < T \leq t_{m_n}^n$ with $\min_{1 \leq i \leq m_n} (t_i^n - t_{i-1}^n) > 3\varepsilon_n$ such that

$$(6.21) \quad \max_{1 \leq i \leq m_n} \sup_{s, t \in [t_{i-1}^n, t_i^n]} r(x(s), x(t)) \leq w'(x, 3\varepsilon_n, T) + \varepsilon_n,$$

and put $m_n^* = \max \{i \geq 0 : t_i^n \leq T\}$ (m_n^* is $m_n - 1$ or m_n). Define $\lambda_n(0) = 0$ and $\lambda_n(t_i^n) = \inf \Gamma(t_i^n, n, \varepsilon_n)$ for $i = 1, \dots, m_n^*$, interpolate linearly on $[0, t_{m_n^*}^n]$, and let $\lambda_n(t) = t - t_{m_n^*}^n + \lambda_n(t_{m_n^*}^n)$ for all $t > t_{m_n^*}^n$. Then $\lambda_n \in \Lambda'$ and $\sup_{t \geq 0} |\lambda_n(t) - t| \leq \varepsilon_n$.

We claim that $\lim_{n \rightarrow \infty} \sup_{0 \leq t \leq T} r(x_n(\lambda_n(t)), x(t)) = 0$ and hence $\lim_{n \rightarrow \infty} d(x_n, x) = 0$ by Proposition 5.3 (see Remark 5.4). To verify the claim, it is enough

to show that if $\{t_n\} \subset [0, T]$, $0 \leq t \leq T$, and $\lim_{n \rightarrow \infty} t_n = t$, then $\lim_{n \rightarrow \infty} r(x_n(\lambda_n(t_n)), x(t_n)) = 0$. If $x(t) = x(t-)$, the result follows from condition (a) since $\lim_{n \rightarrow \infty} \lambda_n(t_n) = t$. Therefore, let us suppose that $x(t) \neq x(t-)$. Then, for each n sufficiently large, $t = t_{i_n}^n$ for some $i_n \in \{1, \dots, m_n^*\}$ by (6.21) and Lemma 6.2(a). To complete the proof, it suffices to consider two cases, $\{t_n\} \subset [t, T]$ and $\{t_n\} \subset [0, t]$. In the first case, $\lambda_n(t_n) \geq \lambda_n(t) = \lambda_n(t_{i_n}^n)$ and $r(x_n(\lambda_n(t_n)), x(t)) \leq \varepsilon_n$ for each n sufficiently large, so $\lim_{n \rightarrow \infty} r(x_n(\lambda_n(t_n)), x(t)) = 0$ by condition (b), and the desired result follows. In the second case, $\lambda_n(t_n) < \lambda_n(t) = \lambda_n(t_{i_n}^n)$ and either $r(x_n(\lambda_n(t_n)-), x(t-)) < \varepsilon_n$ or $r(x_n(\lambda_n(t_n)), x(t-)) \leq \varepsilon_n$ (depending on whether the infimum in the definition of $\lambda_n(t_n)$ is attained or not) for each n sufficiently large. Consequently, for such n , there exists s_n with $\lambda_n(t_n) < s_n \leq \lambda_n(t_{i_n}^n)$ such that $r(x_n(s_n), x(t-)) \leq \varepsilon_n$, and therefore $\lim_{n \rightarrow \infty} r(x_n(\lambda_n(t_n)), x(t-)) = 0$ by condition (c), from which the desired result follows. This completes the proof. \square

7. CONVERGENCE IN DISTRIBUTION IN $D_E[0, \infty)$

As in the previous two sections, (E, r) denotes a metric space. Let \mathcal{S}_E denote the Borel σ -algebra of $D_E[0, \infty)$. We are interested in weak convergence of elements of $\mathcal{P}(D_E[0, \infty))$ and naturally it is important to know more about \mathcal{S}_E . The following result states that \mathcal{S}_E is just the σ -algebra generated by the coordinate random variables.

7.1 Proposition For each $t \geq 0$, define $\pi_t : D_E[0, \infty) \rightarrow E$ by $\pi_t(x) = x(t)$. Then

$$(7.1) \quad \mathcal{S}_E \supset \mathcal{S}'_E \equiv \sigma(\pi_t : 0 \leq t < \infty) = \sigma(\pi_t : t \in D),$$

where D is any dense subset of $[0, \infty)$. If E is separable, then $\mathcal{S}_E = \mathcal{S}'_E$.

Proof. For each $\varepsilon > 0$, $t \geq 0$, and $f \in \bar{C}(E)$,

$$(7.2) \quad f_\varepsilon^t(x) \equiv \frac{1}{\varepsilon} \int_t^{t+\varepsilon} f(\pi_s(x)) ds$$

defines a continuous function f_ε^t on $D_E[0, \infty)$. Since $\lim_{s \rightarrow t} f_\varepsilon^t(x) = f(\pi_t(x))$ for every $x \in D_E[0, \infty)$, we find that $f \circ \pi_t$ is Borel measurable for every $f \in \bar{C}(E)$ and hence for every $f \in B(E)$. Consequently,

$$(7.3) \quad \pi_t^{-1}(\Gamma) = \{x \in D_E[0, \infty) : \chi_\Gamma(\pi_t(x)) = 1\} \in \mathcal{S}_E$$

for all $\Gamma \in B(E)$, and hence $\mathcal{S}_E \supset \mathcal{S}'_E$. For each $t \geq 0$, there exists $\{t_n\} \subset D \cap [t, \infty)$ such that $\lim_{n \rightarrow \infty} t_n = t$. Therefore, $\pi_t = \lim_{n \rightarrow \infty} \pi_{t_n}$ is $\sigma(\pi_s : s \in D)$ -measurable, and hence we have (7.1).

Assume now that E is separable. Let $n \geq 1$, let $0 = t_0 < t_1 < \dots < t_n < t_{n+1} = \infty$, and for $\alpha_0, \alpha_1, \dots, \alpha_n \in E$ define $\eta(\alpha_0, \alpha_1, \dots, \alpha_n) \in D_E[0, \infty)$ by

$$(7.4) \quad \eta(\alpha_0, \alpha_1, \dots, \alpha_n)(t) = \alpha_i, \quad t_i \leq t < t_{i+1}, \quad i = 0, 1, \dots, n.$$

Since

$$(7.5) \quad d(\eta(\alpha_0, \alpha_1, \dots, \alpha_n), \eta(\alpha'_0, \alpha'_1, \dots, \alpha'_n)) \leq \max_{0 \leq i \leq n} r(\alpha_i, \alpha'_i),$$

η is a continuous function from E^{n+1} into $D_E[0, \infty)$. Since each π_i is \mathcal{S}'_E -measurable and E is separable, we have that for fixed $z \in D_E[0, \infty)$, $d(z, \eta \circ (\pi_{t_0}, \pi_{t_1}, \dots, \pi_{t_n}))$ is an \mathcal{S}'_E -measurable function from $D_E[0, \infty)$ into \mathbb{R} . Finally, for $m = 1, 2, \dots$, let η_m be defined as was η with $t_i = i/m$, $i = 0, 1, \dots, n \equiv m^2$. Then

$$(7.6) \quad \lim_{m \rightarrow \infty} d(z, \eta_m(\pi_{t_0}(x), \dots, \pi_{t_{m^2}}(x))) = d(z, x)$$

for every $x \in D_E[0, \infty)$ (see Problem 12), so $d(z, x)$ is \mathcal{S}'_E -measurable in x for fixed $z \in D_E[0, \infty)$. In particular, every open ball $B(z, \varepsilon) = \{x \in D_E[0, \infty) : d(z, x) < \varepsilon\}$ belongs to \mathcal{S}'_E , so since E (and, by Theorem 5.6, $D_E[0, \infty)$) is separable, \mathcal{S}'_E contains all open sets in $D_E[0, \infty)$ and hence contains \mathcal{S}_E . \square

A $D_E[0, \infty)$ -valued random variable is a stochastic process with sample paths in $D_E[0, \infty)$, although the converse need not be true if E is not separable. Let $\{X_\alpha\}$ (where α ranges over some index set) be a family of stochastic processes with sample paths in $D_E[0, \infty)$ (if E is not separable, assume the X_α are $D_E[0, \infty)$ -valued random variables), and let $\{P_\alpha\} \subset \mathcal{P}(D_E[0, \infty))$ be the family of associated probability distributions (i.e., $P_\alpha(B) = P\{X_\alpha \in B\}$ for all $B \in \mathcal{S}_E$). We say that $\{X_\alpha\}$ is *relatively compact* if $\{P_\alpha\}$ is (i.e., if the closure of $\{P_\alpha\}$ in $\mathcal{P}(D_E[0, \infty))$ is compact). Theorem 6.3 gives, through an application of Prohorov's theorem, criteria for $\{X_\alpha\}$ to be relatively compact.

7.2 Theorem Let (E, r) be complete and separable, and let $\{X_\alpha\}$ be a family of processes with sample paths in $D_E[0, \infty)$. Then $\{X_\alpha\}$ is relatively compact if and only if the following two conditions hold:

- (a) For every $\eta > 0$ and rational $t \geq 0$, there exists a compact set $\Gamma_{\eta, t} \subset E$ such that

$$(7.7) \quad \inf_a P\{X_\alpha(t) \in \Gamma_{\eta, t}^a\} \geq 1 - \eta.$$

- (b) For every $\eta > 0$ and $T > 0$, there exists $\delta > 0$ such that

$$(7.8) \quad \sup_a P\{w'(X_\alpha, \delta, T) \geq \eta\} \leq \eta.$$

7.3 Remark In fact, if $\{X_n\}$ is relatively compact, then the stronger *compact containment condition* holds; that is, for every $\eta > 0$ and $T > 0$ there is a compact set $\Gamma_{\eta, T} \subset E$ such that

$$(7.9) \quad \inf_a P\{X_n(t) \in \Gamma_{\eta, T} \text{ for } 0 \leq t \leq T\} \geq 1 - \eta. \quad \square$$

Proof. If $\{X_n\}$ is relatively compact, then Theorems 5.6, 2.2, and 6.3 immediately yield (a) and (b); in fact, $\Gamma_{\eta, T}^n$ can be replaced by $\Gamma_{\eta, T}$ in (7.7).

Conversely, let $\varepsilon > 0$, let T be a positive integer such that $e^{-T} < \varepsilon/2$, and choose $\delta > 0$ such that (7.8) holds with $\eta = \varepsilon/4$. Let $m > 1/\delta$, put $\Gamma = \bigcup_{i=0}^{mT} \Gamma_{\varepsilon 2^{-i-2}, i/m}$, and observe that

$$(7.10) \quad \inf_a P\{X_n(i/m) \in \Gamma^{\varepsilon/4}, \quad i = 0, 1, \dots, mT\} \geq 1 - \frac{\varepsilon}{2}.$$

Using the notation of Lemma 6.1, let $A = A(\Gamma, \delta)$. By the lemma, A has compact closure.

Given $x \in D_E[0, \infty)$ with $w'(x, \delta, T) < \varepsilon/4$ and $x(i/m) \in \Gamma^{\varepsilon/4}$ for $i = 0, 1, \dots, mT$, choose $0 = t_0 < t_1 < \dots < t_{n-1} < T \leq t_n$ such that $\min_{1 \leq i \leq n} (t_i - t_{i-1}) > \delta$ and

$$(7.11) \quad \max_{1 \leq i \leq n} \sup_{s, t \in [t_{i-1}, t_i]} r(x(s), x(t)) < \frac{\varepsilon}{4},$$

and select $\{y_i\} \subset \Gamma$ such that $r(x(i/m), y_i) < \varepsilon/4$ for $i = 0, 1, \dots, mT$. Defining $x' \in A$ by

$$(7.12) \quad x'(t) = \begin{cases} y_{\lfloor mt_{i-1} \rfloor + 1}, & t_{i-1} \leq t < t_i, \quad i = 1, \dots, n-1, \\ y_{\lfloor mt_{n-1} \rfloor + 1}, & t \geq t_{n-1}, \end{cases}$$

we have $\sup_{0 \leq t \leq T} r(x(t), x'(t)) < \varepsilon/2$ and hence $d(x, x') < \varepsilon/2 + e^{-T} < \varepsilon$, implying that $x \in A^\varepsilon$. Consequently, $\inf_a P\{X_n \in A^\varepsilon\} \geq 1 - \varepsilon$, so the relative compactness of $\{X_n\}$ follows from Theorems 5.6 and 2.2. \square

7.4 Corollary Let (E, r) be complete and separable, and let $\{X_n\}$ be a sequence of processes with sample paths in $D_E[0, \infty)$. Then $\{X_n\}$ is relatively compact if and only if the following two conditions hold:

- (a) For every $\eta > 0$ and rational $t \geq 0$, there exists a compact set $\Gamma_{\eta, t} \subset E$ such that

$$(7.13) \quad \lim_{n \rightarrow \infty} P\{X_n(t) \in \Gamma_{\eta, t}\} \geq 1 - \eta.$$

- (b) For every $\eta > 0$ and $T > 0$, there exists $\delta > 0$ such that

$$(7.14) \quad \overline{\lim}_{n \rightarrow \infty} P\{w'(X_n, \delta, T) \geq \eta\} \leq \eta.$$

Proof. Fix $\eta > 0$, rational $t \geq 0$, and $T > 0$. For each $n \geq 1$, there exist by Lemmas 2.1 and 6.2(a) a compact set $\Gamma_n \subset E$ and $\delta_n > 0$ such that $P\{X_n(t) \in \Gamma_n\} \geq 1 - \eta$ and $P\{w'(X_n, \delta_n, T) \geq \eta\} \leq \eta$. By (7.13) and (7.14), there exist a compact set $\Gamma_0 \subset E$, $\delta_0 > 0$, and a positive integer n_0 such that

$$(7.15) \quad \inf_{n \geq n_0} P\{X_n(t) \in \Gamma_0\} \geq 1 - \eta$$

and

$$(7.16) \quad \sup_{n \geq n_0} P\{w'(X_n, \delta_0, T) \geq \eta\} \leq \eta.$$

We can replace n_0 in (7.15) and (7.16) by 1 if we replace Γ_0 by $\Gamma = \bigcup_{n=0}^{n_0-1} \Gamma_n$ and δ_0 by $\delta = \bigwedge_{n=0}^{n_0-1} \delta_n$, so the result follows from Theorem 7.2. \square

7.5 Lemma Let (E, r) be arbitrary, let $\Gamma_1 \subset \Gamma_2 \subset \dots$ be a nondecreasing sequence of compact subsets of E , and define

$$(7.17) \quad S = \{x \in D_E[0, \infty) : x(t) \in \Gamma_n \text{ for } 0 \leq t \leq n, n = 1, 2, \dots\}.$$

Let $\{X_n\}$ be a family of processes with sample paths in S . Then $\{X_n\}$ is relatively compact if condition (b) of Theorem 7.2 holds.

Proof. The proof is similar to that of Theorem 7.2. Let $\varepsilon > 0$, let T be a positive integer such that $e^{-T} < \varepsilon/2$, choose $\delta > 0$ such that (7.8) holds with $\eta = \varepsilon/2$, and let $A = A(\Gamma_T, \delta)$. Given $x \in S$ with $w'(x, \delta, T) < \varepsilon/2$, it is easy to construct $x' \in A \cap S$ with $d(x, x') < \varepsilon$, and hence $x \in (A \cap S)^c$. Consequently, $\inf_n P\{X_n \in (A \cap S)^c\} \geq 1 - \varepsilon$, so the relative compactness of $\{X_n\}$ follows from Lemma 6.1 and Theorem 2.2. Here we are using the fact that (S, d) is complete and separable (Problem 15). \square

7.6 Theorem Let (E, r) be arbitrary, and let $\{X_n\}$ be a family of processes with sample paths in $D_E[0, \infty)$. If the compact containment condition (Remark 7.3) and condition (b) of Theorem 7.2 hold, then the X_n have modifications \tilde{X}_n that are $D_E[0, \infty)$ -valued random variables and $\{\tilde{X}_n\}$ is relatively compact.

Proof. By the compact containment condition there exist compact sets $\Gamma_n \subset E$, $n = 1, 2, \dots$, such that $\inf_n P\{X_n(t) \in \Gamma_n \text{ for } 0 \leq t \leq n\} \geq 1 - n^{-1}$. Let $E_0 = \bigcup_n \Gamma_n$. Note that E_0 is separable and $P\{X_n(t) \in E_0\} = 1$ so X_n has a modification with sample paths in $D_{E_0}[0, \infty)$. Consequently, we may as well assume E is separable. Given $\eta > 0$, we can assume without loss of generality that $\{\Gamma_{n+2-n}\}$ is a nondecreasing sequence of compact subsets of E . Define

$$(7.18) \quad S_\eta = \{x \in D_E[0, \infty) : x(t) \in \Gamma_{n+2-n} \text{ for } 0 \leq t \leq n, n = 1, 2, \dots\},$$

and note that $\inf_n P\{X_n \in S_\eta\} \geq 1 - \eta$. By Lemma 7.4, the family $\{P_n^\eta\} \subset \mathcal{P}(S_\eta)$, defined by

$$(7.19) \quad P_n^\eta(B) = P\{X_n \in B | X_n \in S_\eta\},$$

is relatively compact. The proof proceeds analogously to that of Corollary 2.3. We leave the details to the reader. \square

7.7 Lemma If X is a process with sample paths in $D_E[0, \infty)$, then the complement in $[0, \infty)$ of

$$(7.20) \quad D(X) \equiv \{t \geq 0 : P\{X(t) = X(t-)\} = 1\}$$

is at most countable.

Proof. Let $\varepsilon > 0$, $\delta > 0$, and $T > 0$. If the set

$$(7.21) \quad \{0 \leq t \leq T : P\{r(X(t), X(t-)) \geq \varepsilon\} \geq \delta\}$$

contains a sequence $\{t_n\}$ of distinct points, then

$$(7.22) \quad P\{r(X(t_n), X(t_n-)) \geq \varepsilon \text{ infinitely often}\} \geq \delta > 0,$$

contradicting the fact that, for each $x \in D_E[0, \infty)$, $r(x(t), x(t-)) \geq \varepsilon$ for at most finitely many $t \in [0, T]$. Hence the set in (7.21) is finite, and therefore

$$(7.23) \quad \{t \geq 0 : P\{r(X(t), X(t-)) \geq \varepsilon\} > 0\}$$

is at most countable. The conclusion follows by letting $\varepsilon \rightarrow 0$. \square

7.8 Theorem Let E be separable and let X_n , $n = 1, 2, \dots$, and X be processes with sample paths in $D_E[0, \infty)$.

(a) If $X_n \Rightarrow X$, then

$$(7.24) \quad (X_n(t_1), \dots, X_n(t_k)) \Rightarrow (X(t_1), \dots, X(t_k))$$

for every finite set $\{t_1, \dots, t_k\} \subset D(X)$. Moreover, for each finite set $\{t_1, \dots, t_k\} \subset [0, \infty)$, there exist sequences $\{t_1^n\} \subset [t_1, \infty)$, \dots , $\{t_k^n\} \subset [t_k, \infty)$ converging to t_1, \dots, t_k , respectively, such that $(X_n(t_1^n), \dots, X_n(t_k^n)) \Rightarrow (X(t_1), \dots, X(t_k))$.

(b) If $\{X_n\}$ is relatively compact and there exists a dense set $D \subset [0, \infty)$ such that (7.24) holds for every finite set $\{t_1, \dots, t_k\} \subset D$, then $X_n \Rightarrow X$.

Proof. (a) Suppose that $X_n \Rightarrow X$. By Theorem 1.8, there exists a probability space on which are defined processes Y_n , $n = 1, 2, \dots$, and Y with sample paths in $D_E[0, \infty)$ and with the same distributions as X_n , $n = 1, 2, \dots$, and X , such that $\lim_{n \rightarrow \infty} d(Y_n, Y) = 0$ a.s. If $t \in D(X) = D(Y)$, then, using the

notation of Proposition 7.1, π_n is continuous a.s. with respect to the distribution of Y , so $\lim_{n \rightarrow \infty} Y_n(t) = Y(t)$ a.s. by Corollary 1.9, and the first conclusion follows. We leave it to the reader to show that the second conclusion is a consequence of the first, together with Lemma 7.7.

(b) It suffices to show that every convergent (in distribution) subsequence of $\{X_n\}$ converges in distribution to X . Relabeling if necessary, suppose that $X_n \Rightarrow Y$. We must show that X and Y have the same distribution. Let $\{t_1, \dots, t_k\} \subset D(Y)$ and $f_1, \dots, f_k \in \bar{C}(E)$, and choose sequences $\{t_i^m\} \subset D \cap [t_1, \infty), \dots, \{t_k^m\} \subset D \cap [t_k, \infty)$ converging to t_1, \dots, t_k , respectively, and $n_1 < n_2 < n_3 < \dots$ such that

$$(7.25) \quad \left| E \left[\prod_{i=1}^k f_i(X(t_i^m)) \right] - E \left[\prod_{i=1}^k f_i(X_{n_m}(t_i^m)) \right] \right| < \frac{1}{m}.$$

Then

$$\begin{aligned} (7.26) \quad & \left| E \left[\prod_{i=1}^k f_i(X(t_i)) \right] - E \left[\prod_{i=1}^k f_i(Y(t_i)) \right] \right| \\ & \leq \left| E \left[\prod_{i=1}^k f_i(X(t_i)) \right] - E \left[\prod_{i=1}^k f_i(X(t_i^m)) \right] \right| \\ & \quad + \left| E \left[\prod_{i=1}^k f_i(X(t_i^m)) \right] - E \left[\prod_{i=1}^k f_i(X_{n_m}(t_i^m)) \right] \right| \\ & \quad + \left| E \left[\prod_{i=1}^k f_i(X_{n_m}(t_i^m)) \right] - E \left[\prod_{i=1}^k f_i(Y(t_i)) \right] \right| \end{aligned}$$

for each $m \geq 1$. All three terms on the right tend to zero as $m \rightarrow \infty$, the first by the right continuity of X , the second by (7.25), and the third by the facts that $X_{n_m} \Rightarrow Y$ and $\{t_1, \dots, t_k\} \subset D(Y)$. Consequently,

$$(7.27) \quad E \left[\prod_{i=1}^k f_i(X(t_i)) \right] = E \left[\prod_{i=1}^k f_i(Y(t_i)) \right]$$

for all $\{t_1, \dots, t_k\} \subset [0, \infty)$ (by Lemma 7.7 and right continuity) and all $f_1, \dots, f_k \in \bar{C}(E)$. By Proposition 7.1 and the Dynkin class theorem (Appendix 4), we conclude that X and Y have the same distribution. \square

8. CRITERIA FOR RELATIVE COMPACTNESS IN $D_E[0, \infty)$

Let (E, r) denote a metric space and $q = r \wedge 1$. We now consider a systematic way of selecting the partition in the definition of $w(x, \delta, T)$. Given $\varepsilon > 0$ and $x \in D_E[0, \infty)$, define $\tau_0 = \sigma_0 = 0$ and, for $k = 1, 2, \dots$,

$$(8.1) \quad \tau_k = \inf \left\{ t > \tau_{k-1} : r(x(t), x(\tau_{k-1})) > \frac{\varepsilon}{2} \right\}$$

if $\tau_{k-1} < \infty$, $\tau_k = \infty$ if $\tau_{k-1} = \infty$,

$$(8.2) \quad \sigma_k = \sup \left\{ t \leq \tau_k : r(x(t), x(\tau_k)) \vee r(x(t-), x(\tau_k)) \geq \frac{\varepsilon}{2} \right\}$$

if $\tau_k < \infty$, and $\sigma_k = \infty$ if $\tau_k = \infty$. Given $\delta > 0$ and $T > 0$, observe that $w'(x, \delta, T) < \varepsilon/2$ implies $\min \{ \tau_{k+1} - \sigma_k : \tau_k < T \} > \delta$, for if $\tau_{k+1} - \sigma_k \leq \delta$ and $\tau_k < T$ for some $k \geq 0$, then any interval $[a, b]$ containing τ_k with $b - a > \delta$ must also contain σ_k or τ_{k+1} (or both) in its interior and hence must satisfy $\sup_{s, t \in [a, b]} r(x(s), x(t)) \geq \varepsilon/2$; in this case, $w'(x, \delta, T) \geq \varepsilon/2$.

Letting

$$(8.3) \quad s_k = \frac{\sigma_k + \tau_k}{2}$$

for $k = 0, 1, \dots$, we have $\lim_{k \rightarrow \infty} s_k = \infty$. Observe that, for each $k \geq 0$,

$$(8.4) \quad \sigma_k \leq s_k \leq \tau_k \leq \sigma_{k+1} \leq s_{k+1} \leq \tau_{k+1},$$

and

$$(8.5) \quad s_{k+1} - s_k \geq \frac{\tau_k + \tau_{k+1}}{2} - \frac{\sigma_k + \tau_k}{2} = \frac{\tau_{k+1} - \sigma_k}{2}$$

if $s_k < \infty$, where the middle inequality in (8.4) follows from the fact that $r(x(\tau_k), x(\tau_{k+1})) \geq \varepsilon/2$ if $\tau_{k+1} < \infty$. We conclude from (8.5) that $\min \{ \tau_{k+1} - \sigma_k : \tau_k < T + \delta/2 \} > \delta$ implies

$$(8.6) \quad \min \{ s_{k+1} - s_k : s_k < T \} > \frac{\delta}{2};$$

for if not, there would exist $k \geq 0$ with $s_k < T$, $\tau_k \geq T + \delta/2$, and $s_{k+1} - s_k \leq \delta/2$, a contradiction by (8.4). Finally, (8.6) implies $w'(x, \delta/2, T) \leq \varepsilon$.

Let us now regard τ_k , σ_k , and s_k , $k = 0, 1, \dots$, as functions from $D_E[0, \infty)$ into $[0, \infty]$. Assuming that E is separable (recall Remark 3.4), their \mathcal{S}_E -measurability follows easily from the identities

(8.7)

$$\{\tau_k < u\} = \{\tau_{k-1} < \infty\} \cap \bigcup_{t \in (0, u) \cap Q} \left(\left\{ r(x(t), x(\tau_{k-1})) > \frac{\varepsilon}{2} \right\} \cap \{t > \tau_{k-1}\} \right)$$

and

$$(8.8) \quad \{\sigma_k \geq u\} = \{\tau_k = \infty\} \cup \left(\{\tau_k < \infty\} \cap \left[\left\{ r(x(u-), x(\tau_k)) \geq \frac{\varepsilon}{2} \right\} \cup \bigcap_{n=1}^{\infty} \bigcup_{t \in (u, \infty) \cap Q} \left(\left\{ r(x(t), x(\tau_k)) > \frac{\varepsilon}{2} - \frac{1}{n} \right\} \cap \{t \leq \tau_k\} \right) \right] \right),$$

valid for $0 < u < \infty$ and $k = 1, 2, \dots$. We summarize the implications of the two preceding paragraphs for our purposes in the following lemma.

8.1 Lemma Let (E, r) be separable, and let $\{X_\alpha\}$ be a family of processes with sample paths in $D_E[0, \infty)$. Let $\tau_k^{\alpha, \epsilon}$, $\sigma_k^{\alpha, \epsilon}$, and $s_k^{\alpha, \epsilon}$, $k = 0, 1, \dots$, be defined for given $\epsilon > 0$ and X_α as in (8.1)–(8.3). Then the following are equivalent:

$$(8.9) \quad \liminf_{\delta \rightarrow 0} \liminf_{\alpha} P\{w'(X_\alpha, \delta, T) < \epsilon\} = 1, \quad \epsilon > 0, \quad T > 0.$$

$$(8.10) \quad \liminf_{\delta \rightarrow 0} \liminf_{\alpha} P\{\min\{s_{k+1}^{\alpha, \epsilon} - s_k^{\alpha, \epsilon}; s_k^{\alpha, \epsilon} < T\} \geq \delta\} = 1, \quad \epsilon > 0, \quad T > 0.$$

$$(8.11) \quad \liminf_{\delta \rightarrow 0} \liminf_{\alpha} P\{\min\{\tau_{k+1}^{\alpha, \epsilon} - \sigma_k^{\alpha, \epsilon}; \tau_k^{\alpha, \epsilon} < T\} \geq \delta\} = 1, \quad \epsilon > 0, \quad T > 0.$$

Proof. See the discussion above. \square

8.2. Lemma For each α , let $0 = s_0^\alpha < s_1^\alpha < s_2^\alpha < \dots$ be a sequence of random variables with $\lim_{k \rightarrow \infty} s_k^\alpha = \infty$, define $\Delta_k^\alpha = s_{k+1}^\alpha - s_k^\alpha$ for $k = 0, 1, \dots$, let $T > 0$, and put $K_\alpha(T) = \max\{k \geq 0: s_k^\alpha < T\}$. Define $F: [0, \infty) \rightarrow [0, 1]$ by $F(t) = \sup_\alpha \sup_{k \geq 0} P\{\Delta_k^\alpha < t, s_k^\alpha < T\}$. Then

$$(8.12) \quad F(\delta) \leq \sup_\alpha P\left\{ \min_{0 \leq k \leq K_\alpha(T)} \Delta_k^\alpha < \delta \right\} \leq LF(\delta) + e^T \int_0^\infty L e^{-Lt} F(t) dt$$

for all $\delta > 0$ and $L = 1, 2, \dots$. Consequently,

$$(8.13) \quad \limsup_{\delta \rightarrow 0} P\left\{ \min_{0 \leq k \leq K_\alpha(T)} \Delta_k^\alpha < \delta \right\} = 0$$

if and only if $F(0+) = 0$.

Proof. The first inequality in (8.12) is immediate. As for the second,

$$\begin{aligned} (8.14) \quad P\left\{ \min_{0 \leq k \leq K_\alpha(T)} \Delta_k^\alpha < \delta \right\} &\leq \sum_{k=0}^{L-1} P\{\Delta_k^\alpha < \delta, s_k^\alpha < T\} + P\{K_\alpha(T) \geq L\} \\ &\leq LF(\delta) + e^T E\left[\chi_{\{K_\alpha(T) \geq L\}} \exp\left(-\sum_{k=0}^{L-1} \Delta_k^\alpha\right) \right] \\ &\leq LF(\delta) + e^T \prod_{k=0}^{L-1} \{E[\chi_{\{s_k^\alpha < T\}} \exp(-L\Delta_k^\alpha)]\}^{1/L} \\ &\leq LF(\delta) + e^T \int_0^\infty L e^{-Lt} F(t) dt. \end{aligned}$$

Finally, observe that $F(0+) = 0$ implies that the right side of (8.12) approaches zero as $\delta \rightarrow 0$ and then $L \rightarrow \infty$. \square

8.3 Proposition Under the assumptions of Lemma 8.1, (8.9) is equivalent to

$$(8.15) \quad \lim_{\delta \rightarrow 0} \sup_{\sigma} \sup_{k \geq 0} P\{\tau_{k+1}^{\sigma, \epsilon} - \sigma_k^{\sigma, \epsilon} < \delta, \tau_k^{\sigma, \epsilon} < T\} = 0, \quad \epsilon > 0, \quad T > 0.$$

Proof. The result follows from Lemmas 8.1 and 8.2 together with the inequalities

$$\begin{aligned} (8.16) \quad & P\left\{s_{k+1}^{\sigma, \epsilon} - s_k^{\sigma, \epsilon} < \frac{\delta}{2}, s_k^{\sigma, \epsilon} < T\right\} \\ & \leq P\{\tau_{k+1}^{\sigma, \epsilon} - \sigma_k^{\sigma, \epsilon} < \delta, \tau_k^{\sigma, \epsilon} < T + \delta\} \\ & \leq P\{s_{k+1}^{\sigma, \epsilon} - s_k^{\sigma, \epsilon} < \delta, s_k^{\sigma, \epsilon} < T + \delta\}. \end{aligned} \quad \square$$

The following lemma gives us a means of estimating the probabilities in (8.15). Let $S(T)$ denote the collection of all $\{\mathcal{F}_{t+}^X\}$ -stopping times bounded by T .

8.4 Lemma Let (E, r) be separable, let X be a process with sample paths in $D_E[0, \infty)$, and fix $T > 0$ and $\beta > 0$. Then, for each $\delta > 0$, $\lambda > 0$, and $\tau \in S(T)$,

$$\begin{aligned} (8.17) \quad & P\left\{\sup_{0 \leq u \leq \delta} q(X(\tau + u), X(\tau)) \geq \lambda, \sup_{0 \leq v \leq \delta \wedge \tau} q(X(\tau), X(\tau - v)) \geq \lambda\right\} \\ & \leq \lambda^{-2\beta} [a_\beta + 2a_\beta^2(a_\beta + 4a_\beta^2)]C(\delta) \end{aligned}$$

and

$$\begin{aligned} (8.18) \quad & P\left\{\sup_{0 \leq u \leq \delta} q(X(u), X(0)) \geq \lambda\right\} \\ & \leq \lambda^{-2\beta} \{a_\beta(a_\beta + 4a_\beta^2)C(\delta) + a_\beta E[q^\beta(X(\delta), X(0))]\}, \end{aligned}$$

where

$$(8.19)$$

$$C(\delta) = \sup_{\tau \in S(T+2\delta)} \sup_{0 \leq u \leq 2\delta} E\left[\sup_{0 \leq v \leq 3\delta \wedge \tau} q^\beta(X(\tau + u), X(\tau))q^\beta(X(\tau), X(\tau - v))\right]$$

and $a_\beta = 2^{(\beta-1)\vee 0}$ (and hence $(c+d)^\beta \leq a_\beta(c^\beta + d^\beta)$ for all $c, d \geq 0$).

8.5 Remark (a) In (8.19), $\sup_{\tau \in S(T+2\delta)}$ can be replaced by $\sup_{\tau \in S_0(T+2\delta)}$, where $S_0(T+2\delta)$ is the collection of all discrete $\{\mathcal{F}_t^X\}$ -stopping times bounded by $T+2\delta$. This follows from the fact that for each $\tau \in S(T+2\delta)$ there exists a sequence $\{\tau_n\} \subset S_0(T+2\delta)$ such that $\tau_n \geq \tau$ for each n and $\lim_{n \rightarrow \infty} \tau_n = \tau$; we also need the observation that for fixed $x \in D_E[0, \infty)$, $\sup_{0 \leq v \leq 3\delta \wedge \tau} q^\theta(x(t), x(t-v))$ is right continuous in t ($0 \leq t < \infty$).

(b) If we take $\lambda = \varepsilon/2 \in (0, 1]$ and $\tau = \tau_k \wedge T$ (recall (8.1)) in Lemma 8.4, where $k \geq 1$, then the left side of (8.17) bounds

$$(8.20) \quad P\{\tau_{k+1} - \tau_k \leq \delta, \tau_k - \sigma_k < \delta, \tau_k < T\},$$

which for each $k \geq 1$ bounds $P\{\tau_{k+1} - \sigma_k < \delta, \tau_k < T, \tau_1 > \delta\}$. The left side of (8.18) bounds $P\{\tau_1 \leq \delta\}$, and hence the sum of the right sides of (8.17) and (8.18) bounds $P\{\tau_{k+1} - \sigma_k < \delta, \tau_k < T\}$ for each $k \geq 0$. \square

Proof. Given a $\{\mathcal{F}_{t+}^X\}$ -stopping time τ , let $M_\tau(\delta)$ be the collection of \mathcal{F}_{t+}^X -measurable random variables U satisfying $0 \leq U \leq \delta$. We claim that

$$(8.21) \quad \sup_{\tau \in S(T+\delta)} \sup_{U \in M_\tau(\delta)} E \left[\sup_{0 \leq v \leq 2\delta \wedge \tau} q^\theta(X(\tau+U), X(\tau)) q^\theta(X(\tau), X(\tau-v)) \right] \leq (a_\beta + 4a_\beta^2) C(\delta).$$

To see this, observe that for each $\tau \in S(T+\delta)$ and $U \in M_\tau(\delta)$,

$$(8.22) \quad q^\theta(X(\tau+U), X(\tau)) \leq a_\beta \delta^{-1} \int_\delta^{2\delta} [q^\theta(X(\tau+\theta), X(\tau)) + q^\theta(X(\tau+\theta), X(\tau+U))] d\theta \\ \leq a_\beta \delta^{-1} \left[\int_\delta^{2\delta} q^\theta(X(\tau+\theta), X(\tau)) d\theta + \int_0^{2\delta} q^\theta(X(\tau+U+\theta), X(\tau+U)) d\theta \right],$$

and hence

$$(8.23) \quad \sup_{0 \leq v \leq 2\delta \wedge \tau} q^\theta(X(\tau+U), X(\tau)) q^\theta(X(\tau), X(\tau-v)) \leq a_\beta \delta^{-1} \int_\delta^{2\delta} \sup_{0 \leq v \leq 2\delta \wedge \tau} q^\theta(X(\tau+\theta), X(\tau)) q^\theta(X(\tau), X(\tau-v)) d\theta \\ + a_\beta^2 \delta^{-1} \int_0^{2\delta} \sup_{0 \leq v \leq 2\delta \wedge \tau} q^\theta(X(\tau+U+\theta), X(\tau+U)) \\ \times q^\theta(X(\tau+U), X(\tau-v)) d\theta \\ + a_\beta^2 \delta^{-1} \int_0^{2\delta} q^\theta(X(\tau+U+\theta), X(\tau+U)) q^\theta(X(\tau+U), X(\tau)) d\theta$$

$$\leq a_\beta \delta^{-1} \int_0^{2\delta} \sup_{0 \leq v \leq 2\delta \wedge \tau} q^\theta(X(\tau + \theta), X(\tau)) q^\theta(X(\tau), X(\tau - v)) d\theta \\ + 2a_\beta^2 \delta^{-1} \int_0^{2\delta} \sup_{0 \leq v \leq 3\delta \wedge (\tau+U)} q^\theta(X(\tau + U + \theta), X(\tau + U)) \\ \times q^\theta(X(\tau + U), X(\tau + U - v)) d\theta;$$

also, $\tau + U \in S(T + 2\delta)$, so (8.21) follows from (8.23).

Given $0 < \eta < \lambda$ and $\tau \in S(T)$, define

$$(8.24) \quad \Delta = \inf \{t > 0 : q(X(\tau + t), X(\tau)) > \lambda - \eta\},$$

and observe that

$$(8.25) \quad q^\theta(X(\tau + \Delta \wedge \delta), X(\tau)) q^\theta(X(\tau), X(\tau - v)) \\ \leq a_\beta q^\theta(X(\tau + \delta), X(\tau)) q^\theta(X(\tau), X(\tau - v)) \\ + a_\beta^2 q^\theta(X(\tau + \delta), X(\tau + \Delta \wedge \delta)) q^\theta(X(\tau + \Delta \wedge \delta), X(\tau)) \\ + a_\beta^2 q^\theta(X(\tau + \delta), X(\tau + \Delta \wedge \delta)) q^\theta(X(\tau + \Delta \wedge \delta), X(\tau - v))$$

for $0 \leq v \leq \delta \wedge \tau$. Since $\tau + \Delta \wedge \delta \in S(T + \delta)$, $\delta - \Delta \wedge \delta \in M_{\tau + \Delta \wedge \delta}(\delta)$, and $\Delta \wedge \delta + v \leq 2\delta$, (8.21) and (8.25) imply that

$$(8.26) \quad E \left[\sup_{0 \leq v \leq \delta \wedge \tau} q^\theta(X(\tau + \Delta \wedge \delta), X(\tau)) q^\theta(X(\tau), X(\tau - v)) \right] \\ \leq [a_\beta + 2a_\beta^2(a_\beta + 4a_\beta^2)] C(\delta).$$

But the left side of (8.17) is bounded by $(\lambda - \eta)^{-\theta} \lambda^{-\theta}$ times the left side of (8.26), so (8.17) follows by letting $\eta \rightarrow 0$.

Now define Δ as in (8.24) with $\tau = 0$. Then

$$(8.27) \quad q^{2\theta}(X(\Delta \wedge \delta), X(0)) \leq a_\beta [q^\theta(X(\delta), X(\Delta \wedge \delta)) q^\theta(X(\Delta \wedge \delta), X(0))] \\ + q^\theta(X(\delta), X(0)) q^\theta(X(\Delta \wedge \delta), X(0)) \\ \leq a_\beta q^\theta(X(\delta), X(\Delta \wedge \delta)) q^\theta(X(\Delta \wedge \delta), X(0)) + a_\beta q^\theta(X(\delta), X(0)),$$

so (8.18) follows as above. \square

8.6 Theorem Let (E, r) be complete and separable, and let $\{X_\alpha\}$ be a family of processes with sample paths in $D_E[0, \infty)$. Suppose that condition (a) of Theorem 7.2 holds. Then the following are equivalent:

- (a) $\{X_\alpha\}$ is relatively compact.

(b) For each $T > 0$, there exist $\beta > 0$ and a family $\{\gamma_\alpha(\delta) : 0 < \delta < 1, \text{ all } \alpha\}$ of nonnegative random variables satisfying

$$(8.28) \quad E[q^\theta(X_\alpha(t+u), X_\alpha(t)) | \mathcal{F}_t^x] q^\theta(X_\alpha(t), X_\alpha(t-v)) \leq E[\gamma_\alpha(\delta) | \mathcal{F}_t^x]$$

for $0 \leq t \leq T$, $0 \leq u \leq \delta$, and $0 \leq v \leq \delta \wedge t$, where $\mathcal{F}_t^x \equiv \mathcal{F}_t^{X_\alpha}$; in addition,

$$(8.29) \quad \lim_{\delta \rightarrow 0} \sup_{\alpha} E[\gamma_\alpha(\delta)] = 0$$

and

$$(8.30) \quad \lim_{\delta \rightarrow 0} \sup_{\alpha} E[q^\theta(X_\alpha(\delta), X_\alpha(0))] = 0.$$

(c) For each $T > 0$, there exists $\beta > 0$ such that the quantities

$$(8.31) \quad C_\alpha(\delta) =$$

$$\sup_{\tau \in S_0^x(T)} \sup_{0 \leq u \leq \delta} E \left[\sup_{0 \leq v \leq \delta \wedge \tau} q^\theta(X_\alpha(\tau+u), X_\alpha(\tau)) q^\theta(X_\alpha(\tau), X_\alpha(\tau-v)) \right],$$

defined for $0 < \delta < 1$ and all α , satisfy

$$(8.32) \quad \lim_{\delta \rightarrow 0} \sup_{\alpha} C_\alpha(\delta) = 0;$$

in addition (8.30) holds. (Here $S_0^x(T)$ is the collection of all discrete $\{\mathcal{F}_t^x\}$ -stopping times bounded by T .)

8.7 Remark (a) If, as will typically be the case,

$$(8.33) \quad E[q^\theta(X_\alpha(t+u), X_\alpha(t)) | \mathcal{F}_t^x] \leq E[\gamma_\alpha(\delta) | \mathcal{F}_t^x]$$

in place of (8.28), then $E[q^\theta(X_\alpha(\delta), X_\alpha(0))] \leq E[\gamma_\alpha(\delta)]$ and we need only verify (8.29) in condition (b).

(b) For sequences $\{X_n\}$, one can replace \sup_{α} in (8.29), (8.30), and (8.32) by $\lim_{n \rightarrow \infty}$ as was done in Corollary 7.4. \square

Proof. (a \Rightarrow b) In view of Theorem 7.2, this follows from the facts that

$$\begin{aligned} (8.34) \quad & q(X_\alpha(t+u), X_\alpha(t)) q(X_\alpha(t), X_\alpha(t-v)) \\ & \leq q(X_\alpha(t+u), X_\alpha(t)) \wedge q(X_\alpha(t), X_\alpha(t-v)) \\ & \leq w'(X_\alpha, 2\delta, T+\delta) \wedge 1 \end{aligned}$$

for $0 \leq t \leq T$, $0 \leq u \leq \delta$, and $0 \leq v \leq \delta \wedge t$, and

$$(8.35) \quad q(X_\alpha(\delta), X_\alpha(0)) \leq w'(X_\alpha, \delta, T) \wedge 1.$$

(b \Rightarrow c) Observe that t in (8.28) may be replaced by $\tau \in S_0^x(T)$ (Problem 25 of Chapter 2), and that we may replace the right side of (8.28) by its

supremum over $v \in [0, \delta \wedge \tau] \cap \mathbb{Q}$ and hence over $v \in [0, \delta \wedge \tau]$. Consequently, (8.29) implies (8.32).

(c \Rightarrow a) This follows from Lemma 8.4, Remark 8.5, Proposition 8.3, and Theorem 7.2. \square

The following result gives sufficient conditions for the existence of $\{\gamma_\alpha(\delta): 0 < \delta < 1, \text{ all } \alpha\}$ with the properties required by condition (b) of Theorem 8.6.

8.8 Theorem Let (E, r) be separable, and let $\{X_\alpha\}$ be a family of processes with sample paths in $D_E[0, \infty)$. Fix $T > 0$ and suppose there exist $\beta > 0$, $C > 0$, and $\theta > 1$ such that for all α

$$(8.36) \quad E[q^\theta(X_\alpha(t+h), X_\alpha(t)) \wedge q^\theta(X_\alpha(t), X_\alpha(t-h))] \leq Ch^\theta,$$

$$0 \leq t \leq T+1, 0 \leq h \leq t,$$

which is implied by

$$(8.37) \quad E[q^{\theta/2}(X_\alpha(t+h), X_\alpha(t))q^{\theta/2}(X_\alpha(t), X_\alpha(t-h))] \leq Ch^\theta,$$

$$0 \leq t \leq T+1, 0 \leq h \leq t.$$

Then there exists a family $\{\gamma_\alpha(\delta): 0 < \delta < 1, \text{ all } \alpha\}$ of nonnegative random variables for which (8.29) holds, and

$$(8.38) \quad q^\theta(X_\alpha(t+u), X_\alpha(t))q^\theta(X_\alpha(t), X_\alpha(t-v)) \leq \gamma_\alpha(\delta)$$

for $0 \leq t \leq T$, $0 \leq u \leq \delta$, and $0 \leq v \leq \delta \wedge t$.

8.9 Remark (a) The inequality (8.28) follows by taking conditional expectations on both sides of (8.38).

(b) Let $\varepsilon > 0$, $C > 0$, $\theta > 1$, and $0 < h \leq t$, and suppose that

$$(8.39) \quad P\{r(X_\alpha(t+h), X_\alpha(t)) \geq \lambda, r(X_\alpha(t), X_\alpha(t-h)) \geq \lambda\} \leq \lambda^{-\varepsilon} Ch^\theta$$

for all $\lambda > 0$. Then, letting $\beta = 1 + \varepsilon$,

$$(8.40) \quad E[q^\theta(X_\alpha(t+h), X_\alpha(t)) \wedge q^\theta(X_\alpha(t), X_\alpha(t-h))]$$

$$\begin{aligned} &= \int_0^1 P\{q^\theta(X_\alpha(t+h), X_\alpha(t)) \geq x, q^\theta(X_\alpha(t), X_\alpha(t-h)) \geq x\} dx \\ &\leq \int_0^1 x^{-\varepsilon/\theta} Ch^\theta dx = \beta Ch^\theta. \end{aligned} \quad \square$$

Proof. We prove the theorem in the case $\beta > 1$; the proof for $0 < \beta \leq 1$ is similar and in fact simpler since q^θ is a metric (satisfies the triangle inequality) in this case. In the calculations that follow we drop the subscript α . Define

(8.41)

$$\eta_m = \sum_{1 \leq k \leq 2^m(T+1)-1} q^\beta(X((k+1)2^{-m}), X(k2^{-m})) \wedge q^\beta(X(k2^{-m}), X((k-1)2^{-m}))$$

for $m = 0, 1, \dots$, and fix a nonnegative integer n . We claim that for integers $m \geq n$ and j, k_1, k_2 , and k_3 satisfying

$$(8.42) \quad 0 \leq j2^{-n} \leq k_12^{-m} < k_22^{-m} < k_32^{-m} \leq (j+2)2^{-n} \leq T+1,$$

we have

$$(8.43) \quad q(X(k_32^{-m}), X(k_22^{-m})) \wedge q(X(k_22^{-m}), X(k_12^{-m})) \leq 2 \sum_{i=n}^m \eta_i^{1/\beta}.$$

(If $0 < \beta \leq 1$, replace q by q^β and $\eta_i^{1/\beta}$ by η_i in (8.43).)

We prove the claim by induction. For $m = n$, (8.43) is immediate since (8.42) implies that $k_1 = j$, $k_2 = j+1$, and $k_3 = j+2$, and

$$(8.44) \quad q(X((j+2)2^{-n}), X((j+1)2^{-n})) \wedge q(X((j+1)2^{-n}), X(j2^{-n})) \leq \eta_n^{1/\beta}.$$

Suppose (8.43) holds for some $m \geq n$ and $0 \leq j2^{-n} \leq k_12^{-m-1} < k_22^{-m-1} < k_32^{-m-1} \leq (j+2)2^{-n} \leq T+1$. For $i = 1, 2, 3$, let $\varepsilon_i = q(X(k'_i2^{-m}), X(k_i2^{-m-1}))$, where if k_i is even, $k'_i = k_i/2$, and if k_i is odd, $k'_i = (k_i \pm 1)/2$ as determined by

(8.45)

$$\varepsilon_i = q(X((k_i+1)2^{-m-1}), X(k_i2^{-m-1})) \wedge q(X(k_i2^{-m-1}), X((k_i-1)2^{-m-1})).$$

Note that $\varepsilon_i = 0$ if k_i is even and $\varepsilon_i \leq \eta_{m+1}^{1/\beta}$ otherwise, so the triangle inequality implies that

$$\begin{aligned} (8.46) \quad & q(X(k_32^{-m-1}), X(k_22^{-m-1})) \wedge q(X(k_22^{-m-1}), X(k_12^{-m-1})) \\ & \leq [\varepsilon_3 + q(X(k'_32^{-m}), X(k'_22^{-m})) + \varepsilon_2] \\ & \quad \wedge [\varepsilon_2 + q(X(k'_22^{-m}), X(k'_12^{-m})) + \varepsilon_1] \\ & \leq 2\eta_{m+1}^{1/\beta} + q(X(k'_32^{-m}), X(k'_22^{-m})) \wedge q(X(k'_22^{-m}), X(k'_12^{-m})). \end{aligned}$$

By the definition of k'_i , we still have $0 \leq j2^{-n} \leq k'_12^{-m} \leq k'_22^{-m} \leq k'_32^{-m} \leq (j+2)2^{-n} \leq T+1$, and hence the induction step is verified.

If $0 \leq t_1 < t_2 < t_3 \leq T+\frac{1}{2}$ and t_1, t_2 , and t_3 are dyadic rational with $t_3 - t_1 \leq 2^{-n}$ for some $n \geq 1$, then there exist j, m, k_1, k_2 , and k_3 satisfying (8.42) and $t_i = k_i2^{-m}$. Consequently,

$$(8.47) \quad q(X(t_3), X(t_2)) \wedge q(X(t_2), X(t_1)) \leq 2 \sum_{i=n}^{\infty} \eta_i^{1/\beta} \equiv \varphi_n.$$

By right continuity, (8.47) holds for all $0 \leq t_1 < t_2 < t_3 < T+\frac{1}{2}$ with $t_3 - t_1 \leq 2^{-n}$. If $\delta \geq \frac{1}{4}$, let $\gamma(\delta) = 1$; if $0 < \delta < \frac{1}{4}$, let n_δ be the largest integer n

satisfying $2\delta < 2^{-n}$, and define $\gamma(\delta) = \varphi_{n_\delta}$. Since $ab \leq a \wedge b$ for all $a, b \in [0, 1]$, we conclude that (8.38) holds. Also,

$$(8.48) \quad \begin{aligned} E[\gamma(\delta)] &= 2 \sum_{i=n_\delta}^{\infty} E[\eta_i^{1/\theta}] \leq 2 \sum_{i=n_\delta}^{\infty} E[\eta_i]^{1/\theta} \\ &\leq 2 \sum_{i=n_\delta}^{\infty} [2^i(T+1)C2^{-i\theta}]^{1/\theta}, \end{aligned}$$

so $\lim_{\delta \rightarrow 0} E[\gamma(\delta)] = 0$ (and the limit is uniform in α). \square

8.10 Corollary Let (E, r) be complete and separable, and let X be a process with values in E that is right continuous in probability. Suppose that for each $T > 0$, there exist $\beta > 0$, $C > 0$, and $\theta > 1$ such that

$$(8.49) \quad E[q^\theta(X(t+h_2), X(t)) \wedge q^\theta(X(t), X(t-h_1))] \leq C(h_1 \vee h_2)^\theta$$

whenever $0 \leq t - h_1 \leq t \leq t + h_2 \leq T$. Then X has a version with sample paths in $D_E[0, \infty)$.

Proof. Define the sequence of processes $\{X_n\}$ with sample paths in $D_E[0, \infty)$ by $X_n(t) = X([nt] + 1/n)$. It suffices to show that $\{X_n\}$ is relatively compact, for by Theorem 7.8 and the assumed right continuity in probability of X , the limit in distribution of any convergent subsequence of $\{X_n\}$ has the same finite-dimensional distributions as X .

Given $\eta > 0$ and $t \geq 0$, choose by Lemma 2.1 a compact set $\Gamma_{n,1} \subset E$ such that $P\{X(t) \in \Gamma_{n,1}\} \geq 1 - \eta$. Then (7.13) holds by Theorem 3.1 and the fact that $X_n(t) \Rightarrow X(t)$ in E . Consequently, it suffices to verify condition (b) of Theorem 8.6, and for this we apply Theorem 8.8. By (8.49) with T replaced by $T+2$, there exist $\beta > 0$, $C > 0$, and $\theta > 1$ such that for each n

$$(8.50) \quad \begin{aligned} E[q^\theta(X_n(t+h), X_n(t)) \wedge q^\theta(X_n(t), X_n(t-h))] \\ \leq C \left(\frac{[nt] - [n(t-h)]}{n} \vee \frac{[n(t+h)] - [nt]}{n} \right)^\theta, \\ 0 \leq t \leq T+1, \quad 0 \leq h \leq t. \end{aligned}$$

But the left side of (8.50) is zero if $2h \leq 1/n$ and is bounded by $C(h + n^{-1})^\theta \leq 3^\theta Ch^\theta$ if $2h > 1/n$. Thus, Theorem 8.8 implies that (8.29) holds, and the verification of (8.30) is immediate. \square

9. FURTHER CRITERIA FOR RELATIVE COMPACTNESS IN $D_E[0, \infty)$

We now consider criteria for relative compactness that are particularly useful in approximating by Markov processes. These criteria are based on the following simple result. As usual, (E, r) denotes a metric space.

9.1 Theorem Let (E, r) be complete and separable, and let $\{X_a\}$ be a family of processes with sample paths in $D_E[0, \infty)$. Suppose that the compact containment condition holds. That is, for every $\eta > 0$ and $T > 0$ there exists a compact set $\Gamma_{\eta, T} \subset E$ for which

$$(9.1) \quad \inf_a P\{X_a(t) \in \Gamma_{\eta, T} \text{ for } 0 \leq t \leq T\} \geq 1 - \eta.$$

Let H be a dense subset of $\bar{C}(E)$ in the topology of uniform convergence on compact sets. Then $\{X_a\}$ is relatively compact if and only if $\{f \circ X_a\}$ is relatively compact (as a family of processes with sample paths in $D_R[0, \infty)$) for each $f \in H$.

Proof. Given $f \in \bar{C}(E)$, the mapping $x \rightarrow f \circ x$ from $D_E[0, \infty)$ into $D_R[0, \infty)$ is continuous (Problem 13). Consequently, convergence in distribution of a sequence of processes $\{X_n\}$ with sample paths in $D_E[0, \infty)$ implies convergence in distribution of $\{f \circ X_n\}$, and hence relative compactness of $\{X_n\}$ implies relative compactness of $\{f \circ X_n\}$.

Conversely, suppose that $\{f \circ X_a\}$ is relatively compact for every $f \in H$. It then follows from (9.1), (6.3), and Theorem 7.2 that $\{f \circ X_a\}$ is relatively compact for every $f \in \bar{C}(E)$ and in particular that $\{q(\cdot, z) \circ X_a\}$ is relatively compact for each $z \in E$, where $q = r \wedge 1$. Let $\eta > 0$ and $T > 0$. By the compactness of $\Gamma_{\eta, T}$, there exists for each $\varepsilon > 0$ a finite set $\{z_1, \dots, z_N\} \subset \Gamma_{\eta, T}$ such that $\min_{1 \leq i \leq N} q(x, z_i) < \varepsilon$ for all $x \in \Gamma_{\eta, T}$. If $y \in \Gamma_{\eta, T}$, then, for some $i \in \{1, \dots, N\}$, $q(y, z_i) < \varepsilon$ and hence

$$(9.2) \quad q(x, y) < |q(x, z_i) - q(y, z_i)| + 2\varepsilon$$

for all $x \in E$. Consequently, for $0 \leq t \leq T$, $0 \leq u \leq \delta$, and $0 \leq v \leq \delta \wedge t$,

$$\begin{aligned} (9.3) \quad & q(X_a(t+u), X_a(t))q(X_a(t), X_a(t-v)) \\ & \leq \bigvee_{i=1}^N |q(X_a(t+u), z_i) - q(X_a(t), z_i)| |q(X_a(t), z_i) - q(X_a(t-v), z_i)| \\ & \quad + 4(\varepsilon + \varepsilon^2) + \chi_{\{X_a(s) \notin \Gamma_{\eta, T} \text{ for some } s \in [0, T]\}} \\ & \leq \bigvee_{i=1}^N w'(q(\cdot, z_i) \circ X_a, 2\delta, T + \delta) \wedge 1 \\ & \quad + 4(\varepsilon + \varepsilon^2) + \chi_{\{X_a(s) \notin \Gamma_{\eta, T} \text{ for some } s \in [0, T]\}} \\ & \equiv \gamma_a(\delta), \end{aligned}$$

where $0 < \delta < 1$. Note that N depends on η , T , and ε .

Since $\lim_{\delta \rightarrow 0} \sup_z E[w'(q(\cdot, z) \circ X_n, 2\delta, T + \delta) \wedge 1] = 0$ for each $z \in E$ by Theorem 7.2, we may select η and ε depending on δ in such a way that $\lim_{\delta \rightarrow 0} \sup_z E[\gamma_\varepsilon(\delta)] = 0$. Finally, (9.2) implies that

$$(9.4) \quad q(X_n(\delta), X_n(0)) \leq \bigvee_{i=1}^N |q(X_n(\delta), z_i) - q(X_n(0), z_i)| + 2\varepsilon + \chi_{\{X_n(0) \notin r_{\eta, T}\}}$$

for all $\delta > 0$, so $\lim_{\delta \rightarrow 0} \sup_z E[q(X_n(\delta), X_n(0))] = 0$ by Theorem 8.6. Thus, the relative compactness of $\{X_n\}$ follows from Theorem 8.6. \square

9.2 Corollary Let (E, r) be complete and separable, and let $\{X_n\}$ be a sequence of processes with sample paths in $D_E[0, \infty)$. Let $M \subset \bar{C}(E)$ strongly separate points. Suppose there exists a process X with sample paths in $D_E[0, \infty)$ such that for each finite set $\{g_1, \dots, g_k\} \subset M$,

$$(9.5) \quad (g_1, \dots, g_k) \circ X_n \Rightarrow (g_1, \dots, g_k) \circ X \text{ in } D_{\mathbb{R}^k}[0, \infty).$$

Then $X_n \Rightarrow X$.

Proof. Let H be the smallest algebra containing $M \cup \{1\}$, that is, the algebra of functions of the form $\sum_{i=1}^l a_i f_{i1} f_{i2} \dots f_{im}$, where $l \geq 1$, $m \geq 1$, and $a_i \in \mathbb{R}$ and $f_{ij} \in M \cup \{1\}$ for $i = 1, \dots, l$ and $j = 1, \dots, m$. By the Stone–Weierstrass theorem, H is dense in $\bar{C}(E)$ in the topology of uniform convergence on compact sets. Note that (9.5) holds for every finite set $\{g_1, \dots, g_k\} \subset H$. Thus, by Theorem 9.1, to prove the relative compactness of $\{X_n\}$ we need only verify (9.1).

Let $\Gamma \subset E$ be compact, and let $\delta > 0$. For each $x \in E$, choose $\{h_1^x, \dots, h_{k(x)}^x\} \subset H$ satisfying

$$(9.6) \quad \varepsilon(x) \equiv \inf_{y: r(x, y) \geq \delta} \max_{1 \leq i \leq k(x)} |h_i^x(y) - h_i^x(x)| > 0,$$

and let $U_x = \{y \in E : \max_{1 \leq i \leq k(x)} |h_i^x(y) - h_i^x(x)| < \varepsilon(x)\}$. Then $\Gamma \subset \bigcup_{x \in \Gamma} U_x \subset \Gamma^\delta$, so, since Γ is compact, there exist $x_1, \dots, x_n \in \Gamma$ such that $\Gamma \subset \bigcup_{i=1}^n U_{x_i} \subset \Gamma^\delta$. Define $\sigma: D_{\mathbb{R}}[0, \infty) \rightarrow D_{\mathbb{R}}[0, \infty)$ by $\sigma(x)(t) = \sup_{0 \leq s \leq t} x(s)$ and observe that σ is continuous (by Proposition 5.3). For each n , let

$$(9.7) \quad Y_n(t) = \min_{1 \leq i \leq n} \{g_i(X_n(s)) - \varepsilon(x_i)\}, \quad t \geq 0,$$

where $g_i(x) \equiv \max_{1 \leq s \leq k(x)} |h_i^x(s) - h_i^x(x_i)|$, and put $Z_n = \sigma(Y_n)$. It follows from (9.5) and the continuity of σ that $Z_n \Rightarrow Z$, where Z is defined in terms of X as Z_n is in terms of X_n . Therefore $Z_n(T) \Rightarrow Z(T)$ for all $T \in D(Z)$, and for such T

$$\begin{aligned} (9.8) \quad & \lim_{n \rightarrow \infty} P\{X_n(t) \in \Gamma^\delta \text{ for } 0 \leq t \leq T\} \\ & \geq \lim_{n \rightarrow \infty} P\{Z_n(T) < 0\} \\ & \geq P\{Z(T) < 0\} \\ & \geq P\{X(t) \in \Gamma \text{ for } 0 \leq t \leq T\} \end{aligned}$$

by Theorem 3.1, where the last inequality uses the fact that

$$(9.9) \quad \sup_{x \in \Gamma} \min_{1 \leq i \leq N} \{g_i(x) - a(x_i)\} < 0.$$

Let $\eta > 0$, let $T > 0$ be as above, let $m \geq 1$, and choose a compact set $\Gamma_{0,m} \subset E$ such that

$$(9.10) \quad P\{X(t) \in \Gamma_{0,m} \text{ for } 0 \leq t \leq T\} \geq 1 - \eta 2^{-m-1};$$

this is possible by Theorem 5.6, Lemma 2.1, and Remark 6.4. By (9.8), there exists $n_m \geq 1$ such that

$$(9.11) \quad \inf_{n \geq n_m} P\{X_n(t) \in \Gamma_{0,m}^{1/m} \text{ for } 0 \leq t \leq T\} \geq 1 - \eta 2^{-m}.$$

Finally, for $n = 1, \dots, n_m - 1$, choose a compact set $\Gamma_{n,m} \subset E$ such that

$$(9.12) \quad P\{X_n(t) \in \Gamma_{n,m}^{1/m} \text{ for } 0 \leq t \leq T\} \geq 1 - \eta 2^{-m}.$$

Letting $\Gamma_m = \bigcup_{n=0}^{n_m-1} \Gamma_{n,m}$, we have

$$(9.13) \quad \inf_{n \geq 1} P\{X_n(t) \in \Gamma_m^{1/m} \text{ for } 0 \leq t \leq T\} \geq 1 - \eta 2^{-m},$$

so if we define $\Gamma_{\eta,T}$ to be the closure of $\bigcap_{m \geq 1} \Gamma_m^{1/m}$, then $\Gamma_{\eta,T}$ is compact (being complete and totally bounded) and

$$(9.14) \quad \inf_{n \geq 1} P\{X_n(t) \in \Gamma_{\eta,T} \text{ for } 0 \leq t \leq T\} \geq 1 - \eta.$$

Finally, we note that

$$(9.15) \quad (g_1 \wedge a_1, \dots, g_k \wedge a_k) \circ X_n \Rightarrow (g_1 \wedge a_1, \dots, g_k \wedge a_k) \circ X$$

for all $g_1, \dots, g_k \in H$ and $a_1, \dots, a_k \in \mathbb{R}$. This, together with the fact that H is dense in $\bar{C}(E)$ in the topology of uniform convergence on compact sets, allows one to conclude that the finite-dimensional distributions converge. The details are left to the reader. \square

9.3 Corollary Let E be locally compact and separable, and let E^Δ be its one-point compactification. If $\{X_n\}$ is a sequence of processes with sample paths in $D_E[0, \infty)$ and if $\{f \circ X_n\}$ is relatively compact for every $f \in \bar{C}(E)$ (the space of continuous functions on E vanishing at infinity), then $\{X_n\}$ is relatively compact considered as a sequence of processes with sample paths in $D_{E^\Delta}[0, \infty)$. If, in addition, $(X_n(t_1), \dots, X_n(t_k)) \Rightarrow (X(t_1), \dots, X(t_k))$ for all finite subsets $\{t_1, \dots, t_k\}$ of some dense set $D \subset [0, \infty)$, where X has sample paths in $D_E[0, \infty)$, then $X_n \Rightarrow X$ in $D_E[0, \infty)$.

Proof. If $f \in C(E^A)$, then $f(\cdot) - f(\Delta)$ restricted to E belongs to $\hat{C}(E)$. Consequently, $\{f \circ X_n\}$ is relatively compact for every $f \in C(E^A)$, and the relative compactness of $\{X_n\}$ in $D_{E^A}[0, \infty)$ follows from Theorem 9.1. Under the additional assumptions, $X_n \Rightarrow X$ in $D_E[0, \infty)$ by Theorem 7.7, and hence $X_n \Rightarrow X$ in $D_E[0, \infty)$ by Corollary 3.2. \square

We now consider the problem of verifying the relative compactness of $\{f \circ X_\alpha\}$, where $f \in \bar{C}(E)$ is fixed. First, however, recall the notation and terminology of Section 7 of Chapter 2. For each α , let X_α be a process with sample paths in $D_E[0, \infty)$ defined on a probability space $(\Omega_\alpha, \mathcal{F}^\alpha, P_\alpha)$ and adapted to a filtration $\{\mathcal{F}_t^\alpha\}$. Let \mathcal{L}_α be the Banach space of real-valued $\{\mathcal{F}_t^\alpha\}$ -progressive processes with norm $\|Y\| = \sup_{t \geq 0} E[|Y(t)|] < \infty$. Let

$$(9.16) \quad \hat{\mathcal{A}}_\alpha = \left\{ (Y, Z) \in \mathcal{L}_\alpha \times \mathcal{L}_\alpha : Y(t) - \int_0^t Z(s) ds \text{ is an } \{\mathcal{F}_s^\alpha\}\text{-martingale} \right\},$$

and note that completeness of $\{\mathcal{F}_t^\alpha\}$ is not needed here.

9.4 Theorem Let (E, r) be arbitrary, and let $\{X_\alpha\}$ be a family of processes as above. Let C_α be a subalgebra of $\bar{C}(E)$ (e.g., the space of bounded, uniformly continuous functions with bounded support), and let D be the collection of $f \in \bar{C}(E)$ such that for every $\varepsilon > 0$ and $T > 0$ there exist $(Y_\alpha, Z_\alpha) \in \hat{\mathcal{A}}_\alpha$ with

$$(9.17) \quad \sup_\alpha E \left[\sup_{t \in [0, T] \cap \mathbb{Q}} |Y_\alpha(t) - f(X_\alpha(t))| \right] < \varepsilon$$

and

$$(9.18) \quad \sup_\alpha E[\|Z_\alpha\|_{p, T}] < \infty \quad \text{for some } p \in (1, \infty].$$

($\|h\|_{p, T} = [\int_0^T |h(t)|^p dt]^{1/p}$ if $p < \infty$; $\|h\|_{\infty, T} = \text{ess sup}_{0 \leq t \leq T} |h(t)|$.) If C_α is contained in the closure of D (in the sup norm), then $\{f \circ X_\alpha\}$ is relatively compact for each $f \in C_\alpha$; more generally, $\{(f_1, \dots, f_k) \circ X_\alpha\}$ is relatively compact in $D_{R^k}[0, \infty)$ for all $f_1, f_2, \dots, f_k \in C_\alpha$, $1 \leq k \leq \infty$.

9.5 Remark (a) Taking $p = 1$ in condition (9.18) is not sufficient. For example, let $n \geq 1$ and consider the two-state ($E = \{0, 1\}$) Markov process X_n with infinitesimal matrix

$$(9.19) \quad \begin{pmatrix} -1 & 1 \\ n & -n \end{pmatrix}$$

and $P\{X_n(0) = 0\} = 1$. Given a function f on $\{0, 1\}$, put $Y_n = f \circ X_n$ and $Z_n = (A_n f) \circ X_n$, so that $(Y_n, Z_n) \in \mathcal{A}_n$ and

$$\begin{aligned}(9.20) \quad & E[|A_n f(X_n(t))|] \\&= n|f(0) - f(1)|P\{X_n(t) = 1\} + |f(1) - f(0)|P\{X_n(t) = 0\} \\&= |f(1) - f(0)|(1 + (n-1)(n+1)^{-1}(1 - e^{-(n+1)t})) \\&\leq 2|f(1) - f(0)|,\end{aligned}$$

for all $t \geq 0$. However, observe that the finite-dimensional distributions of X_n converge to those of a process that is identically zero, but $\{X_n\}$ does not converge in distribution.

(b) For sequences $\{X_n\}$, one can replace \sup_a in (9.17) and (9.18) by $\lim_{n \rightarrow \infty}$. \square

Proof. Since D is actually closed, we may assume that $C_\alpha \subset D$. Let $f \in C_\alpha$, $\varepsilon > 0$, and $T > 0$. Then $f^2 \in C_\alpha$ and there exist $(Y_\alpha, Z_\alpha), (Y'_\alpha, Z'_\alpha) \in \mathcal{A}_\alpha$ such that

$$(9.21) \quad \sup_\alpha E \left[\sup_{t \in [0, T+1] \cap \mathbb{Q}} |f(X_\alpha(t)) - Y_\alpha(t)| \right] < \varepsilon,$$

$$(9.22) \quad \sup_\alpha E \left[\sup_{t \in [0, T+1] \cap \mathbb{Q}} |f^2(X_\alpha(t)) - Y'_\alpha(t)| \right] < \varepsilon,$$

$$(9.23) \quad \sup_\alpha E[\|Z_\alpha\|_{p, T+1}] < \infty \quad \text{for some } p \in (1, \infty],$$

$$(9.24) \quad \sup_\alpha E[\|Z'_\alpha\|_{p', T+1}] < \infty \quad \text{for some } p' \in (1, \infty].$$

Let $0 < \delta < 1$. For each $t \in [0, T] \cap \mathbb{Q}$ and $u \in [0, \delta] \cap \mathbb{Q}$,

$$\begin{aligned}(9.25) \quad & E[(f(X_\alpha(t+u)) - f(X_\alpha(t)))^2 | \mathcal{F}_t^\alpha] \\&= E[f^2(X_\alpha(t+u)) - f^2(X_\alpha(t)) | \mathcal{F}_t^\alpha] \\&\quad - 2f(X_\alpha(t))E[f(X_\alpha(t+u)) - f(X_\alpha(t)) | \mathcal{F}_t^\alpha] \\&\leq 2E \left[\sup_{s \in [0, T+1] \cap \mathbb{Q}} |f^2(X_\alpha(s)) - Y'_\alpha(s)| \middle| \mathcal{F}_t^\alpha \right] \\&\quad + 4\|f\|E \left[\sup_{s \in [0, T+1] \cap \mathbb{Q}} |f(X_\alpha(s)) - Y_\alpha(s)| \middle| \mathcal{F}_t^\alpha \right] \\&\quad + E \left[\sup_{0 \leq r \leq T} \int_r^{r+\delta} |Z'_\alpha(s)| ds \middle| \mathcal{F}_t^\alpha \right] \\&\quad + 2\|f\|E \left[\sup_{0 \leq r \leq T} \int_r^{r+\delta} |Z_\alpha(s)| ds \middle| \mathcal{F}_t^\alpha \right].\end{aligned}$$

Let $1/p + 1/q = 1$ and $1/p' + 1/q' = 1$, and note that $\int_r^{r+\delta} |h(s)| ds \leq \delta^{1/q} \|h\|_{p', T+1}$ for $0 \leq r \leq T$. Therefore, if we define

$$(9.26) \quad \begin{aligned} \gamma_a(\delta) &= 2 \sup_{s \in [0, T+1] \cap \mathbb{Q}} |f^2(X_a(s)) - Y'_a(s)| \\ &\quad + 4 \|f\| \sup_{s \in [0, T+1] \cap \mathbb{Q}} |f(X_a(s)) - Y_a(s)| \\ &\quad + \delta^{1/q'} \|Z'_a\|_{p', T+1} + 2 \|f\| \delta^{1/q} \|Z_a\|_{p', T+1}, \end{aligned}$$

then

$$(9.27) \quad E[(f(X_a(t+u)) - f(X_a(t)))^2 | \mathcal{F}_t^\alpha] \leq E[\gamma_a(\delta) | \mathcal{F}_t^\alpha].$$

Note that this holds for all $0 \leq t \leq T$ and $0 \leq u \leq \delta$ (not just for rational t and u) by the right continuity of X_a . Since

$$(9.28) \quad \sup_a E[\gamma_a(\delta)] \leq (2 + 4 \|f\|) \varepsilon + \delta^{1/q'} \sup_a E[\|Z'_a\|_{p', T+1}] + 2 \|f\| \delta^{1/q} \sup_a E[\|Z_a\|_{p', T+1}],$$

we may select ε depending on δ in such a way that (8.29) holds. Therefore, $\{f \circ X_a\}$ is relatively compact by Theorem 8.6 (see Remark 8.7(a)).

Let $1 \leq k < \infty$. Given $f_1, \dots, f_k \in C_a$, define $\gamma_a^J(\delta)$ as in (9.26) and set $\gamma_a(\delta) = \sum_{j=1}^k \gamma_a^j(\delta)$. Then

$$(9.29) \quad E \left[\sum_{j=1}^k (f_j(X_a(t+u)) - f_j(X_a(t)))^2 \middle| \mathcal{F}_t^\alpha \right] \leq E[\gamma_a(\delta) | \mathcal{F}_t^\alpha]$$

for $0 \leq t \leq T$ and $0 \leq u \leq \delta$, and the $\gamma_a^J(\delta)$ can be selected so that (8.29) holds. Finally, relative compactness for $k = \infty$ follows from relative compactness for all $k < \infty$. (See Problem 23.) \square

10. CONVERGENCE TO A PROCESS IN $C_{\mathbb{B}}[0, \infty)$

Let (E, r) be a metric space, and let $C_E[0, \infty)$ be the space of continuous functions $x: [0, \infty) \rightarrow E$. For $x \in D_E[0, \infty)$, define

$$(10.1) \quad J(x) = \int_0^\infty e^{-u} [J(x, u) \wedge 1] du,$$

where

$$(10.2) \quad J(x, u) = \sup_{0 \leq t \leq u} r(x(t), x(t-)).$$

Since the mapping $x \rightarrow J(x, \cdot)$ from $D_E[0, \infty)$ into $D_{(0, \infty)}[0, \infty)$ is continuous (by Proposition 5.3), it follows that J is continuous on $D_E[0, \infty)$. For each $x \in D_E[0, \infty)$, $J(x, \cdot)$ is nondecreasing, so $J(x) = 0$ if and only if $x \in C_E[0, \infty)$.

10.1 Lemma Suppose $\{x_n\} \subset D_E[0, \infty)$, $x \in D_E[0, \infty)$, and $\lim_{n \rightarrow \infty} d(x_n, x) = 0$. Then

$$(10.3) \quad \overline{\lim}_{n \rightarrow \infty} \sup_{0 \leq t \leq u} r(x_n(t), x(t)) \leq J(x, u)$$

for all $u \geq 0$.

Proof. By Proposition 5.3, there exists $\{\lambda_n\} \subset \Lambda$ such that $\lim_{n \rightarrow \infty} \sup_{0 \leq t \leq u} |\lambda_n(t) - t| = 0$ and $\lim_{n \rightarrow \infty} \sup_{0 \leq t \leq u} r(x_n(t), x(\lambda_n(t))) = 0$ for all $u \geq 0$. Therefore

$$\begin{aligned} (10.4) \quad & \overline{\lim}_{n \rightarrow \infty} \sup_{0 \leq t \leq u} r(x_n(t), x(t)) \\ & \leq \overline{\lim}_{n \rightarrow \infty} \sup_{0 \leq t \leq u} r(x_n(t), x(\lambda_n(t))) \\ & \quad + \overline{\lim}_{n \rightarrow \infty} \sup_{0 \leq t \leq u} r(x(\lambda_n(t)), x(t)) \\ & \leq J(x, u) \end{aligned}$$

for all $u \geq 0$. \square

Let $\bar{C}(D_E[0, \infty), d_U)$ be the space of bounded, real-valued functions on $D_E[0, \infty)$ that are continuous with respect to the metric

$$(10.5) \quad d_U(x, y) = \int_0^\infty e^{-u} \sup_{0 \leq t \leq u} [r(x(t), y(t)) \wedge 1] du,$$

that is, continuous in the topology of uniform convergence on compact subsets of $[0, \infty)$. Since $d \leq d_U$, we have $\mathcal{S}_E \equiv \mathcal{B}(D_E[0, \infty), d) \subset \mathcal{B}(D_E[0, \infty), d_U)$.

10.2 Theorem Let X_n , $n = 1, 2, \dots$, and X be processes with sample paths in $D_E[0, \infty)$, and suppose that $X_n \Rightarrow X$. Then (a) X is a.s. continuous if and only if $J(X_n) \Rightarrow 0$, and (b) if X is a.s. continuous, then $f(X_n) \Rightarrow f(X)$ for every \mathcal{S}_E -measurable $f \in \bar{C}(D_E[0, \infty), d_U)$.

Proof. Part (a) follows from the continuity of J on $D_E[0, \infty)$.

By Lemma 10.1, if $\{x_n\} \subset D_E[0, \infty)$, $x \in C_E[0, \infty)$, and $\lim_{n \rightarrow \infty} d(x_n, x) = 0$, then $\lim_{n \rightarrow \infty} d_U(x_n, x) = 0$. Letting $F \subset \mathbb{R}$ be closed and f be as in the statement of the theorem, $f^{-1}(F)$ is d_U -closed. Denoting its d -closure by $\overline{f^{-1}(F)}$, it follows that $\overline{f^{-1}(F)} \cap C_E[0, \infty) = f^{-1}(F) \cap C_E[0, \infty)$. Therefore, if $P_n \Rightarrow P$ on $(D_E[0, \infty), d)$ and $P(C_E[0, \infty)) = 1$,

$$\begin{aligned} (10.6) \quad & \overline{\lim}_{n \rightarrow \infty} P_n f^{-1}(F) \leq \overline{\lim}_{n \rightarrow \infty} P_n (\overline{f^{-1}(F)}) \leq P(\overline{f^{-1}(F)}) \\ & = P(\overline{f^{-1}(F)} \cap C_E[0, \infty)) = P(f^{-1}(F) \cap C_E[0, \infty)) = Pf^{-1}(F) \end{aligned}$$

by Theorem 3.1, so we conclude that $P_n f^{-1} \Rightarrow Pf^{-1}$. This implies (b). \square

The next result provides a useful criterion for a process with sample paths in $D_E[0, \infty)$ to have sample paths in $C_E[0, \infty)$. It can also be used in conjunction with Corollary 8.10.

10.3 Proposition Let (E, r) be separable, and let X be a process with sample paths in $D_E[0, \infty)$. Suppose that for each $T > 0$, there exist $\beta > 0$, $C > 0$, and $\theta > 1$ such that

$$(10.7) \quad E[q^\theta(X(t), X(s))] \leq C(t - s)^\theta$$

whenever $0 \leq s \leq t \leq T$, where $q = r \wedge 1$. Then almost all sample paths of X belong to $C_E[0, \infty)$.

Proof. Let T be a positive integer and observe that

$$(10.8) \quad \sum_{0 < i \leq T} q^\theta(X(i), X(i-)) \leq \lim_{n \rightarrow \infty} \sum_{k=1}^{2^n T} q^\theta(X(k2^{-n}), X((k-1)2^{-n})).$$

By Fatou's lemma and (10.7), the right side of (10.8) has zero expectation, and hence the left side is zero a.s. \square

10.4 Proposition For $n = 1, 2, \dots$, let $\{Y_n(k), k = 0, 1, \dots\}$ be a discrete-parameter \mathbb{R}^d -valued process, let $\alpha_n > 0$, and define

$$(10.9) \quad X_n(t) = Y_n(\lceil \alpha_n t \rceil)$$

and

$$(10.10) \quad Z_n(t) = Y_n(\lceil \alpha_n t \rceil) + (\alpha_n t - \lceil \alpha_n t \rceil)(Y_n(\lceil \alpha_n t \rceil + 1) - Y_n(\lceil \alpha_n t \rceil))$$

for all $t \geq 0$. Note that X_n has sample paths in $D_{\mathbb{R}^d}[0, \infty)$ and Z_n has sample paths in $C_{\mathbb{R}^d}[0, \infty)$. If $\lim_{n \rightarrow \infty} \alpha_n = \infty$ and X is a process with sample paths in $C_{\mathbb{R}^d}[0, \infty)$, then $X_n \Rightarrow X$ if and only if $Z_n \Rightarrow X$.

Proof. We apply Corollary 3.3. It suffices to show that, if either $X_n \Rightarrow X$ or $Z_n \Rightarrow X$, then $d(X_n, Z_n) \rightarrow 0$ in probability. (The two uses of the letter d here should cause no confusion.) For $n = 1, 2, \dots$,

$$(10.11) \quad d(X_n, Z_n) \leq \int_0^\infty e^{-u} \sup_{0 \leq s \leq u + \alpha_n^{-1}} (|X_n(t) - X_n(t-)| \wedge 1) du \\ \leq e^{\alpha_n^{-1}} J(X_n),$$

and

$$(10.12) \quad J(X_n) \leq \int_0^\infty e^{-u} \sup_{0 \leq s \leq u} \sup_{t-\epsilon \wedge s \leq s \leq t} (|Z_n(t) - Z_n(s)| \wedge 1) du$$

provided $\alpha_n^{-1} \leq \varepsilon$. But the function $J_\varepsilon: D_{\mathbb{R}^d}[0, \infty) \rightarrow [0, 1]$, defined for each $\varepsilon > 0$ by

$$(10.13) \quad J_\varepsilon(x) = \int_0^\infty e^{-u} \sup_{0 \leq t \leq u} \sup_{t-\varepsilon \leq s \leq t} (\|x(t) - x(s)\| \wedge 1) du \quad (\geq J(x)),$$

is continuous and satisfies $\lim_{\varepsilon \rightarrow 0} J_\varepsilon(x) = J(x)$ for all $x \in D_{\mathbb{R}^d}[0, \infty)$. Consequently, if $Z_n \Rightarrow X$, then (10.12) and Theorem 3.1 imply that

$$\begin{aligned} (10.14) \quad \overline{\lim}_{n \rightarrow \infty} P\{J(X_n) \geq \delta\} &\leq \lim_{\varepsilon \rightarrow 0} \overline{\lim}_{n \rightarrow \infty} P\{J_\varepsilon(Z_n) \geq \delta\} \\ &\leq \lim_{\varepsilon \rightarrow 0} P\{J_\varepsilon(X) \geq \delta\} = 0 \end{aligned}$$

for all $\delta > 0$, so we conclude that $J(X_n) \rightarrow 0$ in probability. The same conclusion follows from Theorem 10.2 if $X_n \Rightarrow X$. In either case, (10.11) implies that $d(X_n, Z_n) \rightarrow 0$ in probability, as required for Corollary 3.3. \square

11. PROBLEMS

1. Let (S, d) be complete and separable, and let $P, Q \in \mathcal{P}(S)$. Show that there exists $\mu \in \mathcal{M}(P, Q)$ (see Theorem 1.2) such that

$$(11.1) \quad \rho(P, Q) = \inf \{ \varepsilon > 0 : \mu\{(x, y) : d(x, y) \geq \varepsilon\} \leq \varepsilon \}.$$

2. Define $\|f\|_{BL} = \sup_x |f(x)| \vee \sup_{x \neq y} |f(x) - f(y)|/d(x, y)$ for each $f \in C(S)$. Given $P, Q \in \mathcal{P}(S)$, let

$$(11.2) \quad \|P - Q\| = \sup_{\|f\|_{BL}=1} \left| \int f dP - \int f dQ \right|,$$

and show that $\rho^2(P, Q) \leq \|P - Q\| \leq 3\rho(P, Q)$.

Hint: Recall that $\int f dP = \int_0^{\|f\|} P\{f \geq t\} dt$ if $f \geq 0$, and note that $\|(\varepsilon - d(\cdot, F)) \vee 0\|_{BL} \leq 1$ for $0 < \varepsilon < 1$.

3. Show that $\mathcal{P}(S)$ is separable whenever S is.

4. Suppose $\{P_n\} \subset \mathcal{P}(\mathbb{R})$, $P \in \mathcal{P}(\mathbb{R})$, and $P_n \Rightarrow P$. Define

$$(11.3) \quad G_n(x) = \inf \{y \in \mathbb{R} : P_n((-\infty, y]) \geq x\}$$

and

$$(11.4) \quad G(x) = \inf \{y \in \mathbb{R} : P((-\infty, y]) \geq x\}$$

for $0 < x < 1$, and let ξ be uniformly distributed on $(0, 1)$. Show that $G_n(\xi)$ has distribution P_n for each n , $G(\xi)$ has distribution P , and $\lim_{n \rightarrow \infty} G_n(\xi) = G(\xi)$ a.s.

5. Let (S, d) and (S', d') be separable. Let $X_n, n = 1, 2, \dots$, and X be S -valued random variables with $X_n \Rightarrow X$. Suppose there exist Borel measurable mappings $h_k, k = 1, 2, \dots$, and h from S into S' such that:
- For $k = 1, 2, \dots$, h_k is continuous a.s. with respect to the distribution of X .
 - $h_k \rightarrow h$ as $k \rightarrow \infty$ a.s. with respect to the distribution of X .
 - $\lim_{k \rightarrow \infty} \overline{\lim}_{n \rightarrow \infty} P\{d'(h_k(X_n), h(X_n)) > \varepsilon\} = 0$ for every $\varepsilon > 0$.

Show that $h(X_n) \Rightarrow h(X)$. (Note that this generalizes Corollary 1.9.)

6. Let $X_n, n = 1, 2, \dots$, and X be real-valued random variables with finite second moments defined on a common probability space (Ω, \mathcal{F}, P) . Suppose that $\{X_n\}$ converges weakly to X in $L^2(P)$ (i.e., $\lim_{n \rightarrow \infty} E[X_n Z] = E[XZ]$, $Z \in L^2(P)$), and $\{X_n\}$ converges in distribution to some real-valued random variable Y . Give necessary and sufficient conditions for X and Y to have the same distribution.
7. Let X and Y be S -valued random variables defined on a probability space (Ω, \mathcal{F}, P) , and let \mathcal{G} be a sub- σ -algebra of \mathcal{F} . Suppose that $M \subset \bar{C}(S)$ is separating and

$$(11.5) \quad E[f(X)|\mathcal{G}] = f(Y)$$

for every $f \in M$. Show that $X = Y$ a.s.

8. Let $M = \{f \in \bar{C}(\mathbb{R}): f \text{ has period } N \text{ for some positive integer } N\}$. Show that M is separating but not convergence determining.
9. Let $M \subset \bar{C}(S)$ and suppose that for every open set $G \subset S$ there exists a sequence $\{f_n\} \subset M$ with $0 \leq f_n \leq \chi_G$ for each n such that $\text{bp-lim}_{n \rightarrow \infty} f_n = \chi_G$. Show that M is convergence determining.
10. Show that the collection of all twice continuously Frechet differentiable functions with bounded support on a separable Hilbert space is convergence determining.
11. Let S be locally compact and separable. Show that $M \subset \bar{C}(S)$ is convergence determining if and only if M is dense in $\bar{C}(S)$ in the supremum norm.
12. Let $x \in D_E[0, \infty)$, and for each $n \geq 1$, define $x_n \in D_E[0, \infty)$ by $x_n(t) = x(\lfloor nt \rfloor / n) \wedge n$. Show that $\lim_{n \rightarrow \infty} d(x_n, x) = 0$.
13. Let E and F be metric spaces, and let $f: E \rightarrow F$ be continuous. Show that the mapping $x \rightarrow f \circ x$ from $D_E[0, \infty)$ into $D_F[0, \infty)$ is continuous.
14. Show that $D_E[0, \infty)$ is separable whenever E is.

15. Let (E, r) be arbitrary, let $\Gamma_1 \subset \Gamma_2 \subset \dots$ be a nondecreasing sequence of compact subsets of E , and define

$$(11.6) \quad S = \{x \in D_E[0, \infty) : x(t) \in \Gamma_n \text{ for } 0 \leq t \leq n, n = 1, 2, \dots\}.$$

Show that (S, d) is complete and separable, where d is defined by (5.2).

16. Let (E, r) be complete. Show that if A is compact in $D_E[0, \infty)$, then for each $T > 0$ there exists a compact set $\Gamma_T \subset E$ such that $x(t) \in \Gamma_T$ for $0 \leq t \leq T$ and all $x \in A$.

17. Prove the following variation of Proposition 6.5.

Let $\{x_n\} \subset D_E[0, \infty)$ and $x \in D_E[0, \infty)$. Then $\lim_{n \rightarrow \infty} d(x_n, x) = 0$ if and only if whenever $t_n \geq s_n \geq 0$ for each n , $t \geq 0$, $\lim_{n \rightarrow \infty} s_n = t$, and $\lim_{n \rightarrow \infty} t_n = t$, we have

$$(11.7) \quad \lim_{n \rightarrow \infty} [r(x_n(t_n), x(t)) \vee r(x_n(s_n), x(t))] \wedge r(x_n(s_n), x(t-)) = 0$$

and

$$(11.8) \quad \lim_{n \rightarrow \infty} r(x_n(t_n), x(t)) \wedge [r(x_n(t_n), x(t-)) \vee r(x_n(s_n), x(t-))] = 0.$$

18. Let (E, r) be complete and separable. Let $\{X_a\}$ be a family of processes with sample paths in $D_E[0, \infty)$. Suppose that for every $\varepsilon > 0$ and $T > 0$ there exists a family $\{X_a^{t, T}\}$ of processes with sample paths in $D_E[0, \infty)$ (with $X_a^{t, T}$ and X_a defined on the same probability space) such that

$$(11.9) \quad \sup_a P \left\{ \sup_{0 \leq t \leq T} r(X_a^{t, T}(t), X_a(t)) \geq \varepsilon \right\} < \varepsilon$$

and $\{X_a^{t, T}\}$ is relatively compact. Show that $\{X_a\}$ is relatively compact.

19. Let (E, r) be complete and separable. Show that if $\{X_a\}$ is relatively compact in $D_E[0, \infty)$, then the compact containment condition holds.
20. Let $\{N_a\}$ be a family of right continuous counting processes (i.e., $N_a(0) = 0$, $N_a(t) - N_a(t-) = 0$ or 1 for all $t > 0$). For $k = 0, 1, \dots$, let $\tau_k^a = \inf\{t \geq 0 : N_a(t) \geq k\}$ and $\Delta_k^a = \tau_k^a - \tau_{k-1}^a$ (if $\tau_{k-1}^a < \infty$). Use Lemma 8.2 to give necessary and sufficient conditions for the relative compactness of $\{N_a\}$.
21. Let (E, r) be complete and separable, and let $\{X_a\}$ be a family of processes with sample paths in $D_E[0, \infty)$. Show that $\{X_a\}$ is relatively compact if and only if for every $\varepsilon > 0$ there exists a family $\{X_a^\varepsilon\}$ of pure jump processes (with X_a^ε and X_a defined on the same probability space) such that $\sup_a \sup_{t \geq 0} r(X_a^\varepsilon(t), X_a(t)) < \varepsilon$ a.s., $\{X_a^\varepsilon(t)\}$ is relatively compact for each rational $t \geq 0$, and $\{N_a^\varepsilon(t)\}$ is relatively compact, where $N_a^\varepsilon(t)$ is the number of jumps of X_a^ε in $(0, t]$.

22. (a) Give an example in which $\{X_n\}$ and $\{Y_n\}$ are relatively compact in $D_{\mathbb{R}}[0, \infty)$, but $\{(X_n, Y_n)\}$ is not relatively compact in $D_{\mathbb{R}^2}[0, \infty)$.
 (b) Show that if $\{X_n\}$, $\{Y_n\}$, and $\{X_n + Y_n\}$ are relatively compact in $D_{\mathbb{R}}[0, \infty)$, then $\{(X_n, Y_n)\}$ is relatively compact in $D_{\mathbb{R}^2}[0, \infty)$.
 (c) More generally, if $2 \leq r < \infty$, show that $\{(X_n^1, X_n^2, \dots, X_n^r)\}$ is relatively compact in $D_{\mathbb{R}^r}[0, \infty)$ if and only if $\{X_n^k\}$ and $\{X_n^k + X_n^l\}$ ($k, l = 1, \dots, r$) are relatively compact in $D_{\mathbb{R}}[0, \infty)$.
23. Show that $\{(X_n^1, X_n^2, \dots)\}$ is relatively compact in $D_{\mathbb{R}^\infty}[0, \infty)$ (where \mathbb{R}^∞ has the product topology) if and only if $\{(X_n^1, \dots, X_n^r)\}$ is relatively compact in $D_{\mathbb{R}^r}[0, \infty)$ for $r = 1, 2, \dots$.
24. Let (E, r) be complete and separable, and let $\{X_n\}$ be a sequence of processes with sample paths in $D_E[0, \infty)$. Let M be a subspace of $\bar{C}(E)$ that strongly separates points. Show that if the finite-dimensional distributions of X_n converge to those of a process X with sample paths in $D_E[0, \infty)$, and if $\{g \circ X_n\}$ is relatively compact in $D_{\mathbb{R}}[0, \infty)$ for every $g \in M$, then $X_n \Rightarrow X$.
25. Let (E, r) be separable, and consider the metric space $(C_E[0, \infty), d_U)$, where d_U is defined by (10.5). Let \mathcal{B} denote its Borel σ -algebra.
- For each $t \geq 0$, define $\pi_t: C_E[0, \infty) \rightarrow E$ by $\pi_t(x) = x(t)$. Show that $\mathcal{B} = \sigma(\pi_t: 0 \leq t < \infty)$.
 - Show that d_U determines the same topology on $C_E[0, \infty)$ as d (the latter defined by (5.2)).
 - Show that $C_E[0, \infty)$ is a closed subset of $D_E[0, \infty)$, hence it belongs to \mathcal{S}_E , and therefore $\mathcal{B} \subset \mathcal{S}_E$.
 - Suppose that $\{P_n\} \subset \mathcal{P}(D_E[0, \infty))$, $P \in \mathcal{P}(D_E[0, \infty))$, and $P_n(C_E[0, \infty)) = P(C_E[0, \infty)) = 1$ for each n . Define $\{Q_n\} \subset \mathcal{P}(C_E[0, \infty))$ and $Q \in \mathcal{P}(C_E[0, \infty))$ by $Q_n = P_n|_{\mathcal{B}}$ and $Q = P|_{\mathcal{B}}$. Show that $P_n \Rightarrow P$ on $D_E[0, \infty)$ if and only if $Q_n \Rightarrow Q$ on $C_E[0, \infty)$.
26. Show that each of the following functions $f_k: D_{\mathbb{R}}[0, \infty) \rightarrow D_{\mathbb{R}}[0, \infty)$ is continuous:

$$\begin{aligned}
 f_1(x)(t) &= \sup_{s \leq t} x(s), \\
 f_2(x)(t) &= \inf_{s \leq t} x(s), \\
 (11.10) \quad f_3(x)(t) &= \int_0^t x(s) \, ds, \\
 f_4(x)(t) &= \sup_{s \leq t} (x(s) - x(s-)).
 \end{aligned}$$

27. Let $\mathcal{A} \subset \mathcal{B}(S)$ be closed under finite intersections and suppose each open

- set in S is a countable union of sets in \mathcal{A} . Suppose $P, P_n \in \mathcal{P}(S)$, $n = 1, 2, \dots$, and $\lim_{n \rightarrow \infty} P_n(A) = P(A)$ for every $A \in \mathcal{A}$. Show that $P_n \Rightarrow P$.
28. Let (S, d) be complete and separable and let $\mathcal{A} \subset \mathcal{B}(S)$. Suppose for each closed set F and open set U with $F \subset U$, there exists $A \in \mathcal{A}$ such that $F \subset A \subset U$. Show that if $\{P_n\} \subset \mathcal{P}(S)$ is relatively compact and $\lim_{n \rightarrow \infty} P_n(A)$ exists for each $A \in \mathcal{A}$, then there exists $P \in \mathcal{P}(S)$ such that $P_n \Rightarrow P$.

12. NOTES

The standard reference on the topic of convergence of probability measures is Billingsley's (1968) book of that title, where additional historical remarks can be found.

As originally defined, the Prohorov (1956) metric was a symmetrized version of the present ρ . Strassen (1965) noticed that ρ is already symmetric and obtained Theorem 1.2. Lemma 1.3 is essentially due to Dudley (1968). Lemma 1.4 is a modification of the marriage lemma of Hall (1935), and is a special case of a result of Artstein (1983). Prohorov (1956) obtained Theorem 1.7. The Skorohod (1956) representation theorem (Theorem 1.8) originally required that (S, d) be complete; Dudley (1968) removed this assumption. For a recent somewhat stronger result see Blackwell and Dubins (1983). The continuous mapping theorem (Corollary 1.9) can be attributed to Mann and Wald (1943) and Chernoff (1956).

Theorem 2.2 is of course due to Prohorov (1956).

Theorem 3.1 (without (a)) is known as the Portmanteau theorem and goes back to Alexandroff (1940–1943); the equivalence of (a) is due to Prohorov (1956) assuming completeness and to Dudley (1968) in general. Corollary 3.3 is called Slutsky's theorem.

The topology on $D_E[0, \infty)$ is Stone's (1963) analogue of Skorohod's (1956) J_1 topology. Metrizability was first shown by Prohorov (1956). The metric d is analogous to Billingsley's (1968) d_0 on $D[0, 1]$. Theorem 5.6 is essentially due to Kolmogorov (1956).

With a different modulus of continuity, Theorem 6.3 was proved by Prohorov (1956); in its present form, it is due to Billingsley (1968).

Similar remarks apply to Theorem 7.2.

Condition (b) of Theorem 8.6 for relative compactness is due to Kurtz (1975), as are the results preceding it in Section 8; Aldous (1978) is responsible for condition (c). See also Jacod, Mémin, and Métivier (1983). Theorem 8.8 is due to Chenčov (1956).

The results of Section 9 are based on Kurtz (1975).

Proposition 10.4 was proved by Sato (1977).

Problem 5 is due to Lindvall (1974) and can be derived as a consequence of Theorem 4.2 of Billingsley (1968).

4 GENERATORS AND MARKOV PROCESSES

In this chapter we study Markov processes from the standpoint of the generators of their corresponding semigroups. In Section 1 we give the basic definitions of a Markov process, its transition function, and its corresponding semigroup, show that a transition function and initial distribution uniquely determine a Markov process, and verify the important martingale relationship between a Markov process and its generator. Section 2 is devoted to the study of Feller processes and the properties of their sample paths, and Sections 3 through 7 to the martingale problem as a means of characterizing the Markov process corresponding to a given generator. In Section 8 we exploit the characterization of a Markov process by its generator (either through the determination of its semigroup or as the unique solution of a martingale problem) to give general conditions for the weak convergence of a sequence of processes to a Markov process. Stationary distributions are the subject of Section 9. Some conditions under which sums of generators characterize Markov processes are given in Section 10.

Throughout this chapter E is a metric space, $M(E)$ is the collection of all real-valued, Borel measurable functions on E , $B(E) \subset M(E)$ is the Banach space of bounded functions with $\|f\| = \sup_{x \in E} |f(x)|$, and $\bar{C}(E) \subset B(E)$ is the subspace of bounded continuous functions.

1. MARKOV PROCESSES AND TRANSITION FUNCTIONS

Let $\{X(t), t \geq 0\}$ be a stochastic process defined on a probability space (Ω, \mathcal{F}, P) with values in E , and let $\mathcal{F}_t^X = \sigma(X(s): s \leq t)$. Then X is a *Markov process* if

$$(1.1) \quad P\{X(t+s) \in \Gamma | \mathcal{F}_t^X\} = P\{X(t+s) \in \Gamma | X(t)\}$$

for all $s, t \geq 0$ and $\Gamma \in \mathcal{B}(E)$. If $\{\mathcal{G}_t\}$ is a filtration with $\mathcal{F}_t^X \subset \mathcal{G}_t$, $t \geq 0$, then X is a *Markov process with respect to $\{\mathcal{G}_t\}$* if (1.1) holds with \mathcal{F}_t^X replaced by \mathcal{G}_t . (Of course if X is a Markov process with respect to $\{\mathcal{G}_t\}$, then it is a Markov process.) Note that (1.1) implies

$$(1.2) \quad E[f(X(t+s)) | \mathcal{F}_t^X] = E[f(X(t+s)) | X(t)]$$

for all $s, t \geq 0$ and $f \in B(E)$.

A function $P(t, x, \Gamma)$ defined on $[0, \infty) \times E \times \mathcal{B}(E)$ is a *time homogeneous transition function* if

$$(1.3) \quad P(t, x, \cdot) \in \mathcal{P}(E), \quad (t, x) \in [0, \infty) \times E,$$

$$(1.4) \quad P(0, x, \cdot) = \delta_x \quad (\text{unit mass at } x), \quad x \in E,$$

$$(1.5) \quad P(\cdot, \cdot, \Gamma) \in B([0, \infty) \times E), \quad \Gamma \in \mathcal{B}(E),$$

and

$$(1.6) \quad P(t+s, x, \Gamma) = \int P(s, y, \Gamma)P(t, x, dy), \quad s, t \geq 0, x \in E, \Gamma \in \mathcal{B}(E).$$

A transition function $P(t, x, \Gamma)$ is a *transition function for a time-homogeneous Markov process X* if

$$(1.7) \quad P\{X(t+s) \in \Gamma | \mathcal{F}_t^X\} = P(s, X(t), \Gamma)$$

for all $s, t \geq 0$ and $\Gamma \in \mathcal{B}(E)$, or equivalently, if

$$(1.8) \quad E[f(X(t+s)) | \mathcal{F}_t^X] = \int f(y)P(s, X(t), dy)$$

for all $s, t \geq 0$ and $f \in B(E)$.

To see that (1.6), called the *Chapman–Kolmogorov property*, is a reasonable assumption given (1.7), observe that (1.7) implies

$$\begin{aligned} (1.9) \quad P(t+s, X(u), \Gamma) &= P\{X(u+t+s) \in \Gamma | \mathcal{F}_u^X\} \\ &= E[P\{X(u+t+s) \in \Gamma | \mathcal{F}_{u+t}^X\} | \mathcal{F}_u^X] \\ &= E[P(s, X(u+t), \Gamma) | \mathcal{F}_u^X] \\ &= \int P(s, y, \Gamma)P(t, X(u), dy) \end{aligned}$$

for all $s, t, u \geq 0$ and $\Gamma \in \mathcal{B}(E)$.

The probability measure $v \in \mathcal{P}(E)$ given by $v(\Gamma) = P\{X(0) \in \Gamma\}$ is called the *initial distribution* of X .

A transition function for X and the initial distribution determine the finite-dimensional distributions of X by

$$(1.10) \quad P\{X(0) \in \Gamma_0, X(t_1) \in \Gamma_1, \dots, X(t_n) \in \Gamma_n\}$$

$$= \int_{\Gamma_0} \cdots \int_{\Gamma_{n-1}} P(t_n - t_{n-1}, y_{n-1}, \Gamma_n) P(t_{n-1} - t_{n-2}, y_{n-2}, dy_{n-1}) \cdots P(t_1, y_0, dy_1)v(dy_0).$$

In particular we have the following theorem.

1.1 Theorem Let $P(t, x, \Gamma)$ satisfy (1.3)–(1.6) and let $v \in \mathcal{P}(E)$. If for each $t \geq 0$ the probability measure $\int P(t, x, \cdot)v(dx)$ is tight (which will be the case if (E, r) is complete and separable by Lemma 2.1 of Chapter 3), then there exists a Markov process X in E whose finite-dimensional distributions are uniquely determined by (1.10).

Proof. For $I \subset [0, \infty)$, let E^I denote the product space $\prod_{s \in I} E_s$, where for each s , $E_s = E$, and let \mathcal{P}_I denote the collection of probability measures defined on the product σ -algebra $\prod_{s \in I} \mathcal{B}(E_s)$. For $s \in I$, let $X(s)$ denote the coordinate random variable. Note that $\prod_{s \in I} \mathcal{B}(E_s) = \sigma(X(s); s \in I)$.

Let $\{s_i, i = 0, 1, 2, \dots\} \subset [0, \infty)$ satisfy $s_i \neq s_j$ for $i \neq j$ and $s_0 = 0$, and fix $x_0 \in E$. For $n > 1$, let $t_1 < t_2 < \cdots < t_n$ be the increasing rearrangement of s_1, \dots, s_n . Then it follows from Tulcea's theorem (Appendix 9) that there exists $P_n \in \mathcal{P}_{\{s_i\}}$ such that $P_n\{X(0) \in \Gamma_0, X(t_1) \in \Gamma_1, \dots, X(t_n) \in \Gamma_n\}$ equals the right side of (1.10) and $P_n\{X(s_i) = x_0\} = 1$ for $i > n$. Tulcea's theorem gives a measure Q_n on $E^{[s_1, \dots, s_n]}$. Fix $x_0 \in E$ and define $P_n = Q_n \times \delta_{(x_0, x_0, \dots)}$ on $E^I = E^{[s_1, \dots, s_n]} \times E^{[s_n+1, s_n+2, \dots]}$. The sequence $\{P_n\}$ is tight by Proposition 2.4 of Chapter 3, and any limit point will satisfy (1.10) for $\{t_1, \dots, t_n\} \subset \{s_i\}$. Consequently $\{P_n\}$ converges weakly to $P_{\{s_i\}} \in \mathcal{P}_{\{s_i\}}$.

By Problem 27 of Chapter 2 for $B \in \mathcal{B}(E)^{[0, \infty)}$ there exists a countable subset $\{s_i\} \subset [0, \infty)$ such that $B \in \sigma(X(s_i); i = 1, 2, \dots)$, that is, there exists $\hat{B} \in \mathcal{B}(E)^{\{s_i\}}$ such that $B = \{(X(s_1), X(s_2), \dots) \in \hat{B}\}$. Define $P(B) \equiv P^{\{s_i\}}(\hat{B})$. We leave it to the reader to verify the consistency of this definition and to show that X , defined on $(E^{[0, \infty)}, \mathcal{B}(E)^{[0, \infty)}, P)$, is Markov. \square

Let P_x denote the measure on $\mathcal{B}(E)^{[0, \infty)}$ given by Theorem 1.1 with $v = \delta_x$, and let X be the corresponding coordinate process, that is, $X(t, \omega) = \omega(t)$. It follows from (1.5) and (1.10) that $P_x(B)$ is a Borel measurable function of x for $B = \{X(0) \in \Gamma_0, \dots, X(t_n) \in \Gamma_n\}$, $0 < t_1 < t_2 < \cdots < t_n$, $\Gamma_0, \Gamma_1, \dots, \Gamma_n \in \mathcal{B}(E)$. In fact this is true for all $B \in \mathcal{B}(E)^{[0, \infty)}$.

1.2 Proposition Let P_x be as above. Then $P_x(B)$ is a Borel measurable function of x for each $B \in \mathcal{B}(E)^{[0, \infty)}$.

Proof. The collection of B for which $P_x(B)$ is a Borel measurable function is a Dynkin class containing all sets of the form $\{X(0) \in \Gamma_0, \dots, X(t_n) \in \Gamma_n\} \in \mathcal{B}(E)^{[0, \infty)}$, and hence all $B \in \mathcal{B}(E)^{[0, \infty)}$. (See Appendix 4.) \square

Let $\{Y(n), n = 0, 1, 2, \dots\}$ be a discrete-parameter process defined on (Ω, \mathcal{F}, P) with values in E , and let $\mathcal{F}_n^Y = \sigma(Y(k): k \leq n)$. Then Y is a *Markov chain* if

$$(1.11) \quad P\{Y(n+m) \in \Gamma | \mathcal{F}_n^Y\} = P\{Y(n+m) \in \Gamma | Y(n)\}$$

for all $m, n \geq 0$ and $\Gamma \in \mathcal{B}(E)$. A function $\mu(x, \Gamma)$ defined on $E \times \mathcal{B}(E)$ is a *transition function* if

$$(1.12) \quad \mu(x, \cdot) \in \mathcal{P}(E), \quad x \in E,$$

and

$$(1.13) \quad \mu(\cdot, \Gamma) \in \mathcal{B}(E), \quad \Gamma \in \mathcal{B}(E).$$

A transition function $\mu(x, \Gamma)$ is a *transition function for a time-homogeneous Markov chain* Y if

$$(1.14) \quad P\{Y(n+1) \in \Gamma | \mathcal{F}_n^Y\} = \mu(Y(n), \Gamma), \quad n \geq 0, \Gamma \in \mathcal{B}(E).$$

Note that

$$(1.15) \quad P\{Y(n+m) \in \Gamma | \mathcal{F}_n^Y\} = \int \dots \int \mu(y_{m-1}, \Gamma) \mu(y_{m-2}, dy_{m-1}) \dots \mu(Y(n), dy_1).$$

As before, the probability measure $v \in \mathcal{P}(E)$ given by $v(\Gamma) = P\{Y(0) \in \Gamma\}$ is called the initial distribution for Y . The analogues of Theorem 1.1 and Proposition 1.2 are left to the reader.

Let $\{X(t), t \geq 0\}$, defined on (Ω, \mathcal{F}, P) , be an E -valued Markov process with respect to a filtration $\{\mathcal{G}_t\}$ such that X is $\{\mathcal{G}_t\}$ -progressive. Suppose $P(t, x, \Gamma)$ is a transition function for X , and let τ be a $\{\mathcal{G}_t\}$ -stopping time with $\tau < \infty$ a.s. Then X is *strong Markov at τ* if

$$(1.16) \quad P\{X(\tau + t) \in \Gamma | \mathcal{G}_\tau\} = P(t, X(\tau), \Gamma)$$

for all $t \geq 0$ and $\Gamma \in \mathcal{B}(E)$, or equivalently, if

$$(1.17) \quad E[f(X(\tau + t)) | \mathcal{G}_\tau] = \int f(y) P(t, X(\tau), dy)$$

for all $t \geq 0$ and $f \in \mathcal{B}(E)$. X is a *strong Markov process with respect to $\{\mathcal{G}_t\}$* if X is strong Markov at τ for all $\{\mathcal{G}_t\}$ -stopping times τ with $\tau < \infty$ a.s.

1.3 Proposition Let X be E -valued, $\{\mathcal{G}_t\}$ -progressive, and $\{\mathcal{G}_t\}$ -Markov, and let $P(t, x, \Gamma)$ be a transition function for X . Let τ be a discrete $\{\mathcal{G}_t\}$ -stopping time with $\tau < \infty$ a.s. Then X is strong Markov at τ .

Proof. Let t_1, t_2, \dots be the values assumed by τ , and let $f \in B(E)$. If $B \in \mathcal{G}_\tau$, then $B \cap \{\tau = t_i\} \in \mathcal{G}_{t_i}$ and hence for $t \geq 0$,

$$(1.18) \quad \begin{aligned} \int_{B \cap \{\tau = t_i\}} f(X(\tau + t)) dP &= \int_{B \cap \{\tau = t_i\}} f(X(t_i + t)) dP \\ &= \int_{B \cap \{\tau = t_i\}} \int f(y) P(t, X(t_i), dy) dP \\ &= \int_{B \cap \{\tau = t_i\}} \int f(y) P(t, X(\tau), dy) dP. \end{aligned}$$

Summing over i we obtain

$$(1.19) \quad \int_B f(X(\tau + t)) dP = \int_B \int f(y) P(t, X(\tau), dy) dP$$

for all $B \in \mathcal{G}_\tau$, which implies (1.17). \square

1.4 Remark Recall (Proposition 1.3 of Chapter 2) that every stopping time is the limit of a decreasing sequence of discrete stopping times. This fact can frequently be exploited to extend Proposition 1.3 to more-general stopping times. See, for example, Theorem 2.7 below.

1.5 Proposition Let X be E -valued, $\{\mathcal{G}_t\}$ -progressive, and $\{\mathcal{G}_t\}$ -Markov, and let $P(t, x, \Gamma)$ be a transition function for X . Let τ be a $\{\mathcal{G}_t\}$ -stopping time, suppose X is strong Markov at $\tau + t$ for all $t \geq 0$, and let $B \in \mathcal{B}(E)^{[0, \infty)}$. Then

$$(1.20) \quad P\{X(\tau + \cdot) \in B | \mathcal{G}_\tau\} = P_{X(\tau)}(B).$$

Proof. First consider B of the form

$$(1.21) \quad \{x \in E^{[0, \infty)} : x(t_i) \in \Gamma_i, i = 1, \dots, n\}$$

where $0 \leq t_1 \leq t_2 \leq \dots \leq t_n$, $\Gamma_1, \Gamma_2, \dots, \Gamma_n \in \mathcal{B}(E)$. Proceeding by induction on n , for $B = \{x \in E^{[0, \infty)} : x(t_1) \in \Gamma_1\}$ we have

$$(1.22) \quad P\{X(\tau + \cdot) \in B | \mathcal{G}_\tau\} = P\{X(\tau + t_1) \in \Gamma_1 | \mathcal{G}_\tau\} = P(t_1, X(\tau), \Gamma_1)$$

by (1.16), but this is just (1.20). Suppose now that (1.20) holds for all B of the

form (1.21) for some fixed n . Then

$$(1.23) \quad E\left[\prod_{i=1}^n f_i(X(\tau + t_i)) \middle| \mathcal{G}_\tau\right] \\ = \int \cdots \int \prod_{i=1}^n f_i(y_i) P(t_n - t_{n-1}, y_{n-1}, dy_n) \cdots P(t_1, X(\tau), dy_1)$$

for $0 \leq t_1 \leq t_2 \leq \cdots \leq t_n$ and $f_i \in B(E)$. Let B be of the form (1.21) with $n+1$ in place of n . Then

$$(1.24) \quad P\{B | \mathcal{G}_\tau\} = E\left[\prod_{i=1}^{n+1} \chi_{\Gamma_i}(X(\tau + t_i)) \middle| \mathcal{G}_\tau\right] \\ = E\left[E[\chi_{\Gamma_{n+1}}(X(\tau + t_{n+1})) | \mathcal{G}_{\tau+t_n}] \prod_{i=1}^n \chi_{\Gamma_i}(X(\tau + t_i)) \middle| \mathcal{G}_\tau\right] \\ = E\left[P(t_{n+1} - t_n, X(\tau + t_n), \Gamma_{n+1}) \prod_{i=1}^n \chi_{\Gamma_i}(X(\tau + t_i)) \middle| \mathcal{G}_\tau\right] \\ = \int \cdots \int P(t_{n+1} - t_n, y_n, \Gamma_{n+1}) \chi_{\Gamma_n}(y_n) \prod_{i=1}^{n-1} \chi_{\Gamma_i}(y_i) \\ \times P(t_n - t_{n-1}, y_{n-1}, dy_n) \cdots P(t_1, X(\tau), dy_1) \\ = \int_{\Gamma_1} \cdots \int_{\Gamma_n} P(t_{n+1} - t_n, y_n, \Gamma_{n+1}) \\ \times P(t_n - t_{n-1}, y_{n-1}, dy_n) \cdots P(t_1, X(\tau), dy_1) \\ = P_{X(\tau)}(B).$$

Therefore (1.20) holds for all B of the form (1.21). The proposition now follows by the Dynkin class theorem. \square

Ordinarily, formulas for transition functions cannot be obtained, the most notable exception being that of one-dimensional Brownian motion given by

$$(1.25) \quad P(t, x, \Gamma) = \int_{\Gamma} \frac{1}{\sqrt{2\pi t}} \exp\left\{-\frac{(y-x)^2}{2t}\right\} dy.$$

Consequently, directly defining a transition function is usually not a useful method of specifying a Markov process. In this chapter and in Chapters 5 and 6, we consider other methods of specifying processes. In particular, in this chapter we exploit the fact that

$$(1.26) \quad T(t)f(x) \equiv \int f(y)P(t, x, dy)$$

defines a measurable contraction semigroup on $B(E)$ (by the Chapman-Kolmogorov property (1.6)).

Let $\{T(t)\}$ be a semigroup on a closed subspace $L \subset B(E)$. With reference to (1.8), we say that an E -valued Markov process X corresponds to $\{T(t)\}$ if

$$(1.27) \quad E[f(X(t+s)) | \mathcal{F}_t^X] = T(s)f(X(t))$$

for all $s, t \geq 0$ and $f \in L$. Of course if $\{T(t)\}$ is given by a transition function as in (1.26), then (1.27) is just (1.8).

1.6 Proposition Let E be separable. Let X be an E -valued Markov process with initial distribution v corresponding to a semigroup $\{T(t)\}$ on a closed subspace $L \subset B(E)$. If L is separating, then $\{T(t)\}$ and v determine the finite-dimensional distributions of X .

Proof. For $f \in L$ and $t \geq 0$, we have

$$\begin{aligned} (1.28) \quad \int f(y)P\{X(t) \in dy\} &= E[f(X(t))] \\ &= E[E[f(X(t))] | \mathcal{F}_0^X] \\ &= E[T(t)f(X(0))] = \int T(t)f(x)v(dx). \end{aligned}$$

Since L is separating, $v_t(\Gamma) \equiv P\{X(t) \in \Gamma\}$ is determined. Similarly if $f \in L$ and $g \in B(E)$, then for $0 \leq t_1 < t_2$,

$$\begin{aligned} (1.29) \quad E[f(X(t_1))g(X(t_2))] &= E[f(X(t_1))T(t_2 - t_1)g(X(t_1))] \\ &= \int f(x)T(t_2 - t_1)g(x)v_{t_1}(dx) \end{aligned}$$

and the joint distribution of $X(t_1)$ and $X(t_2)$ is determined (cf. Proposition 4.6 of Chapter 3). Proceeding in this manner, the proposition can be proved by induction. \square

Since the finite-dimensional distributions of a Markov process are determined by a corresponding semigroup $\{T(t)\}$, they are in turn determined by its full generator \hat{A} or by a sufficiently large set $A \subset \hat{A}$. One of the best approaches for determining when a set is "sufficiently large" is through the martingale problem of Stroock and Varadhan, which is based on the observation in the following proposition.

1.7 Proposition Let X be an E -valued, progressive Markov process with transition function $P(t, x, \Gamma)$ and let $\{T(t)\}$ and \hat{A} be as above. If $(f, g) \in \hat{A}$ then

$$(1.30) \quad M(t) = f(X(t)) - \int_0^t g(X(s)) \, ds$$

is an $\{\mathcal{F}_t^X\}$ -martingale.

Proof. For each $t, u \geq 0$

$$\begin{aligned} (1.31) \quad E[M(t+u) | \mathcal{F}_t^X] &= \int f(y)P(u, X(t), dy) \\ &\quad - \int_t^{t+u} \int g(y)P(s-t, X(t), dy) \, ds \\ &\quad - \int_0^t g(X(s)) \, ds \\ &= T(u)f(X(t)) - \int_0^u T(s)g(X(t)) \, ds \\ &\quad - \int_0^t g(X(s)) \, ds \\ &= f(X(t)) - \int_0^t g(X(s)) \, ds = M(t). \end{aligned} \quad \square$$

We study the basic properties of the martingale problem in Sections 3–7.

2. MARKOV JUMP PROCESSES AND FELLER PROCESSES

The simplest Markov process to describe is a Markov jump process with a bounded generator. Let $\mu(x, \Gamma)$ be a transition function on $E \times \mathcal{B}(E)$ and let $\lambda \in B(E)$ be nonnegative. Then

$$(2.1) \quad Af(x) = \lambda(x) \int (f(y) - f(x))\mu(x, dy)$$

defines a bounded linear operator A on $B(E)$, and A is the generator for a Markov process in E that can be constructed as follows.

Let $\{Y(k), k = 0, 1, \dots\}$ be a Markov chain in E with initial distribution v and transition function $\mu(x, \Gamma)$. That is, $P\{Y(0) \in \Gamma\} = v(\Gamma)$ and

$$(2.2) \quad P\{Y(k+1) \in \Gamma | Y(0), \dots, Y(k)\} = \mu(Y(k), \Gamma)$$

for all $\Gamma \in \mathcal{B}(E)$ and $k = 0, 1, \dots$. Let $\Delta_0, \Delta_1, \dots$ be independent and exponentially distributed with parameter 1 (and independent of $Y(\cdot)$). Then

$$(2.3) \quad X(t) = \begin{cases} Y(0), & 0 \leq t < \frac{\Delta_0}{\lambda(Y(0))}, \\ Y(k), & \sum_{j=0}^{k-1} \frac{\Delta_j}{\lambda(Y(j))} \leq t < \sum_{j=0}^k \frac{\Delta_j}{\lambda(Y(j))}, \end{cases}$$

defines a Markov process X in E with initial distribution v and generator A . (Note that we allow $\lambda(x) = 0$, taking $\Delta/0 = \infty$.)

To see this, we make use of an even simpler representation. Let $\lambda = \sup_{x \in E} \lambda(x)$, and to avoid trivialities, assume that $\lambda > 0$. Define the transition function $\mu'(x, \Gamma)$ on $E \times \mathcal{B}(E)$ by

$$(2.4) \quad \mu'(x, \Gamma) = \left(1 - \frac{\lambda(x)}{\lambda}\right) \delta_x(\Gamma) + \frac{\lambda(x)}{\lambda} \mu(x, \Gamma),$$

and note that (2.1) can be rewritten as

$$(2.5) \quad Af(x) = \lambda \int (f(y) - f(x)) \mu'(x, dy).$$

Let $\{Y'(k), k = 0, 1, \dots\}$ be a Markov chain in E with initial distribution v and transition function $\mu'(x, \Gamma)$, and let V be an independent Poisson process with parameter λ . Define

$$(2.6) \quad X'(t) = Y'(V(t)), \quad t \geq 0.$$

We leave it to the reader to show that X and X' have the same finite-dimensional distributions (Problem 4).

Observe that

$$(2.7) \quad Pf(x) = \int f(y) \mu'(x, dy)$$

defines a linear contraction P on $B(E)$ and that, by (2.5), $A = \lambda(P - I)$. Consequently, the semigroup $\{T(t)\}$ generated by A is given by

$$(2.8) \quad T(t) = \sum_{k=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^k}{k!} P^k.$$

Let $f \in B(E)$. By the Markov property of Y' (cf. (1.15)),

$$(2.9) \quad E[f(Y'(k+l)) | Y'(0), \dots, Y'(l)] = P^k f(Y'(l))$$

for $k, l = 0, 1, \dots$, and we claim that

$$(2.10) \quad E[f(Y'(k + V(t))) | \mathcal{F}_t] = P^k f(X'(t))$$

for $k = 0, 1, \dots$ and $t \geq 0$, where

$$(2.11) \quad \mathcal{F}_t = \mathcal{F}_t^Y \vee \mathcal{F}_t^X.$$

To see this, let $A \in \mathcal{F}_t^Y$ and $B \in \mathcal{F}_t^X$. Then

$$\begin{aligned} (2.12) \quad & \int_{A \cap B \cap \{V(t) = l\}} f(Y'(k + V(t))) \, dP \\ &= \int_{A \cap B \cap \{V(t) = l\}} f(Y'(k + l)) \, dP \\ &= P(A \cap \{V(t) = l\}) \int_B f(Y'(k + l)) \, dP \\ &= P(A \cap \{V(t) = l\}) \int_B P^k f(Y'(l)) \, dP \\ &= \int_{A \cap B \cap \{V(t) = l\}} P^k f(X'(t)) \, dP. \end{aligned}$$

Since $\{A \cap B \cap \{V(t) = l\} : A \in \mathcal{F}_t^Y, B \in \mathcal{F}_t^X, l = 0, 1, 2, \dots\}$ is closed under finite intersections and generates \mathcal{F}_t , by the Dynkin class theorem (Appendix 4) we have

$$(2.13) \quad \int_A f(Y'(k + V(t))) \, dP = \int_A P^k f(X'(t)) \, dP$$

for all $A \in \mathcal{F}_t$, and (2.10) follows. Finally, since V has independent increments,

$$\begin{aligned} (2.14) \quad & E[f(X'(t + s)) | \mathcal{F}_t] \\ &= E[f(Y'(V(t + s) - V(t) + V(t))) | \mathcal{F}_t] \\ &= \sum_{k=0}^{\infty} e^{-\lambda s} \frac{(\lambda s)^k}{k!} E[f(Y'(k + V(t))) | \mathcal{F}_t] \\ &= \sum_{k=0}^{\infty} e^{-\lambda s} \frac{(\lambda s)^k}{k!} P^k f(X'(t)) \\ &= T(s)f(X'(t)) \end{aligned}$$

for all $s, t \geq 0$. Hence X' is a Markov process in E with initial distribution ν corresponding to the semigroup $\{T(t)\}$ generated by A .

We now assume E is locally compact and consider Markov processes with semigroups that are strongly continuous on the Banach space $C(E)$ of continuous functions vanishing at infinity with norm $\|f\| = \sup_{x \in E} |f(x)|$. Note

that $\hat{C}(E) = C(E)$ if E is compact. Let $\Delta \notin E$ be the point at infinity if E is noncompact and an isolated point if E is compact, and put $E^\Delta = E \cup \{\Delta\}$; in the noncompact case, E^Δ is the one-point compactification of E . We note that if E is also separable, then E^Δ is metrizable. (See, for example, pages 201 and 202 of Cohn (1980).)

A semigroup $\{T(t)\}$ on $\hat{C}(E)$ is said to be *positive* if $T(t)$ is a positive operator for each $t \geq 0$. (A *positive operator* is one that maps nonnegative functions to nonnegative functions.)

An operator A on $\hat{C}(E)$ is said to satisfy the *positive maximum principle* if whenever $f \in \mathcal{D}(A)$, $x_0 \in E$, and $\sup_{x \in E} f(x) = f(x_0) \geq 0$, we have $Af(x_0) \leq 0$.

2.1 Lemma Let E be locally compact. A linear operator A on $\hat{C}(E)$ satisfying the positive maximum principle is dissipative.

Proof. Let $f \in \mathcal{D}(A)$ and $\lambda > 0$. There exists $x_0 \in E$ such that $|f(x_0)| = \|f\|$. Suppose $f(x_0) \geq 0$ (otherwise replace f by $-f$). Since $\sup_{x \in E} f(x) = f(x_0) \geq 0$, $Af(x_0) \leq 0$ and hence

$$(2.15) \quad \|\lambda f - Af\| \geq \lambda f(x_0) - Af(x_0) \geq \lambda f(x_0) = \lambda \|f\|. \quad \square$$

We restate the Hille–Yosida theorem in the present context.

2.2 Theorem Let E be locally compact. The closure \bar{A} of a linear operator A on $\hat{C}(E)$ is single-valued and generates a strongly continuous, positive, contraction semigroup $\{T(t)\}$ on $\hat{C}(E)$ if and only if:

- (a) $\mathcal{D}(A)$ is dense in $\hat{C}(E)$.
- (b) A satisfies the positive maximum principle.
- (c) $\mathcal{D}(\lambda - A)$ is dense in $\hat{C}(E)$ for some $\lambda > 0$.

Proof. The necessity of (a) and (c) follows from Theorem 2.12 of Chapter 1. As for (b), if $f \in \mathcal{D}(A)$, $x_0 \in E$, and $\sup_{x \in E} f(x) = f(x_0) \geq 0$, then

$$(2.16) \quad T(t)f(x_0) \leq T(t)(f^+)(x_0) \leq \|f^+\| = f(x_0)$$

for each $t \geq 0$, so $Af(x_0) \leq 0$.

Conversely, suppose A satisfies (a)–(c). Since (b) implies A is dissipative by Lemma 2.1, \bar{A} is single-valued and generates a strongly continuous contraction semigroup $\{T(t)\}$ by Theorem 2.12 of Chapter 1. To complete the proof, we must show that $\{T(t)\}$ is positive.

Let $f \in \mathcal{D}(\bar{A})$ and $\lambda > 0$, and suppose that $\inf_{x \in E} f(x) < 0$. Choose $\{f_n\} \subset \mathcal{D}(A)$ such that $(\lambda - A)f_n \rightarrow (\lambda - \bar{A})f$, and let $x_n \in E$ and $x_0 \in E$ be points at which f_n and f , respectively, take on their minimum values. Then

$$\begin{aligned} (2.17) \quad \inf_{x \in E} (\lambda - \bar{A})f(x) &\leq \lim_{n \rightarrow \infty} (\lambda - A)f_n(x_n) \\ &\leq \lim_{n \rightarrow \infty} \lambda f_n(x_n) \\ &= \lambda f(x_0) \\ &< 0, \end{aligned}$$

where the second inequality is due to the fact that $\inf_{x \in E} f_n(x) = f_n(x_n) \leq 0$ for n sufficiently large. We conclude that if $f \in \mathcal{D}(\bar{A})$ and $\lambda > 0$, then $(\lambda - \bar{A})f \geq 0$ implies $f \geq 0$, so the positivity of $\{T(t)\}$ is a consequence of Corollary 2.8 of Chapter 1. \square

An operator $A \subset B(E) \times B(E)$ (possibly multivalued) is said to be conservative if $(1, 0)$ is in the bp-closure of A . For example, if $(1, 0)$ is in the full generator of a measurable contraction semigroup $\{T(t)\}$, then $T(t)1 = 1$ for all $t \geq 0$, and conversely. For semigroups given by transition functions, this property is just the fact that $P(t, x, E) = 1$.

A strongly continuous, positive, contraction semigroup on $\hat{C}(E)$ whose generator is conservative is called a *Feller semigroup*. Our aim in this section is to show (assuming in addition that E is separable) that every Feller semigroup on $\hat{C}(E)$ corresponds to a Markov process with sample paths in $D_E[0, \infty)$. First, however, we require several preliminary results, including our first convergence theorem.

2.3 Lemma Let E be locally compact and separable and let $\{T(t)\}$ be a strongly continuous, positive, contraction semigroup on $\hat{C}(E)$. Define the operator $T^\Delta(t)$ on $C(E^\Delta)$ for each $t \geq 0$ by

$$(2.18) \quad T^\Delta(t)f = f(\Delta) + T(t)(f - f(\Delta)).$$

(We do not distinguish notationally between functions on E^Δ and their restrictions to E .) Then $\{T^\Delta(t)\}$ is a Feller semigroup on $C(E^\Delta)$.

Proof. It is easy to verify that $\{T^\Delta(t)\}$ is a strongly continuous semigroup on $C(E^\Delta)$. Fix $t \geq 0$. To show that $T^\Delta(t)$ is a positive operator, we must show that if $\alpha \in \mathbb{R}$, $f \in \hat{C}(E)$, and $\alpha + f \geq 0$, then $\alpha + T^\Delta(t)f \geq 0$. By the positivity of $T(t)$, $T(t)(f^+) \geq 0$ and $T(t)(f^-) \geq 0$. Hence $-\bar{T}(t)f \leq T(t)(f^-)$, and so $(T(t)f)^+ \leq T(t)(f^-)$. Since $T(t)$ is a contraction, $\|T(t)(f^-)\| \leq \|f^-\| \leq \alpha$. Therefore $(T(t)f)^+ \leq \alpha$, so $\alpha + T(t)f \geq 0$.

Next, the positivity of $T^\Delta(t)$ gives $|T^\Delta(t)f| \leq T^\Delta(t)\|f\| = \|f\|$ for all $f \in C(E^\Delta)$, so $\|T^\Delta(t)\| = 1$. Finally, the generator A^Δ of $\{T^\Delta(t)\}$ clearly contains $(1, 0)$. \square

2.4 Proposition Let E be locally compact and separable. Let $\{T(t)\}$ be a strongly continuous, positive, contraction semigroup on $\hat{C}(E)$, and define the semigroup $\{T^\Delta(t)\}$ on $C(E^\Delta)$ as in Lemma 2.3. Let X be a Markov process corresponding to $\{T^\Delta(t)\}$ with sample paths in $D_{E^\Delta}[0, \infty)$, and let $\tau = \inf\{t \geq 0 : X(t) = \Delta \text{ or } X(t-) = \Delta\}$. Then

$$(2.19) \quad P\{\tau < \infty, X(\tau + s) = \Delta \text{ for all } s \geq 0\} = P\{\tau < \infty\}.$$

Let A be the generator of $\{T(t)\}$ and suppose further that A is conservative. If $P\{X(0) \in E\} = 1$, then $P\{X \in D_E[0, \infty)\} = 1$.

Proof. Recalling that E^Δ is metrizable, there exists $g \in C(E^\Delta)$ with $g > 0$ on E and $g(\Delta) = 0$. Put $f = \int_0^\infty e^{-u} T^\Delta(u)g \, du$, and note that $f > 0$ on E and $f(\Delta) = 0$. By the Markov property of X ,

$$\begin{aligned} (2.20) \quad E[e^{-t}f(X(t)) | \mathcal{F}_{s+}^X] &= e^{-t}T^\Delta(t-s)f(X(s)) \\ &= e^{-s} \int_{t-s}^\infty e^{-u} T^\Delta(u)g(X(s)) \, du \\ &\leq e^{-s}f(X(s)), \quad 0 \leq s < t, \end{aligned}$$

so $e^{-t}f(X(t))$ is a nonnegative $\{\mathcal{F}_{s+}^X\}$ -supermartingale. Therefore, (2.19) is a consequence of Proposition 2.15 of Chapter 2. It also follows that $P\{X(t) = \Delta\} = P\{\tau \leq t\}$ for all $t \geq 0$.

Let A^Δ denote the generator of $\{T^\Delta(t)\}$. The assumption that A is conservative (which refers to the bp-closure of A in $B(E) \times B(E)$) implies that $(\chi_E, 0)$ is in the bp-closure of A^Δ (considering A^Δ as a subspace of $B(E^\Delta) \times B(E^\Delta)$). Since the collection of $(f, g) \in B(E^\Delta) \times B(E^\Delta)$ satisfying

$$(2.21) \quad E[f(X(t))] = E[f(X(0))] + E\left[\int_0^t g(X(s)) \, ds\right]$$

is bp-closed and contains A^Δ , for all $t \geq 0$ we have

$$(2.22) \quad P\{\tau > t\} = P\{X(t) \in E\} = P\{X(0) \in E\},$$

and if $P\{X(0) \in E\} = 1$, we conclude that $P\{X \in D_E[0, \infty)\} = P\{\tau = \infty\} = 1$. \square

A converse to the second assertion of Proposition 2.4 is provided by Corollary 2.8.

2.5 Theorem Let E be locally compact and separable. For $n = 1, 2, \dots$ let $\{T_n(t)\}$ be a Feller semigroup on $\hat{C}(E)$, and suppose X_n is a Markov process corresponding to $\{T_n(t)\}$ with sample paths in $D_E[0, \infty)$. Suppose that $\{T(t)\}$ is a Feller semigroup on $\hat{C}(E)$ and that for each $f \in \hat{C}(E)$,

$$(2.23) \quad \lim_{n \rightarrow \infty} T_n(t)f = T(t)f, \quad t \geq 0.$$

If $\{X_n(0)\}$ has limiting distribution $v \in \mathcal{P}(E)$, then there is a Markov process X corresponding to $\{T(t)\}$ with initial distribution v and sample paths in $D_E[0, \infty)$, and $X_n \Rightarrow X$.

Proof. For each $n \geq 1$, let A_n be the generator of $\{T_n(t)\}$. By Theorem 6.1 of Chapter 1, (2.23) implies that for each $f \in \mathcal{D}(A)$, there exist $f_n \in \mathcal{D}(A_n)$ such that $f_n \rightarrow f$ and $A_n f_n \rightarrow Af$. Since $f_n(X_n(t)) - \int_0^t A_n f_n(X_n(s)) ds$ is an $\{\mathcal{F}_s^{X_n}\}$ -martingale for each $n \geq 1$, and since $\mathcal{D}(A)$ is dense in $C(E)$, Chapter 3's Corollary 9.3 and Theorem 9.4 imply that $\{X_n\}$ is relatively compact in $D_E[0, \infty)$.

We next prove the convergence of the finite-dimensional distributions of $\{X_n\}$. For each $n \geq 1$, let $\{T_n^\Delta(t)\}$ and $\{T^\Delta(t)\}$ be the semigroups on $C(E^\Delta)$ defined in terms of $\{T_n(t)\}$ and $\{T(t)\}$ as in Lemma 2.3. Then, for each $f \in C(E^\Delta)$ and $t \geq 0$,

$$(2.24) \quad \lim_{n \rightarrow \infty} E[f(X_n(t))] = \lim_{n \rightarrow \infty} E[T_n^\Delta(t)f(X_n(0))] \\ = \int T^\Delta(t)f(x)v(dx)$$

by the Markov property, the strong convergence of $\{T_n^\Delta(t)\}$, the continuity of $T^\Delta(t)f$, and the convergence in distribution of $\{X_n(0)\}$. Proceeding by induction, let m be a positive integer, and suppose that

$$(2.25) \quad \lim_{n \rightarrow \infty} E[f_1(X_n(t_1)) \cdots f_m(X_n(t_m))]$$

exists for all $f_1, \dots, f_m \in C(E^\Delta)$ and $0 \leq t_1 < \cdots < t_m$. Then

$$(2.26) \quad \lim_{n \rightarrow \infty} E[f_1(X_n(t_1)) \cdots f_m(X_n(t_m))f_{m+1}(X_n(t_{m+1}))] \\ = \lim_{n \rightarrow \infty} E[f_1(X_n(t_1)) \cdots f_m(X_n(t_m))T_n^\Delta(t_{m+1} - t_m)f_{m+1}(X_n(t_m))] \\ = \lim_{n \rightarrow \infty} E[f_1(X_n(t_1)) \cdots f_m(X_n(t_m))T^\Delta(t_{m+1} - t_m)f_{m+1}(X_n(t_m))]$$

exists for all $f_1, \dots, f_{m+1} \in C(E^\Delta)$ and $0 \leq t_1 < \cdots < t_{m+1}$.

It follows that every convergent subsequence of $\{X_n\}$ has the same limit, so there exists a process X with initial distribution v and with sample paths in $D_E[0, \infty)$ such that $X_n \Rightarrow X$. By (2.26), X is a Markov process corresponding to $\{T^\Delta(t)\}$, so by Proposition 2.4, X can be assumed to have sample paths in $D_E[0, \infty)$. Finally, Corollary 9.3 of Chapter 3 implies $X_n \Rightarrow X$ in $D_E[0, \infty)$. \square

2.6 Theorem Let E be locally compact and separable. For $n = 1, 2, \dots$ let $\mu_n(x, \Gamma)$ be a transition function on $E \times \mathcal{B}(E)$ such that T_n , defined by

$$(2.27) \quad T_n f(x) = \int f(y)\mu_n(x, dy),$$

satisfies $T_n : \hat{C}(E) \rightarrow \hat{C}(E)$. Suppose that $\{T(t)\}$ is a Feller semigroup on $\hat{C}(E)$. Let $\varepsilon_n > 0$ satisfy $\lim_{n \rightarrow \infty} \varepsilon_n = 0$ and suppose that for every $f \in \hat{C}(E)$,

$$(2.28) \quad \lim_{n \rightarrow \infty} T_n^{[t/\varepsilon_n]} f = T(t)f, \quad t \geq 0.$$

For each $n \geq 1$, let $\{Y_n(k), k = 0, 1, 2, \dots\}$ be a Markov chain in E with transition function $\mu_n(x, \Gamma)$, and suppose $\{Y_n(0)\}$ has limiting distribution $v \in \mathcal{P}(E)$. Define X_n by $X_n(t) \equiv Y_n([t/\varepsilon_n])$. Then there is a Markov process X corresponding to $\{T(t)\}$ with initial distribution v and sample paths in $D_E[0, \infty)$, and $X_n \Rightarrow X$.

Proof. Following the proof of Theorem 2.5, use Theorem 6.5 of Chapter 1 in place of Theorem 6.1. \square

2.7 Theorem Let E be locally compact and separable, and let $\{T(t)\}$ be a Feller semigroup on $\hat{C}(E)$. Then for each $v \in \mathcal{P}(E)$, there exists a Markov process X corresponding to $\{T(t)\}$ with initial distribution v and sample paths in $D_E[0, \infty)$. Moreover, X is strong Markov with respect to the filtration $\mathcal{G}_t = \mathcal{F}_{t+}^X = \bigcap_{\varepsilon > 0} \mathcal{F}_{t+\varepsilon}^X$.

Proof. Let n be a positive integer, and let

$$(2.29) \quad A_n = A(I - n^{-1}A)^{-1} = n[(I - n^{-1}A)^{-1} - I]$$

be the Yosida approximation of A . Note that since $(I - n^{-1}A)^{-1}$ is a positive contraction on $\hat{C}(E)$, there exists for each $x \in E$ a positive Borel measure $\mu_n(x, \Gamma)$ on E such that

$$(2.30) \quad (I - n^{-1}A)^{-1}f(x) = \int f(y)\mu_n(x, dy)$$

for all $f \in \hat{C}(E)$. It follows that $\mu_n(\cdot, \Gamma)$ is Borel measurable for each $\Gamma \in \mathcal{B}(E)$. For each $(f, g) \in A$, (2.30) implies

$$(2.31) \quad f(x) = \int (f(y) - n^{-1}g(y))\mu_n(x, dy), \quad x \in E.$$

Since the collection of $(f, g) \in B(E) \times B(E)$ satisfying (2.31) is bp-closed, it includes $(1, 0)$ and hence $\mu_n(x, E) = 1$ for each $x \in E$, implying that $\mu_n(x, \Gamma)$ is a transition function on $E \times \mathcal{B}(E)$. Therefore, by the discussion at the beginning of this section, the semigroup $\{T_n(t)\}$ on $\hat{C}(E)$ with generator A_n corresponds to a jump Markov process X_n with initial distribution v and with sample paths in $D_E[0, \infty)$.

Now letting $n \rightarrow \infty$, Proposition 2.7 of Chapter 1 implies that for each $f \in \hat{C}(E)$ and $t \geq 0$, $\lim_{n \rightarrow \infty} T_n(t)f = T(t)f$, so the existence of X follows from Theorem 2.5.

Let τ be a discrete $\{\mathcal{G}_t\}$ -stopping time with $\tau < \infty$ a.s. concentrated on $\{t_1, t_2, \dots\}$. Let $A \in \mathcal{G}_1$, $s > 0$, and $f \in \hat{C}(E)$. Then $A \cap \{\tau = t_i\} \in \mathcal{F}_{t_i+\varepsilon}^X$ for every $\varepsilon > 0$, so

$$(2.32) \quad \begin{aligned} \int_{A \cap \{\tau = t_i\}} f(X(\tau + s)) dP &= \int_{A \cap \{\tau = t_i\}} f(X(t_i + s)) dP \\ &= \int_{A \cap \{\tau = t_i\}} T(s - \varepsilon) f(X(t_i + \varepsilon)) dP \end{aligned}$$

for $0 < \varepsilon \leq s$ and $i = 1, 2, \dots$. Since $\{T(t)\}$ is strongly continuous, $T(s)f$ is continuous on E , and X has right continuous sample paths, we can take $\varepsilon = 0$ in (2.32). This gives

$$(2.33) \quad E[f(X(\tau + s)) | \mathcal{G}_\tau] = T(s)f(X(\tau))$$

for discrete τ .

If τ is an arbitrary $\{\mathcal{G}_t\}$ -stopping time, with $\tau < \infty$ a.s., it is the limit of a decreasing sequence $\{\tau_n\}$ of discrete stopping times (Proposition 1.3 of Chapter 2), so (2.33) follows from the continuity of $T(s)f$ on E and the right continuity of the sample paths of X . (Replace τ by τ_n in (2.33), condition on \mathcal{G}_τ , and then let $n \rightarrow \infty$). \square

2.8 Corollary Let E be locally compact and separable. Let A be a linear operator on $\hat{C}(E)$ satisfying (a)–(c) of Theorem 2.2, and let $\{T(t)\}$ be the strongly continuous, positive, contraction semigroup on $\hat{C}(E)$ generated by \bar{A} . Then there exists for each $x \in E$ a Markov process X_x corresponding to $\{T(t)\}$ with initial distribution δ_x and with sample paths in $D_E[0, \infty)$ if and only if A is conservative.

Proof. The sufficiency follows from Theorem 2.7. As for necessity, let $\{g_n\} \subset \mathcal{R}(I - A)$ satisfy $\text{bp-lim}_{n \rightarrow \infty} g_n = 1$, and define $\{f_n\} \subset \mathcal{D}(A)$ by $f_n = (I - A)^{-1}g_n$. Then

$$(2.34) \quad \lim_{n \rightarrow \infty} f_n(x) = \lim_{n \rightarrow \infty} E \left[\int_0^\infty e^{-t} g_n(X_x(t)) dt \right] = 1$$

for all $x \in E$, so $\text{bp-lim}_{n \rightarrow \infty} f_n = 1$ and $\text{bp-lim}_{n \rightarrow \infty} Af_n = \text{bp-lim}_{n \rightarrow \infty} (f_n - g_n) = 0$. \square

We next give criteria for the continuity of the sample paths of the process obtained in Theorem 2.7. Since we know the process has sample paths in $D_E[0, \infty)$, to show the sample paths are continuous it is enough to show that they have no jumps.

2.9 Proposition Let (E, τ) be locally compact and separable, and let $\{T(t)\}$ be a Feller semigroup on $\hat{C}(E)$. Let $P(t, x, \Gamma)$ be the transition function for $\{T(t)\}$ and suppose for each $x \in E$ and $\varepsilon > 0$,

$$(2.35) \quad \lim_{t \rightarrow 0} t^{-1} P(t, x, B(x, \varepsilon)^c) = 0.$$

Then the process X given by Theorem 2.7 satisfies $P\{X \in C_E[0, \infty)\} = 1$.

2.10 Remark Suppose A is the generator of a Feller semigroup $\{T(t)\}$ on $\hat{C}(E)$ with transition function $P(t, x, \Gamma)$, and that for each $x \in E$ and $\varepsilon > 0$ there exists $f \in \mathcal{D}(A)$ with $f(x) = \|f\|$, $\sup_{y \in B(x, \varepsilon)^c} f(y) \leq M < \|f\|$, and $Af(x) = 0$. Then (2.35) holds. To see this, note that

$$(2.36) \quad (\|f\| - M)P(t, x, B(x, \varepsilon)^c) \leq f(x) - E_x[f(X(t))] \\ = - \int_0^t T(s)Af(x) ds.$$

Divide by t and let $t \rightarrow 0$ to obtain (2.35). \square

Proof. Note that for each $x \in E$ and $t \geq 0$,

$$(2.37) \quad \overline{\lim}_{y \rightarrow x} P(t, y, B(y, \varepsilon)^c) \leq \overline{\lim}_{y \rightarrow x} P\left(t, y, B\left(x, \frac{\varepsilon}{2}\right)^c\right) \leq P\left(t, x, B\left(x, \frac{\varepsilon}{2}\right)^c\right).$$

For each $\delta > 0$ there is a $t(x, \delta) \leq \delta$ such that for $t = t(x, \delta)$ the right side of (2.37) is less than $\delta t(x, \delta)$. Consequently, there is a neighborhood U_x of x such that $y \in U_x$ implies

$$(2.38) \quad P(t(x, \delta), y, B(y, \varepsilon)^c) \leq 2\delta t(x, \delta).$$

Since any compact subset of E can be covered by finitely many such U_x , we can define a Borel measurable function $s(y, \delta) \leq \delta$ such that

$$(2.39) \quad P(s(y, \delta), y, B(y, \varepsilon)^c) \leq 2\delta s(y, \delta),$$

and for each compact $K \subset E$

$$(2.40) \quad \inf_{y \in K} s(y, \delta) > 0.$$

Define $\tau_0 = 0$ and

$$(2.41) \quad \tau_{k+1} = \tau_k + s(X(\tau_k), \delta).$$

Note that $\lim_{k \rightarrow \infty} \tau_k = \infty$ since $\{X(s); s \leq t\}$ has compact closure for each $t \geq 0$.

Let

$$(2.42) \quad N_\delta(n) = \sum_{k=0}^{n-1} \chi_{\{r(X(\tau_{k+1}), X(\tau_k)) \geq \varepsilon\}},$$

and observe that

$$(2.43) \quad M_\delta(n) = N_\delta(n) - \sum_{k=0}^{n-1} P(s(X(\tau_k), \delta), X(\tau_k), B(X(\tau_k), \varepsilon)^c)$$

is a martingale. Let $K \subset E$ be compact, let $T > 0$, and define

$$(2.44) \quad \gamma = \gamma_\delta = \min \{n : N_\delta(n) = 1, \tau_n > T, \text{ or } X(\tau_n) \notin K\}.$$

Then by the optional sampling theorem

$$\begin{aligned} (2.45) \quad E[N_\delta(\gamma_\delta)] &= E\left[\sum_{k=0}^{\gamma-1} P(s(X(\tau_k), \delta), X(\tau_k), B(X(\tau_k), \varepsilon)^c)\right] \\ &\leq E\left[\sum_{k=0}^{\gamma-1} 2\delta s(X(\tau_k), \delta)\right] \\ &\leq E[2\delta\tau_\gamma] \leq 2\delta(T + \delta). \end{aligned}$$

Finally, observe that $\lim_{\delta \rightarrow 0} N_\delta(\gamma_\delta) = 1$ on the set where X has a jump of size larger than ε before T and before leaving K . Consequently, with probability one, no such jump occurs. Since ε , T , and K are arbitrary, we have the desired result. \square

We close this section with two theorems generalizing Theorems 2.5 and 2.6. Much more general results are given in Section 8, but these results can be obtained here using essentially the same argument as in the proof of Theorem 2.5.

2.11 Theorem Let E, E_1, E_2, \dots be metric spaces with E locally compact and separable. For $n = 1, 2, \dots$, let $\eta_n : E_n \rightarrow E$ be measurable, let $\{T_n(t)\}$ be a semigroup on $B(E_n)$ given by a transition function, and suppose Y_n is a Markov process in E_n corresponding to $\{T_n(t)\}$ such that $X_n = \eta_n \circ Y_n$ has sample paths in $D_E[0, \infty)$. Define $\pi_n : B(E) \rightarrow B(E_n)$ by $\pi_n f = f \circ \eta_n$ (cf. Section 6 of Chapter 1). Suppose that $\{T(t)\}$ is a Feller semigroup on $\hat{C}(E)$ and that for each $f \in \hat{C}(E)$ and $t \geq 0$, $T_n(t)\pi_n f \rightarrow T(t)f$ (i.e., $\|T_n(t)\pi_n f - \pi_n T(t)f\| \rightarrow 0$). If $\{X_n(0)\}$ has limiting distribution $v \in \mathcal{P}(E)$, then there is a Markov process X corresponding to $\{T(t)\}$ with initial distribution v and sample paths in $D_E[0, \infty)$, and $X_n \Rightarrow X$.

Proof. Note that

$$\begin{aligned} (2.46) \quad E[f(X_n(t+s)) | \mathcal{F}_t^{Y_n}] &= T_n(s)\pi_n f(Y_n(t)) \\ &\approx \pi_n T(s)f(Y_n(t)) \\ &= T(s)f(X_n(t)). \end{aligned}$$

With this observation the proof is essentially the same as for Theorem 2.5. \square

Finally, we give a similar extension of Theorem 2.6.

2.12 Theorem Let E, E_1, E_2, \dots be metric spaces with E locally compact and separable. For $n = 1, 2, \dots$, let $\eta_n : E_n \rightarrow E$ be measurable, let $\mu_n(x, \Gamma)$ be a transition function on $E_n \times \mathcal{P}(E_n)$, and suppose $\{Y_n(k), k = 0, 1, 2, \dots\}$ is a Markov chain in E_n corresponding to $\mu_n(x, \Gamma)$. Let $\varepsilon_n > 0$ satisfy $\lim_{n \rightarrow \infty} \varepsilon_n = 0$. Define $X_n(t) = \eta_n(Y_n([t/\varepsilon_n]))$.

$$(2.47) \quad T_n f(x) = \int f(y) \mu_n(x, dy), \quad f \in B(E_n),$$

and $\pi_n : B(E) \rightarrow B(E_n)$ by $\pi_n f = f \circ \eta_n$. Suppose that $\{T(t)\}$ is a Feller semi-group on $\hat{C}(E)$ and that for each $f \in \hat{C}(E)$ and $t \geq 0$, $T_n^{[t/\varepsilon_n]} \pi_n f \rightarrow T(t)f$. If $\{X_n(0)\}$ has limiting distribution $v \in \mathcal{P}(E)$, then there is a Markov process X corresponding to $\{T(t)\}$ with initial distribution v and sample paths in $D_E[0, \infty)$, and $X_n \Rightarrow X$.

3. THE MARTINGALE PROBLEM: GENERALITIES AND SAMPLE PATH PROPERTIES

In Proposition 1.7 we observed that, if X is a Markov process with full generator \hat{A} , then

$$(3.1) \quad f(X(t)) - \int_0^t g(X(s)) ds$$

is a martingale for all $(f, g) \in \hat{A}$. In the next several sections we develop the idea of Stroock and Varadhan of using this martingale property as a means of characterizing the Markov process associated with a given generator A . As elsewhere in this chapter, E (or more specifically (E, r)) denotes a metric space. Occasionally we want to allow A to be a multivalued operator (cf. Chapter 1, Section 4), and hence think of A as a subset (not necessarily linear) of $B(E) \times B(E)$. By a *solution of the martingale problem for A* we mean a measurable stochastic process X with values in E defined on some probability space (Ω, \mathcal{F}, P) such that for each $(f, g) \in A$, (3.1) is a martingale with respect to the filtration

$$(3.2) \quad {}^*\mathcal{F}_t^X \equiv \mathcal{F}_t^X \vee \sigma \left(\int_0^s h(X(u)) du : s \leq t, h \in B(E) \right).$$

Note that if X is progressive, in particular if X is right continuous, then ${}^*\mathcal{F}_t^X = \mathcal{F}_t^X$. In general, every event in ${}^*\mathcal{F}_t^X$ differs from an event in \mathcal{F}_t^X by an event of probability zero. See Problem 2 of Chapter 2.

If $\{\mathcal{G}_t\}$ is a filtration with $\mathcal{G}_t \supset {}^*\mathcal{F}_t^X$ for all $t \geq 0$, and (3.1) is a $\{\mathcal{G}_t\}$ -martingale for all $(f, g) \in A$, we say X is a *solution of the martingale*

problem for A with respect to $\{\mathcal{G}_t\}$. When an initial distribution $\mu \in \mathcal{P}(E)$ is specified, we say that a solution X of the martingale problem for A is a *solution of the martingale problem for (A, μ)* if $PX(0)^{-1} = \mu$.

Usually X has sample paths in $D_E[0, \infty)$. It is convenient to call a probability measure $P \in \mathcal{P}(D_E[0, \infty))$ a *solution of the martingale problem for A* (or for (A, μ)) if the coordinate process defined on $(D_E[0, \infty), \mathcal{G}_E, P)$ by

$$(3.3) \quad X(t, \omega) \equiv \omega(t), \quad \omega \in D_E[0, \infty), \quad t \geq 0,$$

is a solution of the martingale problem for A (or for (A, μ)) as defined above.

Note that a measurable process X is a solution of the martingale problem for A if and only if

$$(3.4) \quad \begin{aligned} 0 &= E \left[\left(f(X(t_{n+1})) - f(X(t_n)) - \int_{t_n}^{t_{n+1}} g(X(s)) \, ds \right) \prod_{k=1}^n h_k(X(t_k)) \right] \\ &= E \left[f(X(t_{n+1})) \prod_{k=1}^n h_k(X(t_k)) \right] - E \left[f(X(t_n)) \prod_{k=1}^n h_k(X(t_k)) \right] \\ &\quad - \int_{t_n}^{t_{n+1}} E \left[g(X(s)) \prod_{k=1}^n h_k(X(t_k)) \right] \, ds \end{aligned}$$

whenever $0 \leq t_1 < t_2 < \dots < t_{n+1}$, $(f, g) \in A$, and $h_1, \dots, h_n \in B(E)$ (or equivalently $\mathcal{C}(E)$). Consequently the statement that a (measurable) process is a solution of a martingale problem is a statement about its finite-dimensional distributions. In particular, any measurable modification of a solution of the martingale problem for A is also a solution.

Let A_S denote the linear span of A . Then any solution of the martingale problem for A is a solution for A_S . Note also that, if $A^{(1)} \subset A^{(2)}$, then any solution of the martingale problem for $A^{(2)}$ is also a solution for $A^{(1)}$, but not necessarily conversely. Finally, observe that the set of pairs (f, g) for which (3.1) is a $\{\mathcal{G}_t\}$ -martingale is bp-closed. Consequently, any solution of the martingale problem for A is a solution for the bp-closure of A_S . (See Appendix 3.)

3.1 Proposition Let $A^{(1)}$ and $A^{(2)}$ be subsets of $B(E) \times B(E)$. If the bp-closures of $(A^{(1)})_S$ and $(A^{(2)})_S$ are equal, then X is a solution of the martingale problem for $A^{(1)}$ if and only if it is a solution for $A^{(2)}$.

Proof. This is immediate from the discussion above. \square

The following lemma gives two useful equivalences to (3.1) being a martingale.

3.2 Lemma Let X be a measurable process, $\mathcal{G}_t \supset *F_t^X$, and let $f, g \in B(E)$. Then for fixed $\lambda \in \mathbb{R}$, (3.1) is a $\{\mathcal{G}_t\}$ -martingale if and only if

$$(3.5) \quad e^{-\lambda t} f(X(t)) + \int_0^t e^{-\lambda s} (\lambda f(X(s)) - g(X(s))) \, ds$$

is a $\{\mathcal{G}_t\}$ -martingale. If $\inf_x f(x) > 0$, then (3.1) is a $\{\mathcal{G}_t\}$ -martingale if and only if

$$(3.6) \quad f(X(t)) \exp \left\{ - \int_0^t \frac{g(X(s))}{f(X(s))} ds \right\}$$

is a $\{\mathcal{G}_t\}$ -martingale.

Proof. If (3.1) is a $\{\mathcal{G}_t\}$ -martingale, then by Proposition 3.2 of Chapter 2 (see Problem 22 of the same chapter),

$$\begin{aligned} (3.7) \quad & \left[f(X(t)) - \int_0^t g(X(s)) ds \right] e^{-\lambda t} \\ & + \int_0^t \left[f(X(s)) - \int_0^s g(X(u)) du \right] \lambda e^{-\lambda s} ds \\ & = e^{-\lambda t} f(X(t)) + \int_0^t e^{-\lambda s} \lambda f(X(s)) ds \\ & - \int_0^t g(X(s)) ds e^{-\lambda t} - \int_0^t \int_0^s g(X(u)) du \lambda e^{-\lambda s} ds \\ & = e^{-\lambda t} f(X(t)) + \int_0^t e^{-\lambda s} [\lambda f(X(s)) - g(X(s))] ds \end{aligned}$$

is a $\{\mathcal{G}_t\}$ -martingale. (The last equality follows by Fubini's theorem.) If $\inf_x f(x) > 0$ and (3.1) is a $\{\mathcal{G}_t\}$ -martingale, then

$$\begin{aligned} (3.8) \quad & \left[f(X(t)) - \int_0^t g(X(s)) ds \right] \exp \left\{ - \int_0^t \frac{g(X(s))}{f(X(s))} ds \right\} \\ & + \int_0^t \left[f(X(s)) - \int_0^s g(X(u)) du \right] \frac{g(X(s))}{f(X(s))} \exp \left\{ - \int_0^s \frac{g(X(u))}{f(X(u))} du \right\} ds \\ & = f(X(t)) \exp \left\{ - \int_0^t \frac{g(X(s))}{f(X(s))} ds \right\} \\ & - \int_0^t g(X(s)) ds \exp \left\{ - \int_0^t \frac{g(X(s))}{f(X(s))} ds \right\} \\ & + \int_0^t g(X(s)) \exp \left\{ - \int_0^s \frac{g(X(u))}{f(X(u))} du \right\} ds \\ & - \int_0^t \int_0^s g(X(u)) du \frac{g(X(s))}{f(X(s))} \exp \left\{ - \int_0^s \frac{g(X(u))}{f(X(u))} du \right\} ds \\ & = f(X(t)) \exp \left\{ - \int_0^t \frac{g(X(s))}{f(X(s))} ds \right\} \end{aligned}$$

is a $\{\mathcal{G}_t\}$ -martingale. The converses follow by similar calculations (Problem 14). \square

The above lemma gives the following equivalent formulations of the martingale problem.

3.3 Proposition Let A be a linear subset of $B(E) \times B(E)$ containing $(1, 0)$ and define

$$(3.9) \quad A^+ = \{(f, g) \in A : \inf_x f(x) > 0\}.$$

Let X be a measurable E -valued process and let $\mathcal{G}_t \supset *F_t^X$. Then the following are equivalent:

- (a) X is a solution of the martingale problem for A with respect to $\{\mathcal{G}_t\}$.
- (b) X is a solution of the martingale problem for A^+ with respect to $\{\mathcal{G}_t\}$.
- (c) For each $(f, g) \in A$, (3.5) is a $\{\mathcal{G}_t\}$ -martingale.
- (d) For each $(f, g) \in A^+$, (3.6) is a $\{\mathcal{G}_t\}$ -martingale.

Proof. Since $(A^+)_S = A$, (a) and (b) are equivalent. The other equivalences follow by Lemma 3.2. \square

For right continuous X , the fact that (3.5) is a martingale whenever (3.1) is, is a special case of the following lemma.

3.4 Lemma Let X be a measurable stochastic process on (Ω, \mathcal{F}, P) with values in E . Let $u, v: [0, \infty) \times E \times \Omega \rightarrow \mathbb{R}$ be bounded and $\mathcal{B}[0, \infty) \times \mathcal{B}(E) \times \mathcal{F}$ -measurable, and let $w: [0, \infty) \times [0, \infty) \times E \times \Omega \rightarrow \mathbb{R}$ be bounded and $\mathcal{B}[0, \infty) \times \mathcal{B}[0, \infty) \times \mathcal{B}(E) \times \mathcal{F}$ -measurable. Assume that $v(t, x, \omega)$ is continuous in x for fixed t and ω , that $u(t, X(t))$ is adapted to a filtration $\{\mathcal{G}_t\}$, and that $v(t, X(t))$ and $w(t, t, X(t))$ are $\{\mathcal{G}_t\}$ -progressive. Suppose further that the conditions in either (a) or (b) hold:

- (a) For every $t_2 > t_1 \geq 0$,

$$(3.10) \quad E[u(t_2, X(t_2)) - u(t_1, X(t_1)) | \mathcal{G}_{t_1}] = E\left[\int_{t_1}^{t_2} v(s, X(s)) ds \middle| \mathcal{G}_{t_1}\right]$$

and

$$(3.11) \quad E[u(t_1, X(t_2)) - u(t_1, X(t_1)) | \mathcal{G}_{t_1}] = E\left[\int_{t_1}^{t_2} w(t_1, s, X(s)) ds \middle| \mathcal{G}_{t_1}\right].$$

Moreover, X is right continuous and

$$(3.12) \quad \lim_{\delta \rightarrow 0^+} E[|w(t - \delta, t, X(t)) - w(t, t, X(t))|] = 0, \quad t > 0.$$

(b) For every $t_2 > t_1 \geq 0$,

$$(3.13) \quad E[u(t_2, X(t_2)) - u(t_1, X(t_1)) | \mathcal{G}_{t_1}] = E\left[\int_{t_1}^{t_2} w(t_2, s, X(s)) ds \middle| \mathcal{G}_{t_1}\right]$$

and

$$(3.14) \quad E[u(t_2, X(t_2)) - u(t_1, X(t_1)) | \mathcal{G}_{t_1}] = E\left[\int_{t_1}^{t_2} v(s, X(t_1)) ds \middle| \mathcal{G}_{t_1}\right].$$

Moreover, X is left continuous and

$$(3.15) \quad \lim_{\delta \rightarrow 0^+} E[|w(t + \delta, t, X(t)) - w(t, t, X(t))|] = 0, \quad t \geq 0.$$

Under the above assumptions,

$$(3.16) \quad u(t, X(t)) - \int_0^t \{v(s, X(s)) + w(s, s, X(s))\} ds$$

is a $\{\mathcal{G}_t\}$ -martingale.

Proof. Fix $t_2 > t_1 \geq 0$. For any partition $t_1 = s_0 < s_1 < s_2 < \dots < s_n = t_2$, we have

$$(3.17) \quad \begin{aligned} E[u(t_2, X(t_2)) - u(t_1, X(t_1)) | \mathcal{G}_{t_1}] \\ = E\left[\int_{t_1}^{t_2} \{v(s, X(s'')) + w(s', s, X(s))\} ds \middle| \mathcal{G}_{t_1}\right] \end{aligned}$$

under the assumptions in (a), and

$$(3.18) \quad \begin{aligned} E[u(t_2, X(t_2)) - u(t_1, X(t_1)) | \mathcal{G}_{t_1}] \\ = E\left[\int_{t_1}^{t_2} \{v(s, X(s')) + w(s'', s, X(s))\} ds \middle| \mathcal{G}_{t_1}\right] \end{aligned}$$

under the assumptions in (b), where $s' = s_k$ and $s'' = s_{k+1}$ for $s_k \leq s < s_{k+1}$. Letting $\max |s_{k+1} - s_k| \rightarrow 0$, we obtain

$$(3.19) \quad \begin{aligned} E[u(t_2, X(t_2)) - u(t_1, X(t_1)) | \mathcal{G}_{t_1}] \\ = E\left[\int_{t_1}^{t_2} \{v(s, X(s)) + w(s, s, X(s))\} ds \middle| \mathcal{G}_{t_1}\right]. \quad \square \end{aligned}$$

Clearly, only dissipative operators arise as generators of Markov processes. One consequence of Lemma 3.2 is that we must still restrict our attention to dissipative operators in order to have solutions of the martingale problem.

3.5 Proposition Let A be a linear subset of $B(E) \times B(E)$. If there exists a solution X_x of the martingale problem for (A, δ_x) for each $x \in E$, then A is dissipative.

Proof. Given $(f, g) \in A$ and $\lambda > 0$, (3.5) is a martingale and hence

$$(3.20) \quad f(x) = E \left[\int_0^\infty e^{-\lambda s} (\lambda f(X_x(s)) - g(X_x(s))) ds \right], \quad x \in E.$$

Therefore

$$(3.21) \quad |f(x)| \leq \int_0^\infty e^{-\lambda s} \|\lambda f - g\| ds \leq \lambda^{-1} \|\lambda f - g\|$$

and $\lambda \|f\| \leq \|\lambda f - g\|$. \square

As stated above, we usually are interested in solutions with sample paths in $D_E[0, \infty)$. The following theorem demonstrates that in most cases this is not a restriction.

3.6 Theorem Let E be separable. Let $A \subset \bar{C}(E) \times B(E)$ and suppose that $\mathcal{D}(A)$ is separating and contains a countable subset that separates points. Let X be a solution of the martingale problem for A and assume that for every $\varepsilon > 0$ and $T > 0$, there exists a compact set $K_{\varepsilon, T}$ such that

$$(3.22) \quad P\{X(t) \in K_{\varepsilon, T} \text{ for all } t \in [0, T] \cap \mathbb{Q}\} > 1 - \varepsilon.$$

Then there is a modification of X with sample paths in $D_E[0, \infty)$.

Proof. Let X be defined on (Ω, \mathcal{F}, P) . By assumption, there exists a sequence $\{(f_i, g_i)\} \subset A$ such that $\{f_i\}$ separates points in E . By Proposition 2.9 of Chapter 2, there exists $\Omega' \subset \Omega$ with $P(\Omega') = 1$ such that

$$(3.23) \quad f_i(X(t)) - \int_0^t g_i(X(s)) ds$$

has limits through the rationals from above and below for all $t \geq 0$, all i , and all $\omega \in \Omega'$. By (3.22) there exists $\Omega'' \subset \Omega'$ with $P(\Omega'') = 1$ such that $\{X(t, \omega) : t \in [0, T] \cap \mathbb{Q}\}$ has compact closure for all $T > 0$ and $\omega \in \Omega''$. Suppose $\omega \in \Omega''$. Then for each $t \geq 0$ there exist $s_n \in \mathbb{Q}$ such that $s_n > t$, $\lim_{n \rightarrow \infty} s_n = t$, and $\lim_{n \rightarrow \infty} X(s_n, \omega)$ exists, and hence

$$(3.24) \quad f_i(\lim_{n \rightarrow \infty} X(s_n, \omega)) = \lim_{\substack{s \rightarrow t+ \\ s \in \mathbb{Q}}} f_i(X(s, \omega)),$$

where the limit on the right exists since $\omega \in \Omega'$. Since $\{f_i\}$ separates points we have

$$(3.25) \quad \lim_{\substack{s \rightarrow t^+ \\ s \in Q}} X(s) \equiv Y(t)$$

exists for all $t \geq 0$ and $\omega \in \Omega''$. Similarly

$$(3.26) \quad \lim_{\substack{s \rightarrow t^- \\ s \in Q}} X(s) \equiv Y^-(t)$$

exists for all $t > 0$ and $\omega \in \Omega''$, so Y has sample paths in $D_E[0, \infty)$ by Lemma 2.8 of Chapter 2.

Since X is a solution of the martingale problem, it follows that

$$(3.27) \quad E[f(Y(t)) | \mathcal{F}_t^X] = \lim_{\substack{s \rightarrow t^+ \\ s \in Q}} E[f(X(s)) | \mathcal{F}_t^X] = f(X(t))$$

for every $f \in \mathcal{D}(A)$ and $t \geq 0$. Since $\mathcal{D}(A)$ is separating, $P\{Y(t) = X(t)\} = 1$ for all $t \geq 0$. (See Problem 7 of Chapter 3.) \square

3.7 Corollary Let E be locally compact and separable. Let $A \subset \hat{C}(E) \times B(E)$ and suppose that $\mathcal{D}(A)$ is dense in $\hat{C}(E)$ in the norm topology. Then any solution of the martingale problem for A has a modification with sample paths in $D_{E^\Delta}[0, \infty)$ where E^Δ is the one-point compactification of E .

Proof. Let $A' \subset C(E^\Delta) \times B(E^\Delta)$ be given by

$$(3.28) \quad A' = \{(f, g) : f(\Delta) = g(\Delta) = 0, (f|_E, g|_E) \in A\} \cup \{(1, 0)\}$$

and $A'' = (A')_S$. Then any solution of the martingale problem for A considered as a process with values in E^Δ is a solution of the martingale problem for A'' . Since A'' satisfies the conditions of Theorem 3.6, the corollary follows. \square

In the light of condition (3.22) and in particular Corollary 3.7, it is sometimes useful to first prove the existence of a modification with sample paths in $D_E[0, \infty)$ (where \hat{E} is some compactification of E) and then to prove that the modification actually has sample paths in $D_E[0, \infty)$. With this in mind we prove the following theorem.

3.8 Theorem Let (\hat{E}, r) be a metric space and let $A \subset B(\hat{E}) \times B(\hat{E})$. Let $E \subset \hat{E}$ be open, and suppose that X is a solution of the martingale problem for A with sample paths in $D_E[0, \infty)$. Suppose $(\chi_E, 0)$ is in the bp-closure of $A \cap (\bar{C}(\hat{E}) \times B(\hat{E}))$. If $P\{X(0) \in E\} = 1$, then $P\{X \in D_E[0, \infty)\} = 1$.

Proof. For $m = 1, 2, \dots$, define the $\{\mathcal{F}_{t+}^X\}$ -stopping time

$$(3.29) \quad \tau_m = \inf \left\{ t : \inf_{y \in E-E} r(y, X(t)) < \frac{1}{m} \right\}.$$

Then $\tau_1 \leq \tau_2 \leq \dots$ and $\lim_{m \rightarrow \infty} \tau_m \wedge t \equiv Y(t)$ exists. Note that $Y(t)$ is in $\hat{E} - E$ if and only if $\lim_{m \rightarrow \infty} \tau_m \equiv \tau \leq t$. For $(f, g) \in A \cap (\bar{C}(\hat{E}) \times B(\hat{E}))$,

$$(3.30) \quad f(X(t)) - \int_0^t g(X(s)) \, ds$$

is a right continuous $\{\mathcal{F}_t^X\}$ -martingale, and hence the optional sampling theorem implies that for each $t \geq 0$,

$$(3.31) \quad E[f(X(\tau_m \wedge t))] = E[f(X(0))] + E \left[\int_0^{\tau_m \wedge t} g(X(s)) \, ds \right].$$

Letting $m \rightarrow \infty$, we have

$$(3.32) \quad E[f(Y(t))] = E[f(X(0))] + E \left[\int_0^{t \wedge \tau} g(X(s)) \, ds \right],$$

and this holds for all (f, g) in the bp-closure of $A \cap (\bar{C}(\hat{E}) \times B(\hat{E}))$. Taking $(f, g) = (\chi_E, 0)$, we have

$$(3.33) \quad P\{\tau > t\} = P\{Y(t) \in E\} = 1, \quad t \geq 0.$$

Consequently, with probability 1, X has no limit points in $\hat{E} - E$ on any bounded time interval and therefore has almost all sample paths in $D_E[0, \infty)$. \square

3.9 Proposition Let \hat{E} , A , and X be as above. Let $E \subset \hat{E}$ be open. Suppose there exists $\{(f_n, g_n)\} \subset A \cap (\bar{C}(\hat{E}) \times B(\hat{E}))$ such that

$$(3.34) \quad \text{bp-lim}_{n \rightarrow \infty} f_n = \chi_E,$$

$$(3.35) \quad \inf_n \inf_x g_n(x) > -\infty$$

and $\{g_n\}$ converges pointwise to zero. If $P\{X(0) \in E\} = 1$, then $P\{X \in D_E[0, \infty)\} = 1$.

Proof. Substituting (f_n, g_n) in (3.32) and letting $n \rightarrow \infty$, Fatou's lemma gives

$$(3.36) \quad P\{Y(t) \in E\} \geq P\{X(0) \in E\} = 1. \quad \square$$

3.10 Proposition Let \hat{E} , A , and X be as above. Let E_1, E_2, \dots be open subsets of \hat{E} and let $E = \bigcap_k E_k$. Suppose $(\chi_E, 0)$ is in the bp-closure of $A \cap (\bar{C}(\hat{E}) \times B(\hat{E}))$. If $P\{X(0) \in E\} = 1$, then $P\{X \in D_E[0, \infty)\} = 1$.

Proof. Let τ_m^k be defined as in (3.29) with E replaced by E_k . Then the analogue of (3.32) gives

$$(3.37) \quad P\{\lim_{m \rightarrow \infty} X(\tau_m^k \wedge t) \in E_k\} \geq P\{\lim_{m \rightarrow \infty} X(\tau_m^k \wedge t) \in E\} = 1.$$

Therefore almost all sample paths of X are in $D_{E_k}[0, \infty)$ for every k , and hence in $D_E[0, \infty)$. \square

3.11 Remark In the application of Theorem 3.8 and Propositions 3.9 and 3.10, E might be locally compact and $\hat{E} = E^\Delta$, or $E = \prod_k F_k$, where the F_k are locally compact, $\hat{E} = \prod_k F_k^\Delta$, and $E_k = \prod_{l \leq k} F_l \times \prod_{l > k} F_l^\Delta$. \square

We close this section by showing, under the conditions of Theorem 3.6, that any solution of the martingale problem for A with sample paths in $D_E[0, \infty)$ is *quasi-left continuous*, that is, for every nondecreasing sequence of stopping times τ_n with $\lim_{n \rightarrow \infty} \tau_n = \tau < \infty$ a.s., we have $\lim_{n \rightarrow \infty} X(\tau_n) = X(\tau)$ a.s.

3.12 Theorem Let E be separable. Let $A \subset \bar{C}(E) \times B(E)$ and suppose $\mathcal{D}(A)$ is separating. Let X be a solution of the martingale problem for A with respect to $\{\mathcal{G}_t\}$, having sample paths in $D_E[0, \infty)$. Let $\tau_1 \leq \tau_2 \leq \dots$ be a sequence of $\{\mathcal{G}_t\}$ -stopping times and let $\tau = \lim_{n \rightarrow \infty} \tau_n$. Then

$$(3.38) \quad P\left\{\lim_{n \rightarrow \infty} X(\tau_n) = X(\tau), \tau < \infty\right\} = P\{\tau < \infty\}.$$

In particular, $P\{X(t) = X(t-)\} = 1$ for each $t > 0$.

Proof. Clearly the limit in (3.38) exists. For $(f, g) \in A$ and $t \geq 0$,

$$(3.39) \quad \begin{aligned} \lim_{n \rightarrow \infty} f(X(\tau_n \wedge t)) &= \lim_{n \rightarrow \infty} E\left[f(X(\tau \wedge t)) - \int_{\tau_n \wedge t}^{\tau \wedge t} g(X(s)) ds \mid \mathcal{G}_{\tau_n}\right] \\ &= E[f(X(\tau \wedge t)) \mid \vee \mathcal{G}_{\tau_n}], \end{aligned}$$

and (3.38) follows. (See Problem 7 of Chapter 3.) \square

3.13 Corollary Let (E, r) be separable, and let A and X satisfy the conditions of Theorem 3.12. Let $F \subset E$ be closed and define $\tau = \inf\{t: X(t) \in F \text{ or } X(t-) \in F\}$ and $\sigma = \inf\{t: X(t) \in F\}$. (Note that σ need not be measurable.) Then $\tau = \sigma$ a.s.

Proof. Note that $\{\tau = \sigma\} = \{\tau < \infty, X(\tau) \in F\} \cup \{\tau = \infty\}$. Note that by the right continuity of the martingales, X is a solution of the martingale problem for A with respect to $\{\mathcal{F}_{t+}^X\}$. Let $U_n = \{y: \inf_{x \in F} r(x, y) < 1/n\}$, and define

$\tau_n = \inf \{t: X(t) \in U_n\}$. Then τ_n is an $\{\mathcal{F}_{t+}^X\}$ -stopping time, $\tau_1 \leq \tau_2 \leq \dots$ and $\lim_{n \rightarrow \infty} \tau_n = \tau$. Since $X(\tau_n) \in U_n$, Theorem 3.12 implies

$$(3.40) \quad P\left\{\tau < \infty, X(\tau) = \lim_{n \rightarrow \infty} X(\tau_n) \in F\right\} = P\{\tau < \infty\}. \quad \square$$

4. THE MARTINGALE PROBLEM: UNIQUENESS, THE MARKOV PROPERTY, AND DUALITY

As was observed above, the statement that a measurable process X is a solution of the martingale problem for (A, μ) is a statement about the finite-dimensional distributions of X . Consequently, we say that *uniqueness* holds for solutions of the martingale problem for (A, μ) if any two solutions have the same finite-dimensional distributions. If there exists a solution of the martingale problem for (A, μ) and uniqueness holds, we say that the martingale problem for (A, μ) is *well-posed*. If this is true for all $\mu \in \mathcal{P}(E)$, then the martingale problem for A is said to be well-posed. (Typically, if the martingale problem for (A, δ_x) is well-posed for each $x \in E$, then the martingale problem for (A, μ) is well-posed for each $\mu \in \mathcal{P}(E)$. See Problems 49 and 50.) We say that the martingale problem for (A, μ) is well-posed in $D_E[0, \infty)$ ($C_E[0, \infty)$) if there is a unique solution $P \in \mathcal{P}(D_E[0, \infty))$ ($P \in \mathcal{P}(C_E[0, \infty))$). Note that a martingale problem may be well-posed in $D_E[0, \infty)$ without being well-posed, that is, uniqueness may hold under the restriction that the solution have sample paths in $D_E[0, \infty)$ but not in general. See Problem 21. However, Theorem 3.6 shows that this difficulty is rare. The following theorem says essentially that a Markov process is the unique solution of the martingale problem for its generator.

4.1 Theorem Let E be separable, and let $A \subset B(E) \times B(E)$ be linear and dissipative. Suppose there exists $A' \subset A$, A' linear, such that $\mathcal{R}(\lambda - A') = \mathcal{D}(A') \equiv L$ for some $\lambda > 0$, and L is separating. Let $\mu \in \mathcal{P}(E)$ and suppose X is a solution of the martingale problem for (A, μ) . Then X is a Markov process corresponding to the semigroup on L generated by the closure of A' , and uniqueness holds for the martingale problem for (A, μ) .

Proof. Without loss of generality we can assume A' is closed (it is single-valued by Lemma 4.2 of Chapter 1) and hence, by Theorem 2.6 of Chapter 1, it generates a strongly continuous contraction semigroup $\{T(t)\}$ on L . In particular, by Corollary 6.8 of Chapter 1,

$$(4.1) \quad T(t)f = \lim_{n \rightarrow \infty} (I - n^{-1}A')^{-[nt]}f, \quad f \in L, t \geq 0.$$

We want to show that

$$(4.2) \quad E[f(X(t+u)) | \mathcal{F}_t^X] = T(u)f(X(t))$$

for all $f \in L$, which implies the Markov property, and the uniqueness follows by Proposition 1.6.

If $(f, g) \in A'$ and $\lambda > 0$, then (3.5) in Lemma 3.2 is a martingale and hence

$$(4.3) \quad f(X(t)) = E\left[\int_0^\infty e^{-\lambda s} (\lambda f(X(t+s)) - g(X(t+s))) ds \middle| \mathcal{F}_t^X \right],$$

which gives

$$(4.4) \quad \begin{aligned} (I - n^{-1}A')^{-1}h(X(t)) &= E\left[n \int_0^\infty e^{-ns} h(X(t+s)) ds \middle| \mathcal{F}_t^X \right] \\ &= E\left[\int_0^\infty e^{-s} h(X(t + n^{-1}s)) ds \middle| \mathcal{F}_t^X \right] \end{aligned}$$

for all $h \in L$. Iterating (4.4) gives

$$\begin{aligned} (4.5) \quad (I - n^{-1}A')^{-k}h(X(t)) &= E\left[\int_0^\infty \cdots \int_0^\infty \exp\{-(s_1 + s_2 + \cdots + s_k)\} \right. \\ &\quad \times h(X(t + n^{-1}(s_1 + s_2 + \cdots + s_k))) ds_1 \cdots ds_k \middle| \mathcal{F}_t^X \Big] \\ &= E\left[\int_0^\infty \Gamma(k)^{-1} s^{k-1} e^{-s} h(X(t + n^{-1}s)) ds \middle| \mathcal{F}_t^X \right]. \end{aligned}$$

Suppose $h \in \mathcal{D}(A')$. Then

$$\begin{aligned} (4.6) \quad (I - n^{-1}A')^{-[nu]}h(X(t)) &= E[h(X(t+u)) | \mathcal{F}_t^X] \\ &= E\left[h(X(t+u)) \right. \\ &\quad \left. + E\left[\int_0^\infty \Gamma([nu])^{-1} s^{[nu]-1} e^{-s} \int_u^{s/n} A'h(X(t+v)) dv ds \middle| \mathcal{F}_t^X \right] \right]. \end{aligned}$$

The second term on the right is bounded by

$$\begin{aligned} (4.7) \quad \|A'h\| \int_0^\infty \left| \frac{s}{n} - u \right| \Gamma([nu])^{-1} s^{[nu]-1} e^{-s} ds \\ &= \|A'h\| E\left[\left| n^{-1} \sum_{k=1}^{[nu]} \Delta_k - u \right| \right] \end{aligned}$$

where the Δ_k are independent and exponentially distributed with mean 1. Consequently (4.7) goes to zero as $n \rightarrow \infty$, and we have by (4.1)

$$(4.8) \quad \begin{aligned} T(u)h(X(t)) &= \lim_{n \rightarrow \infty} (I - n^{-1}A')^{-[nu]}h(X(t)) \\ &= E[h(X(t+u)) | \mathcal{F}_t^X]. \end{aligned}$$

Since $\overline{\mathcal{D}(A')} = L$, (4.2) holds for all $f \in L$. \square

Under the conditions of Theorem 4.1, every solution of the martingale problem for A is Markovian. We now show that uniqueness of the solution of the martingale problem always implies the Markov property.

4.2 Theorem Let E be separable, and let $A \subset B(E) \times B(E)$. Suppose that for each $\mu \in \mathcal{P}(E)$ any two solutions X, Y of the martingale problem for (A, μ) have the same one-dimensional distributions, that is, for each $t > 0$,

$$(4.9) \quad P\{X(t) \in \Gamma\} = P\{Y(t) \in \Gamma\}, \quad \Gamma \in \mathcal{B}(E).$$

Then the following hold.

(a) Any solution of the martingale problem for A with respect to a filtration $\{\mathcal{G}_t\}$ is a Markov process with respect to $\{\mathcal{G}_t\}$, and any two solutions of the martingale problem for (A, μ) have the same finite-dimensional distributions (i.e., (4.9) implies uniqueness).

(b) If $A \subset \bar{C}(E) \times B(E)$, X is a solution of the martingale problem for A with respect to a filtration $\{\mathcal{G}_t\}$, and X has sample paths in $D_E[0, \infty)$, then for each a.s. finite $\{\mathcal{G}_t\}$ -stopping time τ ,

$$(4.10) \quad E[f(X(\tau+t)) | \mathcal{G}_\tau] = E[f(X(\tau+t)) | X(\tau)]$$

for all $f \in B(E)$ and $t \geq 0$.

(c) If, in addition to the conditions of part (b), for each $x \in E$ there exists a solution $P_x \in \mathcal{P}(D_E[0, \infty))$ of the martingale problem for (A, δ_x) such that $P_x(B)$ is a Borel measurable function of x for each $B \in \mathcal{S}_E$ (cf. Theorem 4.6), then, defining $T(t)f(x) = \int f(\omega(t))P_x(d\omega)$,

$$(4.11) \quad E[f(X(\tau+t)) | \mathcal{G}_\tau] = T(t)f(X(\tau))$$

for all $f \in B(E)$, $t \geq 0$, and a.s. finite $\{\mathcal{G}_t\}$ -stopping times τ (i.e., X is strong Markov).

Proof. Let X , defined on (Ω, \mathcal{F}, P) , be a solution of the martingale problem for A with respect to a filtration $\{\mathcal{G}_t\}$, fix $r \geq 0$, and let $F \in \mathcal{G}_r$ satisfy $P(F) > 0$. For $B \in \mathcal{F}$ define

$$(4.12) \quad P_r(B) = \frac{E[\chi_F E[X_B | \mathcal{G}_r]]}{P(F)}$$

and

$$(4.13) \quad P_2(B) = \frac{E[\chi_F E[\chi_B | X(r)]]}{P(F)},$$

and set $Y(\cdot) = X(r + \cdot)$. Note that

$$(4.14) \quad P_1\{Y(0) \in \Gamma\} = P_2\{Y(0) \in \Gamma\} = P\{X(r) \in \Gamma | F\}.$$

With (3.4) in mind we set

$$(4.15) \quad \eta(Y) = \left[f(Y(t_{n+1})) - f(Y(t_n)) - \int_{t_n}^{t_{n+1}} g(Y(s)) ds \right] \prod_{k=1}^n h_k(Y(t_k))$$

where $0 \leq t_1 < t_2 < \dots < t_n < t_{n+1}$, $(f, g) \in A$, and $h_k \in B(E)$. Since $E[\eta(X(r + \cdot)) | \mathcal{G}_r] = 0$,

$$(4.16) \quad E_2[\eta(Y)] = \frac{E[\chi_F E[\eta(X(r + \cdot)) | X(r)]]}{P(F)} = 0$$

and similarly for $E_1[\eta(Y)]$. Consequently, Y is a solution of the martingale problem for A on $(\Omega, \mathcal{F}, P_1)$ and $(\Omega, \mathcal{F}, P_2)$. By (4.9), $E_1[f(Y(t))] = E_2[f(Y(t))]$ for each $f \in B(E)$ and $t \geq 0$, and hence

$$(4.17) \quad E[\chi_F E[f(X(r + t)) | \mathcal{G}_r]] = E[\chi_F E[f(X(r + t)) | X(r)]].$$

Since $F \in \mathcal{G}$, is arbitrary, (4.17) implies

$$(4.18) \quad E[f(X(r + t)) | \mathcal{G}_r] = E[f(X(r + t)) | X(r)],$$

which is the Markov property.

Uniqueness is proved in much the same way. Let X and Y be solutions of the martingale problem for (A, μ) defined on (Ω, \mathcal{F}, P) and (Γ, \mathcal{G}, Q) respectively. We want to show

$$(4.19) \quad E^P\left[\prod_{k=1}^m f_k(X(t_k))\right] = E^Q\left[\prod_{k=1}^m f_k(Y(t_k))\right]$$

for all choices of $t_k \in [0, \infty)$ and $f_k \in B(E)$ (cf. Proposition 4.6 of Chapter 3). It is sufficient to consider only $f_k > 0$. For $m = 1$, (4.19) holds by (4.9). Proceeding by induction, assume (4.19) holds for all $m \leq n$, and fix $0 \leq t_1 < t_2 < \dots < t_n$ and $f_1, \dots, f_n \in B(E), f_k > 0$. Define

$$(4.20) \quad \tilde{P}(B) = \frac{E^P[\chi_B \prod_{k=1}^n f_k(X(t_k))]}{E[\prod_{k=1}^n f_k(X(t_k))]}, \quad B \in \mathcal{F},$$

$$(4.21) \quad \tilde{Q}(B) = \frac{E^Q[\chi_B \prod_{k=1}^n f_k(Y(t_k))]}{E[\prod_{k=1}^n f_k(Y(t_k))]}, \quad B \in \mathcal{G},$$

and set $\tilde{X}(t) = X(t_n + t)$ and $\tilde{Y}(t) = Y(t_n + t)$. By the argument used above, \tilde{X} on $(\Omega, \mathcal{F}, \tilde{P})$ and \tilde{Y} on $(\Gamma, \mathcal{G}, \tilde{Q})$ are solutions of the martingale problem for A . Furthermore, by (4.19) with $m = n$,

$$(4.22) \quad \begin{aligned} E^{\tilde{P}}[f(\tilde{X}(0))] &= \frac{E^{\tilde{P}}[f(X(t_n)) \prod_{k=1}^{n-1} f_k(X(t_k))]}{E^{\tilde{P}}[\prod_{k=1}^n f_k(X(t_k))]} \\ &= \frac{E^{\tilde{Q}}[f(Y(t_n)) \prod_{k=1}^{n-1} f_k(Y(t_k))]}{E^{\tilde{Q}}[\prod_{k=1}^n f_k(Y(t_k))]} \\ &= E^{\tilde{Q}}[f(\tilde{Y}(0))], \quad f \in B(E), \end{aligned}$$

so \tilde{X} and \tilde{Y} have the same initial distributions. Consequently, (4.9) applies and

$$(4.23) \quad E^{\tilde{P}}[f(\tilde{X}(t))] = E^{\tilde{Q}}[f(\tilde{Y}(t))], \quad t \geq 0, f \in B(E).$$

As in (4.22), this implies

$$(4.24) \quad E^{\tilde{P}}\left[f(X(t_n + t)) \prod_{k=1}^n f_k(X(t_k))\right] = E^{\tilde{Q}}\left[f(Y(t_n + t)) \prod_{k=1}^n f_k(Y(t_k))\right],$$

and, setting $t_{n+1} = t_n + t$, we have (4.19) for $m = n + 1$.

For part (b), assume that $A \subset \bar{C}(E) \times B(E)$ and that X has sample paths in $D_E[0, \infty)$. Then (3.1) is a right continuous martingale with bounded increments for all $(f, g) \in A$ and the optional sampling theorem (see Problem 11 of Chapter 2) implies

$$(4.25) \quad E[\eta(X(\tau + \cdot)) | \mathcal{G}_\tau] = 0,$$

so (4.10) follows in the same way as (4.18).

Similarly for part (c), if $F \in \mathcal{G}$, and $P(F) > 0$, then

$$(4.26) \quad P_1(B) = \frac{E[\chi_F P_{X(\tau)}(B)]}{P(F)}$$

and

$$(4.27) \quad P_2(B) = \frac{E[\chi_F E[\chi_B(X(\tau + \cdot)) | \mathcal{G}_\tau]]}{P(F)}$$

define solutions of the martingale problem with the same initial distribution and (4.11) follows as before. \square

Since it is possible to have uniqueness among solutions of a martingale problem with sample paths in $D_E[0, \infty)$ without having uniqueness among solutions that are only required to be measurable, it is useful to introduce the terminology $D_E[0, \infty)$ martingale problem and $C_E[0, \infty)$ martingale problem to indicate when we are requiring the designated sample path behavior.

4.3 Corollary Let E be separable, and let $A \subset B(E) \times B(E)$. Suppose that for each $\mu \in \mathcal{P}(E)$, any two solutions X, Y of the martingale problem for (A, μ)

with sample paths in $D_E[0, \infty)$ (respectively, $C_E[0, \infty)$) satisfy (4.9) for each $t \geq 0$. Then for each $\mu \in \mathcal{P}(E)$, any two solutions of the martingale problem for (A, μ) with sample paths in $D_E[0, \infty)$ ($C_E[0, \infty)$) have the same distribution on $D_E[0, \infty)$ ($C_E[0, \infty)$).

Proof. Note that \tilde{X} and \tilde{Y} defined in the proof of Theorem 4.2 have sample paths in $D_E[0, \infty)$ ($C_E[0, \infty)$) if X and Y do. Consequently, the proof that X and Y have the same finite-dimensional distributions is the same as before. Since E is separable, by Proposition 7.1 of Chapter 3, the finite-dimensional distributions of X and Y determine their distributions on $D_E[0, \infty)$ ($C_E[0, \infty)$). \square

4.4 Corollary Let E be separable, and let $A \subset B(E) \times B(E)$ be linear and dissipative. Suppose that for some (hence all) $\lambda > 0$, $\mathcal{R}(\lambda - A) \supset \mathcal{D}(A)$, and that there exists $M \subset B(E)$ such that M is separating and $M \subset \mathcal{R}(\lambda - A)$ for every $\lambda > 0$. Then for each $\mu \in \mathcal{P}(E)$ any two solutions of the martingale problem for (A, μ) with sample paths in $D_E[0, \infty)$ have the same distribution on $D_E[0, \infty)$.

4.5 Remark Note that the significance of this result, in contrast with Theorem 4.1, is that we do not require $\mathcal{D}(A)$ to be separating. See Problem 22 for an example in which $\mathcal{D}(A)$ is not separating. \square

Proof. If X and Y are solutions of the martingale problem for (A, μ) with sample paths in $D_E[0, \infty)$, and if $h \in M$, then by (4.3),

$$(4.28) \quad \begin{aligned} E\left[\int_0^\infty e^{-\lambda t} h(X(t)) dt\right] &= \int (\lambda - A)^{-1} h d\mu \\ &= E\left[\int_0^\infty e^{-\lambda t} h(Y(t)) dt\right] \end{aligned}$$

for every $\lambda > 0$. Since M is separating, the identity

$$(4.29) \quad \int_0^\infty e^{-\lambda t} E[h(X(t))] dt = \int_0^\infty e^{-\lambda t} E[h(Y(t))] dt$$

holds for all $h \in B(E)$ (think of $\int_0^\infty e^{-\lambda t} E[h(X(t))] dt \equiv \int h dv_{\lambda, X}$). By the uniqueness of the Laplace transform, for almost every $t \geq 0$,

$$(4.30) \quad E[h(X(t))] = E[h(Y(t))],$$

and if h is continuous, the right continuity of X and Y imply (4.30) holds for all $t \geq 0$. This in turn implies (4.9) and the uniqueness follows from Corollary 4.3. \square

The following theorem shows that the measurability condition in Theorem 4.2(c) typically holds.

4.6 Theorem Let (E, r) be complete and separable, and let $A \subset \bar{C}(E) \times B(E)$. Suppose there exists a countable subset $A_0 \subset A$ such that A is contained in the bp-closure of A_0 (for example, suppose $A \subset L \times L$ where L is a separable subspace of $\bar{C}(E)$). Suppose that the $D_E[0, \infty)$ martingale problem for A is well-posed. Then, denoting the solution for (A, δ_x) by P_x , $P_x(B)$ is Borel measurable in x for each $B \in \mathcal{G}_E$.

Proof. By Theorems 5.6 and 1.7 of Chapter 3, $(\mathcal{P}(D_E[0, \infty)), \rho)$, where ρ is the Prohorov metric, is complete and separable. By the separability of E and Proposition 4.2 of Chapter 3, there is a countable set $M \subset \bar{C}(E)$ such that M is bp-dense in $B(E)$.

Let H be the collection of functions on $D_E[0, \infty)$ of the form

$$(4.31) \quad \eta = \left(f(X(t_{n+1})) - f(X(t_n)) - \int_{t_n}^{t_{n+1}} g(X(s)) ds \right) \prod_{k=1}^n h_k(X(t_k))$$

where X is the coordinate process, $(f, g) \in A_0$, $h_1, \dots, h_n \in M$, $0 \leq t_1 < t_2 < \dots < t_{n+1}$, and $t_k \in \mathbb{Q}$. Note that since A_0 and M are countable, H is countable, and since f and the h_k are continuous, $P \in \mathcal{P}(D_E[0, \infty))$ is a solution of the martingale problem for A if and only if

$$(4.32) \quad \int \eta \, dP = 0, \quad \eta \in H.$$

Let $\mathcal{M}_A \subset \mathcal{P}(D_E[0, \infty))$ be the collection of all such solutions. Then $\mathcal{M}_A = \bigcap_{\eta \in H} \{P: \int \eta \, dP = 0\}$, and \mathcal{M}_A is a Borel set since H is countable and $\{P: \int \eta \, dP = 0\}$ is a Borel set. (Note that if $\eta \in \bar{C}(D_E[0, \infty))$, then $F_\eta(P) \equiv \int \eta \, dP$ is continuous, hence Borel measurable, and the collection of $\eta \in B(D_E[0, \infty))$ for which F_η is Borel measurable is bp-closed.)

Let $G: \mathcal{P}(D_E[0, \infty)) \rightarrow \mathcal{P}(E)$ be given by $G(P) = P X(0)^{-1}$. Note that G is continuous. The fact that the martingale problem for A is well-posed implies that the restriction of G to \mathcal{M}_A is one-to-one and onto. But a one-to-one Borel measurable mapping of a Borel subset of a complete, separable metric space onto a Borel subset of a complete, separable metric space has a Borel measurable inverse (see Appendix 10), that is, letting P_μ denote the solution of the martingale problem for (A, μ) , the mapping of $\mathcal{P}(E)$ into $\mathcal{P}(D_E[0, \infty))$ given by $\mu \mapsto P_\mu$ is Borel measurable and it follows that the mapping of E into $\mathcal{P}(D_E[0, \infty))$ given by $x \mapsto P_x \equiv P_{\delta_x}$ is also Borel measurable. \square

Theorem 4.2 is the basic tool for proving uniqueness for solutions of a martingale problem. The problem, of course, is to verify (4.9). One approach to doing this which, despite its strange, ad hoc appearance, has found widespread applicability involves the notion of *duality*.

Let (E_1, r_1) and (E_2, r_2) be separable metric spaces. Let $A_1 \subset B(E_1) \times B(E_1)$, $A_2 \subset B(E_2) \times B(E_2)$, $f \in M(E_1 \times E_2)$, $\alpha \in M(E_1)$, $\beta \in M(E_2)$, $\mu_1 \in \mathcal{P}(E_1)$, and $\mu_2 \in \mathcal{P}(E_2)$. Then the martingale problems for (A_1, μ_1) and (A_2, μ_2) are dual

with respect to (f, α, β) if for each solution X of the martingale problem for (A_1, μ_1) and each solution Y for (A_2, μ_2) , $\int_0^t |\alpha(X(s))| ds < \infty$ a.s., $\int_0^t |\beta(Y(s))| ds < \infty$ a.s.,

$$(4.33) \quad \int E \left[\left| f(X(t), y) \exp \left\{ \int_0^t \alpha(X(s)) ds \right\} \right| \right] \mu_2(dy) < \infty,$$

$$(4.34) \quad \int E \left[\left| f(x, Y(t)) \exp \left\{ \int_0^t \beta(Y(s)) ds \right\} \right| \right] \mu_1(dx) < \infty,$$

and

$$(4.35) \quad \begin{aligned} & \int E \left[f(X(t), y) \exp \left\{ \int_0^t \alpha(X(s)) ds \right\} \right] \mu_2(dy) \\ &= \int E \left[f(x, Y(t)) \exp \left\{ \int_0^t \beta(Y(s)) ds \right\} \right] \mu_1(dx) \end{aligned}$$

for every $t \geq 0$. Note that if X and Y are defined on the same sample space and are independent, then (4.35) can be written

$$(4.36) \quad \begin{aligned} & E \left[f(X(t), Y(0)) \exp \left\{ \int_0^t \alpha(X(s)) ds \right\} \right] \\ &= E \left[f(X(0), Y(t)) \exp \left\{ \int_0^t \beta(Y(s)) ds \right\} \right]. \end{aligned}$$

4.7 Proposition Let (E_1, r_1) be complete and separable and let E_2 be separable. Let $A_1 \subset B(E_1) \times B(E_1)$, $A_2 \subset B(E_2) \times B(E_2)$, $f \in M(E_1 \times E_2)$, and $\beta \in M(E_2)$. Let $\mathcal{M} \subset \mathcal{P}(E_1)$ contain $PX(t)^{-1}$ for all solutions X of the martingale problem for A_1 with $PX(0)^{-1}$ having compact support, and all $t \geq 0$. Suppose that (A_1, μ) and (A_2, δ_y) are dual with respect to $(f, 0, \beta)$ (i.e., $\alpha = 0$ in (4.35)) for every $\mu \in \mathcal{P}(E_1)$ with compact support and every $y \in E_2$, and that $\{f(\cdot, y): y \in E_2\}$ is separating on \mathcal{M} . If for every $y \in E_2$ there exists a solution of the martingale problem for (A_2, δ_y) , then for each $\mu \in \mathcal{P}(E_1)$ uniqueness holds for the martingale problem for (A_1, μ) .

4.8 Remark (a) The restriction to μ with compact support in the hypotheses is important since we are not assuming boundedness for f and β . Completeness is needed only so that arbitrary $\mu \in \mathcal{P}(E_1)$ can be approximated by μ with compact support.

(b) The proposition transforms the uniqueness problem for A_1 into an existence problem for A_2 . Existence problems, of course, are typically simpler to handle. \square

Proof. Let Y be a solution of the martingale problem for (A_2, δ_y) . If $\mu \in \mathcal{P}(E_1)$ has compact support and X and \tilde{X} are solutions of the martingale problem for (A_1, μ) , then

$$(4.37) \quad E[f(X(t), y)] = \int E\left[f(x, Y_y(t)) \exp\left\{\int_0^t \beta(Y_y(s)) ds\right\}\right] \mu(dx) \\ = E[f(\tilde{X}(t), y)].$$

Since $\{f(\cdot, y) : y \in E_2\}$ is separating on \mathcal{M} , (4.9) holds for X and \tilde{X} .

Now let $\mu \in \mathcal{P}(E_1)$ be arbitrary. If X and \tilde{X} are solutions of the martingale problem for (A_1, μ) and K is compact with $\mu(K) > 0$, then X conditioned on $\{X(0) \in K\}$ and \tilde{X} conditioned on $\{\tilde{X}(0) \in K\}$ are solutions of the martingale problem for $(A_1, \mu(\cdot \cap K)/\mu(K))$. Consequently,

$$(4.38) \quad P\{X(t) \in \Gamma | X(0) \in K\} = P\{\tilde{X}(t) \in \Gamma | \tilde{X}(0) \in K\}, \quad \Gamma \in \mathcal{B}(E_1).$$

Since K is arbitrary and μ is tight, (4.9) follows, and Theorem 4.2 gives the uniqueness. \square

The next step is to give conditions under which (4.35) holds. For the moment proceeding heuristically, suppose X and Y are independent E_1 - and E_2 -valued processes, $g, h \in M(E_1 \times E_2)$,

$$(4.39) \quad f(X(t), y) - \int_0^t g(X(s), y) ds$$

is an $\{\mathcal{F}_t^X\}$ -martingale for every $y \in E_2$, and

$$(4.40) \quad f(x, Y(t)) - \int_0^t h(x, Y(s)) ds$$

is an $\{\mathcal{F}_t^Y\}$ -martingale for every $x \in E_1$. Then

$$(4.41) \quad \frac{d}{ds} E\left[f(X(s), Y(t-s)) \exp\left\{\int_0^s \alpha(X(u)) du + \int_0^{t-s} \beta(Y(u)) du\right\}\right] \\ = E\left[(g(X(s), Y(t-s)) - h(X(s), Y(t-s)) + (\alpha(X(s)) \right. \\ \left. - \beta(Y(t-s)))f(X(s), Y(t-s))) \right. \\ \left. \times \exp\left\{\int_0^s \alpha(X(u)) du + \int_0^{t-s} \beta(Y(u)) du\right\}\right],$$

which is zero if

$$(4.42) \quad g(x, y) + \alpha(x)f(x, y) = h(x, y) + \beta(y)f(x, y).$$

(Compare this calculation with (2.15) of Chapter 1.) Integrating gives (4.36).

4.9 Example To see that there is some possibility of the above working, suppose $E_1 = (-\infty, \infty)$, $E_2 = \{0, 1, 2, \dots\}$, $A_1 f(x) = f''(x) - xf'(x)$, and $A_2 f(y) = y(y-1)(f(y-2) - f(y))$. Of course A_1 corresponds to an Ornstein-Uhlenbeck process and A_2 to a jump process that jumps down by two until it absorbs in 0 or 1. Let $f(x, y) = x^y$. Let X be a solution of the martingale problem for A_1 . Then

$$(4.43) \quad X(t)^y - \int_0^t (y(y-1)X(s)^{y-2} - yX(s)^y) ds$$

is a martingale provided the appropriate expectations exist; they will if the distribution of $X(0)$ has compact support. Let $g(x, y) = y(y-1)x^{y-2} - yx^y$ and $\alpha(x) = 0$. Then $g(x, y) = A_2 f(x, y) + (y^2 - 2y)x^y$, and we have (4.42) if we set $\beta(y) = y^2 - 2y$. Then, assuming the calculation in (4.41) is justified (and it is in this case), we have

$$(4.44) \quad E[X(t)^{Y(0)}] = E\left[X(0)^{Y(0)} \exp\left\{\int_0^t (Y^2(u) - 2Y(u)) du\right\}\right],$$

and the moments of $X(t)$ are determined. In general, of course, this is not enough to determine the distribution of $X(t)$. However, in this case, (4.44) can be used to estimate the growth rate of the moments and the distribution is in fact determined. (See (4.21) of Chapter 3.)

Note that (4.44) suggests another use for duality. If $Y(0) = y$ is odd, then Y absorbs at 1 and

$$(4.45) \quad \lim_{t \rightarrow \infty} E[X(t)^y] = \lim_{t \rightarrow \infty} E\left[X(0)^{Y(0)} \exp\left\{\int_0^t (Y^2(u) - 2Y(u)) du\right\}\right] \\ = 0,$$

since the integrand in the exponent is -1 after absorption. Note that in order to justify this limit one needs to check that

$$E\left[\exp\left\{\int_0^{\tau_1} (Y^2(u) - 2Y(u)) du\right\}\right] < \infty,$$

where $\tau_1 = \inf\{t: Y(t) = 1\}$. Similarly, if $Y(0) = y$ is even, then Y absorbs at 0 and setting $\tau_0 = \inf\{t: Y(t) = 0\}$,

$$(4.46) \quad \lim_{t \rightarrow \infty} E[X(t)^y] = E\left[\exp\left\{\int_0^{\tau_0} (Y^2(u) - 2Y(u)) du\right\}\right].$$

This identity can be used to determine the moments of the limiting distribution (which is Gaussian). See Problem 23. \square

The next lemma gives the first step in justifying the calculation in (4.41).

4.10 Lemma Suppose $f(s, t)$ on $[0, \infty) \times [0, \infty)$ is absolutely continuous in s for each fixed t and absolutely continuous in t for each fixed s , and, setting $(f_1, f_2) \equiv \nabla f$, suppose

$$(4.47) \quad \int_0^T \int_0^t |f_i(s, t)| ds dt < \infty \quad i = 1, 2, \quad T > 0.$$

Then for almost every $t \geq 0$,

$$(4.48) \quad f(t, 0) - f(0, t) = \int_0^t (f_1(s, t-s) - f_2(s, t-s)) ds.$$

Proof.

$$\begin{aligned} (4.49) \quad & \int_0^T \int_0^t (f_1(s, t-s) - f_2(s, t-s)) ds dt \\ &= \int_0^T \int_0^t f_1(t-s, s) ds dt - \int_0^T \int_0^t f_2(s, t-s) ds dt \\ &= \int_0^T \int_s^T f_1(t-s, s) dt ds - \int_0^T \int_s^T f_2(s, t-s) dt ds \\ &= \int_0^T (f(T-s, s) - f(0, s)) ds - \int_0^T (f(s, T-s) - f(s, 0)) ds \\ &= \int_0^T (f(s, 0) - f(0, s)) ds. \end{aligned}$$

Differentiating with respect to T gives the desired result. \square

The following theorem gives conditions under which the calculation in (4.41) is valid.

4.11 Theorem Let X and Y be independent measurable processes in E_1 and E_2 , respectively. Let $f, g, h \in M(E_1 \times E_2)$, $\alpha \in M(E_1)$, and $\beta \in M(E_2)$. Suppose that for each $T > 0$ there exist an integrable random variable Γ_T and a constant C_T such that

$$\begin{aligned} (4.50) \quad & \sup_{r, s, t \leq T} (|\alpha(X(r))| + 1) |f(X(s), Y(t))| \leq \Gamma_T, \\ & \sup_{r, s, t \leq T} (|\beta(Y(r))| + 1) |f(X(s), Y(t))| \leq \Gamma_T, \\ & \sup_{r, s, t \leq T} (|\alpha(X(r))| + 1) |g(X(s), Y(t))| \leq \Gamma_T, \\ & \sup_{r, s, t \leq T} (|\beta(Y(r))| + 1) |h(X(s), Y(t))| \leq \Gamma_T. \end{aligned}$$

and

$$(4.51) \quad \int_0^T |\alpha(X(u))| du + \int_0^T |\beta(Y(u))| du \leq C_T.$$

Suppose that

$$(4.52) \quad f(X(t), y) - \int_0^t g(X(s), y) ds$$

is an $\{\mathcal{F}_t^X\}$ -martingale for each y , and

$$(4.53) \quad f(x, Y(t)) - \int_0^t h(x, Y(s)) ds$$

is an $\{\mathcal{F}_t^Y\}$ -martingale for each x . (The integrals in (4.52) and (4.53) are assumed to exist.) Then for almost every $t \geq 0$,

$$\begin{aligned} (4.54) \quad & E \left[f(X(t), Y(0)) \exp \left\{ \int_0^t \alpha(X(u)) du \right\} \right] \\ & - E \left[f(X(0), Y(t)) \exp \left\{ \int_0^t \beta(Y(u)) du \right\} \right] \\ & = E \left[\int_0^t (g(X(s), Y(t-s)) - h(X(s), Y(t-s))) \right. \\ & \quad \left. + (\alpha(X(s)) - \beta(Y(t-s))) f(X(s), Y(t-s)) \right] \\ & \quad \times \exp \left\{ \int_0^s \alpha(X(u)) du + \int_0^{t-s} \beta(Y(u)) du \right\} ds. \end{aligned}$$

4.12 Remark Note that (4.54) can frequently be extended to more-general α and β by approximating by bounded α and β . \square

Proof. Since (4.52) is a martingale and Y is independent of X , for $h \geq 0$,

$$\begin{aligned} (4.55) \quad & E \left[f(X(s+h), Y(t)) \exp \left\{ \int_0^{s+h} \alpha(X(u)) du \right\} \exp \left\{ \int_0^t \beta(Y(u)) du \right\} \right] \\ & - E \left[f(X(s), Y(t)) \exp \left\{ \int_0^s \alpha(X(u)) du \right\} \exp \left\{ \int_0^t \beta(Y(u)) du \right\} \right] \\ & = E \left[f(X(s+h), Y(t)) \int_s^{s+h} \alpha(X(r)) \right. \\ & \quad \left. \times \exp \left\{ \int_0^r \alpha(X(u)) du \right\} dr \exp \left\{ \int_0^t \beta(Y(u)) du \right\} \right] \\ & + E \left[\int_s^{s+h} g(X(r), Y(t)) dr \exp \left\{ \int_0^s \alpha(X(u)) du \right\} \exp \left\{ \int_0^t \beta(Y(u)) du \right\} \right] \end{aligned}$$

$$\begin{aligned}
&= \int_s^{s+h} E \left[f(X(r), Y(t)) \alpha(X(r)) \exp \left\{ \int_0^r \alpha(X(u)) du \right\} \exp \left\{ \int_0^t \beta(Y(u)) du \right\} \right] dr \\
&\quad + \int_s^{s+h} E \left[\int_r^{s+h} g(X(v), Y(t)) dv \alpha(X(r)) \right. \\
&\quad \times \exp \left\{ \int_0^r \alpha(X(u)) du \right\} \exp \left\{ \int_0^t \beta(Y(u)) du \right\} \left. \right] dr \\
&\quad + \int_s^{s+h} E \left[g(X(r), Y(t)) \exp \left\{ \int_0^r \alpha(X(u)) du \right\} \exp \left\{ \int_0^t \beta(Y(u)) du \right\} \right] dr \\
&\quad + \int_s^{s+h} E \left[g(X(r), Y(t)) \left(\exp \left\{ \int_0^s \alpha(X(u)) du \right\} \right. \right. \\
&\quad \left. \left. - \exp \left\{ \int_0^r \alpha(X(u)) du \right\} \right) \exp \left\{ \int_0^t \beta(Y(u)) du \right\} \right] dr.
\end{aligned}$$

We use (4.50) and (4.51) to ensure the integrability of the random variables above.

Note that for $t, s+h \leq T$, the absolute values of the second and fourth terms are bounded by

$$(4.56) \quad \frac{1}{2} h^2 E[\Gamma_T] e^{Ct}.$$

Set

$$(4.57) \quad F(s, t) = E \left[f(X(s), Y(t)) \exp \left\{ \int_0^s \alpha(X(u)) du \right\} \exp \left\{ \int_0^t \beta(Y(u)) du \right\} \right].$$

For $0 = s_0 < s_1 < \dots < s_m = s$, write

$$(4.58) \quad F(s, t) - F(0, t) = \sum_{i=1}^m (F(s_i, t) - F(s_{i-1}, t)).$$

Letting $\max_i (s_i - s_{i-1}) \rightarrow 0$, (4.55) and the fact that (4.56) is $O(h^2)$ imply

$$(4.59) \quad F(s, t) - F(0, t)$$

$$\begin{aligned}
&= \int_0^s E \left[(f(X(r), Y(t)) \alpha(X(r)) + g(X(r), Y(t))) \right. \\
&\quad \times \exp \left\{ \int_0^r \alpha(X(u)) du \right\} \exp \left\{ \int_0^t \beta(Y(u)) du \right\} \left. \right] dr.
\end{aligned}$$

A similar identity holds for $F(s, t) - F(s, 0)$ and (4.54) follows from Lemma 4.10. \square

4.13 Corollary If, in addition to the conditions of Theorem 4.11, $g(x, y) + \alpha(x)f(x, y) \equiv h(x, y) + \beta(y)f(x, y)$, then for all $t \geq 0$,

$$(4.60) \quad E\left[f(X(t), Y(0)) \exp\left\{\int_0^t \alpha(X(u)) du\right\}\right] \\ = E\left[f(X(0), Y(t)) \exp\left\{\int_0^t \beta(Y(u)) du\right\}\right].$$

Proof. By (4.54), (4.60) holds for almost every t and extends to all t since $F(t, 0)$ and $F(0, t)$ are continuous (see (4.55)). \square

The estimates in (4.50) may be difficult to obtain, and it may be simpler to work first with the processes stopped at exit times from certain sets and then to take limits through expanding sequences of sets to obtain the desired result.

4.14 Corollary Let $\{\mathcal{F}_t\}$ and $\{\mathcal{G}_t\}$ be independent filtrations. Let X be $\{\mathcal{F}_t\}$ -progressive and let τ be an $\{\mathcal{F}_t\}$ -stopping time. Let Y be $\{\mathcal{G}_t\}$ -progressive and let σ be a $\{\mathcal{G}_t\}$ -stopping time. Suppose that (4.50) and (4.51) hold with X and Y replaced by $X(\cdot \wedge \tau)$ and $Y(\cdot \wedge \sigma)$ and that

$$(4.61) \quad f(X(t \wedge \tau), y) - \int_0^{t \wedge \tau} g(X(s), y) ds$$

is an $\{\mathcal{F}_t\}$ -martingale for each y , and

$$(4.62) \quad f(x, Y(t \wedge \sigma)) - \int_0^{t \wedge \sigma} h(x, Y(s)) ds$$

is a $\{\mathcal{G}_t\}$ -martingale for each x . (The integrals in (4.61) and (4.62) are assumed to exist.) Then for almost every $t \geq 0$,

$$(4.63) \quad E\left[f(X(t \wedge \tau), Y(0)) \exp\left\{\int_0^{t \wedge \tau} \alpha(X(u)) du\right\}\right] \\ - E\left[f(X(0), Y(t \wedge \sigma)) \exp\left\{\int_0^{t \wedge \sigma} \beta(Y(u)) du\right\}\right] \\ = \int_0^t E\left[\left(\chi_{\{s \leq \tau\}} g(X(s), Y((t-s) \wedge \sigma)) - \chi_{\{(t-s) \leq \sigma\}} h(X(s \wedge \tau), Y(t-s))\right.\right. \\ \left.\left.+ (\chi_{\{s \leq \tau\}} \alpha(X(s)) - \chi_{\{(t-s) \leq \sigma\}} \beta(Y(t-s)))f(X(s \wedge \tau), Y((t-s) \wedge \sigma))\right)\right. \\ \times \exp\left\{\int_0^{s \wedge \tau} \alpha(X(u)) du + \int_0^{(t-s) \wedge \sigma} \beta(Y(u)) du\right\}\left.\right] ds.$$

Proof. Note that (4.61), for example, can be rewritten

$$(4.64) \quad f(X(t \wedge \tau), y) - \int_0^t \chi_{\{s \leq \tau\}} g(X(s), y) ds.$$

The proof of (4.63) is then essentially the same as the proof of (4.54). \square

4.15 Corollary Under the conditions of Corollary 4.14, if $g(x, y) + \alpha(x)f(x, y) \equiv h(x, y) + \beta(y)f(x, y)$, then for all $t \geq 0$,

$$\begin{aligned} (4.65) \quad & E \left[f(X(t \wedge \tau), Y(0)) \exp \left\{ \int_0^{t \wedge \tau} \alpha(X(u)) du \right\} \right] \\ & - E \left[f(X(0), Y(t \wedge \sigma)) \exp \left\{ \int_0^{t \wedge \sigma} \beta(Y(u)) du \right\} \right] \\ & = \int_0^t E \left[(\chi_{\{s \leq \tau\}} - \chi_{\{(t-s) \leq \sigma\}}) (g(X(s \wedge \tau), Y((t-s) \wedge \sigma)) \right. \\ & \quad \left. + \alpha(X(s \wedge \tau))f(X(s \wedge \tau), Y((t-s) \wedge \sigma))) \right. \\ & \quad \left. \times \exp \left\{ \int_0^{s \wedge \tau} \alpha(X(u)) du + \int_0^{(t-s) \wedge \sigma} \beta(Y(u)) du \right\} \right] ds. \end{aligned}$$

4.16 Remark As $\tau, \sigma \rightarrow \infty$ in (4.65), the integrand on the right goes to zero. The difficulty in practice is to justify the interchange of limits and expectations. \square

5. THE MARTINGALE PROBLEM: EXISTENCE

In this section we are concerned with the existence of solutions of a martingale problem, in particular with the existence of solutions that are Markov or strong Markov. As a part of this discussion, we also examine the structure of the set \mathcal{M}_A of all solutions of the $D_E[0, \infty)$ martingale problem for a given A , considered as a subset of $\mathcal{P}(D_E[0, \infty))$. One of the simplest ways of obtaining solutions is as weak limits of solutions of approximating martingale problems, as indicated by the following lemma.

5.1 Lemma Let $A \subset \bar{C}(E) \times \bar{C}(E)$ and let $A_n \subset B(E) \times B(E)$, $n = 1, 2, \dots$. Suppose that for each $(f, g) \in A$, there exist $(f_n, g_n) \in A_n$ such that

$$(5.1) \quad \lim_{n \rightarrow \infty} \|f_n - f\| = 0, \quad \lim_{n \rightarrow \infty} \|g_n - g\| = 0.$$

If for each n , X_n is a solution of the martingale problem for A_n with sample paths in $D_E[0, \infty)$, and if $X_n \Rightarrow X$, then X is a solution of the martingale problem for A .

5.2 Remark Suppose that (E, r) is complete and separable and that $\overline{\mathcal{D}(A)}$ contains an algebra that separates points and vanishes nowhere (and hence is dense in $\bar{C}(E)$ in the topology of uniform convergence on compact sets). Then $\{X_n\}$ is relatively compact if (5.1) holds for each $(f, g) \in A$ and if for every $\varepsilon, T > 0$, there exists a compact $K_{\varepsilon, T} \subset E$ such that

$$(5.2) \quad \inf_{n} P\{X_n(t) \in K_{\varepsilon, T} \text{ for } 0 \leq t \leq T\} \geq 1 - \varepsilon.$$

See Theorems 9.1 and 9.4 of Chapter 3. \square

Proof. Let $0 \leq t_i \leq t < s$, $t_i, t, s \in D(X) \equiv \{u : P\{X(u) = X(u-) = 1\}$, and $h_i \in \bar{C}(E)$, $i = 1, \dots, k$. Then for $(f, g) \in A$ and $(f_n, g_n) \in A_n$ satisfying (5.1),

$$(5.3) \quad \begin{aligned} & E\left[\left(f(X(s)) - f(X(t)) - \int_t^s g(X(u)) du\right) \prod_{i=1}^k h_i(X(t_i))\right] \\ &= \lim_{n \rightarrow \infty} E\left[\left(f_n(X_n(s)) - f_n(X_n(t)) - \int_t^s g_n(X_n(u)) du\right) \prod_{i=1}^k h_i(X_n(t_i))\right] \\ &= 0. \end{aligned}$$

By Lemma 7.7 of Chapter 3 and the right continuity of X , the equality holds for all $0 \leq t_i \leq t < s$, and hence X is a solution of the martingale problem for A . \square

We now give conditions under which we can approximate A as in Lemma 5.1.

5.3 Lemma Let E be compact and let A be a dissipative linear operator on $C(E)$ such that $\mathcal{D}(A)$ is dense in $C(E)$ and $(1, 0) \in A$. Then there exists a sequence $\{T_n\}$ of positive contraction operators on $B(E)$ given by transition functions such that

$$(5.4) \quad \lim_{n \rightarrow \infty} n(T_n - I)f = Af, \quad f \in \mathcal{D}(A).$$

Proof. Note that

$$(5.5) \quad \Lambda f \equiv n(n - A)^{-1}f(x)$$

defines a bounded linear functional on $\mathcal{R}(n - A)$ for each $n \geq 1$ and $x \in E$. Since $\Lambda 1 = 1$ and $|\Lambda f| \leq \|f\|$, for $f \geq 0$,

$$(5.6) \quad \|f\| - \Lambda f = \Lambda(\|f\| - f) \leq \|f\|,$$

and hence $\Lambda f \geq 0$. Consequently, Λ is a positive linear functional on $\mathcal{R}(n - A)$ with $\|\Lambda\| = 1$. By the Hahn-Banach theorem Λ extends to a positive linear

functional of norm 1 on all of $C(E)$ and hence by the Riesz representation theorem there exists a measure $\mu \in \mathcal{P}(E)$ (not necessarily unique) such that

$$(5.7) \quad Af = \int f d\mu \quad \text{for all } f \in \mathcal{D}(n - A).$$

Consequently, the set

$$(5.8) \quad M_x^n \equiv \left\{ \mu \in \mathcal{P}(E) : n(n - A)^{-1}f(x) = \int f d\mu \quad \text{for all } f \in \mathcal{D}(n - A) \right\}$$

is nonempty for each $n \geq 1$ and $x \in E$.

If $\lim_{k \rightarrow \infty} x_k = x$ and $\mu_k \in M_{x_k}^n$, then by the compactness of $\mathcal{P}(E)$ (Theorem 2.2 of Chapter 3), there exists a subsequence of $\{\mu_k\}$ that converges in the Prohorov metric to a measure $\mu_\infty \in \mathcal{P}(E)$. Since for all $f \in \mathcal{D}(n - A)$,

$$(5.9) \quad \int f d\mu_\infty = \lim_{k \rightarrow \infty} \int f d\mu_k = \lim_{k \rightarrow \infty} n(n - A)^{-1}f(x_k) = n(n - A)^{-1}f(x),$$

$\mu_\infty \in M_x^n$ and the conditions of the measurable selection theorem (Appendix 10) hold for the mapping $x \rightarrow M_x^n$. Consequently, there exist $\mu_x^n \in M_x^n$ such that the mapping $x \rightarrow \mu_x^n$ is a measurable function from E into $\mathcal{P}(E)$. It follows that $\mu_n(x, \Gamma) \equiv \mu_x^n(\Gamma)$ is a transition function, and hence

$$(5.10) \quad T_n f(x) \equiv \int f(y) \mu_n(x, dy)$$

is a positive contraction operator on $B(E)$.

It remains to verify (5.4). For $f \in \mathcal{D}(A)$,

$$(5.11) \quad \begin{aligned} T_n f &= T_n \left(f - \frac{1}{n} Af \right) + \frac{1}{n} T_n Af \\ &= f + \frac{1}{n} T_n Af \end{aligned}$$

and hence

$$(5.12) \quad \lim_{n \rightarrow \infty} T_n f = f.$$

Since $\mathcal{D}(A)$ is dense in $C(E)$ it follows that (5.12) holds for all $f \in C(E)$. Therefore for $f \in \mathcal{D}(A)$,

$$(5.13) \quad \lim_{n \rightarrow \infty} n(T_n - I)f = \lim_{n \rightarrow \infty} T_n Af = Af,$$

since $Af \in C(E)$. □

If E is locally compact and separable, then we can apply Lemma 5.1 to obtain existence of solutions of the martingale problem for a large class of operators.

5.4 Theorem Let E be locally compact and separable, and let A be a linear operator on $\hat{C}(E)$. Suppose $\mathcal{D}(A)$ is dense in $\hat{C}(E)$ and A satisfies the positive maximum principle (i.e., conditions (a) and (b) of Theorem 2.2 are satisfied). Define the linear operator A^Δ on $C(E^\Delta)$ by

$$(5.14) \quad (A^\Delta f)|_E = A((f - f(\Delta))|_E), \quad A^\Delta f(\Delta) = 0,$$

for all $f \in C(E^\Delta)$ such that $(f - f(\Delta))|_E \in \mathcal{D}(A)$. Then for each $v \in \mathcal{P}(E^\Delta)$, there exists a solution of the martingale problem for (A^Δ, v) with sample paths in $D_E[0, \infty)$.

5.5 Remark If A^Δ satisfies the conditions of Theorem 3.8 (with $\hat{E} = E^\Delta$) and $v(E) = 1$, then the above solution of the martingale problem for (A^Δ, v) will have sample paths in $D_E[0, \infty)$. In particular, this will be the case if E is compact and $(1, 0) \in A$. \square

Proof. Note that $\overline{\mathcal{D}(A^\Delta)} = C(E^\Delta)$ and that $f(x_0) = \sup_y f(y) \geq 0$ implies $f(x_0) - f(\Delta) = \sup_y f(y) - f(\Delta) \geq 0$ so $A^\Delta f(x_0) = A(f - f(\Delta))(x_0) \leq 0$. (If $x_0 = \Delta$ then $A^\Delta f(x_0) = 0$ by definition.) Since $A^\Delta 1 = A0 = 0$, A^Δ satisfies the conditions of Lemma 5.3 and there exists a sequence of transition functions $\mu_n(x, \Gamma)$ on $E^\Delta \times \mathcal{B}(E^\Delta)$ such that

$$(5.15) \quad A_n f \equiv n \int (f(y) - f(\cdot)) \mu_n(\cdot, dy), \quad f \in B(E^\Delta),$$

satisfies

$$(5.16) \quad \lim_{n \rightarrow \infty} A_n f = A^\Delta f \text{ for all } f \in \mathcal{D}(A^\Delta).$$

For every $v \in \mathcal{P}(E^\Delta)$ the martingale problem for (A_n, v) has a solution (a Markov jump process) and hence by Lemma 5.1 and Remark 5.2 there exists a solution of the martingale problem for (A, v) . \square

We now consider the question of the existence of a Markov process that solves a given martingale problem.

Throughout the remainder of this section, X denotes the coordinate process on $D_E[0, \infty)$, \mathcal{F} a collection of nonnegative, bounded, Borel measurable functions on $D_E[0, \infty)$ containing all nonnegative constants, and

$$(5.17) \quad \mathcal{F}_t = \mathcal{F}_t^X \vee \sigma(\{\tau \leq s\}: s \leq t, \tau \in \mathcal{F}).$$

Note that all $\tau \in \mathcal{F}$ are $\{\mathcal{F}_t\}$ -stopping times.

Let $\Gamma \subset \mathcal{P}(D_E[0, \infty))$ and for each $v \in \mathcal{P}(E)$ let $\Gamma_v = \{P \in \Gamma: P X(0)^{-1} = v\}$. Assume $\Gamma_v \neq \emptyset$ for each v and for $f \in B(E)$ define

$$(5.18) \quad \gamma(\Gamma_v, f) = \sup_{P \in \Gamma_v} E^P \left[\int_0^\infty e^{-rt} f(X(t)) dt \right].$$

The following lemma gives an important relationship between γ and the martingale problem.

5.6 Lemma Suppose $f, g \in B(E)$ and

$$(5.19) \quad f(X(t)) - \int_0^t g(X(s)) ds$$

is an $\{\mathcal{F}_t\}$ -martingale for all $P \in \Gamma_v$. Then

$$(5.20) \quad \gamma(\Gamma_v, f - g) = \int f d\nu = E^P \left[\int_0^\infty e^{-t} (f(X(t)) - g(X(t))) dt \right]$$

for all $P \in \Gamma_v$.

Proof. This is immediate from Lemma 3.2. \square

We are interested in the following possible conditions on Γ and \mathcal{F} .

5.7 Conditions

C₁: For $P \in \Gamma$, $\tau \in \mathcal{F}$, $\mu = PX(\tau)^{-1}$, and $P' \in \Gamma_\mu$, there exists $\hat{Q} \in \mathcal{P}(D_E[0, \infty) \times [0, \infty))$ with marginal $Q \in \Gamma$ such that

$$(5.21) \quad \begin{aligned} & E^Q[\chi_B(X(\cdot \wedge \eta), \eta)\chi_C(X(\eta + \cdot))] \\ &= \int E^P[\chi_B(X(\cdot \wedge \tau), \tau) | X(\tau) = x] E^{P'}[\chi_C(X(\cdot)) | X(0) = x] \mu(dx) \end{aligned}$$

for all $B \in \mathcal{S}_E \times \mathcal{B}[0, \infty)$ and $C \in \mathcal{S}_E$, where (X, η) denotes the coordinate random variable on $D_E[0, \infty) \times [0, \infty)$. (Note there can be at most one such \hat{Q} .)

C₂: For $P \in \Gamma$, $\tau \in \mathcal{F}$, and $H \geq 0$ \mathcal{F}_τ -measurable with $0 < E^P[H] < \infty$, the measure $Q \in \mathcal{P}(D_E[0, \infty))$ defined by

$$(5.22) \quad Q(C) = \frac{E^P[H\chi_C(X(\tau + \cdot))]}{E^P[H]}$$

is in Γ .

C₃: Γ is convex.

C₄: For each $h \in \bar{C}(E)$ such that $h \geq 0$, there is a $u \in B(E)$ such that $\gamma(\Gamma_v, h) = \int u d\nu$ for all $v \in \mathcal{P}(E)$.

C₅: Γ_v is compact for all $v \in \mathcal{P}(E)$.

We also use a stronger version of **C₅** and a condition that is implied by **C₂**.

C'₂: For $v, \mu_1, \mu_2 \in \mathcal{P}(E)$ such that $v = \alpha\mu_1 + (1 - \alpha)\mu_2$ for some $\alpha \in (0, 1)$ and for $P \in \Gamma_v$, there exist $Q_1 \in \Gamma_{\mu_1}$ and $Q_2 \in \Gamma_{\mu_2}$ such that $P = \alpha Q_1 + (1 - \alpha)Q_2$.

C_{5'}: $\bigcup_{v \in V} \Gamma_v$ is compact for all compact $V \subset \mathcal{P}(E)$.

5.8 Lemma Condition C_2 implies C_2' .

Proof. Let $h_1 = d\mu_1/dv$ and $h_2 = d\mu_2/dv$, and note that $\alpha h_1 + (1 - \alpha)h_2 = 1$. Then setting $H_i = h_i(X(0))$, $i = 1, 2$,

$$(5.23) \quad \begin{aligned} Q_i(C) &= \frac{E^P[H_i \chi_C(X)]}{E^P[H_i]} \\ &= E^P[H_i \chi_C(X)] \end{aligned}$$

is in Γ_{μ_i} , and $P = \alpha Q_1 + (1 - \alpha)Q_2$. \square

Condition C_5' is important because of its relationship to C_4 . To see this we make use of the following lemma.

5.9 Lemma Let E be separable. Let $\varphi: \mathcal{P}(E) \rightarrow [0, c]$ for some $c > 0$. Suppose φ satisfies

$$(5.24) \quad \varphi(\alpha\mu_1 + (1 - \alpha)\mu_2) = \alpha\varphi(\mu_1) + (1 - \alpha)\varphi(\mu_2)$$

for $\alpha \in (0, 1)$ and $\mu_1, \mu_2 \in \mathcal{P}(E)$ and that φ is upper semicontinuous in the sense that $v_n, v \in \mathcal{P}(E)$, $v_n \Rightarrow v$ implies

$$(5.25) \quad \overline{\lim}_{n \rightarrow \infty} \varphi(v_n) \leq \varphi(v).$$

Then there exists $u \in \dot{B}(E)$ such that

$$(5.26) \quad \varphi(v) = \int u \, dv, \quad v \in \mathcal{P}(E).$$

Proof. By (5.25), $u(x) \equiv \varphi(\delta_x)$ is upper semicontinuous and hence measurable ($\{x: u(x) < a\}$ is open). Let E_i^n , $i = 1, 2, \dots$, be disjoint with diameter less than $1/n$ and $E = \bigcup_i E_i^n$; let $x_i^n \in E_i^n$ satisfy $u(x_i^n) \geq \sup_{x \in E_i^n} u(x) - 1/n$. Fix v and define $u_n \in B(E)$ by

$$(5.27) \quad u_n(x) = \sum_i u(x_i^n) \chi_{E_i^n}(x)$$

and $v_n \in \mathcal{P}(E)$ by

$$(5.28) \quad v_n = \sum_i v(E_i^n) \delta_{x_i^n}.$$

Then $\text{bp-lim}_{n \rightarrow \infty} u_n = u$ and $v_n \Rightarrow v$. Consequently,

$$(5.29) \quad \int u \, dv = \lim_{n \rightarrow \infty} \int u_n \, dv = \lim_{n \rightarrow \infty} \int u \, dv_n = \lim_{n \rightarrow \infty} \varphi(v_n) \leq \varphi(v).$$

To obtain the inequality in the other direction (and hence (5.26)) let $\underline{\mu}_i^n(B) = v(B \cap E_i^n)/v(E_i^n)$ when $v(E_i^n) > 0$ and $v_n(x) = \sum \varphi(\underline{\mu}_i^n) \chi_{E_i^n}(x)$. Note that $\lim_{n \rightarrow \infty} v_n(x) \leq u(x)$ by (5.25), and hence

$$(5.30) \quad \begin{aligned} \varphi(v) &= \sum_i \varphi(\underline{\mu}_i^n) v(E_i^n) = \int v_n \, dv \\ &= \lim_{n \rightarrow \infty} \int v_n \, dv \leq \int \overline{\lim}_{n \rightarrow \infty} v_n \, dv \leq \int u \, dv. \end{aligned}$$

(Note $v_n \leq c$.) □

5.10 Lemma Let (E, r) be complete and separable. Suppose conditions C_2 and C_3 hold. Then for $h \in B(E)$ with $h \geq 0$,

$$(5.31) \quad \gamma(\Gamma_{\alpha\mu_1 + (1-\alpha)\mu_2}, h) = \alpha\gamma(\Gamma_{\mu_1}, h) + (1-\alpha)\gamma(\Gamma_{\mu_2}, h)$$

for all $\alpha \in (0, 1)$ and $\mu_1, \mu_2 \in \mathcal{P}(E)$. If, in addition, C_3 holds, then C_4 holds.

Proof. Condition C_2 implies the right side of (5.31) is greater than or equal to the left while C_3 implies the reverse inequality. If C_3 holds, then for $v_n, v \in \mathcal{P}(E)$, $v_n \Rightarrow v$, we have $\Gamma_v \cup \bigcup_n \Gamma_{v_n}$ compact. Consequently, every sequence $P_n \in \Gamma_{v_n}$ has a subsequence that converges weakly to some $P \in \Gamma_v$. Since, for $h \in \bar{C}(E)$, $\int_0^\infty e^{-t} h(X(t)) \, dt$ is continuous on $D_E[0, \infty)$, it follows that

$$(5.32) \quad \overline{\lim}_{n \rightarrow \infty} \gamma(\Gamma_{v_n}, h) \leq \gamma(\Gamma_v, h).$$

C_4 now follows by Lemma 5.9. □

Let A_0 be the collection of all pairs $(f, g) \in B(E) \times B(E)$ for which (5.19) is an (\mathcal{F}_t) -martingale for all $P \in \Gamma$. Our goal is to produce, under conditions C_1-C_3 , an extension A of A_0 satisfying the conditions of Theorem 4.1 such that for each $v \in \mathcal{P}(E)$, there exists in Γ , a solution (necessarily unique) of the martingale problem for (A, v) . The solution will then be a Markov process by Theorem 4.2. Of course typically one begins with an operator A_0 and seeks a set of solutions Γ rather than the reverse. Therefore, to motivate our consideration of C_1-C_3 we first prove the following theorem.

5.11 Theorem Let (E, r) be complete and separable.

(a) Let $A \subset B(E) \times B(E)$, let $\Gamma = \mathcal{M}_A$ (recall that \mathcal{M}_A is the collection of all solutions of the $D_E[0, \infty)$ martingale problem for A), and let \mathcal{T} be the collection of nonnegative constants. Suppose $\Gamma_v \neq \emptyset$ for all $v \in \mathcal{P}(E)$. Then C_1-C_3 hold.

(b) Let $A \subset \bar{C}(E) \times \bar{C}(E)$, and let $\Gamma = \mathcal{M}_A$ and $\mathcal{T} = \{\tau: \{\tau < t\} \in \mathcal{F}_t^X \text{ for all } t \geq 0, \tau \text{ bounded}\}$. Suppose $\overline{\mathcal{D}(A)}$ contains an algebra that separates points and vanishes nowhere, and suppose for each compact $K \subset E$, $\varepsilon > 0$,

and $T > 0$ there exists a compact $K' \subset E$ such that

$$(5.33) \quad P\{X(t) \in K' \text{ for all } t < T, X(0) \in K\} \\ \geq (1 - \epsilon)P\{X(0) \in K\} \text{ for all } P \in \Gamma.$$

Then C_1, C_2 , and C_3 hold.

(c) In addition to the assumptions of part (b), suppose the $D_E[0, \infty)$ martingale problem for A is well-posed. Then the solutions are Markov processes corresponding to a semigroup that maps $\bar{C}(E)$ into $\bar{C}(E)$.

5.12 Remark Part (b) is the result of primary interest. Before proving the theorem we give a lemma that may be useful in verifying condition (5.33). Of course if E is compact, (5.33) is immediate, and if E is locally compact with $A \subset \bar{C}(E) \times \bar{C}(E)$, one can replace E by its one-point compactification E^Δ and consider the corresponding martingale problem in $D_{E^\Delta}[0, \infty)$. \square

5.13 Lemma Let (E, r) be complete, and let $A \subset \bar{C}(E) \times \bar{B}(E)$. Suppose for each compact $K \subset E$ and $\eta > 0$ there exists a sequence of compact $K_n \subset E$, $K \subset K_n$, and $(f_n, g_n) \in A$ such that for $F_n \equiv \{z : \inf_{x \in K_n} r(x, z) \leq \eta\}$,

$$(5.34) \quad \beta_{n, \eta} \equiv \inf_{y \in K} f_n(y) - \sup_{y \in E - F_n} f_n(y) > 0,$$

$$(5.35) \quad \lim_{n \rightarrow \infty} \beta_{n, \eta}^{-1} \sup_{y \in F_n} g_n^-(y) = 0,$$

and

$$(5.36) \quad \lim_{n \rightarrow \infty} \beta_{n, \eta}^{-1} (\|f_n\| - \inf_{y \in K} f_n(y)) = 0.$$

Then for each compact $K \subset E$, $\epsilon > 0$, and $T > 0$, there exists a compact $K' \subset E$ such that

$$(5.37) \quad P\{X(t) \in K' \text{ for all } t < T, X(0) \in K\} \geq (1 - \epsilon)P\{X(0) \in K\},$$

for all $P \in \mathcal{M}_A$.

5.14 Example Let $E = \mathbb{R}^d$ and $A = \{(f, Gf) : f \in C_c^\infty(\mathbb{R}^d)\}$ where

$$(5.38) \quad Gf = \frac{1}{2} \sum_{i,j} a_{ij} \partial_i \partial_j f + \sum_i b_i \partial_i f$$

and the a_{ij} and b_i are measurable functions satisfying $|a_{ij}(x)| \leq M(1 + |x|^2)$ and $|b_i(x)| \leq M(1 + |x|)$ for some $M > 0$. For compact $K \subset B(0, k) \equiv \{z \in \mathbb{R}^d : |z| < k\}$ and $\eta > 0$, let $K_n = \overline{B(0, k+n)}$ and let $f_n \in C_c^\infty(\mathbb{R}^d)$ satisfy

$$f_n(x) = 1 + \log(1 + (k + n + \eta)^2) - \log(1 + |x|^2)$$

for $|x| \leq k + n + \eta$ and $0 \leq f_n(x) \leq 1$ for $|x| > k + n + \eta$. The calculations are left to the reader. \square

Proof. Given $T > 0$, a compact $K \subset E$, and $\eta > 0$, let F_n be as hypothesized, and define $\tau_n = 0$ if $X(0) \notin K$ and $\tau_n = \inf\{t: X(t) \notin F_n\}$ otherwise. Then for $P \in \mathcal{M}_A$,

$$(5.39) \quad E^P[f_n(X(0)) - f_n(X(\tau_n \wedge T))] = E^P \left[- \int_0^{\tau_n \wedge T} g_n(X(s)) ds \right],$$

and hence

$$\begin{aligned} (5.40) \quad & \beta_{n,\eta} P\{0 < \tau_n \leq T\} \\ & \leq E^P[(f_n(X(0)) - f_n(X(\tau_n))) \chi_{\{0 < \tau_n \leq T\}}] \\ & = E^P \left[- \int_0^{\tau_n \wedge T} g_n(X(s)) ds \right] \\ & \quad + E^P[(f_n(X(T)) - f_n(X(0))) \chi_{\{\tau_n > T\}}] \\ & \leq TP\{X(0) \in K\} \sup_{y \in F_n} g_n^-(y) \\ & \quad + (\|f_n\| - \inf_{y \in K} f_n(y))P\{X(0) \in K\}, \end{aligned}$$

which gives

$$\begin{aligned} (5.41) \quad & P\{X(t) \in F_n \text{ for all } t \leq T, X(0) \in K\} \\ & = P\{X(0) \in K\} - P\{0 < \tau_n \leq T\} \\ & \geq P\{X(0) \in K\} \left(1 - \beta_{n,\eta}^{-1} \left(T \sup_{y \in F_n} g_n^-(y) + \|f_n\| - \inf_{y \in K} f_n(y) \right) \right). \end{aligned}$$

From (5.41), (5.35), and (5.36), it follows that for each $m > 0$ there exists a compact $\tilde{K}_m \subset E$ such that

$$\begin{aligned} (5.42) \quad & P\{X(t) \in \tilde{K}_m^{1/m} \text{ for all } t \leq T, X(0) \in K\} \\ & \geq P\{X(0) \in K\}(1 - \varepsilon 2^{-m}). \end{aligned}$$

Hence taking K' to be the closure of $\bigcap_m \tilde{K}_m^{1/m}$, we have (5.37). \square

In order to be able to verify C_1 we need the following technical lemmas.

5.15 Lemma Let (E, r) , (S_1, ρ_1) , and (S_2, ρ_2) be complete, separable metric spaces, let $P_1 \in \mathcal{P}(S_1)$ and $P_2 \in \mathcal{P}(S_2)$ and suppose that $X_1: S_1 \rightarrow E$ and $X_2: S_2 \rightarrow E$ are Borel measurable and that $\mu \in \mathcal{P}(E)$ satisfies $\mu = P_1 X_1^{-1} =$

$P_2 X_2^{-1}$. Let $\{B_i^m\} \subset \mathcal{B}(E)$, $m = 1, 2, \dots$, be a sequence of countable partitions of E with $\{B_i^{m+1}\}$ a refinement of $\{B_i^m\}$ and $\lim_{m \rightarrow \infty} \sup_i \text{diameter}(B_i^m) = 0$. Define $P^m \in \mathcal{P}(S_1 \times S_2)$ by

$$(5.43) \quad P^m(C) = \sum_i \frac{E^{P_1 \times P_2}[\chi_C \chi_{B_i^m}(X_1) \chi_{B_i^m}(X_2)]}{\mu(B_i^m)}$$

for $C \in \mathcal{B}(S_1 \times S_2)$. Then $\{P^m\}$ converges weakly to a probability measure $P \in \mathcal{P}(S_1 \times S_2)$ satisfying

$$(5.44) \quad P(A_1 \times A_2) = \int E^{P_1}[\chi_{A_1} | X_1 = x] E^{P_2}[\chi_{A_2} | X_2 = x] \mu(dx)$$

for $A_1 \in \mathcal{B}(S_1)$ and $A_2 \in \mathcal{B}(S_2)$. In particular $P(A_1 \times S_2) = P_1(A_1)$ and $P(S_1 \times A_2) = P_2(A_2)$. More generally, if $Z_k \in B(S_k)$, $k = 1, 2$, then

$$(5.45) \quad E^P[Z_1 Z_2] = \int E^{P_1}[Z_1 | X_1 = x] E^{P_2}[Z_2 | X_2 = x] \mu(dx).$$

Proof. For $k = 1, 2$, let $A_k \in \mathcal{B}(S_k)$. Note that $E^{P_k}[\chi_{A_k} | X_k = x]$ is the unique (μ -a.s.) $\mathcal{B}(E)$ -measurable function satisfying

$$(5.46) \quad \int_B E^{P_k}[\chi_{A_k} | X_k = x] \mu(dx) = E^{P_k}[\chi_{A_k} \chi_B(X_k)]$$

for all $B \in \mathcal{B}(E)$.

By the martingale convergence theorem (Problem 26 of Chapter 2),

$$(5.47) \quad \begin{aligned} \lim_{m \rightarrow \infty} \sum_i \frac{E^{P_k}[\chi_{A_k} \chi_{B_i^m}(X_k)] \chi_{B_i^m}(x)}{\mu(B_i^m)} \\ = E^{P_k}[\chi_{A_k} | X_k = x] \end{aligned}$$

μ -a.s. and in $L^2(\mu)$. Consequently,

$$(5.48) \quad \lim_{m \rightarrow \infty} P^m(A_1 \times A_2) = P(A_1 \times A_2)$$

($P(A_1 \times A_2)$ given by (5.44)), and since at most one $P \in \mathcal{P}(S_1 \times S_2)$ can satisfy (5.44), it suffices to show that $\{P^m\}$ is tight (cf. Lemma 4.3 of Chapter 3). Let $\varepsilon > 0$, and let K_1 and K_2 be compact subsets of S_1 and S_2 such that $P_k(K_k) \geq 1 - \varepsilon^2$. Then, since $P_k(K_k) = \int E[\chi_{K_k} | X_k = x] \mu(dx)$,

$$(5.49) \quad \mu\{x: E[\chi_{K_k} | X_k = x] \leq 1 - \varepsilon\} \leq \varepsilon$$

and

$$(5.50) \quad P(K_1 \times K_2) \geq (1 - \varepsilon)^2(1 - 2\varepsilon).$$

Tightness for $\{P^m\}$ now follows easily from (5.48). \square

5.16 Lemma Let (E, r) be complete and separable, and let $A \subset B(E) \times B(E)$. Suppose for each $v \in \mathcal{P}(E)$ there exists a solution of the martingale problem for (A, v) with sample paths in $D_E[0, \infty)$. Let Z be a process with sample paths in $D_E[0, \infty)$ and let τ be a $[0, \infty]$ -valued random variable. Suppose, for $(f, g) \in A$, that

$$(5.51) \quad f(Z(t \wedge \tau)) - \int_0^{t \wedge \tau} g(Z(s)) \, ds$$

is a martingale with respect to $\mathcal{G}_t = \sigma(Z(s \wedge \tau), s \wedge \tau : s \leq t)$. If τ is discrete or if $\mathcal{D}(A) \subset \bar{C}(E)$, then there exists a solution Y of the martingale problem for A with sample paths in $D_E[0, \infty)$ and a $[0, \infty]$ -valued random variable η such that $(Y(\cdot \wedge \eta), \eta)$ has the same distribution as $(Z(\cdot \wedge \tau), \tau)$.

Proof. Let $P_1 \in \mathcal{P}(D_E[0, \infty) \times [0, \infty])$ denote the distribution of (Z, τ) and $\mu \in \mathcal{P}(E)$ the distribution of $Z(\tau)$ (fix $x_0 \in E$ and set $Z(\tau) = x_0$ on $\{\tau = \infty\}$). Let $P_2 \in \mathcal{P}(D_E[0, \infty))$ be a solution of the martingale problem for (A, μ) . By Lemma 5.15 there exists $Q \in \mathcal{P}(D_E[0, \infty) \times [0, \infty] \times D_E[0, \infty))$ such that, for $B \in \mathcal{S}_E \times \mathcal{B}[0, \infty)$ and $C \in \mathcal{S}_E$,

$$(5.52) \quad \begin{aligned} Q(B \times C) &= \int E^{P_1}[\chi_B(X, \eta) | X(\eta) = x] E^{P_2}[\chi_C(X) | X(0) = x] \mu(dx) \\ &= \int E[\chi_B(Z, \tau) | Z(\tau) = x] E^{P_2}[\chi_C(X) | X(0) = x] \mu(dx) \end{aligned}$$

where (X, η) denotes the coordinate random variable on $D_E[0, \infty) \times [0, \infty)$. Let (X_1, η, X_2) denote the coordinate random variable on $\Omega = D_E[0, \infty) \times [0, \infty] \times D_E[0, \infty)$ and define

$$(5.53) \quad Y(t) = \begin{cases} X_1(t) & \text{for } t < \eta \\ X_2(t - \eta) & \text{for } t \geq \eta. \end{cases}$$

Note that on $(\Omega, \mathcal{B}(\Omega), Q)$, $Y(\cdot \wedge \eta) = X_1(\cdot \wedge \eta)$ has the same distribution as $Z(\cdot \wedge \tau)$. It remains to show that Y is a solution of the martingale problem for A .

With reference to (3.4), let $(f, g) \in A$, $h_k \in \bar{C}(E)$, $t_1 < t_2 < t_3 < \dots < t_{n+1}$ and define

$$(5.54) \quad R = \left(f(Y(t_{n+1})) - f(Y(t_n)) - \int_{t_n}^{t_{n+1}} g(Y(s)) \, ds \right) \prod_{k=1}^n h_k(Y(t_k)).$$

We must show $E^Q[R] = 0$. Note that

(5.55)

$$\begin{aligned} R &= \left(f(Y(t_{n+1} \wedge \eta)) - f(Y(t_n \wedge \eta)) - \int_{t_n \wedge \eta}^{t_{n+1} \wedge \eta} g(Y(s)) \, ds \right) \prod_{k=1}^n h_k(Y(t_k)) \\ &\quad + \left(f(Y(t_{n+1} \vee \eta)) - f(Y(t_n \vee \eta)) - \int_{t_n \vee \eta}^{t_{n+1} \vee \eta} g(Y(s)) \, ds \right) \prod_{k=1}^n h_k(Y(t_k)) \\ &\equiv R_1 + R_2. \end{aligned}$$

Since R_1 is zero unless $t_n < \eta$, we have

$$\begin{aligned} (5.56) \quad E^Q[R_1] &= E^Q \left[\left(f(Y(t_{n+1} \wedge \eta)) - f(Y(t_n \wedge \eta)) \right. \right. \\ &\quad \left. \left. - \int_{t_n \wedge \eta}^{t_{n+1} \wedge \eta} g(Y(s)) \, ds \right) \prod_{k=1}^n h_k(Y(t_k \wedge \eta)) \right] \\ &= E^P \left[\left(f(X(t_{n+1} \wedge \eta)) - f(X(t_n \wedge \eta)) \right. \right. \\ &\quad \left. \left. - \int_{t_n \wedge \eta}^{t_{n+1} \wedge \eta} g(X(s)) \, ds \right) \prod_{k=1}^n h_k(X(t_k \wedge \eta)) \right] \\ &= E \left[\left(f(Z(t_{n+1} \wedge \tau)) - f(Z(t_n \wedge \tau)) \right. \right. \\ &\quad \left. \left. - \int_{t_n \wedge \tau}^{t_{n+1} \wedge \tau} g(Z(s)) \, ds \right) \prod_{k=1}^n h_k(Z(t_k \wedge \tau)) \right] \\ &= 0. \end{aligned}$$

It remains to show that $E^Q[R_2] = 0$. Suppose first that $\mathcal{D}(A) \subset \bar{\mathcal{C}}(E)$. Define

$$(5.57) \quad \eta_m = \begin{cases} \frac{[m\eta]}{m} & \text{for } \eta < \infty \\ \infty & \text{for } \eta = \infty \end{cases}$$

and

$$\begin{aligned} (5.58) \quad R_2^m &= \left(f(X_2(t_{n+1} \vee \eta_m - \eta_m)) \right. \\ &\quad \left. - f(X_2(t_n \vee \eta_m - \eta_m)) - \int_{t_n \vee \eta_m}^{t_{n+1} \vee \eta_m} g(X_2(s - \eta_m)) \, ds \right) \\ &\quad \times \prod_{t_k \geq \eta_m} h_k(X_2(t_k - \eta_m)) \prod_{t_k < \eta_m} h_k(X_1(t_k)). \end{aligned}$$

By the right continuity of X_2 and the continuity of f , as $m \rightarrow \infty$ R_2^m converges a.s. to R_2 . Noting that $R_2^m = 0$ unless $\eta_m < t_{n+1}$, we have

$$\begin{aligned}
 (5.59) \quad E^Q[R_2^m] &= \sum_{l < m t_{n+1}} E^Q[R_2^m \chi_{(\eta_m = l/m)}] \\
 &= \sum_{l < m t_{n+1}} \int E^{P_2} \left[\left(f \left(X \left(t_{n+1} - \frac{l}{m} \right) \right) - f \left(X \left(t_n \vee \frac{l}{m} - \frac{l}{m} \right) \right) \right. \right. \\
 &\quad \left. \left. - \int_{t_n \vee l/m}^{t_{n+1}} g \left(X \left(s - \frac{l}{m} \right) \right) ds \right) \prod_{k \geq l/m} h_k \left(X \left(t_k - \frac{l}{m} \right) \right) \middle| X(0) = x \right] \\
 &\quad \times E^{P_1} \left[\chi_{(\eta_m = l/m)} \prod_{t_k < l/m} h_k(X(t_k)) \middle| X(\eta) = x \right] \mu(dx) \\
 &= 0,
 \end{aligned}$$

since P_2 is a solution of the martingale problem for (A, μ) . Letting $m \rightarrow \infty$, we see that $E^Q[R_2] = 0$.

If $\mathcal{D}(A) \neq \bar{\mathcal{C}}(E)$ but η is discrete, then $E^Q[R_2] = 0$ by the same argument as in (5.59). \square

Proof of Theorem 5.11 (a). (C₁) Let $P \in \Gamma$, $\tau \in \mathcal{T}$, $\mu = PX(\tau)^{-1}$, and $P' \in \Gamma_\mu$.

In the construction of Q in the proof of Lemma 5.16 take $P_1(B) = P\{(X, \tau) \in B\}$ for $B \in \mathcal{S}_E \times \mathcal{B}[0, \infty]$ and $P_2 = P'$. Then the desired \bar{Q} is the distribution of (Y, η) defined by (5.53) on $(\Omega, \mathcal{B}(\Omega), Q)$. Note that Lemma 5.16 applies under either the conditions of part (a) or of part (b).

(C₂) Let $P \in \Gamma$, $\tau \in \mathcal{T}$, and $H \geq 0$ and \mathcal{F}_τ -measurable with $0 < E^P[H] < \infty$. Define Q by (5.22). Then for $(f, g) \in A$, $h_k \in B(E)$, and $t_1 < t_2 < \dots < t_{n+1}$,

$$\begin{aligned}
 (5.60) \quad E^Q \left[\left(f(X(t_{n+1})) - f(X(t_n)) - \int_{t_n}^{t_{n+1}} g(X(s)) ds \right) \prod_{k=1}^n h_k(X(t_k)) \right] \\
 = \frac{E^P \left[\left(f(X(\tau + t_{n+1})) - f(X(\tau + t_n)) - \int_{\tau + t_n}^{\tau + t_{n+1}} g(X(s)) ds \right) \prod_{k=1}^n h_k(X(\tau + t_k)) H \right]}{E^P[H]} \\
 = 0,
 \end{aligned}$$

since H is \mathcal{F}_τ -measurable and $P \in \mathcal{M}_A$. (Under the assumptions of part (b), the continuity of f allows the application of the optional sampling theorem.)

(C₃) The set of $P \in \mathcal{P}(D_E[0, \infty))$ for which (3.4) holds is clearly convex.

Proof of Theorem 5.11 (b). To complete the proof of part (b) we need only verify C'_5 (which implies C_5), since C_4 will then follow from Lemma 5.10.

(C₅) Let $V \subset \mathcal{P}(E)$ be compact. Then for $0 < \varepsilon < 1$ and $T > 0$, by Theorem 2.2 of Chapter 3, there exist compact $K \subset E$ such that $v(K) \geq 1 - \varepsilon/2$ for all $v \in V$ and (by (5.33)) compact $K_{\varepsilon, T} \subset E$ such that

$$\begin{aligned} (5.61) \quad P\{X(t) \in K_{\varepsilon, T} \text{ for all } t < T\} \\ \geq P\{X(t) \in K_{\varepsilon, T} \text{ for all } t < T, X(0) \in K\} \\ \geq \left(1 - \frac{\varepsilon}{2}\right) P\{X(0) \in K\} \geq \left(1 - \frac{\varepsilon}{2}\right)^2 \geq 1 - \varepsilon \end{aligned}$$

for all $P \in \bigcup_{v \in V} \Gamma_v$. The following lemma completes the proof of C₅ and hence of part (b).

5.17 Lemma Let (E, r) be complete and separable. For $\varepsilon, T > 0$, let $K_{\varepsilon, T} \subset E$ be compact and define $K_{\varepsilon, T}^* = \{x \in D_E[0, \infty): x(t) \in K_{\varepsilon, T} \text{ for all } t < T\}$. If $A \subset \bar{C}(E) \times B(E)$ and $\mathcal{D}(A)$ contains an algebra that separates points and vanishes nowhere, then

$$(5.62) \quad \{P \in \mathcal{M}_A: P(K_{\varepsilon, T}^*) \geq 1 - \varepsilon \text{ for all } \varepsilon, T > 0\}$$

is relatively compact. If, in addition, $A \subset \bar{C}(E) \times \bar{C}(E)$, then (5.62) is compact.

Proof. The relative compactness follows from Theorems 9.1 and 9.4 of Chapter 3. If $A \subset \bar{C}(E) \times \bar{C}(E)$, then compactness follows from Lemma 5.1 with $A_n = A$ for all n . Note that $K_{\varepsilon, T}^*$ is closed, and hence $P_n \Rightarrow P$ implies $P(K_{\varepsilon, T}^*) \geq \lim_{n \rightarrow \infty} P_n(K_{\varepsilon, T}^*)$. \square

Proof of Theorem 5.11 (c). Let P_x denote the solution of the $D_E[0, \infty)$ martingale problem for (A, δ_x) . By C₅ and uniqueness, P_x is weakly continuous as a function of x , and hence by Theorem 4.2 the solutions are Markov and correspond to a semigroup $\{T(t)\}$. By Theorem 3.12, $P_x\{X(t) = X(t-)\} = 1$ for all t and the weak continuity of P_x implies $T(t): \bar{C}(E) \rightarrow \bar{C}(E)$. \square

We now give a partial converse to Lemma 5.6, which demonstrates the importance of condition C₄.

5.18 Lemma Let $\Gamma \subset \mathcal{P}(D_E[0, \infty))$ and \mathcal{F} satisfy C₂. Suppose $u, h \in B(E)$ and

$$(5.63) \quad \gamma(\Gamma_v, h) = \int u(x) dv = E^P \left[\int_0^\infty e^{-t} h(X(t)) dt \right]$$

for all $P \in \Gamma$ and $v \in \mathcal{P}(E)$. Then for each $P \in \Gamma$,

$$(5.64) \quad u(X(t)) - \int_0^t (u(X(s)) - h(X(s))) ds$$

is an $\{\mathcal{F}_s\}$ -martingale.

Proof. Let $P \in \Gamma$, $t \geq 0$, $B \in \mathcal{F}$, with $P(B) > 0$, and $v(C) = P\{X(t) \in C | B\}$ for all $C \in \mathcal{D}(E)$. Then with Q given by (5.22) for $H = \chi_B$,

$$\begin{aligned}
 (5.65) \quad & e^t \int_B \int_t^\infty e^{-u} h(X(u)) du dP \\
 &= \int_B \int_0^\infty e^{-s} h(X(t+s)) ds dP \\
 &= P(B) E^Q \left[\int_0^\infty e^{-s} h(X(s)) ds \right] \\
 &= P(B) \gamma(\Gamma_v, h) \\
 &= P(B) \int u(x) dv = \int_B u(X(t)) dP.
 \end{aligned}$$

Hence

$$\begin{aligned}
 (5.66) \quad & E^P \left[\int_0^\infty e^{-s} h(X(s)) ds \middle| \mathcal{F}_t \right] \\
 &= e^{-t} u(X(t)) + \int_0^t e^{-s} h(X(s)) ds.
 \end{aligned}$$

Since (5.66) is clearly a martingale, the lemma follows from Lemma 3.2. \square

5.19 Theorem Let (E, r) be complete and separable. Let \mathcal{F} be a collection of nonnegative, bounded Borel measurable functions on $D_E[0, \infty)$ containing all nonnegative constants, and let $\{\mathcal{F}_v\}$ be given by (5.17). Let $\Gamma \subset \mathcal{P}(D_E[0, \infty))$ and suppose $\Gamma_v \neq \emptyset$ for all $v \in \mathcal{P}(E)$. Let A_0 be the set of $(f, g) \in B(E) \times B(E)$ such that

$$(5.67) \quad f(X(t)) - \int_0^t g(X(s)) ds$$

is an $\{\mathcal{F}_v\}$ -martingale for all $P \in \Gamma$. Assuming C_1-C_5 , the following hold:

- (a) There exists a linear dissipative operator $A \supset A_0$ such that $\mathcal{D}(I - A) = B(E)$ (hence by Lemma 2.3 of Chapter 1, $\mathcal{D}(\lambda - A) = B(E)$ for all $\lambda > 0$) and $\mathcal{D}(A)$ is bp-dense in $B(E)$.
- (b) Either Γ_v is a singleton for all v or there exists more than one such extension of A_0 .
- (c) For each $v \in \mathcal{P}(E)$ there exists $P_v \in \Gamma_v$, which is the unique (hence Markovian) solution of the martingale problem for (A, v) , and if P_x is the unique solution for (A, δ_x) , then P_x is a measurable function of x and $P_v = \int P_x v(dx)$.

(d) Every solution P of the martingale problem for A satisfies

$$(5.68) \quad P\{X(\tau + \cdot) \in C | \mathcal{F}_\tau\} = P_{X(\tau)}(C)$$

for all $C \in \mathcal{S}_E$ and $\tau \in \mathcal{T}$.

Proof. Let $f_1, f_2, \dots \in \bar{C}(E)$ be nonnegative and suppose the span of $\{f_k\}$ is bp-dense in $B(E)$ (such a sequence always exists by Proposition 4.2 of Chapter 3). Let $\Gamma^{(0)} = \Gamma$ and $\Gamma_v^{(0)} = \Gamma_v$. Define

$$(5.69) \quad \Gamma_v^{(k+1)} = \left\{ P \in \Gamma_v^{(k)} : E^P \left[\int_0^\infty e^{-t} f_{k+1}(X(t)) dt \right] = \gamma(\Gamma_v^{(k)}, f_{k+1}) \right\}$$

for all $v \in \mathcal{P}(E)$ and set $\Gamma^{(k)} = \bigcup \Gamma_v^{(k)}$. Since $\Gamma_v^{(0)}$ is nonempty and compact and $P \mapsto E^P[\int_0^\infty e^{-t} f_1(X(t)) dt]$ is a continuous function from $\mathcal{P}(D_E[0, \infty))$ to \mathbb{R} , it follows that $\Gamma_v^{(1)}$ is nonempty and compact and similarly that $\Gamma_v^{(k)}$ is nonempty and compact for each k . The key to our proof is the following induction lemma.

5.20 Lemma Fix $k \geq 0$, and let $\Gamma^{(k)}$ be as above. If $\Gamma^{(k)}$ and \mathcal{T} satisfy C_1-C_5 , then $\Gamma^{(k+1)}$ and \mathcal{T} satisfy C_1-C_5 . (We denote these conditions by $C_i^{(k)}$ and $C_i^{(k+1)}$ as they apply to $\Gamma^{(k)}$ and $\Gamma^{(k+1)}$.)

Proof. Let $\mu \in \mathcal{P}(E)$ and $P \in \Gamma_\mu^{(k+1)}$. For $B \in \mathcal{B}(E)$ with $0 < \mu(B) < 1$, $C_2^{(k)}$ implies

$$(5.70) \quad \begin{aligned} \gamma(\Gamma_\mu^{(k)}, f_{k+1}) &= E^P \left[\chi_B(X(0)) \int_0^\infty e^{-t} f_{k+1}(X(t)) dt \right] \\ &\quad + E^P \left[\chi_{B^c}(X(0)) \int_0^\infty e^{-t} f_{k+1}(X(t)) dt \right] \\ &\leq \mu(B)\gamma(\Gamma_{\mu_1}^{(k)}, f_{k+1}) + \mu(B^c)\gamma(\Gamma_{\mu_2}^{(k)}, f_{k+1}), \end{aligned}$$

where $\mu_1(C) = \mu(B \cap C)/\mu(B)$ and $\mu_2(C) = \mu(B^c \cap C)/\mu(B^c)$ for all $C \in \mathcal{B}(E)$, and the inequality holds term by term. But $C_2^{(k)}$, $C_3^{(k)}$ and Lemma 5.10 imply equality holds in (5.70), so by $C_4^{(k)}$ there exists a $u_{k+1} \in B(E)$ such that

$$(5.71) \quad E^P \left[\chi_B(X(0)) \int_0^\infty e^{-t} f_{k+1}(X(t)) dt \right] = \int_B u_{k+1}(x) \mu(dx).$$

Hence

$$(5.72) \quad E^P \left[\int_0^\infty e^{-t} f_{k+1}(X(t)) dt \middle| X(0) = x \right] = u_{k+1}(x) \quad \mu\text{-a.s.}$$

We now verify $C_1^{(k+1)}-C_5^{(k+1)}$.

($C_1^{(k+1)}$) For $P \in \Gamma_v^{(k+1)} \subset \Gamma_v^{(k)}$, $\tau \in \mathcal{F}$, $\mu = PX(\tau)^{-1}$, and $P' \in \Gamma_\mu^{(k+1)} \subset \Gamma_\mu^{(k)}$, there exists \hat{Q} with marginal $Q \in \Gamma_v^{(k)}$ such that (5.21) holds. We must show $Q \in \Gamma_v^{(k+1)}$. Let $\mu^*(dx) = (E^P[e^{-\tau} | X(\tau) = x]/E^P[e^{-\tau}])\mu(dx)$. Then, using (5.21), (5.72), and $C_4^{(k)}$,

$$\begin{aligned}
 (5.73) \quad E^Q \left[\int_0^\infty e^{-t} f_{k+1}(X(t)) dt \right] &= E^Q \left[\int_0^\eta e^{-t} f_{k+1}(X(t)) dt \right] \\
 &\quad + E^Q \left[e^{-\eta} \int_0^\infty e^{-t} f_{k+1}(X(\eta+t)) dt \right] \\
 &= E^P \left[\int_0^\tau e^{-t} f_{k+1}(X(t)) dt \right] \\
 &\quad + \int E^P[e^{-\tau} | X(\tau) = x] E^{P'} \left[\int_0^\infty e^{-t} f_{k+1}(X(t)) dt \mid X(0) = x \right] \mu(dx) \\
 &= E^P \left[\int_0^\tau e^{-t} f_{k+1}(X(t)) dt \right] + E^P[e^{-\tau}] \int u_{k+1}(x) \mu^*(dx) \\
 &= E^P \left[\int_0^\tau e^{-t} f_{k+1}(X(t)) dt \right] + E^P[e^{-\tau}] \gamma(\Gamma_\mu^{(k)}, f_{k+1}).
 \end{aligned}$$

By $C_2^{(k)}$ there exists $P'' \in \Gamma_\mu^{(k)}$ such that

$$\begin{aligned}
 (5.74) \quad \gamma(\Gamma_v^{(k)}, f_{k+1}) &= E^P \left[\int_0^\infty e^{-t} f_{k+1}(X(t)) dt \right] \\
 &= E^P \left[\int_0^\tau e^{-t} f_{k+1}(X(t)) dt \right] \\
 &\quad + E^P[e^{-\tau}] E^{P''} \left[\int_0^\infty e^{-t} f_{k+1}(X(t)) dt \right] \\
 &\leq E^P \left[\int_0^\tau e^{-t} f_{k+1}(X(t)) dt \right] \\
 &\quad + E^P[e^{-\tau}] \gamma(\Gamma_\mu^{(k)}, f_{k+1}) \\
 &= E^Q \left[\int_0^\infty e^{-t} f_{k+1}(X(t)) dt \right].
 \end{aligned}$$

Hence equality must hold in (5.74), so $Q \in \Gamma_v^{(k+1)}$.

($C_2^{(k+1)}$) Let $P \in \Gamma^{(k+1)}$ and $\tau \in \mathcal{F}$, and let μ^* be as above. Then for $B \in \mathcal{F}$, with $0 < P(B) < 1$, $C_2^{(k)}$ and the fact that equality holds in (5.74) imply

$$\begin{aligned}
(5.75) \quad & E^P[e^{-t}]\gamma(\Gamma_{\mu_i^k}^{(k)}, f_{k+1}) \\
&= E^P\left[\int_t^\infty e^{-t}f_{k+1}(X(t)) dt\right] \\
&= E^P\left[\chi_B e^{-t} \int_0^\infty e^{-t}f_{k+1}(X(\tau+t)) dt\right] \\
&\quad + E^P\left[\chi_{B^c} e^{-t} \int_0^\infty e^{-t}f_{k+1}(X(\tau+t)) dt\right] \\
&\leq E^P[\chi_B e^{-t}]\gamma(\Gamma_{\mu_i^k}^{(k)}, f_{k+1}) + E^P[\chi_{B^c} e^{-t}]\gamma(\Gamma_{\mu_i^k}^{(k)}, f_{k+1}),
\end{aligned}$$

where $\mu_1^*(C) = E^P[\chi_B e^{-t}\chi_C(X(\tau))]/E^P[\chi_B e^{-t}]$ and $\mu_2^*(C) = E^P[\chi_{B^c} e^{-t}\chi_C(X(\tau))]/E^P[\chi_{B^c} e^{-t}]$ for all $C \in \mathcal{B}(E)$. As before $C_2^{(k)}$, $C_3^{(k)}$, and Lemma 5.10 imply equality in (5.75), and since the inequality is term by term we must have

$$\begin{aligned}
(5.76) \quad & E^P\left[\chi_B e^{-t} \int_0^\infty e^{-t}f_{k+1}(X(\tau+t)) dt\right] \\
&= E^P[\chi_B e^{-t}] \int u_{k+1}(x)\mu_1^*(dx) \\
&= E^P[\chi_B e^{-t}u_{k+1}(X(\tau))],
\end{aligned}$$

which implies

$$(5.77) \quad E^P\left[\int_0^\infty e^{-t}f_{k+1}(X(\tau+t)) \middle| \mathcal{F}_\tau\right] = u_{k+1}(X(\tau)).$$

Now let $H \geq 0$ be \mathcal{F}_τ -measurable with $0 < E^P[H] < \infty$. Then Q , given by (5.22), is in $\Gamma^{(k)}$, and setting $v = QX(0)^{-1}$, (5.77) implies

$$\begin{aligned}
(5.78) \quad & E^Q\left[\int_0^\infty e^{-t}f_{k+1}(X(t)) dt\right] \\
&= \frac{E^P\left[H \int_0^\infty e^{-t}f_{k+1}(X(\tau+t)) dt\right]}{E^P[H]} \\
&= \frac{E^P[Hu_{k+1}(X(\tau))]}{E^P[H]} \\
&= \int u_{k+1}(x)v(dx) = \gamma(\Gamma_v^{(k)}, f_{k+1}),
\end{aligned}$$

and $Q \in \Gamma^{(k+1)}$.

$(C_3^{(k+1)})$ $C_2^{(k)}$ and $C_3^{(k)}$ imply, by Lemma 5.10,

$$(5.79) \quad \gamma(\Gamma_{\alpha\mu_1 + (1-\alpha)\mu_2}^{(k)}, f_{k+1}) = \alpha\gamma(\Gamma_{\mu_1}^{(k)}, f_{k+1}) + (1-\alpha)\gamma(\Gamma_{\mu_2}^{(k)}, f_{k+1})$$

for $\mu_1, \mu_2 \in \mathcal{P}(E)$ and $0 < \alpha < 1$, which in turn implies the convexity of $\Gamma_v^{(k+1)}$.

($C_4^{(k+1)}$) Let u_{k+1} be as above. By $C_4^{(k)}$, for $h \in \bar{C}(E)$ with $h \geq 0$ and $\varepsilon > 0$, there exists $v_\varepsilon \in B(E)$ such that

$$(5.80) \quad \gamma(\Gamma_v^{(k)}, f_{k+1} + \varepsilon h) = \int v_\varepsilon dv, \quad v \in \mathcal{P}(E).$$

We claim that for each $x \in E$

$$(5.81) \quad v \equiv \lim_{\varepsilon \rightarrow 0} \varepsilon^{-1}(v_\varepsilon(x) - u_{k+1}(x)) = \gamma(\Gamma_{\delta_x}^{(k+1)}, h)$$

and

$$(5.82) \quad \int v dv = \gamma(\Gamma_v^{(k+1)}, h).$$

First observe

$$(5.83) \quad \int v_\varepsilon dv \geq \int u_{k+1} dv + \varepsilon \gamma(\Gamma_v^{(k+1)}, h), \quad v \in \mathcal{P}(E),$$

and in particular $v_\varepsilon \geq u_{k+1}$ and

$$(5.84) \quad \lim_{\varepsilon \rightarrow 0} \varepsilon^{-1}(v_\varepsilon(x) - u_{k+1}(x)) \geq \gamma(\Gamma_{\delta_x}^{(k+1)}, h).$$

For each $\varepsilon > 0$, let $P_\varepsilon \in \Gamma_v^{(k)}$ satisfy

$$(5.85) \quad \int v_\varepsilon dv = E^{P_\varepsilon} \left[\int_0^\infty e^{-t} (f_{k+1}(X(t)) + \varepsilon h(X(t))) dt \right].$$

Clearly $\lim_{\varepsilon \rightarrow 0} \int v_\varepsilon dv = \gamma(\Gamma_v^{(k)}, f_{k+1})$, and by the continuity of $\int_0^\infty e^{-t} f_{k+1}(X(t)) dt$, all limit points of $\{P_\varepsilon\}$ as $\varepsilon \rightarrow 0$ are in $\Gamma_v^{(k+1)}$. Consequently,

$$\begin{aligned} (5.86) \quad & \overline{\lim}_{\varepsilon \rightarrow 0} \int \varepsilon^{-1}(v_\varepsilon - u_{k+1}) dv \\ & \leq \overline{\lim}_{\varepsilon \rightarrow 0} E^{P_\varepsilon} \left[\int_0^\infty e^{-t} h(X(t)) dt \right] \\ & \leq \gamma(\Gamma_v^{(k+1)}, h). \end{aligned}$$

In particular,

$$(5.87) \quad \overline{\lim}_{\varepsilon \rightarrow 0} \varepsilon^{-1}(v_\varepsilon(x) - u_{k+1}(x)) \leq \gamma(\Gamma_{\delta_x}^{(k+1)}, h).$$

Therefore (5.81) holds and since $0 \leq \varepsilon^{-1}(v_\varepsilon - u_{k+1}) \leq \|h\|$, (5.82) follows by the dominated convergence theorem.

($C_5^{(k+1)}$) This was verified above. □

Proof of Theorem 5.19 continued. Let $\Gamma^{(\infty)} = \bigcap_k \Gamma^{(k)}$ and for each $k \geq 0$ let u_{k+1} be as above. Then, by Lemma 5.18, for each $k \geq 0$,

$$(5.88) \quad u_{k+1}(X(t)) - \int_0^t (u_{k+1}(X(s)) - f_{k+1}(X(s))) \, ds$$

is an $\{\mathcal{F}_t\}$ -martingale for all $P \in \Gamma^{(k+1)}$, hence for all $P \in \Gamma^{(\infty)}$. Note that $\Gamma_v^{(\infty)} \neq \emptyset$ for all $v \in \mathcal{P}(E)$ since $\Gamma_v^{(k+1)} \subset \Gamma_v^{(k)}$, $\Gamma_v^{(k)} \neq \emptyset$ for all k , and $\Gamma_v^{(k)}$ is compact. Let A be the collection of $(f, g) \in B(E) \times B(E)$ such that (5.67) is an $\{\mathcal{F}_t\}$ -martingale for all $P \in \Gamma^{(\infty)}$. Then $A \supset A_0$ and since $(u_{k+1}, u_{k+1} - f_{k+1}) \in A$ for $k = 1, 2, \dots$, $\mathcal{R}(I - A)$ contains the linear span of $\{f_k\}$ and hence equals $B(E)$. By Proposition 3.5, A is dissipative, hence by Lemma 2.3 of Chapter 1 $\mathcal{R}(\lambda - A) = B(E)$ for all $\lambda > 0$. Lemma 3.2 implies

$$(5.89) \quad (\lambda - A)^{-1}f(x) = E^P \left[\int_0^\infty e^{-\lambda t} f(X(t)) \, dt \right]$$

for every $P \in \Gamma_{s_x}^{(\infty)}$. Therefore if $f \in \tilde{C}(E)$, then

$$(5.90) \quad \text{bp-lim}_{\lambda \rightarrow \infty} \lambda(\lambda - A)^{-1}f = f,$$

and it follows that $\mathcal{D}(A)$ is bp-dense in $B(E)$, which gives part (a).

Since A satisfies the conditions of Theorem 4.1, the martingale problem for (A, v) , $v \in \mathcal{P}(E)$, has at most one solution, and $\Gamma_v^{(\infty)} \neq \emptyset$ implies it has exactly one solution.

If Γ_v is not a singleton for some $v \in \mathcal{P}(E)$, then there exist $P, P' \in \Gamma_v$ and $k > 0$ such that

$$(5.91) \quad E^P \left[\int_0^\infty e^{-t} f_k(X(t)) \, dt \right] \neq E^{P'} \left[\int_0^\infty e^{-t} f_k(X(t)) \, dt \right].$$

(Otherwise $\Gamma_v = \Gamma_v^{(k)}$ for all k and $\Gamma_v = \Gamma_v^{(\infty)}$.) Therefore replacing f_k by $\|f_k\| - f_k$ for all k in the above procedure would produce a different sequence

$$(5.92) \quad \begin{aligned} \hat{\Gamma}_v^{(k+1)} &= \left\{ P \in \hat{\Gamma}_v^{(k)} : E^P \left[\int_0^\infty e^{-t} (\|f_{k+1}\| - f_{k+1}(X(t))) \, dt \right] \right. \\ &\quad \left. = \gamma(\hat{\Gamma}_v^{(k)}, \|f_{k+1}\| - f_{k+1}) \right\}. \end{aligned}$$

Let k_0 be the smallest k for which there exist v_0 and $P, P' \in \Gamma_{v_0}$ such that (5.91) holds. Then $\Gamma_{v_0}^{(k)} \neq \Gamma_{v_0}^{(k)}$, in fact $\Gamma_{v_0}^{(k)} \cap \Gamma_{v_0}^{(k)} = \emptyset$, for $k > k_0$. Consequently $\Gamma_{v_0}^{(\infty)} \neq \hat{\Gamma}_{v_0}^{(\infty)}$ and the extension of A_0 corresponding to $\|f_k\| - f_k$ differs from A .

Let P_x denote the unique probability measure in $\Gamma_{s_x}^{(\infty)}$. The semigroup $\{T(t)\}$ corresponding to A (defined on $\mathcal{D}(A)$) can be represented by

$$(5.93) \quad T(t)f(x) = E^{P_x}[f(X(t))].$$

Since $T(t): \overline{\mathcal{D}(A)} \rightarrow \overline{\mathcal{D}(A)}$ and $\mathcal{D}(A)$ is bp-dense in $B(E)$, $\{T(t)\}$ can be extended to a semigroup on all of $B(E)$ satisfying (5.93). Consequently,

$$(5.94) \quad P(t, x, \Gamma) \equiv E^P[\chi_\Gamma(X(t))]$$

is a transition function, and by Proposition 1.2 $P_x(B)$ is a Borel measurable function of x for all $B \in \mathcal{B}(E)$. For each $v \in \mathcal{P}(E)$,

$$(5.95) \quad P_v \equiv \int P_x v(dx)$$

is a solution of the martingale problem for (A, v) and hence is the unique element of $\Gamma^{(\omega)}$. This completes the proof of part (c).

Since $\Gamma^{(k)}$ satisfies C_2 for all k , $\Gamma^{(\omega)}$ satisfies C_2 . For $P \in \Gamma^{(\omega)}$, $\tau \in \mathcal{T}$, and $B \in \mathcal{F}$, with $P(B) > 0$, uniqueness implies

$$(5.96) \quad \frac{E^P[\chi_B \chi_C(X(\tau + \cdot))]}{P(B)} = \int P_x(C) \mu(dx)$$

where $\mu(D) = E^P[\chi_B \chi_D(X(\tau))] / P(B)$ for all $D \in \mathcal{B}(E)$. Since B is arbitrary in \mathcal{F} , (5.68) follows. \square

6. THE MARTINGALE PROBLEM: LOCALIZATION

Let $A \subset B(E) \times B(E)$, let U be an open subset of E , and let X be a process with initial distribution $v \in \mathcal{P}(E)$ and sample paths in $D_E[0, \infty)$. Define the $\{\mathcal{F}_t^X\}$ -stopping time

$$(6.1) \quad \tau = \inf \{t \geq 0: X(t) \notin U \text{ or } X(t-) \notin U\}.$$

Then X is a solution of the *stopped martingale problem* for (A, v, U) if $X(\cdot) = X(\cdot \wedge \tau)$ a.s. and

$$(6.2) \quad f(X(t)) - \int_0^{t \wedge \tau} g(X(s)) ds$$

is an $\{\mathcal{F}_t^X\}$ -martingale for all $(f, g) \in A$. (Note that the stopped martingale problem requires sample paths in $D_E[0, \infty)$.)

6.1 Theorem Let (E, r) be complete and separable, and let $A \subset \bar{C}(E) \times B(E)$. If the $D_E[0, \infty)$ martingale problem for A is well-posed, then for each $v \in \mathcal{P}(E)$ and open $U \subset E$ there exists a unique solution of the stopped martingale problem for (A, v, U) .

Proof. Let X be the solution of the $D_E[0, \infty)$ martingale problem for (A, v) , define τ by (6.1), and define $\tilde{X}(\cdot) = X(\cdot \wedge \tau)$. Then \tilde{X} is a solution of the

stopped martingale problem for (A, v, U) by the optional sampling theorem (Theorem 2.13 of Chapter 2).

For uniqueness, fix v and U and let X be a solution of the stopped martingale problem for (A, v, U) . By Lemma 5.16 there exists a solution Y of the $D_E[0, \infty)$ martingale problem for (A, v) and a nonnegative random variable η such that $X (= X(\cdot \wedge \tau))$ has the same distribution as $Y(\cdot \wedge \eta)$. Note that in this case the η constructed in the proof of Lemma 5.16 is $\inf\{\tau \geq 0: Y(\tau) \notin U \text{ or } Y(\tau-) \notin U\}$, and since the distribution of Y is uniquely determined, it follows that the distribution of $Y(\cdot \wedge \eta)$ (and hence of X) is uniquely determined. \square

Our primary interest in this section is in possible converses for the above theorem. That is, we are interested in conditions under which existence and, more importantly, uniqueness of solutions of the stopped martingale problem imply existence and uniqueness for the (unstopped) $D_E[0, \infty)$ martingale problem. Recall that uniqueness for the $D_E[0, \infty)$ martingale problem is typically equivalent to the general uniqueness question (cf. Theorem 3.6) but not necessarily (Problem 21).

6.2 Theorem Let E be separable, and let $A \subset \bar{C}(E) \times B(E)$. Suppose that for each $v \in \mathcal{P}(E)$ there exists a solution of the $D_E[0, \infty)$ martingale problem for (A, v) . If there exist open subsets U_k , $k = 1, 2, \dots$, with $E = \bigcup_{k=1}^{\infty} U_k$ such that for each $v \in \mathcal{P}(E)$ and $k = 1, 2, \dots$ the solution of the stopped martingale problem for (A, v, U_k) is unique, then for each $v \in \mathcal{P}(E)$ the solution of the $D_E[0, \infty)$ martingale problem for (A, v) is unique.

Proof. Let V_1, V_2, \dots be a sequence of open subsets of E such that for each i there exists a k with $V_i = U_k$ and that for each k there exist infinitely many i with $V_i = U_k$. Fix $v \in \mathcal{P}(E)$, and let X be a solution of the martingale problem for (A, v) . Let $\tau_0 = 0$ and for $i \geq 1$

$$(6.3) \quad \tau_i = \inf \{t \geq \tau_{i-1}: X(t) \notin V_i \text{ or } X(t-) \notin V_i\}.$$

We note that $\lim_{i \rightarrow \infty} \tau_i = \infty$. (See Problem 27.) For $f \in \bar{C}(E)$ and $\lambda > 0$,

$$\begin{aligned} (6.4) \quad E \left[\int_0^\infty e^{-\lambda t} f(X(t)) dt \right] \\ = \sum_{i=1}^{\infty} E \left[\int_{\tau_{i-1}}^{\tau_i} e^{-\lambda t} f(X(t)) dt \right] \\ = \sum_{i=1}^{\infty} E \left[e^{-\lambda \tau_{i-1}} \chi_{\{\tau_{i-1} < \infty\}} \int_0^{\tau_i} e^{-\lambda t} f(X(\tau_{i-1} + t)) dt \right], \end{aligned}$$

where on $\{\tau_{i-1} < \infty\}$

$$(6.5) \quad \eta_i = \tau_i - \tau_{i-1} = \inf \{t \geq 0: X(\tau_{i-1} + t) \notin V_i \text{ or } X((\tau_{i-1} + t)-) \notin V_i\}.$$

For $i \geq 1$ such that $P\{\tau_{i-1} < \infty\} > 0$, define

$$(6.6) \quad \mu_i(B) = \frac{E[e^{-\lambda\tau_{i-1}} \chi_{(\tau_{i-1} < \infty)} \chi_B(X(\tau_{i-1}))]}{E[e^{-\lambda\tau_{i-1}} \chi_{(\tau_{i-1} < \infty)}]}$$

for $B \in \mathcal{B}(E)$, and

$$(6.7) \quad P_i(C) = \frac{E[e^{-\lambda\tau_{i-1}} \chi_{(\tau_{i-1} < \infty)} \chi_C(X(\tau_{i-1} + \cdot \wedge \eta_i))]}{E[e^{-\lambda\tau_{i-1}} \chi_{(\tau_{i-1} < \infty)}]}$$

for $C \in \mathcal{S}_E$.

Let Y_i be the coordinate process on $(D_E[0, \infty), \mathcal{S}_E, P_i)$. Then Y_i is a solution of the stopped martingale problem for (A, μ_i, V_i) , and hence, given μ_i , its distribution P_i is uniquely determined. Set

$$(6.8) \quad \gamma_i = \inf \{t: Y_i(t) \notin V_i \text{ or } Y_i(t-) \notin V_i\}.$$

Then for $i \geq 1$ with $P\{\tau_i < \infty\} > 0$

$$(6.9) \quad \begin{aligned} \mu_{i+1}(B) &= \frac{E[e^{-\lambda\tau_{i-1}} \chi_{(\tau_{i-1} < \infty)} e^{-\lambda\eta_i} \chi_{(\eta_i < \infty)} \chi_B(X(\tau_i))]}{\alpha_i} \\ &= \frac{E^{P_i}[e^{-\lambda\gamma_i} \chi_{(\gamma_i < \infty)} \chi_B(Y_i(\gamma_i))] \alpha_{i-1}}{\alpha_i}, \end{aligned}$$

where

$$(6.10) \quad \begin{aligned} \alpha_i &= E[e^{-\lambda\tau_i} \chi_{(\tau_i < \infty)}] = E[e^{-\lambda\tau_{i-1}} \chi_{(\tau_{i-1} < \infty)} e^{-\lambda\eta_i} \chi_{(\eta_i < \infty)}] \\ &= E^{P_i}[e^{-\lambda\gamma_i} \chi_{(\gamma_i < \infty)}] \alpha_{i-1} \\ &= \prod_{k=1}^i E^{P_k}[e^{-\lambda\gamma_k} \chi_{(\gamma_k < \infty)}]. \end{aligned}$$

Consequently, P_1, \dots, P_i determine μ_{i+1} , which in turn determines P_{i+1} . Since $\mu_1 = v$, it follows that the P_i are the same for all solutions of the martingale problem for (A, v) with sample paths in $D_E[0, \infty)$. But the right side of (6.4) can be written

$$(6.11) \quad \sum_{i=1}^{\infty} E^{P_i} \left[\int_0^{\gamma_i} e^{-\lambda t} f(Y_i(t)) dt \right] \alpha_{i-1}$$

so that (6.4) is the same for all solutions of the $D_E[0, \infty)$ martingale problem for (A, v) , and since λ is arbitrary, the uniqueness of the Laplace transform implies $E[f(X(t))]$ is the same for all solutions. (Note $E[f(X(t))]$ is right continuous as a function of t .) Since $f \in \bar{C}(E)$ is arbitrary, the one-dimensional distributions are determined and uniqueness follows by Corollary 4.3. \square

Note that the proof of Theorem 6.2 uses the uniqueness of the solution of the stopped martingale problem for (A, μ, U_k) for every choice of μ . The next result does not require this.

6.3 Theorem Let (E, r) be complete and separable, and let $A \subset \bar{C}(E) \times B(E)$. Let $U_1 \subset U_2 \subset \dots$ be open subsets of E . Fix $v \in \mathcal{P}(E)$, and suppose that for each k there exists a unique solution X_k of the stopped martingale problem for (A, v, U_k) with sample paths in $D_E[0, \infty)$. Setting

$$(6.12) \quad \tau_k = \inf \{t: X_k(t) \notin U_k \text{ or } X_k(t-) \notin U_k\},$$

suppose that for each $t > 0$,

$$(6.13) \quad \lim_{k \rightarrow \infty} P\{\tau_k \leq t\} = 0.$$

Then there exists a unique solution of the $D_E[0, \infty)$ martingale problem for (A, v) .

Proof. Let

$$(6.14) \quad \tau_m^k = \inf \{t: X_m(t) \notin U_k \text{ or } X_m(t-) \notin U_k\}.$$

For $k < m$, $X_m(\cdot \wedge \tau_m^k)$ is a solution of the stopped martingale problem for (A, v, U_k) and hence has the same distribution as X_k . It follows from (6.13) that there exists a process X_∞ such that $X_k \Rightarrow X_\infty$. (In particular, for the metric on $D_E[0, \infty)$ given by (5.2) of Chapter 3, the Prohorov distance between the distributions of X_k and X_∞ is less than $E[e^{-\tau_k \wedge T}]$.) In fact, for any bounded measurable function G on $D_E[0, \infty)$ and any $T > 0$,

$$(6.15) \quad |E[G(X_k(\cdot \wedge T))] - E[G(X_\infty(\cdot \wedge T))]| \leq 2\|G\|P\{\tau_k \leq T\}.$$

Let τ_∞^k be defined as in (6.14). Since the distribution of $X_m(\cdot \wedge \tau_m^k)$ does not depend on $m \geq k$, it follows that the distribution of $X_\infty(\cdot \wedge \tau_\infty^k)$ is the same as that of X_k . Hence

$$(6.16) \quad f(X_\infty(t \wedge \tau_\infty^k)) - \int_0^{t \wedge \tau_\infty^k} g(X_\infty(s)) ds$$

is an $\{\mathcal{F}_s^{X_\infty}\}$ -martingale for each k . Since $\lim_{k \rightarrow \infty} \tau_\infty^k = \infty$ a.s. ($P\{\tau_\infty^k \leq t\} = P\{\tau_k \leq t\}$), we see that X_∞ is a solution of the martingale problem for (A, v) with sample paths in $D_E[0, \infty)$. If X is a solution of the $D_E[0, \infty)$ martingale problem for (A, v) and

$$(6.17) \quad \gamma_k = \inf \{t: X(t) \notin U_k \text{ or } X(t-) \notin U_k\},$$

then $X(\cdot \wedge \gamma_k)$ has the same distribution as X_k , and hence X has the same distribution as X_∞ . \square

6.4 Corollary Let (E, r) be complete and separable. Let A_k , $k = 1, 2, \dots$, $A \subset \bar{C}(E) \times B(E)$ and suppose there exist open subsets $U_1 \subset U_2 \subset \dots$ with $\bigcup_k U_k = E$ such that

$$(6.18) \quad \{(f, \chi_{U_k} g): (f, g) \in A_k\} = \{(f, \chi_{U_k} g): (f, g) \in A\}.$$

(In the single-valued case this is just $\mathcal{D}(A_k) = \mathcal{D}(A)$ and $A_k f|_{U_k} = Af|_{U_k}$ for all $f \in \mathcal{D}(A)$.) If for each k , the $D_E[0, \infty)$ martingale problem for A_k is well-posed, and if for each $v \in \mathcal{P}(E)$, the sequence of solutions $\{X_k\}$ of the $D_E[0, \infty)$ martingale problems for (A_k, v) , $k = 1, 2, \dots$, is relatively compact, then the $D_E[0, \infty)$ martingale problem for A is well-posed.

Proof. Any solution of the stopped martingale problem for (A, v, U_k) is a solution for the stopped martingale problem for (A_k, v, U_k) and hence is unique by Theorem 6.1. Set $\tau_k = \inf \{t: X_k(t) \notin U_k \text{ or } X_k(t-) \notin U_k\}$. Then $\tilde{X}_k = X_k(\cdot \wedge \tau_k)$ is the solution of the stopped martingale problem and (6.13) follows from the relative compactness of $\{X_k\}$. Theorem 6.3 then gives the desired result. \square

The following lemma is useful in obtaining the monotone sequence $\{U_k\}$ in Theorem 6.3.

6.5 Lemma Let E be locally compact and separable, and let U_1, U_2 be open subsets of E with compact closure. Let $A \subset \bar{C}(E) \times B(E)$, and suppose $\mathcal{D}(A)$ separates points. If for each $v \in \mathcal{P}(E)$ and $k = 1, 2$, there exists a solution of the stopped martingale problem for (A, v, U_k) , then for each $v \in \mathcal{P}(E)$ there exists a solution of the stopped martingale problem for $(A, v, U_1 \cup U_2)$.

Proof. Let $V_k = U_1$ for k odd and $V_k = U_2$ for k even, and fix $x_0 \in E$. Lemma 5.15 can be used to construct a process X and stopping times τ_i such that $\tau_0 = 0$, τ_i is given by (6.3), and $X(t) = X_\ell(t - \tau_i)$, $\tau_i \leq t < \tau_{i+1}$, where X_ℓ is a solution of the stopped martingale problem for (A, μ_ℓ, V_ℓ) , $\mu_0 = v$, and $\mu_\ell(\Gamma) = P\{X(\tau_i) \in \Gamma, \tau_i < \infty\} + \delta_{x_0}(\Gamma)P\{\tau_i = \infty\}$. Let $\tau_\infty = \lim_{i \rightarrow \infty} \tau_i$ and note that

$$(6.19) \quad f(X(t \wedge \tau_\infty)) - \int_0^{t \wedge \tau_\infty} g(X(s)) ds$$

is an $\{\mathcal{F}_t^X\}$ -martingale for every $(f, g) \in A$. Either $\tau_\infty = \infty$, $\tau_i = \tau_\infty < \infty$ for $i \geq i_0$ (some i_0) or $\tau_i < \tau_\infty < \infty$ for all $i \geq 0$.

In the second case $\tau_{i+1} = \tau_i$ implies $X(\tau_i) \notin V_i$ and hence $X(\tau_\infty) \notin U_1 \cup U_2$. In the third case, the fact that (6.19) is a martingale implies $\lim_{i \rightarrow \tau_\infty^-} f(X(t))$ exists for $f \in \mathcal{D}(A)$, and the compactness of $\overline{U_1 \cup U_2}$ and the fact that $\mathcal{D}(A)$ is separating imply $\lim_{i \rightarrow \tau_\infty^-} X(t)$ exists. But either $X(\tau_i)$ or $X(\tau_i-) \notin V_i$, and hence $\lim_{i \rightarrow \tau_\infty^-} X(t) \notin U_1 \cup U_2$. Consequently, $\tau \equiv \inf \{t: X(t) \notin U_1 \cup U_2 \text{ or } X(t-) \notin U_1 \cup U_2\} \leq \tau_\infty$, and $X(\cdot \wedge \tau)$ is a solution of the stopped martingale problem for $(A, v, U_1 \cup U_2)$. \square

6.6 Theorem Let E be compact, $A \subset \bar{C}(E) \times B(E)$, and suppose that $\mathcal{D}(A)$ separates points. Suppose that for each $x \in E$ there exists an open set $U \subset E$ with $x \in U$ such that for each $v \in \mathcal{P}(E)$ there exists a solution of the stopped

martingale problem for (A, v, U) . Then for each $v \in \mathcal{P}(E)$ there exists a solution of the $D_E[0, \infty)$ martingale problem for (A, v) .

Proof. This is an immediate consequence of Lemma 6.5. \square

7. THE MARTINGALE PROBLEM: GENERALIZATIONS

A. The Time-dependent Martingale Problem

It is natural to consider processes whose parameters vary in time. With this in mind let $A \subset B(E) \times B(E \times [0, \infty))$. Then a measurable E -valued process X is a solution of the martingale problem for A if, for each $(f, g) \in A$,

$$(7.1) \quad f(X(t)) - \int_0^t g(X(s), s) ds$$

is an $\{\mathcal{F}_t^X\}$ -martingale. As before, X is a solution of the martingale problem for A with respect to $\{\mathcal{G}_t\}$, where $\mathcal{G}_t \supset \mathcal{F}_t^X$, if (7.1) is a $\{\mathcal{G}_t\}$ -martingale for each $(f, g) \in A$. Most of the basic results concerning martingale problems can be extended to the time-dependent case by considering the space-time process $X^0(t) = (X(t), t)$.

7.1 Theorem Let $A \subset B(E) \times B(E \times [0, \infty))$ and define $A^0 \subset B(E \times [0, \infty)) \times B(E \times [0, \infty))$ by

$$(7.2) \quad A^0 = \{(f\gamma, g\gamma + f\gamma'): (f, g) \in A, \gamma \in C_c^1[0, \infty)\}.$$

Then X is a solution of the martingale problem for A with respect to $\{\mathcal{G}_t\}$ if and only if the space-time process X^0 is a solution of the martingale problem for A^0 with respect to $\{\mathcal{G}_t\}$.

Proof. If X is a solution of the martingale problem for A with respect to $\{\mathcal{G}_t\}$, then for $(f, g) \in A$ and $\gamma \in C_c^1[0, \infty)$,

$$(7.3) \quad f(X(t))\gamma(t) - \int_0^t (g(X(s), s)\gamma(s) + f(X(s))\gamma'(s)) ds$$

is a $\{\mathcal{G}_t\}$ -martingale by the argument used to prove Lemma 3.4. The converse follows by considering $\gamma \equiv 1$ on $[0, T]$, $T > 0$. \square

Suppose (7.1) can be written

$$(7.4) \quad f(X(t)) - \int_0^t A(s)f(X(s)) ds$$

where $\{A(t): t \geq 0\}$ is a family of bounded operators. In this case in considering the space-time process the boundedness of the operators is lost. Fortunately, the bounded case is easy to treat directly. Consider generators of the form

$$(7.5) \quad A(t)f(x) = \lambda(t, x) \int (f(y) - f(x))\mu(t, x, dy)$$

where $\lambda \in M([0, \infty) \times E)$ is nonnegative and bounded in x for each fixed t , $\mu(t, x, \cdot) \in \mathcal{P}(E)$ for every $(t, x) \in [0, \infty) \times E$, and $\mu(\cdot, \cdot, \Gamma) \in B([0, \infty) \times E)$ for every $\Gamma \in \mathcal{B}(E)$. We can obtain a (time inhomogeneous) transition function for a Markov process as a solution of the equation

$$(7.6) \quad P(s, t, x, \Gamma) = \delta_x(\Gamma) + \int_s^t \lambda(u, x) \int (P(u, t, y, \Gamma) - P(u, t, x, \Gamma))\mu(u, x, dy) du.$$

7.2 Lemma Suppose there exists a measurable function γ on $[0, \infty)$ such that $\lambda(t, x) \leq \gamma(t)$ for all t and x and that for each $T > 0$

$$(7.7) \quad \int_0^T \gamma(t) dt < \infty.$$

Then (7.6) has a unique solution.

Proof. We first obtain uniqueness. Suppose P and \tilde{P} are solutions of (7.6) and set

$$(7.8) \quad M(s, t) = \sup_{x, \Gamma} \sup_{s \leq u \leq t} |P(u, t, x, \Gamma) - \tilde{P}(u, t, x, \Gamma)|.$$

Since M is nonincreasing in s , it is measurable in s and

$$(7.9) \quad M(s, t) \leq \int_s^t \gamma(u) 2M(u, t) du.$$

A slight modification of Gronwall's inequality (Appendix 5) implies $M(s, t) = 0$.

Existence is obtained by iteration. To see that the solution is a transition function it is simplest to first transform the equation to

$$(7.10) \quad P(s, t, x, \Gamma) = \delta_x(\Gamma) \exp \left\{ - \int_s^t \lambda(u, x) du \right\} \\ + \int_s^t \lambda(u, x) \exp \left\{ - \int_u^t \lambda(r, x) dr \right\} \int P(u, t, y, \Gamma) \mu(u, x, dy) du.$$

(To see that (7.10) is equivalent to (7.6), differentiate both sides with respect to s .) Fix t and set $P^0(s, t, x, \Gamma) = \delta_x(\Gamma)$ and

$$(7.11) \quad P^{n+1}(s, t, x, \Gamma) = \delta_x(\Gamma) \exp \left\{ - \int_s^t \lambda(u, x) du \right\} \\ + \int_s^t \lambda(u, x) \exp \left\{ - \int_s^u \lambda(r, x) dr \right\} \int P^n(u, t, y, \Gamma) \mu(u, x, dy) du.$$

Note that for each n , $P^n(s, t, x, \cdot) \in \mathcal{P}(E)$, and for each $\Gamma \in \mathcal{B}(E)$, $P^n(\cdot, \cdot, \cdot, \Gamma)$ is a Borel measurable function. Let

$$M_n(s, t) = \sup_{x, \Gamma} \sup_{s \leq u \leq t} |P^{n+1}(u, t, x, \Gamma) - P^n(u, t, x, \Gamma)|.$$

Then

$$(7.12) \quad M_n(s, t) \leq \sup_x \int_s^t \lambda(u, x) \exp \left\{ - \int_s^u \lambda(r, x) dr \right\} M_{n-1}(u, t) du \\ \leq \int_s^t \gamma(u) M_{n-1}(u, t) du.$$

Consequently,

$$(7.13) \quad \sum_{k=0}^n M_k(s, t) \leq M_0(s, t) + \int_s^t \gamma(u) \sum_{k=0}^n M_k(u, t) du \\ \leq 2 + \int_s^t \gamma(u) \sum_{k=0}^n M_k(u, t) du,$$

and by Gronwall's inequality,

$$(7.14) \quad \sum_{k=0}^n M_k(s, t) \leq 2 \exp \left\{ \int_s^t \gamma(u) du \right\}.$$

From (7.14) we conclude that $\{P^n(s, t, x, \Gamma)\}$ is a Cauchy sequence whose limit must satisfy (7.6). \square

7.3 Theorem Let λ and μ be as above, and define

$$(7.15) \quad A = \left\{ (f, \lambda(\cdot, \cdot)) \int (f(y) - f(\cdot)) \mu(\cdot, \cdot, dy) : f \in B(E) \right\}.$$

If λ satisfies the conditions of Lemma 7.2, then the martingale problem for A is well-posed.

Proof. The proof is left as Problem 29. \square

B. The Local-Martingale Problem

If we relax the condition that the functions in A be bounded, the natural requirement to place on X is that

$$(7.16) \quad f(X(t)) - \int_0^t g(X(s)) \, ds$$

be a local martingale. (In particular if we drop the boundedness assumption, (7.16) need not be in L^1 .) Consequently, for $A \subset M(E) \times M(E)$ we say that a measurable E -valued process X is a solution of the local-martingale problem for A if for each $(f, g) \in A$

$$(7.17) \quad \int_0^t |g(X(s))| \, ds < \infty \quad \text{a.s.}$$

and (7.16) is an $\{\mathcal{F}_t^X\}$ -local martingale. For example, let

$$(7.18) \quad A = \{(f, \frac{1}{2}f'') : f \in C^2(\mathbb{R})\}.$$

Then the unique solution of the local-martingale problem for (A, v) , $v \in \mathcal{P}(\mathbb{R})$, is just Brownian motion with mean zero, variance parameter 1, and initial distribution v . Itô's formula (Theorem 2.9 of Chapter 5) ensures that (7.16) is a local martingale.

Let $A_1 \subset B(E) \times B(E)$. Let β be nonnegative and measurable (but not necessarily bounded) and set

$$(7.19) \quad A_2 = \{(f, \beta g) : (f, g) \in A_1\}.$$

If Y is a solution of the martingale problem for A_1 and τ satisfies

$$(7.20) \quad \int_0^{\tau(t)} \frac{1}{\beta(Y(s))} \, ds = t, \quad t \geq 0,$$

then $X \equiv Y(\tau(\cdot))$ is a solution of the local-martingale problem for A_2 . (See Chapter 6.)

Many of the results in the previous sections extend easily to local-martingale problems.

C. The Martingale Problem Corresponding to a Given Semigroup

Let $\{T(t)\}$ be a semigroup of operators defined on a closed subspace $L \subset B(E)$. Then X is a solution of the martingale problem for $\{T(t)\}$ if, for each $u > 0$ and $f \in L$,

$$(7.21) \quad T(u-t)f(X(t))$$

is a martingale on the time interval $[0, u]$ with respect to $\{\mathcal{F}_t^X\}$. Of course if L is separating, then X is a Markov process corresponding to $\{T(t)\}$.

8. CONVERGENCE THEOREMS

In Theorem 2.5 we related the weak convergence of a sequence of Feller processes to the convergence of the corresponding semigroups. In this section we give conditions for more general sequences of processes to converge to Markov processes and allow the limiting Markov process to be determined either by its semigroup or as a solution of a martingale problem.

If a sequence of processes $\{X_n\}$ approximates a Markov process, it is reasonable to expect the processes to be approximately Markovian in some sense. One way of expressing this would be to require

$$(8.1) \quad \lim_{n \rightarrow \infty} E[|E[f(X_n(t+s))|\mathcal{F}_t^{X_n}] - T(s)f(X_n(t))|] = 0$$

where $\{T(s)\}$ is the semigroup corresponding to the limiting process. The following lemma shows that a condition weaker than (8.1) is in fact sufficient.

8.1 Lemma Let (E, r) be complete and separable. Let $\{X_n\}$ be a sequence of processes with values in E , and let $\{T(t)\}$ be a semigroup on a closed subspace $L \subset \bar{C}(E)$. Let $M \subset \bar{C}(E)$ be separating, and suppose either L is separating and $\{X_n(t)\}$ is relatively compact for each $t \geq 0$, or L is convergence determining. If X is a Markov process corresponding to $\{T(t)\}$, $X_n(0) \Rightarrow X(0)$, and

$$(8.2) \quad \lim_{n \rightarrow \infty} E\left[(f(X_n(t+s)) - T(s)f(X_n(t))) \prod_{i=1}^k g_i(X_n(t_i))\right] = 0$$

for all $k \geq 0$, $0 \leq t_1 < t_2 < \dots < t_k \leq t < t+s$, $f \in L$, and $g_1, \dots, g_k \in M$, then the finite-dimensional distributions of X_n converge weakly to those of X .

Proof. Let $f \in L$ and $t > 0$. Then, since $X_n(0) \Rightarrow X(0)$ and $T(t)f$ is continuous, (8.2) implies

$$(8.3) \quad \begin{aligned} \lim_{n \rightarrow \infty} E[f(X_n(t))] &= \lim_{n \rightarrow \infty} E[T(t)f(X_n(0))] \\ &= E[T(t)f(X(0))] = E[f(X(t))], \end{aligned}$$

and hence $X_n(t) \Rightarrow X(t)$, using Lemma 4.3 of Chapter 3 under the first conditions on L . Fix $m > 0$ and suppose $(X_n(t_1), \dots, X_n(t_m)) \Rightarrow (X(t_1), \dots, X(t_m))$ for all $0 \leq t_1 < t_2 < \dots < t_m$. Then (8.2) again implies

$$\begin{aligned} (8.4) \quad &\lim_{n \rightarrow \infty} E\left[f(X_n(t_{m+1})) \prod_{i=1}^m g_i(X_n(t_i))\right] \\ &= \lim_{n \rightarrow \infty} E\left[T(t_{m+1} - t_m)f(X_n(t_m)) \prod_{i=1}^m g_i(X_n(t_i))\right] \\ &= E\left[T(t_{m+1} - t_m)f(X(t_m)) \prod_{i=1}^m g_i(X(t_i))\right] \\ &= E\left[f(X(t_{m+1})) \prod_{i=1}^m g_i(X(t_i))\right] \end{aligned}$$

for all $0 \leq t_1 < t_2 < \dots < t_{m+1}$, $f \in L$, and $g_1, \dots, g_m \in M$. Since relative compactness of $\{X_n(t)\}$, $t \geq 0$, implies relative compactness of $\{(X_n(t_1), \dots, X_n(t_{m+1}))\}$, we may apply Lemma 4.3 and Proposition 4.6, both from Chapter 3, to conclude $(X_n(t_1), \dots, X_n(t_{m+1})) \Rightarrow (X(t_1), \dots, X(t_{m+1}))$. \square

The convergence in (8.1) (or (8.2)) can be viewed as a type of semigroup convergence (cf. Chapter 2, Section 7). For $n = 1, 2, \dots$ let $\{\mathcal{G}_i^n\}$ be a complete filtration, and let \mathcal{L}_n be the space of real-valued $\{\mathcal{G}_i^n\}$ -progressive processes ξ satisfying

$$(8.5) \quad \sup_{t \leq T} E[|\xi(t)|] < \infty$$

for each $T > 0$. Let $\hat{\mathcal{A}}_n$ be the collection of pairs $(\xi, \varphi) \in \mathcal{L}_n \times \mathcal{L}_n$ such that

$$(8.6) \quad \xi(t) - \int_0^t \varphi(s) \, ds$$

is a $\{\mathcal{G}_i^n\}$ -martingale. Note that if X_n is a $\{\mathcal{G}_i^n\}$ -progressive solution of the martingale problem for $A_n \subset B(E) \times B(E)$ with respect to $\{\mathcal{G}_i^n\}$, then $(f \circ X_n, g \circ X_n) \in \hat{\mathcal{A}}_n$ for each $(f, g) \in A_n$.

8.2 Theorem Let (E, r) be complete and separable. Let $A \subset \bar{C}(E) \times \bar{C}(E)$ be linear, and suppose the closure of A generates a strongly continuous contraction semigroup $\{T(t)\}$ on $L \equiv \mathcal{D}(A)$. Suppose X_n , $n = 1, 2, \dots$, is a $\{\mathcal{G}_i^n\}$ -progressive E -valued process, X is a Markov process corresponding to $\{T(t)\}$, and $X_n(0) \Rightarrow X(0)$. Let $M \subset \bar{C}(E)$ be separating and suppose either L is separating and $\{X_n(t)\}$ is relatively compact for each $t \geq 0$, or L is convergence determining. Then the following are equivalent:

(a) The finite-dimensional distributions of X_n converge weakly to those of X .

(b) For each $(f, g) \in A$,

$$(8.7)$$

$$\lim_{n \rightarrow \infty} E \left[\left(f(X_n(t+s)) - f(X_n(t)) - \int_t^{t+s} g(X_n(u)) \, du \right) \prod_{i=1}^k h_i(X_n(t_i)) \right] = 0$$

for all $k \geq 0$, $0 \leq t_1 < t_2 < \dots < t_k \leq t < t+s$, and $h_1, \dots, h_k \in M$.

(c) For each $(f, g) \in A$ and $T > 0$, there exist $(\xi_n, \varphi_n) \in \hat{\mathcal{A}}_n$ such that

$$(8.8) \quad \sup_n \sup_{s \leq T} E[|\xi_n(s)|] < \infty,$$

$$(8.9) \quad \sup_n \sup_{s \leq T} E[|\varphi_n(s)|] < \infty,$$

$$(8.10) \quad \lim_{n \rightarrow \infty} E \left[(\xi_n(t) - f(X_n(t))) \prod_{i=1}^k h_i(X_n(t_i)) \right] = 0,$$

and

$$(8.11) \quad \lim_{n \rightarrow \infty} E \left[(\varphi_n(t) - g(X_n(t))) \prod_{i=1}^k h_i(X_n(t_i)) \right] = 0,$$

for all $k \geq 0$, $0 \leq t_1 < t_2 < \dots < t_k \leq t \leq T$, and $h_1, \dots, h_k \in M$.

8.3 Remark (a) Note that (8.10) and (8.11) are implied by

$$(8.12) \quad \lim_{n \rightarrow \infty} E[|\xi_n(t) - f(X_n(t))|] = \lim_{n \rightarrow \infty} E[|\varphi_n(t) - g(X_n(t))|] = 0.$$

If this stronger condition holds, then (8.1) will hold.

(b) Frequently a good choice for (ξ_n, φ_n) will be

$$(8.13) \quad \xi_n(t) = \varepsilon_n^{-1} \int_0^{t_n} E[f(X_n(t+s)) | \mathcal{G}_t^n] ds$$

and

$$(8.14) \quad \varphi_n(t) = \varepsilon_n^{-1} E[f(X_n(t+\varepsilon_n)) - f(X_n(t)) | \mathcal{G}_t^n]$$

for some positive sequence $\{\varepsilon_n\}$ tending to zero. See Proposition 7.5 of Chapter 2.

(c) Conditions (8.9) and (8.11) can be relaxed. In fact it is sufficient to have

$$(8.15) \quad \lim_{n \rightarrow \infty} \int_0^{T-t} \left| E \left[(\varphi_n(t+s) - g(X_n(t+s))) \prod_{i=1}^k h_i(X_n(t_i)) \right] \right| ds = 0$$

for all $k \geq 0$, $0 \leq t_1 < t_2 < \dots < t_k \leq t \leq T$, and $h_1, \dots, h_k \in M$. See (8.23) below.

(d) For the implication $(c \Rightarrow a)$ we may drop the assumption that $\mathcal{D}(A) \subset \tilde{C}(E)$ provided $f \in L$ and $h \in M$ imply $f \cdot h \in L$ and

$$(8.16) \quad \lim_{n \rightarrow \infty} E[f(X_n(0))] = E[f(X(0))], \quad f \in L.$$

Note that (8.16) may be stronger than convergence in distribution since f need not be continuous. \square

Proof. (a \Rightarrow b) This is immediate.

(b \Rightarrow c) Let $(f, g) \in A$, and define ξ_n and φ_n by

$$(8.17) \quad \xi_n(t) = \int_0^\infty e^{-s} E[f(X_n(t+s)) - g(X_n(t+s)) | \mathcal{G}_t^n] ds$$

and

$$(8.18) \quad \varphi_n(t) = \xi_n(t) - f(X_n(t)) + g(X_n(t)).$$

Then $(\xi_n, \varphi_n) \in \hat{\mathcal{A}}_n$ by Theorem 7.1 of Chapter 2. Clearly (8.8) and (8.9) hold, and since $\xi_n - f \circ X_n = \varphi_n - g \circ X_n$, it is enough to verify (8.10).

Integrating by parts gives

$$\begin{aligned} (8.19) \quad \xi_n(t) &= \int_0^\infty e^{-s} E[f(X_n(t+s)) - g(X_n(t+s)) | \mathcal{G}_t^n] ds \\ &= \int_0^\infty e^{-s} E[f(X_n(t+s)) | \mathcal{G}_t^n] ds \\ &\quad - \int_0^\infty e^{-s} E \left[\int_0^s g(X_n(t+u)) du \middle| \mathcal{G}_t^n \right] ds \\ &= f(X_n(t)) + \int_0^\infty e^{-s} E \left[f(X_n(t+s)) \right. \\ &\quad \left. - f(X_n(t)) - \int_0^s g(X_n(t+u)) du \middle| \mathcal{G}_t^n \right] ds \end{aligned}$$

and (8.10) follows from (8.7) and the dominated convergence theorem.

(c \Rightarrow a) Let $(f, g) \in A$, and let (ξ_n, φ_n) be defined by (8.17) and (8.18). We claim that $\{(\xi_n, \varphi_n)\}$ satisfies (8.8)–(8.11) with T replaced by ∞ . As above, it is enough to consider (8.10). Let $T > 0$ and let $(\hat{\xi}_n, \hat{\varphi}_n) \in \hat{\mathcal{A}}_n$ satisfy (8.8)–(8.11) for all $k \geq 0$, $0 \leq t_1 < \dots < t_k \leq t \leq T$, and $h_1, \dots, h_k \in M$. Then

$$(8.20) \quad e^{-t} \hat{\xi}_n(t) + \int_0^t e^{-u} (\hat{\xi}_n(u) - \hat{\varphi}_n(u)) du$$

is a $\{\mathcal{G}_t^n\}$ -martingale (by the same argument as used in the proof of Lemma 3.2), and for $0 \leq t \leq T$,

$$(8.21) \quad \hat{\xi}_n(t) = e^{-(T-t)} E[\hat{\xi}_n(T) | \mathcal{G}_t^n] + \int_t^T e^{-(s-t)} E[\hat{\xi}_n(s) - \hat{\varphi}_n(s) | \mathcal{G}_t^n] ds.$$

Let $k \geq 0$, $0 \leq t_1 < \dots < t_k \leq t \leq T$, and $h_1, \dots, h_k \in M$. Then

$$\begin{aligned} (8.22) \quad &E \left[(\xi_n(t) - f(X_n(t))) \prod_{i=1}^k h_i(X_n(t_i)) \right] \\ &= E \left[(\xi_n(t) - \hat{\xi}_n(t)) \prod_{i=1}^k h_i(X_n(t_i)) \right] \\ &\quad + E \left[(\hat{\xi}_n(t) - f(X_n(t))) \prod_{i=1}^k h_i(X_n(t_i)) \right]. \end{aligned}$$

The first term on the right of (8.22) can be written, by (8.21), as

$$\begin{aligned}
 (8.23) \quad & E \left[\int_{T-t}^{\infty} e^{-s} E[f(X_n(t+s)) - g(X_n(t+s)) | \mathcal{G}_t^n] ds \prod_{i=1}^k h_i(X_n(t_i)) \right] \\
 & = E \left[e^{-(T-t)} (E[\xi_n(T) | \mathcal{G}_t^n] - f(X_n(T))) \prod_{i=1}^k h_i(X_n(t_i)) \right] \\
 & = E \left[e^{-(T-t)} f(X_n(T)) \prod_{i=1}^k h_i(X_n(t_i)) \right] \\
 & + E \left[\int_0^{T-t} e^{-s} (E[f(X_n(t+s)) - g(X_n(t+s)) | \mathcal{G}_t^n] \right. \\
 & \quad \left. - E[\xi_n(t+s) - \hat{\phi}_n(t+s) | \mathcal{G}_t^n]) ds \prod_{i=1}^k h_i(X_n(t_i)) \right].
 \end{aligned}$$

As $n \rightarrow \infty$ the second term on the right of (8.22) goes to zero by (8.10), as do the second and fourth terms in (8.23). (Note that the conditioning may be dropped, and the dominated convergence theorem justifies the interchange of limits and integration in the fourth term. Observe that (8.15) can be used here in place of (8.9) and (8.11).) Consequently,

$$\begin{aligned}
 (8.24) \quad & \overline{\lim}_{n \rightarrow \infty} \left| E \left[(\xi_n(t) - f(X_n(t))) \prod_{i=1}^k h_i(X_n(t_i)) \right] \right| \\
 & \leq e^{-(T-t)} (\|f - g\| + \|f\|) \prod_{i=1}^k \|h_i\|.
 \end{aligned}$$

Since T is arbitrary the limit is in fact zero.

Let \mathcal{L}_n^0 be the Banach space of real-valued $\{\mathcal{G}_t^n\}$ -progressive processes ξ with norm $\|\xi\| = \sup_t E[|\xi(t)|]$. Define $\pi_n: L \rightarrow \mathcal{L}_n^0$ by $\pi_n f(t) = f(X_n(t))$, and for $\xi_n \in \mathcal{L}_n^0$, $n = 1, 2, \dots$, and $f \in L$, define $\text{LIM } \xi_n = f$ if $\sup_n \|\xi_n\| < \infty$ and

$$(8.25) \quad \lim_{n \rightarrow \infty} E \left[(\xi_n(t) - \pi_n f(t)) \prod_{i=1}^k h_i(X_n(t_i)) \right] = 0$$

for all $k \geq 0$, $0 \leq t_1 < \dots < t_k \leq t$, and $h_1, \dots, h_k \in M$.

Let $\{\mathcal{T}_n(s)\}$ denote the semigroup of conditioned shifts on \mathcal{L}_n^0 . Clearly $\text{LIM } \xi_n = 0$ implies $\text{LIM } \mathcal{T}_n(s)\xi_n = 0$ for all $s \geq 0$, and LIM satisfies the conditions of Theorem 6.9 of Chapter 1. For each $(f, g) \in A$, we have shown there exist $(\xi_n, \varphi_n) \in \mathcal{A}_n$ such that $\text{LIM } \xi_n = f$ and $\text{LIM } \varphi_n = g$. Consequently, Theorem 6.9 of Chapter 1 implies

$$(8.26) \quad \text{LIM } \mathcal{T}_n(s)\pi_n f = T(s)f$$

for all $f \in L$ and $s \geq 0$. But (8.26) is just (8.2), and hence Lemma 8.1 implies (a). \square

8.4 Corollary Suppose in Theorem 8.2 that $X_n = \eta_n \circ Y_n$ and $\{\mathcal{G}_t^n\} = \{\mathcal{F}_t^{Y_n}\}$, where Y_n is a progressive Markov process in a metric space E_n corresponding to a measurable contraction semigroup $\{T_n(t)\}$ with full generator \hat{A}_n , and $\eta_n: E_n \rightarrow E$ is Borel measurable. Then (a), (b), and (c) are equivalent to the following:

(d) For each $(f, g) \in A$ and $T > 0$, there exist $(f_n, g_n) \in \hat{A}_n$ such that $\{(\xi_n, \varphi_n)\} \equiv \{(f_n \circ Y_n, g_n \circ Y_n)\}$ satisfies (8.8)–(8.11) for all $k \geq 0$, $0 \leq t_1 < \dots < t_k \leq t \leq T$, and $h_1, \dots, h_k \in M$.

Proof. ($d \Rightarrow c$) It only needs to be observed that $(f_n, g_n) \in \hat{A}_n$ implies $(f_n \circ Y_n, g_n \circ Y_n) \in \hat{\mathcal{A}}_n$.

($c \Rightarrow d$) By the Markov property, (ξ_n, φ_n) defined by (8.17) and (8.18) is of the form $(f_n \circ Y_n, g_n \circ Y_n)$ for some $(f_n, g_n) \in \hat{A}_n$, and (d) follows by (8.24). \square

8.5 Corollary Suppose in Theorem 8.2 that $X_n = \eta_n(Y_n([\alpha_n \cdot]))$ and $\{\mathcal{G}_t^n\} = \{\mathcal{F}_{[\alpha_n t]}^{Y_n}\}$, where $\{Y_n(k), k = 0, 1, 2, \dots\}$ is a Markov chain in a metric space E_n with transition function $\mu_n(x, \Gamma)$, $\eta_n: E_n \rightarrow E$ is Borel measurable, and $\alpha_n \rightarrow \infty$ as $n \rightarrow \infty$. Define $T_n: B(E_n) \rightarrow B(E_n)$ by

$$(8.27) \quad T_n f(x) = \int f(y) \mu_n(x, dy),$$

and let $A_n = \alpha_n(T_n - I)$. Then (a), (b), and (c) are equivalent to the following:

(e) For each $(f, g) \in A$ and $T > 0$, there exist $f_n \in B(E_n)$ such that for $g_n = A_n f_n$, $\{(\xi_n, \varphi_n)\} \equiv \{(f_n(Y_n([\alpha_n \cdot])), g_n(Y_n([\alpha_n \cdot])))\}$ satisfies (8.8)–(8.11) for all $k \geq 0$, $0 \leq t_1 < \dots < t_k \leq t \leq T$, and $h_1, \dots, h_k \in M$.

Proof. ($e \Rightarrow c$) In this case $(\xi_n, \varphi_n) \notin \hat{\mathcal{A}}_n$. Consequently, we define $(\hat{\xi}_n, \hat{\varphi}_n) \in \hat{\mathcal{A}}_n$ by

$$(8.28) \quad \begin{aligned} \hat{\xi}_n(t) &= \alpha_n \int_0^{1/\alpha_n} E[f_n(Y_n([\alpha_n(t+s)])) | \mathcal{G}_s^n] ds \\ &= f_n(Y_n([\alpha_n t])) + \left(\frac{t - [\alpha_n t]}{\alpha_n} \right) g_n(Y_n([\alpha_n t])) \end{aligned}$$

and

$$(8.29) \quad \begin{aligned} \hat{\phi}_n(t) &= \alpha_n E[f_n(Y_n(\lceil \alpha_n t \rceil + 1)) - f_n(Y_n(\lceil \alpha_n t \rceil))] \mid \mathcal{G}_t^n \\ &= g_n(Y_n(\lceil \alpha_n t \rceil)) = \varphi_n(t) \end{aligned}$$

(see Proposition 7.5 of Chapter 2) and note that $E[|\xi_n(t) - \hat{\xi}_n(t)|] \leq E[|\varphi_n(t)|]/\alpha_n$.

(b \Rightarrow e) Define $\pi_n: B(E) \rightarrow B(E_n)$ by $\pi_n f = f \circ \eta_n$. For $(f, g) \in A$, set

$$(8.30) \quad f_n = (1 + \alpha_n)^{-1} \sum_{k=0}^{\infty} \left(\frac{\alpha_n}{1 + \alpha_n} \right)^k T_n^k \pi_n(f - g)$$

and note that

$$(8.31) \quad g_n = A_n f_n = f_n + \pi_n g - \pi_n f.$$

For $t = k/\alpha_n$ ($k \in \mathbb{Z}_+$), (8.30) gives

$$\begin{aligned} (8.32) \quad f_n(Y_n(\lceil \alpha_n t \rceil)) &= E \left[(1 + \alpha_n)^{-1} \alpha_n \int_0^\infty \left(\frac{\alpha_n}{1 + \alpha_n} \right)^{\lfloor \alpha_n s \rfloor} (f(X_n(t+s)) - g(X_n(t+s))) ds \mid \mathcal{G}_t^n \right] \\ &= E \left[(1 + \alpha_n)^{-1} \alpha_n \int_0^\infty \left(\left(\frac{\alpha_n}{1 + \alpha_n} \right)^{\lfloor \alpha_n s \rfloor} - e^{-s} \right) \right. \\ &\quad \times (f(X_n(t+s)) - g(X_n(t+s))) ds \left. \mid \mathcal{G}_t^n \right] \\ &\quad + (1 + \alpha_n)^{-1} \alpha_n \int_0^\infty e^{-s} E[f(X_n(t+s)) - g(X_n(t+s)) \mid \mathcal{G}_t^n] ds. \end{aligned}$$

Since $\lim_{n \rightarrow \infty} (\alpha_n/(1 + \alpha_n))^{\lfloor \alpha_n s \rfloor} = e^{-s}$ uniformly in $s \geq 0$, the result now follows as in the proof of (b \Rightarrow c). \square

8.6 Corollary Suppose in Theorem 8.2 that the X_n and X have sample paths in $D_E[0, \infty)$, and there is an algebra $C_\sigma \subset L$ that separates points. Suppose either that the compact containment condition (7.9) of Chapter 3 holds for $\{X_n\}$ or that C_σ strongly separates points. If $\{(\xi_n, \varphi_n)\}$ in condition (c) can be selected so that

$$(8.33) \quad \lim_{n \rightarrow \infty} E \left[\sup_{t \in \mathbb{Q} \cap [0, T]} |\xi_n(t) - f(X_n(t))| \right] = 0$$

and

$$(8.34) \quad \sup_n E[\|\varphi_n\|_{p, T}] < \infty \quad \text{for some } p \in (1, \infty],$$

then $X_n \Rightarrow X$.

Proof. This follows from Theorems 9.4, 9.1, 7.8 and Corollary 9.2, all of Chapter 3. Note that D in that chapter's Theorem 9.4 contains $\mathcal{D}(A)$. \square

8.7 Corollary Suppose in Theorem 8.2 that the X_n and X have sample paths in $D_E[0, \infty)$, and that there is an algebra $C_n \subset L$ that separates points. Suppose either that the compact containment condition ((7.9) of Chapter 3) holds for $\{X_n\}$ or that C_n strongly separates points. If X_n has the form in Corollary 8.4 and $\pi_n f = f \circ \eta_n$, then either of the following conditions implies $X_n \Rightarrow X$.

(f) For each $(f, g) \in A$ and $T > 0$, there exist $(f_n, g_n) \in \hat{A}_n$ and $G_n \subset E_n$ such that $\{Y_n(t) \in G_n, 0 \leq t \leq T\}$ are events satisfying

$$(8.35) \quad \lim_{n \rightarrow \infty} P\{Y_n(t) \in G_n, 0 \leq t \leq T\} = 1,$$

$\sup_n \|f_n\| < \infty$, and

$$(8.36) \quad \lim_{n \rightarrow \infty} \sup_{y \in G_n} |\pi_n f(y) - f_n(y)| = \lim_{n \rightarrow \infty} \sup_{y \in G_n} |\pi_n g(y) - g_n(y)| = 0.$$

(g) For each $f \in L$ and $T > 0$, there exist $G_n \subset E_n$ such that (8.35) holds and

$$(8.37) \quad \lim_{n \rightarrow \infty} \sup_{y \in G_n} |T_n(t)\pi_n f(y) - \pi_n T(t)f(y)| = 0, \quad 0 \leq t \leq T.$$

8.8 Remark (a) If $G_n = E_n$, then (8.35) is immediate and (f) and (g) are equivalent by Theorem 6.1 of Chapter 1.

(b) If the Y_n are continuous and the G_n are compact, then the assumption that $\sup_n \|f_n\| < \infty$ can be dropped. In general it is needed. For example, with $E_n = E = \{0, 1, 2, \dots\}$ let

$$(8.38) \quad A_n f(k) = n^{-1} \delta_{0k}(f(n) - f(0)).$$

(Clearly, if X_n has generator A_n with $X_n(0) = 0$, then $X_n \Rightarrow X$ where $X \equiv 0$.) Let

$$(8.39) \quad Af(k) = \delta_{0k}(f(1) - f(0)).$$

Set $G_n = \{0, 1, 2, \dots, n-1\}$ and

$$(8.40) \quad f_n(k) = f(k) + \delta_{nk} n(f(1) - f(0)).$$

Then

$$(8.41) \quad A_n f_n(k) = \delta_{0k}(f(1) - f(0) + n^{-1}(f(n) - f(0))),$$

and hence

$$(8.42) \quad \lim_{n \rightarrow \infty} \sup_{k \in G_n} |f_n(k) - f(k)| = \lim_{n \rightarrow \infty} \sup_{k \in G_n} |A_n f_n(k) - Af(k)| = 0$$

suggesting (but not implying!) that $X_n \Rightarrow \tilde{X}$, where \tilde{X} has generator A . Of course $\sup_n \|f_n\| = \infty$. \square

Proof. Assume (g) holds. For $(f, g) \in A$, let

$$(8.43) \quad f_n = \int_0^\infty e^{-t} T_n(t) \pi_n(f - g) dt, \quad g_n = f_n - \pi_n(f - g).$$

Then $(f_n, g_n) \in \hat{A}_n$ and

$$(8.44) \quad |\pi_n f(y) - f_n(y)| = |\pi_n g(y) - g_n(y)|$$

$$\begin{aligned} &= \left| \int_0^\infty e^{-t} T_n(t) \pi_n(f - g)(y) dt - \int_0^\infty e^{-t} \pi_n T(t)(f - g)(y) dt \right| \\ &\leq \int_0^T e^{-t} |T_n(t) \pi_n(f - g)(y) - \pi_n T(t)(f - g)(y)| dt \\ &\quad + 2e^{-T} \|f - g\|. \end{aligned}$$

Using (8.37), the dominated convergence theorem, and the arbitrariness of T , a sequence G_n can be constructed satisfying (8.35) (for each $T > 0$) and (8.36). Consequently (g) implies (f).

To see that (f) implies $X_n \Rightarrow X$, fix $(f, g) \in A$ and $T > 0$. Assuming $(f_n, g_n) \in \hat{A}_n$ and $G_n \subset E_n$ are as in (f), define

$$(8.45) \quad \tau_n = \inf \left\{ t > 0 : \int_0^t |g_n(Y_n(s))|^2 ds \geq t(\|g\|^2 + 1) \right\}.$$

Note that (8.35) and (8.36) imply $\lim_{n \rightarrow \infty} P\{\tau_n < T\} = 0$. Set

$$(8.46) \quad \xi_n(t) = f_n(Y_n(t \wedge \tau_n)), \quad \varphi_n(t) = g_n(Y_n(t)) \chi_{\{t_n > t\}}.$$

Then ξ_n and φ_n satisfy (8.33) and (8.34) with $p = 2$ as well as (8.8), (8.10), and (8.15). \square

8.9 Corollary Suppose in Theorem 8.2 that the X_n and X have sample paths in $D_E[0, \infty)$ and that there is an algebra $C_n \subset L$ that separates points. Suppose either that the compact containment condition ((7.9) of Chapter 3) holds for $\{X_n\}$ or that C_n strongly separates points. If X_n has the form in Corollary 8.5 and $\pi_n f = f \circ \eta_n$, then either of the following conditions implies $X_n \Rightarrow X$:

(b) For each $(f, g) \in A$ and $T > 0$, there exist $f_n \in B(E_n)$ and $G_n \subset E_n$ such that $\{Y_n([\alpha_n t]) \in G_n, 0 \leq t \leq T\}$ are events satisfying

$$(8.47) \quad \lim_{n \rightarrow \infty} P\{Y_n([\alpha_n t]) \in G_n, 0 \leq t \leq T\} = 1,$$

$\sup_n \|f_n\| < \infty$, and

$$(8.48) \quad \lim_{n \rightarrow \infty} \sup_{y \in G_n} |\pi_n f(y) - f_n(y)| = \lim_{n \rightarrow \infty} \sup_{y \in G_n} |\pi_n g(y) - A_n f_n(y)| = 0.$$

- (i) For each $f \in L$ and $T > 0$, there exist $G_n \subset E_n$ such that (8.47) holds and

$$(8.49) \quad \lim_{n \rightarrow \infty} \sup_{y \in G_n} |T_n^{(a_n)} \pi_n f(y) - \pi_n T(t)f(y)| = 0, \quad 0 \leq t \leq T.$$

Proof. The proof is essentially the same as that of Corollary 8.7 using (8.30) in place of (8.43), and (8.28) (appropriately stopped) in place of (8.45). \square

We now give an analogue of Theorem 8.2 in which the assumption that the closure of A generates a strongly continuous semigroup is relaxed to the assumption of uniqueness of solutions of the martingale problem. We must now, however, assume a priori that the sequence $\{X_n\}$ is relatively compact. Note that $\{\mathcal{G}_t^n\}$ and $\hat{\mathcal{A}}_n$ are as in Theorem 8.2.

8.10 Theorem Let (E, r) be complete and separable. Let $A \subset \bar{C}(E) \times \bar{C}(E)$ and $v \in \mathcal{P}(E)$, and suppose that the $D_E[0, \infty)$ martingale problem for (A, v) has at most one solution. Suppose X_n , $n = 1, 2, \dots$, is a $\{\mathcal{G}_t^n\}$ -adapted process with sample paths in $D_E[0, \infty)$, $\{X_n\}$ is relatively compact, $PX_n(0)^{-1} \Rightarrow v$, and $M \subset \bar{C}(E)$ is separating. Then the following are equivalent:

(a') There exists a solution X of the $D_E[0, \infty)$ martingale problem for (A, v) , and $X_n \Rightarrow X$.

(b') There exists a countable set $\Gamma \subset [0, \infty)$ such that for each $(f, g) \in A$

(8.50)

$$\lim_{n \rightarrow \infty} E \left[\left(f(X_n(t+s)) - f(X_n(t)) - \int_t^{t+s} g(X_n(u)) du \right) \prod_{i=1}^k h_i(X_n(t_i)) \right] = 0$$

for all $k \geq 0$, $0 \leq t_1 < t_2 < \dots < t_k \leq t < t+s$ with $t_i, t, t+s \notin \Gamma$, and $h_1, \dots, h_k \in M$.

(c') There exists a countable set $\Gamma \subset [0, \infty)$ such that for each $(f, g) \in A$ and $T > 0$, there exist $(\xi_n, \varphi_n) \in \hat{\mathcal{A}}_n$ such that

$$(8.51) \quad \sup_n \sup_{s \leq T} E[|\xi_n(s)|] < \infty,$$

$$(8.52) \quad \sup_n \sup_{s \leq T} E[|\varphi_n(s)|] < \infty,$$

$$(8.53) \quad \lim_{n \rightarrow \infty} E \left[(\xi_n(t) - f(X_n(t))) \prod_{i=1}^k h_i(X_n(t_i)) \right] = 0,$$

and

$$(8.54) \quad \lim_{n \rightarrow \infty} E \left[(\varphi_n(t) - g(X_n(t))) \prod_{i=1}^k h_i(X_n(t_i)) \right] = 0$$

for all $k \geq 0$, $0 \leq t_1 < t_2 < \dots < t_k \leq t \leq T$ with $t_i, t \notin \Gamma$, and $h_1, \dots, h_k \in M$.

8.11 Remark As in Theorem 8.2, (8.52) and (8.54) can be replaced by (8.15). \square

Proof. (a' \Rightarrow b') Take $\Gamma = [0, \infty) - D(X)$. ($D(X) = \{t \geq 0 : P\{X(t) = X(t-)\} = 1\}$.) By Theorem 7.8 of Chapter 3, $(X_n(t_1), \dots, X_n(t_k)) \Rightarrow (X(t_1), \dots, X(t_k))$ for all finite sets $\{t_1, t_2, \dots, t_k\} \subset D(X)$, and this implies (8.50).

(b' \Rightarrow c') The proof is essentially the same as in Theorem 8.2.

(c' \Rightarrow a') Let Y be a limit point of $\{X_n\}$. Let $(f, g) \in A$ and $T > 0$, and let $\{(\xi_n, \varphi_n)\}$ satisfy the assertions of condition (c'). Let $k \geq 0$, $0 \leq t_1 < \dots < t_k \leq t < t+s \leq T$ with $t_i, t, t+s \in D(Y)$ and $h_1, \dots, h_k \in M$. Since

$$(8.55) \quad E \left[\left(\xi_n(t+s) - \xi_n(t) - \int_t^{t+s} \varphi_n(u) du \right) \prod_{i=1}^k h_i(X_n(t_i)) \right] = 0,$$

it follows that

$$(8.56) \quad \lim_{n \rightarrow \infty} E \left[\left(f(X_n(t+s)) - f(X_n(t)) - \int_t^{t+s} g(X_n(u)) du \right) \prod_{i=1}^k h_i(X_n(t_i)) \right] = 0,$$

and hence

$$(8.57) \quad E \left[\left(f(Y(t+s)) - f(Y(t)) - \int_t^{t+s} g(Y(u)) du \right) \prod_{i=1}^k h_i(Y(t_i)) \right] = 0.$$

By the right continuity of Y , (8.55) holds for all $0 \leq t_1 < \dots < t_k \leq t < t+s$, and hence Y is a solution of the martingale problem for (A, v) . Therefore (a') follows by the assumption of uniqueness. \square

We state analogues of Corollaries 8.4–8.7 and 8.9. Their proofs are similar and are omitted.

8.12 Corollary Suppose in Theorem 8.10 that $X_n = \eta_n \circ Y_n$ and $\{\mathcal{G}_t^n\} = \{\mathcal{F}_{[0,t]}^{Y_n}\}$, where Y_n is a progressive Markov process in a metric space E_n corresponding to a measurable contraction semigroup $\{T_n(t)\}$ with full generator \hat{A}_n , and $\eta_n: E_n \rightarrow E$ is Borel measurable. Then condition (f) of Corollary 8.7 implies (a'), and (a'), (b'), and (c') are equivalent to the following:

(d') There exists a countable set $\Gamma \subset [0, \infty)$ such that for every $(f, g) \in A$ and $T > 0$, there exist $(f_n, g_n) \in \hat{A}_n$ such that $\{(\xi_n, \varphi_n)\} = \{(f_n \circ Y_n, g_n \circ Y_n)\}$ satisfies (8.51)–(8.54) for all $k \geq 0$, $0 \leq t_1 < \dots < t_k \leq T$ with $t_i, t \notin \Gamma$, and $h_1, \dots, h_k \in M$.

8.13 Corollary Suppose in Theorem 8.10 that $X_n = \eta_n \circ Y_n([\alpha_n \cdot])$ and $\{\mathcal{G}_t^n\} = \{\mathcal{F}_{[0,t]}^{Y_n([\alpha_n \cdot])}\}$, where $\{Y_n(k), k = 0, 1, \dots\}$ is a Markov chain in a metric space E_n with transition function $\mu_n(x, \Gamma)$, $\eta_n: E_n \rightarrow E$ is Borel measurable, and $\alpha_n \rightarrow \infty$ as $n \rightarrow \infty$. Define $T_n: B(E_n) \rightarrow B(E_n)$ by (8.27) and let $A_n = \alpha_n(T_n - I)$. Then condition (h) of Corollary 8.9 implies (a'), and (a'), (b'), and (c') are equivalent to the following:

(e') There exists a countable set $\Gamma \subset [0, \infty)$ such that for every $(f, g) \in A$ and $T > 0$, there exist $f_n \in B(E_n)$ such that for $g_n = A_n f_n$, $\{(\xi_n, \varphi_n)\} = \{(f_n(Y_n([\alpha_n \cdot])), g_n(Y_n([\alpha_n \cdot])))\}$ satisfies (8.51)–(8.54) for all $k \geq 0$, $0 \leq t_1 < \dots < t_k \leq T$ with $t_i, t \notin \Gamma$, and $h_1, \dots, h_k \in M$. (Note that we are not claiming $(\xi_n, \varphi_n) \in \mathcal{A}_n$.)

8.14 Remark In the following three corollaries we do not assume a priori that $\{X_n\}$ is relatively compact. We do assume the compact containment condition. The assumption that C_α strongly separates points used in the analogous corollaries to Theorem 8.2 does not suffice. \square

8.15 Corollary Let (E, r) be complete and separable and let $\{X_n\}$ be a sequence of processes with sample paths in $D_E[0, \infty)$. Suppose $PX_n(0)^{-1} \Rightarrow v \in \mathcal{P}(E)$ and the compact containment condition ((7.9) of Chapter 3) holds. Suppose $A \subset \bar{C}(E) \times \bar{C}(E)$, the closure of the linear span of $\mathcal{D}(A)$ contains an algebra that separates points, and the $D_E[0, \infty)$ martingale problem for (A, v) has at most one solution. If $\{(\xi_n, \varphi_n)\}$ in condition (c') of Theorem 8.10 can be selected so that (8.33) and (8.34) hold, then (a') holds.

8.16 Corollary Instead of assuming in Corollary 8.12 that $\{X_n\}$ is relatively compact, suppose that $\{X_n\}$ satisfies the compact containment condition ((7.9) of Chapter 3) and the closure of the linear span of $\mathcal{D}(A)$ contains an algebra that separates points. Then condition (f) of Corollary 8.7 implies condition (a') of Theorem 8.10.

8.17 Corollary Instead of assuming in Corollary 8.13 that $\{X_n\}$ is relatively compact, suppose that $\{X_n\}$ satisfies the compact containment condition ((7.9) of Chapter 3) and that the closure of the linear span of $\mathcal{D}(A)$ contains an algebra that separates points. Then condition (h) of Corollary 8.9 implies condition (a') of Theorem 8.10.

The following proposition may be useful in verifying (8.7) or (8.50) and as a result gives an alternative convergence criterion.

8.18 Proposition Let X_n , A , and $\{\mathcal{G}_t^n\}$ be as in Theorem 8.2. Let $(f, g) \in A$. For $n = 1, 2, \dots$, let $0 = \tau_0^n < \tau_1^n < \dots$ be an increasing sequence of $\{\mathcal{G}_t^n\}$ -stopping times with $\tau_k^n < \infty$ a.s. and $\lim_{k \rightarrow \infty} \tau_k^n = \infty$ a.s. Define

$$(8.58) \quad \tau_n^+(t) = \min \{\tau_k^n : \tau_k^n > t\},$$

$$(8.59) \quad \tau_n^-(t) = \max \{\tau_k^n : \tau_k^n \leq t\},$$

and

$$(8.60) \quad H_n(t) = \sum_{k=0}^{\infty} (E[f(X_n(\tau_{k+1}^n)) - f(X_n(\tau_k^n)) | \mathcal{G}_{\tau_k^n}^n] \\ - g(X_n(\tau_k^n))E[\tau_{k+1}^n - \tau_k^n | \mathcal{G}_{\tau_k^n}^n]) \chi_{\{\tau_k^n \leq t\}}.$$

If

$$(8.61) \quad \lim_{n \rightarrow \infty} E[|f(X_n(\tau_n^+(t))) - f(X_n(t))|] \\ = \lim_{n \rightarrow \infty} E[\tau_n^+(t) - \tau_n^-(t)] = 0, \quad t \geq 0,$$

$$(8.62) \quad \lim_{n \rightarrow \infty} E[|g(X_n(\tau_n^-(t))) - g(X_n(t))|] = 0, \quad \text{a.e. } t \geq 0,$$

and

$$(8.63) \quad \lim_{n \rightarrow \infty} E[|H_n(t)|] = 0, \quad t \geq 0,$$

then

$$(8.64) \quad \lim_{n \rightarrow \infty} E\left[\left| E\left[f(X_n(t+s)) - f(X_n(t)) - \int_t^{t+s} g(X_n(u)) du \middle| \mathcal{G}_t^n \right] \right| \right] = 0$$

for all $s, t \geq 0$, which in turn implies, by (8.19),

$$(8.65) \quad \lim_{n \rightarrow \infty} E[|\zeta_n(t) - f(X_n(t))|] \\ = \lim_{n \rightarrow \infty} E[|\varphi_n(t) - g(X_n(t))|] = 0, \quad t \geq 0,$$

where ζ_n and φ_n are given by (8.17) and (8.18).

Proof. For any $\{\mathcal{G}_t^n\}$ -stopping time τ and any Z such that $E[|Z|] < \infty$,

$$(8.66) \quad \begin{aligned} E[E[Z | \mathcal{G}_t^n] | \mathcal{G}_t^n] \chi_{\{\tau > t\}} \\ = E[Z | \mathcal{G}_t^n] \chi_{\{\tau > t\}} = E[Z \chi_{\{\tau > t\}} | \mathcal{G}_t^n]. \end{aligned}$$

Consequently,

$$(8.67) \quad \begin{aligned} E \left[f(X_n(t+s)) - f(X_n(t)) - \int_t^{t+s} g(X_n(u)) du - (H_n(t+s) - H_n(t)) \middle| \mathcal{G}_t^n \right] \\ = E \left[f(X_n(t+s)) - f(X_n(t)) - \int_t^{t+s} g(X_n(u)) du \middle| \mathcal{G}_t^n \right] \\ - \sum E[(f(X_n(\tau_{k+1}^n)) - f(X_n(\tau_k^n))) \chi_{\{\tau_k^n < \tau_{k+1}^n \leq t+s\}} | \mathcal{G}_t^n] \\ + \sum E[g(X_n(\tau_k^n))(\tau_{k+1}^n - \tau_k^n) \chi_{\{\tau_k^n < \tau_{k+1}^n \leq t+s\}} | \mathcal{G}_t^n] \\ = E[f(X_n(t+s)) - f(X_n(\tau_n^+(t+s))) | \mathcal{G}_t^n] \\ - E[f(X_n(t)) - f(X_n(\tau_n^+(t))) | \mathcal{G}_t^n] \\ - E \left[\int_t^{t+s} g(X_n(u)) du - \int_{\tau_n^+(t)}^{t_n^+(t+s)} g(X_n(\tau_n^-(u))) du \middle| \mathcal{G}_t^n \right] \\ \equiv I_n(t). \end{aligned}$$

By (8.61), (8.62), and the dominated convergence theorem, $I_n(t)$ converges to zero in L^1 . The quantity in (8.64) is bounded by $E(|H_n(t+s)|) + E(|H_n(t)|) + E(|I_n(t)|)$ and the limit follows by (8.63). \square

9. STATIONARY DISTRIBUTIONS

Let $A \subset B(E) \times B(E)$ and suppose the martingale problem for A is well-posed. Then $\mu \in \mathcal{P}(E)$ is a *stationary distribution for A* if every solution X of the martingale problem for (A, μ) is a *stationary process*, that is, if $P\{X(t+s_1) \in \Gamma_1, X(t+s_2) \in \Gamma_2, \dots, X(t+s_k) \in \Gamma_k\}$ is independent of $t \geq 0$ for all $k \geq 1$, $0 \leq s_1 < \dots < s_k$, and $\Gamma_1, \dots, \Gamma_k \in \mathcal{B}(E)$.

The following lemma shows that to check that μ is a stationary distribution it is sufficient to consider only the one-dimensional distributions.

9.1 Lemma Let $A \subset B(E) \times B(E)$ and suppose the martingale problem for A is well-posed. Let $\mu \in \mathcal{P}(E)$ and let X be a solution of the martingale problem for (A, μ) . Then μ is a stationary distribution for A if and only if $X(t)$ has distribution μ for all $t \geq 0$.

Proof. The necessity is immediate. For sufficiency observe that $X_t \equiv X(t + \cdot)$ is a solution of the martingale problem for (A, μ) , and hence, by uniqueness, has the same finite-dimensional distributions as X . \square

9.2 Proposition Suppose A generates a strongly continuous contraction semigroup $\{T(t)\}$ on a closed subspace $L \subset B(E)$, L is separating, and the martingale problem for A is well-posed. If D is a core for A and $\mu \in \mathcal{P}(E)$, then the following are equivalent:

- (a) μ is a stationary distribution for A .
- (b) $\int T(t)f d\mu = \int f d\mu, \quad f \in L, t \geq 0$.
- (c) $\int Af d\mu = 0, \quad f \in D$.

Proof. (a \Rightarrow b) If X is a solution of the martingale problem for (A, μ) , then by (4.2) and (a),

$$(9.1) \quad E[T(t)f(X(0))] = E[f(X(t))] = E[f(X(0))], \quad f \in L, t \geq 0,$$

which is (b).

(b \Rightarrow a) Let X be a solution of the martingale problem for (A, μ) . By (4.2) and (b),

$$(9.2) \quad E[f(X(t))] = E[T(t)f(X(0))] = E[f(X(0))], \quad f \in L, t \geq 0.$$

Since L is separating, $X(t)$ has distribution μ for each $t \geq 0$, and (a) follows by Lemma 9.1.

(b \Rightarrow c) This is immediate from the definition of A .

(c \Rightarrow b) Since A is the closure of A restricted to D , we may as well take $D = \mathcal{D}(A)$. Then, by (c),

$$(9.3) \quad \begin{aligned} \int (T(t)f - f) d\mu &= \iint_0^t A T(s)f ds d\mu \\ &= \int_0^t \int A T(s)f d\mu ds = 0, \end{aligned}$$

for each $f \in \mathcal{D}(A)$ and $t \geq 0$. Since $\mathcal{D}(A)$ is dense in L , (b) follows. \square

If $\{T(t)\}$ is a semigroup on $B(E)$ given by a transition function and condition (b) of Proposition 9.2 holds (with $L = B(E)$), we say that μ is a *stationary distribution for $\{T(t)\}$* .

An immediate question is that of the existence of a stationary distribution. Compactness of the state space is usually sufficient for existence. This observation is a special case of the following theorem.

9.3 Theorem Suppose A generates a strongly continuous contraction semi-group $\{T(t)\}$ on a closed subspace $L \subset \bar{C}(E)$, L is separating, and the martingale problem for A is well-posed. Let X be a solution of the martingale problem for A , and for some sequence $t_n \rightarrow \infty$, define $\{\mu_n\} \subset \mathcal{P}(E)$ by

$$(9.4) \quad \mu_n(\Gamma) = t_n^{-1} \int_0^{t_n} P\{X(s) \in \Gamma\} ds, \quad \Gamma \in \mathcal{B}(E).$$

If μ is the weak limit of a subsequence of $\{\mu_n\}$, then μ is a stationary distribution for A .

9.4 Remark The theorem avoids the question of existence of a weak limit point for $\{\mu_n\}$. Of course, if E is compact, then $\{\mu_n\}$ is relatively compact by Theorem 2.2 of Chapter 3, and existence follows. \square

Proof. Since the sequence $\{t_n\}$ was arbitrary, we may as well assume $\mu_n \Rightarrow \mu$. For $f \in L$ and $t \geq 0$, $T(t)f \in L \subset \bar{C}(E)$, so

$$\begin{aligned} (9.5) \quad \int T(t)f d\mu &= \lim_{n \rightarrow \infty} \int T(t)f d\mu_n \\ &= \lim_{n \rightarrow \infty} t_n^{-1} \int_0^{t_n} E[T(t)f(X(s))] ds \\ &= \lim_{n \rightarrow \infty} t_n^{-1} \int_0^{t_n} \int T(t+s)f dv ds \\ &= \lim_{n \rightarrow \infty} t_n^{-1} \int_1^{t+t_n} \int T(s)f dv ds \\ &= \lim_{n \rightarrow \infty} t_n^{-1} \int_0^{t_n} \int T(s)f dv ds \\ &= \lim_{n \rightarrow \infty} t_n^{-1} \int_0^{t_n} E[f(X(s))] ds \\ &= \int f d\mu, \end{aligned}$$

where $v = PX(0)^{-1}$, and hence μ is a stationary distribution for A by Proposition 9.2. \square

We now turn to the problem of verifying the relative compactness of $\{\mu_n\}$ in Theorem 9.3. Probably the most useful general approach involves the construction of what is called a Lyapunov function. The following lemmas provide one approach to the construction. For diffusion processes, Itô's formula (Theorem 2.9 of Chapter 5) provides a more direct approach. See also Problem 35.

9.5 Lemma Let $A \subset B(E) \times B(E)$ and let $\varphi, \psi \in M(E)$. Suppose there exist $\{f_k, g_k\} \subset A$ and a constant K such that

$$(9.6) \quad 0 \leq f_k \leq \varphi, \quad k = 1, 2, \dots,$$

$$(9.7) \quad \lim_{k \rightarrow \infty} f_k(x) = \varphi(x), \quad x \in E,$$

$$(9.8) \quad g_k \leq K, \quad k = 1, 2, \dots,$$

and

$$(9.9) \quad \lim_{k \rightarrow \infty} g_k(x) = \psi(x), \quad x \in E.$$

If X is a solution of the martingale problem for A with $E[\varphi(X(0))] < \infty$, then

$$(9.10) \quad \varphi(X(t)) - \int_0^t \psi(X(s)) \, ds$$

is an $\{\mathcal{F}_t^X\}$ -supermartingale.

9.6 Remark Note that we only require that the g_k be bounded above. \square

Proof. Since

$$(9.11) \quad E[f_k(X(t))] = E[f_k(X(0))] + E\left[\int_0^t g_k(X(s)) \, ds\right] \leq E[f_k(X(0))] + Kt,$$

it follows that

$$(9.12) \quad E[\varphi(X(t))] = \lim_{k \rightarrow \infty} E[f_k(X(t))] \leq E[\varphi(X(0))] + Kt$$

and

$$(9.13) \quad \lim_{k \rightarrow \infty} E[|\varphi(X(t)) - f_k(X(t))|] = 0.$$

Since g_k is bounded above uniformly in k , Fatou's lemma implies

$$(9.14) \quad \overline{\lim}_{k \rightarrow \infty} E\left[\int_t^{t+s} g_k(X(u)) \, du \middle| \mathcal{F}_t^X\right] \leq E\left[\int_t^{t+s} \psi(X(u)) \, du \middle| \mathcal{F}_t^X\right]$$

Putting (9.13) and (9.14) together we have

$$(9.15) \quad \begin{aligned} 0 &= \lim_{k \rightarrow \infty} E\left[f_k(X(t+s)) - f_k(X(t)) - \int_t^{t+s} g_k(X(u)) \, du \middle| \mathcal{F}_t^X\right] \\ &\geq E\left[\varphi(X(t+s)) - \varphi(X(t)) - \int_t^{t+s} \psi(X(u)) \, du \middle| \mathcal{F}_t^X\right], \end{aligned}$$

and it follows that (9.10) is a supermartingale. \square

9.7 Lemma Let E be locally compact and separable but not compact, and let $E^\Delta = E \cup \{\Delta\}$ be its one-point compactification. Let $\varphi, \psi \in M(E)$ and let X be a measurable E -valued process. Suppose

$$(9.16) \quad \varphi(X(t)) - \int_0^t \psi(X(s)) \, ds$$

is a supermartingale, $\varphi \geq 0$, $\psi \leq C$ for some constant C , and $\lim_{x \rightarrow \Delta} \psi(x) = -\infty$. Then $\{\mu_t : t \geq 1\} \subset \mathcal{P}(E)$, defined by

$$(9.17) \quad \mu_t(\Gamma) = t^{-1} \int_0^t P\{X(s) \in \Gamma\} \, ds,$$

is relatively compact.

Proof. Given $m \geq 1$, select K_m compact so that

$$(9.18) \quad \sup_{x \notin K_m} \psi(x) \leq -m.$$

Then, for each $t \geq 1$,

$$\begin{aligned} (9.19) \quad 0 &\leq E[\varphi(X(t))] \leq E[\varphi(X(0))] + E\left[\int_0^t \psi(X(s)) \, ds\right] \\ &\leq E[\varphi(X(0))] + CE\left[\int_0^t \chi_{K_m}(X(s)) \, ds\right] \\ &\quad - mE\left[\int_0^t \chi_{K_m^c}(X(s)) \, ds\right]. \\ &= E[\varphi(X(0))] + (C + m) \int_0^t P\{X(s) \in K_m\} \, ds - mt, \end{aligned}$$

and therefore

$$\begin{aligned} (9.20) \quad \frac{m}{C+m} - \frac{t^{-1}}{C+m} E[\varphi(X(0))] &\leq t^{-1} \int_0^t P\{X(s) \in K_m\} \, ds \\ &= \mu_t(K_m). \end{aligned}$$

Consequently, the relative compactness follows by Prohorov's theorem (Theorem 2.2 of Chapter 3). \square

9.8 Corollary Let E be locally compact and separable. Let $A \subset B(E) \times B(E)$ and $\varphi, \psi \in M(E)$. Suppose that $\varphi \geq 0$, that $\psi \leq C$ for some constant C and $\lim_{x \rightarrow \Delta} \psi(x) = -\infty$, and that for every solution X of the martingale problem

for A satisfying $E[\varphi(X(0))] < \infty$, (9.16) is a supermartingale. If X is a stationary solution of the martingale problem for A and $\mu = PX(0)^{-1}$, then

$$(9.21) \quad \mu(K_m) \geq \frac{m}{C + m}, \quad m \geq 1,$$

where $K_m = \{x: \psi(x) \geq -m\}$.

Proof. Let $\alpha > 0$ satisfy $P\{\varphi(X(0)) \leq \alpha\} > 0$, and let Y be a solution of the martingale problem for A with $P\{Y \in B\} = P\{X \in B | \varphi(X(0)) \leq \alpha\}$ (cf. (4.12)). By (9.20)

$$(9.22) \quad \begin{aligned} \frac{m}{C + m} P\{\varphi(X(0)) \leq \alpha\} &\leq \lim_{t \rightarrow \infty} t^{-1} \int_0^t P\{Y(s) \in K_m\} ds P\{\varphi(X(0)) \leq \alpha\} \\ &= \lim_{t \rightarrow \infty} t^{-1} \int_0^t P\{X(s) \in K_m, \varphi(X(0)) \leq \alpha\} ds \\ &\leq \mu(K_m). \end{aligned}$$

Since α is arbitrary, (9.21) follows. \square

If X is a Markov process corresponding to a semigroup $T(t): \bar{C}(E) \rightarrow \bar{C}(E)$, then we can relax the conditions on the Lyapunov function φ and still show existence of a stationary distribution even though we do not get relative compactness for $\{\mu_t: t \geq 1\}$.

9.9 Theorem Let E be locally compact and separable. Let $\{T(t)\}$ be a semigroup on $B(E)$ given by a transition function $P(t, x, \Gamma)$ such that $T(t): \bar{C}(E) \rightarrow \bar{C}(E)$ for all $t \geq 0$, and let X be a measurable Markov process corresponding to $\{T(t)\}$. Let $\varphi, \psi \in M(E)$, $\varphi \geq 0$, $\psi \leq C$ for some constant C , and $\lim_{x \rightarrow \Delta} \psi(x) < 0$, and suppose (9.16) is a supermartingale. Then there is a stationary distribution for $\{T(t)\}$.

Proof. Select $\varepsilon > 0$ and K compact so that $\sup_{x \notin K} \psi(x) \leq -\varepsilon$. Then, as in the proof of Lemma 9.7,

$$(9.23) \quad \mu_t(K) \geq \frac{\varepsilon}{C + \varepsilon} - \frac{t^{-1}}{C + \varepsilon} E[\varphi(X(0))],$$

for all $t \geq 1$, where μ_t is given by (9.17). By Theorem 2.2 of Chapter 3, $\{\mu_t\}$ is relatively compact in $\mathcal{P}(E^\Delta)$. Let $v \in \mathcal{P}(E^\Delta)$ be a weak limit point of $\{\mu_t\}$ as $t \rightarrow \infty$, and let v_E be its restriction to E . It follows as in (9.5) that for non-negative $f \in \hat{C}(E)$,

$$(9.24) \quad \int f d\nu_E = \int f dv = \lim_{n \rightarrow \infty} \int T(t)f d\mu_{t_n} \geq \int T(t)f d\nu_E.$$

Note that if $T(t)f \in \hat{C}(E)$, then we have equality in (9.24). By (9.23), $v_E(E) > 0$ so $\mu = v_E/v_E(E) \in \mathcal{P}(E)$ and

$$(9.25) \quad \int f d\mu \geq \int T(t)f d\mu$$

for all nonnegative $f \in \hat{C}(E)$ and $t \geq 0$. By the Dynkin class theorem (Appendix 4), (9.25) holds for all nonnegative $f \in B(E)$, in particular for indicators, so

$$(9.26) \quad \mu(\Gamma) \geq \int P(t, x, \Gamma) \mu(dx),$$

for all $\Gamma \in \mathcal{B}(E)$ and $t \geq 0$. But both sides of (9.26) are probability measures, so we must in fact have equality in (9.26) and hence in (9.25). \square

The results in Section 8 give conditions under which a sequence of processes converge in some sense to a limiting process. If $\{X_n\}$ is a sequence of stationary processes and $X_n \Rightarrow X$ or, more generally, the finite-dimensional distributions of X_n converge weakly to those of X , then X is stationary. Given this observation, if $\{A_n\}$ is a sequence of generators determining Markov processes (i.e., the martingale problem for A_n is well-posed) and if, for each n , μ_n is a stationary distribution for A_n , then, under the hypotheses of one of the convergence theorems of Section 8, one would expect that the sequence $\{\mu_n\}$ would converge weakly to a stationary distribution for the limiting generator A . This need not be the case in general, but the following theorem is frequently applicable.

9.10 Theorem Let (E, r) be complete and separable. Let $\{T_n(t)\}$, $\{T(t)\}$ be contraction semigroups corresponding to Markov processes in E , and suppose that for each n , $\mu_n \in \mathcal{P}(E)$ is a stationary distribution for $\{T_n(t)\}$. Let $L \subset \hat{C}(E)$ be separating and $T(t): L \rightarrow L$ for all $t \geq 0$. Suppose that for each $f \in L$ and compact $K \subset E$,

$$(9.27) \quad \lim_{n \rightarrow \infty} \sup_{x \in K} |T_n(t)f(x) - T(t)f(x)| = 0, \quad t \geq 0.$$

Then every weak limit point of $\{\mu_n\}$ is a stationary distribution for $\{T(t)\}$.

9.11 Remark Note that if $x_n \rightarrow x$ implies $T_n(t)f(x_n) \rightarrow T(t)f(x)$ for all $t \geq 0$, then (9.27) holds. \square

Proof. For simplicity, assume $\mu_n \Rightarrow \mu$. Then for each $f \in L$, $t \geq 0$, and compact $K \subset E$,

$$\begin{aligned}
 (9.28) \quad & \left| \int T(t)f d\mu - \int f d\mu \right| \\
 & \leq \overline{\lim}_{n \rightarrow \infty} \left| \int T(t)f d\mu - \int T(t)f d\mu_n \right| \\
 & \quad + \overline{\lim}_{n \rightarrow \infty} \left| \int (T(t)f - T_n(t)f) d\mu_n \right| \\
 & \quad + \overline{\lim}_{n \rightarrow \infty} \left| \int f d\mu_n - \int f d\mu \right| \\
 & \leq \overline{\lim}_{n \rightarrow \infty} 2\|f\|\mu_n(K^c).
 \end{aligned}$$

But by Prohorov's theorem (Theorem 2.2 of Chapter 3), for every $\varepsilon > 0$ there is a compact $K \subset E$ such that $\mu_n(K^c) < \varepsilon$ for all n . Consequently, the left side of (9.28) is zero. \square

Theorem 9.10 can be generalized considerably.

9.12 Theorem Let (E_n, r_n) , $n = 1, 2, \dots$, and (E, r) be complete, separable metric spaces, let $A_n \subset B(E_n) \times B(E_n)$ and $A \subset B(E) \times B(E)$, and suppose that the martingale problems for the A_n and A are well-posed. For $v_n \in \mathcal{P}(E_n)$ (respectively, $v \in \mathcal{P}(E)$), let $X_n^{v_n}$ (respectively, X^v) denote a solution of the martingale problem for (A_n, v_n) (respectively, (A, v)). Let $\eta_n : E_n \rightarrow E$ be Borel measurable, and suppose that for each choice of $v_n \in \mathcal{P}(E_n)$, $n = 1, 2, \dots$, and any subsequence of $\{v_n \eta_n^{-1}\}$ that converges weakly to $v \in \mathcal{P}(E)$, the finite-dimensional distributions of the corresponding subsequence of $\{\eta_n \circ X_n^{v_n}\}$ converge weakly to those of X^v . If for $n = 1, 2, \dots$, μ_n is a stationary distribution for A_n , then any weak limit point of $\{\mu_n \eta_n^{-1}\}$ is a stationary distribution for A .

Proof. If a subsequence of $\{\mu_n \eta_n^{-1}\}$ converges weakly to μ , then the finite dimensional distributions of the corresponding subsequence of $\{\eta_n \circ X_n^{v_n}\}$ converge weakly to those of X^v . But $\eta_n \circ X_n^{v_n}$ is a stationary process so X^v must also be stationary. \square

The convergence of the finite-dimensional distributions required in the above theorem can be proved using the results of Section 8. The hidden difficulty is that there is no guarantee that $\{\mu_n \eta_n^{-1}\}$ has any convergent subsequences; thus, we need conditions under which $\{\mu_n \eta_n^{-1}\}$ is relatively compact. Corollary 9.8 suggests one set of conditions.

9.13 Lemma Let E_n, E, A_n, A , and η_n be as in Theorem 9.12, and in addition assume that E is locally compact. Let $\varphi_n, \psi_n \in M(E_n)$ and $\psi \in M(E)$. Suppose that $\varphi_n \geq 0, \psi_n \leq \psi \circ \eta_n, \psi \leq C$ for some constant C , that $\lim_{x \rightarrow \Delta} \psi(x) = -\infty$, and that for every solution X_n of the martingale problem for A_n with $E[\varphi_n(X_n(0))] < \infty$,

$$(9.29) \quad \varphi_n(X_n(t)) - \int_0^t \psi_n(X_n(s)) ds$$

is a supermartingale. For $n = 1, 2, \dots$, let μ_n be a stationary distribution for A_n . Then, for each $m, n \geq 1$,

$$(9.30) \quad \mu_n \eta_n^{-1}(K_m) \geq \frac{m}{m + C},$$

where $K_m = \{x: \psi(x) \geq -m\}$, and hence $\{\mu_n \eta_n^{-1}\}$ is relatively compact.

Proof. Let X_n be a solution of the martingale problem for A_n with $E[\varphi_n(X_n(0))] < \infty$. Then, as in (9.19),

$$\begin{aligned} (9.31) \quad 0 &\leq E[\varphi_n(X_n(0))] + E\left[\int_0^t \psi_n(X_n(s)) ds\right] \\ &\leq E[\varphi_n(X_n(0))] + E\left[\int_0^t \psi(\eta_n \circ X_n(s)) ds\right] \\ &\leq E[\varphi_n(X_n(0))] + CE\left[\int_0^t \chi_{K_m}(\eta_n \circ X_n(s)) ds\right] \\ &\quad - mE\left[\int_0^t \chi_{K_m}(\eta_n \circ X_n(s)) ds\right], \end{aligned}$$

and the estimate follows by the same argument used in the proof of Corollary 9.8. \square

Analogues of Theorem 9.12 and Lemma 9.13 hold for sequences of discrete-parameter processes. See Problems 46 and 47.

We give one additional theorem that, while it is not stated in terms of convergence of stationary distributions, typically implies this.

9.14 Theorem Let $\{T_n(t)\}$, $n = 1, 2, \dots$, and $\{T(t)\}$ be strongly continuous semigroups on a closed subspace $L \subset \bar{C}(E)$ corresponding to transition functions $P_n(t, x, \Gamma)$, $n = 1, 2, \dots$, and $P(t, x, \Gamma)$. Suppose for each compact $K_1 \subset E$ and $\varepsilon > 0$ there exists a compact $K_2 \subset E$ such that

$$(9.32) \quad \inf_{x \in K_1} \inf_n \inf_{t \in [0, 1]} P_n(t, x, K_2) \geq 1 - \varepsilon.$$

Suppose that for each $f \in L$, $t_0 > 0$, and compact $K \subset E$,

$$(9.33) \quad \lim_{n \rightarrow \infty} \sup_{x \in K} \sup_{t \leq t_0} |T_n(t)f(x) - T(t)f(x)| = 0,$$

and suppose that there exists an operator $\pi: L \rightarrow L$ such that for each $f \in L$ and compact $K \subset E$,

$$(9.34) \quad \lim_{t \rightarrow \infty} \sup_{x \in K} |T(t)f(x) - \pi f(x)| = 0$$

and

$$(9.35) \quad \lim_{n \rightarrow \infty} \sup_{x \in K} \sup_{0 \leq t < \infty} |T_n(t)\pi f(x) - \pi f(x)| = 0.$$

Then for each $f \in L$ and compact $K \subset E$,

$$(9.36) \quad \lim_{n \rightarrow \infty} \sup_{x \in K} \sup_{0 \leq t < \infty} |T_n(t)f(x) - T(t)f(x)| = 0.$$

9.15 Remark Note that frequently $\pi f(x)$ is a constant independent of x , that is, $\pi f(x) = \int f d\mu$ where μ is the unique stationary distribution for $\{T(t)\}$. This result can be generalized in various ways, for example, to discrete time or to semigroups on different spaces. The proofs are straightforward. \square

Proof. For $t > t_0$,

$$\begin{aligned} (9.37) \quad & T_n(t)f - T(t)f \\ &= T_n(t - t_0)(T_n(t_0)f - T(t_0)f) \\ &\quad + T_n(t - t_0)(T(t_0)f - \pi f) \\ &\quad + T_n(t - t_0)\pi f - \pi f \\ &\quad + \pi f - T(t)f. \end{aligned}$$

Each term on the right can be made small uniformly on compact sets by first taking t_0 sufficiently large and then letting $n \rightarrow \infty$. The details are left to the reader. \square

We now reconsider condition (c) of Proposition 9.2. Note that D is required to be a core for the generator of the semigroup. Consequently, if one only knows that the martingale problem for A is well-posed and not that the closure of A generates a semigroup, then Proposition 9.2 is not applicable. To see that this is more than a technical difficulty, see Problem 40. The next theorem gives conditions under which condition (c) of Proposition 9.2 implies μ is a stationary distribution without requiring that A (or the closure of A) generate a semigroup.

We need the following lemma.

9.16 Lemma Let $A \subset \bar{C}(E) \times \bar{C}(E)$. Suppose for each $v \in \mathcal{P}(E)$ that the martingale problem for (A, v) has a solution with sample paths in $D_E[0, \infty)$. Suppose that φ is continuously differentiable and convex on $G \subset \mathbb{R}^m$, that $(f_1, g_1), \dots, (f_m, g_m) \in A$, $(f_1, \dots, f_m) : E \rightarrow G$, and that $(\varphi(f_1, f_2, \dots, f_m), h) \in A$. Then

$$(9.38) \quad h \geq \nabla \varphi(f_1, f_2, \dots, f_m) \cdot (g_1, g_2, \dots, g_m).$$

Proof. Let $x \in E$ and let X be a solution of the martingale problem for (A, δ_x) . Then by convexity,

$$\begin{aligned} (9.39) \quad E \left[\int_0^t h(X(s)) \, ds \right] &= E[\varphi(f_1(X(t)), \dots, f_m(X(t)))] - \varphi(f_1(x), \dots, f_m(x)) \\ &\geq \nabla \varphi(f_1(x), \dots, f_m(x)) \cdot E[(f_1(X(t)) - f_1(x), \dots, f_m(X(t)) - f_m(x))] \\ &= \nabla \varphi(f_1(x), \dots, f_m(x)) \cdot E \left[\left(\int_0^t g_1(X(s)) \, ds, \dots, \int_0^t g_m(X(s)) \, ds \right) \right], \end{aligned}$$

for all $t > 0$. Dividing by t and letting $t \rightarrow 0$ gives (9.38). \square

9.17 Theorem Let E be locally compact and separable, and let A be a linear operator on $\bar{C}(E)$ satisfying the positive maximum principle such that $\mathcal{D}(A)$ is an algebra and dense in $\bar{C}(E)$. If $\mu \in \mathcal{P}(E)$ satisfies

$$(9.40) \quad \int Af \, d\mu = 0, \quad f \in \mathcal{D}(A),$$

then there exists a stationary solution of the martingale problem for (A, μ) .

Proof. Without loss of generality we may assume E is compact and $(1, 0) \in A$. If not, construct A^Δ as in Theorem 5.4 and extend μ to $\mathcal{B}(E^\Delta)$ by setting $\mu(\{\Delta\}) = 0$. Then A^Δ and μ satisfy the hypotheses of the theorem. If X is a stationary solution of the martingale problem for (A^Δ, μ) , then $P\{X(t) \in E\} = \mu(E) = 1$, and hence X has a modification taking values in E , which is then a stationary solution of the martingale problem for (A, μ) .

Existence of a solution of the martingale problem for (A, μ) is assured by Theorem 5.4, but there is no guarantee that the solution constructed there will be stationary. For $n = 1, 2, \dots$, let

$$(9.41) \quad A_n = \{(f, n[(I - n^{-1}A)^{-1}f - f]): f \in \mathcal{D}(I - n^{-1}A)\}.$$

For $f \in \mathcal{D}(A)$ and $f_n = (I - n^{-1}A)f$, we see that $\|f_n - f\| \rightarrow 0$ and $\|A_n f_n - Af\| \rightarrow 0$ (in fact $A_n f_n = Af$). It follows from Theorem 9.1 of Chapter 3 that if X_n is a solution of the martingale problem for A_n , $n = 1, 2, 3, \dots$, with sample paths in $D_E[0, \infty)$, then $\{X_n\}$ is relatively compact.

By Lemma 5.1, any limit point of $\{X_n\}$ is a solution of the martingale problem for A . Consequently, to complete the proof of the theorem it suffices

to construct, for each $n \geq 1$, a stationary solution of the $D_E[0, \infty)$ martingale problem for (A_n, μ) . Note that for $f \in \mathcal{D}(A_n)$, $f = (I - n^{-1}A)g$ for some $g \in \mathcal{D}(A)$ and

$$(9.42) \quad \int A_n f d\mu = \int Ag d\mu = 0.$$

The key step of the proof is the construction of a measure $v \in \mathcal{P}(E \times E)$ such that

$$(9.43) \quad v(\Gamma \times E) = v(E \times \Gamma) = \mu(\Gamma), \quad \Gamma \in \mathcal{B}(E),$$

and

$$(9.44) \quad \int h(x)g(y)v(dx \times dy) = \int h(x)(I - n^{-1}A)^{-1}g(x)\mu(dx)$$

for all $h \in C(E)$ and $g \in \mathcal{R}(I - n^{-1}A)$. Let $M \subset C(E \times E)$ be the linear space of functions of the form

$$(9.45) \quad F(x, y) = \sum_{i=1}^m h_i(x)g_i(y) + f(y)$$

for $h_1, \dots, h_m, f \in C(E)$, and $g_1, \dots, g_m \in \mathcal{R}(I - n^{-1}A)$, and define a linear functional Λ on M by

$$(9.46) \quad \Lambda F = \int \left[\sum_{i=1}^m h_i(y)(I - n^{-1}A)^{-1}g_i(y) + f(y) \right] \mu(dy).$$

Since $\Lambda 1 = 1$, the Hahn-Banach theorem and the Riesz representation theorem give the existence of a measure $v \in \mathcal{P}(E \times E)$ such that $\Lambda F = \int F dv$ for all $F \in M$ (and hence (9.43) and (9.44)) if we show that $|\Lambda F| \leq \|F\|$. This inequality also implies that if $F \geq 0$, then $\|F\| - \Lambda F = \Lambda(\|F\| - F) \leq \|F\| - F \leq \|F\|$, so $\Lambda F \geq 0$.

Let $f_1, f_2, \dots, f_m \in \mathcal{D}(A)$, let $\alpha_k = \|f_k - n^{-1}Af_k\|$, and let φ be a polynomial on \mathbb{R}^m that is convex on $[-\alpha_1, \alpha_1] \times [-\alpha_2, \alpha_2] \times \dots \times [-\alpha_m, \alpha_m]$. Since $\mathcal{D}(A)$ is an algebra, $\varphi(f_1, \dots, f_m) \in \mathcal{D}(A)$, and by Lemma 9.16,

$$(9.47) \quad A\varphi(f_1, \dots, f_m) \geq \nabla\varphi(f_1, \dots, f_m) \cdot (Af_1, \dots, Af_m).$$

Consequently,

$$\begin{aligned} (9.48) \quad & \varphi((I - n^{-1}A)f_1, \dots, (I - n^{-1}A)f_m) \\ & \geq \varphi(f_1, \dots, f_m) - \frac{1}{n} \nabla\varphi(f_1, \dots, f_m) \cdot (Af_1, \dots, Af_m) \\ & \geq \varphi(f_1, \dots, f_m) - \frac{1}{n} A\varphi(f_1, \dots, f_m), \end{aligned}$$

and hence

$$(9.49) \quad \int \varphi((I - n^{-1}A)f_1, \dots, (I - n^{-1}A)f_m) d\mu \geq \int \varphi(f_1, \dots, f_m) d\mu,$$

or equivalently

$$(9.50) \quad \int \varphi(g_1, \dots, g_m) d\mu \geq \int \varphi((I - n^{-1}A)^{-1}g_1, \dots, (I - n^{-1}A)^{-1}g_m) d\mu$$

for $g_1, \dots, g_m \in \mathcal{D}(I - n^{-1}A)$. Since all convex functions on \mathbb{R}^m can be approximated uniformly on any compact set $K \subset \mathbb{R}^m$ by a polynomial that is convex on K , (9.50) holds for all φ convex on \mathbb{R}^m .

Let F be given by (9.45), and define $\varphi: \mathbb{R}^m \rightarrow \mathbb{R}$ by

$$(9.51) \quad \varphi(u) = \sup_x \sum_{i=1}^m h_i(x)u_i.$$

Note that φ is convex. Then

$$\begin{aligned} (9.52) \quad \Lambda F &= \int \sum_{i=1}^m h_i(y) \left(I - \frac{1}{n} A \right)^{-1} g_i(y) \mu(dy) + \int f(y) \mu(dy) \\ &\leq \int \varphi((I - n^{-1}A)^{-1}g_1, \dots, (I - n^{-1}A)^{-1}g_m) d\mu + \int f d\mu \\ &\leq \int \varphi(g_1, \dots, g_m) d\mu + \int f d\mu \\ &= \int \sup_x \left[\sum_{i=1}^m h_i(x)g_i(y) + f(y) \right] d\mu \\ &\leq \|F\|. \end{aligned}$$

Similarly $-\Lambda F = \Lambda(-F) \leq \| -F \| = \|F\|$, and the existence of the desired v follows.

There exists a transition function $\eta(x, \Gamma)$ such that

$$(9.53) \quad v(A \times B) = \int_A \eta(x, B) \mu(dx), \quad A, B \in \mathcal{B}(E),$$

(see Appendix 8), and hence

$$(9.54) \quad \int_E \eta(x, B) \mu(dx) = v(E \times B) = \mu(B), \quad B \in \mathcal{B}(E).$$

Let $Y(0), Y(1), Y(2), \dots$ be a Markov chain with transition function η and initial distribution μ . By (9.54), $\{Y(k)\}$ is stationary. Since (9.44) holds for all $h \in C(E)$ and $g \in \mathcal{R}(I - n^{-1}A)$, it follows that

$$(9.55) \quad \int g(y)\eta(x, dy) = (I - n^{-1}A)^{-1}g(x) \quad \mu\text{-a.s.}$$

for all $g \in \mathcal{R}(I - n^{-1}A)$. Therefore

$$(9.56) \quad g(Y_k) - \sum_{i=0}^{k-1} n^{-1} A_n g(Y_i)$$

is a martingale with respect to $\{\mathcal{F}_k^Y\}$. Let V be a Poisson process with parameter n and define $X = Y(V(\cdot))$. Then

$$(9.57) \quad g(X(t)) - \int_0^t A_n g(X(s)) ds$$

is an $\{\mathcal{F}_t^X\}$ -martingale for each $g \in \mathcal{R}(I - n^{-1}A)$ (cf. (2.6)). We leave it to the reader (Problem 41) to show that X is stationary. \square

Proposition 9.2 and **Theorem 9.17** are special cases of more-general results. Let $A \subset B(E) \times B(E)$. If X is a solution of the martingale problem for A and v , is the distribution of $X(t)$, then $\{v_t\}$ satisfies

$$(9.58) \quad v_t f + v_0 f + \int_0^t v_s g ds, \quad (f, g) \in A,$$

where $v_t f = \int f d\nu_t$. Of course (9.40) is a special case of (9.58) with $v_t = \mu$ for all $t \geq 0$. We are interested in conditions under which, given v_0 , (9.58) determines v_t for all $t \geq 0$. The first result gives a generalization of Proposition 9.2 (c \Rightarrow a).

9.18 Proposition Suppose $\mathcal{R}(\lambda - A)$ is separating for each $\lambda > 0$. If $\{v_t\}$ and $\{\mu_t\}$ satisfy (9.58), are weakly right continuous, and $v_0 = \mu_0$, then $v_t = \mu_t$ for all $t \geq 0$.

Proof. By (9.58), for $(f, g) \in A$,

$$\begin{aligned} (9.59) \quad \lambda \int_0^\infty e^{-\lambda t} v_t f dt &= v_0 f + \lambda \int_0^\infty e^{-\lambda t} \int_0^t v_s g ds dt \\ &= v_0 f + \lambda \int_0^\infty \int_s^\infty e^{-\lambda t} dt v_s g ds \\ &= v_0 f + \int_0^\infty e^{-\lambda s} v_s g ds. \end{aligned}$$

Consequently,

$$(9.60) \quad \int_0^\infty e^{-\lambda t} v_t(\lambda f - g) dt = v_0 f, \quad (f, g) \in A.$$

Since $\mathcal{D}(\lambda - A)$ is separating, (9.60) implies that v_0 uniquely determines the measure $\int_0^\infty e^{-\lambda t} v_t dt$. Since this holds for each $\lambda > 0$ and $\{v_t\}$ is weakly right continuous, the uniqueness of the Laplace transform implies v_0 determines v_t , $t \geq 0$. \square

We next consider the generalization of Theorem 9.17

9.19 Proposition Let E be locally compact and separable, and let A be a linear operator on $\hat{C}(E)$ satisfying the positive maximum principle such that $\mathcal{D}(A)$ is an algebra and dense in $\hat{C}(E)$. Suppose the martingale problem for A is well-posed. If $\{v_t\} \subset \mathcal{P}(E)$ and $\{\mu_t\} \subset \mathcal{P}(E)$ satisfy (9.58) and $v_0 = \mu_0$, then $v_t = \mu_t$ for all $t \geq 0$.

Proof. Since $\mathcal{D}(A)$ is dense in $\hat{C}(E)$, weak continuity of $\{v_t\}$ and $\{\mu_t\}$ follows from (9.58). We reduce the proof to the case considered in Theorem 9.17. As in the proof of Theorem 9.17, without loss of generality we can assume that E is compact. Let $E_0 = E \times \{-1, 1\}$. Fix $\lambda > 0$. For $f_1 \in \mathcal{D}(A)$ and $f_2 \in B(\{-1, 1\})$, let $f = f_1 f_2$ and define

$$(9.61) \quad Bf(x, n) = f_2(n) Af_1(x) + \lambda \left(f_2(-n) \int f dv_0 - f_1(x) f_2(n) \right).$$

By Theorem 10.3 of the next section, if the martingale problem for A is well-posed, then the martingale problem for B is well-posed. There the new component is a counting process, but essentially the same proof will give uniqueness here.

Define

$$(9.62) \quad \mu = \left(\lambda \int_0^\infty e^{-\lambda t} v_t dt \right) \times \left(\frac{1}{2} \delta_1 + \frac{1}{2} \delta_{-1} \right)$$

Then μ satisfies $\int Bf d\mu = 0$ for all $f \in \mathcal{D}(B)$, and, since the linear extension of B satisfies the conditions of Theorem 9.17, μ is a stationary distribution for B .

We claim there is only one stationary distribution for B . To see this, we observe that any solution of the $D_{E_0}[0, \infty)$ martingale problem for B is a strong Markov process (Theorem 4.2 and Corollary 4.3). Let $\{\eta_t\}$ be the one-dimensional distributions for the solution of the $D_{E_0}[0, \infty)$ martingale problem for $(B, v_0 \times \delta_1)$. Let (Z, N) be any solution of the $D_{E_0}[0, \infty)$ martingale problem for B , and define $\tau_0 = \inf\{t > 0 : N(t) = -1\}$ and $\tau = \inf\{t > \tau_0 : N(t) = 1\}$. Then

$$(9.63) \quad P\{(Z(t), N(t)) \in \Gamma\} = P\{(Z(t), N(t)) \in \Gamma, t < \tau\} + E[\eta_{t-\tau}(\Gamma) \chi_{\{t \leq \tau\}}].$$

Consequently

$$(9.64) \quad \lim_{t \rightarrow \infty} t^{-1} \int_0^t P\{(Z(s), N(s)) \in \Gamma\} ds = \lim_{t \rightarrow \infty} t^{-1} \int_0^t E[\eta_{s-t}(\Gamma) \chi_{\{t < s\}}] ds \\ = \lim_{t \rightarrow \infty} t^{-1} \int_0^t \eta_s(\Gamma) ds,$$

and uniqueness of the stationary distribution follows.

If $\tilde{\mu}$ is defined by (9.62) with $\{v_t\}$ replaced by $\{\mu_t\}$, it is a stationary distribution for B and uniqueness gives

$$(9.65) \quad \int_0^\infty e^{-\lambda t} v_t dt = \int_0^\infty e^{-\lambda t} \mu_t dt.$$

Since $\lambda > 0$ is arbitrary and $\{v_t\}$ and $\{\mu_t\}$ are weakly continuous, it follows that $v_t = \mu_t$ for all $t \geq 0$. \square

10. PERTURBATION RESULTS

Suppose that X_1 is a solution of the martingale problem for $A_1 \subset B(E_1) \times B(E_1)$ and that X_2 is a solution of the martingale problem for $A_2 \subset B(E_2) \times B(E_2)$. If X_1 and X_2 are independent, then (X_1, X_2) is a solution of the martingale problem for $A \subset B(E_1 \times E_2) \times B(E_1 \times E_2)$ given by

$$(10.1) \quad A = \{(f_1, f_2, g_1 f_2 + f_1 g_2) : (f_1, g_1) \in A_1, (f_2, g_2) \in A_2\}.$$

If uniqueness holds for A_1 and A_2 , and if $(1, 0) \in A_i$, $i = 1, 2$, then we also have uniqueness for A .

10.1 Theorem Let (E_1, r_1) , (E_2, r_2) be complete, separable metric spaces. For $i = 1, 2$, let $A_i \subset B(E_i) \times B(E_i)$, $(1, 0) \in A_i$, and suppose that uniqueness holds for the martingale problem for A_i . Then uniqueness holds for the martingale problem for A given by (10.1). In particular, if $X = (X_1, X_2)$ is a solution of the martingale problem for A and $X_1(0)$ and $X_2(0)$ are independent, then X_1 and X_2 are independent.

Proof. Note that $X = (X_1, X_2)$ is a solution of the martingale problem for A if and only if it is a solution for

$$(10.2) \quad \tilde{A} = \{((f_1 + \|f_1\| + 1)f_2 + \|f_2\| + 1),$$

$$g_1(f_2 + \|f_2\| + 1) + (f_1 + \|f_1\| + 1)g_2) : (f_i, g_i) \in A_i, i = 1, 2\},$$

so we may as well assume $f_i \geq 1$ for all $(f_i, g_i) \in A_i$, $i = 1, 2$.

For each $v \in \mathcal{P}(E_i)$ for which a solution Y of the martingale problem for (A_i, v) exists, define $\eta_i(v, t, \Gamma) = P\{Y(t) \in \Gamma\}$, $\Gamma \in \mathcal{B}(E_i)$. By uniqueness, η_i is well-defined.

By Lemma 3.2 and by Problem 23 of Chapter 2, Y is a solution of the martingale problem for A_i if and only if for every bounded, discrete $\{\mathcal{F}_t^Y\}$ -stopping time τ ,

$$(10.3) \quad E\left[f(Y(\tau)) \exp\left\{-\int_0^\tau \frac{g(Y(s))}{f(Y(s))} ds\right\}\right] = E[f(Y(0))]$$

for all $(f, g) \in A_i$.

Let $X = (X_1, X_2)$ be a solution of the martingale problem for A with respect to $\{\mathcal{G}_t\}$ defined on (Ω, \mathcal{F}, P) . For $\Gamma_2 \in \mathcal{B}(E_2)$ with $P\{X_2(0) \in \Gamma_2\} > 0$, define

$$(10.4) \quad Q(B) = \frac{E[\chi_B \chi_{\Gamma_2}(X_2(0))]}{E[\chi_{\Gamma_2}(X_2(0))]}, \quad B \in \mathcal{F}.$$

Then X_1 on (Ω, \mathcal{F}, Q) is a solution of the martingale problem for A_1 , and hence

$$(10.5) \quad E[\chi_{\Gamma_1}(X_1(t)) \chi_{\Gamma_2}(X_2(0))] = \eta_1(v_{\Gamma_2}, t, \Gamma_1) P\{X_2(0) \in \Gamma_2\}$$

where $v_{\Gamma_2}(\Gamma_1) = P\{X_1(0) \in \Gamma_1 \mid X_2(0) \in \Gamma_2\}$.

For $(f_i, g_i) \in A_i$, $i = 1, 2$, define

$$(10.6) \quad M_i(t) = f_i(X_i(t)) \exp\left\{-\int_0^t \frac{g_i(X_i(s))}{f_i(X_i(s))} ds\right\}.$$

Note M_1 , M_2 , and $M_1 M_2$ are $\{\mathcal{G}_t\}$ -martingales. If τ_1 and τ_2 are bounded, discrete $\{\mathcal{G}_t\}$ -stopping times, then

$$\begin{aligned} (10.7) \quad E[M_1(\tau_1)M_2(\tau_2)] &= E[E[M_1(\tau_1)M_2(\tau_2) \mid \mathcal{F}_{\tau_1}]] = E[M_1(\tau_1)M_2(\tau_2 \wedge \tau_1)] \\ &= E[M_1(\tau_2 \wedge \tau_1)M_2(\tau_2 \wedge \tau_1)] \\ &= E[M_1(0)M_2(0)]. \end{aligned}$$

Fix τ_2 and $(f_2, g_2) \in A_2$, and for M_2 as above, define

$$(10.8) \quad \tilde{Q}(B) = \frac{E[\chi_B M_2(\tau_2)]}{E[M_2(\tau_2)]}, \quad B \in \mathcal{F}.$$

For $(f, g) \in A_1$ and any discrete $\{\mathcal{G}_t\}$ -stopping time τ , (10.7) implies

$$\begin{aligned} (10.9) \quad E^{\tilde{Q}}\left[f(X_1(\tau)) \exp\left\{-\int_0^\tau \frac{g(X_1(s))}{f(X_1(s))} ds\right\}\right] \\ &= \frac{E[f(X_1(0))M_2(0)]}{E[M_2(\tau_2)]} \\ &= \frac{E[f(X_1(0))M_2(\tau_2)]}{E[M_2(\tau_2)]} \\ &= E^{\tilde{Q}}[f(X_1(0))]. \end{aligned}$$

Consequently X_1 on $(\Omega, \mathcal{F}, \tilde{Q})$ is a solution of the martingale problem for A_1 , and uniqueness implies that $\tilde{Q}\{X_1(t) \in \Gamma_1\} = \eta_1(v, t, \Gamma_1)$ where $v(\Gamma) = E[\chi_{\Gamma}(X_1(0))f_2(X_2(0))] / E[f_2(X_2(0))]$.

Note that v does not depend on τ_2 , so

$$(10.10) \quad \begin{aligned} E\left[\chi_{\Gamma_1}(X_1(t))f_2(X_2(\tau_2)) \exp\left\{-\int_0^{\tau_2} \frac{g_2(X_2(s))}{f_2(X_2(s))} ds\right\}\right] \\ = \eta_1(v, t, \Gamma_1)E[f_2(X_2(0))] \\ = E[\chi_{\Gamma_1}(X_1(t))f_2(X_2(0))]. \end{aligned}$$

Next, defining

$$(10.11) \quad \hat{Q}(B) = \frac{E[\chi_{\Gamma_1}(X_1(t))\chi_B]}{E[\chi_{\Gamma_1}(X_1(t))]}, \quad B \in \mathcal{F},$$

(10.10) implies $\hat{Q}\{X_2(t) \in \Gamma_2\} = \eta_2(v_2, t, \Gamma_2)$, where

$$(10.12) \quad \begin{aligned} v_2(\Gamma_2) &= \frac{E[\chi_{\Gamma_1}(X_1(t))\chi_{\Gamma_2}(X_2(0))]}{E[\chi_{\Gamma_1}(X_1(t))]} \\ &= \frac{\eta_1(v_{\Gamma_1}, t, \Gamma_1)P\{X_2(0) \in \Gamma_2\}}{P\{X_1(t) \in \Gamma_1\}}, \quad \Gamma_2 \in \mathcal{B}(E_2). \end{aligned}$$

Consequently,

$$(10.13) \quad P\{X_1(t) \in \Gamma_1, X_2(t) \in \Gamma_2\} = \eta_2(v_2, t, \Gamma_2)P\{X_1(t) \in \Gamma_1\}.$$

Since by uniqueness the distribution of $X_1(t)$ is determined by the distribution of $X_1(0)$, v_2 is uniquely determined by the distribution of $(X_1(0), X_2(0))$. Consequently, the right side of (10.13) is uniquely determined by the distribution of $(X_1(0), X_2(0))$. The theorem now follows by Theorem 4.2. \square

Let $\lambda \in B(E)$ be nonnegative and let $\mu(x, \Gamma)$ be a transition function on $E \times \mathcal{B}(E)$. Define B on $B(E)$ by

$$(10.14) \quad Bf(x) = \lambda(x) \int (f(y) - f(x))\mu(x, dy).$$

Let $A \subset B(E) \times B(E)$ be linear and dissipative. If for some $\lambda > 0$, $B(E)$ is the bp-closure of $\mathcal{R}(\lambda - A)$, then $B(E)$ is the bp-closure of $\mathcal{R}(\lambda - (A + B))$ where $A + B = \{(f, g + Bf) : (f, g) \in A\}$. Consequently, Theorem 4.1 and Corollary 4.4 give uniqueness results for $A + B$. Also see Problem 3.

We now want to give existence and uniqueness results without assuming the range condition.

10.2 Proposition Let (E, r) be complete and separable, let $A \subset B(E) \times B(E)$, and let B be given as in (10.14). Suppose that for every $v \in \mathcal{P}(E)$ there exists a solution of the $D_E[0, \infty)$ martingale problem for (A, v) . Then for every $v \in \mathcal{P}(E)$ there exists a solution of the $D_E[0, \infty)$ martingale problem for $(A + B, v)$.

Proof. By the construction in (2.4) we may assume λ is constant.

Let $\Omega = \prod_{k=1}^{\infty} (D_E[0, \infty) \times [0, \infty))$. Let (X_k, Δ_k) denote the coordinate random variables. Define $\mathcal{G}_k = \sigma(X_l, \Delta_l : l \leq k)$ and $\mathcal{G}^k = \sigma(X_l, \Delta_l : l \geq k)$. By an argument similar to the proof of Lemma 5.15, there is a probability distribution on Ω such that for each k , X_k is a solution of the martingale problem for A , Δ_k is independent of $\sigma(X_1, \dots, X_k, \Delta_1, \dots, \Delta_{k-1})$ and exponentially distributed with parameter λ , and for $A_1 \in \mathcal{G}_k$ and $A_2 \in \mathcal{G}^{k+1}$,

$$(10.15) \quad P(A_1 \cap A_2) = E \left[\int P(A_2 | X_{k+1}(0) = x) \mu(X_k(\Delta_k), dx) \chi_{A_1} \right].$$

Define $\tau_0 = 0$, $\tau_k = \sum_{l=1}^k \Delta_l$, and $N(t) = k$ for $\tau_k \leq t < \tau_{k+1}$. Note that N is a Poisson process with parameter λ . Define

$$(10.16) \quad X(t) = X_{k+1}(t - \tau_k), \quad \tau_k \leq t < \tau_{k+1},$$

and $\mathcal{F}_t = \mathcal{F}_t^X \vee \mathcal{F}_t^N$. We claim that X is a solution of the martingale problem for $A + B$ with respect to $\{\mathcal{F}_t\}$. First note that for $(f, g) \in A$,

$$(10.17)$$

$$f(X_{k+1}((t \vee \tau_k) \wedge \tau_{k+1} - \tau_k)) - f(X_{k+1}(0)) - \int_{\tau_k}^{(t \vee \tau_k) \wedge \tau_{k+1}} g(X_{k+1}(s - \tau_k)) ds$$

is an $\{\mathcal{F}_t\}$ -martingale. This follows from the fact that

$$(10.18) \quad E \left[\left(f(X_{k+1}((t_{m+1} \vee \tau_k) \wedge \tau_{k+1} - \tau_k)) - f(X_{k+1}((t_m \vee \tau_k) \wedge \tau_{k+1} - \tau_k)) \right. \right. \\ \left. \left. - \int_{(t_m \vee \tau_k) \wedge \tau_{k+1}}^{(t_{m+1} \vee \tau_k) \wedge \tau_{k+1}} g(X_{k+1}(s - \tau_k)) ds \right) \right. \\ \left. \times \prod_{i=1}^m h_i(X_{k+1}(t_i \vee \tau_k - \tau_k)) \mid \mathcal{G}_k \vee \sigma(\Delta_{k+1}) \right] = 0,$$

which in turn follows from (10.15) and the fact that Δ_{k+1} is independent of X_{k+1} . (See Appendix 4.)

Consequently, summing (10.17) over k ,

$$(10.19) \quad f(X(t)) - f(X(0)) - \int_0^t g(X(s)) ds - \sum_{k=1}^{N(t)} (f(X_{k+1}(0)) - f(X_k(\Delta_k)))$$

is an $\{\mathcal{F}_t\}$ -martingale, as are

$$(10.20) \quad \sum_{k=1}^{N(t)} \left(f(X_{k+1}(0)) - \int f(y)\mu(X_k(\Delta_k), dy) \right)$$

and

$$(10.21) \quad \int_0^t \int (f(y) - f(X(s-)))\mu(X(s-), dy) d(N(s) - \lambda s).$$

Adding (10.20) and (10.21) to (10.19) gives

$$(10.22) \quad f(X(t)) - f(X(0)) - \int_0^t (g(X(s)) + Bf(X(s))) ds,$$

which is therefore an $\{\mathcal{F}_t\}$ -martingale. \square

10.3 Theorem Let $A \subset \bar{C}(E) \times B(E)$, suppose $\mathcal{D}(A)$ is separating, and let B be given by (10.14). Suppose the $D_E[0, \infty)$ martingale problem for A is well-posed, let $P_x \in \mathcal{P}(D_E[0, \infty))$ denote the distribution of the solution for (A, δ_x) , and suppose $x \rightarrow P_x$ is a Borel measurable function from E into $\mathcal{P}(D_E[0, \infty))$ (cf. Theorem 4.6). Then the $D_{E \times \mathbb{Z}_+}[0, \infty)$ martingale problem for $C \subset B(E \times \mathbb{Z}_+) \times B(E \times \mathbb{Z}_+)$, defined by

$$(10.23) \quad C = \left\{ \left(\int f h, gh + \lambda(\cdot) \int (f(y)h(\cdot + 1) - f(\cdot)h(\cdot))\mu(\cdot, dy) \right) : (f, g) \in A, h \in B(\mathbb{Z}_+) \right\},$$

is well-posed.

10.4 Remark Note that if (X, N) is a solution of the martingale problem for C , then X is a solution of the martingale problem for $A + B$. The component N simply counts the "new" jumps. \square

Proof. If the martingale problem for A is well-posed, then by Theorem 10.1 the martingale problem for $A_0 = \{(fh, gh) : (f, g) \in A, h \in B(\mathbb{Z}_+)\}$ is well-posed (for the second component, $N(t) \equiv N(0)$). For $f \in B(E \times \mathbb{Z}_+)$ define

$$(10.24) \quad B_0 f(x, k) = \lambda(x) \int (f(y, k+1) - f(x, k))\mu(x, dy).$$

Then $C = A_0 + B_0$, and the existence of solutions follows by Proposition 10.2.

For $f \in B(E)$ define

$$(10.25) \quad T(t)f(x) = E^{P_x}[f(X(t))],$$

and note that $\{T(t)\}$ is the semigroup corresponding to the solutions of the martingale problem for A . Let (Y, N) be a solution of the $D_{E \times \mathbb{Z}_+}[0, \infty)$ martingale problem for C .

Note that

$$(10.26) \quad \chi_{\{N(t) = N(0)\}} \exp \left\{ \int_0^t \lambda(Y(s)) ds \right\}$$

is a nonnegative mean one martingale, and let $Q \in \mathcal{P}(D_E[0, \infty))$ be determined by

$$(10.27) \quad Q\{X(t_1) \in \Gamma_1, \dots, X(t_m) \in \Gamma_m\}$$

$$= E \left[\prod_{i=1}^m \chi_{\Gamma_i}(Y(t_i)) \chi_{\{N(t_m) = N(0)\}} \exp \left\{ \int_0^{t_m} \lambda(Y(s)) ds \right\} \right],$$

for $0 \leq t_1 < t_2 < \dots < t_m$, $\Gamma_1, \dots, \Gamma_m \in \mathcal{B}(E)$. (Here X is the coordinate process.) Since (Y, N) is a solution of the martingale problem for C , it follows by Lemma 3.4 that

$$(10.28) \quad f(Y(t)) \chi_{\{N(t) = N(0)\}} \exp \left\{ \int_0^t \lambda(Y(s)) ds \right\} - \int_0^t g(Y(s)) \chi_{\{N(s) = N(0)\}} \exp \left\{ \int_0^s \lambda(Y(u)) du \right\} ds$$

is an $\{\mathcal{F}_t^{(Y, N)}\}$ -martingale for $(f, g) \in A$. Since (10.26) and (10.28) are martingales,

$$(10.29) \quad \begin{aligned} & E^Q \left[\left(f(X(t_{n+1})) - f(X(t_n)) - \int_{t_n}^{t_{n+1}} g(X(s)) ds \right) \prod_{k=1}^n h_k(X(t_k)) \right] \\ &= E \left[\left(f(Y(t_{n+1})) - f(Y(t_n)) - \int_{t_n}^{t_{n+1}} g(Y(s)) ds \right) \prod_{k=1}^n h_k(Y(t_k)) \right. \\ & \quad \times \left. \chi_{\{N(t_{n+1}) = N(0)\}} \exp \left\{ \int_0^{t_{n+1}} \lambda(Y(s)) ds \right\} \right] \\ &= 0 \end{aligned}$$

for $t_1 < t_2 < \dots < t_{n+1}$, $(f, g) \in A$, and $h_k \in B(E)$, and it follows that Q is a solution of the martingale problem for A . In particular,

$$(10.30) \quad E[T(t)f(Y(0))] = E \left[f(Y(t)) \exp \left\{ \int_0^t \lambda(Y(s)) ds \right\} \chi_{\{N(t) = N(0)\}} \right].$$

More generally, for $t \geq s$,

$$(10.31) \quad E[T(t-s)f(Y(s))] = E \left[f(Y(t)) \exp \left\{ \int_s^t \lambda(Y(u)) du \right\} \chi_{\{N(t) = N(s)\}} \right].$$

To complete the proof we need the following two lemmas.

10.5 Lemma For $f \in B(E)$ and $t \geq 0$,

$$(10.32) \quad E \left[\int_0^t f(Y(u)) \exp \left\{ \int_0^u \lambda(Y(u)) du \right\} \chi_{\{N(u) = N(s)\}} dN(s) \right] \\ = E \left[\int_0^t T(t-s)f(Y(s)) dN(s) \right].$$

Proof. Proceeding as above, for $0 \leq a < b$ assume $E[N(b) - N(a)] > 0$, and let Q be determined by

$$(10.33) \quad Q\{X(t_1) \in \Gamma_1, \dots, X(t_m) \in \Gamma_m\} \\ = \frac{E \left[\int_a^b \prod_{i=1}^m \chi_{\Gamma_i}(Y(s+t_i)) \exp \left\{ \int_s^{s+t_m} \lambda(Y(u)) du \right\} \chi_{\{N(s+t_m) = N(s)\}} dN(s) \right]}{E[N(b) - N(a)]}.$$

Then Q is a solution of the martingale problem for A with initial distribution given by

$$(10.34) \quad v(\Gamma) = \frac{E \left[\int_a^b \chi_{\Gamma}(Y(s)) dN(s) \right]}{E[N(b) - N(a)]}.$$

Consequently,

$$(10.35) \quad E \left[\int_a^b f(Y(s+t)) \exp \left\{ \int_s^{s+t} \lambda(Y(u)) du \right\} \chi_{\{N(s+t) = N(s)\}} dN(s) \right] \\ = \int_a^b T(t)f d\nu E[N(b) - N(a)] \\ = E \left[\int_a^b T(t)f(Y(s)) dN(s) \right].$$

(Note that if $E[N(b) - N(a)] = 0$, then (10.35) is immediate.)

Since $T(u)f(x)$ is right continuous in u for each $x \in E$ and $f \in \bar{C}(E)$, we have, for $0 = t_0 < \dots < t_{n+1} = t$,

$$(10.36) \quad E \left[\int_0^t T(t-s)f(Y(s)) dN(s) \right] \\ = \lim_{\max(t_{i+1} - t_i) \rightarrow 0} \sum_{i=0}^n E \left[\int_{t_i}^{t_{i+1}} T(t-t_i)f(Y(s)) dN(s) \right] \\ = \lim_{\max(t_{i+1} - t_i) \rightarrow 0} \sum_{i=0}^n E \left[\int_{t_i}^{t_{i+1}} f(Y(s+t-t_i)) \right. \\ \times \exp \left\{ \int_s^{s+t-t_i} \lambda(Y(u)) du \right\} \chi_{\{N(s+t-t_i) = N(s)\}} dN(s) \Big] \\ = E \left[\int_0^t f(Y(t)) \exp \left\{ \int_s^t \lambda(Y(u)) du \right\} \chi_{\{N(t) = N(s)\}} dN(s) \right]. \quad \square$$

10.6 Lemma For $h \in B(E)$ and $t \geq 0$,

$$(10.37) \quad E\left[\int_0^t h(Y(s)) dN(s)\right] = E\left[\int_0^t \lambda(Y(s)) \int h(y) \mu(Y(s), dy) ds\right].$$

Proof. For $(f, g) \in A$,

$$(10.38) \quad f(Y(t)) \chi_{\{N(t)=k\}} - \int_0^t \left(g(Y(s)) \chi_{\{N(s)=k\}} + \lambda(Y(s)) \int f(y) \chi_{\{N(s)+1=k\}} \right. \\ \left. - f(Y(s)) \chi_{\{N(s)=k\}} \right) \mu(Y(s), dy) ds$$

is a right continuous martingale. Consequently, if $\tau_k = \inf\{t: N(t) = k\}$,

$$(10.39) \quad E[f(Y(\tau_k)) \chi_{\{\tau_k \leq t\}}] = E\left[\int_{s_{k-1} \wedge t}^{s_k \wedge t} \lambda(Y(s)) \int f(y) \mu(Y(s), dy) ds\right].$$

Summing over k gives (10.37) with $h = f$. For general h , the result follows from the fact that $\mathcal{D}(A)$ is separating. \square

Proof of Theorem 10.3 continued. From (10.30) and Lemmas 10.5 and 10.6,

$$(10.40) \quad E[f(Y(t))] - E[T(t)f(Y(0))] \\ = E\left[f(Y(t)) \left(1 - \exp\left\{\int_0^t \lambda(Y(u)) du\right\} \chi_{\{N(t)=N(0)\}}\right)\right] \\ = \lim_{\max(s_{i+1} - s_i) \rightarrow 0} E\left[\sum f(Y(t)) \left(\exp\left\{\int_{s_{i+1}}^t \lambda(Y(u)) du\right\} \chi_{\{N(t)=N(s_{i+1})\}} \right. \\ \left. - \exp\left\{\int_{s_i}^t \lambda(Y(u)) du\right\} \chi_{\{N(t)=N(s_i)\}}\right)\right] \\ = E\left[\int_0^t f(Y(t)) \exp\left\{\int_s^t \lambda(Y(u)) du\right\} \chi_{\{N(t)=N(s)\}} dN(s)\right] \\ - E\left[\int_0^t f(Y(t)) \lambda(Y(s)) \exp\left\{\int_s^t \lambda(Y(u)) du\right\} \chi_{\{N(t)=N(s)\}} ds\right] \\ = E\left[\int_0^t \lambda(Y(s)) \int T(t-s)f(y) \mu(Y(s), dy) ds\right] \\ - E\left[\int_0^t \lambda(Y(s)) T(t-s) f(Y(s)) ds\right] \\ = \int_0^t E[BT(t-s)f(Y(s))] ds,$$

that is,

$$(10.41) \quad E[f(Y(t))] = E[T(t)f(Y(0))] + \int_0^t E[BT(t-s)f(Y(s))] ds$$

for every $f \in B(E)$. Iterating this identity gives

$$(10.42) \quad E[f(Y(t))] = E[T(t)f(Y(0))] + \int_0^t E[T(s)BT(t-s)f(Y(0))] ds + \int_0^t \int_0^s E[BT(s-u)BT(t-s)f(Y(u))] du ds,$$

and we see that by repeated iteration the left side is uniquely determined in terms of $Y(0)$, $\{T(t)\}$, and B . Consequently, uniqueness for the martingale problem follows by Theorem 4.2. \square

11. PROBLEMS

1. (a) Show that to verify (1.1) it is enough to show that

$$(11.1) \quad P\{X(u) \in \Gamma | X(t_n), X(t_{n-1}), \dots, X(t_1)\} = P\{X(u) \in \Gamma | X(t)\}$$

for every finite collection $0 \leq t_1 < t_2 < \dots < t_n = t < u$.

- (b) Show that the process constructed in the proof of Theorem 1.1 is Markov.
2. Let X be a progressive Markov process corresponding to a measurable contraction semigroup $\{T(t)\}$ on $B(E)$ with full generator \hat{A} . Let $\Delta_1, \Delta_2, \dots$ be independent random variables with $P\{\Delta_k > t\} = e^{-t}$, $t \geq 0$, and let V be an independent Poisson process with parameter 1. Show that $X(n^{-1} \sum_{k=1}^{V(n)} \Delta_k)$ is a Markov process whose full generator is $A_n = \hat{A}(I - n^{-1} \hat{A})^{-1}$, the Yosida approximation of \hat{A} .
3. Suppose $\{T(t)\}$ is a semigroup on $B(E)$ given by a transition function and has full generator \hat{A} . Let

$$(11.2) \quad Bf(x) = \lambda(x) \int (f(y) - f(x))\mu(x, dy)$$

where $\lambda \in B(E)$ is nonnegative and $\mu(x, \Gamma)$ is a transition function. Show that $\hat{A} + B$ is the full generator of a semigroup on $B(E)$ given by a transition function.

4. Show that X defined by (2.3) has the same finite-dimensional distributions as X' defined by (2.6).
5. Dropping the assumption that λ is bounded in (2.3), show that $X(t)$ is defined for all $t \geq 0$ with probability 1 if and only if $P\{\sum_{k=0}^{\infty} 1/\lambda(Y(k)) = \infty\} = 1$. In particular, show that $P\{\sum_{k=0}^{\infty} \Delta_k/\lambda(Y(k)) = \infty\} \Delta\{\sum_{k=0}^{\infty} 1/\lambda(Y(k)) = \infty\} = 0$.
6. Show that X given by (2.3) is strong Markov.
7. Let X be a Markov process corresponding to a semigroup $\{T(t)\}$ on $B(E)$. Let V be an independent Poisson process with parameter 1. Show that $X(V(nt)/n)$ is a Markov process. What is its generator?
8. Let $E = \{0, 1, 2, \dots\}$. Let $q_{ij} \geq 0$, $i \neq j$, and let $\sum_{j \neq i} q_{ij} = -q_{ii} < \infty$. Suppose for each i there exists a Markov process X^i with sample paths in $D_E[0, \infty)$ such that $X^i(0) = i$ and

$$(11.3) \quad \lim_{\varepsilon \rightarrow 0^+} \varepsilon^{-1} (P\{X^i(t + \varepsilon) = j | X^i(t)\} - \chi_{ij}(X^i(t))) \\ = q_{X^i(t)j}, \quad j \in E, t \geq 0.$$

- (a) Show that X^i is the unique such process.
- (b) For $i \in E$ and $n = 1, 2, \dots$, let X_n^i be a Markov process with sample paths in $D_E[0, \infty)$ satisfying $X_n^i(0) = i$ and

$$(11.4) \quad \lim_{\varepsilon \rightarrow 0^+} \varepsilon^{-1} (P\{X_n^i(t + \varepsilon) = j | X_n^i(t)\} - \chi_{ij}(X_n^i(t))) \\ = q_{X_n^i(t)j}^{(n)}, \quad j \in E, t \geq 0.$$

Show that $X_n^i \Rightarrow X^i$ for all $i \in E$ if and only if

$$(11.5) \quad \lim_{n \rightarrow \infty} q_{ij}^{(n)} = q_{ij}, \quad i, j \in E$$

(cf. Problem 31).

9. Prove Theorem 2.6.
10. Let ξ_1, ξ_2, \dots be independent, identically distributed random variables with mean zero and variance one. Let

$$(11.6) \quad X_n(t) = \frac{1}{\sqrt{n}} \sum_{k=1}^{[nt]} \xi_k.$$

Show that $X_n \Rightarrow X$ where X is standard Brownian motion. (Apply Theorem 6.5 of Chapter 1 and Theorem 2.6 of this chapter, using the fact that $C_c^\infty(\mathbb{R})$ is a core for the generator for X .)

11. Let Y be a Poisson process with parameter λ and define

$$(11.7) \quad X_n(t) = n^{-1}(Y(n^2t) - \lambda n^2 t).$$

Use Theorem 6.1 of Chapter 1 and Theorem 2.5 of this chapter to show that $\{X_n\}$ converges in distribution and identify the limit.

12. Let $E = \mathbb{R}$ and $Af(x) = a(x)f''(x) + b(x)f'(x)$ for $f \in C_c^\infty(\mathbb{R})$, where a and b are locally bounded Borel functions satisfying $0 \leq a(x) \leq K(1 + |x|^2)$ and $xb(x) \leq K(1 + |x|^2)$ for some $K > 0$. Show that A is conservative. Extend this result to higher dimensions.
13. Let $E = \mathbb{R}$ and $Af(x) = x^2(\sin^2 x)f'(x)$ and $Bf = f'$ for $f \in C_c^\infty(\mathbb{R})$.
- Show that A , B , and $A + B$ satisfy the conditions of Theorem 2.2.
 - Show that A and B are conservative but $A + B$ is not.

14. Complete the proof of Lemma 3.2.

15. Let E be locally compact and separable with one-point compactification E^Δ . Suppose that $\lambda \in M(E)$ is nonnegative and bounded on compact sets, and that $\mu(x, \Gamma)$ is a transition function on $E \times \mathcal{B}(E)$. Define X as in (2.3), setting $X(t) = \Delta$ for $t \geq \sum_{k=0}^{\infty} \Delta_k / \lambda(Y(k))$, and let

$$(11.8) \quad Af(x) = \begin{cases} 0 & x = \Delta \\ \lambda(x) \int (f(y) - f(x))\mu(x, dy) & x \in E \end{cases}$$

for each $f \in B(E^\Delta)$ such that

$$(11.9) \quad \sup_{x \in E} \lambda(x) \int |f(y) - f(x)| \mu(x, dy) < \infty.$$

- Show that X is a solution of the martingale problem for A .
- Suppose $B(E^\Delta)$ is the bp-closure of the collection of f satisfying (11.9). Show that X is the unique solution of the martingale problem for (A, v) , where v is the initial distribution for Y , if and only if $P\{\sum_{k=0}^{\infty} 1/\lambda(Y(k)) = \infty\} = 1$.
- Let $E = \mathbb{R}^d$. Suppose $\sup_x \lambda(x)\mu(x, \Gamma) < \infty$ for every compact $\Gamma \subset \mathbb{R}^d$, and

$$(11.10) \quad \lambda(x) \int |y - x| \mu(x, dy) \leq K(1 + |x|), \quad x \in \mathbb{R}^d,$$

for some constant K . Use Theorem 3.8 to show that X has sample paths in $D_E[0, \infty)$, and show that X is the unique solution of the martingale problem for (A, v) .

16. (Discrete-time martingale problem)

- (a) Let $\mu(x, \Gamma)$ be a transition function on $E \times \mathcal{D}(E)$ and let $X(n)$, $n = 0, 1, 2, \dots$ be a sequence of E -valued random variables. Define $A: B(E) \rightarrow B(E)$ by

$$(11.11) \quad Af(x) = \int f(y)\mu(x, dy) - f(x),$$

and suppose

$$(11.12) \quad f(X(n)) - \sum_{k=0}^{n-1} Af(X(k))$$

is an $\{\mathcal{F}_n^X\}$ -martingale for each $f \in B(E)$. Show that X is a Markov chain with transition function $\mu(x, \Gamma)$.

- (b) Let $X(n)$, $n = 0, 1, 2, \dots$, be a sequence of \mathbb{Z} -valued random variables such that for each $n \geq 0$, $|X(n+1) - X(n)| = 1$. Let $g: \mathbb{Z} \rightarrow [-1, 1]$ and suppose that

$$X(n) - \sum_{k=0}^{n-1} g(X(k))$$

is an $\{\mathcal{F}_n^X\}$ -martingale. Show that X is a Markov chain and calculate its transition probabilities in terms of g .

17. Suppose that (E, r) is complete and separable, and that $P(t, x, \Gamma)$ is a transition function satisfying

$$(11.13) \quad \lim_{n \rightarrow \infty} \sup_x P\left(\frac{1}{n}, x, B(x, \varepsilon)^c\right) = 0$$

for each $\varepsilon > 0$.

- (a) Show that each Markov process X corresponding to $P(t, x, \Gamma)$ has a version with sample paths in $D_E[0, \infty)$. (Apply Theorem 8.6 of Chapter 3.)

- (b) Suppose

$$(11.14) \quad \lim_{n \rightarrow \infty} \sup_x nP\left(\frac{1}{n}, x, B(x, \varepsilon)^c\right) = 0$$

for each $\varepsilon > 0$. Show that the version obtained in (a) has sample paths a.s. in $C_E[0, \infty)$ (cf. Proposition 2.9.).

18. Let E be compact, and let A be a linear operator on $C(E)$ with $\mathcal{D}(A)$ dense in $C(E)$. Suppose there exist transition functions $\mu_n(x, \Gamma)$ such that for each $f \in \mathcal{D}(A)$

$$(11.15) \quad Af(x) = \lim_{n \rightarrow \infty} n \int (f(y) - f(x))\mu_n(x, dy)$$

uniformly in x and that

$$(11.16) \quad \lim_{n \rightarrow \infty} \sup_x n \mu_n(x, B(x, \varepsilon)^c) = 0$$

for each $\varepsilon > 0$. Show that for every $v \in \mathcal{P}(E)$ there exists a solution of the martingale problem for (A, v) with continuous sample paths.

19. Let (E, r) be separable, $A \subset B(E) \times B(E)$, $f, g \in M(E \times E)$, and suppose that for each $y \in E$, $(f(\cdot, y), g(\cdot, y)) \in A$. If for each $\varepsilon > 0$ and compact $K \subset E$, $\inf \{f(x, y) - f(y, y) : x, y \in K, r(x, y) \geq \varepsilon\} > 0$ and if for each $x \in E$, $\lim_{y \rightarrow x} g(x, y) = g(x, x) = 0$, then every solution of the martingale problem for A with sample paths in $D_E[0, \infty)$ has almost all sample paths in $C_E[0, \infty)$ (cf. Proposition 2.9 and Remark 2.10).
20. For $i = 1, 2, \dots$, let E_i be locally compact and separable, and let A_i be the generator of a Feller semigroup on $\hat{C}(E_i)$. Let $E = \prod_{i=1}^{\infty} E_i$. For each i , let $\beta_i \in B(E)$ be nonnegative. For $g(x) = \prod_{i=1}^n f_i(x_i)$, $n \geq 1$, $f_i \in \mathcal{D}(A_i)$, define

$$(11.17) \quad Ag(x) = \sum_{i=1}^n \beta_i(x) \left(\prod_{j \neq i} f_j(x_j) \right) A_i f_i(x_i).$$

Show that every solution of the martingale problem for A has a modification with sample paths in $D_E[0, \infty)$.

21. Let E be the set of finite nonnegative integer-valued measures on $\{0, 1, 2, \dots\}$ with the weak topology (which in this case is equivalent to the discrete topology). For $f \in B(E)$ with compact support, define

$$(11.18) \quad Af(\alpha) = \int k^2(f(\alpha + \delta_{k-1} - \delta_k) - f(\alpha))\alpha(dk).$$

- (a) Interpret a solution of the martingale problem for A in terms of particles moving in $\{0, 1, 2, \dots\}$.
- (b) Show that for each $v \in \mathcal{P}(E)$, the $D_E[0, \infty)$ martingale problem for (A, v) has a unique solution, but that uniqueness is lost if the requirement on the sample paths is dropped.

22. Let $E = [0, 1]$ and $A = \{(f, -f') : f \in C^1(E), f(0) = f(1)\}$.
 - (a) Show that A satisfies the conditions of Corollary 4.4.
 - (b) Show that the martingale problem for $(A, \delta_{1/2})$ has more than one solution if the requirement that the sample paths be in $D_E[0, \infty)$ is dropped.
23. Use (4.44) to compute the moments for the limit distribution for X .

Hint: Write the integral as a sum over the intervals on which Y is constant.

24. Let $E_1 = [0, 1]$ and $A_1 = \{(f, x(1-x)f'' + (a-bx)f'): f \in C^2(E_1)\}$, where $0 < a < b$. Use duality to show uniqueness for the martingale problem and to show that, if X is a solution, then $X(t)$ converges in distribution as $t \rightarrow \infty$ with a limiting distribution that does not depend on $X(0)$.
- Hint:* Let $E_2 = \{0, 1, 2, \dots\}$ and $f(x, y) = x^y$.
25. Let $E_1 = E_2 = [0, \infty)$, $A_1 = \{(f, \frac{1}{2}f''): f \in \bar{C}^2(E_1), f''(0) = 0\}$, and $A_2 = \{(f, \frac{1}{2}f''): f \in \bar{C}^2(E_2), f'(0) = 0\}$, that is, let A_1 correspond to absorbing Brownian motion and A_2 to reflecting Brownian motion.
- Let $g \in \bar{C}^2(-\infty, \infty)$ satisfy $g(z) = -g(-z)$. Show that the martingale problems for A_1 and A_2 are dual with respect to $(f, 0, 0)$ where $f(x, y) = g(x+y) + g(x-y)$.
 - Use the result in part (a) to show that $P\{X(t) > y | X(0) = x\} = P\{Y(t) < x | Y(0) = y\}$, where X is absorbing Brownian motion and Y is reflecting Brownian motion.
26. Let $E = [0, 1]$, $A = \{(f, \frac{1}{2}f''): f \in C^2(E), f'(0) = f'(1) = f''(0) = f''(1) = 0\}$, and let $\Gamma \subset \mathcal{P}(D_E[0, \infty))$ be \mathcal{M}_A , the collection of solutions of the martingale problem for A .
- Show that Γ satisfies the conditions of Theorem 5.19.
 - Find a sequence $\{f_k\}$, as in the proof of Theorem 5.19, for which $\Gamma^{(\infty)} = \mathcal{M}_{A_1}$, where
- $$A_1 = \{(f, \frac{1}{2}f''): f \in C^2(E), f''(0) = f''(1) = 0\}.$$
- Find a sequence $\{f_k\}$ for which $\Gamma^{(\infty)} = \mathcal{M}_{A_2}$, where
- $$A_2 = \{(f, \frac{1}{2}f''): f \in C^2(E), f'(0) = f'(1) = 0\}.$$
27. Let U_k , $k = 1, 2, \dots$, be open with $E = \bigcup_{k=1}^{\infty} U_k$. Given $x \in D_E[0, \infty)$, let $S'_k = \inf\{u \geq t: x(u) \notin U_k \text{ or } x(u-) \notin U_k\}$. Show that there exists a sequence of positive integers k_1, k_2, \dots and a sequence $0 = t_1 < t_2 < \dots$ such that $t_{i+1} = S'_{k_i}$ for each $i \geq 1$ and $\lim_{i \rightarrow \infty} t_i = \infty$. In particular every bounded interval $[0, T]$ is contained in a finite union $\bigcup_{i=1}^{\infty} [t_i, S'_{k_i}]$.
- Hint:* Select $k^-(t)$ so that $x(t-) \in U_{k^-(t)}$ and $k^+(t)$ so that $x(t) \in U_{k^+(t)}$, and note that there is an open interval I_t with $t \in I_t$ such that $\{x(s-), x(s): s \in I_t\} \subset U_{k^-(t)} \cup U_{k^+(t)}$.
28. Let (E_k, r_k) , $k = 0, 1, 2, \dots$, be complete, separable metric spaces. Let $E = \bigcup_k E_k$ (think of the E_k as distinct even if they are all the same), and define $r(x, y) = r_k(x, y)$ if $x, y \in E_k$ for some k , and $r(x, y) = 1$ otherwise. (Then (E, r) is complete and separable.) For $k = 0, 1, 2, \dots$, suppose that $A_k \subset \bar{C}(E_k) \times \bar{C}(E_k)$, that the closure of A_k generates a strongly continuous contraction semigroup on $L_k = \overline{\mathcal{D}(A_k)}$, that L_k is separating, and that for each $v \in \mathcal{P}(E_k)$ there exists a solution of the martingale problem

for (A_k, v) with sample paths in $D_{E_k}[0, \infty)$. Let $A \subset \bar{C}(E) \times \bar{C}(E)$ be given by

$$(11.19) \quad A = \left\{ \left(\sum_{k=0}^n \chi_{E_k} f_k, \sum_{k=0}^n \chi_{E_k} A_k f_k \right) : n \geq 0, f_k \in \mathcal{D}(A_k) \right\}.$$

- (a) Show that the closure of A generates a strongly continuous contraction semigroup on $L \equiv \overline{\mathcal{D}(A)}$.
- (b) Show that the martingale problem for A is well-posed.
- (c) Let $\lambda \in C(E)$, $\lambda \geq 0$, and $\sup_{x \in E_k} \lambda(x) < \infty$ for each k . Let $\mu(x, \Gamma)$ be a transition function on $E \times \mathcal{P}(E)$ and define

$$(11.20) \quad Bf(x) = \lambda(x) \int (f(y) - f(x)) \mu(x, dy)$$

for $f \in \bar{C}(E)$. Suppose $B \subset \bar{C}(E) \times C(E)$.

Let

$$(11.21) \quad \lambda_{kl} = \sup_{x \in E_k} \lambda(x) \mu(x, E_l)$$

and suppose for some $a, b \geq 0$,

$$(11.22) \quad \sum_{l>k} l \lambda_{kl} \leq a + bk, \quad k \geq 0.$$

Show that for each $v \in \mathcal{P}(E)$ there exists a unique solution of the local-martingale problem for $(A + B, v)$ with sample paths in $D_E[0, \infty)$.

Remark. The primary examples of the type described above are population models. Let S be the space in which the particles live. Then $E_k = S^k$ corresponds to a population of k particles, A_k describes the motion of the k particles in S , and λ and μ describe the reproduction.

29. Let A be given by (7.15), let λ satisfy the conditions of Lemma 7.2, and define

$$(11.23) \quad U(s, t)f(x) = \int f(y) P(s, t, x, dy),$$

where P is the solution of (7.6).

- (a) Let X be a solution of the martingale problem for A . Show that

$$(11.24) \quad U(s, t)f(X(s)), \quad 0 \leq s \leq t$$

is a martingale.

- (b) Show that the martingale problem for A is well-posed.

30. Let ξ_1, ξ_2, \dots be a stationary process with $E[\xi_{k+1} | \xi_1, \xi_2, \dots, \xi_k] = 0$. Suppose $E[\xi_k^2] = \sigma^2$ and $\lim_{n \rightarrow \infty} n^{-1} \sum_{k=1}^n \xi_k^2 = \sigma^2$ a.s. Apply Theorem 8.2 to show that $\{X_n\}$ given by

$$(11.25) \quad X_n(t) = \frac{1}{\sqrt{n}} \sum_{k=1}^{[nt]} \xi_k$$

converges in distribution.

Hint: Verify relative compactness by estimating $E[(X_n(t+u) - X_n(t))^2 | \mathcal{F}_t^n]$ and applying Theorem 8.6 of Chapter 3.

31. (a) Let (E, r) be complete and separable. Let $\{T_n(t)\}$, $n = 1, 2, \dots$, and $\{T(t)\}$ be semigroups corresponding to Markov processes with state space E . Suppose that $T(t): L \subset \bar{C}(E) \rightarrow L$, where L is convergence determining, and that for each $f \in L$ and compact $K \subset E$,

$$(11.26) \quad \lim_{n \rightarrow \infty} \sup_{x \in K} |T_n(t)f(x) - T(t)f(x)|, \quad t \geq 0.$$

Suppose that for each n , X_n is a Markov process corresponding to $\{T_n(t)\}$, X is a Markov process corresponding to $\{T(t)\}$, and $X_n(0) \Rightarrow X(0)$. Show that the finite-dimensional distributions of X_n converge weakly to those of X .

- (b) Let $E = \{0, 1, 2, \dots\}$. For $f \in B(E)$, define

$$A_n f(k) = \begin{cases} 0 & k = 0 \\ f(n) - f(k) & k \neq 0, n \\ n(f(0) - f(n)) & k = n, \end{cases}$$

$$Af(k) = \begin{cases} 0 & k = 0 \\ f(0) - f(k) & k \neq 0. \end{cases}$$

Show that $T_n(t) \equiv e^{tA_n}$ and $T(t) \equiv e^{tA}$ satisfy (11.26).

- (c) Fix $k > 0$. For each $n \geq 1$, let X_n be a Markov process corresponding to A_n defined in (b) with $X_n(0) = k$. Show that the finite-dimensional distributions of X_n converge weakly but that X_n does not converge in distribution in $D_E[0, \infty)$.

32. Let E be locally compact and separable, and let $\{T(t)\}$ and $\{S(t)\}$ be Feller semigroups on $\bar{C}(E)$ with generators A and B , respectively. Let $\{Y_n(k), k = 0, 1, \dots\}$ be the Markov chain satisfying

$$E[f(Y_n(2k+1)) | Y_n(2k)] = T\left(\frac{1}{n}\right) f(Y_n(2k))$$

and

$$E[f(Y_n(2k)) | Y_n(2k-1)] = S\left(\frac{1}{n}\right) f(Y_n(2k-1)),$$

and set $X_n(t) = Y_n([nt])$. Suppose that $\mathcal{D}(A) \cap \mathcal{D}(B)$ is dense in $\hat{C}(E)$. Show that $\{X_n\}$ is relatively compact in $D_E[0, \infty)$ (E^Δ is the one-point compactification of E), and that any limit point of $\{X_n\}$ is a solution of the martingale problem for $A^\Delta + B^\Delta$ (A^Δ and B^Δ as in Theorem 5.4).

33. Consider a sequence of single server queueing processes. For the n th process, the customer arrivals form a Poisson process with intensity λ_n , and the service time is exponentially distributed with parameter μ_n (only one customer is served at a time).

- (a) Let Y_n be the number of customers waiting to be served. What is the generator corresponding to Y_n ?
- (b) Let $X_n(t) = n^{-1/2} Y_n(nt)$. What is the generator for X_n ?
- (c) Show that if $\{X_n(0)\}$ converges in distribution, $\lim_{n \rightarrow \infty} \lambda_n = \lambda$, and $\lim_{n \rightarrow \infty} n(\lambda_n - \mu_n) = a$, then $\{X_n\}$ converges in distribution. What is the limit?
- (d) What is the result analogous to (c) for two servers serving a single queue?
- (e) What if there are two servers serving separate queues and new arrivals join the shortest queue?

Hint: Make a change of variable. Consider the sum of the two queue lengths and the difference.

34. (a) Let ξ_1, ξ_2, \dots be independent, identically distributed real random variables. For $x_0, a \in \mathbb{R}$, let $Y_n(0) = x_0$ and

$$(11.27) \quad Y_n(k+1) = (1 + n^{-1}a)Y_n(k) + n^{-1/2}\xi_{k+1}, \quad k = 0, 1, \dots$$

For $f \in C_c^2(\mathbb{R})$, calculate

$$(11.28) \quad \lim_{n \rightarrow \infty} nE[f(Y_n(k+1)) - f(Y_n(k)) | Y_n(k) = x],$$

and use this calculation to show that X_n , given by $X_n(t) = Y_n([nt])$, converges in distribution.

- (b) Generalize (a) to d dimensions.

35. Let E be locally compact and separable, let $\lambda: E \rightarrow [0, \infty)$ be measurable and bounded on compact subsets of E , and let $\mu(x, \Gamma)$ be a transition function on $E \times \mathcal{B}(E)$. For $f \in C_c(E)$, define

$$(11.29) \quad Af(x) = \lambda(x) \int (f(y) - f(x))\mu(x, dy).$$

- (a) Let $v \in \mathcal{P}(E)$, and suppose that the local-martingale problem for (A, v) has a solution X with sample paths in $D_E[0, \infty)$ (i.e., does not reach infinity in finite time). Show that the solution is unique.

- (b) Suppose that φ and $\int |\varphi(y)| \mu(\cdot, dy)$ are bounded on compact sets. Suppose that X is a solution of the local-martingale problem for (A, v) with sample paths in $D_E[0, \infty)$. Show that

$$(11.30) \quad \varphi(X(t)) - \int_0^t \lambda(X(s)) \int (\varphi(y) - \varphi(X(s))) \mu(X(s), dy) ds$$

is a local martingale.

- (c) In addition to the assumptions in (b), suppose that $\varphi \geq 0$ and that there exists a constant K such that

$$(11.31) \quad \lambda(x) \int (\varphi(y) - \varphi(x)) \mu(x, dy) \leq K$$

for all x . Show that (11.30) is a supermartingale and that

$$(11.32) \quad \varphi(X(t)) - Kt$$

is a supermartingale.

36. Let $A \subset \bar{C}(E) \times \bar{C}(E)$. Suppose that the martingale problem for A is well-posed, that every solution has a modification with sample paths in $D_E[0, \infty)$, and that there exists $x_0 \in E$ such that every solution (with sample paths in $D_E[0, \infty)$) satisfies $\tau \equiv \inf\{t: X(t) = x_0\} < \infty$ a.s. Show that there is at most one stationary distribution for A .
37. Let $E = \mathbb{R}$, $a, b \in C^2(E)$, $a > 0$, and $A = \{(f, af'' + bf'): f \in C_c^\infty(E)\}$. Suppose there exists $g \in C^2(E)$, $g \geq 0$, satisfying

$$(11.33) \quad \frac{d}{dx} \left[\frac{d}{dx} (ag) - bg \right] = 0$$

and $\int_{-\infty}^{\infty} g dx = 1$. Show that if the martingale problem for A is well-posed, then g is the density for the unique stationary distribution for A .

38. Let $E = \mathbb{R}$ and $A = \{(f, f'' + x^4 f'): f \in C_c^\infty(E)\}$. Show that there exists a stationary solution of the martingale problem for A and describe the behavior of this process.
39. Let $E = [0, 1]$, $a, b \in C(E)$, $a > 0$, and $A = \{(f, af'' + bf'): f \in C^\infty(E)$, $f'(0) = f'(1) = 0\}$. Find the stationary distribution for A .
40. Let $E = [0, 1]$, and $A = \{(f, \frac{1}{2}f''): f \in C^2(E)$, $f'(0) = f'(1) = 0$, and $f'(\frac{1}{3}) = f'(\frac{2}{3})\}$. Show the following:
- (a) $\overline{\mathcal{D}(A)} = C(E)$.
 - (b) $\overline{\mathcal{D}(\lambda - A)} \neq C(E)$ for some (hence all) $\lambda > 0$.
 - (c) The martingale problem for A is well-posed.

(d) $\mu(\Gamma) = 3m(\Gamma \cap [\frac{1}{3}, \frac{2}{3}])$ (where m is Lebesgue measure) satisfies

$$(11.34) \quad \int \frac{1}{2}f'' d\mu = 0, \quad f \in \mathcal{D}(A),$$

but μ is not a stationary distribution for A .

41. Show that X defined in the proof of Theorem 9.17 is stationary.
42. Let (E, r) be complete and separable. If X is a stationary E -valued process, then the ergodic theorem (see, e.g., Lamperti (1977), page 102) ensures that for $h \in B(E)$,

$$(11.35) \quad \lim_{t \rightarrow \infty} t^{-1} \int_0^t h(X(s)) ds$$

exists a.s.

- (a) Let $v \in \mathcal{P}(E)$. Show that if (11.35) equals $\int h dv$ for all $h \in \bar{C}(E)$, then this equality holds for all $h \in B(E)$.
- (b) Let $P(t, x, \Gamma)$ be a transition function such that for some $v \in \mathcal{P}(E)$

$$(11.36) \quad \lim_{t \rightarrow \infty} t^{-1} \int_0^t \int h(y) P(s, x, dy) ds = \int h dv, \quad x \in E, \quad h \in \bar{C}(E).$$

Show that there is at most one such v .

- (c) Let v and $P(t, x, \Gamma)$ be as in (b). Let X be a measurable Markov process corresponding to $P(t, x, \Gamma)$ with initial distribution v (hence X is stationary). Suppose that (11.35) equals $\int h dv$ for all $h \in B(E)$. Show that X is ergodic. (See Lamperti (1977), page 95.)
43. Let (E, r) be complete and separable. Suppose $P(t, x, \Gamma)$ is a transition function with a unique stationary distribution $v \in \mathcal{P}(E)$. Show that if X is a measurable Markov process corresponding to $P(t, x, \Gamma)$ with initial distribution v , then X is ergodic.
44. For $n = 1, 2, \dots$, let X_n be a solution of the martingale problem for $A_n = \{(f, f'' + nb(n \cdot) f'): f \in C_c^2(\mathbb{R})\}$, where b is continuous and in L^1 , and let $\alpha = \int_{-\infty}^{\infty} b(x) dx$. Let X be a solution of the martingale problem for A with

$\mathcal{D}(A) = \{f \in C_c(\mathbb{R}): f'$ and f'' exist and are continuous except

at zero, $f'(0+) = e^{-\alpha} f'(0-)$, and $f''(0+) = f''(0-)\}$,

and Af is the continuous extension of f'' .

- (a) Show that uniqueness holds for the martingale problem for A .
- (b) Show that if $X_n(0) \Rightarrow X(0)$, then $X_n \Rightarrow X$.

Hint: For $f \in \mathcal{D}(A)$, let f_n satisfy $f_n''(x) + nb(nx)f_n'(x) = Af(x)$, and apply the results of Section 8.

- (c) What happens if $\int_{-\infty}^{\infty} b^+(x) dx = \infty$ and $\int_{-\infty}^{\infty} b^-(x) dx < \infty$?
45. Let (E, r) be complete and separable and let $A \subset \mathcal{C}(E) \times \mathcal{B}(E)$. Suppose $\mathcal{D}(A)$ is separating. Let X be a solution of the $D_E[0, \infty)$ martingale problem for A , let $\Gamma \in \mathcal{B}(E)$, and suppose $g(x) = 0$ for every $x \in \Gamma$ and $(f, g) \in A$. For $t \geq 0$, define $\gamma_t = \inf\{u > t: \int_t^u \chi_{\Gamma}(X(s)) ds > 0\}$. Show that $X(u \wedge \gamma_t) = X(t)$ a.s. for all $u > t$, and that with probability one, X is constant on any interval $[t, u]$ for which $\int_t^u \chi_{\Gamma}(X(s)) ds = 0$.
46. Let E be separable. For $n = 1, 2, \dots$, let $\{Y_n(k), k = 0, 1, 2, \dots\}$ be an E -valued discrete-time stationary process, let $\varepsilon_n > 0$, and assume $\varepsilon_n \rightarrow 0$. Define $X_n(t) = Y_n([t/\varepsilon_n])$, and suppose $X_n \Rightarrow X$. Show that X is stationary.
47. Let E be locally compact and separable but not compact, and let $E^\Delta = E \cup \{\Delta\}$ be the one-point compactification. Let $v(x, \Gamma)$ be a transition function on $E \times \mathcal{B}(E)$, and let $\varphi, \psi \in M(E)$. Suppose that $\varphi \geq 0$, that $\psi \leq C$ for some constant C and $\lim_{x \rightarrow \Delta} \psi(x) = -\infty$, and that for every Markov chain $\{Y(k), k = 0, 1, 2, \dots\}$ with transition function $v(x, \Gamma)$ satisfying $E[\varphi(Y(0))] < \infty$,

$$(11.37) \quad \varphi(Y(k)) - \sum_{l=0}^{k-1} \psi(Y(l))$$

is a supermartingale. Suppose Y is stationary. Show that

$$P\{Y(0) \in K_m\} \geq \frac{m}{C+m},$$

where $K_m = \{x: \psi(x) \geq -m\}$.

48. For $i = 1, 2$, let E_i be a locally compact (but not compact) separable metric space and let $E_i^\Delta = E_i \cup \{\Delta_i\}$ be its one-point compactification. Let X be a measurable E_1 -valued process, let Y be a measurable E_2 -valued process, and let $\varphi \in M(E_1)$ and $\psi \in M(E_1 \times E_2)$. For $t > 0$, define

$$(11.38) \quad \mu_t(\Gamma_1) = \frac{1}{t} \int_0^t P\{X(s) \in \Gamma_1\} ds, \quad \Gamma_1 \in \mathcal{B}(E_1),$$

and

$$(11.39) \quad v_t(\Gamma_2) = \frac{1}{t} \int_0^t P\{X(s) \in \Gamma_2\} ds, \quad \Gamma_2 \in \mathcal{B}(E_2).$$

Suppose that

$$(11.40) \quad \varphi(X(t)) - \int_0^t \psi(X(s), Y(s)) ds$$

is a supermartingale, $\varphi \geq 0$, $\psi \leq C$ for some constant C , and that for each compact $K_2 \subset E_2$, $\lim_{x \rightarrow \Delta_1} \sup_{y \in K_2} \psi(x, y) = -\infty$. Show that if $\{\nu_t: t \geq 1\}$ is relatively compact in $\mathcal{P}(E_2)$, then $\{\mu_t: t \geq 1\}$ is relatively compact in $\mathcal{P}(E_1)$. (See Chapter 12.)

49. (a) Let E be compact and $A \subset C(E) \times C(E)$ with $\mathcal{D}(A)$ dense in $C(E)$. Suppose the martingale problem for (A, δ_x) is well-posed for each $x \in E$. Show that the martingale problem for A is well-posed (i.e., the martingale problem for (A, μ) is well-posed for each $\mu \in \mathcal{P}(E)$).
- (b) Extend the result in (a) to E and A satisfying the conditions of Theorem 5.11(b).
50. Let $E_1 = E_2 = [0, 1]$, and set $E = E_1 \times E_2$. Let

$$(11.41) \quad \begin{aligned} A_1 &= \{(f_1, f_2, f_1'' f_2): f_2 \in C(E_2), f_1 \in C^2(E_1), f_1'(0) = \\ &\quad f_1''(0) = f_1'(1) = f_1''(1) = 0\}, \\ A_2 &= \{(f_1 \chi_{[\alpha]}, f_1'' \chi_{[\alpha]}): \alpha \in E_2, f_1 \in C^2(E_1), f_1'(0) = \\ &\quad f_1'(1) = 0\}, \end{aligned}$$

and $A = A_1 \cup A_2$. Show that the martingale problem for $(A, \delta_{(x,y)})$ is well-posed for each $(x, y) \in E$ but that the martingale problem for (A, μ) has more than one solution if μ is absolutely continuous (cf. Problem 26).

12. NOTES

The basic reference for the material in Sections 1 and 2 is Dynkin (1965). Theorem 2.5 originally appeared in Mackevicius (1974) and Kurtz (1975).

Levy (1948) (see Doob (1953)) characterized standard Brownian motion as the unique continuous process W such that $W(t)$ and $W(t)^2 - t$ are martingales. Watanabe (1964) characterized the unit Poisson process as the unique counting process N such that $N(t) - t$ is a martingale. The systematic development of the martingale problem began with Stroock and Varadhan (1969) (see Stroock and Varadhan (1979)) for diffusion processes and was extended to other classes of processes in Stroock and Varadhan (1971), Stroock (1975), Anderson (1976), Holley and Stroock (1976, 1978).

The primary significance of Corollary 4.4 is its applicability to Ray processes. See Williams (1979) for a discussion of this class of processes. Theorem 4.6 is essentially Exercise 6.7.4 of Stroock and Varadhan (1979).

The notion of duality given by (4.36) was developed first in the context of infinite particle systems by Vasershtein (1969), Vasershtein and Leontovich (1970), Spitzer (1970), Holley and Liggett (1975), Harris (1976), Liggett (1977), Holley and Stroock (1979). It has also found application to birth and death processes (Siegmund (1976)), to diffusion processes, particularly those arising

in genetics (Holley, Stroock, and Williams (1977), Shiga (1980, 1981), Cox and Rösler (1982)) (see Problem 25), and to measure-valued processes (Dawson and Kurtz (1982), Ethier and Kurtz (1986)).

Lemma 5.3 is due to Roth (1976). Theorem 5.19 is a refinement of a result of Krylov (1973). The presentation here is in part motivated by an unpublished approach of Gray and Griffléath (1977b). See also the presentation in Stroock and Varadhan (1979).

The use of semigroup approximation theorems to prove convergence to Markov processes began with Trotter (1958) and Skorohod (1958), although work on diffusion approximations by Khintchine (1933) is very much in this spirit. These techniques were refined in Kurtz (1969, 1975). Use of the martingale problem to prove limit theorems began with the work of Stroock and Varadhan (1969) and was developed further in Morkvenas (1974), Papanicolaou, Stroock, and Varadhan (1977), Kushner (1980), and Rebolledo (1979) (cf. Theorem 4.1 of Chapter 7). Proposition 8.18 abstracts an approach of Helland (1981). The recent book of Kushner (1984) gives another development of the convergence theory with many applications.

The results on existence of stationary distributions are due to Khasminskii (1960, 1980), Wonham (1966), Benes (1968), and Zakai (1969). Similar convergence results can be found in Blankenship and Papanicolaou (1978), Constantini, Gerardi, and Nappo (1982), and Kushner (1982). Theorem 9.14 is due to Norman (1977). Theorem 9.17 is due to Echeverria (1982) and has been extended by Weiss (1981).

Problem 25 is from Cox and Rösler (1982). Problem 40 gives an example of a well-posed martingale problem with a compact state space for which the closure of A is not a generator. The first such example was given by Gray and Griffléath (1977a). Problem 44 is due to Rosenkrantz (1975).

5

STOCHASTIC INTEGRAL EQUATIONS

The emphasis in this chapter is on existence and uniqueness of solutions of stochastic integral equations, and the relationship to existence and uniqueness of solutions of the corresponding martingale problems. These results comprise Section 3. Section 1 introduces d -dimensional Brownian motion, while Section 2 defines stochastic integrals with respect to continuous, local martingales and includes Itô's formula.

1. BROWNIAN MOTION

Let ξ_1, ξ_2, \dots be a sequence of independent, identically distributed, \mathbb{R}^d -valued random variables with mean vector 0 and covariance matrix I_d , the $d \times d$ identity matrix. Think of the process

$$(1.1) \quad X_n(t) = \frac{1}{\sqrt{n}} \sum_{k=1}^{\lfloor nt \rfloor} \xi_k, \quad t \geq 0,$$

as specifying for fixed $n \geq 1$ the position at time t of a particle subjected to independent, identically distributed, random displacements of order $1/\sqrt{n}$ at times $1/n, 2/n, 3/n, \dots$. Now let $n \rightarrow \infty$. In view of the (multivariate) central limit theorem, the existence of a limiting process (specified in terms of its finite-dimensional distributions) is clear. If such a process also has continuous sample paths, it is called a d -dimensional Brownian motion.

More precisely, a process $W = \{W(t), t \geq 0\}$ with values in \mathbb{R}^d is said to be a (standard) d -dimensional $\{\mathcal{F}_t\}$ -Brownian motion if:

- (a) $W(0) = 0$ a.s.
- (b) W is adapted to the filtration $\{\mathcal{F}_t\}$, and \mathcal{F}_t is independent of $\sigma(W(u) - W(t); u \geq t)$ for each $t \geq 0$.
- (c) $W(t) - W(s)$ is $N(0, (t-s)I_d)$ (i.e., normal with mean vector 0 and covariance matrix $(t-s)I_d$) for every $t > s \geq 0$.
- (d) W has sample paths in $C_{\mathbb{R}^d}[0, \infty)$.

When $\{\mathcal{F}_t\} = \{\mathcal{F}_t^W\}$ in the above definition, W is said to be a (standard) d -dimensional Brownian motion.

Note that if W is a d -dimensional $\{\mathcal{F}_t\}$ -Brownian motion defined on a probability space (Ω, \mathcal{F}, P) , then W is a d -dimensional $\{\tilde{\mathcal{F}}_t\}$ -Brownian motion on $(\Omega, \tilde{\mathcal{F}}, P)$, where $\tilde{\mathcal{F}}$, and $\tilde{\mathcal{F}}$ denote the P -completions of \mathcal{F} , and \mathcal{F} , and P denotes its own extension to $\tilde{\mathcal{F}}$. (If \mathcal{G} is a sub- σ -algebra of \mathcal{F} , the P -completion of \mathcal{G} is defined to be the smallest σ -algebra containing $\mathcal{G} \cup \{A \subset \Omega : A \subset N \text{ for some } N \in \mathcal{F} \text{ with } P(N) = 0\}$.)

The existence of a d -dimensional Brownian motion can be proved in a number of ways. The approach taken here, while perhaps not as efficient as others, provides an application of the results of Chapter 4, Section 2.

We begin by constructing the Feller semigroup $\{T(t)\}$ on $\hat{C}(\mathbb{R}^d)$ corresponding to W . The interpretation above suggests that $\{T(t)\}$ should satisfy

$$(1.2) \quad T(t)f(x) = \lim_{n \rightarrow \infty} E \left[f \left(x + \frac{1}{\sqrt{n}} \sum_{k=1}^{[nt]} \xi_k \right) \right]$$

for all $f \in \hat{C}(\mathbb{R}^d)$, $x \in \mathbb{R}^d$, and $t \geq 0$. By the central limit theorem, (1.2) is equivalent to

$$(1.3) \quad T(t)f(x) = E[f(x + \sqrt{t}Z)],$$

where Z is $N(0, I_d)$. We take (1.3) as our definition of the semigroup $\{T(t)\}$ on $\hat{C}(\mathbb{R}^d)$.

1.1 Proposition Equation (1.3) defines a Feller semigroup $\{T(t)\}$ on $\hat{C}(\mathbb{R}^d)$. Its generator A is an extension of

$$(1.4) \quad \{(f, \frac{1}{2}\Delta_d f) : f \in \hat{C}^2(\mathbb{R}^d)\},$$

where $\Delta_d \equiv \sum_{i=1}^d \partial_i^2$. Moreover, $C_c^\infty(\mathbb{R}^d)$ is a core for A .

Proof. For each $t \geq 0$, $T(t) : \hat{C}(\mathbb{R}^d) \rightarrow \hat{C}(\mathbb{R}^d)$ by the dominated convergence theorem, so $T(t)$ is a positive linear contraction on $\hat{C}(\mathbb{R}^d)$. Let Z' be an independent copy of Z . Then, by Fubini's theorem,

$$\begin{aligned} (1.5) \quad T(s)T(t)f(x) &= E[T(t)f(x + \sqrt{s}Z)] \\ &= E[f(x + \sqrt{s}Z + \sqrt{t}Z')] \\ &= E[f(x + \sqrt{s+t}Z)] \\ &= T(s+t)f(x) \end{aligned}$$

for all $f \in \hat{C}(\mathbb{R}^d)$, $x \in \mathbb{R}^d$, and $s, t \geq 0$. Since $T(0) = I$, this implies that $\{T(t)\}$ is a semigroup. Observe that each $f \in \hat{C}(\mathbb{R}^d)$ is uniformly continuous (with respect to the Euclidean metric), and let $w(f, \delta)$ denote its modulus of continuity, defined for $\delta \geq 0$ by

$$(1.6) \quad w(f, \delta) = \sup \{ |f(y) - f(x)| : x, y \in \mathbb{R}^d, |y - x| \leq \delta \}.$$

Then $\|T(t)f - f\| \leq E[w(f, \sqrt{t}Z)]$ for all $t \geq 0$, so by the dominated convergence theorem, $\{T(t)\}$ is strongly continuous.

To show that the generator A of $\{T(t)\}$ extends (1.4), fix $f \in \hat{C}^2(\mathbb{R}^d)$. By Taylor's theorem,

$$\begin{aligned} (1.7) \quad T(t)f(x) - f(x) &= E[f(x + \sqrt{t}Z) - f(x)] \\ &= E \left[\sum_{i=1}^d \sqrt{t}Z_i \partial_i f(x) + \frac{1}{2} \sum_{i,j=1}^d t Z_i Z_j \partial_i \partial_j f(x) \right. \\ &\quad \left. + \int_0^1 (1-u) \sum_{i,j=1}^d t Z_i Z_j \{\partial_i \partial_j f(x + u\sqrt{t}Z) - \partial_i \partial_j f(x)\} du \right] \\ &= t \{ \frac{1}{2} \Delta_d f(x) + \varepsilon(t, x) \}, \end{aligned}$$

for all $x \in \mathbb{R}^d$ and $t \geq 0$, where

$$\begin{aligned} (1.8) \quad |\varepsilon(t, x)| &\leq \int_0^1 (1-u) \sum_{i,j=1}^d \{E[Z_i^2]E[Z_j^2 w(\partial_i \partial_j f, u\sqrt{t}Z)^2]\}^{1/2} du \\ &\leq \frac{1}{2} \sum_{i,j=1}^d \{E[Z_i^2 w(\partial_i \partial_j f, \sqrt{t}Z)^2]\}^{1/2}. \end{aligned}$$

We conclude that $f \in \mathcal{D}(A)$ and $Af = \frac{1}{2} \Delta_d f$.

Observe next that (1.3) can be rewritten as

$$(1.9) \quad T(t)f(x) = \int_{\mathbb{R}^d} f(y)(2\pi t)^{-d/2} \exp \{-|y-x|^2/2t\} dy,$$

provided $t > 0$. It follows easily that

$$(1.10) \quad T(t): \hat{C}(\mathbb{R}^d) \rightarrow \hat{C}^\infty(\mathbb{R}^d), \quad t > 0,$$

where $\hat{C}^\infty(\mathbb{R}^d) = \bigcap_{k \geq 1} \hat{C}^k(\mathbb{R}^d)$. By Proposition 3.3 of Chapter 1, $\hat{C}^\infty(\mathbb{R}^d)$ is a core for A . Now choose $h \in C_c^\infty(\mathbb{R}^d)$ such that $\chi_{\{x: |x| \leq 1\}} \leq h \leq \chi_{\{x: |x| \leq 2\}}$, and define $\{h_n\} \subset C_c^\infty(\mathbb{R}^d)$ by $h_n(x) = h(x/n)$. Given $f \in \hat{C}^\infty(\mathbb{R}^d)$, observe that $f h_n \rightarrow f$ and $A(f h_n) = (Af)h_n + f A h_n + \nabla f \cdot \nabla h_n \rightarrow Af$ uniformly as $n \rightarrow \infty$, implying that $C_c^\infty(\mathbb{R}^d)$ is a core for A . Finally, since $\text{bp-lim}_{n \rightarrow \infty} (h_n, Ah_n) = (1, 0)$, A is conservative. \square

The main result of this section proves the existence of a d -dimensional Brownian motion and describes several of its properties.

1.2 Theorem A d -dimensional Brownian motion exists. Let W be a d -dimensional $\{\mathcal{F}_t\}$ -Brownian motion. Then the following hold:

- (a) W is a strong Markov process with respect to $\{\mathcal{F}_t\}$ and corresponds to the semigroup $\{T(t)\}$ of Proposition 1.1.
- (b) Writing $W = (W_1, \dots, W_d)$, each W_i is a continuous, square-integrable, $\{\mathcal{F}_t\}$ -martingale, and $\langle W_i, W_j \rangle_t = \delta_{ij}t$ for $i, j = 1, \dots, d$ and all $t \geq 0$.
- (c) With $\{X_n\}$ defined as in the first paragraph of this section, $X_n \Rightarrow W$ in $D_{\mathbb{R}^d}[0, \infty)$ as $n \rightarrow \infty$.

Proof. By Proposition 1.1 of this chapter and Theorem 2.7 of Chapter 4, there exists a probability space (Ω, \mathcal{F}, P) on which is defined a process $W = \{W(t), t \geq 0\}$ with $W(0) = 0$ a.s. and sample paths in $D_{\mathbb{R}^d}[0, \infty)$ satisfying part (a) of the theorem with $\{\mathcal{F}_t\} = \{\mathcal{F}_t^W\}$. In fact, we may assume that W has sample paths in $C_b[0, \infty)$ by Proposition 2.9 of Chapter 4. For $t > s \geq 0$ and $f \in \hat{C}(\mathbb{R}^d)$, the Markov property and (1.3) give

$$(1.11) \quad E[f(W(t) - z) | \mathcal{F}_s^W] = E[f(\sqrt{t-s}Z + x - z)]|_{x=W(s)}.$$

Consequently, since $W(s)$ is \mathcal{F}_s^W -measurable,

$$(1.12) \quad E[f(W(t) - W(s)) | \mathcal{F}_s^W] = E[f(\sqrt{t-s}Z)].$$

It follows that \mathcal{F}_s^W is independent of $\sigma(W(u) - W(s); u \geq s)$ and that $W(t) - W(s)$ is $N(0, (t-s)I_d)$ for all $t > s \geq 0$. In particular, W is a d -dimensional Brownian motion.

Let W be a d -dimensional $\{\mathcal{F}_t\}$ -Brownian motion where $\{\mathcal{F}_t\}$ need not be $\{\mathcal{F}_t^W\}$. To prove part (a), let τ be an $\{\mathcal{F}_t\}$ -stopping time concentrated on $\{t_1\}$,

$t_2, \dots} \subset [0, \infty)$. Let $A \in \mathcal{F}_t$, $s > 0$, and $f \in \hat{C}(\mathbb{R}^d)$. Then $A \cap \{\tau = t_i\} \in \mathcal{F}_{t_i}$, so

$$\begin{aligned}
 (1.13) \quad & \int_{A \cap \{\tau = t_i\}} f(W(\tau + s)) dP \\
 &= \int_{A \cap \{\tau = t_i\}} E[f(W(t_i + s)) - W(t_i) + W(t_i)] | \mathcal{F}_{t_i} dP \\
 &= \int_{A \cap \{\tau = t_i\}} E[f(\sqrt{s}Z + x)]|_{x=W(t_i)} dP \\
 &= \int_{A \cap \{\tau = t_i\}} T(s)f(W(t_i)) dP.
 \end{aligned}$$

The verification of (a) is completed as in the proof of Theorem 2.7 of Chapter 4.

Applying the Markov property (with respect to $\{\mathcal{F}_t\}$), we have

$$(1.14) \quad E[f(W(t)) | \mathcal{F}_s] = E[f(x + \sqrt{t-s}Z)]|_{x=W(s)}$$

for all $f \in \hat{C}(\mathbb{R}^d)$ and $t > s \geq 0$, hence for all $f \in C(\mathbb{R}^d)$ with polynomial growth. Taking $f(x) = x_i$ and then $f(x) = x_i x_j$, we conclude that W_i is a continuous, square-integrable, $\{\mathcal{F}_t\}$ -martingale, and

$$(1.15) \quad E[W_i(t)W_j(t) | \mathcal{F}_s] = W_i(s)W_j(s) + \delta_{ij}(t-s), \quad t > s \geq 0,$$

for $i, j = 1, \dots, d$. This implies (b).

Part (c) follows from Theorems 6.5 of Chapter 1 and 2.6 of Chapter 4, provided we can show that, for every $f \in C_c^\infty(\mathbb{R}^d)$,

$$(1.16) \quad nE\left[f\left(x + \frac{1}{\sqrt{n}}\xi_1\right) - f(x)\right] \rightarrow \frac{1}{2}\Delta_d f(x)$$

as $n \rightarrow \infty$, uniformly in $x \in \mathbb{R}^d$. Observe, however, that this follows immediately from (1.7) and (1.8) if we replace t and Z by $1/n$ and ξ_1 . \square

2. STOCHASTIC INTEGRALS

Let (Ω, \mathcal{F}, P) be a complete probability space with a filtration $\{\mathcal{F}_t\}$ such that \mathcal{F}_0 contains all P -null sets of \mathcal{F} . Throughout this section, $\{\mathcal{F}_t\}$ implicitly prefixes each of the following terms: martingale, progressive, adapted, stopping time, local martingale, and Brownian motion.

Let \mathcal{M}_c be the space of continuous, square-integrable martingales M with $M(0) = 0$ a.s. Given $M \in \mathcal{M}_c$, denote its increasing process (see Chapter 2,

Section 6) by $\langle M \rangle$, and let $L^2(\langle M \rangle)$ be the space of all real-valued, progressive processes X such that

$$(2.1) \quad E\left[\int_0^t X^2 d\langle M \rangle\right] < \infty, \quad t \geq 0.$$

In this section we define the stochastic integral

$$(2.2) \quad \int_0^t X dM$$

for each $X \in L^2(\langle M \rangle)$ as an element of \mathcal{M}_c itself. Actually, (2.2) is uniquely determined only up to indistinguishability. As in the case with conditional expectations and increasing processes, this indeterminacy is inherent. Therefore we adopt the convention of suppressing the otherwise pervasive phrase "almost surely" whenever it is needed only because of this indeterminacy.

Since the sample paths of M are typically of unbounded variation on every nondegenerate interval, we cannot in general define (2.2) in a pathwise sense. However, if the sample paths of X are of bounded variation on bounded intervals, then we can define (2.2) pathwise as a Stieltjes integral, and integrating by parts gives

$$(2.3) \quad \int_0^t X dM = X(t)M(t) - \int_0^t M dX, \quad t \geq 0.$$

In particular, when X belongs to the space S of real-valued, bounded, adapted, right-continuous step functions, that is, when X is a real-valued, bounded process for which there exist $0 = t_0 < t_1 < t_2 < \dots$ with $t_n \rightarrow \infty$ such that

$$(2.4) \quad X(t) = \sum_{i=0}^{\infty} X(t_i) \chi_{[t_i, t_{i+1})}(t), \quad t \geq 0,$$

and $X(t_i)$ is \mathcal{F}_{t_i} -measurable for each $i \geq 0$, we have

$$(2.5) \quad \int_0^t X dM = \sum_{\substack{i \geq 0 \\ t_{i+1} \leq t}} X(t_i)(M(t_{i+1}) - M(t_i)) + X(t_{m(t)})(M(t) - M(t_{m(t)}))$$

for all $t \geq 0$, where $m(t) = \max \{i \geq 0: t_i \leq t\}$. Observe that (2.5) is linear in X and in M .

2.1 Lemma If $M \in \mathcal{M}_c$ and $X \in S$, then (2.5) defines a process $\int_0^t X dM \in \mathcal{M}_c$ and

$$(2.6) \quad \left\langle \int_0^t X dM \right\rangle_t = \int_0^t X^2 d\langle M \rangle, \quad t \geq 0.$$

If, in addition, $N \in \mathcal{M}_c$ and $Y \in S$, then

$$(2.7) \quad \left\langle \int_0^t X dM, \int_0^t Y dN \right\rangle_t = \int_0^t XY d\langle M, N \rangle, \quad t \geq 0.$$

Proof. Clearly, (2.5) is continuous and adapted, and it is square-integrable because X is bounded and $M \in \mathcal{M}_c$. Fix $t \geq s \geq 0$. We can assume that the partition $0 = t_0 < t_1 < \dots$ associated with X as in (2.4) is also associated with Y and that s and t belong to it. Letting

$$(2.8) \quad \int_s^t X \, dM = \int_0^t X \, dM - \int_0^s X \, dM,$$

we have

$$(2.9) \quad E\left[\int_s^t X \, dM \mid \mathcal{F}_s\right] = E\left[\sum_i X(t_i)(M(t_{i+1}) - M(t_i)) \mid \mathcal{F}_s\right]$$

$$= \sum_i E[X(t_i)E[M(t_{i+1}) - M(t_i) \mid \mathcal{F}_{t_i}] \mid \mathcal{F}_s] = 0$$

and

$$(2.10) \quad E\left[\int_0^t X \, dM \int_0^t Y \, dN - \int_0^t XY \, d\langle M, N \rangle \right.$$

$$\quad \left. - \left\{ \int_0^s X \, dM \int_0^s Y \, dN - \int_0^s XY \, d\langle M, N \rangle \right\} \mid \mathcal{F}_s\right]$$

$$= E\left[\left(\int_s^t X \, dM\right)\left(\int_s^t Y \, dN\right) - \int_s^t XY \, d\langle M, N \rangle \mid \mathcal{F}_s\right]$$

$$= E\left[\sum_i \sum_j X(t_i)Y(t_j)(M(t_{i+1}) - M(t_i))(N(t_{j+1}) - N(t_j)) \right.$$

$$\quad \left. - \sum_i X(t_i)Y(t_i)(\langle M, N \rangle_{t_{i+1}} - \langle M, N \rangle_{t_i}) \mid \mathcal{F}_s\right]$$

$$= 0,$$

where sums over i range over $\{i \geq 0: t_i \geq s, t_{i+1} \leq t\}$, and similarly for sums over j . The final equality in (2.10) follows by conditioning the (i, j) th term in the first sum on $\mathcal{F}_{t_{i \wedge t_j}}$ and the i th term in the second sum on \mathcal{F}_{t_i} (as in (2.9)). This gives (2.7) and, as a special case, (2.6). \square

To define (2.2) more generally, we need the following approximation result.

2.2 Lemma If $M \in \mathcal{M}_c$ and $X \in L^2(\langle M \rangle)$, then there exists a sequence $\{X_n\} \subset S$ such that

$$(2.11) \quad \lim_{n \rightarrow \infty} E\left[\int_0^t (X_n - X)^2 \, d\langle M \rangle\right] = 0, \quad t \geq 0.$$

Proof. By the dominated convergence theorem, (2.11) holds with

$$(2.12) \quad X_n(t) = X(t)\chi_{[t-n, n]}(X(t)),$$

which for each n is bounded and progressive.

Thus we can assume that X is bounded. We claim that (2.11) then holds by the dominated convergence theorem with

$$(2.13) \quad X_n(t) = \{ \langle M \rangle_t - \langle M \rangle_{t-n^{-1} \wedge t} + n^{-1} \}^{-1} \int_{t-n^{-1} \wedge t}^t X(u) d(\langle M \rangle_u + u),$$

which for each n is bounded, adapted, and continuous. Here we use the fact that if $h \in B[0, \infty)$ and μ is a positive Borel measure on $[0, \infty)$ without atoms such that $0 < \mu((s, t]) < \infty$ whenever $0 \leq s < t < \infty$, then

$$(2.14) \quad \lim_{\epsilon \rightarrow 0^+} \mu((t - \epsilon \wedge t, t])^{-1} \int_{t-\epsilon \wedge t}^t h \, d\mu = h(t) \quad \mu\text{-a.e.}$$

Of course, this is well known when μ is Lebesgue measure, in which case ϵ is allowed to depend on t . In the general case, it suffices to write the left side of (2.14) as

$$(2.15) \quad \lim_{\epsilon \rightarrow 0^+} (F(t) - F(t - \epsilon \wedge t))^{-1} \int_{F(t-\epsilon \wedge t)}^{F(t)} h(F^{-1}(u)) \, du,$$

where $F(t) \equiv \mu((0, t])$, and to apply the Lebesgue case.

Thus, we can assume that X is bounded and continuous. It then follows that (2.11) holds with

$$(2.16) \quad X_n(t) = X\left(\frac{\lfloor nt \rfloor}{n}\right),$$

which for each n belongs to S . □

The following result defines the stochastic integral (2.2) for each $M \in \mathcal{M}_c$ and $X \in L^2(\langle M \rangle)$.

2.3 Theorem Let $M \in \mathcal{M}_c$ and $X \in L^2(\langle M \rangle)$. Then there exists a unique (up to indistinguishability) process $\int_0^t X \, dM \in \mathcal{M}_c$ such that whenever $\{X_n\} \subset S$ satisfies

$$(2.17) \quad \sum_n \left\{ E \left[\int_0^n (X_n - X)^2 \, d\langle M \rangle \right] \right\}^{1/2} < \infty,$$

we have

$$(2.18) \quad \sup_{0 \leq t \leq T} \left| \int_0^t X_n \, dM - \int_0^t X \, dM \right| \rightarrow 0, \quad T > 0,$$

a.s. and in $L^2(P)$ as $n \rightarrow \infty$. Moreover, (2.6) holds, and

$$(2.19) \quad E \left[\left(\int_0^t X \, dM \right)^2 \right] = E \left[\int_0^t X^2 \, d\langle M \rangle \right].$$

If, in addition, $N \in \mathcal{M}_c$ and $Y \in L^2(\langle N \rangle)$, then

$$(2.20) \quad \int_0^t |XY| d\langle M, N \rangle \leq \left\{ \int_0^t X^2 d\langle M \rangle \int_0^t Y^2 d\langle N \rangle \right\}^{1/2}$$

for all $t \geq 0$, and (2.7) holds.

Proof. Choose $\{X_n\} \subset S$ satisfying (2.17); such a sequence exists by Lemma 2.2. Then, for each $T > 0$,

$$\begin{aligned} (2.21) \quad & E \left[\sum_n \sup_{0 \leq t \leq T} \left| \int_0^t X_{n+1} dM - \int_0^t X_n dM \right| \right] \\ & \leq \sum_n \left\{ E \left[\sup_{0 \leq t \leq T} \left| \int_0^t (X_{n+1} - X_n) dM \right|^2 \right] \right\}^{1/2} \\ & \leq \sum_n \left\{ 4E \left[\left| \int_0^T (X_{n+1} - X_n) dM \right|^2 \right] \right\}^{1/2} \\ & = 2 \sum_n \left\{ E \left[\int_0^T (X_{n+1} - X_n)^2 d\langle M \rangle \right] \right\}^{1/2} < \infty \end{aligned}$$

by Proposition 2.16 of Chapter 2 and by Lemma 2.1 of this Chapter. In particular, the sum inside the expectation on the left side of (2.21) is finite a.s. for every $T > 0$, implying that there exists $A \in \mathcal{F}$ with $P(A) = 0$ such that, for every $\omega \in \Omega$, $\{\chi_A, \int_0^t X_n dM\}$ converges uniformly on bounded time intervals as $n \rightarrow \infty$. By Lemma 2.1, the limiting process, which we denote by $\int_0^t X dM$, is continuous, square-integrable, and adapted. (Note that $A \in \mathcal{F}_0$ by the assumption on $\{\mathcal{F}_t\}$ made at the beginning of this section.) Clearly, (2.18) holds a.s.

Moreover,

$$\begin{aligned} (2.22) \quad & E \left[\sup_{0 \leq t \leq T} \left| \int_0^t X_n dM - \int_0^t X dM \right|^2 \right] \\ & = E \left[\liminf_{m \rightarrow \infty} \sup_{0 \leq t \leq T} \left| \int_0^t X_n dM - \int_0^t X_m dM \right|^2 \right] \\ & \leq \liminf_{m \rightarrow \infty} E \left[\sup_{0 \leq t \leq T} \left| \int_0^t X_n dM - \int_0^t X_m dM \right|^2 \right] \\ & \leq \liminf_{m \rightarrow \infty} 4E \left[\left| \int_0^T (X_n - X_m) dM \right|^2 \right] \\ & = \liminf_{m \rightarrow \infty} 4E \left[\int_0^T (X_n - X_m)^2 d\langle M \rangle \right] \\ & = 4E \left[\int_0^T (X_n - X)^2 d\langle M \rangle \right] \end{aligned}$$

for each $T > 0$ and each n , so (2.18) holds in $L^2(P)$. If $\{X'_n\} \subset S$ also satisfies (2.7) (with X_n replaced by X'_n), then

$$(2.23) \quad E \left[\sup_{0 \leq t \leq T} \left| \int_0^t X_n dM - \int_0^t X'_n dM \right|^2 \right] \\ \leq 4E \left[\int_0^T (X_n - X'_n)^2 d\langle M \rangle \right]$$

for each $T > 0$ and each n . Together with (2.22) this implies the uniqueness (up to indistinguishability) of $\int_0^t X dM$.

To show that $\int_0^t X dM$ belongs to \mathcal{M}_c and satisfies (2.7), it is enough to check that

$$(2.24a) \quad E \left[\int_s^t X dM \middle| \mathcal{F}_s \right] = 0$$

and

$$(2.24b) \quad E \left[\left(\int_s^t X dM \right)^2 \middle| \mathcal{F}_s \right] = E \left[\int_s^t X^2 d\langle M \rangle \middle| \mathcal{F}_s \right]$$

whenever $t \geq s \geq 0$, where we use the notation (2.8). But these follow immediately from the fact that they hold with X replaced by X_n and the fact that (2.18) holds in $L^2(P)$.

Suppose, in addition, that $N \in \mathcal{M}_c$ and $Y \in L^2(\langle N \rangle)$, fix $t \geq 0$, and let $\Gamma \in \mathcal{B}[0, t]$. Since $\langle M + \alpha N \rangle = \langle M \rangle + 2\alpha \langle M, N \rangle + \alpha^2 \langle N \rangle$,

$$(2.25) \quad \int_s^t \chi_\Gamma d\langle M \rangle + 2\alpha \int_s^t \chi_\Gamma d\langle M, N \rangle + \alpha^2 \int_s^t \chi_\Gamma d\langle N \rangle \geq 0, \quad 0 \leq s \leq t,$$

for all $\alpha \in \mathbb{R}$, and hence

$$(2.26) \quad \left| \int_s^t \chi_\Gamma d\langle M, N \rangle \right| \leq \left\{ \int_s^t \chi_\Gamma d\langle M \rangle \int_s^t \chi_\Gamma d\langle N \rangle \right\}^{1/2}, \quad 0 \leq s \leq t.$$

From (2.26) and the Schwarz inequality, we readily obtain (2.20) in the case in which X and Y are simple functions (that is, linear combinations of indicator functions). A standard approximation procedure then gives (2.20) in general.

To complete the proof, we must check that

$$(2.27) \quad E \left[\left(\int_s^t X dM \right) \left(\int_s^t Y dN \right) - \int_s^t XY d\langle M, N \rangle \middle| \mathcal{F}_s \right] = 0$$

for all $t \geq s \geq 0$. Let $\{Y_n\} \subset S$ be chosen by analogy with (2.17). Then by (2.10), (2.27) holds with X and Y replaced by X_n and Y_n . Applying (2.18) (in $L^2(P)$) and its analogue for Y as well as (2.20) with XY replaced by $(X_n - X)Y_n$ and by $X(Y_n - Y)$ (which sum to $X_n Y_n - XY$), we obtain the desired result by passing to the limit. \square

Before considering further generalizations, we state two simple lemmas. For the first one, given $M \in \mathcal{M}_c$, and a stopping time τ , define M^τ and $\langle M \rangle^\tau$ by

$$(2.28) \quad M^\tau(t) = M(t \wedge \tau), \quad \langle M \rangle_t^\tau = \langle M \rangle_{t \wedge \tau}, \quad t \geq 0,$$

and observe that $M^\tau \in \mathcal{M}_c$ and

$$(2.29) \quad \langle M^\tau \rangle = \langle M \rangle^\tau.$$

2.4 Lemma If $M \in \mathcal{M}_c$, $X \in L^2(\langle M \rangle)$, and τ is a stopping time, then

$$(2.30) \quad \int_0^t X \, dM^\tau = \int_0^{t \wedge \tau} X \, dM, \quad t \geq 0.$$

Proof. Fix $t \geq 0$. Observe first that

$$(2.31) \quad E\left[\left(\int_0^t X \, dM^\tau\right)^2\right] = E\left[\int_0^t X^2 \, d\langle M^\tau \rangle\right] = E\left[\int_0^{t \wedge \tau} X^2 \, d\langle M \rangle\right]$$

by Theorem 2.3 and (2.29). Second,

$$(2.32) \quad E\left[\left(\int_0^{t \wedge \tau} X \, dM\right)^2\right] = E\left[\int_0^{t \wedge \tau} X^2 \, d\langle M \rangle\right],$$

also by Theorem 2.3. Finally,

$$\begin{aligned} (2.33) \quad E\left[\left(\int_0^t X \, dM^\tau\right)\left(\int_0^{t \wedge \tau} X \, dM\right)\right] &= E\left[\left(\int_0^{t \wedge \tau} X \, dM^\tau\right)\left(\int_0^{t \wedge \tau} X \, dM\right)\right] \\ &= E\left[\int_0^{t \wedge \tau} X^2 \, d\langle M^\tau, M \rangle\right] = E\left[\int_0^{t \wedge \tau} X^2 \, d\langle M \rangle\right], \end{aligned}$$

where the first equality is obtained by conditioning on $\mathcal{F}_{t \wedge \tau}$, the second depends on (2.7), and the third follows from the fact that $\langle M^\tau, M \rangle_{t \wedge \tau} = \langle M \rangle_{t \wedge \tau}$. We conclude that

$$(2.34) \quad E\left[\left(\int_0^t X \, dM^\tau - \int_0^{t \wedge \tau} X \, dM\right)^2\right] = 0,$$

which suffices for the proof. \square

2.5 Lemma If $M \in \mathcal{M}_c$, X is progressive, $Y \in L^2(\langle M \rangle)$, and $XY \in L^2(\langle M \rangle)$, then $X \in L^2(\int_0^t Y dM)$ and

$$(2.35) \quad \int_0^t X d\left(\int_0^\cdot Y dM\right) = \int_0^t XY dM, \quad t \geq 0.$$

Proof. See Problem 11. □

The integrability assumptions of the preceding results can be removed by extending the definition of the stochastic integral (2.2). Let $\mathcal{M}_{c, loc}$ be the space of continuous local martingales M with $M(0) = 0$ a.s. Given $M \in \mathcal{M}_{c, loc}$, denote its increasing process (see Chapter 2, Section 6) by $\langle M \rangle$, and let $L^2_{loc}(\langle M \rangle)$ be the space of all real-valued progressive processes X such that

$$(2.36) \quad \int_0^t X^2 d\langle M \rangle < \infty \text{ a.s.,} \quad t \geq 0.$$

If τ is a stopping time, define M^τ and $\langle M \rangle^\tau$ by (2.28), and observe that $M^\tau \in \mathcal{M}_{c, loc}$ and (2.29) holds.

2.6 Theorem Let $M \in \mathcal{M}_{c, loc}$ and $X \in L^2_{loc}(\langle M \rangle)$. Then there exists a unique (up to indistinguishability) process $\int_0^\cdot X dM \in \mathcal{M}_{c, loc}$ such that whenever τ is a stopping time satisfying $M^\tau \in \mathcal{M}_c$ and $X \in L^2(\langle M^\tau \rangle)$, we have

$$(2.37) \quad \int_0^{t \wedge \tau} X dM = \int_0^t X dM^\tau, \quad t \geq 0.$$

(The right side of (2.37) is defined in Theorem 2.3.) Moreover, (2.6) holds. If, in addition, $N \in \mathcal{M}_{c, loc}$ and $Y \in L^2_{loc}(\langle N \rangle)$, then (2.20) and (2.7) hold.

Proof. Given $M \in \mathcal{M}_{c, loc}$ and $X \in L^2_{loc}(\langle M \rangle)$, there exist stopping times $\tau_1 \leq \tau_2 \leq \dots$ with $\tau_n \rightarrow \infty$ such that for each $n \geq 1$, $M^{\tau_n} \in \mathcal{M}_c$ and $\int_0^\cdot X^2 d\langle M \rangle \leq n$, implying

$$(2.38) \quad E\left[\int_0^t X^2 d\langle M^{\tau_n} \rangle\right] = E\left[\int_0^{t \wedge \tau_n} X^2 d\langle M \rangle\right] \leq n, \quad t \geq 0,$$

and hence $X \in L^2(\langle M^{\tau_n} \rangle)$. By Lemma 2.4,

$$(2.39) \quad \int_0^{t \wedge \tau_n} X dM^{\tau_n} = \int_0^t X dM^{\tau_n \wedge \tau_n} = \int_0^{t \wedge \tau_n} X dM^{\tau_n}, \quad t \geq 0,$$

for all $m, n \geq 1$, and existence and uniqueness of $\int_0^\cdot X dM$ follow. The conclusions (2.6), (2.20), and (2.7) follow easily from Theorem 2.3 and (2.37). □

We need the analogues of Lemmas 2.4 and 2.5. The extended lemmas are immediate consequences of the earlier ones and (2.37).

2.7 Lemma If $M \in \mathcal{M}_{c, loc}$, $X \in L^2_{loc}(\langle M \rangle)$, and τ is a stopping time, then (2.30) holds.

2.8 Lemma If $M \in \mathcal{M}_{c, loc}$, X is progressive, $Y \in L^2_{loc}(\langle M \rangle)$, and $XY \in L^2_{loc}(\langle \int_0^\cdot Y dM \rangle)$, then $X \in L^2_{loc}(\langle \int_0^\cdot Y dM \rangle)$ and (2.35) holds.

The next result is known as Itô's formula.

2.9 Theorem For $i = 1, \dots, d$, let V_i be a real-valued, continuous, adapted process of bounded variation on bounded intervals with $V_i(0) = 0$, let $M_i \in \mathcal{M}_{c, loc}$, and suppose that X_i is a real-valued process such that $X_i(0)$ is \mathcal{F}_0 -measurable and

$$(2.40) \quad X_i(t) = X_i(0) + V_i(t) + M_i(t), \quad t \geq 0.$$

Put $X = (X_1, \dots, X_d)$ and let $f \in C^{1,2}([0, \infty) \times \mathbb{R}^d)$, that is, f, f_t, f_{x_i} , and $f_{x_i x_j}$ exist and belong to $C([0, \infty) \times \mathbb{R}^d)$ for $i, j = 1, \dots, d$. Then

$$(2.41) \quad \begin{aligned} f(t, X(t)) - f(0, X(0)) \\ = & \int_0^t f_t(s, X(s)) ds + \sum_{i=1}^d \int_0^t f_{x_i}(s, X(s)) dV_i(s) \\ & + \sum_{i=1}^d \int_0^t f_{x_i}(s, X(s)) dM_i(s) \\ & + \frac{1}{2} \sum_{i,j=1}^d \int_0^t f_{x_i x_j}(s, X(s)) d\langle M_i, M_j \rangle_s, \quad t \geq 0. \end{aligned}$$

2.10 Remark Itô's formula (2.41) is often written in the easy-to-remember form

$$(2.42) \quad \begin{aligned} df(t, X(t)) = & f_t(t, X(t)) dt + \sum_{i=1}^d f_{x_i}(t, X(t)) dX_i(t) \\ & + \frac{1}{2} \sum_{i,j=1}^d f_{x_i x_j}(t, X(t)) dX_i(t) dX_j(t), \end{aligned}$$

where $dX_i(t) = dV_i(t) + dM_i(t)$ and $dX_i(t) dX_j(t)$ is evaluated using the "multiplication table"

	$dV_j(t)$	$dM_j(t)$
$dV_i(t)$	0	0
$dM_i(t)$	0	$d\langle M_i, M_j \rangle_t$

□

Proof. Denoting by $|V_i|(t)$ the total variation of V_i on $[0, t]$, let

$$(2.44) \quad \tau_n = \inf \left\{ t \geq 0 : \max_{1 \leq i \leq d} (|X_i(0)| + |V_i|(t) + |M_i(t)|) \geq n \right\}$$

for each n , and note that $\tau_n \rightarrow \infty$. Thus it suffices to verify (2.41) with t replaced by $t \wedge \tau_n$ for each n . But this is equivalent (by Lemma 2.7) to proving (2.41) with X_i , V_i , and M_i replaced by X_i^n , V_i^n , and M_i^n for each n . We conclude therefore that it involves no loss of generality to assume that $X_i(t)$, $|V_i|(t)$, $M_i(t)$, and $\langle M_i \rangle_t$ are uniformly bounded in $t \geq 0$, $\omega \in \Omega$, and $i = 1, \dots, d$.

With this assumption, we can require that f have compact support. Fix $t \geq 0$, and let $0 = t_0 < t_1 < \dots < t_m = t$ be a partition of $[0, t]$. For the remainder of the proof, we use the notation that for a given process Y (real- or \mathbb{R}^d -valued), $\Delta_k Y = Y(t_{k+1}) - Y(t_k)$ for $k = 0, \dots, m-1$. By Taylor's theorem,

$$\begin{aligned}
 (2.45) \quad & f(t, X(t)) - f(0, X(0)) \\
 &= \sum_{k=0}^{m-1} \{f(t_{k+1}, X(t_{k+1})) - f(t_k, X(t_{k+1}))\} \\
 &\quad + \sum_{k=0}^{m-1} \{f(t_k, X(t_{k+1})) - f(t_k, X(t_k))\} \\
 &= \sum_{k=0}^{m-1} \int_{t_k}^{t_{k+1}} f_u(u, X(t_{k+1})) du \\
 &\quad + \sum_{i=1}^d \sum_{k=0}^{m-1} f_{x_i}(t_k, X(t_k)) \Delta_k X_i \\
 &\quad + \frac{1}{2} \sum_{i,j=1}^d \sum_{k=0}^{m-1} f_{x_i x_j}(t_k, \xi_k) \Delta_k X_i \Delta_k X_j,
 \end{aligned}$$

where $|\xi_k - X(t_k)| \leq |X(t_{k+1}) - X(t_k)|$.

The proof now consists of showing that, as the mesh of the partition tends to zero, the right side of (2.45) converges in probability to the right side of (2.41).

Convergence of the sum of the first two terms in (2.45) to the sum of the first three terms in (2.41) is straightforward (see Problem 12).

Note that by Proposition 3.4 of Chapter 2 and the continuity and bounded variation of the V_i ,

$$\begin{aligned}
 (2.46) \quad & \lim_{\max(t_{k+1} - t_k) \rightarrow 0} \sum_k \Delta_k X_i \Delta_k X_j \\
 &= \lim_{\max(t_{k+1} - t_k) \rightarrow 0} \sum_k \Delta_k M_i \Delta_k M_j \\
 &= \langle M_i, M_j \rangle_t
 \end{aligned}$$

in probability, and that

$$(2.47) \quad \lim_{\max(t_{k+1} - t_k) \rightarrow 0} \max_k |f_{x_i x_j}(t_k, \xi_k) - f_{x_i x_j}(t_k, X(t_k))| = 0.$$

Observing that $2 \Delta_k X_i \Delta_k X_j = (\Delta_k X_i + \Delta_k X_j)^2 - (\Delta_k X_i)^2 - (\Delta_k X_j)^2$ (i.e., $\Delta_k X_i \Delta_k X_j$ is a linear combination of positive quantities), the convergence of

the last term in (2.45) to the last term in (2.41) is a consequence of the following lemma. \square

2.11 Lemma Let f be continuous, and let F be nonnegative, nondecreasing, and right continuous on $[0, \infty)$. For $n = 1, 2, \dots$, let $0 = t_0^n < t_1^n < t_2^n < \dots$, with $t_k^n \rightarrow \infty$. Suppose for each $t > 0$ that $\max_{t_k^n < t} (t_{k+1}^n - t_k^n) \rightarrow 0$ as $n \rightarrow \infty$, and suppose that f_n and a_n satisfy $a_n \geq 0$,

$$(2.48) \quad \lim_{n \rightarrow \infty} \max_{t_k^n \leq t} |f_n(t_k^n) - f(t_k^n)| = 0, \quad t \geq 0,$$

and

$$(2.49) \quad \lim_{n \rightarrow \infty} \sum_{t_k^n \leq t} a_n(t_k^n) = F(t)$$

for each t at which F is continuous. Then

$$(2.50) \quad \lim_{n \rightarrow \infty} \sum_{t_k^n \leq t} f_n(t_k^n) a_n(t_k^n) = \int_0^t f(s) dF(s)$$

for each t at which F is continuous.

Proof. Clearly

$$(2.51) \quad \lim_{n \rightarrow \infty} \left[\sum_{t_k^n \leq t} f_n(t_k^n) a_n(t_k^n) - \sum_{t_k^n \leq t} f(t_k^n) a_n(t_k^n) \right] = 0.$$

Suppose t is a continuity point of F and $F(t) > 0$. Let μ_n and μ be the probability measures on $[0, t]$ given by

$$(2.52) \quad \mu_n[0, s] = \frac{\sum_{t_k^n \leq s} a_n(t_k^n)}{\sum_{t_k^n \leq t} a_n(t_k^n)}, \quad 0 \leq s \leq t,$$

and $\mu[0, s] = F(s)/F(t)$, $0 \leq s \leq t$. Then $\mu_n \Rightarrow \mu$, and hence

$$\begin{aligned} (2.53) \quad & \lim_{n \rightarrow \infty} \sum_{t_k^n \leq t} f(t_k^n) a_n(t_k^n) \\ &= \lim_{n \rightarrow \infty} \left[\sum_{t_k^n \leq t} a_n(t_k^n) \right] \int_0^t f d\mu_n \\ &= F(t) \int_0^t f d\mu \\ &= \int_0^t f dF. \end{aligned}$$

\square

We conclude this section by applying Itô's formula to give an important characterization of Brownian motion, which is essentially the converse of Theorem 1.2(b).

2.12 Theorem Suppose that $X_1, \dots, X_d \in \mathcal{M}_{c, loc}$ satisfy $\langle X_i, X_j \rangle_t = \delta_{ij}t$ for $i, j = 1, \dots, d$ and all $t \geq 0$. Then $X = (X_1, \dots, X_d)$ is a d -dimensional Brownian motion.

Proof. Let $\theta \in \mathbb{R}^d$ be arbitrary, and define $f: [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{C}$ by

$$(2.54) \quad f(t, x) = \exp \{i\theta \cdot x + \frac{1}{2}|\theta|^2 t\},$$

where $i = \sqrt{-1}$. By Theorem 2.9, $\{f(t, X(t)), t \geq 0\}$ is a complex-valued, continuous, local martingale, bounded on bounded time intervals, so

$$(2.55) \quad E[f(t, X(t)) | \mathcal{F}_s] = f(s, X(s))$$

for all $t \geq s \geq 0$, that is,

$$(2.56) \quad E[\exp \{i\theta \cdot (X(t) - X(s))\} | \mathcal{F}_s] = \exp \{-\frac{1}{2}|\theta|^2(t-s)\}.$$

Consequently, X is a d -dimensional Brownian motion. \square

3. STOCHASTIC INTEGRAL EQUATIONS

Let $\sigma: [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}^d \otimes \mathbb{R}^d$ (the space of real, $d \times d$ matrices) and $b: [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ be locally bounded (i.e., bounded on each compact set) and Borel measurable. In this section we consider the stochastic integral equation

$$(3.1) \quad X(t) = X(0) + \int_0^t \sigma(s, X(s)) dW(s) + \int_0^t b(s, X(s)) ds, \quad t \geq 0,$$

where W is a d -dimensional Brownian motion independent of $X(0)$ and $\int_0^t \sigma(s, X(s)) dW(s)$ denotes the \mathbb{R}^d -valued process whose i th component is given by

$$(3.2) \quad \sum_{j=1}^d \int_0^t \sigma_{ij}(s, X(s)) dW_j(s).$$

Observe that (3.2) is well-defined (and is a continuous, local martingale) if X is a continuous, \mathbb{R}^d -valued, $\{\mathcal{F}_t\}$ -adapted process, where $\mathcal{F}_t = \mathcal{F}_t^W \vee \sigma(X(0))$. In the classical approach of Itô, W and $X(0)$ are given, and one seeks such a solution X . For our purposes, however, it is convenient to regard W as part of the solution and to allow $\{\mathcal{F}_t\}$ to be an arbitrary filtration with respect to which W is a d -dimensional Brownian motion.

Let $\mu \in \mathcal{P}(\mathbb{R}^d)$. We say that $(\Omega, \mathcal{F}, P, \{\mathcal{F}_t\}, W, X)$ is a solution of the stochastic integral equation corresponding to (σ, b, μ) (respectively, (σ, b)) if:

- (a) (Ω, \mathcal{F}, P) is a probability space with a filtration $\{\mathcal{F}_t\}$, and W and X are \mathbb{R}^d -valued processes on (Ω, \mathcal{F}, P) .
- (b) W is a d -dimensional $\{\mathcal{F}_t\}$ -Brownian motion.
- (c) X is $\{\mathcal{F}_t\}$ -adapted, where \mathcal{F}_t denotes the P -completion of \mathcal{F}_{t_0} , that is, the smallest σ -algebra containing $\mathcal{F}_{t_0} \cup \{A \subset \Omega : A \subset N \text{ for some } N \in \mathcal{F} \text{ with } P(N) = 0\}$.
- (d) $PX(0)^{-1} = \mu$ and X has sample paths in $C_{\mathbb{R}^d}[0, \infty)$ (respectively, X has sample paths in $C_{\mathbb{R}^d}[0, \infty)$).
- (e) (3.1) holds a.s.

The definition of the stochastic integral is as in Theorem 2.6, the roles of (Ω, \mathcal{F}, P) and $\{\mathcal{F}_t\}$ being played by (Ω, \mathcal{F}, P) and $\{\mathcal{F}_t\}$. Because it is defined only up to indistinguishability, we continue our convention of suppressing the phrase "almost surely" whenever it is needed only because of this indeterminacy.

There are two types of uniqueness of solutions of (3.1) that are considered:

Pathwise uniqueness is said to hold for solutions of the stochastic integral equation corresponding to (σ, b, μ) if, whenever $(\Omega, \mathcal{F}, P, \{\mathcal{F}_t\}, W, X)$ and $(\Omega, \mathcal{F}, P, \{\mathcal{F}_t\}, W, X')$ are solutions (with the same probability space, filtration, and Brownian motion), $P\{X(0) = X'(0)\} = 1$ implies $P\{X(t) = X'(t) \text{ for all } t \geq 0\} = 1$.

Distribution uniqueness is said to hold for solutions of the stochastic integral equation corresponding to (σ, b, μ) if, whenever $(\Omega, \mathcal{F}, P, \{\mathcal{F}_t\}, W, X)$ and $(\Omega', \mathcal{F}', P', \{\mathcal{F}'_t\}, W', X')$ are solutions (not necessarily with the same probability space), we have $PX^{-1} = P'(X')^{-1}$ (as elements of $\mathcal{P}(C_{\mathbb{R}^d}[0, \infty))$).

Let

$$(3.3) \quad A = \{(f, Gf) : f \in C_c^\infty(\mathbb{R}^d)\},$$

where

$$(3.4) \quad G = \frac{1}{2} \sum_{i,j=1}^d a_{ij}(t, x) \partial_i \partial_j + \sum_{i=1}^d b_i(t, x) \partial_i$$

and

$$(3.5) \quad a = \sigma \sigma^T.$$

Observe that if $(\Omega, \mathcal{F}, P, \{\mathcal{F}_t\}, W, X)$ is a solution of the stochastic integral equation corresponding to (σ, b, μ) , then, by Itô's formula (Theorem 2.9),

(3.6)

$$f(X(t)) - \int_0^t Gf(s, X(s)) ds = f(X(0)) + \sum_{i, j=1}^d \int_0^t f_{x_i}(X(s)) \sigma_{ij}(s, X(s)) dW_j(s)$$

for all $t \geq 0$ and $f \in C_c^\infty(\mathbb{R}^d)$, so X is a solution of the $C_{\mathbb{R}^d}[0, \infty)$ martingale problem for (A, μ) with respect to $\{\mathcal{F}_t\}$.

In what sense is the converse of this result true? We consider first the nondegenerate case.

3.1 Proposition Let $\sigma: [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}^d \otimes \mathbb{R}^d$ and $b: [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ be locally bounded and Borel measurable, and let $\mu \in \mathcal{P}(\mathbb{R}^d)$. Suppose $\sigma(t, x)$ is nonsingular for each $(t, x) \in [0, \infty) \times \mathbb{R}^d$ and σ^{-1} is locally bounded. Define A by (3.3)–(3.5). If X is a solution of the $C_{\mathbb{R}^d}[0, \infty)$ martingale problem for (A, μ) with respect to a filtration $\{\mathcal{F}_t\}$ on a probability space (Ω, \mathcal{F}, P) , then there exists a d -dimensional $\{\mathcal{F}_t\}$ -Brownian motion W such that $(\Omega, \mathcal{F}, P, \{\mathcal{F}_t\}, W, X)$ is a solution of the stochastic integral equation corresponding to (σ, b, μ) .

Proof. Since

$$(3.7) \quad f(X(t)) - f(X(0)) - \int_0^t Gf(s, X(s)) ds$$

belongs to \mathcal{M}_c for every $f \in C_c^\infty(\mathbb{R}^d)$, it follows easily that (3.7) belongs to $\mathcal{M}_{c, loc}$ for every $f \in C^\infty(\mathbb{R}^d)$. In particular,

$$(3.8) \quad M_i(t) \equiv X_i(t) - X_i(0) - \int_0^t b_i(s, X(s)) ds$$

belongs to $\mathcal{M}_{c, loc}$ and, by (3.6) and Theorem 2.6 (see (2.7)),

$$(3.9) \quad \langle M_i, M_j \rangle_t = \int_0^t a_{ij}(s, X(s)) ds, \quad t \geq 0.$$

We claim that

$$(3.10) \quad W(t) \equiv \int_0^t \sigma^{-1}(s, X(s)) dM(s)$$

defines a d -dimensional $\{\mathcal{F}_t\}$ -Brownian motion. This follows from Theorem 2.12 since

$$\begin{aligned} (3.11) \quad \langle W_i, W_j \rangle_t &= \sum_{k, l=1}^d \left\langle \int_0^t (\sigma^{-1})_{ik}(s, X(s)) dM_k(s), \int_0^t (\sigma^{-1})_{jl}(s, X(s)) dM_l(s) \right\rangle, \\ &= \sum_{k, l=1}^d \int_0^t [(\sigma^{-1})_{ik} a_{kl} ((\sigma^T)^{-1})_{lj}] (s, X(s)) ds \\ &= \delta_{ij} t, \quad t \geq 0, \end{aligned}$$

where the second equality depends on Theorem 2.6. Consequently, by Lemma 2.8,

$$(3.12) \quad \int_0^t \sigma(s, X(s)) dW(s) = \int_0^t dM(t) = X(t) - X(0) - \int_0^t b(s, X(s)) ds$$

for all $t \geq 0$, which is to say that (3.1) holds. \square

We turn to the general case, in which σ may be singular. Here the conclusion of Proposition 3.1 need not hold because X may not have enough randomness in terms of which to construct a Brownian motion. However, by suitably enlarging the probability space, we can obtain a solution of (3.1).

It will be convenient to separate out a preliminary result concerning matrix-valued functions.

3.2 Lemma Let $\sigma: [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}^d \otimes \mathbb{R}^d$ be Borel measurable, and put $a = \sigma\sigma^T$. Then there exist Borel measurable functions $\rho, \eta: [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}^d \otimes \mathbb{R}^d$ such that

$$(3.13) \quad \rho a \rho^T + \eta \eta^T = I_d,$$

$$(3.14) \quad \sigma \eta = 0,$$

$$(3.15) \quad (I_d - \sigma \rho)a(I_d - \sigma \rho)^T = 0.$$

Proof. Suppose first that σ is a constant function. Since $a \in S_d$ (the set of real, symmetric, nonnegative-definite, $d \times d$ matrices), there exists a real, orthogonal matrix U (i.e., $UU^T = U^T U = I_d$) and a diagonal matrix Λ with non-negative entries such that $a = U^T \Lambda U$. Moreover, a has a unique S_d -valued square root $a^{1/2}$, and $a^{1/2} = U^T \Lambda^{1/2} U$. Let Γ be a diagonal matrix with diagonal entries 1 or 0 depending on whether the corresponding entry of Λ is positive or 0, and let Λ_1 be diagonal with $\Lambda_1 \Lambda = \Gamma$. Then, since $\sigma\sigma^T = U^T \Lambda U$, we have $(\Lambda_1^{1/2} U \sigma)(\Lambda_1^{1/2} U \sigma)^T = \Gamma$. By the Gram-Schmidt orthogonalization procedure, we can therefore construct a real, orthogonal matrix V such that $\Lambda_1^{1/2} U \sigma = \Gamma V$, and hence $\Gamma U \sigma = \Lambda^{1/2} V$. It follows that $U \sigma = \Lambda^{1/2} V$, from which we conclude $\sigma = a^{1/2} U^T V$. It is now easily checked that (3.13)-(3.15) hold with $\rho = V^T \Lambda_1^{1/2} U$ and $\eta = V^T (I_d - \Gamma) U$.

To complete the proof, it suffices to show that the measurable selection theorem (Appendix 10) is applicable to the multivalued function taking $\sigma \in \mathbb{R}^d \otimes \mathbb{R}^d$ to the set of pairs (U, V) as above. The details of this step are left to the reader. \square

3.3 Theorem Let $\sigma: [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}^d \otimes \mathbb{R}^d$ and $b: [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ be locally bounded and Borel measurable, and let $\mu \in \mathcal{P}(\mathbb{R}^d)$. Define A by (3.3)-(3.5), and suppose X is a solution of the $C_{\mathbb{R}}[0, \infty)$ martingale problem for (A, μ) with respect to a filtration $\{\mathcal{F}_t\}$ on a probability space (Ω, \mathcal{F}, P) . Let W' be a d -dimensional $\{\mathcal{F}'_t\}$ -Brownian motion on a probability space $(\Omega', \mathcal{F}', P')$,

P') and define $\tilde{\Omega} = \Omega \times \Omega'$, $\tilde{\mathcal{F}} = \overline{\mathcal{F} \times \mathcal{F}'}$, $\tilde{P} = P \times P'$, $\tilde{\mathcal{F}}_t = \overline{\mathcal{F}_t \times \mathcal{F}'_t}$, and $\tilde{X}(t, \omega, \omega') = X(t, \omega)$. Then there exists a d -dimensional $\{\tilde{\mathcal{F}}_t\}$ -Brownian motion \tilde{W} such that $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P}, \{\tilde{\mathcal{F}}_t\}, \tilde{W}, \tilde{X})$ is a solution of the stochastic integral equation corresponding to (σ, b, μ) .

Proof. Define M by (3.8), $\tilde{M}(t, \omega, \omega') \equiv M(t, \omega)$, and $\tilde{W}'(t, \omega, \omega') \equiv W'(t, \omega')$. Using the notation of Lemma 3.2, we claim that

$$(3.16) \quad \tilde{W}(t) \equiv \int_0^t \rho(s, \tilde{X}(s)) d\tilde{M}(s) + \int_0^t \eta(s, \tilde{X}(s)) d\tilde{W}'(s)$$

defines a d -dimensional $\{\tilde{\mathcal{F}}_t\}$ -Brownian motion. Again, this is a consequence of Theorem 2.12 since

$$\begin{aligned} (3.17) \quad \langle \tilde{W}_i, \tilde{W}_j \rangle_t &= \sum_{k, l=1}^d \left\langle \int_0^t \rho_{ik}(s, \tilde{X}(s)) d\tilde{M}_k(s), \int_0^t \rho_{jl}(s, \tilde{X}(s)) d\tilde{M}_l(s) \right\rangle, \\ &\quad + \sum_{k, l=1}^d \left\langle \int_0^t \eta_{ik}(s, \tilde{X}(s)) d\tilde{W}'_k(s), \int_0^t \eta_{jl}(s, \tilde{X}(s)) d\tilde{W}'_l(s) \right\rangle, \\ &= \sum_{k, l=1}^d \int_0^t (\rho_{ik} a_{kl} \rho_{lj})(s, \tilde{X}(s)) ds \\ &\quad + \sum_{k, l=1}^d \int_0^t (\eta_{ik} \delta_{kl} \eta_{jl})(s, \tilde{X}(s)) ds \\ &= \int_0^t (\rho \rho^T + \eta \eta^T)_{ij}(s, \tilde{X}(s)) ds \\ &= \delta_{ij} t, \quad t \geq 0, \end{aligned}$$

where the first equality uses the fact that $\langle \tilde{M}_k, \tilde{W}_l \rangle = 0$ for $k, l = 1, \dots, d$, the second depends on Theorem 2.6, and the fourth on (3.13). By Lemma 2.8, (3.14), and (3.15),

$$\begin{aligned} (3.18) \quad \int_0^t \sigma(s, \tilde{X}(s)) d\tilde{W}(s) &= \int_0^t (\sigma \rho)(s, \tilde{X}(s)) d\tilde{M}(s) \\ &\quad + \int_0^t (\sigma \eta)(s, \tilde{X}(s)) d\tilde{W}'(s) \\ &= \int_0^t (\sigma \rho)(s, \tilde{X}(s)) d\tilde{M}(s) \\ &= \tilde{M}(t) - \int_0^t (I_d - \sigma \rho)(s, \tilde{X}(s)) d\tilde{M}(s) \\ &= \tilde{M}(t) \\ &= \tilde{X}(t) - \tilde{X}(0) - \int_0^t b(s, \tilde{X}(s)) ds, \quad t \geq 0, \end{aligned}$$

where the next-to-last equality is a consequence of (3.15) and

$$(3.19) \quad \begin{aligned} & \left\langle \sum_{j=1}^d \int_0^t (I_d - \sigma\rho)_j(s, X(s)) d\tilde{M}_j(s) \right\rangle_t \\ &= \sum_{k, l=1}^d \int_0^t [(I_d - \sigma\rho)_{ik} a_{kl} (I_d - \sigma\rho)_{jl}](s, X(s)) ds \\ &= 0, \end{aligned}$$

and the desired result follows. \square

3.4 Corollary Let $\sigma: [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}^d \otimes \mathbb{R}^d$ and $b: [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ be locally bounded and Borel measurable, and let $\mu \in \mathcal{P}(\mathbb{R}^d)$. Define A by (3.3)–(3.5). Then there exists a solution of the stochastic integral equation corresponding to (σ, b, μ) if and only if there exists a solution of the $C_{\mathbb{R}^d}[0, \infty)$ martingale problem for (A, μ) . Moreover, distribution uniqueness holds for solutions of the stochastic integral equation corresponding to (σ, b, μ) if and only if uniqueness holds for solutions of the $C_{\mathbb{R}^d}[0, \infty)$ martingale problem for (A, μ) .

3.5 Proposition Suppose, in addition to the hypotheses of Corollary 3.4, that there exists a constant K such that

$$(3.20) \quad |a(t, x)| \leq K(1 + |x|^2), \quad x \cdot b(t, x) \leq K(1 + |x|^2), \quad t \geq 0, x \in \mathbb{R}^d.$$

Then every solution of the martingale problem for (A, μ) has a modification with sample paths in $C_{\mathbb{R}^d}[0, \infty)$. Consequently, the phrase “ $C_{\mathbb{R}^d}[0, \infty)$ martingale problem” in the conclusions of Corollary 3.4 can be replaced by the phrase “martingale problem.”

Proof. Let X be a solution of the martingale problem for (A, μ) . By Theorem 7.1 of Chapter 4, $X^0(t) \equiv (t, X(t))$ is a solution of the martingale problem for A^0 , where

$$(3.21) \quad A^0 = \{(\gamma f, \gamma Gf + \gamma' f) : f \in C_c^\infty(\mathbb{R}^d), \gamma \in C_c^1[0, \infty)\}.$$

By Corollary 3.7 of Chapter 4, X^0 has a modification Y^0 with sample paths in $D_{([0, \infty) \times \mathbb{R}^d)}[0, \infty)$. Letting

$$(3.22) \quad (A^0)^\Delta = \{(f, g) \in C(([0, \infty) \times \mathbb{R}^d)^\Delta) \times B(([0, \infty) \times \mathbb{R}^d)^\Delta) : (f, g)|_{[0, \infty) \times \mathbb{R}^d} \in A^0, f(\Delta) = g(\Delta) = 0\},$$

$$(f, g)|_{[0, \infty) \times \mathbb{R}^d} \in A^0, f(\Delta) = g(\Delta) = 0\},$$

it follows that Y^0 is a solution of the martingale problem for $(A^0)^\Delta$. Choose $\varphi \in C_c^\infty[0, \infty)$ with $\chi_{[0, 1]} \leq \varphi \leq \chi_{[0, 2]}$ and $\varphi' \leq 0$. Then the sequence

$\{f_n, g_n\} \subset (A^0)^A$ given by $f_n(t, x) = \varphi(t/n)\varphi(|x|^2/n^2)$, $f_n(\Delta) = 0$, and $g_n = (A^0)^A f_n$ satisfies $\text{bp-lim}_{n \rightarrow \infty} f_n = \chi_{[0, \infty) \times \mathbb{R}^d}$, $g_n \rightarrow 0$ pointwise, and $\sup_n \|g_n\| < \infty$. By Proposition 3.9 of Chapter 4, Y^0 has almost all sample paths in $D_{[0, \infty) \times \mathbb{R}^d}[0, \infty)$, and therefore, by Problem 19 of Chapter 4, in $C_{[0, \infty) \times \mathbb{R}^d}[0, \infty)$. Define $\eta: [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ by $\eta(t, y) = y$. Then $\eta \circ Y^0$ is a solution of the martingale problem for (A, μ) with almost all sample paths in $C_{\mathbb{R}^d}[0, \infty)$, and the first conclusion follows.

The second conclusion is an immediate consequence of this. \square

3.6 Theorem Let $\sigma: [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}^d \otimes \mathbb{R}^d$ and $b: [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ be locally bounded and Borel measurable, and let $\mu \in \mathcal{P}(\mathbb{R}^d)$. Then pathwise uniqueness of solutions of the stochastic integral equation corresponding to (σ, b, μ) implies distribution uniqueness.

Proof. Let $(\Omega, \mathcal{F}, P, \{\mathcal{F}_t\}, W, X)$ and $(\Omega', \mathcal{F}', P', \{\mathcal{F}'_t\}, W', X')$ be two solutions. We apply Lemma 5.15 of Chapter 4 with $E = C_{\mathbb{R}^d}[0, \infty) \times \mathbb{R}^d$, $S_1 = S_2 = C_{\mathbb{R}^d}[0, \infty) \times C_{\mathbb{R}^d}[0, \infty)$, $P_1 = P(W, X)^{-1}$, $P_2 = P'(W', X')^{-1}$, and $X_1(\omega_1, \omega_2) = X_2(\omega_1, \omega_2) \equiv (\omega_1, \omega_2(0))$. Letting $\lambda = PW^{-1} = P'(W')^{-1}$, we have $P(W, X(0))^{-1} = P'(W', X'(0))^{-1} = \lambda \times \mu$. We conclude that there exists a probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$ on which are defined $C_{\mathbb{R}^d}[0, \infty)$ -valued random variables \tilde{W} , \tilde{X} , and \tilde{X}' satisfying $\tilde{P}(\tilde{W}, \tilde{X})^{-1} = P(W, X)^{-1}$, $\tilde{P}(\tilde{W}, \tilde{X}')^{-1} = P'(W', X')^{-1}$, and $\tilde{P}\{\tilde{X}(0) = \tilde{X}'(0)\} = 1$. Moreover, for all $f, g \in B(C_{\mathbb{R}^d}[0, \infty) \times C_{\mathbb{R}^d}[0, \infty))$,

$$(3.23) \quad \begin{aligned} E^P[f(\tilde{W}, \tilde{X})g(\tilde{W}, \tilde{X}')] \\ = \iint E^P[f(W, X)|(W, X(0)) = (\beta, x)]E^{P'}[g(W', X')|(W', X'(0)) \\ = (\beta, x)]\lambda(d\beta)\mu(dx). \end{aligned}$$

Let $0 \leq s_1 \leq \dots \leq s_k \leq t_1 \leq \dots \leq t_k$ and $f_{ij} \in B(E)$ for $i = 1, \dots, k$ and $j = 0, 1, 2, 3$. Then

$$(3.24) \quad \begin{aligned} E^P\left[\prod_{i=1}^k f_{i0}(\tilde{W}(t_i) - \tilde{W}(s_i))f_{i1}(\tilde{W}(s_i))f_{i2}(\tilde{X}(s_i))f_{i3}(\tilde{X}'(s_i))\right] \\ = \iint \prod_{i=1}^k f_{i0}(\beta(t_i) - \beta(s))E^P\left[\prod_{i=1}^k f_{i1}(W(s_i))f_{i2}(X(s_i)) \middle| (W, X(0)) = (\beta, x)\right] \\ \cdot E^{P'}\left[\prod_{i=1}^k f_{i3}(X'(s_i)) \middle| (W', X'(0)) = (\beta, x)\right]\lambda(d\beta)\mu(dx) \end{aligned}$$

$$\begin{aligned}
&= \int \prod_{i=1}^k f_{i0}(\beta(t_i) - \beta(s)) \lambda(d\beta) \\
&\quad \cdot \iint E^P \left[\prod_{i=1}^k f_{i1}(W(s_i)) f_{i2}(X(s_i)) \middle| (W, X(0)) = (\beta, x) \right] \\
&\quad \cdot E^P \left[\prod_{i=1}^k f_{i3}(X'(s_i)) \middle| (W', X'(0)) = (\beta, x) \right] \lambda(d\beta) \mu(dx) \\
&= E^P \left[\prod_{i=1}^k f_{i0}(\tilde{W}(t_i) - \tilde{W}(s)) \right] E^P \left[\prod_{i=1}^k f_{i1}(\tilde{W}(s_i)) f_{i2}(\tilde{X}(s_i)) f_{i3}(\tilde{X}'(s_i)) \right],
\end{aligned}$$

where the second equality depends on the fact that the two conditional expectations on its left side are functions only of $(\beta(\cdot \wedge s), x)$. It follows that \tilde{W} is a d -dimensional $\{\mathcal{F}_s\}$ -Brownian motion, where

$$(3.25) \quad \tilde{\mathcal{F}}_t = \sigma(\tilde{W}(s), \tilde{X}(s), \tilde{X}'(s); 0 \leq s \leq t),$$

and hence $(\Omega, \mathcal{F}, P, \{\mathcal{F}_s\}, \tilde{W}, \tilde{X})$ and $(\Omega, \mathcal{F}, \tilde{P}, \{\mathcal{F}_s\}, \tilde{W}, \tilde{X}')$ are solutions of the stochastic integral equation corresponding to (σ, b, μ) . By pathwise uniqueness, $\tilde{P}\{\tilde{X}(t) = \tilde{X}'(t) \text{ for all } t \geq 0\} = 1$, so $PX^{-1} = \tilde{P}\tilde{X}^{-1} = \tilde{P}(\tilde{X}')^{-1} = P(X')^{-1}$. Thus distribution uniqueness holds. \square

The next result gives sufficient conditions for pathwise uniqueness of solutions of (3.1).

3.7 Theorem Let $\sigma: [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}^d \otimes \mathbb{R}^d$ and $b: [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ be locally bounded and Borel measurable. Let $U \subset \mathbb{R}^d$ be open, let $T > 0$, and suppose that there exists a constant K such that

$$(3.26) \quad |\sigma(t, x) - \sigma(t, y)| \vee |b(t, x) - b(t, y)| \leq K|x - y|,$$

$$0 \leq t \leq T, \quad x, y \in U.$$

Given two solutions $(\Omega, \mathcal{F}, P, \{\mathcal{F}_s\}, W, X)$ and $(\Omega, \mathcal{F}, P, \{\mathcal{F}_s\}, W, Y)$ of the stochastic integral equation corresponding to (σ, b) , let

$$(3.27) \quad \tau = \inf \{t \geq 0: X(t) \notin U \text{ or } Y(t) \notin U\}.$$

Then $P\{X(0) = Y(0)\} = 1$ implies $P\{X(t \wedge \tau) = Y(t \wedge \tau) \text{ for } 0 \leq t \leq T\} = 1$.

Proof. For $0 \leq t \leq T$,

$$\begin{aligned}
 (3.28) \quad & E[|X(t \wedge \tau) - Y(t \wedge \tau)|^2] \\
 & \leq 2E\left[\left|\int_0^{t \wedge \tau} (\sigma(s, X(s)) - \sigma(s, Y(s))) dW(s)\right|^2\right] \\
 & \quad + 2E\left[\left|\int_0^{t \wedge \tau} (b(s, X(s)) - b(s, Y(s))) ds\right|^2\right] \\
 & \leq 2E\left[\int_0^{t \wedge \tau} |\sigma(s, X(s)) - \sigma(s, Y(s))|^2 ds\right] \\
 & \quad + 2tE\left[\int_0^{t \wedge \tau} |b(s, X(s)) - b(s, Y(s))|^2 ds\right] \\
 & \leq 2K^2(1+t)E\left[\int_0^{t \wedge \tau} |X(s) - Y(s)|^2 ds\right] \\
 & \leq 2K^2(1+T)E\left[\int_0^t |X(s \wedge \tau) - Y(s \wedge \tau)|^2 ds\right],
 \end{aligned}$$

and hence the desired result follows from Gronwall's inequality. \square

In particular, if $\sigma(t, x)$ and $b(t, x)$ are locally Lipschitz continuous in x , uniformly in t in bounded intervals (i.e., for every bounded open set $U \subset \mathbb{R}^d$ and $T > 0$, (3.26) holds for some K), then we have pathwise uniqueness. This condition suffices for many applications. However, in some cases, $a = \sigma\sigma^T$ is a smooth function but σ is not. In general this causes serious difficulties, but not when $d = 1$.

3.8 Theorem In the case $d = 1$, Theorem 3.7 is valid with (3.26) replaced by

$$\begin{aligned}
 (3.29) \quad & |\sigma(t, x) - \sigma(t, y)|^2 \vee |b(t, x) - b(t, y)| \leq K|x - y|, \\
 & 0 \leq t \leq T, \quad x, y \in U.
 \end{aligned}$$

3.9 Remark If $\sigma \geq 0$, (3.29) is implied by

$$\begin{aligned}
 (3.30) \quad & |\sigma^2(t, x) - \sigma^2(t, y)| \vee |b(t, x) - b(t, y)| \leq K|x - y|, \\
 & 0 \leq t \leq T, \quad x, y \in U. \quad \square
 \end{aligned}$$

Proof. For each $\varepsilon > 0$, define $\varphi_\varepsilon \in C^2(\mathbb{R})$ by $\varphi_\varepsilon(u) = (u^2 + \varepsilon)^{1/2}$ and $\psi_\varepsilon \in C(\mathbb{R})$ by $\psi_\varepsilon(u) = \varepsilon|u|/(u^2 + \varepsilon)^{3/2}$. For $0 \leq t \leq T$, we have, by Itô's formula,

$$\begin{aligned}
 (3.31) \quad & E[\varphi_\varepsilon(X(t \wedge \tau)) - Y(t \wedge \tau)] \\
 &= \varphi_\varepsilon(0) + E\left[\int_0^{t \wedge \tau} \{\varphi_\varepsilon'(X(s))b(s, X(s)) - b(s, Y(s))\right. \\
 &\quad \left. + \frac{1}{2}\varphi_\varepsilon''(X(s))(s, X(s)) - s, Y(s))^2\} ds \right] \\
 &\leq \varphi_\varepsilon(0) + E\left[\int_0^{t \wedge \tau} \{K|X(s) - Y(s)| + \frac{1}{2}K\psi_\varepsilon(X(s) - Y(s))\} ds \right].
 \end{aligned}$$

Noting that $\psi_\varepsilon(u) \leq \sup_{y \in \mathbb{R}} |y|/(y^2 + 1)^{3/2}$ for all $u \in \mathbb{R}$ and $\varepsilon > 0$, we let $\varepsilon \rightarrow 0$ and conclude from the dominated convergence theorem that

$$\begin{aligned}
 (3.32) \quad & E[|X(t \wedge \tau) - Y(t \wedge \tau)|] \leq KE\left[\int_0^{t \wedge \tau} |X(s) - Y(s)| ds \right] \\
 &\leq K \int_0^t E[|X(s \wedge \tau) - Y(s \wedge \tau)|] ds
 \end{aligned}$$

for $0 \leq t \leq T$, and the result again follows from Gronwall's inequality. \square

We turn finally to the question of existence of solutions of (3.1). We take two approaches. The first is based on Corollary 3.4 and results in Chapter 4. The second is the classical iteration method.

3.10 Theorem Let $\sigma: [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}^d \otimes \mathbb{R}^d$ and $b: [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ be continuous and satisfy

$$\begin{aligned}
 (3.33) \quad & |\sigma(t, x)|^2 \leq K(1 + |x|^2), \quad x \cdot b(t, x) \leq K(1 + |x|^2), \\
 & t \geq 0, \quad x \in \mathbb{R}^d,
 \end{aligned}$$

for some constant K , and let $\mu \in \mathcal{P}(\mathbb{R}^d)$. Then there exists a solution of the stochastic integral equation corresponding to (σ, b, μ) .

Proof. It suffices by Corollary 3.4 and Proposition 3.5 to prove the existence of a solution of the martingale problem for (A, μ) , where A is defined by (3.3)–(3.5). By Theorem 7.1 of Chapter 4 it suffices to prove the existence of a solution of the martingale problem for $(A^0, \delta_0 \times \mu)$, where A^0 is defined by (3.21). Noting that $A^0 \subset \hat{C}([0, \infty) \times \mathbb{R}^d) \times \hat{C}([0, \infty) \times \mathbb{R}^d)$ and A^0 satisfies the positive maximum principle, Theorem 5.4 of Chapter 4 guarantees a solution of the $D_{[0, \infty) \times \mathbb{R}^d}, [0, \infty)$ martingale problem for $((A^0)^A, \delta_0 \times \mu)$, where $(A^0)^A$ is defined by (3.22). Arguing as in the proof of Proposition 3.5, we complete the proof using Proposition 3.9 and Problem 19, both of Chapter 4. \square

3.11 Theorem Let $\sigma: [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}^d \otimes \mathbb{R}^d$ and $b: [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ be locally bounded and Borel measurable. Suppose that for each $T > 0$ and $n \geq 1$ there exist constants K_T and $K_{T,n}$ such that

$$(3.34) \quad |\sigma(t, x)|^2 \leq K_T(1 + |x|^2), \quad x \cdot b(t, x) \leq K_T(1 + |x|^2), \\ 0 \leq t \leq T, \quad x \in \mathbb{R}^d,$$

and

$$(3.35) \quad |\sigma(t, x) - \sigma(t, y)| \vee |b(t, x) - b(t, y)| \leq K_{T,n}|x - y|, \\ 0 \leq t \leq T, \quad |x| \vee |y| \leq n.$$

Given a d -dimensional Brownian motion W and an independent \mathbb{R}^d -valued random variable ξ on a probability space (Ω, \mathcal{F}, P) such that $E[|\xi|^2] < \infty$, there exists a process X with $X(0) = \xi$ a.s. such that $(\Omega, \mathcal{F}, P, \{\mathcal{F}_t\}, W, X)$ is a solution of the stochastic integral equation corresponding to (σ, b) , where $\mathcal{F}_t = \mathcal{F}_t^W \vee \sigma(\xi)$.

Proof. We first give the proof in the case that (3.34) and (3.35) are replaced by

$$(3.36) \quad |\sigma(t, x)| \vee |b(t, x)| \leq K_T, \quad 0 \leq t \leq T, \quad x \in \mathbb{R}^d$$

and

$$(3.37) \quad |\sigma(t, x) - \sigma(t, y)| \vee |b(t, x) - b(t, y)| \leq K_T|x - y|, \\ 0 \leq t \leq T, \quad x, y \in \mathbb{R}^d.$$

Let $X_0(t) \equiv \xi$. Having defined X_0, \dots, X_k , let

$$(3.38) \quad X_{k+1}(t) \equiv \xi + \int_0^t \sigma(s, X_k(s)) dW(s) + \int_0^t b(s, X_k(s)) ds,$$

and note that $E[|X_{k+1}(t)|^2] < \infty$ for each $t \geq 0$ by (3.36). For $k = 0, 1, \dots$, let $\varphi_k(t) \equiv E[|X_{k+1}(t) - X_k(t)|^2]$. Given $T > 0$, (3.37) implies that

$$(3.39) \quad \varphi_k(t) \leq 2K_T^2(1 + T) \int_0^t \varphi_{k-1}(s) ds, \quad 0 \leq t \leq T.$$

Since $\varphi_0(t) \leq 2K_T^2(1 + T)t$ for $0 \leq t \leq T$ by (3.36), we have by induction that, for $k = 0, 1, \dots$,

$$(3.40) \quad \varphi_k(t) \leq \frac{[2K_T^2(1 + T)]^{k+1}t^{k+1}}{(k+1)!}, \quad 0 \leq t \leq T.$$

It follows that

$$\begin{aligned}
 (3.41) \quad & E \left[\sup_{0 \leq t \leq T} |X_{k+1}(t) - X_k(t)|^2 \right] \\
 & \leq 2E \left[\sup_{0 \leq t \leq T} \left| \int_0^t (\sigma(s, X_k(s)) - \sigma(s, X_{k-1}(s))) dW(s) \right|^2 \right] \\
 & \quad + 2E \left[\sup_{0 \leq t \leq T} \left| \int_0^t (b(s, X_k(s)) - b(s, X_{k-1}(s))) ds \right|^2 \right] \\
 & \leq 8E \left[\left| \int_0^T (\sigma(s, X_k(s)) - \sigma(s, X_{k-1}(s))) dW(s) \right|^2 \right] \\
 & \quad + 2TE \left[\int_0^T |b(s, X_k(s)) - b(s, X_{k-1}(s))|^2 ds \right] \\
 & \leq 2K_T^2(4 + T) \int_0^T E[|X_k(s) - X_{k-1}(s)|^2] ds \\
 & \leq \frac{2K_T^2(4 + T)[2K_T^2(1 + T)]^k T^{k+1}}{(k + 1)!} \equiv \varepsilon_k(T),
 \end{aligned}$$

and therefore

$$(3.42) \quad \sum_{k=0}^{\infty} P \left\{ \sup_{0 \leq t \leq T} |X_{k+1}(t) - X_k(t)| \geq 2^{-k} \right\} \leq \sum_{k=0}^{\infty} 4^k \varepsilon_k(T) < \infty.$$

By the Borel-Cantelli lemma, $\sup_{0 \leq t \leq T} |X_{k+1}(t) - X_k(t)| < 2^{-k}$ for all $k \geq k(\omega)$ for almost all ω . Now T was arbitrary, so there exists $A \in \mathcal{F}$ with $P(A) = 0$ such that, for every $\omega \in \Omega$, $\{\chi_A \cdot X_k\}$ converges uniformly on bounded time intervals. Letting X be the limiting process, we conclude from (3.42) that $X(0) = \xi$ a.s. and $(\Omega, \mathcal{F}, P, \{\mathcal{F}_t\}, W, X)$ is a solution of the stochastic integral equation corresponding to (σ, b) .

We now want to obtain the conclusion of the theorem under the original hypotheses ((3.34) and (3.35) instead of (3.36) and (3.37)). For each $n \geq 1$, define $\rho_n: [0, \infty) \times \mathbb{R}^d \rightarrow [0, \infty) \times \mathbb{R}^d$ by $\rho_n(t, x) = (t, (1 \wedge (n/|x|))x)$, and let $\sigma_n = \sigma \circ \rho_n$ and $b_n = b \circ \rho_n$. By the first part of the proof there exists a solution $(\Omega, \mathcal{F}, P, \{\mathcal{F}_t\}, W, X_n)$ of the stochastic integral equation corresponding to (σ_n, b_n) . Letting $\tau_n = \inf\{t \geq 0: |X_n(t)| \geq n\}$, Theorem 3.7 guarantees that $X_n(t) = X_m(t)$ whenever $0 \leq t \leq \tau_n \wedge \tau_m$ and $m, n \geq 1$. Thus, we can define $X(t) = X_n(t)$ for $0 \leq t \leq \tau_n$, $n \geq 1$.

To complete the proof, it suffices to show that $\tau_n \rightarrow \infty$ a.s. By Itô's formula and (3.34),

$$(3.43) \quad E[\log(1 + |X_n(t \wedge \tau_n)|^2)]$$

is bounded above in n for fixed $t \geq 0$. The same is therefore true of $\log(1 + n^2)P\{\tau_n \leq t\}$, so $P\{\tau_n \leq t\} \rightarrow 0$ for each $t \geq 0$. Since $\tau_1 \leq \tau_2 \dots$, the desired conclusion follows. \square

4. PROBLEMS

1. Let W be a d -dimensional $\{\mathcal{F}_t\}$ -Brownian motion, and let τ be an $\{\mathcal{F}_t\}$ -stopping time with $\tau < \infty$ a.s. Show that $W^*(\cdot) \equiv W(\tau + \cdot) - W(\tau)$ ($\equiv 0$ if $\tau = \infty$) is a d -dimensional Brownian motion, and that $\mathcal{F}_t^{W^*}$ is independent of \mathcal{F}_τ for each $t \geq 0$.
2. Let W be a d -dimensional $\{\mathcal{F}_t\}$ -Brownian motion. Show that

$$(4.1) \quad X_\theta(t) \equiv \exp \left\{ \theta \cdot W(t) - \frac{|\theta|^2 t}{2} \right\}$$

is an $\{\mathcal{F}_t\}$ -martingale. For $d = 1$, $\alpha > 0$, and $\beta > 0$, show that $P\{\sup_{0 \leq s \leq t} (W(s) - \alpha s/2) > \beta\} \leq e^{-\alpha\beta}$.

3. Let W be a one-dimensional Brownian motion. Evaluate the stochastic integral $\int_0^t W^2 dW$ directly from its definition (Theorem 2.3). Check your result using Itô's formula.

4. Let $M \in \mathcal{M}_{c, loc}$ and $X, Y, X_1, X_2, \dots \in L_{loc}^2(\langle M \rangle)$. Suppose that $|X_n| \leq Y$ for each $n \geq 1$ and $X_n(t) \rightarrow X(t)$ a.s. for each $t \geq 0$. Show that for every $T > 0$,

$$(4.2) \quad \sup_{0 \leq t \leq T} \left| \int_0^t X_n dM - \int_0^t X dM \right| \xrightarrow{P} 0.$$

5. Let W be a d -dimensional $\{\mathcal{F}_t\}$ -Brownian motion (with $\{\mathcal{F}_t\}$ a complete filtration), and let $\sigma: [0, \infty) \times \Omega \rightarrow \mathbb{R}^d \otimes \mathbb{R}^d$ be $\{\mathcal{F}_t\}$ -progressive and satisfy $\sigma\sigma^T = I_d$. Show that $\tilde{W} \equiv \int_0^t \sigma(s) dW(s)$ is a d -dimensional $\{\mathcal{F}_t\}$ -Brownian motion.

6. Show that the spherical coordinates

$$(4.3) \quad \begin{aligned} \rho &= |B| = (B_1^2 + B_2^2 + B_3^2)^{1/2}, \\ \varphi &= \cos^{-1}(B_3/\rho) = \text{colatitude}, \\ \theta &= \tan^{-1}(B_2/B_1) = \text{longitude} \end{aligned}$$

of a three-dimensional Brownian motion $B = (B_1, B_2, B_3)$ evolve according to the stochastic differential equations

$$(4.4) \quad \begin{aligned} d\rho &= dW_1 + \rho^{-1} dt, \\ d\varphi &= \rho^{-1} dW_2 + \frac{1}{2}\rho^{-2} \cot \varphi dt, \\ d\theta &= \rho^{-1} \csc \varphi dW_3 \end{aligned}$$

with a new three-dimensional Brownian motion $W = (W_1, W_2, W_3)$:

$$(4.5) \quad \begin{aligned} W_1 &= \int_0^t \rho^{-1} (B_1 dB_1 + B_2 dB_2 + B_3 dB_3), \\ W_2 &= \int_0^t \rho^{-2} (\csc \varphi) B_3 (B_1 dB_1 + B_2 dB_2) - \int_0^t \sin \varphi dB_3, \\ W_3 &= \int_0^t \rho^{-1} (\csc \varphi) (B_1 dB_2 - B_2 dB_1). \end{aligned}$$

7. Let $\sigma: [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}^d \otimes \mathbb{R}^d$ and $b: [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ be locally bounded and Borel measurable and suppose that $(\Omega, \mathcal{F}, P, \{\mathcal{F}_t\}, W, X)$ is a solution of the stochastic integral equation corresponding to (σ, b) . Let $c: [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}$ be bounded and Borel measurable. Show that if $f \in C_c^{1,2}([0, \infty) \times \mathbb{R}^d)$, then

$$(4.6) \quad \begin{aligned} E \left[f(t, X(t)) \exp \left\{ \int_0^t c(s, X(s)) ds \right\} \right] \\ = E[f(0, X(0))] + E \left[\int_0^t (Gf + cf)(s, X(s)) \exp \left\{ \int_0^s c(r, X(r)) dr \right\} ds \right] \end{aligned}$$

for all $t \geq 0$, where G is defined by (3.4) and (3.5).

8. Let $\Phi: \mathbb{R}^d \rightarrow \mathbb{R}^d$ be a C^2 -diffeomorphism (that is, Φ is one-to-one, onto, and twice continuously differentiable, as is its inverse Φ^{-1}). Let $\sigma: \mathbb{R}^d \rightarrow \mathbb{R}^d \otimes \mathbb{R}^d$ and $b: \mathbb{R}^d \rightarrow \mathbb{R}^d$ be locally bounded and Borel measurable, and suppose the stochastic integral equation corresponding to (σ, b) has a solution $(\Omega, \mathcal{F}, P, \{\mathcal{F}_t\}, W, X)$. Observe that then there exist $\hat{\sigma}: \mathbb{R}^d \rightarrow \mathbb{R}^d \otimes \mathbb{R}^d$ and $\hat{b}: \mathbb{R}^d \rightarrow \mathbb{R}^d$ locally bounded and Borel measurable such that $(\Omega, \mathcal{F}, P, \{\mathcal{F}_t\}, W, \Phi \circ X)$ is a solution of the stochastic integral equation corresponding to $(\hat{\sigma}, \hat{b})$. Define G in terms of σ and b and \hat{G} in terms of $\hat{\sigma}$ and \hat{b} as in (3.4) and (3.5). Show that $Gf = [\hat{G}(f \circ \Phi^{-1})] \circ \Phi$ for all $f \in C_c^2(\mathbb{R}^d)$. Thus the relationship between a stochastic integral equation and its associated differential operator is invariant under diffeomorphism.
9. Let $a: \mathbb{R}^d \rightarrow S_d$ and $b: \mathbb{R}^d \rightarrow \mathbb{R}^d$ be locally bounded and Borel measurable, and define A and G by (3.3) and (3.4). Let $\varphi \in C^2(\mathbb{R}^d)$ and suppose that for each $n \geq 1$ there exists a constant $K_n \geq 0$ such that

$$(4.7) \quad \max \{ \nabla \varphi \cdot a \nabla \varphi, G\varphi \} \chi_{\{x \in \mathbb{R}^d: \varphi(x) > 0, |x| \leq n\}} \leq K_n \varphi.$$

Show that if X is a solution of the $C_{\mathbb{R}^d}[0, \infty)$ martingale problem for A , then $P\{\varphi(X(0)) \leq 0\} = 1$ implies $P\{\varphi(X(t)) \leq 0 \text{ for all } t \geq 0\} = 1$.

Hint: Show that Gronwall's inequality applies to $E[\varphi^+(X(t \wedge \tau_n))]$, where $\tau_n = \inf \{t \geq 0: |X(t)| \geq n\}$, by approximating φ^+ by a sequence of the form $\{h_n \circ \varphi\}$.

10. Define $\int_0^t \sigma(s, X(s)) dW(s)$ for $\sigma: [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}^{d \times m}$ (the space of real $d \times m$ matrices) locally bounded and Borel measurable with W and X as before by defining $\tilde{\sigma}: [0, \infty) \times \mathbb{R}^{d \times m} \rightarrow \mathbb{R}^{d \times m} \otimes \mathbb{R}^{d \times m}$ in terms of σ in the obvious way. Check to see which of the results of Section 3 extend to nonsquare σ .
11. (a) Let $M \in \mathcal{M}_c$ and $X \in L^2(\langle M \rangle)$, and let $s \geq 0$ and Z be a bounded \mathcal{F}_s -measurable random variable. Show that

$$(4.8) \quad \int_s^t ZX \, dM = Z \int_s^t X \, dM, \quad t > s.$$

- (b) Prove Lemma 2.5.

Hint: First consider $X \in S$.

12. Let $M \in \mathcal{M}_{c, loc}$ and let X be continuous and adapted. Show that for $0 = t_0 < t_1 < \dots < t_m = t$,

$$(4.9) \quad \int_0^t X \, dM = \lim_{\max(t_{k+1} - t_k) \rightarrow 0} \sum_k X(t_k)(M(t_{k+1}) - M(t_k)).$$

13. Let W be a one-dimensional Brownian motion, and let $X(t) = W(t) + t$. Find a function φ such that $\varphi(X(t))$ is a martingale. (Use Itô's formula.) Let $\tau = \inf\{t: X(t) = -a \text{ or } b\}$. Use $\varphi(X(t))$ to find $P\{X(\tau) = b\}$. What is $E[\tau]$?

14. Let X be a solution in \mathbb{R} of

$$X(t) = x + \int_0^t bX(s) \, ds + \int_0^t \sigma X(s) \, dW(s)$$

and let $Y = X^2$.

- (a) Use Itô's formula to find the stochastic integral equation satisfied by Y .
- (b) Use the equation in (a) to find $E[X^2]$.
- (c) Extend the above argument to find $E[X^k]$, $k = 1, 2, 3, \dots$.
15. Let W be a one-dimensional Brownian motion.

- (a) Let $X = (X_1, X_2)$ satisfy

$$\begin{aligned} X_1(t) &= x_1 + \int_0^t X_2(s) \, ds \\ X_2(t) &= x_2 - \int_0^t X_1(s) \, ds + \int_0^t cX_1(s) \, dW(s). \end{aligned}$$

Define $m_1(t) = E[X^2(t)]$, $m_2(t) = E[X(t)Y(t)]$, and $m_3(t) = E[Y^2(t)]$. Find a system of three linear differential equations satisfied by m_1 ,

m_2 , and m_3 . Show that the expected "total energy" ($E[X^2(t) + Y^2(t)]$) is asymptotic to $ke^{\lambda t}$ for some $\lambda > 0$ and $k > 0$.

- (b) Let $X = (X_1, X_2)$ satisfy

$$X_1(t) = X_1(0) + \int_0^t X_2(s) dW(s)$$

$$X_2(t) = X_2(0) - \int_0^t X_1(s) dW(s).$$

Show that $X_1^2(t) + X_2^2(t) = (X_1^2(0) + X_2^2(0)) e^t$.

16. Let W be a one-dimensional Brownian motion.

- (a) For $x \geq 0$, let $X(t, x) = x + \int_0^t \lambda X(s, x) ds + \int_0^t \sqrt{X(s, x)} dW(s)$ and $\tau_x = \inf\{t: X(t, x) = 0\}$. Calculate $P\{\tau_x < \infty\}$ as a function of λ .
- (b) For $x > 0$, let $X(t, x) = x - \int_0^t \lambda X(s, x) ds + \int_0^t X(s, x) dW(s)$ with $\lambda > 0$, and let τ_x be defined as above. Show that $P\{\tau_x < \infty\} = 0$, but that $P\{\lim_{t \rightarrow \infty} X(t, x) = 0\} = 1$.
- (c) For $x > 0$, let $X(t, x) = x + \int_0^t \sigma(X(s, x)) dW(s)$, and let τ_x be defined as above. Give conditions on σ that imply $E[\tau_x] < \infty$.
- (d) For $x > 0$, let $X(t, x) = x + \int_0^t \lambda ds + \int_0^t \sqrt{X(s, x)} dW(s)$, and let τ_x be defined as above. For what values of $\lambda > 0$ is $P\{\tau_x < \infty\} > 0$? For these values, show that $P\{\tau_x < \infty\} = 1$, but that $E[\tau_x] = \infty$.

5. NOTES

There are many general references on stochastic integration and stochastic integral equations. These include McKean (1969), Gihman and Skorohod (1972), Friedman (1975), Ikeda and Watanabe (1981), Elliot (1982), Métivier (1982), and Chung and Williams (1983). Our treatment is heavily influenced by Priouret (1974).

Stochastic integrals with respect to square integrable martingales go back to Doob (1953), page 437, and were developed by Courrege (1963) and Kunita and Watanabe (1967). The extension to local martingales is due to Meyer (1967) and Doleans-Dade and Meyer (1970). Itô's formula goes back, of course, to Itô (1951).

Theorem 3.3 is due to Stroock and Varadhan (1972), Theorems 3.6 and 3.8 to Yamada and Watanabe (1971), and Theorem 3.10 to Skorohod (1965). Theorems 3.7 and 3.11 are the classical uniqueness and existence theorems of Itô (1951).

Problems 6 and 8 were borrowed from McKean (1969) and Friedman (1975), respectively.

6

RANDOM TIME CHANGES

In this chapter we continue the study of stochastic equations that determine Markov processes. These equations involve random time changes of other Markov processes and frequently reduce stochastic problems to problems in analysis. Section 1 considers random time changes of a single process. The multiparameter analogue is developed in Section 2. Section 3 gives convergence results based on the random time changes. Sections 4 and 5 give time change equations for large classes of Markov chains and diffusion processes.

1. ONE-PARAMETER RANDOM TIME CHANGES

Let Y be a process with sample paths in $D_E[0, \infty)$, and let β be a nonnegative Borel measurable function on E . Suppose that $\beta \circ Y$ is a.s. bounded on bounded time intervals. We are interested in solutions of

$$(1.1) \quad Z(t) = Y \left(\int_0^t \beta(Z(s)) \, ds \right).$$

Observe that if Z is a solution and we set

$$(1.2) \quad \tau(t) = \int_0^t \beta(Z(s)) \, ds = \int_0^t \beta(Y(\tau(s))) \, ds,$$

then (see Problem 11)

$$(1.3) \quad t \geq \lim_{\varepsilon \rightarrow 0^+} \int_0^t \frac{\beta(Z(s))}{\beta(Z(s)) \vee \varepsilon} ds = \lim_{\varepsilon \rightarrow 0^+} \int_0^{\tau(t)} \frac{1}{\beta(Y(u)) \vee \varepsilon} du \\ = \int_0^{\tau(t)} \frac{1}{\beta(Y(u))} du,$$

and equality holds if and only if the Lebesgue measure of $\{s \leq t: \beta(Z(s)) = 0\}$ is zero. Conversely if

$$(1.4) \quad t = \int_0^{\tau(t)} \frac{1}{\beta(Y(u))} du$$

has a solution for all t (of necessity unique), then $\tau(t)$ is locally absolutely continuous (in fact locally Lipschitz) with

$$(1.5) \quad \dot{\tau}(t) = \beta(Y(\tau(t))), \quad \text{a.e. } t,$$

(differentiate both sides of (1.4)), and hence $Z(t) \equiv Y(\tau(t))$ is a solution of (1.1). More generally, let

$$(1.6) \quad \tau_1 = \inf \left\{ s: \int_0^s \frac{1}{\beta(Y(u))} du = \infty \right\}$$

and suppose $\int_0^\infty (1/\beta(Y(u))) du = \infty$. If $\tau \leq \tau_1$ and $\beta(Y(\tau)) = 0$ when $\tau < \infty$, let $\tau(t)$ satisfy (1.4) for

$$(1.7) \quad t \leq \int_0^t \frac{1}{\beta(Y(u))} du \equiv t_0$$

and

$$(1.8) \quad \tau(t) = \tau, \quad t > t_0.$$

Then $Z(t) \equiv Y(\tau(t))$ is a solution of (1.1).

1.1 Theorem Let Y , β , and τ_1 be as above. Define

$$(1.9) \quad \tau_0 = \inf \{s: \beta(Y(s)) = 0\}$$

and

$$(1.10) \quad \tau_2 = \liminf_{\varepsilon \rightarrow 0^+} \{s: \beta(Y(s)) < \varepsilon\}.$$

(a) If $\tau_0 = \tau_1$ and $\beta(Y(\tau_0)) = 0$ when $\tau_0 < \infty$, then (1.1) has a unique solution $Z(t)$.

(b) Suppose β is continuous. If $\tau_0 = \tau_1 = \tau_2$, then there is a unique locally absolutely continuous function $\gamma(t)$ satisfying

$$(1.11) \quad \begin{aligned} & \beta(Y(\gamma(t))) \wedge \beta(Y^-(\gamma(t))) \\ & \leq \gamma'(t) \leq \beta(Y(\gamma(t))) \vee \beta(Y^-(\gamma(t))), \quad \text{a.e. } t, \end{aligned}$$

where $Y^-(u) = \lim_{v \rightarrow u^-} Y(v)$, $u > 0$, and $Y^-(0) = Y(0)$, and $Z(t) = Y(\gamma(t))$, that is, $\tau(t) = \gamma(t)$.

1.2 Remark Note that τ_0 , τ_1 , and τ_2 may be infinite. \square

Proof. Existence follows from the construction in (1.7) and (1.8). By (1.3), any solution, $Z(t)$, with $\tau(t)$ defined by (1.2), must satisfy $\tau(t) \leq \tau_1$. If $\tau(t) < \tau_0$, then $\beta(Z(s)) \neq 0$ for all $s \leq t$, and (1.4) uniquely determines $\tau(t)$. If $\tau_1 = \tau_0$, $\tau(t)$ is uniquely determined for all t .

If $\gamma(t)$ satisfies (1.11), then as above

$$(1.12) \quad \begin{aligned} t & \geq \lim_{\epsilon \rightarrow 0^+} \int_0^t \frac{\beta(Y(\gamma(s))) \vee \beta(Y^-(\gamma(s)))}{\beta(Y(\gamma(s))) \wedge \beta(Y^-(\gamma(s))) \vee \epsilon} ds \\ & \geq \int_0^{\gamma(t)} \frac{1}{\beta(Y(u)) \vee \beta(Y^-(u))} du \\ & = \int_0^{\gamma(t)} \frac{1}{\beta(Y(u))} du, \end{aligned}$$

since $Y(u) = Y^-(u)$ for almost every u , and $\gamma(t) \leq \tau_1$. Since $\tau_2 = \tau_1$, $\beta(Y(u)) \wedge \beta(Y^-(u)) \neq 0$ for $u < \tau_1$, and hence for $\gamma(t) < \tau_1$,

$$(1.13) \quad t = \int_0^t \frac{\beta(Y(\gamma(s))) \wedge \beta(Y^-(\gamma(s)))}{\beta(Y(\gamma(s))) \wedge \beta(Y^-(\gamma(s)))} ds \leq \int_0^{\gamma(t)} \frac{1}{\beta(Y(u))} du,$$

and (1.12) and (1.13) imply $Z(\gamma(t))$ is the unique solution of (1.1). \square

We now relate solutions of (1.1) to solutions of a martingale problem.

1.3 Theorem Let $A \subset \bar{C}(E) \times B(E)$ and suppose Y is a solution of the $D_E[0, \infty)$ martingale problem for A .

If the conditions of Theorem 1.1(a) hold for almost every sample path, then the solution of (1.1) is a solution of the martingale problem for $\beta A \cap (B(E) \times B(E))$, where

$$(1.14) \quad \beta A \equiv \{(f, \beta g): (f, g) \in A\}.$$

Proof. Note that $\{\tau(s) \leq t\} = \{\int_0^s (1/\beta(Y(u))) du \geq s\} \cup \{\tau_1 \leq t\} \in \mathcal{F}_{t,+}^Y$, so $\tau(s)$ is an $\{\mathcal{F}_{t,+}^Y\}$ -stopping time. For $(f, g) \in A$

$$(1.15) \quad f(Y(t)) - \int_0^t g(Y(s)) ds$$

is an $\{\mathcal{F}_{t,+}^Y\}$ -martingale, hence the optional sampling theorem implies

$$(1.16) \quad \begin{aligned} f(Y(\tau(t))) - \int_0^{\tau(t)} g(Y(s)) ds \\ = f(Z(t)) - \int_0^t \beta(Z(s))g(Z(s)) ds \end{aligned}$$

is an $\{\mathcal{F}_{\tau(t),+}^Y\}$ -martingale. \square

We have the following converse to Theorem 1.3.

1.4 Theorem Let (E, r) be complete and separable, and let $A \subset \bar{C}(E) \times B(E)$. Suppose $\mathcal{D}(A)$ is separating, the $D_E[0, \infty)$ martingale problem for A is well-posed, and $\beta \in M(E)$, $\beta \geq 0$, is such that $\beta A \subset B(E) \times B(E)$. If Z is a solution of the $D_E[0, \infty)$ martingale problem for βA , then there is a version of Z satisfying (1.1) for a process Y that is a solution of the martingale problem for A .

Proof. First suppose

$$(1.17) \quad \tau(\infty) \equiv \int_0^\infty \beta(Z(s)) ds = \infty \quad \text{a.s.}$$

and define

$$(1.18) \quad \gamma(t) = \inf \left\{ u: \int_0^u \beta(Z(s)) ds > t \right\}.$$

Then $Y(t) \equiv Z(\gamma(t))$ is a solution of the martingale problem for A , that is, for $(f, g) \in A$,

$$(1.19) \quad f(Z(\gamma(t))) - \int_0^{\gamma(t)} \beta(Z(s))g(Z(s)) ds = f(Y(t)) - \int_0^t g(Y(u)) du$$

is a martingale by the optional sampling theorem. (See Problem 12 for the equality in (1.19).) Let $\gamma_+(t) = \lim_{s \rightarrow t+} \gamma(s)$. We claim that Z is constant on the interval $[\gamma(t), \gamma_+(t)]$ (see Problem 45 of Chapter 4), and since $\gamma(\tau(t)) \leq t \leq \gamma_+(\tau(t))$,

$$(1.20) \quad Z(t) = Z(\gamma(\tau(t))) = Y \left(\int_0^t \beta(Z(s)) ds \right).$$

If $\tau(\infty) < \infty$, then $Y(t) = Z(\gamma(t))$ for $t \leq \tau(\infty)$ and Y must be extended past $\tau(\infty)$ (on an enlarged sample space using Lemma 5.16 of Chapter 4). \square

1.5 Theorem Suppose Y is as above, β is continuous, $\{Y_n\}$ is a sequence of processes with sample paths in $D_E[0, \infty)$ such that $Y_n \Rightarrow Y$, and $\{\beta_n\}$ is a sequence of nonnegative, Borel measurable functions on E such that

$$(1.21) \quad \lim_{n \rightarrow \infty} \sup_{x \in K} |\beta_n(x) - \beta(x)| = 0$$

for every compact K . Let $h_n > 0$ and $\lim_{n \rightarrow \infty} h_n = 0$. Suppose Z , Z_n , and W_n satisfy

$$(1.22) \quad Z(t) = Y \left(\int_0^t \beta(Z(s)) ds \right),$$

$$(1.23) \quad Z_n(t) = Y_n \left(\int_0^t \beta_n(Z_n(s)) ds \right),$$

and

$$(1.24) \quad W_n(t) = Y_n \left(\int_0^{(t/h_n)h_n} \beta_n(W_n(s)) ds \right)$$

for all $t \geq 0$. If $\tau_0 = \tau_1 = \tau_2$ a.s., then $Z_n \Rightarrow Z$ and $W_n \Rightarrow Z$.

1.6 Remark If Y is quasi-left continuous (see Theorem 3.12 of Chapter 4), in particular if Y is a Feller process, then $\tau_0 = \tau_2$ a.s. Observe that, for $\varepsilon > 0$, $\tau^{(\varepsilon)} = \inf \{s: \beta(Y(s)) \leq \varepsilon\} \leq \tau_0$, and by quasi-left continuity, $Y(\tau_2) = \lim_{\varepsilon \rightarrow 0} Y(\tau^{(\varepsilon)})$ on the set $\{\tau_2 < \infty\}$. Consequently, on $\{\tau_2 < \infty\}$, $\lim_{\varepsilon \rightarrow 0} \beta(Y(\tau^{(\varepsilon)})) = \beta(Y(\tau_2)) = 0$ a.s., and hence $\tau_2 = \tau_0$. Of course if $\inf_x \beta(x) > 0$, then $\tau_0 = \tau_1 = \tau_2 = \infty$.

Proof. By Theorem 1.8 of Chapter 3 we can, without loss of generality, assume Y_n and Y are defined on the same sample space (Ω, \mathcal{F}, P) and that $\lim_{n \rightarrow \infty} d(Y_n, Y) = 0$ a.s. Fix $\omega \in \Omega$ for which $\lim_{n \rightarrow \infty} d(Y_n, Y) = 0$ and $\tau_0 = \tau_1 = \tau_2$. We first assume $\bar{\beta} \equiv \sup_{x, n} \beta_n(x)$ is finite. Then

$$(1.25) \quad w'(Z_n, \delta, T) \leq w'(Y_n, \delta \bar{\beta}, T \bar{\beta}),$$

$$(1.26) \quad w'(W_n, \delta, T) \leq w'(Y_n, (\delta + h_n) \bar{\beta}, T \bar{\beta}),$$

$$(1.27) \quad \{Z_n(s): s \leq T\} \subset \{Y_n(u): u \leq T \bar{\beta}\},$$

and

$$(1.28) \quad \{W_n(s): s \leq T\} \subset \{Y_n(u): u \leq T \bar{\beta}\}.$$

Since $\{Y_n\}$ is convergent in $D_E[0, \infty)$, it follows that $\{Z_n\}$ and $\{W_n\}$ are relatively compact in $D_E[0, \infty)$. We show that any convergent subsequence of $\{Z_n\}$ (or $\{W_n\}$) converges to Z , and hence $\lim_{n \rightarrow \infty} d(Z_n, Z) = 0$ (and $\lim_{n \rightarrow \infty} d(W_n,$

$Z) = 0$). If $\lim_{k \rightarrow \infty} d(Z_{n_k}, \hat{Z}) = 0$, then

$$(1.29) \quad \lim_{k \rightarrow \infty} \int_0^t \beta_{n_k}(Z_{n_k}(s)) ds = \int_0^t \beta(\hat{Z}(s)) ds \equiv \gamma(t)$$

and

$$(1.30) \quad \gamma'(t) = \beta(\hat{Z}(t)) = \lim_{k \rightarrow \infty} \beta\left(Y_{n_k}\left(\int_0^t \beta_{n_k}(Z_{n_k}(s)) ds\right)\right)$$

for almost every t . The right side is either $\beta(Y(\gamma(t)))$ or $\beta(Y^-(\gamma(t)))$. Therefore $\gamma(t)$ satisfies (1.11) and hence $\hat{Z} = Z$. Similarly, if $\lim_{k \rightarrow \infty} d(W_{n_k}, \hat{Z}) = 0$, then

$$(1.31) \quad \lim_{k \rightarrow \infty} \int_0^{[t/h_n]h_n} \beta_{n_k}(W_{n_k}(s)) ds = \int_0^t \beta(\hat{Z}(s)) ds$$

and the proof follows as for $\{Z_n\}$.

Now dropping the assumption that the β_n are bounded, let Z^M , Z_n^M , and W_n^M be as above, but with β and β_n replaced by $M \wedge \beta$ and $M \wedge \beta_n$, $M > 0$. Since we are not assuming the solution of (1.23) is unique, take $Z_n^M = Z_n(\theta_n^M(t))$ where

$$(1.32) \quad \int_0^{\theta_n^M(t)} 1 \vee \left(\frac{\beta_n(Z_n(s))}{M} \right) ds = t.$$

As above, fix $\omega \in \Omega$ such that $\lim_{n \rightarrow \infty} d(Y_n, Y) = 0$ and $\tau_0 = \tau_1 = \tau_2$. Then $\lim_{n \rightarrow \infty} d(Z_n^M, Z^M) = 0$, and $\lim_{n \rightarrow \infty} d(W_n^M, Z^M) = 0$. Fix $t > 0$, and let $M > \sup_{s \leq t} \beta(Z(s))$. Note that $Z^M(s) = Z(s)$ for $s \leq t$. We claim that for n sufficiently large, $M > \sup_{s \leq t} \beta_n(Z_n^M(s))$ and hence $Z_n^M(s) = Z_n(s)$ for $s \leq t$. (Similarly for W_n^M .) To see this, suppose not. Then there exist $0 \leq s_n \leq t$ (which we can assume satisfy $s_n \rightarrow s_0$) such that $\lim_{n \rightarrow \infty} \beta_n(Z_n^M(s_n)) \geq M$. But $\{Z_n^M(s_n)\}$ is relatively compact with limit points in $\{Z^M(s_0^-), Z^M(s_0)\} = \{Z(s_0^-), Z(s_0)\}$. Consequently,

$$(1.33) \quad \overline{\lim}_{n \rightarrow \infty} \beta_n(Z_n^M(s_n)) = \overline{\lim}_{n \rightarrow \infty} \beta(Z_n^M(s_n)) \leq \beta(Z(s_0^-)) \vee \beta(Z(s_0)) < M.$$

Recall that if $Z^M(s) = Z(s)$ for $s \leq t$, then $d(Z^M, Z) \leq e^{-t}$. Since t is arbitrary, it follows that $d(Z_n, Z) \rightarrow 0$. \square

2. MULTIPARAMETER RANDOM TIME CHANGES

We now consider a system of random time changes analogous to (1.1). For $k = 1, 2, \dots$, let (E_k, r_k) be a complete, separable metric space, and let Y_k be a process with sample paths in $D_{E_k}[0, \infty)$ defined on a complete probability

space (Ω, \mathcal{F}, P) . Let $\beta_k: \prod_i E_i \rightarrow [0, \infty)$ be nonnegative Borel measurable functions. We are interested in solutions of the system

$$(2.1) \quad Z_k(t) = Y_k \left(\int_0^t \beta_k(Z(s)) ds \right),$$

where $Z = (Z_1, Z_2, \dots)$. (Similarly we set $Y = (Y_1, Y_2, \dots)$.)

We begin with the following technical lemma.

2.1 Lemma If for almost every $\omega \in \Omega$ a solution Z of (2.1) exists and is unique, then Z is a stochastic process.

Proof. Let $S = \prod_i D_{E_i}[0, \infty)$ and define $\gamma: S \times S \rightarrow S$ by

$$(2.2) \quad \gamma_k(y, z) = y_k \left(\int_0^t \beta_k(z(s)) ds \right).$$

Then γ_k is Borel measurable and hence

$$(2.3) \quad \Gamma = \{(y, z): z = \gamma(y, z)\}$$

is a Borel measurable subset of $S \times S$, as is

$$(2.4) \quad \Gamma_{k, t, B} = \{(y, z): z = \gamma(y, z), z_k(t) \in B\}$$

for $B \in \mathcal{B}(E_k)$. Therefore $\pi \Gamma_{k, t, B} = \{y: (y, z) \in \Gamma_{k, t, B}\}$ is an analytic subset of S and

$$(2.5) \quad \{Z_k(t) \in B\} = \{Y \in \pi \Gamma_{k, t, B}\} \in \mathcal{F}$$

by the completeness of (Ω, \mathcal{F}, P) . (See Appendix 11.) □

In the one-dimensional case we noted that $\tau(t)$ was a stopping time with respect to \mathcal{F}_{t+}^Y , at least in the case $\tau_0 = \tau_1, \beta(Y(\tau_0)) = 0$ for $\tau_0 < \infty$.

To determine the analogue of this observation in the multiparameter case we define

$$(2.6) \quad \mathcal{F}_u^0 = \sigma(Y_k(s_k); s_k \leq u_k), \quad u \in [0, \infty]^\infty,$$

and

$$(2.7) \quad \mathcal{F}_u = \bigcap_n \sigma(\mathcal{N}) \vee \mathcal{F}_{u^{(n)}}^0$$

where $\mathcal{N} \subset \mathcal{F}$ is the collection of all sets of probability zero, and $u^{(n)}$ is defined by $u_k^{(n)} = u_k + 1/n$, $k \leq n$, and $u_k^{(n)} = \infty$, $k > n$. A random variable $\tau = (\tau_1, \tau_2, \dots)$ with values in $[0, \infty)^\infty$ is an $\{\mathcal{F}_u\}$ -stopping time if $\{\tau \leq u\} \equiv \{\tau_1 \leq u_1, \tau_2 \leq u_2, \dots\} \in \mathcal{F}_u$, for all $u \in [0, \infty)^\infty$. (See Chapter 2, Section 8, for details concerning multiparameter stopping times.)

2.2 Theorem (a) For $u \in [0, \infty]^\infty$, $t > 0$, let $H_{u,t}$ be the set of $\omega \in \Omega$ such that there exists $z \in S \equiv \prod_{E_t} D_E [0, \infty)$ satisfying

$$(2.8) \quad z_k(r) = Y_k \left(\int_0^r \beta_k(z(s)) \, ds, \omega \right), \quad r \leq t, \quad k = 1, 2, \dots,$$

and

$$(2.9) \quad \int_0^t \beta_k(z(s)) \, ds \leq u_k, \quad k = 1, 2, \dots.$$

Then $H_{u,t} \in \mathcal{F}_u$.

(b) Suppose a solution of (2.1) exists and is unique in the sense that for each $t > 0$ and almost every $\omega \in \Omega$, if z^1 and z^2 satisfy (2.8), then $z^1(r) = z^2(r)$, $r \leq t$. Then for all $t \geq 0$, $\tau(t) \equiv (\tau_1(t), \tau_2(t), \dots)$, with

$$(2.10) \quad \tau_k(t) = \int_0^t \beta_k(Z(s)) \, ds,$$

is an $\{\mathcal{F}_u\}$ -stopping time.

Proof. (a) Proceeding as in the proof of Lemma 2.1, let $\Gamma_{u,t} \subset S \times S$ be the set of (y, z) such that $z_k(r) = y_k(\int_0^r \beta_k(z(s)) \, ds)$, $r \leq t$, and (2.9) is satisfied. Then

$$H_{u,t} = \{Y \in \pi \Gamma_{u,t}\} \in \mathcal{F}_u.$$

(b) By the uniqueness assumption $P(\{\tau(t) \leq u\} \Delta H_{u,t}) = 0$, and hence by the completeness of \mathcal{F}_u , $\{\tau(t) \leq u\} \in \mathcal{F}_u$. \square

2.3 Remark If we drop the assumption of uniqueness, then there will in general (even in the one-dimensional case) be solutions for which $\tau(t)$ is not an $\{\mathcal{F}_u\}$ -stopping time. See Problem 1. \square

Given $Y = (Y_1, Y_2, \dots)$ on (Ω, \mathcal{F}, P) , we say (2.1) has a *weak solution* if there exists a probability space $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{P})$ on which are defined stochastic processes $\hat{Y} = (\hat{Y}_1, \hat{Y}_2, \dots)$ and $\hat{Z} = (\hat{Z}_1, \hat{Z}_2, \dots)$ such that \hat{Y} is a version of Y and

$$(2.11) \quad \hat{Z}_k(t) = \hat{Y}_k \left(\int_0^t \beta_k(\hat{Z}(s)) \, ds \right), \quad \hat{P}\text{-a.s.}$$

2.4 Proposition If (2.1) has a weak solution, then for almost every $\omega \in \Omega$ (2.1) has a solution.

Proof. As in the proof of Lemma 2.1, let $S = \prod_k D_{E_k}[0, \infty)$ and let $\Gamma \subset S \times S$ be given by (2.3). Let \hat{Y} be as above, and let $\pi\Gamma = \{y: (y, z) \in \Gamma\}$. Then

$$(2.12) \quad P\{Y \in \pi\Gamma\} = \hat{P}\{\hat{Y} \in \pi\Gamma\} = 1,$$

that is, (2.1) has a solution for almost every $\omega \in \Omega$. \square

2.5 Remark In general it may not be possible to define a version of \hat{Z} on (Ω, \mathcal{F}, P) . For example, let Ω consist of a single point, let $Y(t) = t$, and let $\beta(z) = \sqrt{z}$. Let ξ , defined on $(\Omega, \mathcal{F}, \hat{P})$, be uniformly distributed on $[0, 1]$ and define

$$(2.13) \quad \hat{Z}(t) = \begin{cases} 0, & t < \xi, \\ \frac{1}{4}(t - \xi)^2, & t \geq \xi. \end{cases}$$

Then for $\hat{Y}(t) = t$,

$$(2.14) \quad \hat{Z}(t) = \hat{Y}\left(\int_0^t \sqrt{\hat{Z}(s)} \, ds\right) = \int_0^t \sqrt{\hat{Z}(s)} \, ds,$$

but a version of \hat{Z} cannot be defined on (Ω, \mathcal{F}, P) . \square

The condition that $\tau(t)$ is in some sense a stopping time plays an important role as we examine the relationship between random time changes and corresponding martingale problems. With this in mind, we say that a stochastic process Z defined on (Ω, \mathcal{F}, P) and satisfying (2.1) is a *nonanticipating solution* of (2.1) if there exists a filtration $\{\mathcal{G}_u\}$ indexed by $u \in [0, \infty)^\omega$ such that $\mathcal{F}_u \subset \mathcal{G}_u \subset \mathcal{F}$ (\mathcal{F}_u given by (2.7)).

$$(2.15) \quad P\{(Y_1(u_1 + \cdot), Y_2(u_2 + \cdot), \dots) \in B | \mathcal{G}_u\} \\ = P\{(Y_1(u_1 + \cdot), Y_2(u_2 + \cdot), \dots) \in B | \mathcal{F}_u\}$$

for all Borel subsets B of $\prod_k D_{E_k}[0, \infty)$, and if $\tau(t)$, given by (2.10), is a $\{\mathcal{G}_u\}$ -stopping time for each $t \geq 0$.

We have three notions of solution, and hence three different notions of uniqueness. We say that *strong uniqueness* holds if for almost every $\omega \in \Omega$, (2.1) has at most one solution; we say that *weak uniqueness* holds if any two weak solutions have the same finite-dimensional distributions; and we say that we have *weak uniqueness for nonanticipating solutions* if any two weak, non-anticipating solutions have the same finite-dimensional distributions.

We turn now to the analogue of Theorem 1.3. Let Y_k , $k = 1, 2, \dots$, be independent Markov processes corresponding to semigroups $\{T_k(t)\}$. Suppose $\{T_k(t)\}$ is strongly continuous on a closed subspace $L_k \subset \bar{C}(E_k)$, and let A_k be the (strong) generator for $\{T_k(t)\}$. We assume that L_k is separating, contains the constants, and is an algebra, and that the $D_{E_k}[0, \infty)$ martingale problem for A is well-posed. By analogy with the one-dimensional case, a solution of (2.1) should be a solution of the martingale problem for

$$(2.16) \quad A \equiv \left\{ \left(\prod_{i \in I} f_i, \sum_{j \in I} \beta_j A_j f_j \prod_{i \in I - \{j\}} f_i \right) : I \subset \{1, 2, \dots\}, I \text{ finite}, f_i \in \mathcal{D}(A_i) \right\}.$$

2.6 Lemma Let Y_1, Y_2, \dots be independent Markov processes (as above) defined on (Ω, \mathcal{F}, P) . Then a stochastic process Z satisfying (2.1) is a nonanticipating solution if and only if for every $t \geq 0$, $\tau(t)$ is a $\{\mathcal{G}_u\}$ -stopping time for some $\{\mathcal{G}_u\}$ satisfying

$$(2.17) \quad E \left[\prod_{k \in I} f_k(Y_k(u_k + v_k)) \middle| \mathcal{G}_u \right] = \prod_{k \in I} T_k(v_k) f_k(Y_k(u_k))$$

for all finite $I \subset \{1, 2, \dots\}$, $f_k \in L_k$, and $u_k, v_k \geq 0$, or, setting $H_k f_k = A_k f_k / f_k$,

$$(2.18) \quad E \left[\prod_{k \in I} f_k(Y_k(u_k + v_k)) \exp \left\{ - \int_{u_k}^{u_k + v_k} H_k f_k(Y_k(s)) ds \right\} \middle| \mathcal{G}_u \right] \\ = \prod_{k \in I} f_k(Y_k(u_k))$$

for all finite $I \subset \{1, 2, \dots\}$, $f_k \in \mathcal{D}^+(A_k)$, and $u_k, v_k \geq 0$. ($\mathcal{D}^+(A_k) = \{f \in \mathcal{D}(A_k) : \inf_{x \in E_k} f(x) > 0\}$.)

Proof. The equivalence of (2.17) and (2.15) follows from the Markov property and the independence of the Y_k . The equivalence of (2.18) and (2.15) follows from the uniqueness for the martingale problem for A_k and the independence of the Y_k . If \mathcal{G}_u does not contain \mathcal{F}_u , then $\mathcal{G}_u \vee \mathcal{F}_u$ still satisfies (2.17) and (2.18). In particular, \mathcal{G}_u can be replaced by \mathcal{G}_u^0 where \mathcal{G}_u^0 is obtained from \mathcal{G}_u as \mathcal{F}_u is obtained from \mathcal{F}_u^0 . See (2.7). \square

2.7 Lemma Let Y_1, Y_2, \dots be independent Markov processes (as above). A stochastic process Z satisfying (2.1) is a nonanticipating solution if and only if

$$(2.19) \quad E \left[\prod_{k \in I} f_k(Y_k(u_k + v_k)) \prod_i g_{ik}(Y_k(s_{ik})) \prod_j h_j \left(\min_{t \in I} (u_t - \tau_i(t_j)) \vee 0 \right) \right] \\ = E \left[\prod_{k \in I} T_k(v_k) f_k(Y_k(u_k)) \prod_i g_{ik}(Y_k(s_{ik})) \prod_j h_j \left(\min_{t \in I} (u_t - \tau_i(t_j)) \vee 0 \right) \right]$$

for all $u_k, v_k \geq 0$, $0 \leq s_{ik} \leq u_k$, $t_j \geq 0$, finite $I \subset \{1, 2, \dots\}$, $f_k \in L_k$, $g_{ik} \in \bar{C}(E_k)$, and $h_j \in \hat{C}[0, \infty)$, or

$$(2.20) \quad E \left[\prod_{k \in I} f_k(Y_k(u_k + v_k)) \exp \left\{ - \int_{u_k}^{u_k + v_k} H_k f_k(Y_k(s)) ds \right\} \prod_i g_{ik}(Y_k(s_{ik})) \right. \\ \times \left. \prod_j h_j \left(\min_{t \in I} (u_t - \tau_i(t_j)) \vee 0 \right) \right] \\ = E \left[\prod_{k \in I} f_k(Y_k(u_k)) \prod_i g_{ik}(Y_k(s_{ik})) \prod_j h_j \left(\min_{t \in I} (u_t - \tau_i(t_j)) \vee 0 \right) \right]$$

for all $u_k, v_k \geq 0$, $0 \leq s_{ik} \leq u_k$, $t_j \geq 0$, finite $I \subset \{1, 2, \dots\}$, $f_k \in \mathcal{D}^+(A_k)$, $g_{ik} \in \bar{C}(E_k)$, and $h_j \in \bar{C}[0, \infty)$.

Proof. The necessity of (2.19) and (2.20) is immediate from Lemma 2.6. Define

$$(2.21) \quad \mathcal{G}_{u, I}^0 = \sigma(Y_k(s_k) : s_k \leq u_k, k \in I) \vee \sigma\left(\min_{k \in I} (u_k - \tau_k(t)) \vee 0 : t \geq 0\right).$$

Then (2.19) implies

$$(2.22) \quad E\left[\prod_{k \in I} f_k(Y_k(u_k + v_k)) \middle| \mathcal{G}_{u, I}^0\right] = \prod_{k \in I} T_k(v_k) f_k(Y_k(u_k)).$$

Fix I . If $I' \supset I$, then, by taking $f_k \equiv 1$ for $k \in I' - I$, we can replace $\mathcal{G}_{u, I}^0$ in (2.22) by $\mathcal{G}_{u, I'}^0$. If $u_k = \infty$ for $k \notin I$, then, for $I' \supset I$, $\min_{k \in I'} (u_k - \tau_k(t)) \vee 0 = \min_{k \in I} (u_k - \tau_k(t)) \vee 0$, and hence $\mathcal{G}_{u, I'}^0$, with $I' \supset I$, is increasing in I' . For u satisfying $u_k = \infty$, $k \notin I$, we define

$$(2.23) \quad \mathcal{G}_u^0 = \bigvee_{I' \supset I} \mathcal{G}_{u, I'}^0,$$

and we note that we can replace $\mathcal{G}_{u, I}^0$ in (2.22) by \mathcal{G}_u^0 . For arbitrary u define

$$(2.24) \quad \mathcal{G}_u = \bigcap_n \mathcal{G}_{u^{(n)}}^0$$

where $u_k^{(n)} = u_k + 1/n$ for $k \leq n$, $u_k = \infty$ for $k > n$. If $I \subset \{1, 2, \dots, n\}$, we have

$$(2.25) \quad E\left[\prod_{k \in I} f_k\left(Y_k\left(u_k + \frac{1}{n} + v_k\right)\right) \middle| \mathcal{G}_{u^{(n)}}^0\right] = \prod_{k \in I} T_k(v_k) f_k\left(Y_k\left(u_k + \frac{1}{n}\right)\right).$$

The right continuity of Y_k , the continuity of f_k and $T_k(v_k)f_k$, and the fact that $\mathcal{G}_{u^{(n)}}^0$ is decreasing in n imply (2.17). A similar argument shows that (2.20) implies (2.18). Finally $\{\tau(t) \leq u\} = \bigcap_I \bigcap_n \{\min_{k \in I} (u_k^{(n)} - \tau_k(t)) > 0\} \in \mathcal{G}_u$. \square

2.8 Theorem Let Y_k , $k = 1, 2, \dots$, be independent Markov processes (as above). Let β_k , $k = 1, 2, \dots$ be nonnegative bounded Borel functions on $E \equiv \prod_I E_I$, and let A be given by (2.16).

(a) If Z is a nonanticipating solution of (2.1), then Z is a solution of the martingale problem for A .

(b) If Z is a solution of the $D_E[0, \infty)$ martingale problem for A , then there is a version of Z that is a (weak) nonanticipating solution of (2.1).

2.9 Remark (a) If $\inf_z \beta_k(z) > 0$, $k = 1, 2, \dots$, in (b), then Z itself is a nonanticipating solution of (2.1).

(b) The hypothesis that $\sup_z \beta_k(z) < \infty$ is used to ensure that $A \subset \bar{C}(E) \times B(E)$. There are two approaches toward eliminating this hypothesis. One

would be to restrict the domain of A so that $A \subset \bar{C}(E) \times B(E)$. (See Problems 2, 3.) The other would be to develop the notion of a "local-martingale problem." (See Chapter 4, Section 7, and Proposition 2.10.)

Proof. (a) Let $I \subset \{1, 2, \dots\}$ be finite. For $k \in I$, let $f_k \in \mathcal{D}(A_k)$, $\inf_{x \in E_k} f_k(x) > 0$, and set $H_k f_k = A_k f_k / f_k$. Define $f = \prod_{k \in I} f_k$, $g = \sum_{k \in I} \beta_k A_k f_k / \prod_{j \in I - \{k\}} f_j$, and

$$(2.26) \quad M(u) = \prod_{k \in I} f_k(Y_k(u_k)) \exp \left\{ - \int_0^{u_k} H_k f_k(Y_k(s)) ds \right\}.$$

By (2.17), for fixed u , $Y_k^{(u)}(v_k) = Y_k(u_k + v_k)$, $k = 1, 2, \dots$, are Markov processes corresponding to A_k and are conditionally independent given \mathcal{G}_u . Therefore

$$\begin{aligned} (2.27) \quad E[M(u+v)|\mathcal{G}_u] &= \prod_{k \in I} E \left[f_k(Y_k(u_k + v_k)) \exp \left\{ - \int_0^{u_k + v_k} H_k f_k(Y_k(s)) ds \right\} \middle| \mathcal{G}_u \right] \\ &= \prod_{k \in I} f_k(Y_k(u_k)) \exp \left\{ - \int_0^{u_k} H_k f_k(Y_k(s)) ds \right\} \\ &= M(u), \end{aligned}$$

and hence $M(u)$ is a $\{\mathcal{G}_u\}$ -martingale.

By assumption, $\tau(t)$ is a $\{\mathcal{G}_u\}$ -stopping time and the optional sampling theorem (Theorem 8.7 of Chapter 2) implies $M(\tau(t))$ is a $\mathcal{G}_{\tau(t)}$ -martingale. But

$$(2.28) \quad M(\tau(t)) = \prod_{k \in I} f_k(Z_k(t)) \exp \left\{ - \int_0^t \beta_k(Z(s)) H_k f_k(Z_k(s)) ds \right\}$$

and

$$(2.29) \quad f(Z(t)) - \int_0^t g(Z(s)) ds$$

is a martingale by Lemma 3.2 of Chapter 4.

(b) The basic idea of the proof is to let

$$(2.30) \quad \gamma_k(u) = \inf \left\{ t: \int_0^t \beta_k(Z(s)) ds > u \right\}$$

and define

$$(2.31) \quad Y_k(u) = Z_k(\gamma_k(u)).$$

Note that then, as in the one-dimensional case, the fact that $\mathcal{D}(A_k)$ is separating implies Z satisfies (2.1).

Two difficulties arise. First, $\gamma_k(u)$ need not be defined for all $u \geq 0$, and second, even if γ_k is defined for all $u \geq 0$, it is not immediately clear that the Y_k are independent. Note that for each $t > 0$ and $(f, g) \in A_k$,

$$(2.32) \quad M(u) = f(Z_k(\gamma_k(u) \wedge t)) - \int_0^{\gamma_k(u) \wedge t} \beta_k(Z(s))g(Z_k(s)) ds \\ = f(Y_k(u \wedge \tau_k(t))) - \int_0^{u \wedge \tau_k(t)} g(Y_k(s)) ds$$

is an $\{\mathcal{F}_{\gamma_k(u)}^Z\}$ -martingale. By Lemma 5.16 of Chapter 4, there exists a solution $\tilde{Y}_{k,t}$ of the martingale problem for A_k and a nonnegative random variable $\eta_k(t)$ such that $Y_k(\cdot \wedge \tau_k(t))$ has the same distribution as $\tilde{Y}_{k,t}(\cdot \wedge \eta_k(t))$. Letting $t \rightarrow \infty$, $(\tilde{Y}_{k,t}, \eta_k(t))$ converges in distribution in $D_{E_k}[0, \infty) \times [0, \infty]$ (at least through a sequence of t 's) to $(\tilde{Y}_{k,\infty}, \eta_k(\infty))$ and $Y_k(\cdot \wedge \tau_k(\infty))$ has the same distribution as $\tilde{Y}_{k,\infty}(\cdot \wedge \eta_k(\infty))$. In particular,

$$(2.33) \quad Z_k(\infty) \equiv \lim_{t \rightarrow \infty} Z_k(t) = \lim_{u \rightarrow \infty} Y_k(u \wedge \tau_k(\infty))$$

exists on $\{\tau_k(\infty) < \infty\}$. Fix $y_k \in E_k$ and set $Z_k(\infty) = y_k$ on $\{\tau_k(\infty) = \infty\}$.

Let $\Omega' = \Omega \times \prod_k D_{E_k}[0, \infty)$ and define

$$(2.34) \quad Q(C \times B_1 \times B_2 \times B_3 \times \dots) \\ = \int_C \prod_k P_{Z_k(\infty)}^{(k)}(B_k) dP$$

for $C \in \mathcal{F}$ and $B_k \in \mathcal{B}(D_{E_k}[0, \infty))$, where $P_y^{(k)}$ is the distribution of the Markov process with generator A_k starting from y . Then Q extends to a measure on $\mathcal{F} \times \prod_k \mathcal{B}(D_{E_k}[0, \infty))$.

Defining Z on Ω' by $Z(t, (\omega, \omega_1, \omega_2, \dots)) \equiv Z(t, \omega)$, we see that Z on $(\Omega', \mathcal{F} \times \prod_k \mathcal{B}(D_{E_k}[0, \infty)), Q)$ is a version of Z on (Ω, \mathcal{F}, P) . Let W_k denote the coordinate process in $D_{E_k}[0, \infty)$, that is,

$$(2.35) \quad W_k(t, (\omega, \omega_1, \omega_2, \omega_3, \dots)) = \omega_k(t).$$

Set

$$(2.36) \quad \tau_k(t) = \int_0^t \beta_k(Z(s)) ds,$$

allowing $t = \infty$, and define

$$(2.37) \quad Y_k(t) = \begin{cases} Z_k(\gamma_k(t)), & t < \tau_k(\infty), \\ W_k(t - \tau_k(\infty)), & t \geq \tau_k(\infty). \end{cases}$$

We must show that there is a family of σ -algebras $\{\mathcal{G}_u\}$ such that $\tau(t)$ is a $\{\mathcal{G}_u\}$ -stopping time and the Y_k satisfy (2.17).

Let $f_k \in \mathcal{D}(A_k)$, $f_k > 0$. Let $\psi_k(x_k, t)$, $(\partial/\partial t)\psi_k(x_k, t) \in \bar{\mathcal{C}}(E_k \times [0, \infty))$ with $\psi_k \geq 0$, and suppose $q_k(x_k, t)$ satisfies

$$(2.38) \quad \frac{\partial}{\partial t} q_k(x_k, t) = A_k q_k(x_k, t) - \psi_k(x_k, t) q_k(x_k, t),$$

and $q_k(x_k, 0) = f_k(x_k)$. For $u \in [0, \infty)^\infty$ define

$$(2.39) \quad h_k(x_k, t) = \begin{cases} q_k(x_k, u_k - t), & t \leq u_k, \\ f_k(x_k), & t > u_k, \end{cases}$$

and set $Kh_k = (\partial/\partial t)h_k/h_k$. Setting

$$(2.40) \quad M_k(t) = h_k(Z_k(t), \tau_k(t)) \exp \left\{ - \int_0^t \beta_k(Z(s)) [Kh_k(Z_k(s), \tau_k(s)) \right. \\ \left. + H_k h_k(Z_k(s), \tau_k(s))] ds \right\},$$

Lemmas 3.2 and 3.4 of Chapter 4 imply that

$$(2.41) \quad \prod_{k \in I} M_k(t)$$

is a martingale for any finite $I \subset \{1, 2, \dots\}$, with respect to $\{\mathcal{F}_t^Z\}$. Defining $\gamma_k(u)$ by (2.30) for $u < \tau_k(\infty)$ and setting $\gamma_k(u) = \infty$ for $u \geq \tau_k(\infty)$, Problem 24 of Chapter 2 implies

$$(2.42) \quad E \left[\prod_{k \in I} M_k(\gamma_k(v_k)) \middle| \mathcal{F}_{\gamma_I(u)}^Z \right] = \prod_{k \in I} M_k(\gamma_I(u))$$

for $u \leq v$, where $\gamma_I(u) = \bigwedge_{k \in I} \gamma_k(u_k)$. In particular, from the definition of h_k and q_k ,

$$(2.43) \quad E \left[\prod_{k \in I} h_k(Y_k(u_k \wedge \tau_k(\infty)), v_k \wedge \tau_k(\infty)) \exp \left\{ - \int_{u_k \wedge \tau_k(\infty)}^{v_k \wedge \tau_k(\infty)} H_k f_k(Y_k(s)) ds \right\} \right. \\ \times \exp \left\{ - \int_0^{u_k \wedge \tau_k(\infty)} \psi_k(Y_k(s), u_k \wedge \tau_k(\infty) - s) ds \right\} \left. \middle| \mathcal{F}_{\gamma_I(u)}^Z \right] \\ = E \left[\prod_{k \in I} h_k(Y_k(u_k \wedge \tau_k(\infty)), u_k \wedge \tau_k(\infty)) \right. \\ \times \exp \left\{ - \int_0^{u_k \wedge \tau_k(\infty)} \psi_k(Y_k(s), u_k \wedge \tau_k(\infty) - s) ds \right\} \left. \middle| \mathcal{F}_{\gamma_I(u)}^Z \right].$$

Observing that

$$\begin{aligned}
 (2.44) \quad & E \left[\prod_{k \in I} f_k(Y_k(v_k)) \exp \left\{ - \int_{u_k}^{v_k} H_k f_k(Y_k(s)) ds \right\} \right. \\
 & \times \exp \left\{ - \int_0^{u_k} \psi_k(Y_k(s), u_k - s) ds \right\} \Big| \mathcal{F}_\infty^Z \\
 & = E \left[\prod_{k \in I} h_k(Y_k(v_k \wedge \tau_k(\infty)), v_k \wedge \tau_k(\infty)) \exp \left\{ - \int_{u_k \wedge \tau_k(\infty)}^{v_k \wedge \tau_k(\infty)} H_k f_k(Y_k(s)) ds \right\} \right. \\
 & \times \exp \left\{ - \int_0^{u_k \wedge \tau_k(\infty)} \psi_k(Y_k(s), u_k \wedge \tau_k(\infty) - s) ds \right\} \Big| \mathcal{F}_\infty^Z,
 \end{aligned}$$

we see that we can drop the " $\wedge \tau_k(\infty)$ " on both sides of (2.43).

Let $\varphi_{ik}(x_k) \in L_k$ satisfy $0 < \varphi_{ik} \leq 1$, and let $0 \leq s_{ik} \leq u_k$. Let $\rho \geq 0$ be continuously differentiable with compact support in $(0, \infty)$ and $\int_0^\infty \rho(s) ds = 1$. Replace ψ_k in (2.38) by

$$(2.45) \quad \psi_k^{(n)}(x_k, t) = - \sum_{i=1}^m n \rho((u_k - t - s_{ik})n) \varphi_{ik}(x_k).$$

Since L_k is an algebra, $B_n(t)f \equiv \psi_k^{(n)}(t, \cdot)f$ defines a bounded linear operator on L_k , and the differentiability of ρ ensures the existence of q_k (see Problem 23 of Chapter 1). Letting $n \rightarrow \infty$ in (2.43) gives

$$\begin{aligned}
 (2.46) \quad & E \left[\prod_{k \in I} f_k(Y_k(v_k)) \exp \left\{ - \int_{u_k}^{v_k} H_k f_k(Y_k(s)) ds - \sum_i \varphi_{ik}(Y_k(s_{ik})) \right\} \right| \mathcal{F}_{Y_I(u)}^Z \\
 & = E \left[\prod_{k \in I} f_k(Y_k(u_k)) \exp \left\{ - \sum_i \varphi_{ik}(Y_k(s_{ik})) \right\} \right| \mathcal{F}_{Y_I(u)}^Z.
 \end{aligned}$$

Setting

$$(2.47) \quad \mathcal{G}_u = \sigma(Y_k(s_k) : s_k \leq u_k) \vee \bigcap_k \mathcal{F}_{Y_k(u_k)}^Z,$$

we note that (2.46) implies (2.18) and that

$$(2.48) \quad \{\tau(t) \leq u\} = \left\{ \bigwedge_k \gamma_k(u_k) \geq t \right\} \in \bigcap_k \mathcal{F}_{Y_k(u_k)}^Z \subset \mathcal{G}_u.$$

Part (b) now follows by Lemma 2.6. \square

The following proposition is useful in reducing the study of (2.1) with unbounded β_k to the bounded case.

2.10 Proposition Let α be measurable, and suppose $\inf_z \alpha(z) > 0$. Let Z be an E -valued stochastic process, let η satisfy

$$(2.49) \quad \int_0^{\eta(t)} \alpha(Z(s)) \, ds = t,$$

$\lim_{t \rightarrow \infty} \eta(t) = \infty$ a.s., and define

$$(2.50) \quad Z^\alpha(t) = Z(\eta(t)).$$

Then Z is a nonanticipating solution of (2.1) if and only if Z^α is a nonanticipating solution of

$$(2.51) \quad Z_k^\alpha(t) = Y_k \left(\int_0^t \frac{\beta_k(Z^\alpha(s))}{\alpha(Z^\alpha(s))} \, ds \right).$$

Proof. If Z satisfies (2.1), a simple change of variable verifies that Z^α satisfies (2.51). Assume Z is a nonanticipating solution, and let $\{\mathcal{G}_u\}$ be the family of σ -algebras in Lemma 2.6. Since the $\tau(t)$ form an increasing family of $\{\mathcal{G}_u\}$ -stopping times, and for each s , $\eta(s)$ is a $\{\mathcal{G}_{\tau(s)}\}$ -stopping time, Proposition 8.6 of Chapter 2 gives that $\tau^\alpha(s) = \tau(\eta(s))$ is a $\{\mathcal{G}_u\}$ -stopping time. Consequently, by Lemma 2.6, Z^α is a nonanticipating solution of (2.51).

The converse is proved similarly. \square

3. CONVERGENCE

We now consider criteria for convergence of a sequence of processes $Z^{(n)}$ satisfying

$$(3.1) \quad Z_k^{(n)}(t) = Y_k^{(n)} \left(\int_0^t \beta_k^{(n)}(Z^{(n)}(s)) \, ds \right), \quad k = 1, 2, \dots$$

where $Y_k^{(n)}$ is a process with sample paths in $D_{E_k}[0, \infty)$. We continue to assume that the (E_k, r_k) are complete and separable. Relative compactness for sequences of this form is frequently quite simple.

3.1 Proposition Let $Z^{(n)}$ satisfy (3.1). If $\{Y_k^{(n)}\}$ is relatively compact in $D_{E_k}[0, \infty)$ and $\bar{\beta}_k \equiv \sup_n \sup_z \beta_k^{(n)}(z) < \infty$, then $\{Z_k^{(n)}\}$ is relatively compact in $D_{E_k}[0, \infty)$, and hence if $\{Y^{(n)}\}$ is relatively compact in $\prod_k D_{E_k}[0, \infty)$ and $\sup_n \sup_z \beta_k^{(n)}(z) < \infty$ for each k , then $\{Z^{(n)}\}$ is relatively compact in $\prod_k D_{E_k}[0, \infty)$.

Proof. The proposition follows immediately from the fact that

$$(3.2) \quad w'(Z_k^{(n)}, \delta, T) \leq w'(Y_k^{(n)}, \bar{\beta}_k \delta, \bar{\beta}_k T)$$

and

$$(3.3) \quad \{Z_k^{(n)}(t) \in K \text{ for all } t \leq T\} \supset \{Y_k^{(n)}(t) \in K \text{ for all } t \leq \beta T\}.$$

(Recall that we are assuming the (E_k, r_k) are complete and separable.) \square

We would prefer, of course, to have relative compactness in $D_E[0, \infty)$ where $E = \prod_k E_k$, but relative compactness of $\{Y^{(n)}\}$ in $D_E[0, \infty)$ and the boundedness of the $\beta_k^{(n)}$ do not necessarily imply the relative compactness of $\{Z^{(n)}\}$ in $D_E[0, \infty)$. We do note the following.

3.2 Proposition Let $\{Z^{(n)}\}$ be a sequence of processes with sample paths in $D_E[0, \infty)$, $E = \prod_k E_k$. If $Z^{(n)} \Rightarrow Z$ in $\prod_k D_{E_k}[0, \infty)$ and if no two components of Z have simultaneous jumps (i.e., if $P\{Z_k(t) \neq Z_k(t-) \text{ and } Z_l(t) \neq Z_l(t-) \text{ for some } t \geq 0\} = 0$ for all $k \neq l$), then $Z^{(n)} \Rightarrow Z$ in $D_E[0, \infty)$.

Proof. The result follows from Proposition 6.5 of Chapter 3. Details are left to the reader (Problem 5). \square

We next give the analogue of Theorem 1.5.

3.3 Theorem Suppose that for $k = 1, 2, \dots$, Y_k , defined on (Ω, \mathcal{F}, P) , has sample paths in $D_{E_k}[0, \infty)$, β_k is nonnegative, bounded, and continuous on $E = \prod_k E_k$, and either Y_k is continuous or $\beta_k(z) > 0$ for all $z \in E$. Suppose that for almost every $\omega \in \Omega$,

$$(3.4) \quad Z_k(t) = Y_k \left(\int_0^t \beta_k(Z(s)) ds \right)$$

has a unique solution. Let $\{Y^{(n)}\}$ satisfy $Y^{(n)} \Rightarrow Y$ in $\prod_k D_{E_k}[0, \infty)$, and for $k = 1, 2, \dots$, let $\beta_k^{(n)}$ be nonnegative Borel measurable functions satisfying $\sup_n \sup_{z \in K} \beta_k^{(n)}(z) < \infty$ and

$$(3.5) \quad \lim_{n \rightarrow \infty} \sup_{z \in K} |\beta_k^{(n)}(z) - \beta_k(z)| = 0$$

for each compact $K \subset E$. Suppose that $Z^{(n)}$ satisfies

$$(3.6) \quad Z_k^{(n)}(t) = Y_k^{(n)} \left(\int_0^t \beta_k^{(n)}(Z^{(n)}(s)) ds \right)$$

and that $W^{(n)}$ satisfies

$$(3.7) \quad W_k^{(n)}(t) = Y_k^{(n)} \left(\int_0^{(t/h_n)h_n} \beta_n(W^{(n)}(s)) ds \right),$$

where $h_n > 0$ and $\lim_{n \rightarrow \infty} h_n = 0$. Then $Z^{(n)} \Rightarrow Z$ and $W^{(n)} \Rightarrow Z$ in $\prod_k D_{E_k}[0, \infty)$.

Proof. The proof is essentially the same as for Theorem 1.3, so we only give a sketch. We may assume $\lim_{n \rightarrow \infty} Y^{(n)} = Y$ a.s. The estimates in (3.2) and (3.3) imply that if $\{Y^{(n)}(\omega)\}$ is convergent in $\prod_k D_{E_k}[0, \infty)$ for some $\omega \in \Omega$, then $\{Z^{(n)}(\omega)\}$ and $\{W^{(n)}(\omega)\}$ are relatively compact in $\prod_k D_{E_k}[0, \infty)$. The continuity and positivity of β_k imply that any limit point $\hat{Z}(\omega)$ of $\{Z^{(n)}(\omega)\}$ or $\{W^{(n)}(\omega)\}$ must satisfy

$$(3.8) \quad \hat{Z}_k(\omega, t) = Y_k\left(\omega, \int_0^t \beta_k(\hat{Z}(\omega, s)) ds\right).$$

(If Y_k is not continuous, then the positivity of β_k implies $Y_k(\omega, \int_0^t \beta_k(\hat{Z}(\omega, s)) ds) = Y_k(\omega, \int_0^t \beta_k(\hat{Z}(\omega, s)) ds)$ for almost every $t \geq 0$. See Problem 6.) Since the solution of (3.4) is almost surely unique, it follows that $\lim_{n \rightarrow \infty} Z^{(n)} = Z$ and $\lim_{n \rightarrow \infty} W^{(n)} = Z$ in $\prod_k D_{E_k}[0, \infty)$ a.s. \square

The proof of Theorem 3.3 is typical of proofs of weak convergence: compactness is verified, it is shown that any possible limit must possess certain properties, and finally it is shown (or in this case assumed) that those properties uniquely determine the possible limit. The uniqueness used above was strong uniqueness. Unfortunately, there are many situations in which weak uniqueness for nonanticipating solutions is known but not strong uniqueness. Consequently we turn now to convergence criteria in which the limiting process is characterized as the unique weak, nonanticipating solution.

We want to cover not only sequences of the form (3.6) and (3.7) but also solutions of equations of the form

$$(3.9) \quad Z_k^{(n)}(t) = Y_k^{(n)}\left(\int_0^t \beta_k^{(n)}(Z^{(n)}(s), \xi^{(n)}(s)) ds\right)$$

where $\xi^{(n)}$ is a rapidly fluctuating process that "averages" $\beta_k^{(n)}$ in the sense that

$$(3.10) \quad \left| \int_0^t \beta_k^{(n)}(Z^{(n)}(s), \xi^{(n)}(s)) ds - \int_0^t \beta_k(Z^{(n)}(s)) ds \right| \rightarrow 0.$$

The following theorem provides conditions for convergence that apply to all three of these situations.

3.4 Theorem Let $Y^{(n)}$, $n = 1, 2, \dots$, have values in $\prod_k D_{E_k}[0, \infty)$, let $\{\mathcal{F}_u^{(n)}\}$ be a filtration indexed by $[0, \infty)^{\omega}$ satisfying $\mathcal{F}_u^{(n)} \supset \sigma(Y_k^{(n)}(s_k); s_k \leq u_k, k = 1, 2, \dots)$, and let $\tau^{(n)}(t)$, $t \geq 0$, be a nondecreasing (componentwise) family of $\{\mathcal{F}_u^{(n)}\}$ -stopping times that is right continuous in t . Define

$$(3.11) \quad Z_k^{(n)}(t) = Y_k^{(n)}(\tau_k^{(n)}(t)).$$

Suppose for $k = 1, 2, \dots$ that $\{T_k(t)\}$ is a strongly continuous semigroup on $L_k \subset \bar{C}(E_k)$ corresponding to a Markov process Y_k , and L_k is convergence determining, that $\beta_k: E \rightarrow [0, \infty)$ is continuous, and that either $\beta_k > 0$ or Y_k is continuous.

Assume

$$(3.12) \quad \lim_{n \rightarrow \infty} E \left[\left| E \left[\prod_{k \in I} f_k(Y_k^{(n)}(u_k + v_k)) \middle| \mathcal{G}_u^{(n)} \right] - \prod_{k \in I} T_k(v_k) f_k(Y_k^{(n)}(u_k)) \right| \right] = 0$$

for $f_k \in L_k$, finite $I \subset \{1, 2, \dots\}$, and $u, v \in [0, \infty)^\omega$, and assume

$$(3.13) \quad \left| \tau_k^{(n)}(t) - \int_0^t \beta_k(Z^{(n)}(s)) ds \right|^p \rightarrow 0$$

for each $k = 1, 2, \dots$ and $t \geq 0$.

(a) If $(Y^{(n)}, Z^{(n)}) \Rightarrow (Y, Z)$ in $\prod_k D_{E_k}[0, \infty) \times \prod_k D_{E_k}[0, \infty)$, then Z is a nonanticipating solution of (2.1).

(b) Suppose that for each $\varepsilon, T > 0$ and $k = 1, 2, \dots$ there exists a compact $K_{t, T}^k \subset E_k$ such that

$$(3.14) \quad \inf_n P\{\tau_k^{(n)}(t) \in K_{t, T}^k \text{ for all } t \leq T\} \geq 1 - \varepsilon.$$

If $Y^{(n)} \Rightarrow Y$ in $\prod_k D_{E_k}[0, \infty)$, and (2.1) has a weakly unique nonanticipating solution Z , then $Z^{(n)} \Rightarrow Z$ in $\prod_k D_{E_k}[0, \infty)$.

3.5 Remark (a) Note that (3.12) implies that the finite-dimensional distributions of $Y^{(n)}$ converge and that the Y_k are conditionally independent given $Y(0)$. See Remark 8.3(a) of Chapter 4 for conditions implying (3.12).

(b) If the $Y_k^{(n)}$ are Markov processes satisfying

$$(3.15) \quad E \left[\prod_{k \in I} f_k(Y_k^{(n)}(u_k + v_k)) \middle| \mathcal{G}_u^{(n)} \right] = \prod_{k \in I} T_k^{(n)}(v_k) f_k(Y_k^{(n)}(u_k)),$$

then (3.12) is implied by

$$(3.16) \quad \lim_{n \rightarrow \infty} E[|T_k^{(n)}(t)f_k(Y_k^{(n)}(u)) - T_k(t)f_k(Y_k^{(n)}(u))|] = 0$$

for all $t, u \geq 0$ and $k = 1, 2, 3, \dots$. \square

Proof. (a) If $(Y^{(n)}, Z^{(n)}) \Rightarrow (Y, Z)$, then (3.13) and the continuity of the β_k imply $(Y^{(n)}, Z^{(n)}, \tau^{(n)}) \Rightarrow (Y, Z, \tau)$ in $\prod_k D_{E_k}[0, \infty) \times \prod_k D_{E_k}[0, \infty) \times [D_{[0, \infty)}[0, \infty)]^\omega$, where τ_k is as usual

$$(3.17) \quad \tau_k(t) = \int_0^t \beta_k(Z(s)) ds.$$

It follows that $Z_k(t) = Y_k(\tau_k(t))$ or $Y_k^-(\tau_k(t))$. We need $Z_k(t) = Y_k(\tau_k(t))$. If Y_k is continuous, then (2.1) is satisfied; or if $\beta_k > 0$, then the fact that τ_k is

(strictly) increasing and $Z_k(t)$ and $Y_k(\tau_k(t))$ are right continuous implies (2.1) is satisfied.

To see that Z is nonanticipating, note that with the parameters as in (2.19)

$$\begin{aligned}
 (3.18) \quad & E \left[\prod_{k \in I} f_k(Y_k(u_k + v_k)) \prod_i g_{ik}(Y_k(s_{ik})) \right. \\
 & \quad \cdot \left. \prod_j h_j \left(\min_{t \in I} (u_t - \tau_k(t_j)) \vee 0 \right) \right] \\
 = & \lim_{n \rightarrow \infty} E \left[\prod_{k \in I} f_k(Y_k^{(n)}(u_k + v_k)) \prod_i g_{ik}(Y_k^{(n)}(s_{ik})) \right. \\
 & \quad \cdot \left. \prod_j h_j \left(\min_{t \in I} (u_t - \tau_k^{(n)}(t_j)) \vee 0 \right) \right] \\
 = & \lim_{n \rightarrow \infty} E \left[E \left[\prod_{k \in I} f_k(Y_k^{(n)}(u_k + v_k)) \middle| \mathcal{G}_u^{(n)} \right] \prod_i g_{ik}(Y_k^{(n)}(s_{ik})) \right. \\
 & \quad \cdot \left. \prod_j h_j \left(\min_{t \in I} (u_t - \tau_k^{(n)}(t_j)) \vee 0 \right) \right] \\
 = & \lim_{n \rightarrow \infty} E \left[\prod_{k \in I} T_k(v_k) f_k(Y_k^{(n)}(u_k)) \prod_i g_{ik}(Y_k^{(n)}(s_{ik})) \right. \\
 & \quad \cdot \left. \prod_j h_j \left(\min_{t \in I} (u_t - \tau_k^{(n)}(t_j)) \vee 0 \right) \right] \\
 = & E \left[\prod_{k \in I} T_k(v_k) f_k(Y_k(u_k)) \prod_i g_{ik}(Y_k(s_{ik})) \right. \\
 & \quad \cdot \left. \prod_j h_j \left(\min_{t \in I} (u_t - \tau_k(t_j)) \vee 0 \right) \right].
 \end{aligned}$$

Observe that the τ_i are continuous and that $P\{Y_k(t) = Y_k(t-)\} = 1$ (cf. Theorem 3.12 of Chapter 4) for all t . Consequently all the finite-dimensional distributions of $(Y^{(n)}, \tau^{(n)})$ converge to those of (Y, τ) . By Lemma 2.7, Z is a nonanticipating solution of (2.1).

(b) By part (a), it is enough to show that $\{(Y^{(n)}, Z^{(n)})\}$ is relatively compact, since any convergent subsequence must converge to the unique nonanticipating solution of (2.1). By Proposition 2.4 of Chapter 3, it is enough to verify the relative compactness of $\{Z_k^{(n)}\}$. Let

$$(3.19) \quad \gamma_k^{(n)}(t) = \int_0^t \beta_k(Z^{(n)}(s)) \, ds.$$

The monotonicity of $\gamma_k^{(n)}$ and $\tau_k^{(n)}$ and (3.14) imply the convergence in (3.13) is uniform in t on bounded intervals. For $\delta, T > 0$, let

$$(3.20) \quad \eta_k^{(n)}(\delta, T) = \sup_{t \leq T} (\tau_k^{(n)}(t + \delta) - \tau_k^{(n)}(t)) + \sup_{t \leq T} (\tau_k^{(n)}(t) - \tau_k^{(n)}(t -)).$$

Note that by the uniformity in t in (3.13) and (3.14), as $n \rightarrow \infty$ and $\delta \rightarrow 0$, $\eta_k^{(n)}(\delta, T) \xrightarrow{P} 0$. Finally

$$(3.21) \quad w'(Z_k^{(n)}, \delta, T) \leq w'(Y_k^{(n)}, \eta_k^{(n)}(\delta, T), \tau_k^{(n)}(T))$$

(see Problem 7), and hence for $\varepsilon > 0$ the relative compactness of $\{Y_k^{(n)}\}$ implies

$$\begin{aligned} (3.22) \quad & \lim_{\delta \rightarrow 0} \overline{\lim}_{n \rightarrow \infty} P\{w'(Z_k^{(n)}, \delta, T) > \varepsilon\} \\ & \leq \lim_{\delta \rightarrow 0} \overline{\lim}_{n \rightarrow \infty} P\{w'(Y_k^{(n)}, \eta_k^{(n)}(\delta, T), \tau_k^{(n)}(T)) > \varepsilon\} \\ & = 0 \end{aligned}$$

and the relative compactness of $\{Z_k^{(n)}\}$ follows. \square

3.6 Corollary Let Y_1, Y_2, \dots be independent Markov processes (as above), let $\beta_k: E \rightarrow [0, \infty)$ be continuous and bounded, and assume either $\beta_k > 0$ or Y_k is continuous. Then (2.1) has a weak, nonanticipating solution.

Proof. Let $Y^{(n)} = Y$ and $W^{(n)}$ satisfy (3.7) with $h_n = 1/n$. Then $\{W_k^{(n)}\}$ is relatively compact by essentially the same estimates as in the proof of Proposition 3.1. Any limit point of $\{W^{(n)}\}$ is a nonanticipating solution of (2.1). \square

4. MARKOV PROCESSES IN \mathbf{Z}^d

Let E be the one-point compactification of the d -dimensional integer lattice \mathbf{Z}^d , that is, $E = \mathbf{Z}^d \cup \{\Delta\}$. Let $\beta_l: \mathbf{Z}^d \rightarrow [0, \infty)$, $l \in \mathbf{Z}^d$, $\sum_l \beta_l(k) < \infty$ for each $k \in \mathbf{Z}^d$, and for f vanishing off a finite subset of \mathbf{Z}^d , set

$$(4.1) \quad Af(x) = \begin{cases} \sum_l \beta_l(x)(f(x + l) - f(x)), & x \in \mathbf{Z}^d, \\ 0, & x = \Delta. \end{cases}$$

Let Y_l , $l \in \mathbf{Z}^d$, be independent Poisson processes, let $X(0)$ be nonrandom, and suppose X satisfies

$$(4.2) \quad X(t) = X(0) + \sum_l l Y_l \left(\int_0^t \beta_l(X(s)) ds \right), \quad t < \tau_\infty,$$

and

$$(4.3) \quad X(t) = \Delta, \quad t \geq \tau_\infty,$$

where $\tau_\infty = \inf \{t: X(t-) = \Delta\}$.

4.1 Theorem (a) Given $X(0)$, the solution of (4.2) and (4.3) is unique.

(b) X is a solution of the local-martingale problem for A . (Cf. Chapter 4, Section 7. Note, we have not assumed Af is bounded for each $f \in \mathcal{D}(A)$. If this is true, then X is a solution of the martingale problem for A .)

(c) If \tilde{X} is a solution of the local-martingale problem for A with sample paths in $D_E[0, \infty)$ satisfying $\tilde{X}(t) = \Delta$ for $t \geq \tau_\infty$ (τ_∞ as above), then there is a version X of \tilde{X} satisfying (4.2) and (4.3).

Proof. (a) Let $X_0(t) \equiv X(0)$ and set

$$(4.4) \quad X_k(t) = X(0) + \sum_l l Y_l \left(\int_0^t \beta_l(X_{k-1}(s)) ds \right).$$

Then if τ_k is the k th jump time of X_k , $X_k(t) = X_{k-1}(t)$ for $t < \tau_k$. Therefore

$$(4.5) \quad X(t) = \lim_{k \rightarrow \infty} X_k(t), \quad t < \lim_{k \rightarrow \infty} \tau_k,$$

exists and X satisfies (4.2). We leave the proof of uniqueness and the fact that $\lim_{k \rightarrow \infty} \tau_k = \tau_\infty$ to the reader.

(b) Let $\alpha(x) = 1 + \sum_l \beta_l(x)$ and

$$(4.6) \quad \int_0^{\eta(t)} \alpha(X(s)) ds = t$$

(cf. Proposition 2.10). Then $X^0(t) \equiv X(\eta(t))$ is a solution of

$$(4.7) \quad X^0(t) = X(0) + \sum_l l Y_l \left(\int_0^t \beta_l^0(X^0(s)) ds \right),$$

where $\beta_l^0 \equiv \beta_l/\alpha$. Note $\sum_l \beta_l^0 < 1$. If X^0 is a solution of the martingale problem for A^0 (defined as in (4.1) using the β_l^0), then by inverting the time change, we have that X is a solution of the local-martingale problem for A . For $z \in (\mathbb{Z}_+)^{\mathbb{Z}^d}$, let

$$(4.8) \quad \beta_k^1(z) = \begin{cases} \beta_k^0 \left(X(0) + \sum_l l z_l \right), & \sum_l |l| z_l < \infty, \\ 0, & \sum_l |l| z_l = \infty, \end{cases}$$

and set

$$(4.9) \quad Z_l(t) = Y_l \left(\int_0^t \beta_l^0(X^0(s)) ds \right) = Y_l \left(\int_0^t \beta_l^1(Z(s)) ds \right).$$

Since Z is the unique solution of (4.9), it is nonanticipating by Theorem 2.2. Consequently, by Theorem 2.8(a), Z is a solution of the martingale problem for

$$B = \left\{ \left(\prod_{l \in I} f_l, \sum_k \beta_k^l (f_k(\cdot + e_k) - f_k) \prod_{l \neq k} f_l \right) : I \subset \mathbb{Z}^d, I \text{ finite}, f_l \in B(\mathbb{Z}_+) \right\}.$$

The bp-closure of B contains $(f, \sum_{l \in I} \beta_l^l (f(\cdot + e_l) - f))$ where f is any bounded function depending only on the coordinates with indices in I (I finite). Consequently,

$$(4.10) \quad f \left(X(0) + \sum_{l \in I} l Z_l(t) \right) - \int_0^t \sum_{k \in I} \beta_k^0(X(s)) \left(f \left(X(0) + \sum_{l \in I} l Z_l(s) + k \right) - f \left(X(0) + \sum_{l \in I} l Z_l(s) \right) \right) ds$$

is a martingale for any finite I and any $f \in \mathcal{D}(A^0)$. Letting I increase to all of \mathbb{Z}^d , we see that X^0 is a solution of the martingale problem for A^0 .

(c) As before let

$$(4.11) \quad \int_0^{\tilde{\eta}(t)} \alpha(\tilde{X}(s)) ds = t.$$

Then $\tilde{X}^0(t) \equiv \tilde{X}(\tilde{\eta}(t))$ is a solution of the martingale problem for A^0 . But A^0 is bounded so the solution is unique for each $\tilde{X}(0)$. Consequently, if $X(0) = \tilde{X}(0)$, then by part (b), X^0 must be a version of \tilde{X}^0 and X must be a version of \tilde{X} . \square

5. DIFFUSION PROCESSES

Let $E = \mathbb{R}^d \cup \{\Delta\}$ be the one-point compactification of \mathbb{R}^d . For $k = 1, 2, \dots$, let $\beta_k: \mathbb{R}^d \rightarrow [0, \infty)$ be measurable, $\alpha_k \in \mathbb{R}^d$, and suppose that for each compact $K \subset \mathbb{R}^d$, $\sup_{x \in K} \sum_{k=1}^{\infty} |\alpha_k|^2 \beta_k(x) < \infty$. Thinking of the α_k as column vectors, define

$$(5.1) \quad G(x) = ((G_{ij}(x))) = \sum_{k=1}^{\infty} \alpha_k \alpha_k^T \beta_k(x).$$

Let $F: \mathbb{R}^d \rightarrow \mathbb{R}^d$ be measurable and bounded on compact sets. For $f \in C_c^{\infty}(\mathbb{R}^d)$, extend f to E by setting $f(\Delta) = 0$, and define

$$(5.2) \quad Af(x) = \begin{cases} \frac{1}{2} \sum_{i,j} G_{ij}(x) \partial_i \partial_j f(x) + \sum_i F_i(x) \partial_i f(x), & x \neq \Delta, \\ 0, & x = \Delta. \end{cases}$$

Let W_i , $i = 1, 2, \dots$, be independent standard Brownian motions, let $X(0)$ be nonrandom, and suppose X satisfies

$$(5.3) \quad X(t) = X(0) + \sum_{i=1}^{\infty} \alpha_i W_i \left(\int_0^t \beta_i(X(s)) ds \right) + \int_0^t F(X(s)) ds, \quad t < \tau_{\infty},$$

and

$$(5.4) \quad X(t) = \Delta, \quad t \geq \tau_{\infty},$$

where $\tau_{\infty} = \inf \{t: X(t-) = \Delta\}$. The solution of (5.3) and (5.4) is not in general unique, so we again employ the notion of a nonanticipating solution. In this context X is nonanticipating if for each $t \geq 0$, $W'_i \equiv W_i(\tau_i(t) + \cdot) - W_i(\tau_i(t))$, $i = 1, 2, \dots$, are independent standard Brownian motions that are independent of \mathcal{F}_t^X .

5.1 Theorem If X is a nonanticipating solution of (5.3) and (5.4), then X is a solution of the martingale problem for A .

5.2 Remark (a) Note that uniqueness for the martingale problem for A implies uniqueness of nonanticipating solutions of (5.3) and (5.4).

(b) A converse for Theorem 5.1 can be obtained from Theorem 5.3 below and Theorem 3.3 of Chapter 5. \square

Proof. The proof is essentially the same as for Theorem 4.1(b). \square

To simplify the statement of the next result, we assume $\tau_{\infty} = \infty$ in (5.3) and (5.4).

5.3 Theorem (a) If X is a nonanticipating solution of (5.3) for all $t < \infty$ (i.e., $\tau_{\infty} = \infty$), then there is a version of X satisfying the stochastic integral equation

$$(5.5) \quad Y(t) = Y(0) + \sum_{i=1}^{\infty} \alpha_i \int_0^t \sqrt{\beta_i(Y(s))} dB_i(s) + \int_0^t F(Y(s)) ds.$$

(b) If Y is a solution of (5.5) for all $t < \infty$, then there is a version of Y that is a nonanticipating solution of (5.3).

Proof. (a) Since X is a solution of the martingale problem for A , (a) follows from Theorem 3.3 of Chapter 5.

(b) Let \tilde{W}_i , $i = 1, 2, \dots$, be independent standard Brownian motions, independent of the B_i and Y . (It may be necessary to enlarge the sample space to obtain the \tilde{W}_i . See the proof of Theorem 3.3 in Chapter 5.)

Let

$$(5.6) \quad \tau_i(t) = \int_0^t \beta_i(Y(s)) \, ds,$$

and let

$$(5.7) \quad \gamma_i(u) = \inf \left\{ t: \int_0^t \beta_i(Y(s)) \, ds > u \right\}, \quad u \leq \tau_i(\infty).$$

Define

$$(5.8) \quad W_i(u) = \begin{cases} \int_0^{\gamma_i(u)} \sqrt{\beta_i(Y(s))} \, dB_i(s), & u \leq \tau_i(\infty), \\ \tilde{W}_i(u - \tau_i(\infty)) + W_i(\tau_i(\infty)), & \tau_i(\infty) < u < \infty. \end{cases}$$

Since $\gamma_i(u)$ is a stopping time, W_i is a martingale by the optional sampling theorem, as is $W_i^2(u) - u$. Consequently, W_i is a standard Brownian motion (Theorem 2.11 of Chapter 5). The independence of the W_i and the stopping properties of the τ_i follow by much the same argument as in the proof of the independence of the Y_k in Theorem 2.8(b). Finally, since $\int_0^t \sqrt{\beta_i(Y(s))} \, dB_i(s)$ is constant on any interval on which $\beta_i(Y(s))$ is zero, it follows that Y is a solution of (5.3). \square

The representations in Section 4 and in the present section combine to give a natural approach to diffusion approximations.

5.4 Theorem. Let $\beta_i^{(n)}: \mathbb{R}^d \rightarrow [0, \infty)$, $\alpha_i \in \mathbb{R}^d$, $i = 1, 2, \dots$, satisfy

$$(5.9) \quad \sup_n \sup_{x \in K} \sum_i (1 \vee |\alpha_i|^2) \beta_i^{(n)}(x) < \infty$$

for each compact $K \subset \mathbb{R}^d$, and let $\lambda_n > 0$ satisfy $\lim_{n \rightarrow \infty} \lambda_n = \infty$. Let Y_i , $i = 1, 2, \dots$, be independent unit Poisson processes and suppose X_n satisfies

$$(5.10) \quad X_n(t) = X_n(0) + \sum_i \lambda_n^{-1/2} \alpha_i Y_i \left(\lambda_n \int_0^t \beta_i^{(n)}(X_n(s)) \, ds \right).$$

Define $W_i^{(n)}(u) = \lambda_n^{-1/2} (Y_i(\lambda_n u) - \lambda_n u)$ and

$$(5.11) \quad F_n(x) = \lambda_n^{1/2} \sum_i \alpha_i \beta_i^{(n)}(x).$$

Let $\beta_i: \mathbb{R}^d \rightarrow [0, \infty)$, $i = 1, 2, \dots$, let $F: \mathbb{R}^d \rightarrow \mathbb{R}^d$ be continuous, and suppose for each compact $K \subset \mathbb{R}^d$ that

$$(5.12) \quad \lim_{n \rightarrow \infty} \sup_{x \in K} |\beta_i^{(n)}(x) - \beta_i(x)| = 0, \quad i = 1, 2, \dots,$$

$$(5.13) \quad \lim_{n \rightarrow \infty} \sup_{x \in K} |F_n(x) - F(x)| = 0,$$

and

$$(5.14) \quad \lim_{m \rightarrow \infty} \overline{\lim}_{n \rightarrow \infty} \sup_{x \in K} \sum_{l \geq m} |\alpha_l|^2 \beta_l^{(n)}(x) = 0.$$

Suppose that (5.3) and (5.4) have a unique nonanticipating solution and that $X_n(0) \rightarrow X(0)$. Let $\tau_a^n = \inf \{t : |X_n(t)| \geq a \text{ or } |X_n(t-)| \geq a\}$ and $\tau_a = \inf \{t : |X(t)| \geq a\}$. Then for all but countably many $a \geq 0$,

$$(5.15) \quad X_n(\cdot \wedge \tau_a^n) \Rightarrow X(\cdot \wedge \tau_a).$$

If $\lim_{a \rightarrow \infty} \tau_a = \infty$, then $X_n \Rightarrow X$.

5.5 Remark More-general results of this type can be obtained as corollaries to Theorem 3.4. \square

Proof. Note that

$$(5.16) \quad X_n(t) = X_n(0) + \sum_i \alpha_i W_i^{(n)} \left(\int_0^t \beta_i^{(n)}(X_n(s)) ds \right) + \int_0^t F_n(X_n(s)) ds.$$

It follows from (5.12), (5.13), (5.14), the relative compactness of $\{W_i^{(n)}\}$, and (5.16), that $\{X_n(\cdot \wedge \tau_a^n)\}$ is relatively compact (cf. Proposition 3.1). Furthermore, if for $a_0 > 0$ and some subsequence $\{n_i\}$, $X_{n_i}(\cdot \wedge \tau_{a_0}^{n_i}) \Rightarrow Y_{a_0}$, then setting $\eta_a = \inf \{t : |Y_{a_0}(t)| \geq a \text{ or } |Y_{a_0}(t-)| \geq a\}$, $(X_{n_i}(\cdot \wedge \tau_{a_0}^{n_i}), \tau_{a_0}^{n_i}) \Rightarrow (Y_{a_0}(\cdot \wedge \eta_a), \eta_a)$ in $D_R[0, \infty) \times [0, \infty]$ for all $a < a_0$ such that

$$(5.17) \quad P \left\{ \lim_{b \rightarrow a} \eta_b = \eta_a \right\} = 1.$$

Note that the monotonicity of η_a implies (5.17) holds for all but countably many a .

Since a_0 is arbitrary, we can select the subsequence so that $\{(X_{n_i}(\cdot \wedge \tau_{a_0}^{n_i}), \tau_{a_0}^{n_i})\}$ converges in distribution for all but countably many a , and the limit has the form $(Y(\cdot \wedge \eta_a), \eta_a)$ for a fixed process Y with sample paths in $D_E[0, \infty)$ (η_a as before). (We may assume that $Y(t) = \Delta$ implies $Y(s) = \Delta$ for all $s > t$.) By the continuous mapping theorem (Corollary 1.9 of Chapter 3), Y satisfies

$$(5.18) \quad Y(t \wedge \eta_a) = Y(0) + \sum_i \alpha_i W_i \left(\int_0^{t \wedge \eta_a} \beta_i(Y(s)) ds \right) + \int_0^{t \wedge \eta_a} F(Y(s)) ds.$$

Here (5.14) allows the interchange of summation and limits. It follows as in the proof of Theorem 3.4 that Y is a nonanticipating solution of (5.3) and (5.4) and hence Y has the same distribution as X . The uniqueness of the possible limit point gives (5.15) for all a such that $\lim_{b \rightarrow a} \eta_b = \eta_a$ a.s. The final statement of the theorem is left to the reader. \square

Equations of the form of (5.10) ordinarily arise after renormalization of space and time. For example, suppose

$$(5.19) \quad U_n(t) = U_n(0) + \sum_l l Y_l \left(\int_0^t \beta_l^{(n)}(U_n(s)) ds \right),$$

and set $X_n(t) = n^{-1/2} U_n(nt)$. Then X_n satisfies

$$(5.20) \quad X_n(t) = X_n(0) + \sum_l l W_l^{(n)} \left(\int_0^t \beta_l^{(n)}(n^{1/2} X_n(s)) ds \right) + \int_0^t F_n(X_n(s)) ds,$$

where $W_l^{(n)}(u) = n^{-1/2}(Y_l(nu) - nu)$ and

$$(5.21) \quad F_n(x) = n^{1/2} \sum_l l \beta_l^{(n)}(n^{1/2} x).$$

6. PROBLEMS

1. Let W be standard Brownian motion.

- (a) Show that for $0 < \alpha < 1$,

$$(6.1) \quad \int_0^t \frac{1}{|W(s)|^\alpha} ds < \infty \quad \text{a.s.,} \quad t \geq 0,$$

and for $\alpha \geq 1$,

$$(6.2) \quad \int_0^t \frac{1}{|W(s)|^\alpha} ds = \infty \quad \text{a.s.,} \quad t > 0.$$

- (b) Show that for $\alpha \geq 1$ the solution of

$$(6.3) \quad X(t) = X(0) + W \left(\int_0^t |X(s)|^\alpha ds \right)$$

is unique, but it is not unique if $0 < \alpha < 1$.

- (c) Let $0 < \alpha < 1$ and $\gamma_0 = \sup \{t < 100: W(t) = 0\}$. Let $\tau(t)$ satisfy

$$(6.4) \quad \begin{aligned} \int_0^{\tau(t)} \frac{1}{|W(s)|^\alpha} ds &= t, & t < \int_0^{\gamma_0} \frac{1}{|W(s)|^\alpha} ds, \\ \tau(t) &= \gamma_0, & t \geq \int_0^{\gamma_0} \frac{1}{|W(s)|^\alpha} ds. \end{aligned}$$

Show that $X(t) = W(\tau(t))$ satisfies (6.3), but that it is not a solution of the martingale problem for $A = \{(f, \frac{1}{2}|x|^2 f''): f \in C_c^\infty(\mathbb{R})\}$.

2. Let Y_1 and Y_2 be independent standard Brownian motions. Let β_1 and β_2 be nonnegative, measurable functions on \mathbb{R}^2 satisfying $\beta_i(x, y) \leq$

$K(1 + x^2 + y^2)$. Show that the random time change problem

$$(6.5) \quad Z_t(t) = Y_t \left(\int_0^t \beta_i(Z(s)) \, ds \right)$$

is equivalent to the martingale problem for A given by

$$(6.6) \quad A = \{(f, \beta_1 f_{xx} + \beta_2 f_{yy}); f \in C_c^2(\mathbb{R}^2)\},$$

that is, any nonanticipating solution of (6.5) is a solution of the martingale problem for A , and any solution of the martingale problem for A has a version that is a weak, nonanticipating solution of (6.5).

3. State and prove a result analogous to that in Problem 2 in which Y_1 and Y_2 are Poisson processes.

4. Let Y be Brownian motion,

$$(6.7) \quad \beta_1(x) = \begin{cases} 1, & |x| \leq 1, \\ 0, & |x| > 1, \end{cases}$$

and

$$(6.8) \quad \beta_2(x) = \begin{cases} 1, & |x| < 1, \\ 0, & |x| \geq 1. \end{cases}$$

Show that

$$(6.9) \quad Z(t) = Y \left(\int_0^t \beta_1(Z(s)) \, ds \right)$$

has no solution but that

$$(6.10) \quad Z(t) = Y \left(\int_0^t \beta_2(Z(s)) \, ds \right)$$

does. In the second case, what is the (strong) generator corresponding to Z ?

5. Prove Proposition 3.2.

6. For $Y_1(t) = [t]$ and $Y_2(t) = t$, let $(Z_1^{(n)}, Z_2^{(n)})$ satisfy

$$(6.11) \quad \begin{aligned} Z_1^{(n)}(t) &= Y_1 \left(\int_0^t (1 - n^{-1}) \sqrt{(1 - Z_2^{(n)}(s)) \vee 0} \, ds \right), \\ Z_2^{(n)}(t) &= Y_2 \left(\int_0^t (Z_1^{(n)}(s) + \sqrt{(1 - Z_2^{(n)}(s)) \vee 0}) \, ds \right), \end{aligned}$$

and let (Z_1, Z_2) satisfy

$$(6.12) \quad \begin{aligned} Z_1(t) &= Y_1 \left(\int_0^t \sqrt{(1 - Z_2(s)) \vee 0} \, ds \right), \\ Z_2(t) &= Y_2 \left(\int_0^t (Z_1(s) + \sqrt{(1 - Z_2(s)) \vee 0}) \, ds \right). \end{aligned}$$

that $\lim_{n \rightarrow \infty} (Z_1^{(n)}, Z_2^{(n)}) \neq (Z_1, Z_2)$.

7. Let $\tau(t)$ be nonnegative, nondecreasing, and right continuous. Let $y \in D_E[0, \infty)$ and $z = y(\tau(\cdot))$. Define

$$(6.13) \quad \eta(\delta, T) = \sup_{t \leq T} (\tau(t + \delta) - \tau(t)) + \sup_{t \leq T} (\tau(t) - \tau(t -)),$$

and $\gamma(t) = \inf \{u : \tau(u) \geq t\}$. Show that if $0 \leq t_1 < t_2$ and $t_2 - t_1 > \eta(\delta, T)$, then $\gamma(t_2) - \gamma(t_1) > \delta$, and that

$$(6.14) \quad w'(z, \delta, T) \leq w'(y, \eta(\delta, T), \tau(T)).$$

8. Suppose in (4.2) that $\sum |l| \beta_l(x) \leq A + B|x|$. Show that $\tau_\infty = \infty$.

9. Let W and Y be independent, W a standard Brownian motion and Y a unit Poisson process. Show that

$$(6.15) \quad Z_n(t) = W \left(\int_0^t (2 + (-1)^{Y(ns)}) \, ds \right)$$

and

$$(6.16) \quad \tilde{Z}_n(t) = \int_0^t \sqrt{2 + (-1)^{Y(ns)}} \, dW(s)$$

have the same distribution, that $\{Z_n\}$ converges a.s., but $\{\tilde{Z}_n\}$ does not converge a.s.

10. Let $E = \{(x, y) : x, y \geq 0, x + y \leq 1\}$. For $f \in C^\infty(E)$, define $Af = x(1-x)f_{xx} - 2xyf_{xy} + y(1-y)f_{yy}$. Show that if X is a solution of the martingale problem for A , then X satisfies (5.3) with $\alpha_i = 0$, $i \geq 4$.

11. Let f and g be locally absolutely continuous on \mathbb{R} .

- (a) Show that if h is bounded and Borel measurable, then

$$(6.17) \quad \int_a^b h(g(z))g'(z) \, dz = \int_{g(a)}^{g(b)} h(u) \, du, \quad a, b \in \mathbb{R},$$

with the usual convention that $\int_a^b f(z) \, dz = -\int_b^a f(z) \, dz$ if $b < a$.

Hint: Check (6.9) first for continuous h by showing both sides are locally absolutely continuous as functions of b and differentiating. Then apply a monotone class argument (see Appendix 4).

- (b) Show that if $A \in \mathcal{B}(\mathbb{R})$ has Lebesgue measure zero, then

$$(6.18) \quad \int_a^b \chi_{\{g(z) \in A\}} g'(z) dz = 0.$$

In particular, for each α , $m(\{g'(z) \neq 0\} \cap \{g(z) = \alpha\}) = 0$.

- (c) Show that if g is nondecreasing, then $f \circ g$ is locally absolutely continuous.

- (d) Define

$$(6.19) \quad h(z) = \begin{cases} f'(g(z))g'(z) & \text{on } \{z: f'(g(z)) \text{ and } g'(z) \text{ exist}\}, \\ 0 & \text{otherwise.} \end{cases}$$

(Note that $m(\mathbb{R} - \{z: f'(g(z)) \text{ and } g'(z) \text{ exist}\} \cup \{z: g'(z) = 0\}) = 0$.) Show that $f \circ g$ is locally absolutely continuous if and only if h is locally L^1 , and that under those conditions

$$(6.20) \quad \frac{d}{dz} f(g(z)) = h(z) \quad \text{a.e.}$$

- (e) Let $f(t) = \sqrt{|t|}$ and $g(t) = t^2 \cos^2(1/t)$. Show that f and g are locally absolutely continuous, but $f \circ g$ is not.

Hint: Show that $f \circ g$ does not have bounded variation.

12. Let β be a nonnegative Borel measurable function on $[0, \infty)$ that is locally L^1 . Define $\gamma(t) = \inf\{u: \int_0^u \beta(s) ds > t\}$.

- (a) Show that γ is right continuous.

- (b) Show that

$$(6.21) \quad \int_a^b \beta(s) ds = \int_0^\infty \chi_{\{a \leq \gamma(t) < b\}} dt$$

for all $0 \leq a < b$.

- (c) Show that if g is Borel measurable and βg is locally L^1 , then

$$(6.22) \quad \int_0^{\gamma(t)} \beta(s)g(s) ds = \int_0^t g(\gamma(u)) du.$$

7. NOTES

Volkonski (1958) introduced the one-parameter random time change for Markov processes. See also Lamperti (1967b). Helland (1978) gave results similar to Theorem 1.5 with applications to branching processes (see Chapter 9, Section 1).

The multiparameter time changes were introduced by Helms (1974) and developed in Kurtz (1980a). Holley and Stroock (1976) use a slightly different approach.

Applications of multiparameter time changes to convergence theorems are given in Kurtz (1978a, 1981c, 1982). See Chapters 9 and 11.

Any diffusion with a uniformly elliptic generator with bounded coefficients can be obtained as a nonanticipating solution of an equation of the form of (5.3) with only finitely many nonzero α_i . See Kurtz (1980a).

7

INvariance PRINCIPLES AND
DIFFUSION APPROXIMATIONS

Let ξ_1, ξ_2, \dots be independent, identically distributed random variables with mean zero and variance one. Define

$$(0.1) \quad X_n(t) = n^{-1/2} \sum_{k=1}^{\lfloor nt \rfloor} \xi_k.$$

A simple application of Theorem 2.6 of Chapter 4 and Theorem 6.5 of Chapter 1 gives Donsker's (1951) invariance principle, Theorem 1.2(c) of Chapter 5, that is, that $X_n \Rightarrow W$ where W is standard Brownian motion. One noteworthy property of X_n is that it is a martingale. In Section 1 we show that the invariance principle can be extended to very general sequences of martingales.

Another direction in which the invariance principle has been extended is to processes satisfying mixing conditions, that is, some form of asymptotic independence. A large number of such conditions have been introduced. We consider some of these in Section 2 and give examples of related invariance principles in Section 3.

Section 4 is devoted to an extension of the results of Section 1 allowing the limiting process to be an arbitrary diffusion process.

Section 5 contains recent refinements of the invariance principle due to Komlós, Major, and Tusnády (1975, 1976) who showed how to construct X_n and W on the same sample space in such a way that

$$(0.2) \quad \sup_{t \leq T} |X_n(t) - W(t)| = O\left(\frac{\log n}{\sqrt{n}}\right).$$

1. THE MARTINGALE CENTRAL LIMIT THEOREM

In this section we give the extension of Donsker's invariance principle to sequences of martingales in \mathbb{R}^d . The convergence results are based on the following martingale characterization of processes with independent increments.

1.1 Theorem Let $C = ((c_{ij}))$ be a continuous, symmetric, $d \times d$ matrix-valued function, defined on $[0, \infty)$, satisfying $C(0) = 0$ and

$$(1.1) \quad \sum (c_{ij}(t) - c_{ij}(s)) \xi_i \xi_j \geq 0, \quad \xi \in \mathbb{R}^d, \quad t > s \geq 0.$$

Then there exists a unique (in distribution) process X with sample paths in $C_{\mathbb{R}^d}[0, \infty)$ such that X_i , $i = 1, 2, \dots, d$, and $X_i X_j - c_{ij}$, $i, j = 1, 2, \dots, d$, are (local) martingales with respect to $\{\mathcal{F}_t^X\}$. The process X has independent Gaussian increments.

Proof. As in the proof of Theorem 2.12 of Chapter 5, if X is such a process, then for $\theta \in \mathbb{R}^d$

$$(1.2) \quad f(t, X) = \exp \{i\theta \cdot X(t) + \frac{1}{2}\theta \cdot C(t)\theta\}$$

is a martingale, and hence

$$(1.3) \quad E[\exp \{i\theta \cdot (X(t) - X(s))\} | \mathcal{F}_s] = \exp \{-\frac{1}{2}\theta \cdot (C(t) - C(s))\},$$

which implies X has independent Gaussian increments and determines the finite-dimensional distributions of X .

To obtain such an X set

$$(1.4) \quad \gamma(t) = \sum_{i=1}^d c_{ii}(t).$$

Note that (1.1) implies

$$(1.5) \quad \begin{aligned} |c_{ij}(t) - c_{ij}(s)| &\leq c_{ii}(t) - c_{ii}(s) + c_{jj}(t) - c_{jj}(s) \\ &\leq \gamma(t) - \gamma(s) \end{aligned}$$

(take $\xi_i = 1$, $\xi_j = \pm 1$, and $\xi_i = 0$ otherwise), and hence c_{ij} is of bounded variation and c_{ij} can be written as

$$(1.6) \quad c_{ij}(t) = \int_0^t d_{ij}(s) d\gamma(s),$$

where $D(s) = ((d_{ij}(s)))$ is nonnegative definite. Let $D^{1/2}(s)$ denote the symmetric nonnegative-definite square root of $D(s)$, let W be d -dimensional standard Brownian motion, and set

$$(1.7) \quad M(t) = W(\gamma(t)).$$

Then

$$(1.8) \quad X(t) = \int_0^t D \, dM$$

is the desired process. \square

1.2 Theorem Let C be as in Theorem 1.1. Suppose that X is a measurable process and that, for each $\theta \in \mathbb{R}^d$ and $f \in C_c^\infty(\mathbb{R})$,

$$(1.9) \quad f(\theta \cdot X(t)) - \int_0^t \frac{1}{2} f''(\theta \cdot X(s)) \, dc_\theta(s)$$

is an $\{\mathcal{F}_t^X\}$ -martingale, where

$$(1.10) \quad c_\theta(t) = \theta \cdot C(t)\theta.$$

Then X has independent Gaussian increments with mean zero and

$$(1.11) \quad E[X(t)X(t)^T] = C(t).$$

1.3 Remark Note that it is crucial that (1.9) be a martingale with respect to $\{\mathcal{F}_t^X\}$ and not just with respect to $\{\mathcal{F}_t^{\theta \cdot X}\} \equiv \sigma\{\theta \cdot X(s): s \leq t\}$. See Problem 2. \square

Proof. The collection of f for which (1.9) is an $\{\mathcal{F}_t^X\}$ -martingale is closed under bp-convergence. Consequently,

$$(1.12) \quad \exp\{i\theta \cdot X(t)\} + \int_0^t \frac{1}{2} \exp\{i\theta \cdot X(s)\} \, dc_\theta(s)$$

is an $\{\mathcal{F}_t^X\}$ -martingale, and hence, by Itô's formula, Theorem 2.9 of Chapter 5,

$$(1.13) \quad \exp\{i\theta \cdot X(t) + \frac{1}{2}c_\theta(t)\}$$

is an $\{\mathcal{F}_t^X\}$ -martingale. The theorem follows as in the proof of Theorem 1.1. \square

1.4 Theorem For $n = 1, 2, \dots$, let $\{\mathcal{F}_t^n\}$ be a filtration and let M_n be an $\{\mathcal{F}_t^n\}$ -local martingale with sample paths in $D_{\mathbb{R}^d}[0, \infty)$ and $M_n(0) = 0$. Let $A_n = ((A_n^{ij}))$ be symmetric $d \times d$ matrix-valued processes such that A_n^{ij} has sample paths in $D_{\mathbb{R}}[0, \infty)$ and $A_n(t) - A_n(s)$ is nonnegative definite for $t > s \geq 0$. Assume one of the following conditions holds:

(a) For each $T > 0$,

$$(1.14) \quad \lim_{n \rightarrow \infty} E \left[\sup_{t \leq T} |M_n(t) - M_n(t-)| \right] = 0$$

and

$$(1.15) \quad A_n^U = [M_n^I, M_n^J].$$

(b) For each $T > 0$ and $i, j = 1, 2, \dots, d$,

$$(1.16) \quad \lim_{n \rightarrow \infty} E \left[\sup_{t \leq T} |A_n^U(t) - A_n^U(t-)| \right] = 0,$$

$$(1.17) \quad \lim_{n \rightarrow \infty} E \left[\sup_{t \leq T} |M_n(t) - M_n(t-)|^2 \right] = 0,$$

and for $i, j = 1, 2, \dots, d$,

$$(1.18) \quad M_n^I(t)M_n^J(t) - A_n^U(t)$$

is an $\{\mathcal{F}_t^n\}$ -local martingale.

Suppose that C satisfies the conditions of Theorem 1.1 and that, for each $t \geq 0$ and $i, j = 1, 2, \dots, d$,

$$(1.19) \quad A_n^U(t) \rightarrow c_{ij}(t)$$

in probability. Then $M_n \Rightarrow X$, where X is the process with independent Gaussian increments given by Theorem 1.1.

1.5 Remark In the discrete-time case, let $\{\zeta_k^n: k = 1, 2, \dots\}$ be a collection of \mathbb{R}^d -valued random variables and define

$$(1.20) \quad M_n(t) = \sum_{k=1}^{\lfloor \alpha_n t \rfloor} \zeta_k^n$$

for some $\alpha_n \rightarrow \infty$. In condition (a)

$$(1.21) \quad A_n(t) = \sum_{k=1}^{\lfloor \alpha_n t \rfloor} \zeta_k^n \zeta_k^{nT}$$

(considering the ζ_k^n as column vectors), and for condition (b) one can take

$$(1.22) \quad A_n(t) = \sum_{k=1}^{\lfloor \alpha_n t \rfloor} E[\zeta_k^n \zeta_k^{nT} | \mathcal{F}_{k-1}^n]$$

where $\mathcal{F}_k^n = \sigma(\zeta_l^n: l \leq k)$. Of course M_n is a martingale if $E[\zeta_k^n | \mathcal{F}_{k-1}^n] = 0$. \square

Proof. Without loss of generality, we may assume the M_n are martingales. If not, there exist stopping times τ_n with $P[\tau_n < n] \leq n^{-1}$ such that $M_n(\cdot \wedge \tau_n)$ is a martingale and $A_n(\cdot \wedge \tau_n)$ satisfies the conditions of the theorem with M_n

replaced by $M_n(\cdot \wedge \tau_n)$. Similarly, under condition (b) we may assume the processes in (1.18) are martingales.

(a) Assume condition (a). Let

$$(1.23) \quad \eta_n = \inf \{t: A_n^{ii}(t) > c_{ii}(t) + 1 \text{ for some } i \in \{1, 2, \dots, d\}\}.$$

Since (1.19) implies $\eta_n \rightarrow \infty$ in probability, the convergence of M_n is equivalent to the convergence of $\tilde{M}_n = M_n(\cdot \wedge \eta_n)$.

Fix $\theta \in \mathbb{R}^d$ and define

$$(1.24) \quad Y_n(t) = \theta \cdot \tilde{M}_n(t),$$

$$(1.25) \quad A_n^\theta(t) = \sum_{i,j} A_n^{ij}(t) \theta_i \theta_j,$$

and

$$(1.26) \quad c_\theta(t) = \sum_{i,j} c_{ij}(t) \theta_i \theta_j.$$

Let $f \in C_c^\infty(\mathbb{R})$, $0 \leq t = t_0 < t_1 < \dots < t_m = t + s$, and $\xi_k = Y_n(t_{k+1}) - Y_n(t_k)$. Then

$$(1.27) \quad E[f(Y_n(t+s)) - f(Y_n(t))] \mid \mathcal{F}_t^n \\ = E \left[\sum_{k=0}^{m-1} (f(Y_n(t_{k+1})) - f(Y_n(t_k)) - f'(Y_n(t_k)) \xi_k) \mid \mathcal{F}_t^n \right].$$

Let

$$(1.28) \quad \gamma = \max \{k: t_k < \eta_n \wedge (t+s)\}$$

and

$$(1.29) \quad \zeta = \max \left\{ k: t_k < \eta_n \wedge (t+s), \sum_{i=0}^k \xi_i^2 \leq d \sum_{i=0}^k c_{ii}(t+s) \theta_i^2 + 2d |\theta|^2 \right\}.$$

Note that by the definition of η_n , $\gamma = \zeta$ for $\max (t_{k+1} - t_k)$ sufficiently small.

Then by (1.27),

$$(1.30) \quad E[f(Y_n(t+s)) - f(Y_n(t))] \mid \mathcal{F}_t^n \\ = E \left[\sum_{k=\zeta}^{\gamma} (f(Y_n(t_{k+1})) - f(Y_n(t_k)) - f'(Y_n(t_k)) \xi_k) \mid \mathcal{F}_t^n \right] \\ + E \left[\sum_{k=0}^{\zeta-1} (f(Y_n(t_{k+1})) - f(Y_n(t_k)) - f'(Y_n(t_k)) \xi_k - \frac{1}{2} f''(Y_n(t_k)) \xi_k^2) \mid \mathcal{F}_t^n \right] \\ + E \left[\sum_{k=0}^{\zeta-1} \frac{1}{2} f''(Y_n(t_k)) \xi_k^2 \mid \mathcal{F}_t^n \right].$$

Setting $\Delta Y_n(u) = Y_n(u) - Y_n(u-)$, and letting $\max (t_{k+1} - t_k) \rightarrow 0$,

$$\begin{aligned}
 (1.31) \quad & E[f(Y_n(t+s)) - f(Y_n(t)) \mid \mathcal{F}_t^n] \\
 &= E \left[f(Y_n((t+s) \wedge \eta_n)) - f(Y_n((t+s) \wedge \eta_n-)) \right. \\
 &\quad \left. - f'(Y_n((t+s) \wedge \eta_n-)) \Delta Y_n((t+s) \wedge \eta_n) \middle| \mathcal{F}_t^n \right] \\
 &+ E \left[\sum_{t < u < (t+s) \wedge \eta_n} (f(Y_n(u)) - f(Y_n(u-))) \right. \\
 &\quad \left. - f'(Y_n(u-)) \Delta Y_n(u) - \frac{1}{2} f''(Y_n(u-)) (\Delta Y_n(u))^2 \middle| \mathcal{F}_t^n \right] \\
 &+ E \left[\int_{(t, (t+s) \wedge \eta_n)} \frac{1}{2} f''(Y_n(u-)) dA_n^0(u) \middle| \mathcal{F}_t^n \right].
 \end{aligned}$$

Note that the second term on the right is bounded by

$$(1.32) \quad 6^{-1} \|f'''\| E \left[\sup_{t < u < (t+s)} |\Delta Y_n(u)| d \sum_i (c_{ii}(t+s) + 1) \theta_i^2 \middle| \mathcal{F}_t^n \right]$$

(by the definition of η_n), and hence

$$\begin{aligned}
 (1.33) \quad & |E[f(Y_n(t+s)) - f(Y_n(t)) \mid \mathcal{F}_t^n]| \\
 &\leq C_f E \left[\sup_{t < u \leq t+s} |\Delta Y_n(u)| \left(1 + d \sum_i (c_{ii}(t+s) + 1) \theta_i^2 \right) \right. \\
 &\quad \left. + A_n^0((t+s) \wedge \eta_n-) - A_n^0(t \wedge \eta_n-) \middle| \mathcal{F}_t^n \right],
 \end{aligned}$$

where C_f depends only on $\|f'\|$, $\|f''\|$, and $\|f'''\|$. In particular (1.33) can be extended to all f (including unbounded f) whose first three derivatives are bounded.

Let φ be convex with $\varphi(0) = 0$, $\lim_{x \rightarrow \infty} \varphi(x) = \infty$, and φ' , φ'' , and φ''' bounded. Then

$$(1.34) \quad P \left\{ \sup_{t \leq T} |Y_n(t)| > K \right\} \leq \frac{E[\varphi(Y_n(t))]}{\varphi(K)}$$

$$\leq C_\varphi E \left[\sup_{u \leq T} |\Delta Y_n(u)| \left(1 + d \sum_i (c_{ii}(T) + 1) \theta_i^2 \right) \right.$$

$$\left. + d \sum_i (c_{ii}(T) + 1) \theta_i^2 \right] / \varphi(K).$$

Furthermore, for $\delta > 0$, $0 \leq t \leq T$, and $0 \leq u \leq \delta$,

$$(1.35) \quad E[(f(Y_n(t+u)) - f(Y_n(t)))^2 | \mathcal{F}_t^n]$$

$$= E[f^2(Y_n(t+u)) - f^2(Y_n(t)) | \mathcal{F}_t^n]$$

$$- 2f(Y_n(t))E[f(Y_n(t+u)) - f(Y_n(t)) | \mathcal{F}_t^n]$$

$$\leq E[\gamma_n(\delta) | \mathcal{F}_t^n],$$

where

$$(1.36) \quad \gamma_n(\delta) = (C_{f^2} + 2\|f\|C_f) \left[\sup_{s \leq T+\delta} |\Delta Y_n(s)| \left(1 + d \sum_i (c_{ii}(T+\delta) + 1) \theta_i^2 \right) \right.$$

$$\left. + \sup_{t \leq T} (A_n^\theta((t+\delta) \wedge \eta_n-) - A_n^\theta(t \wedge \eta_n-)) \right].$$

By (1.14) and (1.19),

$$(1.37) \quad \lim_{\delta \rightarrow 0} \lim_{n \rightarrow \infty} E[\gamma_n(\delta)] = \lim_{\delta \rightarrow 0} (C_{f^2} + 2\|f\|C_f) \sup_{t \leq T} (c_\theta(t+\delta) - c_\theta(t))$$

$$= 0,$$

so for each $f \in C_c^\infty(\mathbb{R}^d)$, $\{f(Y_n)\}$ is relatively compact by Theorem 8.6 of Chapter 3. Consequently, since we have (1.34), the relative compactness of $\{Y_n\} = \{\theta \cdot \tilde{M}_n\}$ follows by Theorem 9.1 of Chapter 3. Since θ is arbitrary, $\{\tilde{M}_n\}$ must be relatively compact (cf. Problem 22 of Chapter 3).

The continuity of $c_\theta(t)$ and (1.19) imply

$$(1.38) \quad \int_{(t, (t+s) \wedge \eta_n)} \frac{1}{2} f''(y(u-)) dA_n^\theta(u) - \int_t^{t+s} \frac{1}{2} f''(y(u)) dc_\theta(u) \rightarrow 0$$

in probability uniformly for y in compact subsets of $D_R[0, \infty)$. Consequently the relative compactness of $\{Y_n\}$ implies

$$(1.39) \quad E \left[\left| \int_{(t, (t+s) \wedge \eta_n)} \frac{1}{2} f''(Y_n(u-)) dA_n^\theta(u) - \int_t^{t+s} \frac{1}{2} f''(Y_n(u)) dc_\theta(u) \right| \right] \rightarrow 0.$$

The first two terms on the right of (1.32) go to zero in L^1 and it follows easily (cf. the proof of Theorem 8.10 in Chapter 4 and Problem 7 in this chapter) that if X is a limit point of $\{\tilde{M}_n\}$, then (1.9) is an $\{\mathcal{F}_t^X\}$ -martingale. Since this uniquely characterizes X , the convergence of $\{\tilde{M}_n\}$ follows.

(b) The proof is similar to that of part (a). With η_n and \tilde{M}_n defined as in (1.23), we have

$$(1.40) \quad A_n^{ii}(t \wedge \eta_n) \leq c_{ii}(t) + 1 + \sup_{s \leq t} (A_n^{ii}(s) - A_n^{ii}(s-)),$$

and the third term on the right goes to zero in L^1 by (1.16).

Setting $\tilde{A}_n^{ij}(t) = A_n^{ij}(t \wedge \eta_n)$, note that

$$(1.41) \quad E[|\tilde{M}_n(t+h) - \tilde{M}_n(t)|^2 | \mathcal{F}_t^n] = E\left[\sum_{i=1}^d (\tilde{A}_n^{ii}(t+h) - \tilde{A}_n^{ii}(t)) \middle| \mathcal{F}_t^n\right]$$

and to apply Theorem 8.6 of Chapter 3, fix $T > 0$ and define

$$(1.42) \quad \gamma_n(\delta) = \sup_{t \leq T} \sum_{i=1}^d (\tilde{A}_n^{ii}(t+\delta) - \tilde{A}_n^{ii}(t)).$$

Since

$$(1.43) \quad \gamma_n(\delta) \leq \sum_{i=1}^d \left(c_{ii}(T+\delta) + 1 + \sup_{s \leq T+\delta} (A_n^{ii}(s) - A_n^{ii}(s-)) \right),$$

and since (1.16) implies the right side of (1.43) is convergent in L^1 , we conclude that

$$(1.44) \quad \lim_{\delta \rightarrow 0} \lim_{n \rightarrow \infty} E[\gamma_n(\delta)] = \lim_{\delta \rightarrow 0} \sup_{t \leq T} \sum_{i=1}^d (c_{ii}(t+\delta) - c_{ii}(t)) = 0.$$

Let X be any limit point of $\{\tilde{M}_n\}$. By (1.17), X is continuous. Since for each $T > 0$, $\sup_n E[|\tilde{M}_n(T)|^2] < \infty$, $\{\tilde{M}_n(T)\}$ is uniformly integrable, and hence X must be a martingale (see Problem 7). Since $X_i X_j - c_{ij}$ is the limit in distribution of a subsequence $\{\tilde{M}_{n_k}^i \tilde{M}_{n_k}^j - \tilde{A}_{n_k}^{ij}\}$, we can conclude that it is also a martingale if we show that $\{\tilde{M}_{n_k}^i(T) \tilde{M}_{n_k}^j(T) - \tilde{A}_{n_k}^{ij}(T)\}$ is uniformly integrable for each T . Since (1.40) and (1.16) imply that $\{\tilde{A}_{n_k}^{ij}(T)\}$ is uniformly integrable (recall $|\tilde{A}_n^{ij}(T)| \leq \frac{1}{2}(\tilde{A}_n^{ii}(T) + \tilde{A}_n^{jj}(T))$), it is enough to consider $\{\tilde{M}_{n_k}^i(T) \tilde{M}_{n_k}^j(T)\}$, and since $|\tilde{M}_{n_k}^i(T) \tilde{M}_{n_k}^j(T)| \leq \frac{1}{2}(\tilde{M}_{n_k}^i(T)^2 + \tilde{M}_{n_k}^j(T)^2)$, it is enough to consider $\{\tilde{M}_{n_k}^i(T)^2\}$. Since $\tilde{M}_{n_k}^i(T)^2 \Rightarrow X_i(T)^2$, $\{\tilde{M}_{n_k}^i(T)^2\}$ is uniformly integrable if (and only if) $E[\tilde{M}_{n_k}^i(T)^2] \rightarrow E[X_i(T)^2]$, that is, if

$$(1.45) \quad E[X_i(T)^2] = c_{ii}(T).$$

Let

$$(1.46) \quad \tau_n^\alpha = \inf \{t: \tilde{M}_n^i(t)^2 > \alpha\}$$

and

$$(1.47) \quad \tau^* = \inf \{t: X_i(t)^2 > \alpha\}.$$

Since

$$(1.48) \quad \tilde{M}_n^i(T \wedge \tau_n^*)^2 \leq 2 \left(\alpha + \sup_{s \leq T} |\tilde{M}_n^i(s) - \tilde{M}_n^i(s-)|^2 \right),$$

$\{\tilde{M}_n^i(T \wedge \tau_n^*)\}$ is uniformly integrable by (1.17). For all but countably many α and T , $(\tau_{n_k}^*, \tilde{M}_{n_k}^i(T \wedge \tau_{n_k}^*)) \Rightarrow (\tau^*, X_i^i(T \wedge \tau^*))$ and, excluding the countably many α and T ,

$$\begin{aligned} (1.49) \quad E[X_i(T \wedge \tau^*)^2] &= \lim_{k \rightarrow \infty} E[\tilde{M}_{n_k}^i(T \wedge \tau_{n_k}^*)^2] \\ &= \lim_{k \rightarrow \infty} E[\tilde{A}_{n_k}^{ii}(T \wedge \tau_{n_k}^*)] \\ &= E[c_{ii}(T \wedge \tau^*)]. \end{aligned}$$

Letting $\alpha \rightarrow \infty$ we have (1.45), and it follows that $X_i X_j - c_{ij}$ are martingales for $i, j = 1, \dots, d$, and that X is the unique process characterized in Theorem 1.1. Therefore $M_n \Rightarrow X$. \square

2. MEASURES OF MIXING

Measures of mixing are measures of the degree of independence of two σ -algebras. Let (Ω, \mathcal{F}, P) be a probability space, and let \mathcal{G} and \mathcal{H} be sub- σ -algebras of \mathcal{F} . Two kinds of measures of mixing are commonly used. The measure of *uniform mixing* is given by

$$\begin{aligned} (2.1) \quad \varphi(\mathcal{G} | \mathcal{H}) &\equiv \sup_{A \in \mathcal{G}} \sup_{\substack{B \in \mathcal{H} \\ P(B) > 0}} |P(A | B) - P(A)| \\ &= \sup_{A \in \mathcal{G}} \|P(A | \mathcal{H}) - P(A)\|_\infty, \end{aligned}$$

where $\|\cdot\|_\infty$ denotes the norm for $L^p(\Omega, \mathcal{F}, P)$. The proof of equality of the two expressions is left as a problem. The measure of *strong mixing* is given by

$$\begin{aligned} (2.2) \quad \alpha(\mathcal{G}, \mathcal{H}) &= \sup_{A \in \mathcal{G}} \sup_{B \in \mathcal{H}} |P(AB) - P(A)P(B)| \\ &= \frac{1}{2} \sup_{A \in \mathcal{G}} E[|P(A | \mathcal{H}) - P(A)|] \\ &= \frac{1}{2} \sup_{B \in \mathcal{H}} E[|P(B | \mathcal{G}) - P(B)|] \\ &= \frac{1}{2} \sup_{A \in \mathcal{G}} \|P(A | \mathcal{H}) - P(A)\|_1. \end{aligned}$$

Again the equality of the four expressions is left as a problem.

A comparison of the right sides of (2.1) and (2.2) suggests the following general definition. For $1 \leq p \leq \infty$ set

$$(2.3) \quad \varphi_p(\mathcal{G} | \mathcal{H}) = \sup_{A \in \mathcal{G}} \|P(A | \mathcal{H}) - P(A)\|_p.$$

Note that $\varphi = \varphi_\infty$ and $\alpha = \frac{1}{2}\varphi_1$.

Let E_1 and E_2 be separable metric spaces. Let X be E_1 -valued and \mathcal{G} -measurable, and let Y be E_2 -valued and \mathcal{H} -measurable. The primary application of measures of mixing is to estimate differences such as

$$(2.4) \quad \left| E[\psi(X, Y)] - \int \psi(x, y) \mu_X(dx) \mu_Y(dy) \right|$$

where μ_X and μ_Y are the distributions of X and Y . Of course if X and Y are independent then (2.4) is zero. We need the following lemma.

2.1 Lemma Let μ_1 and μ_2 be measures on \mathcal{G} , and let $\|\mu_1 - \mu_2\|$ denote the total variation of $\mu_1 - \mu_2$. Let $r, s \in [1, \infty]$, $r^{-1} + s^{-1} = 1$. Then for g in $L^r(\Omega, \mathcal{G}, \mu_1 + \mu_2)$,

$$(2.5) \quad \left| \int g \, d\mu_1 - \int g \, d\mu_2 \right| \leq \|\mu_1 - \mu_2\|^{1/r} \left[\int |g|^s \, d\mu_1 + \int |g|^s \, d\mu_2 \right]^{1/s}.$$

Proof. Let f_i be the Radon–Nikodym derivative of μ_i with respect to $\mu_1 + \mu_2$. Then

$$(2.6) \quad \begin{aligned} \|\mu_1 - \mu_2\| &\equiv \sup_{A \in \mathcal{G}} (\mu_1(A) - \mu_2(A)) + \sup_{A \in \mathcal{G}} (\mu_2(A) - \mu_1(A)) \\ &= \int |f_1 - f_2| \, d(\mu_1 + \mu_2), \end{aligned}$$

and

$$\begin{aligned} (2.7) \quad \left| \int g \, d\mu_1 - \int g \, d\mu_2 \right| &= \left| \int g(f_1 - f_2) \, d(\mu_1 + \mu_2) \right| \\ &\leq \left[\int |g|^s |f_1 - f_2| \, d(\mu_1 + \mu_2) \right]^{1/s} \left[\int |f_1 - f_2| \, d(\mu_1 + \mu_2) \right]^{1/s} \\ &\leq \left[\int |g|^s \, d\mu_1 + \int |g|^s \, d\mu_2 \right]^{1/s} \|\mu_1 - \mu_2\|^{1/r}. \end{aligned} \quad \square$$

2.2 Proposition Let $1 \leq r, s, p, q \leq \infty$, $r^{-1} + s^{-1} = 1$, $p^{-1} + q^{-1} = 1$. Then for real-valued Y, Z with Y in $L^q(\Omega, \mathcal{H}, P)$ and Z in $L^p(\Omega, \mathcal{G}, P)$,

$$(2.8) \quad |E[ZY] - E[Z]E[Y]| \leq 2^{q+2} \varphi_p^{1/r}(\mathcal{G} | \mathcal{H}) \|Z\|_{sp} \|Y\|_q.$$

2.3 Remark Note that for $q > 2$ we may select $s = q/p$ so that (2.8) becomes

$$(2.9) \quad |E[ZY] - E[Z]E[Y]| \leq 4\varphi_p^{(q-p)/q}(\mathcal{G}|\mathcal{H})\|Z\|_q\|Y\|_q. \quad \square$$

Proof. First assume $Y \geq 0$. For $A \in \mathcal{G}$, define $\mu_1(A) = E[\chi_A Y]$ and $\mu_2(A) = P(A)E[Y]$. Then

$$\begin{aligned} (2.10) \quad & \|\mu_1 - \mu_2\| \leq 2 \sup_{A \in \mathcal{G}} |\mu_1(A) - \mu_2(A)| \\ &= 2 \sup_{A \in \mathcal{G}} |E[(E[\chi_A|\mathcal{H}] - P(A))Y]| \\ &\leq 2 \sup_{A \in \mathcal{G}} \|P(A|\mathcal{H}) - P(A)\|_p \|Y\|_q \\ &= 2\varphi_p(\mathcal{G}|\mathcal{H})\|Y\|_q. \end{aligned}$$

By Lemma 2.1,

$$\begin{aligned} (2.11) \quad & |E[ZY] - E[Z]E[Y]| = \left| \int Z d\mu_1 - \int Z d\mu_2 \right| \\ &\leq 2^{1/r}\varphi_p^{1/r}(\mathcal{G}|\mathcal{H})\|Y\|_q^{1/r}(E[|Z|^s Y] + E[|Z|^s]E[Y])^{1/s} \\ &\leq 2\varphi_p^{1/r}(\mathcal{G}|\mathcal{H})\|Z\|_s\|Y\|_q, \end{aligned}$$

since both $E[|Z|^s Y]$ and $E[|Z|^s]E[Y]$ are bounded by $E[|Z|^{sp}]^{1/p}\|Y\|_q$.

For general Y , apply (2.11) to Y^+ and Y^- and add to obtain (2.8). Note that $\|Y^+\|_q + \|Y^-\|_q \leq 2\|Y\|_q$ for all q , and for $q < 2$ this can be improved to $\|Y^+\|_q + \|Y^-\|_q \leq 2^{q-1}\|Y\|_q$. \square

2.4 Corollary Let $1 \leq r, s \leq \infty$, $r^{-1} + s^{-1} = 1$. Then for real-valued Z in $L(\Omega, \mathcal{G}, P)$,

$$(2.12) \quad E[|E[Z|\mathcal{H}] - E[Z]|] \leq 8\varphi_r^{1/r}(\mathcal{G}|\mathcal{H})\|Z\|_r.$$

Proof. Let Y_1 be the indicator of the event $\{E[Z|\mathcal{H}] - E[Z] \geq 0\}$ and $Y_2 = 1 - Y_1$. Then

$$\begin{aligned} (2.13) \quad & E[|E[Z|\mathcal{H}] - E[Z]|] \\ &= E[E[Z|\mathcal{H}]Y_1 - E[Z]Y_1] + E[E[Z|\mathcal{H}]Y_2 - E[Z]Y_2] \\ &= E[ZY_1] - E[Z]E[Y_1] + |E[ZY_2] - E[Z]E[Y_2]| \end{aligned}$$

and (2.12) follows from (2.8). \square

2.5 Corollary Let $1 \leq u, v, w \leq \infty$, $u^{-1} + v^{-1} + w^{-1} = 1$. Then for real-valued Y, Z with Y in $L^w(\Omega, \mathcal{H}, P)$ and Z in $L^v(\Omega, \mathcal{G}, P)$,

$$(2.14) \quad |E[ZY] - E[Z]E[Y]| \leq 2^{u+w+2+1}\alpha^{1/u}(\mathcal{G}, \mathcal{H})\|Z\|_v\|Y\|_w.$$

Proof. By the symmetry of α in \mathcal{G} and \mathcal{H} , it is enough to consider the case $w \leq v$. Let $q = w$, $p = q/(q-1)$, $s = v/p$, and $r = s/(s-1)$. Note that since $v^{-1} + w^{-1} = v^{-1} + q^{-1} \leq 1$ we must have $v \geq p$ and hence $s \geq 1$, and that $u = pr$.

By (2.8),

$$(2.15) \quad |E[ZY] - E[Z]E[Y]| \leq 2^{w^{-1}}\varphi_p^{p/w}(\mathcal{G}|\mathcal{H})\|Z\|_v\|Y\|_w.$$

Finally note that

$$(2.16) \quad \varphi_p^{p/w}(\mathcal{G}|\mathcal{H}) = \sup_{A \in \mathcal{G}} E[|P(A|\mathcal{H}) - P(A)|^p]^{1/p}$$

is a decreasing function of p . Replacing p by 1 in (2.15) and φ_1 by 2α gives (2.14). \square

In the uniform mixing case ($p = \infty$) much stronger results are possible. Note that for each $A \in \mathcal{G}$

$$(2.17) \quad |P(A|\mathcal{H}) - P(A)| \leq \varphi_\infty(\mathcal{G}|\mathcal{H}) \quad \text{a.s.}$$

where $\varphi_\infty(\mathcal{G}|\mathcal{H})$ is, of course, a constant. We relax this requirement by assuming the existence of an \mathcal{H} -measurable random variable Φ such that

$$(2.18) \quad |P(A|\mathcal{H}) - P(A)| \leq \Phi \quad \text{a.s.}$$

for each $A \in \mathcal{G}$. (See Problem 9.) To see why this generalization is potentially useful, consider a Markov process X with values in a complete, separable metric space, transition function $P(t, x, \Gamma)$, and initial distribution v . Let $\mathcal{G} = \mathcal{F}^{t+s} = \sigma(X(u): u \geq t+s)$ and $\mathcal{H} = \mathcal{F}_t = \sigma(X(u): u \leq t)$. By the Markov property, for $A \in \mathcal{F}^{t+s}$ there is a function h_A such that $E[\chi_A | \mathcal{F}_{t+s}] = h_A(X_{t+s})$. Therefore for $A \in \mathcal{F}^{t+s}$,

$$\begin{aligned} (2.19) \quad & |P(A|\mathcal{F}_t) - P(A)| \\ &= |E[h_A(X_{t+s})|\mathcal{F}_t] - E[h_A(X_{t+s})]| \\ &= \left| \int h_A(y)P(s, X_t, dy) - \iint h_A(y)P(t+s, x, dy)v(dx) \right| \\ &\leq \psi_{t,s}(X_t) \end{aligned}$$

where

$$(2.20) \quad \psi_{t,s}(z) = \sup_{\Gamma} \left| P(s, z, \Gamma) - \int P(t+s, x, \Gamma)v(dx) \right|.$$

For examples, see Problem 10.

2.6 Proposition Let $1 < s \leq \infty$ and $r^{-1} + s^{-1} = 1$. Suppose that Φ is \mathcal{H} -measurable and satisfies (2.18). Then for real-valued Z in $L^r(\Omega, \mathcal{G}, P)$,

$$(2.21) \quad |E[Z|\mathcal{H}] - E[Z]| \leq 2^{1/r}\Phi^{1/r}(E[|Z|^s|\mathcal{H}] + E[|Z|^r])^{1/s},$$

$$(2.22) \quad \|E[Z|\mathcal{H}] - E[Z]\|_s \leq 2 \max (\|\Phi^{1/r}Z\|_s, \|\Phi^{1/r}\|_s\|Z\|_s),$$

and for $1 \leq p \leq \infty$,

$$(2.23) \quad \|E[Z|\mathcal{H}] - E[Z]\|_p \leq 2\|\Phi\|_p^{1/r}\|E[|Z|^s|\mathcal{H}]\|_p^{1/s}.$$

Proof. Fix $B \in \mathcal{H}$ with $P(B) > 0$, and take $\mu_1(A) = P(A|B)$ and $\mu_2(A) = P(A)$, $A \in \mathcal{G}$. Then noting that $\|\mu_1 - \mu_2\| \leq 2P(B)^{-1} \int_B \Phi dP$, Lemma 2.1 gives

$$(2.24) \quad \left| P(B)^{-1} \int_B E[Z|\mathcal{H}] dP - E[Z] \right| \\ \leq 2^{1/r}P(B)^{-1/r} \left[\int_B \Phi dP \right]^{1/r} \left(P(B)^{-1} \int_B E[|Z|^2|\mathcal{H}] dP + E[|Z|^r] \right)^{1/s}.$$

For $\alpha, \beta > 0$, let $B = \{E[Z|\mathcal{H}] - E[Z] > 2^{1/r}\alpha\beta, \Phi \leq \alpha^r, E[|Z|^s|\mathcal{H}] + E[|Z|^r] \leq \beta^s\}$. If $P(B) > 0$, then (2.24) is violated. Consequently, $P(B) = 0$ for all choices of α and β , which implies

$$(2.25) \quad E[Z|\mathcal{H}] - E[Z] \leq 2^{1/r}\Phi^{1/r}(E[|Z|^s|\mathcal{H}] + E[|Z|^r])^{1/s}.$$

A similar argument gives the estimate for $E[Z] - E[Z|\mathcal{H}]$. Finally (2.21) and the Hölder inequality give (2.22) and (2.23). \square

2.7 Corollary Let $1 < s \leq \infty$ and $r^{-1} + s^{-1} = 1$. Suppose that Φ is \mathcal{H} -measurable and satisfies (2.18). Then for real-valued Y, Z , with Y in $L^r(\Omega, \mathcal{H}, P)$ and Z in $L^r(\Omega, \mathcal{G}, P)$,

$$(2.26) \quad |E[ZY] - E[Z]E[Y]| \leq 2 \max (\|\Phi^{1/r}Z\|_s, \|\Phi^{1/r}\|_s\|Z\|_s)\|Y\|_r,$$

and

$$(2.27) \quad |E[ZY] - E[Z]E[Y]| \leq 2\|Y\|_r\|Z\|_s.$$

In particular

$$(2.28) \quad |E[ZY] - E[Z]E[Y]| \leq 2\varphi_\infty^{1/r}(\mathcal{G}|\mathcal{H})\|Z\|_s\|Y\|_r.$$

Proof. Use (2.21) to estimate $E[(E[Z|\mathcal{H}] - E[Z])Y]$ and apply the Hölder inequality. \square

2.8 Corollary Let $1 < s \leq \infty$ and $r^{-1} + s^{-1} = 1$. Suppose that Φ is \mathcal{H} -measurable and satisfies (2.18). Let E_1 and E_2 be separable. Let X be \mathcal{G} -measurable and E_1 -valued, let Y be \mathcal{H} -measurable and E_2 -valued, and let μ_X

and μ_Y denote the distributions of X and Y . Then for ψ in $L^r(E_1 \times E_2, \mathcal{B}(E_1 \times E_2), \mu_X \times \mu_Y)$ such that $\psi(X, Y)$ is in $L^s(\Omega, \mathcal{F}, P)$,

$$(2.29) \quad \left| E[\psi(X, Y)] - \iint \psi(x, y) \mu_X(dx) \mu_Y(dy) \right| \\ \leq 2(E[\Phi])^{1/r} \max \left(\|\psi(X, Y)\|_s, \left[\iint |\psi(x, y)|^s \mu_X(dx) \mu_Y(dy) \right]^{1/s} \right).$$

Proof. Since we can always approximate ψ by $\psi_n = n \wedge (\psi \vee (-n))$, we may as well assume that ψ is bounded, and since the collection of bounded ψ satisfying (2.28) is bp-closed, we may as well assume that ψ is continuous (see Appendix 4). Finally, if ψ is bounded and continuous, we can obtain (2.28) for arbitrary Y by approximating by discrete Y (recall E_2 is separable), so we may as well assume that Y is discrete.

There exists a $\mathcal{B}(E_2) \times \mathcal{H}$ -measurable function $\varphi(y, \omega)$ such that $E[\psi(X, y) | \mathcal{H}] = \varphi(y, \cdot)$ for each y and $E[\psi(X, Y) | \mathcal{H}] = \varphi(Y(\cdot), \cdot)$. (See Appendix 4.) By (2.21)

$$(2.30) \quad \left| \varphi(y, \cdot) - \int \psi(x, y) \mu_X(dx) \right| \\ \leq 2^{1/r} \Phi^{1/r} \left(E[|\psi(X, y)|^s | \mathcal{H}] + \int |\psi(x, y)|^s \mu_X(dx) \right)^{1/s},$$

and hence, since Y is discrete,

$$(2.31) \quad \left| E[\psi(X, Y) | \mathcal{H}] - \int \psi(x, Y) \mu_X(dx) \right| \\ \leq 2^{1/r} \Phi^{1/r} \left(E[|\psi(X, Y)|^s | \mathcal{H}] + \int |\psi(x, Y)|^s \mu_X(dx) \right)^{1/s}.$$

Taking expectations in (2.31) and applying the Hölder inequality gives (2.29). \square

3. CENTRAL LIMIT THEOREMS FOR STATIONARY SEQUENCES

In this section we apply the martingale central limit theorem of Section 1 to extend the invariance principle to stationary sequences of random variables. Let $\{Y_k, k \in \mathbb{Z}\}$ be \mathbb{R} -valued and stationary, and define $\mathcal{F}_n = \sigma(Y_k : k \leq n)$ and $\mathcal{F}^n = \sigma(Y_k : k \geq n)$. For $m \geq 0$, let

$$(3.1) \quad \varphi_p(m) = \varphi_p(\mathcal{F}^{n+m} | \mathcal{F}_n).$$

The stationarity of $\{Y_k\}$ implies that the right side of (3.1) is independent of n . (See Problem 12.)

We are interested in

$$(3.2) \quad X_n(t) = \frac{1}{\sqrt{n}} \sum_{k=1}^{\lfloor nt \rfloor} Y_k.$$

3.1 Theorem Let $\{Y_k, k \in \mathbb{Z}\}$ be stationary with $E[Y_k] = 0$, and for some $\delta > 0$ suppose $E[|Y_k|^{2+\delta}] < \infty$. Let $p = (2 + \delta)/(1 + \delta)$ and suppose

$$(3.3) \quad \sum_m [\varphi_p(m)]^{\delta/(1+\delta)} < \infty.$$

Then the series

$$(3.4) \quad \sigma^2 = E[Y_1^2] + 2 \sum_{k=2}^{\infty} E[Y_1 Y_k]$$

is convergent and $X_n \Rightarrow X$, where X is Brownian motion with mean zero and variance parameter σ^2 .

3.2 Remark (a) The assumption that $\{Y_k\}$ is indexed by all $k \in \mathbb{Z}$ is a convenience. Any stationary sequence indexed by $k = 1, 2, \dots$ has a version that is part of a stationary sequence indexed by \mathbb{Z} . Specifically, given $\{X_k, k \geq 1\}$, if $\{X_k\}$ is stationary, then

$$(3.5) \quad P\{Y_{l+1} \in \Gamma_1, Y_{l+2} \in \Gamma_2, \dots, Y_{l+m} \in \Gamma_m\} \\ = P\{X_1 \in \Gamma_1, X_2 \in \Gamma_2, \dots, X_m \in \Gamma_m\}, \quad l \in \mathbb{Z}, m = 1, 2, \dots, \Gamma_k \in \mathcal{P}(\mathbb{R}),$$

determines a consistent family of finite-dimensional distributions.

(b) By (2.16), for $p = (2 + \delta)/(1 + \delta)$,

$$(3.6) \quad \sum_m [\varphi_p(m)]^{\delta/(1+\delta)} \leq \sum_m [\varphi_1(m)]^{\delta/(2+\delta)}.$$

Proof. By Corollary 2.4,

$$(3.7) \quad E[|E[Y_{n+m} | \mathcal{F}_n]|] \leq 8\varphi_1^{(1+\delta)/(2+\delta)}(m)\|Y_{n+m}\|_{2+\delta} \\ \leq 8\varphi_p^{(1+\delta)/(2+\delta)}(m)\|Y_1\|_{2+\delta} \\ \leq 8\varphi_p^{\delta/(1+\delta)}(m)\|Y_1\|_{2+\delta}.$$

Consequently the sum on the right of

$$(3.8) \quad M(l) = \sum_{k=1}^l Y_k + \sum_{m=1}^{\infty} E[Y_{l+m} | \mathcal{F}_l]$$

is convergent, and M is a martingale.

The convergence of the series in (3.4) follows from (2.9), which gives

$$(3.9) \quad E[Y_1 Y_k] \leq 4\varphi_p^{\delta/(1+\delta)}(k-1)\|Y_1\|_{2+\delta}\|Y_k\|_{2+\delta}.$$

Note that

$$(3.10) \quad M(l) - M(l-1) = Y_l + \sum_{m=1}^{\infty} E[Y_{l+m} | \mathcal{F}_l] - \sum_{m=1}^{\infty} E[Y_{l-1+m} | \mathcal{F}_{l-1}] \\ = \sum_{m=0}^{\infty} E[Y_{l+m} | \mathcal{F}_l] - \sum_{m=0}^{\infty} E[Y_{l+m} | \mathcal{F}_{l-1}]$$

is a stationary sequence. The sequence

$$(3.11) \quad G_N = \sum_{m=0}^N E[Y_{l+m} | \mathcal{F}_l] - \sum_{m=0}^N E[Y_{l+m} | \mathcal{F}_{l-1}]$$

converges in L^2 (check that it is Cauchy), so

$$(3.12) \quad E[(M(l) - M(l-1))^2] = \lim_{N \rightarrow \infty} E[G_N^2] \\ = \lim_{N \rightarrow \infty} \left(E\left[\left(\sum_{m=0}^N E[Y_{l+m} | \mathcal{F}_l] \right)^2 \right] - E\left[\left(\sum_{m=0}^N E[Y_{l+m} | \mathcal{F}_{l-1}] \right)^2 \right] \right) \\ = \lim_{N \rightarrow \infty} \left(E\left[\left(\sum_{m=0}^N E[Y_{l+m} | \mathcal{F}_l] \right)^2 \right] - E\left[\left(\sum_{m=1}^{N+1} E[Y_{l+m} | \mathcal{F}_l] \right)^2 \right] \right) \\ = \lim_{N \rightarrow \infty} \left(E[Y_l^2] + 2 \sum_{m=1}^N E[Y_l Y_{l+m}] - E[E[Y_{l+N+1} | \mathcal{F}_l]^2] \right. \\ \left. - 2 \sum_{m=1}^N E[E[Y_{l+N+1} | \mathcal{F}_l] E[Y_{l+m} | \mathcal{F}_l]] \right) \\ = \sigma^2.$$

We have used the stationarity here and the fact that

$$(3.13) \quad |E[E[Y_{l+N+1} | \mathcal{F}_l] E[Y_{l+m} | \mathcal{F}_l]]| \\ = |E[Y_{l+N+1} E[Y_{l+m} | \mathcal{F}_l]]| \\ \leq 4\varphi_p^{\delta/(1+\delta)}(N+1) \|Y_{l+N+1}\|_{2+\delta} \|Y_{l+m}\|_{2+\delta}.$$

The fact that $(N+1)\varphi_p^{\delta/(1+\delta)}(N+1) \rightarrow 0$ as $N \rightarrow \infty$ follows from (3.3) and the monotonicity of $\varphi_p(m)$. Since $\{Y_k\}$ is mixing, it is ergodic (see, e.g., Lamperti (1977), page 96), and the ergodic theorem gives

$$(3.14) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{l=1}^{[nt]} (M(l) - M(l-1))^2 = \sigma^2 t \quad \text{a.s.}$$

Define $M_n(t) = n^{-1/2} M([nt])$. Then (3.14) gives (1.19).

To obtain (1.14), the stationarity of $\{M(l) - M(l-1)\}$ implies that for each $\epsilon > 0$,

$$\begin{aligned}
 (3.15) \quad & E \left[\sup_{t \leq T} |M_n(t) - M_n(t-1)| \right] \\
 &= \int_0^\infty P \left\{ \sup_{t \leq T} |M_n(t) - M_n(t-1)| > x \right\} dx \\
 &\leq \epsilon + \int_\epsilon^\infty [nT] P \{ |M(1) - M(0)| > \sqrt{nx} \} dx, \\
 &\leq \epsilon + T\epsilon^{-1} E[|M(1) - M(0)|^2 \chi_{\{|M(1) - M(0)| > \sqrt{n}\epsilon\}}].
 \end{aligned}$$

By Theorem 1.4(a), $M_n \Rightarrow X$.

Finally note that $\sup_{t \leq T} |X_n(t) - M_n(t)| \rightarrow 0$ in probability by the same type of estimate used in (3.15), so $X_n \Rightarrow X$. \square

Now let $\Phi_n(m)$ be a random variable satisfying

$$(3.16) \quad |P(A | \mathcal{F}_n) - P(A)| \leq \Phi_n(m) \quad \text{a.s.}$$

for each $A \in \mathcal{F}^{n+m}$. Without loss of generality we can assume that for each m , $\{\Phi_n(m)\}$ is stationary and $\Phi_n(m) \leq 1$ a.s.

3.3 Theorem Let $\{Y_k, k \in \mathbb{Z}\}$ be stationary with $E[Y_k] = 0$. Let $1 < s \leq \infty$ and $s^{-1} + r^{-1} = 1$. Suppose $E[|Y_k|^{r+s}] < \infty$,

$$(3.17) \quad \sum_{m=1}^{\infty} \|\Phi_0^{1/r}(m) Y_m\|_s < \infty,$$

and

$$(3.18) \quad \sum_{m=1}^{\infty} \|\Phi_0^{1/r}(m)\|_s < \infty.$$

Then the series in (3.4) converges, and $X_n \Rightarrow X$, where X is Brownian motion with mean zero and variance parameter σ^2 .

3.4 Remark If

$$(3.19) \quad \sum_{m=1}^{\infty} \varphi_{\infty}^{1/r}(m) < \infty,$$

then (3.17) and (3.18) hold. \square

Proof. The proof is essentially the same as for Theorem 3.1 using (2.22) to estimate the left side of (3.7) and (2.26) to estimate the left sides of (3.9) and (3.13). \square

4. DIFFUSION APPROXIMATIONS

We now give conditions analogous to those of Theorem 1.4 for the convergence to general diffusion processes.

4.1 Theorem Let $a = ((a_{ij}))$ be a continuous, symmetric, nonnegative definite, $d \times d$ matrix-valued function on \mathbb{R}^d and let $b: \mathbb{R}^d \rightarrow \mathbb{R}^d$ be continuous. Let

$$A = \{(f, Gf \equiv \frac{1}{2} \sum a_{ij} \partial_i \partial_j f + \sum b_i \partial_i f): f \in C_c^\infty(\mathbb{R}^d)\},$$

and suppose that the $C_{\mathbb{R}^d}[0, \infty)$ martingale problem for A is well-posed.

For $n = 1, 2, \dots$, let X_n and B_n be processes with sample paths in $D_{\mathbb{R}^d}[0, \infty)$, and let $A_n = ((A_n^{ij}))$ be a symmetric $d \times d$ matrix-valued process such that A_n^{ij} has sample paths in $D_{\mathbb{R}}[0, \infty)$ and $A_n(t) - A_n(s)$ is nonnegative definite for $t > s \geq 0$. Set $\mathcal{F}_t^n = \sigma(X_n(s), B_n(s), A_n(s): s \leq t)$.

Let $\tau'_n = \inf \{t: |X_n(t)| \geq r \text{ or } |X_n(t-)| \geq r\}$, and suppose that

$$(4.1) \quad M_n \equiv X_n - B_n$$

and

$$(4.2) \quad M_n^i M_n^j - A_n^{ij}, \quad i, j = 1, 2, \dots, d,$$

are $\{\mathcal{F}_t^n\}$ -local martingales, and that for each $r > 0$, $T > 0$, and $i, j = 1, \dots, d$,

$$(4.3) \quad \lim_{n \rightarrow \infty} E \left[\sup_{t \leq T \wedge \tau'_n} |X_n(t) - X_n(t-)|^2 \right] = 0,$$

$$(4.4) \quad \lim_{n \rightarrow \infty} E \left[\sup_{t \leq T \wedge \tau'_n} |B_n(t) - B_n(t-)|^2 \right] = 0,$$

$$(4.5) \quad \lim_{n \rightarrow \infty} E \left[\sup_{t \leq T \wedge \tau'_n} |A_n^{ij}(t) - A_n^{ij}(t-)| \right] = 0,$$

$$(4.6) \quad \sup_{t \leq T \wedge \tau'_n} \left| B_n^i(t) - \int_0^t b_i(X_n(s)) ds \right|^p \rightarrow 0,$$

and

$$(4.7) \quad \sup_{t \leq T \wedge \tau'_n} \left| A_n^{ij}(t) - \int_0^t a_{ij}(X_n(s)) ds \right|^p \rightarrow 0.$$

Suppose that $PX_n(0)^{-1} \Rightarrow v \in \mathcal{P}(\mathbb{R}^d)$. Then $\{X_n\}$ converges in distribution to the solution of the martingale problem for (A, v) .

Proof. Let

$$(4.8) \quad \eta'_n = \inf \left\{ t: A_n^{ii}(t) > t \sup_{|x| \leq r} a_{ii}(x) + 1 \text{ for some } i \right\},$$

and set $\tilde{M}_n' = M_n(\cdot \wedge \eta_n' \wedge \tau_n')$. Relative compactness of $\{\tilde{M}_n'\}$ follows as in the proof of Theorem 1.4(b), which in turn implies relative compactness for $\{X_n(\cdot \wedge \tau_n')\}$ since $\{B_n(\cdot \wedge \tau_n')\}$ is relatively compact by (4.6). Fix $r_0 > 0$ and let $\{X_{n_k}(\cdot \wedge \tau_{n_k}')\}$ be a convergent subsequence with limit X^{r_0} . For all but countably many $r < r_0$,

$$(X_{n_k}(\cdot \wedge \tau_{n_k}'), \tau_{n_k}') \Rightarrow (X^{r_0}(\cdot \wedge \tau'), \tau'),$$

where $\tau' = \inf \{t : |X^{r_0}(t)| \geq r\}$ (i.e., for all $r < r_0$ such that $P\{\lim_{s \rightarrow r} \tau^s = \tau'\} = 1$).

Again as in the proof of Theorem 1.4(b),

$$(4.9) \quad M'^{r_0}(t \wedge \tau_r) \equiv X^{r_0}(t \wedge \tau') - \int_0^{t \wedge \tau'} b(X^{r_0}(s)) ds$$

and

$$(4.10) \quad M'^{r_0}_i(t \wedge \tau') M'^{r_0}_j(t \wedge \tau') - \int_0^{t \wedge \tau'} a_{ij}(X^{r_0}(s)) ds$$

are martingales, and by Itô's formula, Theorem 2.9 of Chapter 5,

$$(4.11) \quad f(X^{r_0}(t \wedge \tau')) - \int_0^{t \wedge \tau'} Gf(X^{r_0}(s)) ds$$

is a martingale for each $f \in C_c^\infty(\mathbb{R}^d)$. Uniqueness for the martingale problem for (A, v) implies uniqueness for the stopped martingale problem for $(A, v, \{x : |x| < r\})$. Consequently, if X is a solution of the martingale problem for (A, v) , then $X_n(\cdot \wedge \tau_n') \Rightarrow X(\cdot \wedge \tau')$ for all r such that $P\{\lim_{s \rightarrow r} \tau^s = \tau'\} = 1$ (here $\tau' = \inf \{t : |X(t)| \geq r\}$). But $\tau_r \rightarrow \infty$ as $r \rightarrow \infty$ (since X has sample paths in $C_{\mathbb{R}^d}[0, \infty)$), so $X_n \Rightarrow X$. \square

4.2 Corollary Let a , b , and A be as in Theorem 4.1, and suppose the martingale problem for (A, v) has a unique solution for each $v \in \mathcal{P}(\mathbb{R}^d)$. Let $\mu_n(x, \Gamma)$, $n = 1, 2, \dots$, be a transition function on \mathbb{R}^d , and set

$$(4.12) \quad b_n(x) = n \int_{|y-x| \leq 1} (y-x)\mu_n(x, dy)$$

and

$$(4.13) \quad a_n(x) = n \int_{|y-x| \leq 1} (y-x)(y-x)^T \mu_n(x, dy).$$

Suppose for each $r > 0$ and $\varepsilon > 0$,

$$(4.14) \quad \sup_{|x| \leq r} |a_n(x) - a(x)| \rightarrow 0,$$

$$(4.15) \quad \sup_{|x| \leq r} |b_n(x) - b(x)| \rightarrow 0,$$

and

$$(4.16) \quad \sup_{|x| \leq r} n\mu_n(x, \{y: |y - x| \geq \varepsilon\}) \rightarrow 0.$$

Let Y_n be a Markov chain with transition function $\mu_n(x, \Gamma)$ and define $X_n(t) = Y_n(\lfloor nt \rfloor)$. If $PY_n(0)^{-1} \Rightarrow v$, then $\{X_n\}$ converges in distribution to the solution of the martingale problem (A, v) .

Proof. Let τ_n^r be as in the proof of Theorem 4.1, and let $\gamma_n = \inf\{t: |X_n(t) - X_n(t-)| > 1\}$. Then (4.16) implies $P\{\gamma_n < \tau_n^r \wedge T\} \rightarrow 0$ for each $r > 0$ and $T > 0$. Therefore (see Problem 13), we may as well assume $\mu_n(x, \{y: |y - x| > 1\}) = 0$. Let

$$(4.17) \quad B_n(t) = \int_0^{\lfloor nt \rfloor/n} b_n(X_n(s)) ds$$

and

$$(4.18) \quad A_n(t) = \int_0^{\lfloor nt \rfloor/n} (a_n(X_n(s)) - n^{-1}b_n(X_n(s))b_n^T(X_n(s))) ds,$$

and (4.3)–(4.7) follow easily. \square

5. STRONG APPROXIMATION THEOREMS

In this section we present, without proof, strong approximation theorems of Komlós, Major, and Tusnády for sums of independent identically distributed random variables. We obtain as a corollary a result on the approximation of the Poisson process by Brownian motion. To understand the significance of a strong approximation theorem, it may be useful to recall Theorem 1.2 of Chapter 3. This theorem can be restated to say that if $\mu, v \in \mathcal{P}(S)$ and $\rho(\mu, v) < \varepsilon$, then there exist a probability space (Ω, \mathcal{F}, P) and random variables X and Y defined on (Ω, \mathcal{F}, P) , X with distribution μ , Y with distribution v , such that $P\{d(X, Y) \geq \varepsilon\} \leq \varepsilon$.

5.1 Theorem Let $\mu \in \mathcal{P}(\mathbb{R})$ satisfy $\int e^{\alpha x} \mu(dx) < \infty$ for $|\alpha| \leq \alpha_0$, some $\alpha_0 > 0$. Then there exist a sequence of independent identically distributed random variables $\{\zeta_k\}$ with distribution μ , a Brownian motion W with $m \equiv E[W(1)] = E[\zeta_1]$ and $\sigma^2 \equiv \text{var}(W(1)) = \text{var}(\zeta_1)$ (defined on the same sample space), and positive constants C, K , and λ depending only on μ , such that

$$(5.1) \quad P \left\{ \max_{1 \leq k \leq n} |S_k - W(k)| > C \log n + x \right\} < Ke^{-\lambda x}$$

for each $n \geq 1$ and $x > 0$, where $S_k = \sum_{i=1}^k \zeta_i$.

Proof. See Komlós, Major, and Tusnády (1975, 1976). \square

5.2 Corollary Let $\{\xi_k\}$ and W be as in Theorem 5.1. Set $X_n(t) = n^{-1/2} \sum_{k=1}^{\lfloor nt \rfloor} (\xi_k - E[\xi_k])$ and $W_n(t) = n^{-1/2}(W(nt) - E[W(nt)])$. (Note that $W_n(t)$ is a Brownian motion with mean zero and $\text{var}(W_n(t)) = t \text{ var}(\xi_1)$.) Then there exist positive constants C , K , γ , and λ , depending only on μ , such that for $T \geq 1$, $n \geq 1$, and $x > 0$,

$$(5.2) \quad P \left\{ \sup_{t \leq T} |X_n(t) - W_n(t)| > Cn^{-1/2} \log n + x \right\} \leq KT^\gamma e^{-\lambda\sqrt{nx}}.$$

It follows that there exists a $\beta > 0$ such that for $n \geq 2$, $\rho(PX_n^{-1}, PW_n^{-1}) \leq \beta n^{-1/2} \log n$, where ρ is the Prohorov metric on $\mathcal{P}(D_R[0, \infty))$.

Proof. Let C_1 , K_1 , and λ_1 be the C , K , and λ guaranteed by Theorem 5.1 and set $C = 2C_1$. Then defining $\tilde{W}(t) = W(t) - t$, the left side of (5.2) is bounded by

$$(5.3) \quad P \left\{ \sup_{k \leq nT} |S_k - W(k)| > C_1 \log [nT] - C_1 \log T + \frac{\sqrt{nx}}{2} \right\} \\ + P \left\{ \sup_{k \leq nT} \sup_{0 \leq s \leq 1} |\tilde{W}(k+s) - \tilde{W}(k)| > C_1 \log n + \frac{\sqrt{nx}}{2} \right\}.$$

The second term in (5.3) is bounded by

$$(5.4) \quad nTP \left\{ \sup_{0 \leq s \leq 1} |\tilde{W}(s)| > C_1 \log n + \frac{\sqrt{nx}}{2} \right\},$$

and for any $\alpha > 0$,

$$(5.5) \quad P \left\{ \sup_{0 \leq s \leq 1} |\tilde{W}(s)| > z \right\} \leq 2e^{(1/2)\alpha^2 n^2} e^{-\alpha z}$$

(see Problem 17). Selecting $\alpha > \lambda_1$ so that $\alpha C_1 > 1$, (5.2) is bounded by

$$(5.6) \quad K_1 \exp \left\{ -\lambda_1 \left(-C_1 \log T + \frac{\sqrt{nx}}{2} \right) \right\} + K_2 T \exp \left\{ \frac{-\lambda_1 \sqrt{nx}}{2} \right\} \\ \leq KT^\gamma \exp \{ -\lambda_1 \sqrt{nx} \},$$

for $\gamma = (\lambda_1 C_1) \vee 1$, $\lambda = \lambda_1/2$, and $K = K_1 + K_2$.

For $a_k > 0$, $k = 1, 2, \dots$, with $\sum_k a_k = 1$, $A > 0$ to be determined, and $n \geq 2$,

$$\begin{aligned}
 (5.7) \quad P\{d(X_n, W_n) > An^{-1/2} \log n\} \\
 &\leq P\left\{\int_0^\infty e^{-t} \sup_{s \leq t} |X_n(s) - W_n(s)| dt > An^{-1/2} \log n\right\} \\
 &\leq \sum_k P\left\{\int_{k-1}^k e^{-t} \sup_{s \leq t} |X_n(s) - W_n(s)| dt > a_k An^{-1/2} \log n\right\} \\
 &\leq \sum_k P\left\{\sup_{s \leq k} |X_n(s) - W_n(s)| > e^{k-1} a_k An^{-1/2} \log n\right\} \\
 &\leq \sum_k K k^\gamma \exp\{-\lambda(e^{k-1} a_k A - C) \log n\} \\
 &\leq n^{-1} \sum_k K k^\gamma \exp\{-\lambda(e^{k-1} a_k A - C - \lambda^{-1}) \log 2\},
 \end{aligned}$$

provided $e^{k-1} a_k A - C - \lambda^{-1} > 0$ and the sum is finite. Note the a_k and A can be selected so these conditions are satisfied. Finally, select $\beta \geq A$ so that $\beta n^{-1/2} \log n$ bounds the right side for all $n \geq 2$. \square

5.3 Corollary Let $\mu \in \mathcal{P}(\mathbb{R})$ be infinitely divisible and $\int e^{\alpha x} \mu(dx) < \infty$ for $|\alpha| \leq \alpha_0$, some $\alpha_0 > 0$. Then there exists a process X with stationary independent increments, $X(1)$ with distribution μ , a Brownian motion W with the same mean and variance as X , and positive constants C , K , and λ depending only on μ such that for $T \geq 1$ and $x > 0$,

$$(5.8) \quad P\left\{\sup_{t \leq T} |X(t) - W(t)| > C \log T + x\right\} \leq K e^{-\lambda x}.$$

5.4 Remark Note that if we replace x by $x + \gamma \log T$, then (5.8) becomes

$$(5.9) \quad P\left\{\sup_{t \leq T} |X(t) - W(t)| > (C + \gamma) \log T + x\right\} \leq K T^{-\gamma \lambda} e^{-\lambda x}. \quad \square$$

Proof. Let $\{\xi_k\}$ and W be as in Theorem 5.1. (Note that the C , K , and λ of the corollary differ from those of the theorem.) Let $\{X_k\}$ be independent processes with stationary independent increments with the distribution of $X_k(1)$ being μ . Since the distribution on \mathbb{R}^∞ of $\{X_k(1)\}$ is the same as the distribution of $\{\xi_k\}$, by Lemma 5.15 of Chapter 4 we may assume $\{X_k\}$, $\{\xi_k\}$, and W are defined on the same sample space and that $X_k(1) = \xi_k$. Finally define

$$(5.10) \quad X(t) = \sum_{i=1}^{k-1} \xi_i + X_k(t - k + 1), \quad k - 1 \leq t < k,$$

and note that the left side of (5.8) is bounded by

$$(5.11) \quad P \left\{ \max_{k \leq T} |S_k - W(k)| > 3^{-1}C \log [T] + 3^{-1}x \right\}$$

$$+ P \left\{ \max_{k \leq T} \sup_{s \leq 1} |\tilde{X}_k(s)| > 3^{-1}C \log T + 3^{-1}x \right\}$$

$$+ P \left\{ \max_{k \leq T} \sup_{s \leq 1} |\tilde{W}(s+k) - \tilde{W}(k)| > 3^{-1}C \log T + 3^{-1}x \right\}$$

where $\tilde{W}(t) = W(t) - E[W(t)]$ and $\tilde{X}(t) = X(t) - E[X(t)]$. The result follows from (5.1) and Problem 17. \square

5.5 Corollary Let X and W be as in Corollary 5.3. Then

$$(5.12) \quad \sup_t \frac{|X(t) - W(t)|}{\log(2 \vee t)} < \infty \quad \text{a.s.}$$

Proof. Take $\gamma = 1$ in (5.9). Then

$$(5.13) \quad P \left\{ \sup_t \frac{|X(t) - W(t)|}{\log(2 \vee t)} > (C+1)2 \log 2 + \frac{x}{\log 2} \right\}$$

$$\leq \sum_{n=2}^{\infty} P \left\{ \sup_{t \leq 2^n} \frac{|X(t) - W(t)|}{\log(2^{n-1} \vee t)} > \frac{(C+1) \log 2^n}{\log 2^{n-1}} + \frac{x}{\log 2} \right\}$$

$$\leq \sum_{n=2}^{\infty} P \left\{ \sup_{t \leq 2^n} |X(t) - W(t)| > (C+1) \log 2^n + x \right\}$$

$$\leq K(1 - 2^{-\lambda})^{-1} e^{-\lambda x}. \quad \square$$

The construction in Corollary 5.5 is best possible in the following sense.

5.6 Theorem Suppose X is a process with stationary independent increments, sample paths in $D_R[0, \infty)$, and $X(0) = 0$ a.s., W is a Brownian motion, and

$$(5.14) \quad \lim_{t \rightarrow \infty} \frac{|X(t) - W(t)|}{\log(2 \vee t)} = 0 \quad \text{a.s.}$$

Then X is a Brownian motion with the same mean and variance as W .

Proof. See Bártfai (1966). \square

6. PROBLEMS

1. (a) Let N be a counting process (i.e., N is right continuous and constant except for jumps of +1) with $N(0) = 0$. Suppose that C is a continuous nondecreasing function with $C(0) = 0$ and that $N(t) - C(t)$ is a martingale. Show that N has independent Poisson distributed increments, and that the distribution of N is uniquely determined.
- (b) Let $\{N_n\}$ be a sequence of counting processes, with $N_n(0) = 0$, and let A_n , $n = 1, 2, \dots$, be a process with nondecreasing sample paths such that $\sup_t (A_n(t) - A_n(t-)) \leq 1$ and $N_n - A_n$ is a martingale. Let C be a continuous nondecreasing function with $C(0) = 0$, and suppose for each $t \geq 0$ that $A_n(t) \xrightarrow{P} C(t)$. Show that $N_n \Rightarrow N$ where N is the process characterized in part (a).

Remark. In regard to the condition $\sup_t (A_n(t) - A_n(t-)) \leq 1$, consider the following example. Let Y_1 , Y_2 , and A be independent processes, Y_1 and Y_2 Poisson with parameter one and A nondecreasing. Let $N_n(t) = Y_1(A(t))$ and $A_n(t) = nY_2(A(t)/n)$. Then $N_n - A_n$ is a martingale and $A_n(t) \rightarrow 0$ in probability.

2. Let W_1 and W_2 be independent standard Brownian motions and define

$$(6.1) \quad X(t) = \begin{pmatrix} W_1\left(\frac{t}{3}\right) - W_2\left(\frac{2t}{3}\right) \\ W_1\left(\frac{2t}{3}\right) + W_2\left(\frac{t}{3}\right) \end{pmatrix}.$$

Show that for each $\theta \in \mathbb{R}^2$, $\theta \cdot X(t)$ is Brownian motion with variance $|\theta|^2 t$ and hence

$$(6.2) \quad f(\theta \cdot X(t)) - \int_0^t \frac{1}{2} |\theta|^2 f''(\theta \cdot X(s)) ds$$

is a martingale with respect to $\mathcal{F}_t^{\theta \cdot X} = \sigma(\theta \cdot X(s); s \leq t)$ (cf. Theorem 1.2), but that (6.2) is not (in general) an $\{\mathcal{F}_t^X\}$ -martingale.

3. Let N be a Poisson process with parameter 1, and define $V(t) = (-1)^{N(t)}$. For $n = 1, 2, \dots$, let

$$(6.3) \quad X_n(t) = n^{-1} \int_0^{n^2 t} V(s) ds.$$

Show that

$$(6.4) \quad M(t) = V(t) + 2 \int_0^t V(s) ds$$

is a martingale and use this fact to show $X_n \Rightarrow W$ where W is standard Brownian motion.

4. Develop and prove the analogue of the result in Problem 3 in which V is an Ornstein–Uhlenbeck process, that is, V is a diffusion in \mathbb{R} with generator $Af = \frac{1}{2}af'' - bxf'$, $a, b > 0, f \in C_c^\infty(\mathbb{R})$.
5. Let ξ_1, ξ_2, \dots be independent and identically distributed with $\xi_k \geq 0$ a.s., $E[\xi_k] = \mu > 0$, and $\text{var } (\xi_k) = \sigma^2 < \infty$. Let $N(t) = \max \{k : \sum_{i=1}^k \xi_i \leq t\}$, and define

$$(6.5) \quad X_n(t) = n^{-1/2} \left(N(nt) - \frac{nt}{\mu} \right).$$

- (a) Show that

$$(6.6) \quad M(t) = \sum_{k=1}^{N(t)+1} \xi_k - (N(t) + 1)\mu$$

is a martingale.

- (b) Apply Theorem 1.4 to show that $X_n \Rightarrow X$, where X is a Brownian motion with mean zero and variance parameter σ^2/μ^3 .

6. Let E be the unit sphere in \mathbb{R}^3 . Let $\mu(x, \Gamma)$ be a transition function on $E \times \mathcal{P}(E)$ satisfying

$$(6.7) \quad \int y \mu(x, dy) = \rho x, \quad x \in E,$$

for some $\rho \in (-1, 1)$.

Define $Tf(x) = \int f(y) \mu(x, dy)$. Suppose there exists $v \in \mathcal{P}(E)$ such that

$$(6.8) \quad \lim_{n \rightarrow \infty} n^{-1} \sum_{k=1}^{[nt]} T^k f(x) = \int f dv$$

for each $x \in E$ and $f \in C(E)$. Let $\{Y(k), k = 0, 1, 2, \dots\}$ be a Markov chain with transition function $\mu(x, \Gamma)$ and define

$$(6.9) \quad X_n(t) = \frac{1}{\sqrt{n}} \sum_{k=1}^{[nt]} Y_k.$$

Show that

$$(6.10) \quad M_n(t) = X_n(t) + \rho(1 - \rho)^{-1} Y_{[nt]}/\sqrt{n}$$

is a martingale, and use Theorem 1.4 to prove $X_n \Rightarrow X$ where X is a three-dimensional Brownian motion with mean zero and covariance

$$(6.11) \quad C(t) = t \frac{1 + \rho}{1 - \rho} \left(\left(\int y_i y_j v(dy) \right) \right).$$

7. For $n = 1, 2, \dots$, let X_n be a process with sample paths in $D_E[0, \infty)$, and let M_n be a process with sample paths in $D_R[0, \infty)$. Suppose that M_n is an $\{\mathcal{F}_t^{X_n}\}$ -martingale, and that $(X_n, M_n) \Rightarrow (X, M)$. Show that if M is $\{\mathcal{F}_t^X\}$ -adapted and for each $t \geq 0$ $\{M_n(t)\}$ is uniformly integrable, then M is an $\{\mathcal{F}_t^X\}$ -martingale.
8. Verify the identities in (2.1) and (2.2).
9. Let Γ be a collection of nonnegative random variables, and suppose there is a random variable η such that for each $\xi \in \Gamma$, $\xi \leq \eta$ a.s. Show that there is a minimal such η , that is, show that there is a random variable η_0 such that for each $\xi \in \Gamma$, $\xi \leq \eta_0$ a.s. and there exist $\xi_i \in \Gamma$, $i = 1, 2, \dots$, such that $\eta_0 = \sup_i \xi_i$.

Hint: First assume $E[\eta] < \infty$ and consider

$$(6.12) \quad \sup_{\{\xi_i\} \subset \Gamma} E \left[\sup_i \xi_i \right].$$

10. (a) Let $E = \{1, 2, \dots, d\}$. Let $\{Y(k), k = 0, 1, 2, \dots\}$ be a Markov chain in E with transition matrix $P = ((p_{ij}))$, that is, $P\{Y(k+1) = j | Y(k) = i\} = p_{ij}$. Suppose P is irreducible and aperiodic, that is, P^k has all elements positive for k sufficiently large. Let $\mathcal{F}_n = \sigma\{Y(k): k \leq n\}$ and $\mathcal{F}^n = \sigma\{Y(k): k \geq n\}$, and define $\varphi(m) = \sup_n \varphi_\infty(\mathcal{F}^{n+m} | \mathcal{F}_n)$. Show that $\lim_{m \rightarrow \infty} m^{-1} \log \varphi(m) = \alpha < 0$ exists, and characterize α in terms of P .
- (b) Let X be an Ornstein–Uhlenbeck process with generator $Af = \frac{1}{2}af'' - bxf'$, $a, b > 0$. Suppose $PX(0)^{-1} = v$ is the stationary distribution for X . Compute $\psi_{t,s} = \psi_{0,s}$ given by (2.20).
- (c) Let X be a reflecting Brownian motion on $[0, 1]$ with generator $Af = \frac{1}{2}\sigma^2 f''$. Suppose $PX(0)^{-1} = m$ (Lebesgue measure) and compute $\psi_{t,s} = \psi_{0,s}$ given by (2.20).

11. For $n = 1, 2, \dots$, let $\{\xi_k^n, k = 1, 2, \dots\}$ be a stationary sequence of random variables with $P\{\xi_k^n = 1\} = p_n$ and $P\{\xi_k^n = 0\} = 1 - p_n$. Let $\mathcal{F}_k^n = \sigma(\xi_i^n: i \leq k)$, $\mathcal{G}_k^n = \sigma(\xi_i^n: i \geq k)$, and define $\varphi_p^n(m) = \sup_k \varphi_p(\mathcal{G}_{k+m}^n | \mathcal{F}_k^n)$. Suppose $np_n \rightarrow \lambda > 0$ and $\max_{k \leq n} P\{\xi_{k+1}^n = 1 | \mathcal{F}_k^n\} \rightarrow 0$ as $n \rightarrow \infty$, and define

$$(6.13) \quad N_n(t) = \sum_{k=1}^{\lfloor nt \rfloor} \xi_k^n.$$

Give conditions on $\varphi_p^n(m)$ that imply $N_n \Rightarrow N$, where N is a Poisson process with parameter λ .

12. Let $\{Y_k, k \in \mathbb{Z}\}$ be stationary and define $\mathcal{F}_n^0 = \sigma(Y_k: 1 \leq k \leq n)$ and $\mathcal{F}^n = \sigma(Y_k: k \geq n)$. Show that for each m , $\varphi_p(\mathcal{F}^{n+m} | \mathcal{F}_n^0)$ is a nondecreasing function of n and $\varphi_p(m) = \lim_{n \rightarrow \infty} \varphi_p(\mathcal{F}^{n+m} | \mathcal{F}_n^0)$, where $\varphi_p(m)$ is given by (3.1).

13. For $n = 1, 2, \dots$, let $\mu_n(x, \Gamma)$ and $v_n(x, \Gamma)$ be transition functions on $\mathbb{R}^d \times \mathcal{B}(\mathbb{R}^d)$. Suppose $\mu_n(x, \Gamma \cap B(x, 1)) = v_n(x, \Gamma \cap B(x, 1))$, $x \in \mathbb{R}^d$, $\Gamma \in \mathcal{B}(\mathbb{R}^d)$, and $\lim_{n \rightarrow \infty} \sup_x n\mu_n(x, B(x, 1)^c) = 0$. Show that for each $v \in \mathcal{P}(\mathbb{R}^d)$ there exist Markov chains $\{Y_n(k), k = 0, 1, 2, \dots\}$ and $\{Z_n(k), k = 0, 1, 2, \dots\}$ with $PY_n(0)^{-1} = PX_n(0)^{-1} = v$, such that Y_n corresponds to $\mu_n(x, \Gamma)$ and Z_n corresponds to $v_n(x, \Gamma)$, and for each $K > 0$, $\lim_{n \rightarrow \infty} P\{Y_n(k) \neq Z_n(k) \text{ for some } k \leq nK\} = 0$.
14. For $n = 1, 2, \dots$ let Y_n be a Markov chain in $E_n = \{k/n : k = 0, 1, \dots, n\}$ with transition function given by

$$(6.14) \quad P \left\{ Y_n(k+1) = \frac{j}{n} \mid Y_n(k) = x \right\} = \binom{n}{j} x^j (1-x)^{n-j}.$$

Apply Corollary 4.2 to obtain a diffusion approximation for $Y_n([nt])$ (cf. Chapter 10).

15. Let $E = \{0, 1, 2, \dots\}$, and let Z_n be a continuous-time branching process with generator

$$(6.15) \quad Af(k) = \lambda k \sum_{l=1}^{\infty} p_l (f(k+l-1) - f(k))$$

for $f \in C_c(E)$, where $p_l \geq 0$, $l = 0, 1, 2, \dots$, $\lambda > 0$, and $\sum_{l=1}^{\infty} p_l = 1$. Define $X_n(t) = Z_n(nt)/n$, and assume $PX_n(0)^{-1} \Rightarrow v \in \mathcal{P}([0, \infty))$. Suppose $\sum_{l=0}^{\infty} lp_l = 1$ and $\sum_{l=0}^{\infty} l^2 p_l < \infty$. Apply Theorem 4.1 to obtain a diffusion approximation for X_n (cf. Chapter 9).

16. Let N_1 and N_2 be independent Poisson processes and let $F, G \in C^1(\mathbb{R})$. Apply Theorem 4.1 to obtain a diffusion approximation for X_n satisfying

$$(6.16) \quad \dot{X}_n(t) = (-1)^{N_1(n^2t)} nF(X_n(t)) + (-1)^{N_2(n^2t)} nG(X_n(t))$$

(cf. Chapter 12).

Hint: Find the analogue of the martingale defined in (6.4).

17. Let X be a process with stationary independent increments satisfying $E[X(t)] = 0$ for $t \geq 0$, and suppose there exists $\alpha_0 > 0$ such that $e^{\psi(\alpha)} \equiv E[e^{\alpha X(1)}] < \infty$ for $|\alpha| \leq \alpha_0$. Show that $\exp\{\alpha X(t) - t\psi(\alpha)\}$ is a martingale for each α , $|\alpha| \leq \alpha_0$, and that for $0 < \alpha \leq \alpha_0$

$$(6.17) \quad P \left\{ \sup_{s \leq t} |X(s)| \geq x \right\} \leq [\exp\{t\psi(\alpha)\} + \exp\{t\psi(-\alpha)\}] \exp\{-\alpha x\}.$$

7. NOTES

The invariance principle for independent random variables was given by Donsker (1951). For discrete time, the fact that the conditions of Theorem 1.4(b) with A_n given by (1.22) imply asymptotic normality was observed by Lévy (see Doob (1953), page 383) and developed by Dvoretzky (1972). Brown (1971) gave the corresponding invariance principle. McLeish (1974) gave the discrete-time version of Theorem 1.4(a). Various authors have extended and refined these results, for example Rootzen (1977, 1980) and Gänssler and Häusler (1979). Rebolledo (1980) extended the results to continuous time. The version we have presented is not quite the most general. See also Hall and Heyde (1980) and the survey article by Helland (1982).

Uniform mixing was introduced by Ibragimov (1959) and strong mixing by Rosenblatt (1956). For $p = r = 1$, (2.8) is due to Volkonskii and Rozanov (1959), and (2.14) is due to Davydov (1968). For $\Phi = \varphi_\infty(\mathcal{G}|\mathcal{H})$, (2.26) appears in Ibragimov (1962). A variety of other mixing conditions are discussed in Withers (1981) and Peligrad (1982).

A vast literature exists on central limit theorems and invariance principles under mixing conditions. See Hall and Heyde (1980), Chapter 5, for a recent survey and Ibragimov and Linnik (1971). Central limit theorems under the hypotheses of Theorems 3.1 and 3.3 (assuming (3.19)) were given by Ibragimov (1962). Weak convergence assuming (3.19) was established by Billingsley (1968). The proof given here is due to Heyde (1974).

Theorem 4.1, in the form given here, is due to Rebolledo (1979). Corollary 4.2 is due to Stroock and Varadhan (1969). See Stroock and Varadhan (1979), page 266. Skorohod (1965), Borovkov (1970), and Kushner (1974) give other approaches to diffusion approximations.

Theorem 5.1 is due to Komlós, Major, and Tusnády (1975, 1976). See also Major (1976) and Csörgő and Révész (1981). Theorem 5.6 is due to Bártfai (1966).

The characterization of the Poisson process given in Problem 1(a) is due to Watanabe (1964). Various authors have given results along the lines of Problem 1(b), Brown (1978), Kabanov, Lipster, and Shiryaev (1980), Grigelionis and Mikuličios (1981), and Kurtz (1982).

The example in Problem 2 is due to Hardin (1985).

There is also a vast literature on central limit theorems and related invariance principles for Markov processes (Problems 3, 4, and 6). The martingale approach to these results has been taken by Maigret (1978), Bhattacharya (1982), and Kurtz (1981b).

8

EXAMPLES OF GENERATORS

The purpose of this chapter is to list conditions under which certain linear operators, corresponding (at least intuitively) to specific Markov processes, generate Feller semigroups or have the property that the associated martingale problem is well-posed. In contrast to other chapters, here we reference other sources wherever possible.

Section 1 contains results for nondegenerate diffusions. These include classical, one-dimensional diffusions with local boundary conditions, diffusions in bounded regions with absorption or oblique reflection at the boundary, diffusions in \mathbb{R}^d with Hölder continuous coefficients, and diffusions in \mathbb{R}^d with continuous, time-dependent coefficients.

Section 2 concerns degenerate diffusions. Results are given for one-dimensional diffusions with smooth coefficients and with Lipschitz continuous, time-dependent coefficients, diffusions in \mathbb{R}^d with smooth diffusion matrix and with diffusion matrix having a Lipschitz continuous, time-dependent square root, and a class of diffusions in a subset of \mathbb{R}^d occurring in population genetics.

In Section 3, other processes are considered. Results are included for jump processes with unbounded generators, processes with Lévy generators including independent-increment processes, and two classes of infinite particle systems, namely, spin-flip systems and exclusion processes.

1. NONDEGENERATE DIFFUSIONS

We begin with the one-dimensional case, where one can explicitly characterize the generator of the semigroup. Given $-\infty \leq r_0 < r_1 \leq \infty$, let I be the closed interval in \mathbb{R} with endpoints r_0 and r_1 , I° its interior, and \bar{I} its closure in $[-\infty, \infty]$. In other words,

$$(1.1) \quad I = [r_0, r_1] \cap \mathbb{R}, \quad I^\circ = (r_0, r_1), \quad \bar{I} = [r_0, r_1].$$

We identify $C(\bar{I})$ with the space of $f \in C(I^\circ)$ for which $\lim_{x \rightarrow r_i} f(x)$ exists and is finite for $i = 0, 1$. Suppose $a, b \in C(I^\circ)$ and $a > 0$ on I° . Then there is at least one restriction of

$$(1.2) \quad G = \frac{1}{2}a(x) \frac{d^2}{dx^2} + b(x) \frac{d}{dx}$$

acting on $\{f \in C(\bar{I}) \cap C^2(I^\circ) : Gf \in C(\bar{I})\}$ that generates a Feller semigroup on $C(\bar{I})$.

To specify the appropriate restrictions, we need to introduce Feller's boundary classification. Fix $r \in (r_0, r_1)$ and define $B, m, p \in C(I^\circ)$ by

$$(1.3) \quad B(x) = \int_r^x 2a(y)^{-1}b(y) dy,$$

$$(1.4) \quad p(x) = \int_r^x e^{-B(y)} dy, \quad m(x) = \int_r^x 2a(y)^{-1}e^{B(y)} dy,$$

so that

$$(1.5) \quad G = \frac{1}{2}a(x)e^{-B(x)} \frac{d}{dx} \left(e^{B(x)} \frac{d}{dx} \right) = \frac{d}{dm(x)} \left(\frac{d}{dp(x)} \right).$$

Define u and v on \bar{I} by

$$(1.6) \quad u(x) = \int_r^x m dp, \quad v(x) = \int_r^x p dm.$$

Then, for $i = 0, 1$, the boundary r_i is said to be

$$(1.7) \quad \begin{aligned} &\text{regular} && \text{if } u(r_i) < \infty \text{ and } v(r_i) < \infty, \\ &\text{exit} && \text{if } u(r_i) < \infty \text{ and } v(r_i) = \infty, \\ &\text{entrance} && \text{if } u(r_i) = \infty \text{ and } v(r_i) < \infty, \\ &\text{natural} && \text{if } u(r_i) = \infty \text{ and } v(r_i) = \infty. \end{aligned}$$

Regular and exit boundaries are said to be *accessible*; entrance and natural boundaries are termed *inaccessible*. See Problem 1 for some motivation for this terminology.

Let

$$(1.8) \quad \mathcal{D} = \{f \in C(\bar{I}) \cap C^2(I^\circ) : Gf \in C(\bar{I})\},$$

and for $i = 0, 1$, define

$$(1.9) \quad \mathcal{D}_i = \mathcal{D}, \quad r_i \text{ inaccessible},$$

$$(1.10) \quad \mathcal{D}_i = \left\{ f \in \mathcal{D} : \lim_{x \rightarrow r_i} Gf(x) = 0 \right\}, \quad r_i \text{ exit},$$

and

$$(1.11) \quad \mathcal{D}_i = \left\{ f \in \mathcal{D} : q_i \lim_{x \rightarrow r_i} Gf(x) = (-1)^i (1 - q_i) \lim_{x \rightarrow r_i} e^{B(x)} f'(x) \right\},$$

r_i regular,

where $q_i \in [0, 1]$ and B is given by (1.3).

1.1 Theorem Given $-\infty \leq r_0 < r_1 \leq \infty$, define I , I° , and \bar{I} by (1.1). Suppose $a, b \in C(I^\circ)$ with $a > 0$ on I° , and define G , \mathcal{D}_0 , and \mathcal{D}_1 by (1.2) and (1.8)–(1.11), where $q_i \in [0, 1]$ is fixed if r_i is regular, $i = 0, 1$. Then $\{(f, Gf) : f \in \mathcal{D}_0 \cap \mathcal{D}_1\}$ generates a Feller semigroup on $C(\bar{I})$.

Proof. See Mandl (1968) (except for the exit case). \square

1.2 Corollary Suppose, in addition to the hypotheses of Theorem 1.1, that infinite boundaries of I are natural. Then $\{(f, Gf) : f \in \hat{C}(I) \cap \mathcal{D}_0 \cap \mathcal{D}_1, Gf \in \hat{C}(I)\}$ generates a Feller semigroup on $\hat{C}(I)$.

Proof. See Problem 1. \square

1.3 Remark We note a simple, sufficient (but not necessary) condition for the extra hypothesis in Corollary 1.2. If there exists a constant K such that

$$(1.12) \quad a(x) \leq K(1 + x^2), |b(x)| \leq K(1 + |x|), \quad x \in I^\circ,$$

then infinite boundaries of I are natural. The proof is left to the reader (Problem 2). \square

For some applications, it is useful to find a core (such as $C_c^\infty(I)$) for the generator in Corollary 1.2. In Section 2 we do this under certain assumptions on the coefficients a and b .

We turn next to the case of a bounded, connected, open set $\Omega \subset \mathbb{R}^d$, where $d \geq 2$. Before stating any results, we need to introduce a certain amount of notation and terminology.

Let $0 < \mu \leq 1$. A function $f \in C(\bar{\Omega})$ is said to satisfy a Hölder condition with exponent μ and Hölder constant M if

$$(1.13) \quad \sup_{y \in V} \rho^{-\mu} \left\{ \sup_{y \in V} f(y) - \inf_{y \in V} f(y) \right\} = M,$$

where the supremum is over $0 < \rho \leq \rho_0$ (ρ_0 fixed), $x \in \mathbb{R}^d$, and components V of $\Omega \cap B(x, \rho)$. We denote M by $|f|_\mu$.

For $m = 0, 1, \dots$, we define

$$(1.14) \quad C^{m,\mu}(\bar{\Omega}) = \{f \in C^m(\Omega) : |D^\alpha f|_\mu < \infty \text{ whenever } |\alpha| = m\},$$

where $C^0(\Omega) = C(\bar{\Omega})$, $\alpha \in (\mathbb{Z}_+)^d$, $D^\alpha = \partial_1^{\alpha_1} \cdots \partial_d^{\alpha_d}$, and $|\alpha| = \alpha_1 + \cdots + \alpha_d$. Observe that functions in $C^{0,\mu}(\bar{\Omega})$ need not have continuous extensions to $\bar{\Omega}$, as such extensions might have to be multivalued on $\partial\Omega$.

Regarding elements of \mathbb{R}^d as column vectors, a mapping of \mathbb{R}^d onto \mathbb{R}^d of the form $y = U(x - x_0)$, where $x_0 \in \partial\Omega$ and U is an orthogonal matrix ($UU^T = I_d$), is said to be a local Cartesian coordinate system with origin at x_0 if the outer normal to $\partial\Omega$ at x_0 is mapped onto the nonnegative y_d axis. For $m = 1, 2, \dots$, we say that $\partial\Omega$ is of class $C^{m,\mu}$ if there exists $\rho > 0$ such that for every $x_0 \in \partial\Omega$, $\partial\Omega \cap B(x_0, \rho)$ is a connected surface that, in terms of the local Cartesian coordinate system (y_1, \dots, y_d) with origin at x_0 , is of the form $y_d = v(y_1, \dots, y_{d-1})$, where $v \in C^{m,\mu}(\bar{D})$, D being the projection of $\partial\Omega \cap B(x_0, \rho)$ (in the local Cartesian coordinate system) onto $y_d = 0$. Assuming $\partial\Omega$ is of class $C^{m,\mu}$, a function $\varphi: \partial\Omega \rightarrow \mathbb{R}$ is said to belong to $C^{m,\mu}(\partial\Omega)$ if, for every $x_0 \in \partial\Omega$, φ as a function of (y_1, \dots, y_{d-1}) belongs to $C^{m,\mu}(\bar{D})$, where the notation is as above. Note that if $\partial\Omega$ is of class $C^{m,\mu}$ and if $0 \leq k \leq m$, then each function in $C^{k,\mu}(\bar{\Omega})$ has a (unique) continuous extension to $\bar{\Omega}$, and its restriction to $\partial\Omega$ belongs to $C^{k,\mu}(\partial\Omega)$.

We consider

$$(1.15) \quad G = \frac{1}{2} \sum_{i,j=1}^d a_{ij}(x) \partial_i \partial_j + \sum_{i=1}^d b_i(x) \partial_i,$$

treating separately the cases of absorption and oblique reflection at the boundary. S_d denotes the space of $d \times d$ nonnegative-definite matrices.

1.4 Theorem Let $d \geq 2$ and $0 < \mu \leq 1$, and let $\Omega \subset \mathbb{R}^d$ be bounded, connected, and open with $\partial\Omega$ of class $C^{2,\mu}$. Suppose $a: \Omega \rightarrow S_d$, $b: \Omega \rightarrow \mathbb{R}^d$, $a_{ij}, b_i \in C^{0,\mu}(\bar{\Omega})$ for $i, j = 1, \dots, d$, and

$$(1.16) \quad \inf_{x \in \Omega} \inf_{|\theta|=1} \theta \cdot a(x)\theta > 0.$$

Then, with G defined by (1.15), the closure of

$$(1.17) \quad A \equiv \{(f, Gf) : f \in C^{2,\mu}(\bar{\Omega}), Gf = 0 \text{ on } \partial\Omega\}$$

is single-valued and generates a Feller semigroup on $C(\bar{\Omega})$.

Proof. We apply Theorem 2.2 of Chapter 4. Clearly, A satisfies the positive maximum principle. If $\lambda > 0$ and $g \in C^{2,\mu}(\bar{\Omega})$, then, by Theorem 3.13 of Ladyzhenskaya and Ural'tseva (1968), the equation $\lambda f - Gf = g$ has a solution $f \in C^{2,\mu}(\bar{\Omega})$ with $f = \lambda^{-1}g$ on $\partial\Omega$. It follows that $Gf = 0$ on $\partial\Omega$, so $f \in \mathcal{D}(A)$, proving that $\mathcal{R}(\lambda - A) \supset C^{2,\mu}(\bar{\Omega})$ for every $\lambda > 0$.

To show that $\mathcal{D}(A)$ is dense in $C(\bar{\Omega})$, let $f \in C^{2,\mu}(\bar{\Omega})$. For each $\lambda > 0$, choose $h_\lambda \in \mathcal{D}(A)$ such that $(\lambda - A)h_\lambda = \lambda f - Gf$. Then, as $\lambda \rightarrow \infty$,

$$(1.18) \quad \|f - h_\lambda\| = \sup_{x \in \partial\Omega} |f(x) - h_\lambda(x)| = \sup_{x \in \partial\Omega} \lambda^{-1} |Gf(x)| \rightarrow 0,$$

where the first equality is due to the weak maximum principle for elliptic operators (Friedman (1975), Theorem 6.2.2). \square

1.5 Theorem Suppose, in addition to the hypotheses of Theorem 1.4, that $c: \partial\Omega \rightarrow \mathbb{R}^d$, $c_i \in C^{1,\mu}(\partial\Omega)$ for $i = 1, \dots, d$, and

$$(1.19) \quad \inf_{x \in \partial\Omega} c(x) \cdot n(x) > 0,$$

where $n(x)$ denotes the outward unit normal to $\partial\Omega$ at x . Then, with G defined by (1.15), the closure of

$$(1.20) \quad A \equiv \{(f, Gf): f \in C^{2,\mu}(\bar{\Omega}), c \cdot \nabla f = 0 \text{ on } \partial\Omega\}$$

is single-valued and generates a Feller semigroup on $C(\bar{\Omega})$.

Proof. Again, we apply Theorem 2.2 of Chapter 4. Because $c(x) \cdot n(x) \neq 0$ for all $x \in \partial\Omega$, A satisfies the positive maximum principle. (If $f \in \mathcal{D}(A)$ has a positive maximum at $x \in \partial\Omega$, then $\nabla f(x) = 0$.) By Theorem 3.3.2 of Ladyzhenskaya and Ural'tseva (1968), there exists $\lambda > 0$ such that for every $g \in C^{2,\mu}(\bar{\Omega})$ the equation $\lambda f - Gf = g$ has a solution $f \in C^{2,\mu}(\bar{\Omega})$ with $c \cdot \nabla f = 0$ on $\partial\Omega$. Thus $\mathcal{R}(\lambda - A) \supset C^{2,\mu}(\bar{\Omega})$.

It remains to show that $\mathcal{D}(A)$ is dense in $C(\bar{\Omega})$, or equivalently (by Remark 3.2 of Chapter 1), that the bp-closure of $\mathcal{D}(A)$ contains $C(\bar{\Omega})$. By Stroock and Varadhan (1971) there exists for each $x \in \bar{\Omega}$ a solution X_x of the $C_0[0, \infty)$ martingale problem for (A, δ_x) . Consequently, for each $f \in C(\bar{\Omega})$,

$$(1.21) \quad f(x) = \text{bp-lim}_{n \rightarrow \infty} n \int_0^\infty e^{-nt} E[f(X_x(t))] dt.$$

Since the right side of (1.21) belongs to the bp-closure of $\mathcal{D}(A)$, the proof is complete. \square

Let us now consider the case in which $\Omega = \mathbb{R}^d$.

1.6 Theorem Let $a: \mathbb{R}^d \rightarrow S_d$ and $b: \mathbb{R}^d \rightarrow \mathbb{R}^d$ be bounded, $0 < \mu \leq 1$, and $K > 0$, and suppose that

$$(1.22) \quad |a(x) - a(y)| + |b(x) - b(y)| \leq K|x - y|^\mu, \quad x, y \in \mathbb{R}^d,$$

and

$$(1.23) \quad \inf_{x \in \mathbb{R}^d} \inf_{|\theta|=1} \theta \cdot a(x)\theta > 0.$$

Then, with G defined by (1.15), the closure of $\{(f, Gf): f \in C_c^\infty(\mathbb{R}^d)\}$ is single-valued and generates a Feller semigroup on $\hat{C}(\mathbb{R}^d)$.

Proof. According to Theorem 5.11 of Dynkin (1965), there exists a strongly continuous, positive, contraction semigroup $\{T(t)\}$ on $\hat{C}(\mathbb{R}^d)$ whose generator A extends $G|_{C_c^2(\mathbb{R}^d)}$ (and therefore $G|_{\hat{C}^2(\mathbb{R}^d)}$). It suffices to show that A is conservative and that $C_c^\infty(\mathbb{R}^d)$ is a core for A .

Given $f \in \hat{C}(\mathbb{R}^d)$ and $t > 0$, the estimate in part 2° of the proof of Theorem 5.11 of Dynkin (1965), with (0.41) and (0.42) in place of (0.40), implies that $\partial_i T(t)f$ and $\partial_i \partial_j T(t)f$ exist and belong to $\hat{C}(\mathbb{R}^d)$. Thus, $T(t): \hat{C}^2(\mathbb{R}^d) \rightarrow \hat{C}^2(\mathbb{R}^d)$ for all $t \geq 0$, so $\hat{C}^2(\mathbb{R}^d)$ is a core for A by Proposition 3.3 of Chapter 1. Let $h \in C_c^2(\mathbb{R}^d)$ satisfy $\chi_{B(0, 1)} \leq h \leq \chi_{B(0, 2)}$, and approximate $f \in C^2(\mathbb{R}^d)$ by $\{fh_n\} \subset C_c^2(\mathbb{R}^d)$, where $h_n(x) = h(x/n)$, to show that $C_c^2(\mathbb{R}^d)$ is a core for A . Similarly, using the sequence $\{(h_n, Gh_n)\}$, we find that A is conservative. Finally, choose $\varphi \in C_c^\infty(\mathbb{R}^d)$ with $\varphi \geq 0$ and $\int \varphi(x) dx = 1$, and approximate $f \in C_c^2(\mathbb{R}^d)$ by $\{f * \varphi_n\} \subset C_c^\infty(\mathbb{R}^d)$, where $\varphi_n(x) = n^d \varphi(nx)$, to show that $C_c^\infty(\mathbb{R}^d)$ is a core for A . (Note that $f * \varphi_n$ has compact support because both f and φ_n do.) \square

If one is satisfied with uniqueness of solutions of the martingale problem, the assumptions of Theorem 1.6 can be weakened considerably. Moreover, time-dependent coefficients are permitted. Consider therefore

$$(1.24) \quad G = \frac{1}{2} \sum_{i, j=1}^d a_{ij}(t, x) \partial_i \partial_j + \sum_{i=1}^d b_i(t, x) \partial_i.$$

1.7 Theorem Let $a: [0, \infty) \times \mathbb{R}^d \rightarrow S_d$ and $b: [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ be locally bounded and Borel measurable. Suppose that, for every $x \in \mathbb{R}^d$ and $t_0 > 0$,

$$(1.25) \quad \inf_{0 \leq t \leq t_0} \inf_{|\theta|=1} \theta \cdot a(t, x)\theta > 0$$

and

$$(1.26) \quad \lim_{y \rightarrow x} \sup_{0 \leq t \leq t_0} |a(t, y) - a(t, x)| = 0.$$

Suppose further that there exists a constant K such that

$$(1.27) \quad |a(t, x)| \leq K(1 + |x|^2), \quad t \geq 0, x \in \mathbb{R}^d,$$

and

$$(1.28) \quad x \cdot b(t, x) \leq K(1 + |x|^2), \quad t \geq 0, x \in \mathbb{R}^d.$$

Then, with G defined by (1.24), the martingale problem for $\{(f, Gf): f \in C_c^\infty(\mathbb{R}^d)\}$ is well-posed.

Proof. Recalling Proposition 3.5 of Chapter 5, the result follows from Theorem 10.2.2 of Stroock and Varadhan (1979) and the discussion following their Corollary 10.1.2. \square

2. DEGENERATE DIFFUSIONS

Again we treat the one-dimensional case separately. We begin by obtaining sufficient conditions for $C_c^\infty(I)$ to be a core for the generator in Corollary 1.2.

2.1 Theorem Given $-\infty \leq r_0 < r_1 \leq \infty$, define I and I° by (1.1). Suppose $a \in C^2(I)$, $a \geq 0$, a'' is bounded, $b: I \rightarrow \mathbb{R}$ is Lipschitz continuous (that is, $\sup_{x, y \in I, x \neq y} |b(y) - b(x)|/|x - y| < \infty$), and

$$(2.1) \quad a(r_i) = 0 \leq (-1)^i b(r_i) \text{ if } |r_i| < \infty, \quad i = 0, 1.$$

Then with G defined by (1.2), the closure of $\{(f, Gf): f \in C_c^\infty(I)\}$ is single-valued and generates a Feller semigroup $\{T(t)\}$ on $\hat{C}(I)$. If $a > 0$ on I° , then $\{T(t)\}$ coincides with the semigroup defined in terms of a and b by Corollary 1.2 (with $q_i = 0$ if r_i is regular, $i = 0, 1$).

The proof depends on a lemma, which involves the function space $\hat{C}_{-,\gamma}^m(I)$, defined for $m = 0, 1, \dots$ and $\gamma \geq 0$ by

$$(2.2) \quad \hat{C}_{-,\gamma}^m(I) = \{f \in C^m(I): \varphi_\gamma f^{(k)} \in \hat{C}(I), k = 0, \dots, m\},$$

where $\varphi_\gamma(x) = (1 + x^2)^{\gamma/2}$. Note that $\hat{C}_0^m(I) = \hat{C}^m(I)$.

2.2 Lemma Assume, in addition to the hypotheses of Theorem 2.1, that $b \in C^2(I)$ and b'' is bounded. Then there exists a (unique) strongly continuous, positive, contraction semigroup $\{T(t)\}$ on $\hat{C}(I)$ whose generator A is an extension of $\{(f, Gf): f \in \hat{C}(I) \cap C^2(I), Gf \in \hat{C}(I)\}$; moreover:

- (a) $T(t): \hat{C}_{-,\gamma}^m(I) \rightarrow \hat{C}_{-,\gamma}^m(I)$ for all $t \geq 0$, $m = 1, 2$, and $\gamma \geq 0$.
- (b) $\|(T(t)f)'\| \leq \exp(\|b'\|t)\|f'\|$ for all $f \in \hat{C}^1(I)$ and $t \geq 0$.

Proof. This is a special case of a result of Ethier (1978). \square

Proof of Theorem 2.1. $\hat{C}_{-2}^2(I) \subset \hat{C}(I) \cap C^2(I) \cap A^{-1}\hat{C}(I)$, so under the additional assumptions of Lemma 2.2, $\hat{C}_{-2}^2(I)$ is a core for A by Proposition 3.3 of

Chapter 1. To obtain this conclusion without the additional assumptions, choose $\{b_n\} \subset C^2(I)$ such that each b_n satisfies the conditions of Lemma 2.2, $\lim_{n \rightarrow \infty} \|b_n - b\| = 0$, and $\sup_n \|b'_n\| \leq M < \infty$. This can be done via convolutions. For each n , let $\{T_n(t)\}$ be associated with a and b_n as in Lemma 2.2. We apply Proposition 3.7 of Chapter 1 with $D_0 = \mathcal{D}(A) = \bar{C}_{-2}(I)$, $D_1 = \bar{C}^1(I)$, $\|f\| = \|f\| + \|f'\|$, $\omega = M$, and $\varepsilon_n = \|b_n - b\|$, concluding that $\bar{C}_{-2}(I)$ is a core for A under the assumptions of the theorem. The remainder of the proof that $C_c^\infty(I)$ is a core for A (and the proof that A is conservative) is analogous to that of Theorem 1.6. For the proof of the second conclusion of the theorem, see Ethier (1978). \square

Of course, one can get uniqueness of solutions of the martingale problem under conditions that are much weaker than those of Theorem 2.1. One of the assumptions in Theorem 2.1 that is often too restrictive in applications is the requirement (when $b \in C^1(I)$) that b' be bounded, because, in the context of Theorem 1.1, infinite boundaries are often entrance. We permit time-dependent coefficients and so we consider

$$(2.3) \quad G = \frac{1}{2}a(t, x) \frac{d^2}{dx^2} + b(t, x) \frac{d}{dx}.$$

2.3 Theorem Given $-\infty \leq r_0 < r_1 \leq \infty$, define I by (1.1), and let a and b be real-valued, locally bounded, and Borel measurable on $[0, \infty) \times I$ with $a \geq 0$. Suppose that, for each $n \geq 1$ and $t_0 > 0$, a and b are Lipschitz continuous in $|x| \leq n$, uniformly in $0 \leq t \leq t_0$. Suppose further that there exists a constant K such that (1.27) and (1.28) hold with \mathbb{R}^d replaced by I , and

$$(2.4) \quad a(t, r_i) = 0 \leq (-1)^i b(t, r_i) \quad \text{if } |r_i| < \infty \quad t \geq 0, i = 0, 1.$$

Then, with G defined by (2.3), the martingale problem for $\{(f, Gf) : f \in C_c^\infty(I)\}$ is well-posed.

Proof. Existence of solutions follows from Theorem 3.10 of Chapter 5, together with (2.4). (Extend a to be zero outside of I and b by setting $b(t, x) = b(t, r_0)$, $x < r_0$, and $b(t, x) = b(t, r_1)$, $x > r_1$.) Uniqueness is a consequence of Theorem 3.8, Proposition 3.5, and Corollary 3.4, all from Chapter 5. \square

Unfortunately, the extent to which Theorem 2.1 can be generalized to d dimensions is unknown. However, results can be obtained in a few special cases.

2.4 Proposition Let $\sigma: \mathbb{R}^d \rightarrow \mathbb{R}^d \otimes \mathbb{R}^d$ and $b: \mathbb{R}^d \rightarrow \mathbb{R}^d$ satisfy $\sigma_{ij}, b_i \in \bar{C}^2(\mathbb{R}^d)$ for $i, j = 1, \dots, d$, and put $a = \sigma\sigma^T$. Then, with G defined by (1.15), the closure of $\{(f, Gf) : f \in C_c^\infty(\mathbb{R}^d)\}$ is single-valued and generates a Feller semigroup on $\bar{C}(\mathbb{R}^d)$.

Proof. A proof has been outlined by Roth (1977). The details are left to the reader (Problem 4). \square

The following result generalizes Proposition 2.4.

2.5 Theorem Let $a: \mathbb{R}^d \rightarrow S_d$ satisfy $a_{ij} \in C^2(\mathbb{R}^d)$ with $\partial_k \partial_l a_{ij}$ bounded for $i, j, k, l = 1, \dots, d$, and let $b: \mathbb{R}^d \rightarrow \mathbb{R}^d$ be Lipschitz continuous (i.e., $\sup_{x, y \in \mathbb{R}^d, x \neq y} |b(x) - b(y)|/|x - y| < \infty$). Then, with G defined by (1.15), the closure of $\{(f, Gf): f \in C_c^\infty(\mathbb{R}^d)\}$ is single-valued and generates a Feller semigroup on $\hat{C}(\mathbb{R}^d)$.

Proof. We simply outline the proof, leaving it to the reader to fill in a number of details. First, some additional notation is needed. For $\gamma \geq 0$ define $\varphi_\gamma: \mathbb{R}^d \rightarrow (0, \infty)$ by $\varphi_\gamma(x) = (1 + |x|^2)^{\gamma/2}$ and

$$(2.5) \quad \hat{C}_{-\gamma}(\mathbb{R}^d) = \{f \in C(\mathbb{R}^d): \varphi_\gamma f \in \hat{C}(\mathbb{R}^d)\}.$$

For $m = 1, 2, \dots$ and $\gamma \geq 0$ define

$$(2.6) \quad \hat{C}_{-\gamma}^m(\mathbb{R}^d) = \{f \in C^m(\mathbb{R}^d): \varphi_\gamma D^\alpha f \in \hat{C}(\mathbb{R}^d) \text{ if } |\alpha| \leq m\}.$$

A useful norm on $\hat{C}_{-\gamma}^m(\mathbb{R}^d)$ is given by

$$(2.7) \quad \|f\|_{\hat{C}_{-\gamma}^m(\mathbb{R}^d)} = \left\| \varphi_\gamma \left\{ \sum_{1 \leq |\alpha| \leq m} (D^\alpha f)^2 \right\}^{1/2} \right\|.$$

Finally, we define

$$(2.8) \quad \hat{C}_{-\gamma}^{0, m}([0, T] \times \mathbb{R}^d) = \{f \in C^{0, m}([0, T] \times \mathbb{R}^d): \varphi_\gamma D^\alpha f \in \hat{C}([0, T] \times \mathbb{R}^d) \text{ if } |\alpha| \leq m\}.$$

Suppose, in addition to the hypotheses of the theorem, that $b_i \in C^2(\mathbb{R}^d)$ and $\partial_k b_i, \partial_k \partial_l b_i$ are bounded for $i, k, l = 1, \dots, d$. Then there exist sequences $\sigma^{(n)}: \mathbb{R}^d \rightarrow \mathbb{R}^d \otimes \mathbb{R}^d$ and $b^{(n)}: \mathbb{R}^d \rightarrow \mathbb{R}^d$ with the following properties, where $a^{(n)} = \sigma^{(n)}(\sigma^{(n)})^T: \sigma_{ij}^{(n)}, b_i^{(n)} \in \hat{C}^\infty(\mathbb{R}^d)$, $a_{ij}^{(n)} \rightarrow a_{ij}$ and $b_i^{(n)} \rightarrow b_i$ uniformly on compact sets, and $a_{ij}^{(n)}/\varphi_2, b_i^{(n)}/\varphi_1, \partial_k \partial_l a_{ij}^{(n)}, \partial_k b_i^{(n)}, \partial_k \partial_l b_i^{(n)}$ are uniformly bounded, $i, j, k, l = 1, \dots, d$.

Fix n for now. Letting G_n be defined in terms of $a^{(n)}$ and $b^{(n)}$ as in (1.15), one can show as in Problem 5(b) that the closure of $\{(f, G_n f): f \in C_c^\infty(\mathbb{R}^d)\}$ is single-valued and generates a Feller semigroup $\{T_n(t)\}$ on $\hat{C}(\mathbb{R}^d)$. (This also follows from Proposition 2.4.) Moreover, a slight extension of this argument shows that for each $m \geq 1$, $T_n(t): \hat{C}^m(\mathbb{R}^d) \rightarrow \hat{C}^m(\mathbb{R}^d)$ for all $t \geq 0$ and $\{T_n(t)\}$ restricted to $\hat{C}^m(\mathbb{R}^d)$ is strongly continuous in the norm $\|\cdot\|_{\hat{C}_c^m(\mathbb{R}^d)}$ (recall (2.7)). A simple argument involving Gronwall's inequality implies that for each $\gamma > 0$, there exists $\lambda_{n, \gamma} > 0$, such that $T_n(t)\varphi_{-\gamma} \leq \exp(\lambda_{n, \gamma} t)\varphi_{-\gamma}$ for all $t \geq 0$. Using the fact that $\hat{C}^3(\mathbb{R}^d) \cap \hat{C}_{-\gamma}(\mathbb{R}^d) \subset \hat{C}_{-\gamma}^2(\mathbb{R}^d)$ for γ sufficiently large, we conclude that if $f \in C_c^\infty(\mathbb{R}^d)$, $t > 0$, and $u_n(s, x) \equiv T_n(t-s)f(x)$, then $u_n \in \hat{C}_{-\gamma}^{0, 2}([0, t] \times \mathbb{R}^d) \cap \hat{C}^{0, 4}([0, t] \times \mathbb{R}^d)$ and $\partial u_n/\partial s + G_n u_n = 0$.

We can therefore apply Oleinik's (1966) a priori estimate (or actually a simple extension of it to $C_c^\infty(\mathbb{R}^d)$) to conclude that there exists $\omega \geq 0$, depending (continuously) only on the uniform bounds on $a_{ij}^{(n)}/\varphi_2$, $b_i^{(n)}/\varphi_1$, $\partial_k \partial_i a_{ij}^{(n)}$, $\partial_k b_i^{(n)}$, and $\partial_k \partial_i b_i^{(n)}$, such that

$$(2.9) \quad \|T_n(t)f\|_{C_c^2(\mathbb{R}^d)} \leq e^{\omega t} \|f\|_{C_c^2(\mathbb{R}^d)}$$

for all $f \in C_c^\infty(\mathbb{R}^d)$, all n , and all $t \geq 0$. (The proof is essentially as in Stroock and Varadhan (1979), Theorem 3.2.4.) It follows that for each n and $t \geq 0$, $T_n(t): \hat{C}_{-3}^2(\mathbb{R}^d) \rightarrow \hat{C}_{-3}^2(\mathbb{R}^d)$ and (2.9) holds for all $f \in \hat{C}_{-3}^2(\mathbb{R}^d)$. Since

$$(2.10) \quad \begin{aligned} \|(G_n - G)f\| &\leq \frac{1}{2} \sum_{i,j=1}^d \left\| \frac{a_{ij}^{(n)} - a_{ij}}{\varphi_3} \right\| \|\varphi_3 \partial_i \partial_j f\| \\ &\quad + \sum_{i=1}^d \left\| \frac{b_i^{(n)} - b_i}{\varphi_3} \right\| \|\varphi_3 \partial_i f\| \\ &\leq \varepsilon_n \|f\|_{C_c^2(\mathbb{R}^d)} \end{aligned}$$

for all $f \in \hat{C}_{-3}^2(\mathbb{R}^d)$ and all n , where $\lim_{n \rightarrow \infty} \varepsilon_n = 0$, the stated conclusion follows from Proposition 3.7 of Chapter 1 with $D_0 = C_c^\infty(\mathbb{R}^d)$ and $\mathcal{D}(A) = D_1 = \hat{C}_{-3}^2(\mathbb{R}^d)$, at least under the additional hypotheses noted in the second paragraph of the proof. But these are removed just as in the proof of Theorem 2.1, the analogue of Lemma 2.2 following as in (2.9) from Oleinik's result. \square

To get uniqueness of solutions of the martingale problem in general, we need to assume that a has a Lipschitz continuous square root.

2.6 Theorem Let $\sigma: [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}^d \otimes \mathbb{R}^d$ and $b: [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ satisfy the conditions of Theorem 3.10 of Chapter 5, and put $a = \sigma\sigma^T$. Then, with G defined by (1.24), the martingale problem for $\{(f, Gf): f \in C_c^\infty(\mathbb{R}^d)\}$ is well-posed.

Proof. The result is a consequence of Theorems 3.10 and 3.6, Proposition 3.5, and Corollary 3.4, all from Chapter 5. \square

2.7 Remark Typically, it is a , rather than σ , that is given, and with this in mind we give sufficient conditions for $a^{1/2}$, the S_d -valued square root of a , to be Lipschitz continuous. Let $a: [0, \infty) \times \mathbb{R}^d \rightarrow S_d$ be Borel measurable, and assume either of the following conditions:

- (a) There exist $C > 0$ and $\delta > 0$ such that

$$(2.11) \quad |a(t, y) - a(t, x)| \leq C|y - x|, \quad t \geq 0, x, y \in \mathbb{R}^d,$$

and

$$(2.12) \quad \inf_{(t, x) \in [0, \infty) \times \mathbb{R}^d} \inf_{|\theta|=1} \theta \cdot a(t, x) \theta > \delta.$$

(b) $a_{ij}(t, \cdot) \in C^2(\mathbb{R}^d)$ for $i, j = 1, \dots, d$ and all $t \geq 0$, and there exists $\lambda > 0$ such that

$$(2.13) \quad \sup_{(t, x) \in [0, \infty) \times \mathbb{R}^d} \max_{1 \leq i \leq d} \sup_{|\theta| = 1} |\theta \cdot (\partial_i^2 a)(t, x)\theta| \leq \lambda.$$

Then $a^{1/2}$ is Borel measurable and

$$(2.14) \quad |a^{1/2}(t, y) - a^{1/2}(t, x)| \leq K|y - x|, \quad t \geq 0, x, y \in \mathbb{R}^d,$$

where $K = C/(2\delta^{1/2})$ if (a) holds and $K = d(2\lambda)^{1/2}$ if (b) holds. See Section 5.2 of Stroock and Varadhan (1979). \square

We conclude this section by considering a special class of generators, which arise in population genetics (see Chapter 10).

2.8 Theorem Let

$$(2.15) \quad K_d = \left\{ x \in [0, 1]^d : \sum_{i=1}^d x_i \leq 1 \right\},$$

define $a: K_d \rightarrow S_d$ by $a_{ij}(x) = x_i(\delta_{ij} - x_j)$, and let $b: K_d \rightarrow \mathbb{R}^d$ be Lipschitz continuous and satisfy

$$(2.16) \quad \begin{aligned} b_i(x) &\geq 0 \quad \text{if } x \in K_d \quad \text{and} \quad x_i = 0, \quad i = 1, \dots, d, \\ \sum_{i=1}^d b_i(x) &\leq 0 \quad \text{if } x \in K_d \quad \text{and} \quad \sum_{i=1}^d x_i = 1. \end{aligned}$$

Then, with G defined by (1.15), the closure of $\{(f, Gf) : f \in C^2(K_d)\}$ is single-valued and generates a Feller semigroup on $C(K_d)$. Moreover, the space of polynomials on K_d is a core for the generator.

The proof is quite similar to that of Theorem 2.1. It depends on the following lemma from Ethier (1976).

2.9 Lemma Assume, in addition to the hypotheses of Theorem 2.8, that $b_1, \dots, b_d \in C^4(K_d)$. Then the first conclusion of the theorem holds, as do the following two assertions:

- (a) $T(t): C^m(K_d) \rightarrow C^m(K_d)$ for all $t \geq 0$ and $m = 1, 2$.
- (b) $\sum_{i=1}^d \|\partial_i T(t)f\| \leq e^{\lambda t} \sum_{i=1}^d \|\partial_i f\|$ for all $f \in C^1(K_d)$ and $t \geq 0$, where

$$(2.17) \quad \lambda = \max_{1 \leq i \leq d} \sum_{j=1}^d \|\partial_j b_i\|.$$

Proof of Theorem 2.8. Choose $b^{(n)}: K_d \rightarrow \mathbb{R}^d$ for $n = 1, 2, \dots$ satisfying the conditions of Lemma 2.9 such that $\lim_{n \rightarrow \infty} \|b^{(n)} - b\| = 0$ and

$$(2.18) \quad \sup_{n \geq 1} \max_{1 \leq i \leq d} \sum_{j=1}^d \|\partial_j b_i^{(n)}\| < \infty.$$

The latter two conditions follow using convolutions. To get (2.16), it may be necessary to add $\varepsilon_n(1 - (d+1)x_i)$ to the $b_i^{(n)}(x)$ thus obtained, where $\varepsilon_n \rightarrow 0+$. The first conclusion now follows from Proposition 3.7 of Chapter 1 in the same way that Theorem 2.1 did. The second conclusion is a consequence of the fact that the space of polynomials on K_d is dense in $C^2(K_d)$ with respect to the norm

$$(2.19) \quad \|f\|_{C^2(K_d)} = \sum_{|\alpha| \leq 2} \|D^\alpha f\|$$

(see Appendix 7). □

3. OTHER PROCESSES

We begin by considering jump Markov processes in a locally compact, separable metric space E , the generators for which have the form

$$(3.1) \quad Af(x) = \lambda(x) \int (f(y) - f(x)) \mu(x, dy),$$

where $\lambda \in B_{loc}(E)$ is nonnegative and $\mu(x, \Gamma)$ is a transition function on $E \times \mathcal{B}(E)$. We assume, among other things, that λ and the mapping $x \rightarrow \mu(x, \cdot)$ are continuous on E . Thus, if E is compact, then A is a bounded linear operator on $C(E)$ and generates a Feller semigroup on $C(E)$. We can therefore assume without further loss of generality that E is noncompact. The case in which $E = \{0, 1, \dots\}$ is treated separately as a corollary.

3.1 Theorem Let E be a locally compact, noncompact, separable metric space and let $E^\Delta = E \cup \{\Delta\}$ be its one-point compactification. Let $\lambda \in C(E)$ be nonnegative and let $\mu(x, \Gamma)$ be a transition function on $E \times \mathcal{B}(E)$ such that the mapping $x \rightarrow \mu(x, \cdot)$ of E into $\mathcal{P}(E)$ is continuous. Let γ and η be positive functions in $C(E)$ such that $1/\gamma$ and $1/\eta$ belong to $\hat{C}(E)$ and

$$(3.2) \quad \sup_{x \in E} \frac{\lambda(x)}{\gamma(x)} = C_1 < \infty,$$

$$(3.3) \quad \lim_{x \rightarrow \Delta} \lambda(x) \mu(x, K) = 0 \quad \text{for every compact } K \subset E,$$

$$(3.4) \quad \sup_{x \in E} \lambda(x) \int \frac{\gamma(x) - \gamma(y)}{\gamma(y)} \mu(x, dy) \equiv C_2 < \infty,$$

$$(3.5) \quad \sup_{x \in E} \lambda(x) \int \frac{\eta(y) - \eta(x)}{\eta(x)} \mu(x, dy) \equiv C_3 < \infty.$$

Then, with A defined by (3.1), the closure of $\{(f, Af) : f \in \hat{C}(E), \gamma f \in \bar{C}(E), Af \in \hat{C}(E)\}$ is single-valued and generates a Feller semigroup on $\hat{C}(E)$. Moreover, $C_c(E)$ is a core for this generator.

Proof. Consider A as a linear operator on $\hat{C}(E)$ with domain $\mathcal{D}(A) = \{f \in \hat{C}(E) : \gamma f \in \bar{C}(E)\} \subset \hat{C}(E)$. To see that $A : \mathcal{D}(A) \rightarrow \bar{C}(E)$, let $f \in \mathcal{D}(A)$ and observe that $Af \in C(E)$ and

$$\begin{aligned} (3.6) \quad |Af(x)| &\leq \lambda(x) \left\{ \int \frac{1}{\gamma(y)} \mu(x, dy) + \frac{1}{\gamma(x)} \right\} \|\gamma f\| \\ &\leq \left\{ \frac{C_2 + \lambda(x)}{\gamma(x)} + C_1 \right\} \|\gamma f\| \\ &\leq (C_2 \|1/\gamma\| + 2C_1) \|\gamma f\|, \quad x \in E, \end{aligned}$$

by (3.2) and (3.4). Using the idea of Lemma 2.1 of Chapter 4, we also find that A is dissipative.

We claim that $\overline{\mathcal{D}(\alpha - A)} \supset \mathcal{D}(A)$ for all α sufficiently large. Given $n \geq 1$, define A_n on $\hat{C}(E)$ as in (3.1) but with $\lambda(x)$ replaced by $\lambda(x) \wedge n$. By (3.3), $A_n : C_c(E) \rightarrow \hat{C}(E)$, and hence A_n is a bounded linear operator on $\hat{C}(E)$ satisfying the positive maximum principle. It therefore generates a strongly continuous, positive, contraction semigroup $\{T_n(t)\}$ on $\hat{C}(E)$. By (3.4), there exists $\omega \geq 0$ not depending on n such that $\gamma A_n(1/\gamma) \leq \omega$, so

$$(3.7) \quad e^{-\omega t} T_n(t) \frac{1}{\gamma} - \frac{1}{\gamma} = \int_0^t e^{-\omega s} T_n(s) \left(A_n \frac{1}{\gamma} - \frac{\omega}{\gamma} \right) ds \leq 0$$

for all $t \geq 0$. Let $f \in \mathcal{D}(A)$. By (3.7), $T_n(t)f \in \mathcal{D}(A)$ and

$$(3.8) \quad \|\gamma T_n(t)f\| \leq \left\| \gamma T_n(t) \frac{1}{\gamma} \right\| \|\gamma f\| \leq e^{\omega t} \|\gamma f\|$$

for all $t \geq 0$. Hence, if $\alpha > \omega$, then $(\alpha - A_n)^{-1}f \in \mathcal{D}(A)$ and

$$(3.9) \quad \|\gamma(\alpha - A_n)^{-1}f\| \leq (\alpha - \omega)^{-1}\|\gamma f\|,$$

so

$$(3.10) \quad \begin{aligned} \|(A_n - A)(\alpha - A_n)^{-1}f\| &\leq \left\| \frac{\lambda \wedge n - \lambda}{(\lambda \wedge n)\gamma} \right\| \|\gamma A_n(\alpha - A_n)^{-1}f\| \\ &\leq \frac{C_1}{n} \left(\frac{\alpha}{\alpha - \omega} + 1 \right) \|\gamma f\| \end{aligned}$$

by (3.2). Since f and n were arbitrary, we conclude from Lemma 3.6 of Chapter 1 that $\mathcal{D}(\lambda - A) \supset \mathcal{D}(A)$.

Thus, by Theorem 4.3 of Chapter 1,

$$(3.11) \quad A_0 \equiv \{(f, g) \in \bar{A}: g \in \hat{C}(E)\}$$

is single-valued and generates a strongly continuous, positive, contraction semigroup $\{T(t)\}$ on $\hat{C}(E)$. Clearly, if $f \in \mathcal{D}(A_0)$, then $A_0 f$ is given by the right side of (3.1) and $A_n f \rightarrow A_0 f$ as $n \rightarrow \infty$. It follows from Theorem 6.1 of Chapter 1 that $T_n(t)f \rightarrow T(t)f$ for all $f \in \hat{C}(E)$ and $t \geq 0$. In particular, by (3.7), $T(t)(1/\gamma) \leq e^{\omega t}(1/\gamma)$ for all $t \geq 0$, so $T(t): \mathcal{D}(A) \rightarrow \mathcal{D}(A)$ for every $t \geq 0$. We conclude from Proposition 3.3 of Chapter 1 that $\mathcal{D}(A_0) \cap \mathcal{D}(A)$ is a core for A_0 , that is, the closure of $\{(f, Af): f \in \hat{C}(E), \gamma f \in \hat{C}(E), Af \in \hat{C}(E)\}$ generates $\{T(t)\}$.

Let $f \in \mathcal{D}(A_0) \cap \mathcal{D}(A)$ and choose $\{h_n\} \subset C_c(E)$ such that $\chi_{\{x \in E: |f(x)| > 1/n\}} \leq h_n \leq 1$ for each n , and observe that $\{f h_n\} \subset C_c(E)$, $f h_n \rightarrow f$ uniformly, and $A(f h_n) \rightarrow Af$ boundedly (by (3.7)) and pointwise. Recalling Remark 3.2 of Chapter 1, this implies that $C_c(E)$ is a core for A_0 .

It remains to show that A_0 is conservative. Fix $x \in E$, and let X be a Markov process corresponding to $\{T^A(t)\}$ (see Lemma 2.3 of Chapter 4) with sample paths in $D_E[0, \infty)$ and initial distribution δ_x . Extend η from E to E^A by setting $\eta(\Delta) = \infty$. Let $n \geq 1$ and define

$$(3.12) \quad \tau_n = \inf \{t \geq 0: \eta(X(t)) > n\}.$$

Then, approximating η monotonically from below by functions in $C_c(E)$, we find that

$$(3.13) \quad \begin{aligned} E[\eta(X(t \wedge \tau_n))] &= \eta(x) + E \left[\int_0^{t \wedge \tau_n} A\eta(X(s)) ds \right] \\ &\leq \eta(x) + C_3 E \left[\int_0^{t \wedge \tau_n} \eta(X(s)) ds \right] \\ &\leq \eta(x) + C_3 tn, \end{aligned}$$

for all $t \geq 0$ by (3.5), so $E[\eta(X(t \wedge \tau_n))]$ is bounded in t on bounded intervals. By the first inequality in (3.13),

$$(3.14) \quad E[\eta(X(t \wedge \tau_n))] \leq \eta(x) + C_3 \int_0^t E[\eta(X(s \wedge \tau_n))] ds,$$

and thus

$$(3.15) \quad nP\{\tau_n \leq t\} \leq E[\eta(X(t \wedge \tau_n))] \leq \eta(x)e^{C_3 t}$$

for all $t \geq 0$ by Gronwall's inequality. It follows that $\lim_{n \rightarrow \infty} \tau_n = \infty$ a.s. and hence X has almost all sample paths in $D_E[0, \infty)$. By Corollary 2.8 of Chapter 4, we conclude that A_0 is conservative. \square

3.2 Corollary

Let $E = \{0, 1, \dots\}$ and

$$(3.16) \quad Af(i) = \sum_{j \geq 0} q_{ij} f(j),$$

where the matrix $(q_{ij})_{i, j \geq 0}$ has nonnegative off-diagonal entries and row sums equal to zero. Assume also

$$(3.17) \quad \sup_{i \geq 0} \frac{|q_{ii}|}{i+1} < \infty,$$

$$(3.18) \quad \lim_{i \rightarrow \infty} q_{ij} = 0, \quad j \geq 0,$$

$$(3.19) \quad \sup_{i \geq 0} \sum_{j \geq 0} \frac{i+1}{j+1} q_{ij} < \infty,$$

$$(3.20) \quad \sup_{i \geq 0} \frac{1}{i+1} \sum_{j \geq 0} (j-i) q_{ij} < \infty.$$

Then the closure of $\{(f, Af) : f \in C_c(E)\}$ is single-valued and generates a Feller semigroup on $\hat{C}(E)$.

Proof. Apply Theorem 3.1 with $\lambda(i)\mu(i, \{j\}) = q_{ij}$ for $i \neq j$, $\mu(i, \{i\}) = 0$, and $\gamma(i) = \eta(i) = i + 1$. \square

We next state a uniqueness result for processes in \mathbb{R}^d with Lévy generators, that is, generators of the form

$$(3.21) \quad Gf(x) = \frac{1}{2} \sum_{i, j=1}^d a_{ij}(t, x) \partial_i \partial_j f(x) + \sum_{i=1}^d b_i(t, x) \partial_i f(x) \\ + \int_{\mathbb{R}^d} \left(f(x+y) - f(x) - \frac{y \cdot \nabla f(x)}{1+|y|^2} \right) \mu(t, x; dy).$$

3.3 Theorem Let $a: [0, \infty) \times \mathbb{R}^d \rightarrow S_d$ be bounded, continuous, and positive-definite-valued, $b: [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ bounded and Borel measurable, and

$\mu: [0, \infty) \times \mathbb{R}^d \rightarrow \mathcal{M}(\mathbb{R}^d)$ such that $\int_{\Gamma} |y|^2(1 + |y|^2)^{-1}\mu(t, x; dy)$ is bounded and continuous in (t, x) for every $\Gamma \in \mathcal{B}(\mathbb{R}^d)$. Then, with G defined by (3.21), the martingale problem for $\{(f, Gf); f \in C_c^\infty(\mathbb{R}^d)\}$ is well-posed.

Proof. By Corollary 3.7 and Theorem 3.8, both from Chapter 4, every solution of the martingale problem has a modification with sample paths in $D_{\mathbb{R}^d}[0, \infty)$. The result therefore follows from Stroock (1975). \square

When a , b , and μ in (3.22) are independent of (t, x) , A becomes

$$(3.22) \quad Gf(x) = \frac{1}{2} \sum_{i,j=1}^d a_{ij} \partial_i \partial_j f(x) + \sum_{i=1}^d b_i \partial_i f(x) \\ + \int_{\mathbb{R}^d} \left(f(x+y) - f(x) - \frac{y \cdot \nabla f(x)}{1+|y|^2} \right) \mu(dy).$$

Every stochastically continuous process in \mathbb{R}^d with homogeneous, independent increments has a generator of this form, where $a \in S_d$, $b \in \mathbb{R}^d$, $\mu \in \mathcal{M}(\mathbb{R}^d)$, and $\int_{\mathbb{R}^d} |y|^2(1 + |y|^2)^{-1}\mu(dy) < \infty$ (see Gihman and Skorohod (1969), Theorem VI.3.3). In this case we can strengthen the conclusion of Theorem 3.3.

3.4 Theorem Let $a \in S_d$, $b \in \mathbb{R}^d$, and $\mu \in \mathcal{M}(\mathbb{R}^d)$, and assume that $\int_{\mathbb{R}^d} |y|^2(1 + |y|^2)^{-1}\mu(dy) < \infty$. Then, with G defined by (3.22), the closure of $\{(f, Gf); f \in \hat{C}^2(\mathbb{R}^d)\}$ is single-valued and generates a Feller semigroup on $\hat{C}(\mathbb{R}^d)$. Moreover, $C_c^\infty(\mathbb{R}^d)$ is a core for this generator.

Proof. If a is positive definite, then by Theorem 3.3, the martingale problem for $\{(f, Gf); f \in C_c^\infty(\mathbb{R}^d)\}$ is well-posed. For each $x \in \mathbb{R}^d$, denote by X_x a solution with initial distribution δ_x , and note that since $(Gf)^x = G(f^x)$ for all $f \in C_c^\infty(\mathbb{R}^d)$, where $f^x(y) \equiv f(x+y)$, we have

$$(3.23) \quad E[f(X_x(t))] = E[f(x + X_0(t))]$$

for all $f \in B(E)$ and $t \geq 0$. It follows that we can define a strongly continuous, positive, contraction semigroup $\{T(t)\}$ on $\hat{C}(\mathbb{R}^d)$ by letting $T(t)f(x)$ be given by (3.23). Denoting the generator of $\{T(t)\}$ by A , we have $\{(f, Gf); f \in C_c^\infty(\mathbb{R}^d)\} \subset A$, hence $\{(f, Gf); f \in \hat{C}^2(\mathbb{R}^d)\} \subset A$. Moreover, by (3.23), $T(t): \hat{C}^\infty(\mathbb{R}^d) \rightarrow \hat{C}^\infty(\mathbb{R}^d)$ for all $t \geq 0$, so $\hat{C}^\infty(\mathbb{R}^d)$ is a core for A by Proposition 3.3 of Chapter 1.

Let $h \in C_c^\infty(\mathbb{R}^d)$ satisfy $\chi_{B(0, 1)} \leq h \leq \chi_{B(0, 2)}$, and approximate $f \in \hat{C}^\infty(\mathbb{R}^d)$ by $\{fh_n\} \subset C_c^\infty(\mathbb{R}^d)$, where $h_n(x) = h(x/n)$, to show that $C_c^\infty(\mathbb{R}^d)$ is a core for A . (To check that $\text{bp-lim}_{n \rightarrow \infty} A(fh_n) = Af$, it suffices to split the integral in (3.22) into two parts, $|y| \leq 1$ and $|y| > 1$.) Similarly, using $\{(h_n, Ah_n)\}$, we find that A is conservative. The case in which a is only nonnegative definite can be handled by approximating a by $a + \varepsilon I$, $\varepsilon > 0$. \square

We conclude this section with two results from the area of infinite particle systems. The first concerns spin-flip systems and the second exclusion processes. For further background on these processes, see Liggett (1977, 1985).

3.5 Theorem Let S be a countable set, and give $\{-1, 1\}$ the discrete topology and $E \equiv \{-1, 1\}^S$ the product topology. For each $i \in S$, define the difference operator Δ_i on $C(E)$ by $\Delta_i f(\eta) = f(\cdot, \eta) - f(\eta)$, where $(\cdot, \eta)_j = (1 - 2\delta_{ij})\eta_j$ for all $j \in S$. For each $i \in S$, let $c_i \in C(E)$ be nonnegative, and assume that

$$(3.24) \quad \sup_{i \in S} \|c_i\| < \infty, \quad \sup_{i \in S} \sum_{j \in S} \|\Delta_j c_i\| < \infty.$$

Then, with

$$(3.25) \quad A = \sum_{i \in S} c_i(\eta) \Delta_i,$$

the closure of $\{(f, Af) : f \in C(E), \sum_{i \in S} \|\Delta_i f\| < \infty\}$ is single-valued and generates a Feller semigroup on $C(E)$. Moreover, the space of (continuous) functions on E depending on only finitely many coordinates is a core for this generator.

Proof. The first assertion is essentially a special case of a more general result of Liggett (1972). The second is left to the reader (Problem 8). \square

3.6 Theorem Let S be a countable set, and give $\{0, 1\}$ the discrete topology and $E \equiv \{0, 1\}^S$ the product topology. For each $i, j \in S$, define the difference operator Δ_{ij} on $C(E)$ by $\Delta_{ij} f(\eta) = f(\cdot, \eta) - f(\eta)$, where

$$(3.26) \quad (\cdot, \eta)_k = \begin{cases} \eta_k, & k \neq i, j \\ \eta_j, & k = i \\ \eta_i, & k = j. \end{cases}$$

For each $i, j \in S$, let $c_{ij} \in C(E)$ be nonnegative and γ_{ij} be a nonnegative number, and assume that $c_{ii} \equiv 0$, $c_{ij} \equiv c_{ji}$, $c_{ij} \leq \gamma_{ij}$, and $\gamma_{ij} = \gamma_{ji}$ for all $i, j \in S$,

$$(3.27) \quad \sup_{i \in S} \sum_{j \in S} \gamma_{ij} < \infty,$$

and

$$(3.28) \quad \sum_{k \in S} \sup_{\eta \in E} |c_{ik}(\cdot, \eta) - c_{ik}(\eta)| \leq K\gamma_{ij}, \quad i, j \in S,$$

where $(\cdot, \eta)_l = \delta_{kl} + (1 - 2\delta_{kl})\eta_l$ for all $l \in S$ and K is a constant. Then, with

$$(3.29) \quad A = \sum_{i, j \in S} c_{ij}(\eta) \Delta_{ij},$$

the closure of $\{(f, Af) : f \in C(E), \sum_{i, j \in S} \gamma_{ij} \|\Delta_{ij} f\| < \infty\}$ is single-valued and generates a Feller semigroup on $C(E)$. Moreover, the space of (continuous) functions on E depending on only finitely many coordinates is a core for this generator.

Proof. The references given for the preceding theorem apply here. \square

4. PROBLEMS

1. For each $x \in I = [r_0, r_1]$, let $P_x \in \mathcal{P}(C_I[0, \infty))$ be the distribution of the diffusion process in I with initial distribution δ_x corresponding to the semigroup of Theorem 1.1. Let X be the coordinate process on $C_I[0, \infty)$, and define $\tau_y = \inf\{t \geq 0: X(t) = y\}$ for $y \in I$.
 - (a) Show that r_1 is accessible if and only if there exist $x \in I^\circ$ and $t > 0$ such that

$$(4.1) \quad \inf_{y \in (x, r_1)} P_x\{\tau_y \leq t\} > 0.$$
 - (b) Suppose r_1 is inaccessible. Show that r_1 is entrance if and only if there exist $y \in I^\circ$ and $t > 0$ such that

$$(4.2) \quad \inf_{x \in (y, r_1)} P_x\{\tau_y \leq t\} > 0.$$
 - (c) Prove Corollary 1.2.
2. Suppose, in addition to the hypotheses of Theorem 1.1, that there exists a constant K such that (1.12) holds. Show that infinite boundaries of I are natural.
3. Use Proposition 3.4 of Chapter 1 to establish Theorem 2.1 in the special case in which $I = [0, \infty)$ and

$$(4.3) \quad a(x) = ax, \quad b(x) = bx, \quad x \in I,$$

where $0 < a < \infty$ and $-\infty < b < \infty$. (The resulting diffusion occurs in Chapter 9.)

Hint: Look for solutions of the form $u(t, x) = e^{-\lambda(t)x}$.

4. Assume the hypotheses of Proposition 2.4, and for each $t \geq 0$ define the linear contraction $S(t)$ on $\hat{C}(\mathbb{R}^d)$ by

$$(4.4) \quad S(t)f(x) = E[f(x + \sqrt{t}\sigma(x)Z + tb(x))],$$

where Z is $N(0, I_d)$. ($\{S(t)\}$ is not necessarily a semigroup.) Given $t \geq 0$ and a partition $\pi = (0 = t_0 \leq t_1 \leq \dots \leq t_n = t)$ of $[0, t]$, define $\mu(\pi) = \max_{1 \leq i \leq n} (t_i - t_{i-1})$ and

$$(4.5) \quad S_\pi = S(t_n - t_{n-1}) \cdots S(t_1 - t_0),$$

and note that $S_\pi: \hat{C}^2(\mathbb{R}^d) \rightarrow \hat{C}^2(\mathbb{R}^d)$. Define the norm $\|\cdot\|$ on $\hat{C}^2(\mathbb{R}^d)$ by

$$(4.6) \quad \|f\| = \left\| \left\{ \sum_{i=1}^d (\partial_i f)^2 + \sum_{1 \leq i < j \leq d} (\partial_i \partial_j f)^2 \right\}^{1/2} \right\|.$$

Prove Proposition 2.4 by verifying each of the following assertions:

(a) There exists $K > 0$ such that

$$(4.7) \quad \|S(t)S(s)f - S(t+s)f\| \leq Ks\sqrt{t}\|f\|$$

for all $f \in \hat{C}^2(\mathbb{R}^d)$ and $s, t \in [0, 1]$.

(b) There exists $K > 0$ such that

$$(4.8) \quad \|S(t)f\| \leq (1 + Kt)\|f\|$$

for all $f \in \hat{C}^2(\mathbb{R}^d)$ and $0 \leq t \leq 1$.

(c) By parts (a) and (b), there exists $K > 0$ such that

$$(4.9) \quad \|S_{\pi_1}f - S_{\pi_2}f\| \leq Kt\sqrt{\mu(\pi_1) \vee \mu(\pi_2)}\|f\|$$

for all $f \in \hat{C}^2(\mathbb{R}^d)$, $0 \leq t \leq 1$, and partitions π_1, π_2 of $[0, t]$.

(d) Choose $\varphi \in C_c^\infty(\mathbb{R}^d)$ with $\varphi \geq 0$ and $\int \varphi(x) dx = 1$, and define $\{\varphi_n\} \subset C_c^\infty(\mathbb{R}^d)$ by $\varphi_n(x) = n^d \varphi(nx)$. Then there exists $K > 0$ such that

$$(4.10) \quad \|S(t)(f * \varphi_n) - (S(t)f) * \varphi_n\| \leq \frac{Kt\|f\|}{n}$$

for all $f \in \hat{C}^2(\mathbb{R}^d)$, $0 \leq t \leq 1$, and n .

(e) By parts (b)–(d), for each $f \in \hat{C}(\mathbb{R}^d)$ and $t \geq 0$,

$$(4.11) \quad T(t)f = \lim_{n \rightarrow \infty} S\left(\frac{t}{n}\right)^n f$$

exists and defines a Feller semigroup $\{T(t)\}$ on $\hat{C}(\mathbb{R}^d)$ whose generator is the closure of $\{(f, Gf) : f \in C_c^\infty(\mathbb{R}^d)\}$, where G is given by (1.15).

5. (a) Use Corollary 3.8 of Chapter 1 to prove the following result.

Let E be a closed convex set in \mathbb{R}^d with nonempty interior, let $a: E \rightarrow S_d$ and $b: E \rightarrow \mathbb{R}^d$ be bounded and continuous, and for every $x \in E$ let $\xi(x)$ be an \mathbb{R}^d -valued random variable with mean vector 0 and covariance matrix $a(x)$. Suppose that $E[|\xi(x)|^3]$ is bounded in x and that, for some $t_0 > 0$,

$$(4.12) \quad x + \sqrt{t}\xi(x) + tb(x) \in E \quad \text{a.s.}$$

whenever $x \in E$ and $0 \leq t \leq t_0$. Suppose further that, for $0 \leq t \leq t_0$, the equation

$$(4.13) \quad S(t)f(x) = E[f(x + \sqrt{t}\zeta(x) + tb(x))]$$

defines a linear contraction $S(t)$ on $\hat{C}(E)$ that maps $\hat{C}^3(E)$ into $\hat{C}^3(E)$, and that there exists $K > 0$ and a norm $\|\cdot\|$ on $\hat{C}^3(E)$ with respect to which it is a Banach space such that

$$(4.14) \quad \|S(t)f\| \leq (1 + Kt)\|f\|$$

for all $f \in \hat{C}^3(E)$ and $0 \leq t \leq t_0$. Then, with G defined by (1.15), the closure of $\{(f, Gf) : f \in C_c^\infty(E)\}$ is single-valued and generates a Feller semigroup on $\hat{C}(E)$.

- (b) Use part (a) to prove Proposition 2.4 under the additional assumption that $\sigma_{ij}, b_i \in \hat{C}^3(\mathbb{R}^d)$ for $i, j = 1, \dots, d$.
- (c) Use part (a) to prove Theorem 2.8 under the additional assumption that $b_1, \dots, b_d \in C^3(K_d)$.
- (d) Use part (a) to prove Theorem 2.1 under the additional assumptions that $-\infty < r_0 < r_1 < \infty$, $a, b \in C^3(I)$, and $a = \sigma_0 \sigma_1$, where $\sigma_i \in C^3(I)$, $\sigma_i(r_i) = 0$, and $\sigma_i > 0$ on I° for $i = 0, 1$, and $\sigma_0/(\sigma_0 + \sigma_1)$ is nondecreasing on I° and extends to an element of $C^3(I)$.

6. Fix integers $r, s \geq 2$ and index the coordinates of elements of \mathbb{R}^{rs-1} by

$$(4.15) \quad J = \{(i, j) : i = 1, \dots, r, j = 1, \dots, s, (i, j) \neq (r, s)\}.$$

Fix $\gamma \geq 0$ and, using the notation (2.15), define $G : C^2(K_{rs-1}) \rightarrow C(K_{rs-1})$ by

$$(4.16) \quad G = \frac{1}{2} \sum_{(i, j), (k, l) \in J} x_{ij} (\delta_{ik} \delta_{jl} - x_{kl}) \partial_{ij} \partial_{kl} - \gamma \sum_{(i, j) \in J} (x_{ij} - x_{i.} x_{.j}) \partial_{ij},$$

where $x_{i.} = \sum_{j=1}^s x_{ij}$, $x_{.j} = \sum_{i=1}^r x_{ij}$, and $x_{rs} = 1 - \sum_{(i, j) \in J} x_{ij}$. It follows from Theorem 2.8 that the closure of $\{(f, Gf) : f \in C^2(K_{rs-1})\}$ is single-valued and generates a Feller semigroup on $C(K_{rs-1})$. Use Proposition 3.5 of Chapter 1 to give a direct proof of this result.

Hint: Make the change of variables

$$(4.17) \quad p_i = x_{i.}, \quad q_j = x_{.j}, \quad D_{ij} = x_{ij} - x_{i.} x_{.j},$$

where $i = 1, \dots, r-1$ and $j = 1, \dots, s-1$, and define

$$(4.18) \quad \begin{aligned} \text{degree} \left(\prod_{i=1}^{r-1} p_i^{k_i} \prod_{j=1}^{s-1} q_j^{l_j} \prod_{i=1}^{r-1} \prod_{j=1}^{s-1} D_{ij}^{m_{ij}} \right) \\ = \sum_{i=1}^{r-1} k_i + \sum_{j=1}^{s-1} l_j + 2 \sum_{i=1}^{r-1} \sum_{j=1}^{s-1} m_{ij}. \end{aligned}$$

Let L_n be the space of polynomials of "degree" less than or equal to n .

7. Let S be a countable set, give $[0, 1]^S$ the product topology, and define

$$(4.19) \quad K = \left\{ x \in [0, 1]^S : \sum_{i \in S} x_i \leq 1 \right\}.$$

Suppose the matrix $(q_{ij})_{i, j \in S}$ has nonnegative off-diagonal entries and row sums equal to zero, and

$$(4.20) \quad \sup_{f \in S} \sum_{i \in S} |q_{ij}| < \infty.$$

Show that, with

$$(4.21) \quad G = \frac{1}{2} \sum_{i, j \in S} x_i (\delta_{ij} - x_j) \partial_i \partial_j + \sum_{i \in S} \left(\sum_{j \in S} q_{ji} x_j \right) \partial_i,$$

the closure of $\{(f, Gf) : f \in C(K), f \text{ depends on only finitely many coordinates and is twice continuously differentiable}\}$ is single-valued and generates a Feller semigroup on $C(K)$.

8. Use Problem 8 of Chapter 1 to prove Theorems 3.5 and 3.6.

5. NOTES

Theorem 1.1 is a very special case of Feller's (1952) theory of one-dimensional diffusions. (Our treatment follows Mandl (1968).) Theorems 1.4, 1.5, and 1.6 are based, respectively, on partial differential equation results of Schauder (1934), Fiorenza (1959), and Il'in, Kalashnikov, and Oleinik (1962). The first two of these results are presented in Ladyzhenskaya and Ural'tseva (1968) and the latter in Dynkin (1965). Theorem 1.7 is due to Stroock and Varadhan (1979).

Essentially Theorem 2.1 appears in Ethier (1978). Theorem 2.3 is due primarily to Yamada and Watanabe (1971). Roth (1977) is responsible for Proposition 2.4, while Theorem 2.5 is based on Oleinik (1966). Remark 2.7 is due to Freidlin (1968) and Phillips and Sarason (1968). Theorem 2.8 is a slight improvement of a result of Ethier (1976).

Theorem 3.3 was obtained by Stroock (1975), and Theorems 3.5 and 3.6 by Liggett (1972).

Problem 4 is Roth's (1977) proof. Problem 5(c) generalizes Norman (1971) and Problem 5(d) is due to Norman (1972). Problem 6 can be traced to Littler (1972) and Serant and Villard (1972). Problem 7 is due to Ethier (1981).

9

BRANCHING PROCESSES

Because of their independence properties, branching processes provide a rich source of weak convergence results. Here we give four examples. Section 1 considers the classical Galton–Watson process and the Feller diffusion approximation. Section 2 gives an analogous result for two-type branching models, and Section 3 does the same for a branching process in random environments. In Section 4 conditions are given for a sequence of branching Markov processes to converge to a measure-valued process.

1. GALTON–WATSON PROCESSES

In this section we consider approximations for the Galton–Watson branching process, which can be described as follows: Let independent, nonnegative integer-valued random variables $Z_0, \xi_k^n, k, n = 1, 2, \dots$, be given, and assume the ξ_k^n are identically distributed. Define Z_1, Z_2, \dots recursively by

$$(1.1) \quad Z_n = \sum_{k=1}^{Z_{n-1}} \xi_k^n.$$

Then Z_n gives the number of particles in the n th generation of a population in which individual particles reproduce independently and in the same manner. The distribution of ξ is called the *offspring distribution*, and that of Z_0 the

initial distribution. We are interested in approximations of this process when Z_0 is large. The first such approximations are given by the law of large numbers and the central limit theorem.

1.1 Theorem Let Z_n, ξ_k^n be as above and assume $E[\xi_k^n] \equiv m < \infty$. Then

$$(1.2) \quad \lim_{Z_0 \rightarrow \infty} \frac{Z_n}{Z_0} = m^n \quad \text{a.s.}$$

In addition let $\text{var } (\xi_k^n) = E[(\xi_k^n - m)^2] \equiv \sigma^2 < \infty$. Then as $Z_0 \rightarrow \infty$ the joint distributions of

$$(1.3) \quad W_n = Z_0^{-1/2}(Z_n - m^n Z_0)$$

converge to those of

$$(1.4) \quad W_n^\infty = \sum_{l=1}^n m^{n-(l+1)/2} V_l$$

where the V_l are independent normal random variables with $E[V_l] = 0$ and $\text{var } (V_l) = \sigma^2$.

1.2 Remark Note that

$$(1.5) \quad W_n^\infty = m W_{n-1}^\infty + m^{(n-1)/2} V_n. \quad \square$$

Proof. The limit in (1.2) is obtained by induction. The law of large numbers gives

$$(1.6) \quad \lim_{Z_0 \rightarrow \infty} \frac{Z_1}{Z_0} = \lim_{Z_0 \rightarrow \infty} Z_0^{-1} \sum_{k=1}^{Z_0} \xi_k^1 = m \quad \text{a.s.,}$$

and assuming $\lim_{Z_0 \rightarrow \infty} Z_{n-1}/Z_0 = m^{n-1}$ a.s.,

$$(1.7) \quad \lim_{Z_0 \rightarrow \infty} \frac{Z_n}{Z_0} = \left(\frac{Z_{n-1}}{Z_0} \right) Z_{n-1}^{-1} \sum_{k=1}^{Z_{n-1}} \xi_k^n = m^n \quad \text{a.s.}$$

Rewriting W_n as

$$\begin{aligned} (1.8) \quad W_n &= Z_0^{-1/2}(Z_n - m^n Z_0) \\ &= Z_0^{-1/2} \sum_{l=1}^n (m^{n-l} Z_l - m^{n-l+1} Z_{l-1}) \\ &= \sum_{l=1}^n m^{n-l} \left(\frac{Z_{l-1}}{Z_0} \right)^{1/2} Z_{l-1}^{-1/2} \sum_{k=1}^{Z_{l-1}} (\xi_k^l - m), \end{aligned}$$

we see from (1.8) that it is enough to show the random variables

$$(1.9) \quad U_l = Z_{l-1}^{-1/2} \sum_{k=1}^{Z_{l-1}} (\xi_k^l - m)$$

converge in distribution to independent $N(0, \sigma^2)$ random variables. Let $\mathcal{F}_n = \sigma(Z_0, \xi_k^l : 1 \leq l \leq n, 1 \leq k < \infty)$. Then as in the usual proof of the central limit theorem

$$(1.10) \quad \lim_{Z_0 \rightarrow \infty} E[\exp \{i\theta U_l\} | \mathcal{F}_{l-1}] \\ = \lim_{Z_{l-1} \rightarrow \infty} E_\xi [\exp \{i\theta Z_{l-1}^{-1/2}(\xi - m)\}]^{Z_{l-1}} = \exp \{-\frac{1}{2}\sigma^2\theta^2\} \quad \text{a.s.}$$

where the expectation on the right is with respect to ξ . Therefore

$$(1.11) \quad \lim_{Z_0 \rightarrow \infty} E \left[\prod_{l=1}^n \exp \{i\theta_l U_l\} \right] = \prod_{l=1}^n \exp \{-\frac{1}{2}\sigma^2\theta_l^2\}$$

and the convergence in distribution follows from the convergence of the characteristic functions. \square

Of more interest is the Feller diffusion approximation.

1.3 Theorem Let $E[\xi_k^n] = 1$ (the critical case) and let $\text{var}(\xi_k^n) = \sigma^2 < \infty$. Let $Z_k^{(m)}$ be a sequence of Galton–Watson processes defined as in (1.1), and suppose $Z_0^{(m)}/m$ converges in distribution. Then W_m defined by

$$(1.12) \quad W_m(t) \equiv \frac{Z_{\lfloor mt \rfloor}^{(m)}}{m}$$

converges in distribution in $D_{[0, \infty)}[0, \infty)$ to a diffusion with generator

$$(1.13) \quad Af(x) = \frac{1}{2}\sigma^2 x f''(x), \quad f \in C_c^\infty([0, \infty)).$$

Proof. By Theorem 2.1 of Chapter 8, $C_c^\infty([0, \infty))$ is a core for A . Note that $Z_n^{(m)}/m$ is a Markov chain with values in $E_m = \{l/m : l = 0, 1, 2, \dots\}$, and define

$$(1.14) \quad T_m f(x) = E \left[f \left(m^{-1} \sum_{k=1}^{mx} \xi_k \right) \right],$$

where the ξ_k are independent and distributed as the ξ_k^n . By Theorem 6.5 of Chapter 1 and Corollary 8.9 of Chapter 4, it is sufficient to show

$$(1.15) \quad \lim_{m \rightarrow \infty} \sup_{x \in E_m} |m(T_m f(x) - f(x)) - \frac{1}{2}\sigma^2 x f''(x)| = 0, \quad f \in C_c^\infty([0, \infty)).$$

For $x \in E_m$, define

$$(1.16) \quad \begin{aligned} \varepsilon_m(x) &= m(T_m f(x) - f(x)) - \frac{1}{2}\sigma^2 x f''(x) \\ &= mE \left[f \left(m^{-1} \sum_{k=1}^{mx} \xi_k \right) - f(x) \right] - \frac{1}{2}E \left[\left(\frac{1}{\sqrt{m}} \sum_{k=1}^{mx} (\xi_k - 1) \right)^2 \right] f''(x) \\ &= E \left[\int_0^1 S_{mx}^2 x(1-v) \left(f'' \left(x + v \sqrt{\frac{x}{m}} S_{mx} \right) - f''(x) \right) dv \right]. \end{aligned}$$

where $S_n = n^{-1/2} \sum_{k=1}^n (\xi_k - 1)$. Suppose the support of f is contained in $[0, c]$. Since

$$(1.17) \quad x + v \sqrt{\frac{x}{m}} S_{mx} = x + v \frac{1}{m} \sum_{k=1}^{mx} (\xi_k - 1) \geq x(1 - v),$$

the integrand on the right of (1.16) is zero if $v \leq 1 - c/x$. Consequently,

$$\begin{aligned} (1.18) \quad & \left| \int_0^1 S_{mx}^2 x(1-v) \left(f'' \left(x + v \sqrt{\frac{x}{m}} S_{mx} \right) - f''(x) \right) dv \right| \\ & \leq \int_{0 \vee (1-c/x)}^1 S_{mx}^2 x(1-v) 2 \|f''\| dv \\ & = x \|f''\| ((c/x) \wedge 1)^2 S_{mx}^2. \end{aligned}$$

To show that $\lim_{m \rightarrow \infty} \sup_x \varepsilon_m(x) = 0$, it is enough to show $\lim_{m \rightarrow \infty} \varepsilon_m(x_m) = 0$ for every convergent sequence x_m , where we allow $\lim_{m \rightarrow \infty} x_m = \infty$. Since $E[S_{mx}^2] = \sigma^2$ for all m, x , inequality (1.18) implies $\lim_{m \rightarrow \infty} \varepsilon_m(x_m) = 0$ if either $\lim_{m \rightarrow \infty} x_m = 0$ or $\lim_{m \rightarrow \infty} x_m = \infty$. Therefore we need only consider the case $x_m \rightarrow x$, for $0 < x < \infty$. Replacing x by x_m in (1.18), $S_{mx_m} \Rightarrow V$, where V is $N(0, \sigma^2)$, and hence the left side of (1.18) converges in probability to zero and the right side converges in distribution to $x \|f''\| ((c/x) \wedge 1)^2 V^2$. Since

$$(1.19) \quad \lim_{m \rightarrow \infty} E \left[x_m \|f''\| \left(\left(\frac{c}{x_m} \right) \wedge 1 \right)^2 S_{mx_m}^2 \right] = E \left[x \|f''\| \left(\left(\frac{c}{x} \right) \wedge 1 \right)^2 V^2 \right],$$

the dominated convergence theorem implies $\lim_{m \rightarrow \infty} \varepsilon_m(x_m) = 0$. \square

Let ξ_k be a sequence of nonnegative integer-valued, independent, identically distributed random variables. Given Z_0 , independent of the ξ_k , define

$$(1.20) \quad S(l) = Z_0 + \sum_{k=1}^l (\xi_k - 1)$$

and for $n > 0$,

$$(1.21) \quad Z_n = S \left(\sum_{i=0}^{n-1} Z_i \right).$$

Let

$$(1.22) \quad Y_n = \sum_{i=0}^{n-1} Z_i.$$

Since

$$(1.23) \quad Z_n - Z_{n-1} = \sum_{k=Y_{n-1}+1}^{Y_n} (\xi_k - 1) = \sum_{k=1}^{Z_{n-1}} \xi_{Y_{n-1}+k} - Z_{n-1},$$

we see that Z_n is a Galton-Watson branching process (i.e., the joint distribution of the Z_n defined by (1.21) is the same as in (1.1) for the same offspring and initial distributions). We use the representation in (1.21) and Theorem 1.5 of Chapter 6 to give a generalization of Theorem 1.3.

Let $Z_n^{(m)}$ be a sequence of branching processes represented as in (1.21) with offspring distributions that may depend on m . Let $c_m > 0$ be a sequence of constants with $\lim_{m \rightarrow \infty} c_m = \infty$ and define

$$(1.24) \quad Y_m(u) = c_m^{-1} \left(\sum_{k=1}^{\lfloor mc_m u \rfloor} (\xi_k^m - 1) + Z_0^{(m)} \right)$$

and

$$(1.25) \quad W_m(t) = c_m^{-1} Z_{\lfloor mt \rfloor}^{(m)}.$$

Then

$$(1.26) \quad W_m(t) = Y_m \left(\int_0^{\lfloor mt \rfloor / m} W_m(s) \, ds \right) = Y_m \left(\int_0^{\lfloor mt \rfloor / m} \beta(W_m(s)) \, ds \right),$$

where $\beta(x) = x \vee 0$.

1.4 Theorem Suppose (for simplicity) that $Y_m(0)$ is a constant, that $\lim_{m \rightarrow \infty} Y_m(0) = Y(0) > 0$, and that $\{Y_m(1)\}$ converges in distribution. Then $Y_m \Rightarrow Y$ where Y is a process with independent increments, and $W_m \Rightarrow W$ in $D_{[0, \infty]}[0, \infty)$ where W satisfies

$$(1.27) \quad W(t) = Y \left(\int_0^t \beta(W(s)) \, ds \right)$$

for $t < \tau_\infty \equiv \lim_{n \rightarrow \infty} \inf \{s : W(s) > n\}$ and $W(t) = \infty$ for $t \geq \tau_\infty$.

1.5 Remark If $W(t) < \infty$ for all $t \geq 0$ a.s., then $W_m \Rightarrow W$ in $D_{[0, \infty]}[0, \infty)$. \square

Proof. For simplicity we treat only the case where $\alpha = \sup_m E[|Y_m(1)|] < \infty$, which in particular implies $E[W_m(t)] \leq Y_m(0)e^{\alpha t}$ and hence $\tau_\infty = \infty$.

Let $X_m(t) = Y_m(t) - Y_m(0)$. Since $\{Y_m(1)\}$ converges in distribution and $Y_m(0) \rightarrow Y(0)$, $\{X_m(1)\}$ converges in distribution, and we have

$$(1.28) \quad \begin{aligned} \lim_{m \rightarrow \infty} E[\exp \{i\theta X_m(m^{-1}c_m^{-1})\}]^{mc_m} &= \lim_{m \rightarrow \infty} E[\exp \{i\theta X_m(1)\}] \\ &\equiv \psi(\theta). \end{aligned}$$

It follows that

$$(1.29) \quad \begin{aligned} \lim_{m \rightarrow \infty} E[\exp \{i\theta X_m(t)\}] &= \lim_{m \rightarrow \infty} E[\exp \{i\theta X_m(m^{-1}c_m^{-1})\}]^{\lfloor mc_m t \rfloor} \\ &= \psi(\theta)^t. \end{aligned}$$

The independence of the ξ_k implies the finite-dimensional distributions of

X_m converge to those of the process X having independent increments and satisfying $E[e^{i\theta X(t)}] = \psi(\theta)$. To see that the sequence $\{X_m\}$ is relatively compact, define $\mathcal{F}_t^{(m)} = \sigma(X_m(s): s \leq t)$ and note that the independence implies

$$(1.30) \quad E[|X_m(t+u) - X_m(t)| \wedge 1 | \mathcal{F}_t^{(m)}]$$

$$\leq E[|X_m(u)| \wedge 1] \vee E\left[\left|X_m\left(u + \frac{1}{m}\right)\right| \wedge 1\right],$$

and the relative compactness follows from Theorem 8.6 of Chapter 3.

Under the assumption that $\sup_m E[|Y_m(1)|] < \infty$, the theorem follows from Theorem 1.5 of Chapter 6 if we verify that $\tau_0 = \tau_1$ a.s. (defined as in (1.6) and (1.9), both of Chapter 6).

Note that $X_m(t) - X_m(t-) \geq -c_m^{-1}$ and it follows from Theorem 10.2 of Chapter 3 that $M(t) \equiv \inf_{s \leq t} X(s)$ is continuous (see Problem 26 of Chapter 3). Consequently $X(\tau_0) = 0$ if $\tau_0 < \infty$, and by the strong Markov property the following lemma implies $\tau_0 = \tau_1$ a.s. and completes the proof. \square

1.6 Lemma Let X be a right continuous process with stationary, independent increments (in particular, the distribution of $X(t+u) - X(t)$ does not depend on u) and let $X(0) = 0$. Suppose $\inf_{s \leq t} X(s)$ is continuous. Then

$$(1.31) \quad P\left\{\int_0^t (0 \vee X(t))^{-1} dt = \infty\right\} = 1$$

for all $\varepsilon > 0$.

Proof. The process X may be written as a sum $X = X_1 + X_2$, where X_1 and X_2 are independent, X_2 is a compound Poisson process, and X_1 has finite expectation. Then (1.31) is equivalent to

$$(1.32) \quad P\left\{\int_0^t (0 \vee X_1(t))^{-1} dt = \infty\right\} = 1$$

for all $\varepsilon > 0$, which in turn is equivalent to showing

$$(1.33) \quad \tau(t) \equiv \inf \left\{s: \int_0^s (0 \vee X_1(u))^{-1} du \geq t\right\} = 0 \quad \text{a.s.}$$

for all $t \geq 0$. Let $Z(t) = X_1(\tau(t))$. Then (cf. (1.4) of Chapter 6)

$$(1.34) \quad \tau(t) = \int_0^t 0 \vee Z(s) ds = \int_0^t Z(s) ds$$

since Z is nonnegative by the continuity of $\inf_{s \leq t} X(s)$. Let $c = E[X_1(1)]$. Since $\tau(t)$ is a stopping time and $X_1(t) - ct$ is a martingale, we have

$$(1.35) \quad \begin{aligned} E[\tau(t)] &= \int_0^t E[Z(s)] ds = \int_0^t E[X_1(\tau(s))] ds \\ &= \int_0^t cE[\tau(s)] ds, \end{aligned}$$

and it follows that $E[\tau(t)] = 0$. Hence we have (1.33). \square

2. TWO-TYPE MARKOV BRANCHING PROCESSES

We alter the model considered in Section 1 in two ways. First, we assume that there are two types of particles (designated type 1 and type 2) with each particle capable, at death, of producing particles not only of its type but of the other type as well. Second, we assume each particle lives a random length of time that has an exponential distribution with parameter λ_i , $i = 1, 2$, depending on the type of the particle. Let p_{kl}^i be the probability that at death a particle of type i produces k offspring of type 1 and l offspring of type 2.

The generator for the two-type process then has the form

$$(2.1) \quad \begin{aligned} Bf(z_1, z_2) &= \lambda_1 z_1 \sum_{k,l} p_{kl}^1 (f(z_1 - 1 + k, z_2 + l) - f(z_1, z_2)) \\ &\quad + \lambda_2 z_2 \sum_{k,l} p_{kl}^2 (f(z_1 + k, z_2 - 1 + l) - f(z_1, z_2)). \end{aligned}$$

Let (γ_1, γ_2) have joint distribution p_{kl}^1 and let (ψ_1, ψ_2) have joint distribution p_{kl}^2 . Assume that $E[\gamma_i^3] < \infty$ and $E[\psi_i^3] < \infty$, $i = 1, 2$. Let $m_{ij} = E[\gamma_j]$ and $m_{2j} = E[\psi_j]$. We assume $m_{ij} > 0$ for all i, j ; that is, there is positive probability of offspring of each type regardless of the type of the parent. We also assume that the process is critical; in other words the matrix

$$(2.2) \quad M = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix}$$

has largest eigenvalue 1. This implies that the matrix

$$(2.3) \quad \Lambda = \begin{pmatrix} \lambda_1(m_{11} - 1) & \lambda_1 m_{12} \\ \lambda_2 m_{21} & \lambda_2(m_{22} - 1) \end{pmatrix}$$

has eigenvalues 0 and $-\eta$ for some $\eta > 0$. Let $(v_1, v_2)^T$ denote the eigenvector corresponding to 0 and $(\mu_1, \mu_2)^T$ the eigenvector corresponding to $-\eta$. We can take $v_1, v_2 > 0$ and μ_1 and μ_2 will have opposite signs.

Let (z_1, z_2) be fixed. We consider a sequence of processes $\{(Z_1^{(n)}, Z_2^{(n)})\}$ with generator B and $(Z_1^{(n)}(0), Z_2^{(n)}(0)) = ([nz_1], [nz_2])$. Define

$$(2.4) \quad X_n(t) = \frac{v_1 Z_1^{(n)}(nt) + v_2 Z_2^{(n)}(nt)}{n}$$

and

$$(2.5) \quad Y_n(t) = \frac{\mu_1 Z_1^{(n)}(nt) + \mu_2 Z_2^{(n)}(nt)}{n}.$$

Then X_n and $Y_n e^{-n\eta t}$ are martingales (Problem 8). Since for $t > 0$, $e^{n\eta t} \rightarrow \infty$ as $n \rightarrow \infty$, the fact that $Y_n e^{-n\eta t}$ is a martingale suggests that $Y_n(t) \rightarrow 0$ and, consequently, that $Z_2^{(n)}(nt)/n \approx -\mu_1 \mu_2^{-1} Z_1^{(n)}(nt)/n$ so that $Z_1^{(n)}(nt)/n \approx \mu_2 X_n(t)/(v_1 \mu_2 - v_2 \mu_1)$ and $Z_2^{(n)}(nt)/n \approx \mu_1 X_n(t)/(v_2 \mu_1 - v_1 \mu_2)$. This is indeed the case, so the limiting behavior of $\{X_n\}$ gives the limiting behavior of $(Z_1^{(n)}(nt)/n, Z_2^{(n)}(nt)/n)$ for $t > 0$.

We describe the limiting behavior of $\{Y_n\}$ in terms of

$$(2.6) \quad W_n(t) = Y_n(t) + \int_0^t n\eta Y_n(s) ds,$$

which is also a martingale.

Define

$$(2.7) \quad \begin{aligned} \xi_1 &= v_1(\gamma_1 - 1) + v_2 \gamma_2, & \xi_2 &= \mu_1(\gamma_1 - 1) + \mu_2 \gamma_2, \\ \varphi_1 &= v_1 \psi_1 + v_2(\psi_2 - 1), & \varphi_2 &= \mu_2 \psi_1 + \mu_1(\psi_2 - 1). \end{aligned}$$

Let $\alpha_{ij}^1 = E[\xi_i \xi_j]$ and $\alpha_{ij}^2 = E[\varphi_i \varphi_j]$, and note that $E[\xi_1] = E[\varphi_1] = 0$, $E[\xi_2] = -\eta \mu_1 \lambda_1^{-1}$, and $E[\varphi_2] = -\eta \mu_2 \lambda_2^{-1}$.

2.1 Theorem (a) The sequence $\{(X_n, W_n)\}$ converges in distribution to a diffusion process (X, W) with generator

$$(2.8) \quad Af(x, w) = \frac{x}{2} (a_{11} f_{xx}(x, w) + 2a_{12} f_{xw}(x, w) + a_{22} f_{ww}(x, w))$$

where $a_{ij} = (\lambda_1 \mu_2 \alpha_{ij}^1 - \lambda_2 \mu_1 \alpha_{ij}^2)/(v_1 \mu_2 - v_2 \mu_1)$.

(b) For $T > 0$, $\sup_{t \leq T} |Y_n(t) - Y_n(0)e^{-n\eta t}|$ converges to zero in probability, and hence for $0 < t_1 < t_2$, $\int_{t_1}^{t_2} n\eta Y_n(s) ds$ converges in distribution to $W(t_2) - W(t_1)$.

(c) For $T > 0$,

$$\sup_{t \leq T} |Z_1^{(n)}(nt)/n - (\mu_2 X_n(t) - v_2 Y_n(0)e^{-n\eta t})/(v_1 \mu_2 - v_2 \mu_1)|$$

converges to zero in probability.

Proof. It is not difficult to show that $C_c^\infty([0, \infty) \times (-\infty, \infty))$ is a core for A (see Problem 3). With reference to Corollary 8.6 of Chapter 4, let $f \in C_c^\infty([0, \infty) \times (-\infty, \infty))$ and set

$$(2.9) \quad \tilde{f}_n(t) = f(X_n(t), W_n(t)).$$

To find \tilde{g}_n so that $(\tilde{f}_n, \tilde{g}_n) \in \mathcal{A}_n$ we calculate $\lim_{\epsilon \rightarrow 0} \epsilon^{-1} E[\tilde{f}_n(t + \epsilon) - \tilde{f}_n(t) | \mathcal{F}_t^n]$ and obtain

$$(2.10)$$

$$\begin{aligned} \tilde{g}_n(t) &= n\lambda_1 Z_1^{(n)}(nt)E_\xi \left[f\left(X_n(t) + \frac{1}{n}\xi_1, W_n(t) + \frac{1}{n}\xi_2\right) - f(X_n(t), W_n(t)) \right] \\ &\quad + n\lambda_2 Z_2^{(n)}(nt)E_\varphi \left[f\left(X_n(t) + \frac{1}{n}\varphi_1, W_n(t) + \frac{1}{n}\varphi_2\right) - f(X_n(t), W_n(t)) \right] \\ &\quad + n\eta Y_n(t)f_w(X_n(t), W_n(t)), \end{aligned}$$

where E_ξ and E_φ only involve the ξ_i and φ_i . Recalling that $E[\xi_1] = E[\varphi_1] = 0$, $E[\xi_2] = -\eta\mu_1\lambda_1^{-1}$, $E[\varphi_2] = -\eta\mu_2\lambda_2^{-1}$, and $Y_n(t) = (\mu_1 Z_1^{(n)}(nt) + \mu_2 Z_2^{(n)}(nt))/n$, we see that if we expand f in a Taylor series about $(X_n(t), W_n(t))$, then the terms involving f_x have zero coefficients and the terms involving f_w cancel. Hence we have

$$\begin{aligned} (2.11) \quad \tilde{g}_n(t) &= \lambda_1 2^{-1}n^{-1}Z_1^{(n)}(nt)[\alpha_{11}^1 f_{xx}(X_n(t), W_n(t)) \\ &\quad + 2\alpha_{12}^1 f_{xw}(X_n(t), W_n(t)) + \alpha_{22}^1 f_{ww}(X_n(t), W_n(t))] \\ &\quad + \lambda_2 2^{-1}n^{-1}Z_2^{(n)}(nt)[\alpha_{11}^2 f_{xx}(X_n(t), W_n(t)) \\ &\quad + 2\alpha_{12}^2 f_{xw}(X_n(t), W_n(t)) + \alpha_{22}^2 f_{ww}(X_n(t), W_n(t))] \\ &\quad + O(n^{-1}X_n(t)) \\ &= Af(X_n(t), W_n(t)) + Y_n(t)h(X_n(t), W_n(t)) + O(n^{-1}X_n(t)), \end{aligned}$$

where h is smooth and does not depend on n . The last step follows by replacing $n^{-1}Z_1^{(n)}(nt)$ by $(\mu_2 X_n(t) - v_1 Y_n(t))/(v_1 \mu_1 - v_2 \mu_2)$ and $n^{-1}Z_2^{(n)}(nt)$ by $(\mu_1 X_n(t) - v_1 Y_n(t))/(v_2 \mu_1 - v_1 \mu_2)$ and collecting terms. The error is $O(n^{-1}X_n(t))$ since $n^{-1}Z_1^{(n)}(nt)$ and $n^{-1}Z_2^{(n)}(nt)$ are bounded by a constant times $X_n(t)$.

Finally, setting

$$(2.12) \quad f_n(t) = f(X_n(t), W_n(t)) + n^{-1}\eta^{-1}Y_n(t)h(X_n(t), W_n(t)),$$

and calculating g_n so that $(f_n, g_n) \in \mathcal{A}_n$, we obtain

$$\begin{aligned}
(2.13) \quad g_n(t) &= \tilde{g}_n(t) + \eta^{-1} \lambda_1 Z_1^{(n)}(nt) \\
&\times E_{\xi} \left[\left(Y_n(t) + \frac{1}{n} \xi_2 \right) h \left(X_n(t) + \frac{1}{n} \xi_1, W_n(t) + \frac{1}{n} \xi_2 \right) \right. \\
&\quad \left. - Y_n(t) h(X_n(t), W_n(t)) \right] \\
&+ \eta^{-1} \lambda_2 Z_2^{(n)}(nt) \\
&\times E_{\varphi} \left[\left(Y_n(t) + \frac{1}{n} \varphi_2 \right) h \left(X_n(t) + \frac{1}{n} \varphi_1, W_n(t) + \frac{1}{n} \varphi_2 \right) \right. \\
&\quad \left. - Y_n(t) h(X_n(t), W_n(t)) \right] \\
&+ Y_n^2(t) h_w(X_n(t), W_n(t)) \\
&= \tilde{g}_n(t) - n^{-1} Z_1^{(n)}(nt) \mu_1 h(X_n(t), W_n(t)) \\
&\quad - n^{-1} Z_2^{(n)}(nt) \mu_2 h(X_n(t), W_n(t)) + O(n^{-1} X_n^2(t)) \\
&= Af(X_n(t), W_n(t)) + O(n^{-1} X_n(t), n^{-1} X_n^2(t)).
\end{aligned}$$

Since X_n is a martingale and

$$\begin{aligned}
(2.14) \quad E[X_n^2(t)] &= X_n^2(0) + E \left[\int_0^t (\lambda_1 n^{-1} Z_1^{(n)}(ns) E_{\xi}[\xi_1^2] \right. \\
&\quad \left. + \lambda_2 n^{-1} Z_2^{(n)}(ns) E_{\varphi}[\varphi_1^2]) ds \right] \\
&\leq X_n^2(0) + (\lambda_1 v_1^{-1} + \lambda_2 v_2^{-1}) \int_0^t E[X_n(s)] ds \\
&= X_n^2(0) + t(\lambda_1 v_1^{-1} + \lambda_2 v_2^{-1}) X_n(0),
\end{aligned}$$

we have

$$(2.15) \quad E \left[\sup_{t \leq T} X_n^2(t) \right] \leq 4E[X_n^2(T)] < \infty.$$

Consequently,

$$(2.16) \quad \lim_{n \rightarrow \infty} E \left[\sup_{t \leq T} |f(X_n(t), W_n(t)) - f_n(t)| \right] = 0$$

and

$$(2.17) \quad \lim_{n \rightarrow \infty} E \left[\sup_{t \leq T} |Af(X_n(t), W_n(t)) - g_n(t)| \right] = 0,$$

so part (a) follows by Corollary 8.6 of Chapter 4. Observe that (2.6) can be inverted to give

$$(2.18) \quad Y_n(t) - e^{-n\eta t} Y_n(0) = \int_0^t n\eta e^{-n\eta(t-s)} (W_n(t) - W_n(s)) ds \\ + e^{-n\eta t} (W_n(t) - W_n(0)).$$

Let $U_n(s) = \sup_{t \leq T} |W_n(t+s) - W_n(t)|$. Then for $t \leq T$,

$$(2.19) \quad |Y_n(t) - e^{-n\eta t} Y_n(0)| \leq \int_0^t n\eta e^{-n\eta s} U_n(s) ds + e^{-n\eta t} U_n(t).$$

Part (b) follows from the fact that $U_n \Rightarrow U$ ($U(s) = \sup_{t \leq T} |W(t+s) - W(t)|$) and $\lim_{s \rightarrow 0} U(s) = 0$. Part (c) follows from part (b) and the definitions of X_n and Y_n . \square

3. BRANCHING PROCESSES IN RANDOM ENVIRONMENTS

We now consider continuous-time processes in which the splitting intensities are themselves stochastic processes. Let X be the population size and Λ_k , $k \geq 0$, be nonnegative stochastic processes. We want

$$(3.1) \quad P\{X(t + \Delta t) = X(t) - 1 + k \mid \mathcal{F}_t\} = E \left[\int_t^{t+\Delta t} \Lambda_k(s) X(s) ds \mid \mathcal{F}_t \right] + o(\Delta t),$$

that is (essentially), we want $\Lambda_k(t) \Delta t$ to be the probability that a given particle dies and is replaced by k particles. The simplest way to construct such a process is to take independent standard Poisson processes Y_k , independent of the Λ_k , and solve

$$(3.2) \quad X(t) = X(0) + \sum_{k=0}^{\infty} (k-1) Y_k \left(\int_0^t \Lambda_k(s) X(s) ds \right).$$

We assume that $\sum k \int_0^t \Lambda_k(s) ds < \infty$ a.s. for all $t > 0$ to assure that a solution of (3.2) exists for all time. In fact, we take (3.2) to define X rather than (3.1). We leave the verification that (3.1) holds for X satisfying (3.2) as a problem (Problem 4).

By analogy with the results of Sections 1 and 2, we consider a sequence of processes X_n with corresponding intensity processes $\Lambda_k^{(n)}$ and define $Z_n(t) = X_n(nt)/n$. Assuming $X_n(0) = n$ and defining $A_n(t) = \sum_{k=0}^{\infty} (k-1)\Lambda_k^{(n)}(t)$, we get

$$\begin{aligned} (3.3) \quad Z_n(t) &= 1 + \sum_{k=0}^{\infty} (k-1)n^{-1} Y_k \left(n^2 \int_0^t \Lambda_k^{(n)}(ns) Z_n(s) ds \right) \\ &= 1 + \sum_{k=0}^{\infty} (k-1)n^{-1} \tilde{Y}_k \left(n^2 \int_0^t \Lambda_k^{(n)}(ns) Z_n(s) ds \right) \\ &\quad + \int_0^t n A_n(ns) Z_n(s) ds \\ &\equiv U_n(t) + \int_0^t n A_n(ns) Z_n(s) ds. \end{aligned}$$

Set

$$(3.4) \quad B_n(t) = \int_0^t n A_n(ns) ds.$$

Then

$$(3.5) \quad Z_n(t)e^{-B_n(t)} = 1 + \int_0^t e^{-B_n(s)} dU_n(s).$$

Note that U_n is (at least) a local martingale. However, since B_n is continuous and U_n has bounded variation, no special definition of the stochastic integral is needed.

3.1 Theorem Let $D_n(t) = \int_0^t \sum_{k=0}^{\infty} (k-1)^2 \Lambda_k^{(n)}(ns) ds$. Suppose that $(B_n, D_n) \Rightarrow (B, D)$ and that there exist α_n satisfying $\alpha_n/n \rightarrow 0$ and

$$(3.6) \quad \lim_{n \rightarrow \infty} \int_0^t \sum_{k > \alpha_n} (k-1)^2 \Lambda_k^{(n)}(ns) ds = 0 \text{ a.s.}$$

for all $t > 0$. Then Z_n converges in distribution to the unique solution of

$$(3.7) \quad Z(t) = e^{B(t)} \left[1 + W \left(\int_0^t e^{-2B(s)} Z(s) dD(s) \right) \right],$$

where W is standard Brownian motion independent of B and D .

Proof. We begin by verifying the uniqueness of the solution of (3.7). Let $\gamma(t)$ satisfy

$$(3.8) \quad \int_0^{\gamma(t)} e^{-B(s)} dD(s) = t$$

for $t < \Gamma = \int_0^\infty e^{-B(s)} dD(s)$. Then

$$(3.9) \quad \begin{aligned} Z(\gamma(t)) &= e^{B(\gamma(t))} \left[1 + W \left(\int_0^{\gamma(t)} e^{-2B(s)} Z(s) dD(s) \right) \right] \\ &= e^{B(\gamma(t))} \left[1 + W \left(\int_0^t e^{-2B(\gamma(s))} Z(\gamma(s)) ds \right) \right]. \end{aligned}$$

It follows that $Z(t) = e^{B(t)} \tilde{Z}(\int_0^t e^{-B(s)} dD(s))$, where \tilde{Z} is the unique solution of

$$(3.10) \quad \tilde{Z}(u) = 1 + W \left(\int_0^u \tilde{Z}(s) ds \right).$$

Note that \tilde{Z} is the diffusion arising in Theorem 1.3, with $\sigma^2 = 1$. See Theorem 1.1 of Chapter 6. By Corollary 1.9 of Chapter 3, we may assume for all $t > 0$,

$$(3.11) \quad \begin{aligned} \lim_{n \rightarrow \infty} \sup_{s \leq t} |B_n(s) - B(s)| &= 0 \quad \text{a.s.}, \\ \lim_{n \rightarrow \infty} \sup_{s \leq t} |D_n(s) - D(s)| &= 0 \quad \text{a.s.}, \\ \lim_{n \rightarrow \infty} \int_0^t \sum_{k > a_n} (k-1)^2 \Lambda_k^n(ns) ds &= 0 \quad \text{a.s.} \end{aligned}$$

Since the Λ_k^n are independent of the Y_k , it is enough to prove the theorem under the assumption that the Λ_k^n are deterministic and satisfy (3.11). This amounts to conditioning on the Λ_k^n . With this assumption we have that

$$(3.12) \quad E[Z_n(t)] = e^{B_n(t)}$$

and $V_n \equiv \int_0^t e^{-B_n(s)} dU_n(s)$ is a square integrable martingale with

$$(3.13) \quad \langle V_n, V_n \rangle_t = \int_0^t e^{-2B_n(s)} Z_n(s) dD_n(s).$$

Fix $T > 0$. Let $\tau_n(t)$ satisfy

$$(3.14) \quad \int_0^{\tau_n(t)} e^{-2B_n(s)} Z_n(s) dD_n(s) = t$$

for $t < \Gamma_n = \int_0^T e^{-2B_n(s)} Z_n(s) dD_n(s)$, let W_0 be a Brownian motion independent of all the other processes (we can always enlarge the sample space to obtain W_0), and define

$$(3.15) \quad W_n(t) = \begin{cases} V_n(\tau_n(t)), & t < \Gamma_n, \\ W_0(t - \Gamma_n) + V_n(T), & \Gamma_n \leq t < \infty. \end{cases}$$

Then W_n is a square integrable martingale with $\langle W_n, W_n \rangle_t = t$, and

$$(3.16) \quad Z_n(t) e^{-B_n(t)} = 1 + W_n \left(\int_0^t e^{-2B_n(s)} Z_n(s) dD_n(s) \right), \quad t \leq T.$$

Since $\langle W_n, W_n \rangle_t = t$ for all n , to show that $W_n \Rightarrow W$ using Theorem 1.4(b) of Chapter 7 we need only verify

$$(3.17) \quad \lim_{n \rightarrow \infty} E \left[\sup_{s \leq t} |W_n(s) - W_n(s-)|^2 \right] = 0$$

for all $t > 0$. Setting $b_n = \sup_{0 \leq s \leq T} |B_n(s)|$, we have

$$\begin{aligned} (3.18) \quad & E \left[\sup_s |W_n(s) - W_n(s-)|^2 \right] \\ &= E \left[\sup_{0 \leq s \leq T} |V_n(s) - V_n(s-)|^2 \right] \\ &\leq e^{2b_n} \alpha_n^2 n^{-2} + e^{2b_n} E \left[\sum_{k > \alpha_n} \frac{(k-1)^2}{n^2} Y_k \left(n^2 \int_0^T \Lambda_k^{(n)}(ns) Z_n(s) ds \right) \right] \\ &= e^{2b_n} \alpha_n^2 n^{-2} + e^{2b_n} \int_0^T \sum_{k > \alpha_n} (k-1)^2 \Lambda_k^{(n)}(ns) E[Z_n(s)] ds \\ &\leq e^{2b_n} \alpha_n^2 n^{-2} + e^{3b_n} \int_0^T \sum_{k > \alpha_n} (k-1)^2 \Lambda_k^{(n)}(ns) ds. \end{aligned}$$

The right side goes to zero by (3.11) and the hypotheses on α_n . Since $Z_n(t)e^{-B_n(t)}$ is a martingale,

$$(3.19) \quad P \left\{ \sup_{s \leq T} Z_n(s)e^{-B_n(s)} > z \right\} \leq z^{-1},$$

and relative compactness for $\{Z_n\}$ (restricted to $[0, T]$) follows easily from (3.16) and the relative compactness for $\{W_n\}$. If a subsequence $\{Z_{n_k}\}$ converges in distribution to Z , then a subsequence of

$$\{(W_{n_k}, \int_0^t \exp \{-2B_{n_k}(s)\} Z_{n_k}(s) dD_{n_k}(s))\}$$

converges in distribution to $(W, \int_0^t \exp \{-2B(s)\} Z(s) dD(s))$, and (3.16) and the continuous mapping theorem (Corollary 1.9 of Chapter 3) imply

$$(3.20) \quad Z(t) = e^{B(t)} \left[1 + W \left(\int_0^t e^{-2B(s)} Z(s) dD(s) \right) \right]$$

for $t \leq T$. The theorem now follows from the uniqueness of the solution of (3.20) and the fact that T is arbitrary. \square

3.2 Example Let $\zeta(t)$ be a standard Poisson process. Let $\Lambda_0^n(t) \equiv 1$, $\Lambda_2^n(t) \equiv 1 + n^{-1/2}(-1)^{\zeta(t)}$, and $\Lambda_k^n = 0$ for $k \neq 0, 2$. This gives

$$(3.21) \quad B_n(t) = \int_0^t n^{1/2}(-1)^{\zeta(ns)} ds$$

and

$$(3.22) \quad D_n(t) = \int_0^t (2 + n^{-1/2}(-1)^{\zeta(ns)}) ds.$$

Then $(B_n, D_n) \Rightarrow (B, D)$ where B is a standard Brownian motion and $D(t) = 2t$. (See Problem 3 of Chapter 7). The limit Z then satisfies

$$(3.23) \quad Z(t) = e^{B(t)} \left[1 + W \left(\int_0^t 2e^{-2B(s)} Z(s) ds \right) \right].$$

Note that B and $M(\cdot) = W(\int_0^\cdot 2e^{-2B(s)} Z(s) ds)$ are martingales with $\langle B \rangle_t = t$, $\langle B, M \rangle_t = 0$, and $\langle M \rangle_t = \int_0^t 2e^{-2B(s)} Z(s) ds$. For $f \in C_c^2(\mathbb{R})$ define $g(x, y) = f(e^x(1+y))$. Then by Itô's formula,

$$\begin{aligned} (3.24) \quad f(Z(t)) &= g(B(t), M(t)) \\ &= f(1) + \int_0^t g_x(B(s), M(s)) dB(s) \\ &\quad + \int_0^t g_y(B(s), M(s)) dM(s) \\ &\quad + \int_0^t (\frac{1}{2}g_{xx}(B(s), M(s)) + e^{-2B(s)} Z(s) g_{yy}(B(s), M(s))) ds \\ &= f(1) + \tilde{M}(t) \\ &\quad + \int_0^t (\frac{1}{2}Z(s)f'(Z(s)) + (Z(s) + \frac{1}{2}Z(s)^2)f''(Z(s))) ds, \end{aligned}$$

and we see that Z is a solution of the martingale problem for A with $Af(z) = \frac{1}{2}zf'(z) + (z + \frac{1}{2}z^2)f''(z)$. \square

4. BRANCHING MARKOV PROCESSES

We begin with an example of the type of process we are considering. Take the number of particles $\{N(t), t \geq 0\}$ in a collection to be a continuous-time Markov branching process; that is, each particle lives an exponentially distributed lifetime with some parameter α and at death is replaced by a random number of offspring, where the lifetimes and numbers of offspring are independent random variables. Note that N has generator (on an appropriate domain)

$$(4.1) \quad Af(k) = \sum_l \alpha k p_l (f(k-1+l) - f(k))$$

where p_l is the probability that a particle has l offspring.

In addition, we assume each particle has a location in \mathbb{R}^d and moves as a Brownian motion with generator $\frac{1}{2}\Delta$, and the motions are taken to be indepen-

dent and independent of the lifetimes and numbers of offspring. We also assume that the offspring are initially given their parent's location at the time of birth. The state space for the process can be described as

$$(4.2) \quad E = \{(k, x_1, x_2, \dots, x_k) : k = 0, 1, 2, \dots, x_i \in \mathbb{R}^d\};$$

that is, k is the number of particles and the x_i are the locations. However, we modify this description later.

Of course it may not be immediately clear that such a process exists or that the above conditions uniquely determine the behavior of the process. Consequently, in order to make the above description precise, we specify the generator for the process on functions of the form $f(k, x_1, x_2, \dots, x_k) = \prod_{i=1}^k g(x_i)$ where $g \in \mathcal{D}(\Delta)$ and $\|g\| < 1$. If the particles were moving without branching, then the generator would be

$$(4.3) \quad A_1 = \left\{ \left(\prod_{i=1}^k g(x_i), \frac{1}{2} \sum_{j=1}^k \Delta g(x_j) \prod_{i \neq j} g(x_i) \right) : g \in \mathcal{D}(\Delta), \|g\| < 1 \right\}.$$

If there were branching but not motion, then the generator would be

$$(4.4) \quad A_2 = \left\{ \left(\prod_{i=1}^k g(x_i), \sum_{j=1}^k \alpha(\varphi(g(x_j)) - g(x_j)) \prod_{i \neq j} g(x_i) \right) : \|g\| < 1 \right\}$$

where $\varphi(z) = \sum_{l=0}^{\infty} p_l z^l$, that is, φ is the generating function of the offspring distribution. The assumption that the motion and branching are independent suggests that the desired generator is $A_1 + A_2$.

More generally we consider processes in which the particles are located in a separable, locally compact, metric space E_0 , move as Feller processes with generator B , die with a location dependent intensity $\alpha \in \bar{C}(E_0)$ (that is, a particle located at x at time t dies before time $t + \Delta t$ with probability $\alpha(x) \Delta t + o(\Delta t)$), and in which the offspring distribution is location dependent (that is, a particle that dies at x produces l offspring with probability $p_l(x)$). We assume that $p_l \in \bar{C}(E_0)$ and define

$$(4.5) \quad \varphi(z) = \sum_l p_l z^l, \quad |z| \leq 1.$$

Note that for fixed z , $\varphi(z) \in \bar{C}(E_0)$. We also assume $\sum_l l p_l \in \bar{C}(E_0)$, that is, the mean number of offspring is finite and depends continuously on the location of the parent. We denote $(\partial/\partial z)\varphi(z)$ by $\varphi'(z)$. In particular $\varphi'(1) = \sum_l l p_l$.

The order of the x_i in the state $(k, x_1, x_2, \dots, x_k)$ is not really important and causes some notational difficulty. Consequently, we take for the state space, not (4.2), but the space of measures

$$(4.6) \quad E = \left\{ \sum_{i=1}^k \delta_{x_i} : k = 0, 1, 2, \dots, x_i \in E_0 \right\}$$

where δ_x denotes the measure with mass one at x . Of course E is a subset of the space $\mathcal{M}(E_0)$ of finite, positive, Borel measures on E_0 . We topologize

$\mathcal{M}(E_0)$ (and hence E) with the weak topology. In other words, $\lim_{n \rightarrow \infty} \mu_n = \mu$ if and only if

$$(4.7) \quad \lim_{n \rightarrow \infty} \int f d\mu_n = \int f d\mu$$

for every $f \in \bar{\mathcal{C}}(E_0)$. The weak topology is metrizable by a natural extension of the Prohorov metric (Problem 6). Note that in E , convergence in the weak topology just means the number and locations of the particles converge.

Let $\bar{\mathcal{C}}^+(E_0) = \{f \in \bar{\mathcal{C}}(E_0) : \inf f > 0\}$. Define

$$(4.8) \quad \langle g, \mu \rangle = \int g d\mu, \quad g \in \bar{\mathcal{C}}(E_0), \quad \mu \in E,$$

and note that for $\mu = \sum_{i=1}^k \delta_{x_i}$ and $g \in \bar{\mathcal{C}}^+(E_0)$,

$$(4.9) \quad \prod_{i=1}^k g(x_i) = e^{\langle \log g, \mu \rangle}.$$

Extend B to the span of the constants and the original domain, by defining $B1 = 0$, so that the martingale problem for $\{(f, Bf) : f \in \mathcal{D}(B) \cap \bar{\mathcal{C}}^+(E_0)\}$ is still well-posed. With reference to (4.3) and (4.4), the generator for the process will be

$$(4.10) \quad A = \left\{ \left(e^{\langle \log g, \mu \rangle}, e^{\langle \log g, \mu \rangle} \left\langle \frac{Bg + \alpha(\varphi(g) - g)}{g}, \mu \right\rangle \right) : g \in \mathcal{D}(B) \cap \bar{\mathcal{C}}^+(E_0), \|g\| < 1 \right\}.$$

Let $\{S(t)\}$ denote the semigroup generated by B . By Lemma 3.4 of Chapter 4, if X is a solution of the martingale problem for A , then for g satisfying the conditions in (4.10)

$$(4.11) \quad \exp \{ \langle \log S(T-t)g, X(t) \rangle \} - \int_0^t \exp \{ \langle \log S(T-s)g, X(s) \rangle \} \\ \times \left\langle \frac{\alpha(\varphi(S(T-s)g) - S(t-s)g)}{S(T-s)g}, X(s) \right\rangle ds$$

is a martingale for $0 \leq t \leq T$. Note that

$$(4.12) \quad \frac{d}{dt} \exp \{ \langle \log S(T-t)g, \mu \rangle \} = \exp \{ \langle \log S(T-t)g, \mu \rangle \} \\ \times \left\langle -\frac{BS(T-t)g}{S(T-t)g}, \mu \right\rangle.$$

4.1 Lemma Let X be a solution of the martingale problem for (A, δ_μ) . Then setting $|X(t)| = \langle 1, X(t) \rangle$ (i.e., $|X|$ is the total population size),

$$(4.13) \quad E[|X(t)|] \leq |\mu| \exp\{t\|\alpha(\varphi'(1) - 1)\|\},$$

and

$$(4.14) \quad P\left\{\sup_t |X(t)| \exp\{-t\|\alpha(\varphi'(1) - 1)\|\} > x\right\} \leq \frac{|\mu|}{x}.$$

Proof. Let $\lambda > 0$ be a constant. Take $g = e^{-\lambda}$. Then

$$(4.15) \quad M_\lambda(t) \equiv e^{-\lambda|X(t)|} - \int_0^t e^{-\lambda|X(s)|} \langle \alpha(\varphi(e^{-\lambda}) - e^{-\lambda})e^\lambda, X(s) \rangle ds$$

is a martingale, and hence

$$(4.16) \quad E[e^{-\lambda|X(t)|}] = e^{-\lambda|\mu|} + \int_0^t E[e^{-\lambda|X(s)|} \langle \alpha(\varphi(e^{-\lambda}) - e^{-\lambda})e^\lambda, X(s) \rangle] ds,$$

so

$$\begin{aligned} (4.17) \quad E[e^{-\lambda|X(t)|}\lambda|X(t)|] &\leq E[1 - e^{-\lambda|X(t)|}] \\ &= 1 - e^{-\lambda|\mu|} + \int_0^t E[e^{-\lambda|X(s)|} \langle \alpha(1 - e^\lambda\varphi(e^{-\lambda})), X(s) \rangle] ds \\ &\leq 1 - e^{-\lambda|\mu|} + \int_0^t E[e^{-\lambda|X(s)|} \langle \alpha(\varphi'(1) - \varphi(e^{-\lambda}))\lambda, X(s) \rangle] ds \\ &\leq 1 - e^{-\lambda|\mu|} + \int_0^t \|\alpha(\varphi'(1) - \varphi(e^{-\lambda}))\| E[e^{-\lambda|X(s)|}\lambda|X(s)|] ds. \end{aligned}$$

By Gronwall's inequality

$$\begin{aligned} (4.18) \quad E[\exp\{-\lambda|X(t)|\}|X(t)|] \\ &\leq \lambda^{-1}(1 - \exp\{-\lambda|\mu|\}) \exp\{t\|\alpha(\varphi'(1) - \varphi(e^{-\lambda}))\|\}. \end{aligned}$$

Letting $\lambda \rightarrow 0$ gives (4.13).

Let

$$(4.19) \quad M(t) = \lim_{\lambda \rightarrow 0} \lambda^{-1}(1 - M_\lambda(t)) = |X(t)| - \int_0^t \langle \alpha(\varphi'(1) - 1), X(s) \rangle ds.$$

From (4.13) it follows that the convergence in (4.19) is in L^1 and hence M is a martingale and

$$(4.20) \quad |X(t)| \exp\{-t\|\alpha(\varphi'(1) - 1)\|\}$$

is a supermartingale. Consequently (4.14) follows from Proposition 2.16 of Chapter 2. \square

4.2 Theorem Let B , α , and φ be as above, and let A be given by (4.10). Then for $v \in \mathcal{P}(E)$ the martingale problem for (A, v) has a unique solution.

Proof. Existence is considered in Problem 7. To obtain uniqueness, we apply Theorem 4.2 of Chapter 4. Let X be a solution of the martingale problem for (A, v) and define

$$(4.21) \quad u(t, g) = E[\exp \{\langle \log g, X(t) \rangle\}].$$

Note that $u(t, \cdot)$ is a bounded continuous function on $E' = \{g \in \bar{C}^+(E_0) : \|g\| < 1\}$. For $H \in \bar{C}(E')$ define

$$(4.22) \quad \Gamma H(g) = \lim_{\varepsilon \rightarrow 0+} \varepsilon^{-1} (H(e^{-\alpha\varepsilon}g + (1 - e^{-\alpha\varepsilon})\varphi(g)) - H(g))$$

if the limit exists uniformly in g . Observe that Γ is dissipative, since $\Gamma H = \lim_{\varepsilon \rightarrow 0+} \varepsilon^{-1} (Q_\varepsilon - I)H$ where Q_ε is a contraction. We claim that $u(t, \cdot) \in \mathcal{D}(\Gamma)$ and

$$(4.23) \quad \Gamma u(t, g) = E \left[\exp \{\langle \log g, X(t) \rangle\} \left\langle \frac{\alpha(\varphi(g) - g)}{g}, X(t) \right\rangle \right].$$

To see this write

$$\begin{aligned} (4.24) \quad & \varepsilon^{-1} (u(t, (e^{-\alpha\varepsilon}g + (1 - e^{-\alpha\varepsilon})\varphi(g))) - u(t, g)) \\ &= \varepsilon^{-1} \int_0^t E \left[\exp \{\langle \log (e^{-\alpha\varepsilon}g + (1 - e^{-\alpha\varepsilon})\varphi(g)), X(s) \rangle\} \right. \\ & \quad \times \left. \left\langle \frac{\alpha(\varphi(g) - g)}{g + (e^{\alpha s} - 1)\varphi(g)}, X(s) \right\rangle \right] ds. \end{aligned}$$

The expression inside the expectation on the right of (4.24) is dominated by

$$(4.25) \quad \|\alpha(\varphi(g) - g)\| |X(t)| \leq 2\|\alpha\| |X(t)|,$$

so by (4.13) and the dominated convergence theorem it is enough to show

$$\begin{aligned} (4.26) \quad & \left| \exp \{\langle \log (e^{-\alpha\varepsilon}g + (1 - e^{-\alpha\varepsilon})\varphi(g)), \mu \rangle\} \left\langle \frac{\alpha(\varphi(g) - g)}{g + (e^{\alpha s} - 1)\varphi(g)}, \mu \right\rangle \right. \\ & \quad \left. - \exp \{\langle \log g, \mu \rangle\} \left\langle \frac{\alpha(\varphi(g) - g)}{g}, \mu \right\rangle \right| \end{aligned}$$

converges to zero as $s \rightarrow 0$ uniformly in g for each $\mu \in E$. To check that this convergence holds, calculate the derivative of (4.26) and show that it is bounded. Finally, define

$$(4.27) \quad \mathcal{S}(t)H(g) = H(S(t)g)$$

and note that $\{\mathcal{S}(t)\}$ is a contraction semigroup on $\bar{C}(E')$. The fact that (4.11) is a martingale gives (for $T = t$)

$$(4.28) \quad u(t, g) = E[\exp \{\langle \log S(t)g, X(0) \rangle\}] + \int_0^t \Gamma u(s, S(t-s)g) ds \\ = \mathcal{S}(t)u_0(g) + \int_0^t \mathcal{S}(t-s)\Gamma u(s, g) ds.$$

By Proposition 5.4 of Chapter 1 there is at most one solution of this equation, so $E[\exp \{\langle \log g, X(t) \rangle\}]$ is uniquely determined. Since the linear space generated by functions of the form $\exp \{\langle \log g, \mu \rangle\}$ for $g \in \bar{C}^+(E_0)$ is an algebra that separates points in $\mathcal{M}(E_0)$, it follows that the distribution of $X(t)$ is determined, and since v was arbitrary, the solution of the martingale problem for (A, v) is unique by Theorem 4.2 of Chapter 4. \square

We now consider a sequence of branching Markov processes X_n , $n = 1, 2, 3, \dots$, with death intensities α_n , and offspring generating functions φ_n , in which the particles move as Feller processes in E_0 with generators B_n , extended as before. We define

$$(4.29) \quad Z_n = n^{-1}X_n.$$

Note that the state space for Z_n is

$$(4.30) \quad E_n = \left\{ \mu \in \mathcal{M}(E_0) : \mu = n^{-1} \sum_{i=1}^k \delta_{x_i}, x_i \in E_0 \right\},$$

and that Z_n is a solution of the martingale problem for

(4.31)

$$A_n = \left\{ \left(\exp \{\langle n \log g, \mu \rangle\}, \exp \{\langle n \log g, \mu \rangle\} \left\langle \frac{nB_n g + n\alpha_n(\varphi_n(g) - g)}{g}, \mu \right\rangle \right) : \right. \\ \left. g \in \mathcal{D}(B_n), \inf g > 0, \|g\| < 1 \right\}.$$

Define

$$(4.32) \quad F_n(h) = n\alpha_n(1 - \varphi_n(1 - n^{-1}h) - n^{-1}h)$$

for $h \in \bar{C}^+(E_0)$, $\|h\| < n$. If $h \in \mathcal{D}(B_n) \cap \bar{C}^+(E_0)$ and $\|h\| < n$, then setting $g = 1 - n^{-1}h$ we have

$$(4.33) \quad \left(\exp \{\langle n \log (1 - n^{-1}h), \mu \rangle\}, \exp \{\langle n \log (1 - n^{-1}h), \mu \rangle\} \times \left\langle \frac{-B_n h - F_n(h)}{1 - n^{-1}h}, \mu \right\rangle \right) \in A_n.$$

For simplicity, we assume E_0 is compact (otherwise replace E_0 by its one-point compactification).

4.3 Theorem Let E_0 be compact. Let B be the generator for a Feller semi-group extended as before, and let $F(\cdot): \bar{C}^+(E_0) \rightarrow \bar{C}(E_0)$. Suppose

$$(4.34) \quad \text{ex-lim}_{n \rightarrow \infty} B_n = B,$$

$$(4.35) \quad \sup_n \|\alpha_n(\varphi'_n(1) - 1)\| < \infty,$$

and for each $k > 0$,

$$(4.36) \quad \lim_{n \rightarrow \infty} \sup_{\substack{h \in \bar{C}^+(E_0) \\ \|h\| \leq k}} \|F_n(h) - F(h)\| = 0.$$

If $\{Z_n(0)\}$ has limiting distribution $v \in \mathcal{P}(\mathcal{M}(E_0))$, then $Z_n \Rightarrow Z$ where Z is the unique solution of the martingale problem for (A, v) with

$$(4.37)$$

$$A = \{\exp\{-\langle h, \mu \rangle\}, \exp\{-\langle h, \mu \rangle\} \langle -Bh - F(h), \mu \rangle : h \in \mathcal{D}(B) \cap \bar{C}^+(E_0)\}.$$

4.4 Remark From Taylor's formula it follows that

$$(4.38) \quad F_n(h) = \alpha_n(\varphi'_n(1) - 1)h - n^{-1}\alpha_n \int_0^1 (1-v)\varphi''_n(1-n^{-1}hv) dv \cdot h^2,$$

so typically $F(h) = ah - bh^2$, where

$$(4.39) \quad a = \lim_{n \rightarrow \infty} \alpha_n(\varphi'_n(1) - 1)$$

and

$$(4.40) \quad b = \lim_{n \rightarrow \infty} \frac{\alpha_n}{2n} \varphi''_n(1).$$

In particular, if $\alpha_n \equiv n$ and $\varphi_n(z) \equiv \frac{1}{2} + \frac{1}{2}z^2$, then $F(h) = -\frac{1}{2}h^2$. Since the integral expression multiplying h^2 in (4.38) is decreasing in h , (4.35) and the existence of the limit in (4.36) imply there exist positive constants C_k , $k = 1, 2, 3$, such that

$$(4.41) \quad -(C_1 h + C_2 h^2) \leq F(h) \leq C_3 h. \quad \square$$

Proof. We apply Corollary 8.16 of Chapter 4. For $h \in \mathcal{D}(B) \cap \bar{C}^+(E_0)$, there exist $h_n \in \mathcal{D}(B_n) \cap \bar{C}^+(E_0)$ such that $\lim_{n \rightarrow \infty} h_n = h$ and $\lim_{n \rightarrow \infty} B_n h_n = Bh$.

For n sufficiently large, $\|h_n\| < n$ and $h_n \geq \frac{1}{2} \inf h \equiv \varepsilon > 0$. Consequently, taking $g = (1 - n^{-1}h_n)$ in A_n ,

$$(4.42) \quad \begin{aligned} & \sup_{\mu \in E_n} |\exp \{\langle n \log (1 - n^{-1}h_n), \mu \rangle\} - \exp \{-\langle h, \mu \rangle\}| \\ & \leq \sup_{\mu \in E_n} \exp \{-\varepsilon \langle 1, \mu \rangle\} \langle 1, \mu \rangle \|n \log (1 - n^{-1}h_n) - h\| \\ & \leq \varepsilon^{-1} \|n \log (1 - n^{-1}h_n) - h\| \end{aligned}$$

and

$$(4.43) \quad \begin{aligned} & \sup_{\mu \in E_n} \left| \exp \{\langle n \log (1 - n^{-1}h_n), \mu \rangle\} \left\langle \frac{-B_n h_n - F_n(h_n)}{1 - n^{-1}h_n}, \mu \right\rangle \right. \\ & \quad \left. - e^{-\langle h, \mu \rangle} \langle -Bh - F(h), \mu \rangle \right| \\ & \leq \sup_{\mu \in E_n} \exp \{-\varepsilon \langle 1, \mu \rangle\} \langle 1, \mu \rangle \left(\left\| \frac{-B_n h_n - F_n(h_n)}{1 - n^{-1}h_n} + Bh + F(h) \right\| \right. \\ & \quad \left. + \| -Bh - F(h) \| \|n \log (1 - n^{-1}h_n) - h\| \right). \end{aligned}$$

Therefore, condition (f) of Corollary 8.7 in Chapter 4 is satisfied with $G_n = E_n$.

The compact containment condition follows from (4.14), and it remains only to verify uniqueness for the martingale problem. Uniqueness can be obtained by the same argument used in the proof of Theorem 4.2, in this case defining

$$(4.44) \quad \Gamma H(h) = \lim_{\varepsilon \rightarrow 0^+} \varepsilon^{-1} (H((h + \varepsilon F(h)) \vee 0) - H(h)).$$

The estimates in (4.41) ensure that the limit

$$(4.45) \quad \begin{aligned} & \Gamma E[\exp \{-\langle h, X(t) \rangle\}] \\ & = \lim_{\varepsilon \rightarrow 0^+} \varepsilon^{-1} E[\exp \{-\langle (h + \varepsilon F(h)) \vee 0, X(t) \rangle\} - \exp \{-\langle h, X(t) \rangle\}] \end{aligned}$$

exists uniformly in h . □

5. PROBLEMS

- State and prove an analogue of Theorem 1.3 for a Galton–Watson process in independent random environments. That is, let η_1, η_2, \dots be independent and uniform on $[0, 1]$. Suppose the ξ_k^n are conditionally independent given $\mathcal{F} = \sigma(\eta_i, i = 1, 2, \dots)$ and $P\{\xi_k^n = l | \mathcal{F}\} = P_l(\eta_n)$. Define Z_n as in (1.1).

Consider a sequence of such processes $\{Z^{(m)}\}$ determined by $\{P_i^{(m)}\}$ and $Z^{(m)}(0)$, and give conditions under which $Z^{(m)}([m \cdot])/m$ converges in distribution.

2. Let $\{X_n\}$ be as in Section 2. Represent $(Z_1^{(n)}, Z_2^{(n)})$ using multiple random time changes (see Chapter 6), and use the representation to prove the convergence of $\{X_n\}$.

3. Show that $D = C_c^\infty([0, \infty) \times (-\infty, \infty))$ is a core for A given by (2.8).

Hint: Begin by looking for solutions of $u_t = Au$ of the form $e^{-at}x \sin(b(t)x + cy)$ and $e^{-at}x \cos(b(t)x + cy)$. Show that the bounded pointwise closure of $A|_D$ contains (f, Af) for $f = e^{-ax} \sin(bx + cy)$ and $f = e^{-ax} \cos(bx + cy)$, and the bp-closure of $\mathcal{R}(\lambda - A|_D)$ contains $e^{-ax} \sin(bx + cy)$ and $e^{-ax} \cos(bx + cy)$, and hence all of $\bar{C}([0, \infty) \times (-\infty, \infty))$. See Chapter 1, Section 3.

4. In (3.2), assume $\sum k \int_0^t \Lambda_k(s) ds < \infty$ a.s. for all $t > 0$.

- (a) Show that the solution of (3.2) exists for all time.
 (b) Show that the solution of (3.2) satisfies (3.1).

5. (a) In (3.23) suppose B is a Brownian motion with generator $\frac{1}{2}af'' + bf'$. Show that Z is a Markov process and find its generator.

- (b) In (3.23) suppose B is a diffusion process with generator $\frac{1}{2}\sigma^2(x)f'' + m(x)f'$. Show that (Z, B) is a Markov process and find its generator.

6. Let $\mathcal{M}(E_0)$ be the space of finite, positive Borel measures on a metric space E_0 . Let $|\mu| = \mu(E_0)$ and define $\tilde{\rho}(\mu, v) = \rho(\mu/|\mu|, v/|v|) + ||\mu| - |v||$ where ρ is the Prohorov metric. Show that $\tilde{\rho}$ is a metric on $\mathcal{M}(E_0)$ giving the weak topology and that $(\mathcal{M}(E_0), \tilde{\rho})$ is complete and separable if (E_0, r) is.

7. Let B , α , and φ be as in (4.10), and let $\varepsilon > 0$. Let $B_\varepsilon \equiv B(I - \varepsilon B)^{-1}$ be the Yosida approximation of B and let $|\mu| = \mu(E_0)$, that is, the total number of particles. Set

$$A_\varepsilon = \left\{ \left(e^{\langle \log g, \mu \rangle}, e^{\langle \log g, \mu \rangle} \left\langle \frac{B_\varepsilon g + \alpha e^{-\varepsilon|\mu|}(\varphi(g) - g)}{g}, \mu \right\rangle \right): g \in \bar{C}(E_0), \inf g > 0, \|g\| < 1 \right\}.$$

- (a) Show that A_ε extends to an operator of the form of (2.1) in Chapter 4 and hence for each $\mu \in E$, the martingale problem for $(A_\varepsilon, \delta_\mu)$ has a unique solution. Describe the behavior of this process.
 (b) Let $\mu \in E$ and let X_ε be a solution of the martingale problem for $(A_\varepsilon, \delta_\mu)$ with sample paths in $D_E[0, \infty)$. Show that $\{X_\varepsilon, 0 < \varepsilon < 1\}$ is relatively compact and that any limit in distribution of a sequence

$\{X_{\varepsilon_n}\}$, $\varepsilon_n \rightarrow 0$, is a solution of the martingale problem for (A, δ_μ) , A given by (4.10).

8. For X_n and Y_n defined by (2.4) and (2.5) show that X_n and $Y_n e^{nrt}$ are martingales.

6. NOTES

For a general introduction to branching processes see Athreya and Ney (1972). The diffusion approximation for the Galton–Watson process was formulated by Feller (1951) and proved by Jířina (1969) and Lindvall (1972). These results have been extended to the age-dependent case by Jagers (1971). Theorem 1.4 is due to Grimvall (1974). The approach taken here is from Helland (1978). Work of Lamperti (1967a) is closely related.

Theorem 2.1 is from Kurtz (1978b) and has been extended by Joffe and Metivier (1984). Keiding (1975) formulated a diffusion approximation for a Galton–Watson process in a random environment that was made rigorous by Helland (1981). The Galton–Watson analogue of Theorem 3.1 is in Kurtz (1978b). See also Barbour (1976).

Branching Markov processes were extensively studied by Ikeda, Nagasawa, and Watanabe (1968, 1969). The measure diffusion approximation was given by Watanabe (1968) and Dawson (1975). Also see Wang (1982b). The limiting measure-valued process has been studied by Dawson (1975, 1977, 1979), Dawson and Hochberg (1979), and Wang (1982b).

10 | GENETIC MODELS

Diffusion processes have been used to approximate discrete stochastic models in population genetics for over fifty years. In this chapter we describe several such models and show how the results of earlier chapters can be used to justify these approximations mathematically.

In Section 1 we give a fairly careful formulation of the so-called Wright–Fisher model, defining the necessary terminology from genetics; we then obtain a diffusion process as a limit in distribution. Specializing to the case of two alleles in Section 2, we give three applications of this diffusion approximation, involving stationary distributions, mean absorption times, and absorption probabilities. Section 3 is concerned with more complicated genetic models, in which the gene-frequency process may be non-Markovian. Nevertheless limiting diffusions are obtained as an application of Theorem 7.6 of Chapter 1. Finally, in Section 4, we consider the infinitely-many-neutral-alleles model with uniform mutation, and we characterize the stationary distribution of the limiting (measure-valued) diffusion process. We conclude with a derivation of Ewens’ sampling formula.

1. THE WRIGHT-FISHER MODEL

We begin by introducing a certain amount of terminology from population genetics.

Every organism is initially, at the time of conception, just a single cell. It is this cell, called a *zygote* (and others formed subsequently that have the same genetic makeup), that contains all relevant genetic information about an individual and influences that of its offspring. Thus, when discussing the genetic composition of a population, it is understood that by the genetic properties of an individual member of the population one simply means the genetic properties of the zygote from which the individual developed.

Within each cell are a certain fixed number of *chromosomes*, threadlike objects that govern the inheritable characteristics of an organism. Arranged in linear order at certain positions, or *loci*, on the chromosomes, are *genes*, the fundamental units of heredity. At each locus there are several alternative types of genes that can occur; the various alternatives are called *alleles*.

We restrict our attention to *diploid* organisms, those for which the chromosomes occur in *homologous* pairs, two chromosomes being homologous if they have the same locus structure. An individual's genetic makeup with respect to a particular locus, as indicated by the unordered pair of alleles situated there (one on each chromosome), is referred to as its *genotype*. Thus, if there are r alleles, A_1, \dots, A_r , at a given locus, then there are $r(r + 1)/2$ possible genotypes, $A_i A_j$, $1 \leq i \leq j \leq r$.

We also limit our discussion to *monoecious* populations, those in which each individual can act as either a male or a female parent. While many populations (e.g., plants) are in fact monoecious, this is mainly a simplifying assumption. Several of the problems at the end of the chapter deal with models for *dioecious* populations, those in which individuals can act only as male or as female parents.

To describe the Wright-Fisher model, we first propose a related model. Let A_1, \dots, A_r be the various alleles at a particular locus in a population of N adults. We assume, in effect, that generations are nonoverlapping. Let P_{ij} be the (relative) frequency of $A_i A_j$ genotypes just prior to reproduction, $1 \leq i \leq j \leq r$. Then

$$(1.1) \quad p_i = P_{ii} + \frac{1}{2} \sum_{j: j > i} P_{ij} + \frac{1}{2} \sum_{j: j < i} P_{ji}$$

is the frequency of the allele A_i , $1 \leq i \leq r$.

For our purposes, the reproductive process can be roughly described as follows. Each individual has a large number of *germ* cells, cells of the same genotype (neglecting mutation) as that of the zygote. These germ cells split into *gametes*, cells containing one chromosome from each homologous pair in the original cell, thus half the usual number. We assume that the gametes are produced without *fertility* differences, that is, that all genotypes have equal

probabilities of transmitting gametes in this way. The gametes then fuse at random, forming the zygotes of the next generation. We suppose that the number of such zygotes is (effectively) infinite, and so the genotypic frequencies among zygotes are $(2 - \delta_{ij})p_i p_j$, where δ_{ij} is the Kronecker delta. These are the so-called *Hardy-Weinberg proportions*.

Typically, certain individuals have a better chance than others of survival to reproductive age. Letting w_{ij} denote the *viability* of $A_i A_j$ individuals, that is, the relative likelihood that an $A_i A_j$ zygote will survive to maturity, we find that, after taking into account this viability selection, the genotypic frequencies become

$$(1.2) \quad P_{ij}^* = \frac{(2 - \delta_{ij})w_{ij}p_i p_j}{\sum_{k \leq i} (2 - \delta_{ki})w_{ki}p_k p_i},$$

and the allelic frequencies have the form

$$(1.3) \quad p_i^* = \frac{\sum_j w_{ij}p_i p_j}{\sum_{k, l} w_{kl}p_k p_l},$$

where $w_{ji} \equiv w_{ij}$ for $1 \leq i < j \leq r$. The population size remains infinite.

We next consider the possibility of mutations. Letting u_{ij} denote the probability that an A_i gene mutates to an A_j gene ($u_{ii} \equiv 0$), and assuming that the two genes carried by an individual mutate independently, we find that the genotypic frequencies after mutation are given by

$$(1.4) \quad P_{ij}^{**} = (1 - \frac{1}{2}\delta_{ij}) \sum_{k \leq i} (u_{ki}^* u_{ij}^* + u_{kj}^* u_{ii}^*) P_{ki}^*,$$

where

$$(1.5) \quad u_{ij}^* = \left(1 - \sum_k u_{ik}\right)\delta_{ij} + u_{ij},$$

the latter denoting the probability that an A_i gene in a zygote appears as A_j in a gamete. The corresponding allelic frequencies have the form

$$(1.6) \quad p_i^{**} = \sum_k u_{ki}^* p_k^*,$$

as shown by the calculation

$$\begin{aligned} (1.7) \quad p_i^{**} &= \sum_j \frac{1}{2}(1 + \delta_{ij})P_{i \wedge j, i \vee j}^* \\ &= \frac{1}{2} \sum_j \sum_{k \leq i} (u_{ki}^* u_{ij}^* + u_{kj}^* u_{ii}^*) P_{ki}^* \\ &= \frac{1}{2} \sum_{k \leq i} (u_{ki}^* + u_{ii}^*) \frac{1}{2}(1 + \delta_{ki}) P_{k \wedge i, k \vee i}^* \\ &= \frac{1}{2} \sum_{k, l} (u_{ki}^* + u_{ii}^*) \frac{1}{2}(1 + \delta_{ki}) P_{k \wedge i, k \vee i}^* \end{aligned}$$

$$\begin{aligned}
 &= \sum_{k,l} u_{kl}^* \frac{1}{2}(1 + \delta_{kl}) P_{k+l,k+l}^* \\
 &= \sum_k u_{ki}^* p_k^*.
 \end{aligned}$$

Again, the population size remains infinite.

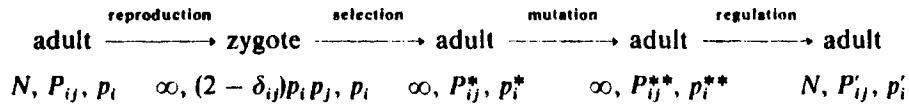
Finally, we provide for chance fluctuations in genotypic frequencies, known as *random genetic drift*, by reducing the population to its original size N through random sampling. The genotypic frequencies P'_{ij} in the next generation just prior to reproduction have the joint distribution specified by

$$(1.8) \quad (P'_{ij})_{i \leq j} \sim N^{-1} \text{ multinomial}(N, (P_{ij}^{**})_{i \leq j}).$$

This is simply a concise notation for the statement that $(NP'_{ij})_{i \leq j}$ has a multinomial distribution with sample size N and mean vector $(NP_{ij}^{**})_{i \leq j}$. In terms of probability generating functions,

$$(1.9) \quad E\left[\prod_{i \leq j} \zeta_{ij}^{NP'_{ij}}\right] = \left(\sum_{i \leq j} P_{ij}^{**} \zeta_{ij}\right)^N.$$

We summarize our description of the model in the following diagram:



Observe that (1.8), (1.4), (1.5), (1.2), and (1.1) define the transition function of a homogeneous Markov chain, in that the distribution of $(P'_{ij})_{i \leq j}$ is completely specified in terms of $(P_{ij})_{i \leq j}$. We have more to say about this chain in Section 3. For now we simply note that if the frequencies P_{ij}^{**} are in Hardy-Weinberg form, that is, if

$$(1.10) \quad P_{ij}^{**} = (2 - \delta_{ij})p_i^{**}p_j^{**}$$

for all $i \leq j$, then

$$\begin{aligned}
 (1.11) \quad E\left[\prod_i \tau_i^{2Np_i}\right] &= E\left[\prod_{i \leq j} (\tau_i \tau_j)^{NP_{ij}^{**}}\right] \\
 &= \left(\sum_{i \leq j} (2 - \delta_{ij})p_i^{**}p_j^{**}\tau_i \tau_j\right)^N \\
 &= \left(\sum_i p_i^{**}\tau_i\right)^{2N}
 \end{aligned}$$

by (1.9), implying that

$$(1.12) \quad (p'_1, \dots, p'_r) \sim (2N)^{-1} \text{ multinomial}(2N, (p_1^{**}, \dots, p_r^{**})).$$

One can check that (1.10) holds (for all $(P_{ij})_{i \leq j}$) in the absence of selection (i.e., $w_{ij} = 1$ for all $i \leq j$) and, more generally, when viabilities are multiplicative (i.e., there exist v_1, \dots, v_r such that $w_{ij} = v_i v_j$ for all $i \leq j$), but not in general.

Nevertheless, whether or not (1.10) necessarily holds, (1.12), (1.6), (1.5), and (1.3) define the transition function of a homogeneous Markov chain, in that the distribution of (p'_1, \dots, p'_{r-1}) is completely specified in terms of (p_1, \dots, p_{r-1}) . (Note that $p_r = 1 - \sum_{i=1}^{r-1} p_i$.) This chain, which may or may not be related to the previously described chain by (1.1), is known as the Wright-Fisher model. Although its underlying biological assumptions are somewhat obscure, the Wright-Fisher model is probably the best-known discrete stochastic model in population genetics. Nevertheless, because of the complicated nature of its transition function, it is impractical to study this Markov chain directly. Instead, it is typically approximated by a diffusion process. Before indicating in the next section the usefulness of such an approach, we formulate the diffusion approximation precisely.

Put $\mathbb{Z}_+ = \{0, 1, \dots\}$ and

$$(1.13) \quad K_N = \left\{ (2N)^{-1} \alpha : \alpha \in (\mathbb{Z}_+)^{r-1}, \sum_{i=1}^{r-1} \alpha_i \leq 2N \right\}.$$

Given constants $\mu_{ij} \geq 0$ (with $\mu_{ii} = 0$) and σ_{ij} ($= \sigma_{ji}$) real for $i, j = 1, \dots, r$, let $\{Z^N(k), k = 0, 1, \dots\}$ be a homogeneous Markov chain in K_N whose transition function, starting at $(p_1, \dots, p_{r-1}) \in K_N$, is specified by (1.12), (1.6), (1.5), (1.3), and

$$(1.14) \quad u_{ij} = [(2N)^{-1} \mu_{ij}] \wedge r^{-1}, \quad w_{ij} = [1 + (2N)^{-1} \sigma_{ij}] \vee \frac{1}{2},$$

and let T_N be the associated transition operator on $C(K_N)$, that is,

$$(1.15) \quad T_N f(p_1, \dots, p_{r-1}) = E[f(p'_1, \dots, p'_{r-1})].$$

Let

$$(1.16) \quad K = \left\{ p = (p_1, \dots, p_{r-1}) \in [0, 1]^{r-1} : \sum_{i=1}^{r-1} p_i \leq 1 \right\},$$

and form the differential operator

$$(1.17) \quad G = \frac{1}{2} \sum_{i,j=1}^{r-1} a_{ij}(p) \frac{\partial^2}{\partial p_i \partial p_j} + \sum_{i=1}^{r-1} b_i(p) \frac{\partial}{\partial p_i},$$

where

$$(1.18) \quad a_{ij}(p) = p_i(\delta_{ij} - p_j)$$

and

$$(1.19) \quad b_i(p) = - \sum_{j=1}^r \mu_{ij} p_i + \sum_{j=1}^r \mu_{ji} p_j + p_i \left(\sum_{j=1}^r \sigma_{ij} p_j - \sum_{k,l=1}^r \sigma_{kl} p_k p_l \right).$$

Let $\{T(t)\}$ be the Feller semigroup on $C(K)$ generated by the closure of $A \equiv \{(f, Gf) : f \in C^2(K)\}$ (see Theorem 2.8 of Chapter 8), and let X be a diffusion process in K with generator A (i.e., a Markov process with sample paths in $C_K[0, \infty)$ corresponding to $\{T(t)\}$).

Finally, let X^N be the process with sample paths in $D_K[0, \infty)$ defined by

$$(1.20) \quad X^N(t) = Z^N([2Nt]).$$

1.1 Theorem. Under the above conditions,

$$(1.21) \quad \lim_{N \rightarrow \infty} \sup_{0 \leq t \leq t_0} \sup_{p \in K_N} |T_N^{(2Nt)} f(p) - T(t)f(p)| = 0$$

for every $f \in C(K)$ and $t_0 \geq 0$. Consequently, if $X^N(0) \Rightarrow X(0)$ in K , then $X^N \Rightarrow X$ in $D_K[0, \infty)$.

Proof. To prove (1.21), it suffices by Theorem 6.5 of Chapter 1 to show that

$$(1.22) \quad \lim_{N \rightarrow \infty} \sup_{p \in K_N} |2N(T_N - I)f(p) - Gf(p)| = 0$$

for all $f \in C^2(K)$. By direct calculation,

$$(1.23) \quad 2NE[p'_i - p_i] = b_i(p) + O(N^{-1}),$$

$$(1.24) \quad 2N \operatorname{cov}(p'_i, p'_j) = a_{ij}(p) + O(N^{-1}),$$

$$(1.25) \quad 2NE[(p'_i - p_i)^4] = O(N^{-1}),$$

and hence

$$(1.26) \quad 2NE[(p'_i - p_i)(p'_j - p_j)] = a_{ij}(p) + O(N^{-1}),$$

$$(1.27) \quad 2NP\{|p'_i - p_i| > \varepsilon\} = O(N^{-1}),$$

as $N \rightarrow \infty$, uniformly in $p \in K_N$, for $i, j = 1, \dots, r - 1$ and $\varepsilon > 0$. We leave to the reader the proof that (1.23), (1.26), and (1.27) imply (1.22) (Problem 1).

The second assertion of the theorem is a consequence of (1.21) and Corollary 8.9 of Chapter 4. \square

2. APPLICATIONS OF THE DIFFUSION APPROXIMATION

In this section we describe three applications of Theorem 1.1. We obtain diffusion approximations of stationary distributions, mean absorption times, and absorption probabilities of the one-locus, two-allele Wright–Fisher model. Moreover, we justify these approximations mathematically by proving appropriate limit theorems.

Let $\{Z^N(k), k = 0, 1, \dots\}$ be a homogeneous Markov chain in

$$(2.1) \quad K_N = \left\{ \frac{i}{2N} : i = 0, 1, \dots, 2N \right\}$$

whose transition function, starting at $p \in K_N$, is specified by (1.12), (1.6), (1.5), (1.3), and (1.14) in the special case $r = 2$. Concerning the parameters $\mu_{12}, \mu_{21}, \sigma_{11}, \sigma_{12}$, and σ_{22} in (1.14), we assume that $\sigma_{12} = 0$ and relabel the remaining parameters as μ_1, μ_2, σ_1 , and σ_2 to reduce the number of subscripts. (Since all viabilities can be multiplied by a constant without affecting (1.3), it involves no real loss of generality to take $w_{12} = 1$, i.e., $\sigma_{12} = 0$.) Then Z^N satisfies

$$(2.2) \quad P\left\{ Z^N(k+1) = \frac{j}{2N} \mid Z^N(k) = p \right\} = \binom{2N}{j} (p^{**})^j (1 - p^{**})^{2N-j},$$

where

$$(2.3) \quad p^{**} = (1 - u_1)p^* + u_2(1 - p^*),$$

$$(2.4) \quad p^* = \frac{w_1 p^2 + p(1-p)}{w_1 p^2 + 2p(1-p) + w_2(1-p)^2},$$

and

$$(2.5) \quad u_i = [(2N)^{-1}\mu_i] \wedge \frac{1}{2}, \quad w_i = [1 + (2N)^{-1}\sigma_i] \vee \frac{1}{2}, \quad i = 1, 2.$$

Recalling the other notation that is needed, T_N is the transition operator on $C(K_N)$ defined by (1.15),

$$(2.6) \quad K = [0, 1],$$

and X^N is the process with sample paths in $D_K[0, \infty)$ defined by (1.20). Finally, $\{T(t)\}$ is the strongly continuous semigroup on $C(K)$ generated by the closure of $A \equiv \{(f, Gf) : f \in C^2(K)\}$, where

$$(2.7) \quad G = \frac{1}{2}a(p) \frac{\partial^2}{\partial p^2} + b(p) \frac{\partial}{\partial p},$$

$$(2.8) \quad a(p) = p(1-p),$$

and

$$(2.9) \quad b(p) = -\mu_1 p + \mu_2(1-p) + p(1-p)[\sigma_1 p - \sigma_2(1-p)],$$

and X is a diffusion process in K with generator A . Clearly, the conclusions of Theorem 1.1 are valid in this special case.

As a first application, we consider the problem of approximating stationary distributions. Note that, if $\mu_1, \mu_2 > 0$, then $0 < p^{**} < 1$ for all $p \in K_N$, so Z^N is an irreducible, finite Markov chain. Hence it has a unique stationary distribution $v_N \in \mathcal{P}(K_N)$. (Of course, we may also regard v_N as an element of $\mathcal{P}(K)$.) Because v_N cannot be effectively evaluated, we approximate it by the stationary distribution of X .

2.1 Lemma Let $\mu_1, \mu_2 > 0$. Then X has one and only one stationary distribution $v \in \mathcal{P}(K)$. Moreover, v is absolutely continuous with respect to Lebesgue measure on K , and its density h_0 is the unique $C^2(0, 1)$ solution of the equation

$$(2.10) \quad \frac{1}{2}(ah_0)'' - (bh_0)' = 0$$

with $\int_0^1 h_0(p) dp = 1$. Consequently, there is a constant $\beta > 0$ such that

$$(2.11) \quad h_0(p) = \beta p^{2\mu_2-1} (1-p)^{2\mu_1-1} \exp \{ \sigma_1 p^2 + \sigma_2 (1-p)^2 \}$$

for $0 < p < 1$.

Proof. We first prove existence. Define h_0 by (2.11), where β is such that $\int_0^1 h_0(p) dp = 1$, and define $v \in \mathcal{P}(K)$ by $v(dp) = h_0(p) dp$. Since $(ah_0)(0+) = (ah_0)(1-) = 0$ and (2.10) holds, integration by parts yields

$$(2.12) \quad \int_K Gf dv = 0, \quad f \in C^2(K).$$

It follows that $\int_K \bar{A}f dv = 0$ for all $f \in \mathcal{D}(\bar{A})$, and hence

$$(2.13) \quad \int_K T(t)f dv = \int_K f dv, \quad f \in C(K), t \geq 0.$$

Thus, v is a stationary distribution for X . (See Chapter 4, Section 9.)

Turning to uniqueness, let $v \in \mathcal{P}(K)$ be a stationary distribution for X , and define

$$(2.14) \quad c(p) = \int_{[0, p]} b(q)v(dq).$$

Since (2.13) holds, so does (2.12). In particular, $c(1) = 0$ (take $f(p) \equiv p$), so

$$\begin{aligned} (2.15) \quad 0 &= \int_K \frac{1}{2}af'' dv - \int_K b(q) \left[\int_q^1 f''(p) dp \right] v(dp) \\ &= \int_K \frac{1}{2}af'' dv - \int_0^1 f''(p) \left[\int_{[0, p]} b(q)v(dq) \right] dp \\ &= \int_K f''(p) [\frac{1}{2}a(p)v(dp) - c(p) dp] \end{aligned}$$

for every $f \in C^2(K)$. Therefore,

$$(2.16) \quad \frac{1}{2}a(p)v(dp) = c(p) dp$$

as Borel measures on K . Since $a > 0$ on $(0, 1)$, we have $v(dp) \ll dp$ on $(0, 1)$, so by (2.14), c is continuous on $[0, 1]$. By (2.16), $v(dp)/dp$ has a continuous version h_0 on $(0, 1)$, and $\frac{1}{2}ah_0 = c$ there. Thus, by (2.14),

$$(2.17) \quad \frac{1}{2}a(p)h_0(p) = c(0) + \int_0^p b(q)h_0(q) dq$$

for $0 < p < 1$. It follows that $h_0 \in C^2(0, 1)$ and (2.10) holds, so since h_0 is Lebesgue integrable, it is easily verified that h_0 has the form (2.11) for some constant β . To verify that β is such that $\int_0^1 h_0(p) dp = 1$, and to complete the proof that v is uniquely determined, it suffices to show that $v(\{0\}) = v(\{1\}) = 0$. By (2.11), $(ah_0)(0+) = 0$, so by (2.17), $c(0) = 0$. Since $b(0) = \mu_2 > 0$, we have $v(\{0\}) = 0$ by (2.14). Similarly, $c(1-) = 0$, so $v(\{1\}) = 0$, completing the proof. \square

We now show that v_N , the stationary distribution of Z^N , can be approximated by v , the stationary distribution of X .

2.2 Theorem. Let $\mu_1, \mu_2 > 0$. Then $v_N \Rightarrow v$ on K .

Proof We could essentially quote Theorem 9.10 of Chapter 4, but instead we give a self-contained proof. By Prohorov's theorem, $\{v_N\}$ is relatively compact in $\mathcal{P}(K)$, so every subsequence of $\{v_N\}$ has a further subsequence $\{v_{N'}\}$ that converges weakly to some limit $\tilde{v} \in \mathcal{P}(K)$. Consequently, for all $f \in C(K)$ and $t \geq 0$,

$$\begin{aligned} (2.18) \quad \int_K T(t)f d\tilde{v} &= \lim_{N' \rightarrow \infty} \int_{K_{N'}} T(t)f dv_{N'} \\ &= \lim_{N' \rightarrow \infty} \int_{K_{N'}} T_N^{(2N')t} f dv_{N'} \\ &= \lim_{N' \rightarrow \infty} \int_{K_{N'}} f dv_{N'} \\ &= \int_K f d\tilde{v}, \end{aligned}$$

so \tilde{v} is a stationary distribution for X . (This gives, incidentally, an alternative proof of the existence of a stationary distribution for X .) By Lemma 2.1, $\tilde{v} = v$. Hence the original sequence converges weakly to v . \square

2.3 Remark If $\sigma_1 = \sigma_2 = 0$, then the stationary distribution of Theorem 2.2 belongs to the beta family. In particular, its mean is $\mu_2/(\mu_1 + \mu_2)$, which also happens to be the stable equilibrium of the corresponding deterministic model $p = -\mu_1 p + \mu_2(1 - p)$. \square

If $\mu_1 = 0$ (respectively, if $\mu_2 = 0$), then 1 (respectively, 0) is an absorbing state for the Markov chain Z^N , so interest centers on the time until absorption and (if $\mu_1 = \mu_2 = 0$) the probability of ultimate absorption at 1. Of the three cases, $\mu_1 = \mu_2 = 0$, $\mu_1 = 0 < \mu_2$, and $\mu_1 > 0 = \mu_2$, it will suffice (by symmetry) to treat the first two. Let

$$(2.19) \quad \begin{aligned} F &= \{0, 1\} && \text{if } \mu_1 = \mu_2 = 0, \\ F &= \{1\} && \text{if } \mu_1 = 0, \mu_2 > 0. \end{aligned}$$

Then F is the set of absorbing states of Z^N (and hence X^N), and it is easily seen (by uniqueness, e.g.) that F is also the set of absorbing states of the diffusion X .

Define $\zeta: D_K[0, \infty) \rightarrow [0, \infty]$ by

$$(2.20) \quad \zeta(x) = \inf \{t \geq 0: x(t) \in F \text{ or } x(t-) \in F\}$$

where $\inf \emptyset = \infty$. Then ζ is Borel measurable, so we can define

$$(2.21) \quad \tau_N = \zeta(X^N), \quad \tau = \zeta(X).$$

In order to study the mean absorption time $E[\tau_N]$ and (if $\mu_1 = \mu_2 = 0$) the absorption probability $P\{X^N(\tau_N) = 1\}$, we regard $E[\tau]$ and $P\{X(\tau) = 1\}$ as approximations, the latter two quantities being quite easy to evaluate. The following theorem is used to provide a justification for these approximations.

2.4 Theorem Let $\mu_1 = 0, \mu_2 \geq 0$. If $X^N(0) \Rightarrow X(0)$ in K , then $(X^N, \tau_N) \Rightarrow (X, \tau)$ in $D_K[0, \infty) \times [0, \infty]$, and the sequence $\{\tau_N\}$ is uniformly integrable.

We postpone the proof to the end of this section.

2.5 Remark It follows from Theorem 2.4 that, for each $N \geq 1$, τ_N and τ have finite expectations, hence they are a.s. finite, and therefore $X^N(\tau_N)$ and $X(\tau)$ are defined a.s. and equal to 0 or 1 a.s. These facts are needed in the corollaries that follow. It is also worth noting that the first assertion of Theorem 2.4 is not a consequence of Corollary 1.9 of Chapter 3 because the function $x \mapsto (x, \zeta(x))$ on $D_K[0, \infty)$ is discontinuous at every $x \in D_K[0, \infty)$ for which $\zeta(x) < \infty$, hence discontinuous a.s. with respect to the distribution of X . \square

2.6 Corollary Let $\mu_1 = 0, \mu_2 \geq 0$. If $X^N(0) \Rightarrow X(0)$, then

$$(2.22) \quad \lim_{N \rightarrow \infty} E[\tau_N] = E[\tau].$$

Proof. By Theorem 2.4, $\tau_N \Rightarrow \tau$, so (2.22) follows from the uniform integrability of $\{\tau_N\}$. \square

2.7 Corollary Let $\mu_1 = \mu_2 = 0$. If $X^N(0) \Rightarrow X(0)$, then

$$(2.23) \quad \lim_{N \rightarrow \infty} P\{X^N(\tau_N) = 1\} = P\{X(\tau) = 1\}.$$

Proof. Define $\xi: D_K[0, \infty] \times [0, \infty] \rightarrow K$ by $\xi(x, t) = x(t)$ for $0 \leq t < \infty$ and $\xi(x, \infty) = \frac{1}{2}$, say. Then ξ is continuous at each point of $C_K[0, \infty] \times [0, \infty)$, hence continuous a.s. with respect to the distribution of (X, τ) . By Theorem 2.4 of this chapter and Corollary 1.9 of Chapter 3, $X^N(\tau_N) \Rightarrow X(\tau)$, so (2.23) follows from Remark 2.5. \square

To evaluate the right sides of equations (2.22) and (2.23), we introduce the notation $P_p\{\cdot\}$ and $E_p[\cdot]$, where the subscript p denotes the starting point of the process involved in the probability or expectation.

2.8 Proposition Suppose first that $\mu_1 = \mu_2 = 0$. Let f_0 be the unique $C^2(K)$ solution of the differential equation $Gf_0 = 0$ with boundary conditions $f_0(0) = 0, f_0(1) = 1$. Then $P_p\{X(\tau) = 1\} = f_0(p)$ for all $p \in K$. Consequently,

$$(2.24) \quad P_p\{X(\tau) = 1\} = \frac{\int_0^1 e^{-\lambda(q)} dq}{\int_0^1 e^{-\lambda(q)} dq},$$

where $\lambda(q) = \sigma_1 q^2 + \sigma_2(1 - q)^2$.

Now suppose that $\mu_1 = 0, \mu_2 \geq 0$. Let g_0 be the unique $C(K) \cap C^2(K - F)$ solution of the differential equation $Gg_0 = -1$ with boundary conditions

$$(2.25) \quad g_0(0) = g_0(1) = 0, \quad \text{if } \mu_2 = 0, \\ g'_0(0+) \text{ finite, } g_0(1) = 0, \quad \text{if } \mu_2 > 0.$$

Then $E_p[\tau] = g_0(p)$ for all $p \in K$. Consequently, if $\mu_2 = 0$, then

$$(2.26) \quad E_p[\tau] = \int_0^p e^{-\lambda(q)} \int_q^{1/2} \frac{2e^{\lambda(r)}}{r(1-r)} dr dq \\ - f_0(p) \int_0^1 e^{-\lambda(q)} \int_q^{1/2} \frac{2e^{\lambda(r)}}{r(1-r)} dr dq,$$

and, if $\mu_2 > 0$, then

$$(2.27) \quad E_p[\tau] = \int_p^1 q^{-2\mu_2} e^{-\lambda(q)} \int_0^q 2r^{2\mu_2-1} (1-r)^{-1} e^{\lambda(r)} dr dq.$$

Proof. For each $f \in C^2(K)$, $p \in K$, and $t \geq 0$, the optional sampling theorem implies that

$$(2.28) \quad E_p[f(X(\tau \wedge t))] = f(p) + E_p\left[\int_0^{\tau \wedge t} Gf(X(s)) ds\right].$$

Replacing f by f_0 and letting $t \rightarrow \infty$, we get $P_p\{X(\tau) = 1\} = f_0(p)$ for all $p \in K$. Here we are using Remark 2.5.

Given $h \in C(K)$, let g be the unique $C(K) \cap C^2(K - F)$ solution of the differential equation $Gg = -h$ with boundary conditions analogous to (2.25). Then $g = Bh$, where

$$(2.29) \quad \begin{aligned} Bh(p) &= \int_0^p e^{-\lambda(q)} \int_q^{1/2} \frac{2e^{\lambda(r)}}{r(1-r)} h(r) dr dq \\ &\quad - f_0(p) \int_0^1 e^{-\lambda(q)} \int_q^{1/2} \frac{2e^{\lambda(r)}}{r(1-r)} h(r) dr dq \end{aligned}$$

if $\mu_2 = 0$, and

$$(2.30) \quad Bh(p) = \int_p^1 q^{-2\mu_2} e^{-\lambda(q)} \int_0^q 2r^{2\mu_2-1}(1-r)^{-1} e^{\lambda(r)} h(r) dr dq$$

if $\mu_2 > 0$. Consequently, $Bh \in C^2(K)$ if $h \in C^1(K)$ and $h = 0$ on F . Thus, we choose $\{h_n\} \subset C^1(K)$ with $h_n \geq 0$ and $h_n \nearrow 1 - \chi_F$. Replacing f by $g_n = Bh_n$ in (2.28), and noting that $\text{bp-lim } g_n = g_0$, we obtain

$$(2.31) \quad E_p[g_0(X(\tau \wedge t))] = g_0(p) - E_p[\tau \wedge t]$$

for all $p \in K$ and $t \geq 0$. Letting $t \rightarrow \infty$ gives $E_x[\tau] = g_0(p)$ by (2.25).

We leave it to the reader to verify (2.24), (2.26), and (2.27). \square

2.9 Remark As simple special cases of Proposition 2.8, one can check that

$$(2.32) \quad P_p\{X(\tau) = 1\} = \begin{cases} p, & \sigma = 0, \\ \frac{1 - e^{-2\sigma p}}{1 - e^{-2\sigma}}, & \sigma \neq 0, \end{cases}$$

if $\mu_1 = \mu_2 = 0$ and $\sigma_1 = -\sigma_2 = \sigma$, and

$$(2.33) \quad E_p[\tau] = -2[p \log p + (1-p) \log(1-p)]$$

if $\mu_1 = \mu_2 = \sigma_1 = \sigma_2 = 0$. To get some idea of the effect that selection can have, observe that, when $p = \frac{1}{2}$, the right side of (2.32) becomes $1/(1 + e^{-\sigma})$, and in view of (2.5), $|\sigma|$ may differ significantly from zero. When interpreting (2.33) (or, more generally, (2.26) or (2.27)), one must keep in mind that, because of (1.20), time is measured in units of $2N$ generations. \square

In order to prove Theorem 2.4, we need the following lemma.

2.10 Lemma Let $\mu_1 = 0$, $\mu_2 \geq 0$, and define the function g_0 as in Proposition 2.8. Then there exist positive integers, κ and N_0 , depending only on μ_2 , σ_1 , and σ_2 , such that

$$(2.34) \quad E_p[\tau_N] \leq \kappa g_0(p), \quad p \in K_N, N \geq N_0.$$

Proof. Define the operator G_N on $C(K_N)$ by

$$(2.35) \quad G_N f(p) = 2N \{E_p[f(Z^N(1))] - f(p)\}.$$

For $0 \leq \varepsilon < \frac{1}{2}$, let

$$(2.36) \quad V(\varepsilon) = \begin{cases} (\varepsilon, 1 - \varepsilon) & \text{if } \mu_2 = 0 \\ [0, 1 - \varepsilon) & \text{if } \mu_2 > 0, \end{cases}$$

and put $V_N(\varepsilon) = K_N \cap V(\varepsilon)$. (Note that $V(0) = K - F$.) The first step in the proof is to show that

$$(2.37) \quad \lim_{m \rightarrow \infty} \overline{\lim}_{N \rightarrow \infty} \sup_{p \in V_N(m/2N)} G_N g_0(p) \leq -1.$$

A fourth-order Taylor expansion yields

$$(2.38) \quad G_N g_0(p) = 2N \left\{ \sum_{k=1}^3 \frac{1}{k!} E_p[(Z^N(1) - p)^k] g_0^{(k)}(p) + \frac{1}{3!} E_p \left[(Z^N(1) - p)^4 \int_0^1 (1-t)^3 g_0^{(4)}(p + t(Z^N(1) - p)) dt \right] \right\}$$

for all $p \in V_N(0)$ and $N \geq 1$. (We note that the integral under the fourth expectation exists, as does the expectation itself.) Expanding each of the moments about p^{**} , which we temporarily denote by γ , we obtain

$$(2.39) \quad \begin{aligned} G_N g_0(p) &= 2N(\gamma - p)g'_0(p) + \frac{2N}{2} \left[\frac{\gamma(1-\gamma)}{2N} + (\gamma - p)^2 \right] g''_0(p) \\ &\quad + \frac{2N}{6} \left[\frac{\gamma(1-\gamma)(1-2\gamma)}{(2N)^2} + \frac{3\gamma(1-\gamma)(\gamma-p)}{2N} + (\gamma-p)^2 \right] g'''_0(p) \\ &\quad + \theta_{N,p} \frac{2N}{6} \left[\frac{3\gamma^2(1-\gamma)^2}{(2N)^2} + \frac{\gamma(1-\gamma)(1-6\gamma+6\gamma^2)}{(2N)^3} \right. \\ &\quad \left. + \frac{4\gamma(1-\gamma)(1-2\gamma)(\gamma-p)}{(2N)^2} + \frac{6\gamma(1-\gamma)(\gamma-p)^2}{2N} + (\gamma-p)^2 \right] \\ &\quad \times \int_0^1 (1-t)^3 \sup_{0 \leq \varepsilon \leq 1} |g_0^{(4)}(p + t(\gamma - p))| dt \end{aligned}$$

for all $p \in V_N(0)$ and $N \geq 1$, where $|\theta_{N,p}| \leq 1$. Now one can easily check that

$$(2.40) \quad 2N(p^{**} - p) = b(p)(1 + O(N^{-1})) + O(\mu_2(1-p)^2 N^{-1})$$

and

$$(2.41) \quad p^{**}(1 - p^{**}) = a(p)(1 + O(N^{-1})) + O(\mu_2(1-p)^2 N^{-1})$$

as $N \rightarrow \infty$, uniformly in $p \in K_N$. Also, by direct calculation, there exist constants M_1, \dots, M_4 , depending only on μ_2, σ_1 , and σ_2 , such that

$$(2.42) \quad \begin{aligned} |g'_0(p)| &\leq M_1 \log \left[\frac{1 + \mu_2}{(p + \mu_2)(1 - p)} \right], \\ |g_0^{(k)}(p)| &\leq M_k \left[\frac{1 + \mu_2}{(p + \mu_2)(1 - p)} \right]^{k-1}, \end{aligned}$$

for all $p \in V(0)$ and $k = 2, 3, 4$. Finally, we note that

$$(2.43) \quad \min_{0 \leq t \leq 1} \frac{[p + t(q - p) + \mu_2][1 - p - t(q - p)]}{1 + \mu_2} \geq (1 - t) \frac{(p + \mu_2)(1 - p)}{1 + \mu_2}$$

since the minimum occurs at $q = 0$ or $q = 1$, and therefore

$$(2.44) \quad \int_0^1 (1 - t)^3 \sup_{0 \leq t \leq 1} |g_0^{(4)}(p + t(q - p))| dt \leq M_4 \left[\frac{1 + \mu_2}{(p + \mu_2)(1 - p)} \right]^3$$

for all $p \in V(0)$. By (2.39)–(2.42) and (2.44), we have

$$(2.45) \quad \begin{aligned} G_N g_0(p) &\leq \frac{1}{2} a(p) g_0''(p) + b(p) g_0'(p) \\ &\quad + (\frac{1}{12} M_3 + \frac{1}{4} M_4) \left[\frac{N(p + \mu_2)(1 - p)}{1 + \mu_2} \right]^{-1} \\ &\quad + \frac{1}{24} M_4 \left[\frac{N(p + \mu_2)(1 - p)}{1 + \mu_2} \right]^{-2} + O(N^{-1}) \end{aligned}$$

as $N \rightarrow \infty$, uniformly in $p \in V_N(0)$, which implies (2.37) since $Gg_0 = -1$.

Next, we show that

$$(2.46) \quad \overline{\lim}_{N \rightarrow \infty} G_N g_0 \left(1 - \frac{m}{2N} \right) < 0, \quad m = 1, 2, \dots$$

Fix $m \geq 1$, and let $p_N = (1 - m/2N) \vee 0$. Since g_0'' is bounded on $(0, \frac{1}{2})$ if $\mu_2 > 0$, there exists a constant M_0 , depending only on μ_2, σ_1 , and σ_2 , such that

$$(2.47) \quad \begin{aligned} g_0''(p) &= \frac{-2}{p(1-p)} - \left[\frac{2\mu_2}{p} + 2(\sigma_1 p - \sigma_2(1-p)) \right] g_0'(p) \\ &\leq \frac{-2}{1-p} + M_0 \log \left[\frac{1 + \mu_2}{(p + \mu_2)(1 - p)} \right] \end{aligned}$$

for all $p \in V(0)$. Consequently, a second-order Taylor expansion yields

$$\begin{aligned}
 (2.48) \quad G_N g_0(p_N) &= 2NE_{p_N}[Z^N(1) - p_N]g'_0(p_N) \\
 &\quad + 2NE_{p_N}\left[(Z^N(1) - p_N)^2 \int_0^1 (1-t)g''_0(Y) dt\right] \\
 &\leq 2NE_{p_N}[Z^N(1) - p_N]g'_0(p_N) \\
 &\quad + 2NM_0 E_{p_N}\left[(Z^N(1) - p_N)^2 \int_0^1 (1-t)\right. \\
 &\quad \times \log\left(\frac{1+\mu_2}{(Y+\mu_2)(1-Y)}\right) dt\left.\right] \\
 &\quad - 4NE_{p_N}\left[(Z^N(1) - p_N)^2 \int_0^1 (1-t)(1-Y)^{-1} dt\right]
 \end{aligned}$$

for each $N \geq 1$, where $Y = p_N + t(Z^N(1) - p_N)$. Using (2.42) and (2.43), the first two terms on the right side of (2.48) are $O(N^{-1} \log N)$ as $N \rightarrow \infty$, so (2.46) is equivalent to

$$(2.49) \quad \lim_{N \rightarrow \infty} NE_{p_N}\left[(Z^N(1) - p_N)^2 \int_0^1 (1-t)(1-Y)^{-1} dt\right] > 0.$$

Denoting p^{**} by p_N^{**} when $p = p_N$, the expectation in (2.49) can be expressed as

$$\begin{aligned}
 (2.50) \quad &\sum_{l=0}^{2N} \left(\frac{l}{2N} - \frac{m}{2N}\right)^2 \int_0^1 (1-t)\left(\frac{m}{2N} + t\left(\frac{l}{2N} - \frac{m}{2N}\right)\right)^{-1} dt \\
 &\times \binom{2N}{l} (1-p_N^{**})^l (p_N^{**})^{2N-l},
 \end{aligned}$$

and since $1 - p_N^{**} = m/2N + O(N^{-2})$, an application of Fatou's lemma shows that the left side of (2.49) is at least as large as

$$(2.51) \quad \sum_{l=0}^{\infty} (l-m)^2 \int_0^1 (1-t)[m + t(l-m)]^{-1} dt \frac{m^l e^{-m}}{l!},$$

which of course is positive; here we have used the familiar Poisson approximation of the binomial distribution. This proves (2.46), and, by symmetry,

$$(2.52) \quad \overline{\lim}_{N \rightarrow \infty} G_N g_0\left(\frac{m}{2N}\right) < 0, \quad m = 1, 2, \dots,$$

if $\mu_2 = 0$. Combining (2.37), (2.46), and (if $\mu_2 = 0$) (2.52), we conclude that there exist κ and N_0 such that

$$(2.53) \quad G_N g_0(p) \leq -\frac{1}{\kappa}, \quad p \in V(0), N \geq N_0.$$

Finally, to complete the proof of the lemma, we note that

$$(2.54) \quad \left\{ g_0(Z^N(n)) - \frac{1}{2N} \sum_{m=0}^{n-1} G_N g_0(Z^N(m)), \quad n = 0, 1, \dots \right\}$$

is a martingale, so by the optional sampling theorem and (2.53),

$$(2.55) \quad \begin{aligned} E_p[g_0(X^N(\tau_N \wedge t))] &= g_0(p) + E_p\left[\frac{1}{2N} \sum_{m=0}^{2N(\tau_N \wedge t)-1} G_N g_0\left(X^N\left(\frac{m}{2N}\right)\right)\right] \\ &\leq g_0(p) - \frac{1}{\kappa} E_p[\tau_N \wedge t] \end{aligned}$$

for all $p \in V_N(0)$, $t = 0, 1/2N, 2/2N, \dots$, and $N \geq N_0$, and this implies (2.34). \square

Proof of Theorem 2.4. For $0 \leq \varepsilon < \frac{1}{2}$, define $\zeta^\varepsilon: D_K[0, \infty) \rightarrow [0, \infty]$ by

$$(2.56) \quad \zeta^\varepsilon(x) = \inf \{t \geq 0: x(t) \notin V(\varepsilon) \text{ or } x(t-) \notin V(\varepsilon)\}$$

where $V(\varepsilon)$ is given by (2.36). (Note that $\zeta^0 = \zeta$; see (2.20).) Then $\zeta^\varepsilon(x) \rightarrow \zeta(x)$ and $\varepsilon \rightarrow 0$ for every $x \in C_K[0, \infty)$, hence a.s. with respect to the distribution of X . In addition, we leave it to the reader to show that ζ^ε is continuous a.s. with respect to the distribution of X for $0 < \varepsilon < \frac{1}{2}$ (Problem 3).

We apply the result of Problem 5 of Chapter 3 with $S = D_K[0, \infty)$, $S' = D_K[0, \infty) \times [0, \infty]$, $h(x) = (x, \zeta(x))$, $h_k(x) = (x, \zeta^{\varepsilon_k}(x))$, where $0 < \varepsilon_k < \frac{1}{2}$ and $\varepsilon_k \rightarrow 0$ as $k \rightarrow \infty$. To conclude that $h(X^N) \Rightarrow h(X)$, that is, $(X^N, \tau_N) \Rightarrow (X, \tau)$, we need only show that

$$(2.57) \quad \lim_{\varepsilon \rightarrow 0} \overline{\lim}_{N \rightarrow \infty} P\{\rho(\zeta^\varepsilon(X^N), \zeta(X^N)) > \delta\} = 0$$

for every $\delta > 0$, where $\rho(t, t') = |\tan^{-1} t - \tan^{-1} t'|$. By the strong Markov property, the inequality $|\tan^{-1} t - \tan^{-1} (t+s)| \leq s$ for $s, t \geq 0$, and Lemma 2.10, we have

$$\begin{aligned} (2.58) \quad P\{\rho(\zeta^\varepsilon(X^N), \zeta(X^N)) &> \delta\} \\ &\leq E[\chi_{\{\tau_N < \infty\}} P_{X^N(\tau_N)} \{\tau_N > \delta\}] \\ &\leq \delta^{-1} \sup_{p \in K_N \cap V(\varepsilon K)} E_p[\tau_N] \\ &\leq \delta^{-1} \kappa \sup_{p \in K \cap V(\varepsilon K)} g_0(p) \end{aligned}$$

for all $\delta > 0$, $N \geq N_1$, and $0 < \varepsilon < \frac{1}{2}$, where $\tau_N^\varepsilon = \zeta^\varepsilon(X^N)$. Since $g_0 = 0$ on F , (2.57) follows from (2.58).

Finally, we claim that the uniform integrability of $\{\tau_N\}$ is also a consequence of Lemma 2.10. Let g_0 , κ , and N_0 be as in that lemma. Then

$$(2.59) \quad \sup_{p \in K_N} P_p\{\tau_N > t\} \leq t^{-1} \sup_{p \in K_N} E_p[\tau_N] \leq t^{-1} \kappa \sup_{p \in K} g_0(p)$$

for all $N \geq N_0$ and $t > 0$, so there exist $t_0 > 0$ and $\eta < 1$ such that

$$(2.60) \quad \sup_{N \geq 1} \sup_{p \in K_N} P_p\{\tau_N > t_0\} < \eta.$$

Letting $E_m^N = \{\tau_N > m/2N\}$, we conclude from the strong Markov property that, if $n \geq [2Nt_0]$, then

$$(2.61) \quad P_p(E_{m+n}^N) = E_p[\chi_{E_m^N} P_{X^{N(m/2N)}}(E_n^N)] \leq \eta P_p(E_m^N)$$

for each $m \geq 0$, $p \in K_N$, and $N \geq 1$. Consequently, for arbitrary $l > 1$,

$$\begin{aligned} (2.62) \quad E_p[(\tau_N)^l] &= \sum_{k=0}^{\infty} \sum_{j=kN+1}^{(k+1)N} \left(\frac{j}{2N}\right)^l P_p\left\{\tau_N = \frac{j}{2N}\right\} \\ &\leq \sum_{k=0}^{\infty} (k+1)^l t_0^l P_p(E_{kn}^N) \\ &\leq t_0^l \sum_{k=0}^{\infty} (k+1)^l \eta^k < \infty \end{aligned}$$

for all $p \in K_N$ and $N \geq 1$, where $n = [2Nt_0]$. Since the bound in (2.62) is uniform in p and N , the uniform integrability of $\{\tau_N\}$ follows, and the proof is complete. \square

3. GENOTYPIC-FREQUENCY MODELS

There are several one-locus genetic models in which the successive values (from generation to generation, typically) of the vector $(P_{ij})_{i \leq j}$ of genotypic frequencies form a Markov chain, but the successive values of the vector (p_1, \dots, p_{r-1}) of allelic frequencies do not; nevertheless, the genotypic frequencies rapidly converge to Hardy-Weinberg proportions, while, at a slower rate, the allelic frequencies converge to a diffusion process. Thus, in this section, we formulate a limit theorem for diffusion approximations of Markov chains with two "time scales," and we apply it to two models. Further applications are mentioned in the problems.

Let K and H be compact, convex subsets of \mathbb{R}^m and \mathbb{R}^n , respectively, having nonempty interiors, and assume that $0 \in H$. We begin with two lemmas involving first-order differential and difference equations, in which the zero solution is globally asymptotically stable.

3.1 Lemma Let $c: K \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ be of class C^2 and such that the solution $Y(t, x, y)$ of the differential equation

$$(3.1) \quad \frac{d}{dt} Y(t, x, y) = c(x, Y(t, x, y)), \quad Y(0, x, y) = y,$$

exists for all $(t, x, y) \in [0, \infty) \times K \times H$ and satisfies

$$(3.2) \quad \lim_{t \rightarrow \infty} \sup_{(x, y) \in K \times H} |Y(t, x, y)| = 0.$$

Then there exists a compact set E , with $K \times H \subset E \subset K \times \mathbb{R}^n$, such that $(x, y) \in E$ implies $(x, Y(t, x, y)) \in E$ for all $t \geq 0$, and the formula

$$(3.3) \quad S(t)h(x, y) = h(x, Y(t, x, y))$$

defines a strongly continuous semigroup $\{S(t)\}$ on $C(E)$ (with sup norm). The generator B of $\{S(t)\}$ has $C^2(E) \equiv \{f|_E : f \in C^2(\mathbb{R}^m \times \mathbb{R}^n)\}$ as a core, and

$$(3.4) \quad Bh(x, y) = \sum_{i=1}^n c_i(x, y) \frac{\partial}{\partial y_i} h(x, y) \text{ on } K \times H, \quad h \in C^2(E).$$

Finally,

$$(3.5) \quad \lim_{t \rightarrow \infty} \sup_{(x, y) \in E} |S(t)h(x, y) - h(x, 0)| = 0, \quad h \in C(E).$$

Proof. Let $E = \{(x, Y(t, x, y)) : (t, x, y) \in [0, \infty) \times K \times H\}$. By (3.2), E is bounded, and E is easily seen to be closed. If $(x, y) \in E$, then $y = Y(s, x, y_0)$ for some $s \geq 0$ and $y_0 \in H$. Hence $(x, Y(t, x, y)) = (x, Y(t + s, x, y_0)) \in E$ for all $t \geq 0$, and

$$(3.6) \quad \lim_{t \rightarrow \infty} \sup_{(x, y) \in E} |Y(t, x, y)| \leq \lim_{t \rightarrow \infty} \sup_{s \geq 0} \sup_{(x, y_0) \in K \times H} |Y(t + s, x, y_0)| = 0$$

by (3.2). It is straightforward to check that $\{S(t)\}$ is a strongly continuous semigroup on $C(E)$. By the mean value theorem, $C^2(E) \subset \mathcal{D}(B)$ and (3.4) holds. Since $S(t): C^2(E) \rightarrow C^2(E)$ for all $t \geq 0$, $C^2(E)$ is a core for B . Finally, (3.5) is a consequence of (3.6). \square

3.2 Remark If $c(x, y) = \varphi(x)y$ for all $(x, y) \in K \times \mathbb{R}^n$, where $\varphi: K \rightarrow \mathbb{R}^n \otimes \mathbb{R}^n$ is of class C^2 , and if for each $x \in K$ all eigenvalues of $\varphi(x)$ have negative real parts, then c satisfies the hypotheses of Lemma 3.1. In this case, $Y(t, x, y) \equiv e^{t\varphi(x)}y$. \square

3.3 Lemma Given $\delta_\infty > 0$, let $c: K \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ be continuous, such that the solution $Y(k, x, y)$ of the difference equation

$$(3.7) \quad \delta_\infty^{-1}\{Y(k+1, x, y) - Y(k, x, y)\} = c(x, Y(k, x, y)), \quad Y(0, x, y) = y,$$

which exists for all $(k, x, y) \in \mathbb{Z}_+ \times K \times H$, satisfies

$$(3.8) \quad \lim_{k \rightarrow \infty} \sup_{(x, y) \in K \times H} |Y(k, x, y)| = 0.$$

Then there exists a compact set E , with $K \times H \subset E \subset K \times \mathbb{R}^n$, such that $(x, y) \in E$ implies $(x, Y(k, x, y)) \in E$ for $k = 0, 1, \dots$, and the formula

$$(3.9) \quad S(t)h(x, y) = E[h(x, Y(V(t), x, y))],$$

where V is a Poisson process with parameter δ_∞^{-1} , defines a strongly continuous semigroup $\{S(t)\}$ on $C(E)$. The generator B of $\{S(t)\}$ is the bounded linear operator

$$(3.10) \quad B = \delta_\infty^{-1}(Q - I),$$

where Q is defined on $C(E)$ by $Qh(x, y) = h(x, y + \delta_\infty c(x, y))$. Finally,

$$(3.11) \quad \lim_{k \rightarrow \infty} \sup_{(x, y) \in E} |Q^k h(x, y) - h(x, 0)| = 0, \quad h \in C(E).$$

Proof. Let $E = \{(x, Y(k, x, y)): (k, x, y) \in \mathbb{Z}_+ \times K \times H\}$. The details of the proof are left to the reader. \square

3.4 Remark If $c(x, y) = \varphi(x)y$ for all $(x, y) \in K \times \mathbb{R}^n$, where $\varphi: K \rightarrow \mathbb{R}^n \otimes \mathbb{R}^n$ is continuous, and if for each $x \in K$ all eigenvalues of $\varphi(x)$ belong to $\{\zeta \in \mathbb{C}: |\zeta + \delta_\infty^{-1}| < \delta_\infty^{-1}\}$, then c satisfies the hypotheses of Lemma 3.3. In this case, $Y(k, x, y) \equiv (I + \delta_\infty \varphi(x))^k y$. \square

The preceding lemmas allow us to state the following theorem. Recall the assumptions on K and H in the second paragraph of this section.

3.5 Theorem For $N = 1, 2, \dots$, let $\{Z^N(k), k = 0, 1, \dots\}$ be a Markov chain in a metric space E_N with a transition function $\mu_N(z, \Gamma)$, and denote $\int f(z')\mu_N(z, dz')$ by $E_z[f(Z^N(1))]$. Suppose both $\Phi_N: E_N \rightarrow K$ and $\Psi_N: E_N \rightarrow H$ are Borel measurable, define $X^N(k) = \Phi_N(Z^N(k))$ and $Y^N(k) = \Psi_N(Z^N(k))$ for each $k \geq 0$, and let $\varepsilon_N > 0$ and $\delta_N > 0$. Assume that $\lim_{N \rightarrow \infty} \delta_N = \delta_\infty \in [0, \infty)$ and $\lim_{N \rightarrow \infty} \varepsilon_N / \delta_N = 0$. Let each of the functions $a: K \times \mathbb{R}^n \rightarrow \mathbb{R}^m \otimes \mathbb{R}^n$, $b: K \times \mathbb{R}^n \rightarrow \mathbb{R}^m$, and $c: K \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ be continuous, and suppose that, for $i, j = 1, \dots, m$ and $l = 1, \dots, n$,

$$(3.12) \quad \varepsilon_N^{-1} E_z[X_i^N(1) - x_i] = b_i(x, y) + o(1),$$

$$(3.13) \quad \varepsilon_N^{-1} E_z[(X_i^N(1) - x_i)(X_j^N(1) - x_j)] = a_{ij}(x, y) + o(1),$$

$$(3.14) \quad \varepsilon_N^{-1} E_z[(X_i^N(1) - x_i)^4] = o(1),$$

$$(3.15) \quad \delta_N^{-1} E_z[Y_i^N(1) - y_i] = c_i(x, y) + o(1),$$

$$(3.16) \quad \delta_N^{-1} E_z[(Y_i^N(1) - E_z[Y_i^N(1)])^2] = o(1),$$

as $N \rightarrow \infty$, uniformly in $z \in E_N$, where $x = \Phi_N(z)$ and $y = \Psi_N(z)$. Let

$$(3.17) \quad G = \frac{1}{2} \sum_{i, j=1}^m a_{ij}(x, 0) \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^m b_i(x, 0) \frac{\partial}{\partial x_i},$$

and assume that the closure of $\{(f, Gf): f \in C^2(K)\}$ is single-valued and generates a Feller semigroup $\{U(t)\}$ on $C(K)$ corresponding to a diffusion process X in K . Suppose further that c satisfies the hypotheses of Lemma 3.1 if $\delta_\infty = 0$ and of Lemma 3.3 if $\delta_\infty > 0$. Then the following conclusions hold:

- (a) If $X^N(0) \Rightarrow X(0)$ in K , then $X^N([\cdot/\varepsilon_N]) \Rightarrow X$ in $D_K[0, \infty)$.
 (b) If $\{t_N\} \subset [0, \infty)$ satisfies $\lim_{N \rightarrow \infty} t_N = \infty$, then $Y^N([t_N/\delta_N]) \Rightarrow 0$ in H .

3.6 Remark (a) Observe that (3.12)–(3.14) are analogous to (1.23), (1.26), and (1.25), except that the right sides of (3.12) and (3.13) depend on y . But because of (3.15), (3.16), and the conditions on c , it is clear (at least intuitively) that conclusion (b) holds, and hence that $Y^N([t/\varepsilon_N]) \Rightarrow 0$ for each $t > 0$. Thus, in the “slow” time scale (i.e., t/ε_N), the Y^N process is “approximately” zero, and therefore the limiting generator has the form (3.17).

(b) We note that (3.14) implies

$$(3.18) \quad \varepsilon_N^{-1} P_z \{ |X_i^N(1) - x_i| > \gamma \} = o(1), \quad \gamma > 0,$$

for $i = 1, \dots, m$. (Here and below, we omit the phrase, “as $N \rightarrow \infty$, uniformly in $z \in E_N$, where $x = \Phi_N(z)$ and $y = \Psi_N(z)$.”) In fact the latter condition suffices in the proof. \square

Proof. Let E be as in Lemma 3.1 if $\delta_\infty = 0$ and as in Lemma 3.3 if $\delta_\infty > 0$, and apply Theorem 7.6(b) of Chapter 1 with $L_N = B(E_N)$ (with sup norm), $L = C(E)$, and $\pi_N: L \rightarrow L_N$ defined by $\pi_N f(z) = f(x, y)$, where $x = \Phi_N(z)$ and $y = \Psi_N(z)$. Define the linear operator A on L by

$$(3.19) \quad A = \frac{1}{2} \sum_{i, j=1}^m a_{ij}(x, y) \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^m b_i(x, y) \frac{\partial}{\partial x_i}, \quad \mathcal{D}(A) = C^2(E),$$

and let B be the generator of the semigroup $\{S(t)\}$ on L defined in Lemma 3.1 if $\delta_\infty = 0$ and in Lemma 3.3 if $\delta_\infty > 0$. Define P on L by $Ph(x, y) = h(x, 0)$, and let $D = \mathcal{D}(A) \cap \mathcal{R}(P)$ and $D' = C^2(E)$. By the lemmas, D' is a core for B , and (7.15) of Chapter 1 holds if $\delta_\infty = 0$, while (7.16) of that chapter holds (where Q is as in Lemma 3.3) if $\delta_\infty > 0$. Let $A_N = \varepsilon_N^{-1}(T_N - I)$, where T_N is defined on L_N by $T_N f(z) = E_z[f(Z^N(1))]$, and let $\alpha_N = \delta_N/\varepsilon_N$.

Given $f \in D$,

$$\begin{aligned} (3.20) \quad A_N \pi_N f(z) &= \varepsilon_N^{-1} E_z[f(X^N(1), y) - f(x, y)] \\ &= \sum_{i=1}^m \varepsilon_N^{-1} E_z[X_i^N(1) - x_i] f_{x_i}(x, y) \\ &\quad + \frac{1}{2} \sum_{i, j=1}^m \varepsilon_N^{-1} E_z[(X_i^N(1) - x_i)(X_j^N(1) - x_j)] f_{x_i x_j}(x, y) \\ &\quad + \sum_{i, j=1}^m \varepsilon_N^{-1} E_z \left[(X_i^N(1) - x_i)(X_j^N(1) - x_j) \right. \\ &\quad \left. \cdot \int_0^1 (1-u) \{ f_{x_i x_j}(x + u(X^N(1) - x), y) - f_{x_i x_j}(x, y) \} \, du \right] \\ &= Af(x, y) + o(1), \end{aligned}$$

where the first equality uses the fact that $f \in \mathcal{R}(P)$, the second uses the convexity of K , and the third depends on (3.12), (3.13), and (3.18). (To show that the remainder term in the Taylor expansion is $o(1)$, integrate separately over $|X^N(1) - x| \leq \gamma$ and $|X^N(1) - x| > \gamma$, using the Schwarz inequality, (3.13), and (3.18).) This implies (7.17) of Chapter 1.

Given $h \in D'$,

$$\begin{aligned}
 (3.21) \quad & \delta_N^{-1} E_z [h(X^N(1), Y^N(1)) - h(x, E_z[Y^N(1)])] \\
 &= \sum_{i=1}^n \delta_N^{-1} E_z [X_i^N(1) - x_i] h_{x_i}(x, y) \\
 &\quad + \frac{1}{2} \sum_{i,j=1}^n \delta_N^{-1} E_z [(X_i^N(1) - x_i)(X_j^N(1) - x_j)] \eta^N(h_{x_i x_j}, z, Z^N(1)) \\
 &\quad + \sum_{i=1}^n \sum_{j=1}^n \delta_N^{-1} E_z [(X_i^N(1) - x_i)(Y_j^N(1) - E_z[Y_j^N(1)])] \eta^N(h_{x_i y_j}, z, Z^N(1)) \\
 &\quad + \frac{1}{2} \sum_{i,j=1}^n \delta_N^{-1} E_z [(Y_i^N(1) - E_z[Y_i^N(1)]) \\
 &\quad \quad \times (Y_j^N(1) - E_z[Y_j^N(1)])] \eta^N(h_{y_i y_j}, z, Z^N(1))
 \end{aligned}$$

where

$$\begin{aligned}
 (3.22) \quad & \eta^N(g, z, Z^N(1)) \\
 &\equiv \int_0^1 (1-u)g(x+u(X^N(1)-x), E_z[Y^N(1)] + u(Y^N(1) - E_z[Y^N(1)])) du.
 \end{aligned}$$

(Here the convexity of H and of $K \times H$ is used.) But the right side of (3.21) is $o(1)$ by the Schwarz inequality, (3.12), (3.13), and (3.16). Consequently,

$$\begin{aligned}
 (3.23) \quad \alpha_N^{-1} A_N \pi_N h(z) &= \delta_N^{-1} \{h(x, E_z[Y^N(1)]) - h(x, y)\} + o(1) \\
 &= Bh(x, y) + o(1)
 \end{aligned}$$

by (3.15) and either (3.4) or (3.10). This implies (7.18) of Chapter 1.

Finally, define $\rho: K \rightarrow E$ by $\rho(x) = (x, 0)$, and observe that $G(f \circ \rho) = (PAf) \circ \rho$ for all $f \in D$. Since the closure of $\{(f, Gf): f \in C^2(K)\}$ is single-valued and generates a Feller semigroup $\{U(t)\}$ on $C(K)$, the Feller semigroup $\{T(t)\}$ on $\bar{D} = \mathcal{R}(P)$ satisfying

$$(3.24) \quad U(t)(f \circ \rho) = [T(t)f] \circ \rho, \quad f \in \bar{D}, t \geq 0,$$

is generated by the closure of $PA|_D$. Theorem 7.6(b) of Chapter 1, together with Corollary 8.9 of Chapter 4, yields conclusion (a) of the theorem, and Corollary 7.7 of Chapter 1 (with $h(x, y) = |y|$) yields conclusion (b). \square

3.7 Remark (a) Since $\lim_{N \rightarrow \infty} \varepsilon_N = 0$, (3.12) implies that (3.13) is equivalent to

$$(3.25) \quad \varepsilon_N^{-1} E_z[(X_i^N(1) - E_z[X_i^N(1)])(X_j^N(1) - E_z[X_j^N(1)])] = a_{ij}(x, y) + o(1)$$

for $i, j = 1, \dots, m$ and that (3.14) is equivalent to

$$(3.26) \quad \varepsilon_N^{-1} E_z[(X_i^N(1) - E_z[X_i^N(1)])^4] = o(1)$$

for $i = 1, \dots, m$. It is often more convenient to verify (3.25) and (3.26). We note also that, if $\lim_{N \rightarrow \infty} \delta_N = 0$, then (3.15) implies that (3.16) is equivalent to

$$(3.27) \quad \delta_N^{-1} E_z[(Y_l^N(1) - y_l)^2] = o(1)$$

for $l = 1, \dots, n$.

(b) It is sometimes possible to avoid explicit calculation of (3.16) by using the following inequalities. Let ξ and η be real random variables with means $\bar{\xi}$ and $\bar{\eta}$ such that $|\xi| \leq M$ a.s. and $|\eta| \leq M$ a.s. Then

$$(3.28) \quad \text{var } (\xi + \eta) \leq 2(\text{var } \xi + \text{var } \eta)$$

and

$$\begin{aligned} (3.29) \quad \text{var } (\xi\eta) &\leq E[(\xi\eta - \bar{\xi}\bar{\eta})^2] \\ &\leq 2E[(\xi - \bar{\xi})^2\eta^2] + 2\bar{\xi}^2E[(\eta - \bar{\eta})^2] \\ &\leq 2M^2(\text{var } \xi + \text{var } \eta). \end{aligned} \quad \square$$

In the remainder of this section we consider two genetic models in detail, showing that Theorem 3.5 is applicable to both of them. Although the models differ substantially, they have several features in common, and it may be worthwhile pointing these out explicitly beforehand.

Adopting the convention that coordinates of elements of $\mathbb{R}^{r(r+1)/2}$ are to be indexed by $\{(i, j): 1 \leq i \leq j \leq r\}$, the state space E_N of the underlying Markov chain Z^N in both cases is the space of genotypic frequencies

$$(3.30) \quad E_N = \left\{ N^{-1}v: v \in (\mathbb{Z}_+)^{r(r+1)/2}, \sum_{i \leq j} v_{ij} = N \right\}.$$

In applying Theorem 3.5, the transformations $\Phi_N: E \rightarrow \mathbb{R}^{r-1}$ and $\Psi_N: E_N \rightarrow \mathbb{R}^{r(r+1)/2}$ are given by

$$(3.31) \quad \Phi_N((P_{ij})_{i \leq j}) = (p_1, \dots, p_{r-1})$$

and

$$(3.32) \quad \Psi_N((P_{ij})_{i \leq j}) = (Q_{ij})_{i \leq j},$$

where p_i is the allelic frequency (1.1) and Q_{ij} is the Hardy-Weinberg deviation

$$(3.33) \quad Q_{ij} = P_{ij} - (2 - \delta_{ij})p_i p_j.$$

Observe that $\Phi_N(E_N) \subset K$, where K is defined by (1.16). As we see, in both of our examples, the functions $a: K \times \mathbb{R}^{r(r+1)/2} \rightarrow \mathbb{R}^{r-1} \otimes \mathbb{R}^{r-1}$ and $b: K \times \mathbb{R}^{r(r+1)/2} \rightarrow \mathbb{R}^{r-1}$ are such that $a_i(p, 0)$ and $b_i(p, 0)$ are given by the right sides of (1.18) and (1.19). Consequently, the condition on G in Theorem 3.5 is satisfied. In addition, the function $c: K \times \mathbb{R}^{r(r+1)/2} \rightarrow \mathbb{R}^{r(r+1)/2}$ is seen to trivially satisfy the conditions of either Remark 3.2 or Remark 3.4. (Hence H can be taken to be an arbitrary compact, convex set containing $\bigcup_{N \geq 1} \Psi_N(E_N)$.)

Thus, to apply Theorem 3.5, it suffices in each case to specify the transition function, starting at $(P_{ij})_{i \leq j} \in E_N$, of the Markov chain Z^N , and to verify the five moment conditions (3.12)–(3.16) for appropriately chosen sequences $\{\varepsilon_N\}$ and $\{\delta_N\}$.

Before proceeding, we introduce a useful computational device, which already appeared in (1.7) without explicit mention. Given $(d_{ij})_{i \leq j} \in \mathbb{R}^{r(r+1)/2}$, we define

$$(3.34) \quad \tilde{d}_{ij} = \frac{1}{2}(1 + \delta_{ij}) d_{i \wedge j, i \vee j}, \quad i, j = 1, \dots, r.$$

We apply this symmetrization to P_{ij} , P_{ij}^* , P_{ij}^{**} , P'_{ij} , Q_{ij} , Q'_{ij} , and so on. The point is that (1.1) can be expressed more concisely as $p_i = \sum_j \tilde{P}_{ij}$. For later reference, we isolate the following simple identity. With $(d_{ij})_{i \leq j}$ as above,

$$\begin{aligned} (3.35) \quad & \sum_{k, l} (\delta_{ij} \delta_{kl} \vee \delta_{il} \delta_{jk})(1 + \delta_{jl}) \tilde{d}_{lk} \\ &= \sum_{k, l} [\delta_{ij} \delta_{kl} + \delta_{il} \delta_{jk}(1 - \delta_{ij} \delta_{kl})](1 + \delta_{jl}) \tilde{d}_{lk} \\ &= \delta_{ij} \sum_k \tilde{d}_{lk} + \tilde{d}_{ij}, \quad i, j = 1, \dots, r. \end{aligned}$$

3.8 Example We consider first the multinomial-sampling model described in Section 1. The transition function of Z^N , starting at $(P_{ij})_{i \leq j} \in E_N$, is specified by (1.8), (1.4), (1.5), (1.2), (1.1), and (1.14).

Since $E[P'_{ij}] = P_{ij}^{**}$, we have

$$(3.36) \quad 2NE[p'_i - p_i] = 2N(p_i^{**} - p_i) = b_i(p) + O(N^{-1}),$$

where $b: K \rightarrow \mathbb{R}^{r-1}$ is given by (1.19). (Throughout, all O and o terms are uniform in the genotypic frequencies.) The relation $\text{cov}(P'_{ik}, P'_{jl}) = N^{-1}P_{ik}^{**}(\delta_{ij} \delta_{kl} - P_{jl}^{**})$ implies

$$(3.37) \quad \text{cov}(\tilde{P}'_{ik}, \tilde{P}'_{jl}) = N^{-1}[\frac{1}{2}(\delta_{ij} \delta_{kl} \vee \delta_{il} \delta_{jk})(1 + \delta_{jl})\tilde{P}_{ik}^{**} - \tilde{P}_{ik}^{**}\tilde{P}_{jl}^{**}],$$

and therefore, by (3.35),

$$\begin{aligned} (3.38) \quad 2N \text{cov}(p'_i, p'_j) &= \sum_{k, l} 2N \text{cov}(\tilde{P}'_{ik}, \tilde{P}'_{jl}) \\ &= \delta_{ij} p_i^{**} + \tilde{P}_{ij}^{**} - 2p_i^{**}p_j^{**} \\ &= p_i^{**}(\delta_{ij} - p_j^{**}) + \tilde{P}_{ij}^{**} - p_i^{**}p_j^{**}. \end{aligned}$$

This shows, incidentally, that (1.10) is not only sufficient for (1.12) but necessary as well. Now observe that $p_i^{**} = p_i + O(N^{-1})$ by (3.36) and

$$(3.39) \quad \begin{aligned} \tilde{P}_{ij}^{**} &= \frac{1}{2} \sum_{k \leq l} (\delta_{kl} \delta_{ij} + \delta_{kj} \delta_{il})(2 - \delta_{kl})p_k p_l + O(N^{-1}) \\ &= p_i p_j + O(N^{-1}), \end{aligned}$$

so

$$(3.40) \quad 2N \operatorname{cov}(p'_i, p'_j) = p_i(\delta_{ij} - p_j) + O(N^{-1}).$$

Next, we note that

$$(3.41) \quad \begin{aligned} E[Q'_{ij} - Q_{ij}] &= P_{ij}^{**} - (2 - \delta_{ij})[\operatorname{cov}(p'_i, p'_j) + p_i^{**} p_j^{**}] - Q_{ij} \\ &= -Q_{ij} + O(N^{-1}) \end{aligned}$$

by (3.39) and (3.40).

Finally,

$$(3.42) \quad \begin{aligned} 2NE[(p'_i - E[p'_i])^4] &\leq 2Nr^3 \sum_{j=1}^r E[(\tilde{P}'_{ij} - E[\tilde{P}'_{ij}])^4] \\ &= O(N^{-1}) \end{aligned}$$

since $P'_{ij} \sim N^{-1}$ binomial ($N, E[P'_{ij}]$) for each $i \leq j$, and the fourth central moment of N^{-1} binomial (N, p) is $O(N^{-2})$, uniformly in p . Also,

$$(3.43) \quad \begin{aligned} \operatorname{var}(Q'_{ij}) &\leq 2 \operatorname{var}(P'_{ij}) + 2(2 - \delta_{ij})^2 \operatorname{var}(p'_i p'_j) \\ &\leq O(N^{-1}) + 4(2 - \delta_{ij})^2(\operatorname{var}(p'_i) + \operatorname{var}(p'_j)) \\ &= O(N^{-1}) \end{aligned}$$

by (3.28) and (3.29). This completes the verification of conditions (3.12)–(3.16) of Theorem 3.5 (see Remark 3.7(a)) with $\epsilon_N = (2N)^{-1}$ and $\delta_N = 1$.

We note that the limits as $N \rightarrow \infty$ of the right sides of (3.36) and (3.40) depend only on p_1, \dots, p_{r-1} . (For this reason, Theorem 3.5 could easily be avoided here.) However, this is not typical, as other examples suggest. \square

3.9 Example The next genetic model we consider is a generalization of a model due to Moran. Its key feature is that, in contrast to the multinomial-sampling model of Example 3.8, generations are overlapping. A single step of the Markov chain here corresponds to the death of an individual and its replacement by the birth of another.

Suppose the genotype $A_i A_j$ produces gametes with fertility $w_{ij}^{(1)}$ and has mortality rate $w_{ij}^{(2)}$. If $(P_{ij})_{i \leq j} \in E_N$ is the initial vector of genotypic frequencies, then the probability that an $A_i A_j$ individual dies is

$$(3.44) \quad \gamma_{ij} = \frac{w_{ij}^{(2)} P_{ij}}{\sum_{k \leq l} w_{ki}^{(2)} P_{kl}}.$$

The frequency of A_i in gametes before mutation reads

$$(3.45) \quad p_i^* = \frac{\sum_j w_{ij}^{(1)} \tilde{P}_{ij}}{\sum_{k,l} w_{kl}^{(1)} \tilde{P}_{kl}},$$

where $w_{ji}^{(1)} \equiv w_{ij}^{(1)}$ for $1 \leq i < j \leq r$. With mutation rates u_{ij} (where $u_{ii} = 0$), this becomes

$$(3.46) \quad p_i^{**} = \left(1 - \sum_j u_{ij} \right) p_i^* + \sum_j u_{ji} p_j^*$$

after mutation, so the probability that an $A_i A_j$ individual is born has the form

$$(3.47) \quad \beta_{ij} = (2 - \delta_{ij}) p_i^{**} p_j^{**}.$$

Consequently, the joint distribution of genotypic frequencies P'_{ij} after a birth-death event is specified as follows. For each $(k, l) \neq (m, n)$ (with $1 \leq k \leq l \leq r, 1 \leq m \leq n \leq r$),

$$(3.48) \quad P'_{ij} = \begin{cases} P_{ij} + N^{-1} & \text{if } (i, j) = (m, n) \\ P_{ij} - N^{-1} & \text{if } (i, j) = (k, l) \\ P_{ij} & \text{otherwise} \end{cases}$$

with probability $\gamma_{kl} \beta_{mn}$, and $P'_{ij} = P_{ij}$ for all $i \leq j$ with probability $\sum_{k \leq l} \gamma_{kl} \beta_{kl}$. If we further require that

$$(3.49) \quad u_{ij} = (N^{-1} \mu_{ij}) \wedge r^{-1}, \quad w_{ij}^{(k)} = (1 + N^{-1} \sigma_{ij}^{(k)}) \vee \frac{1}{2}, \quad k = 1, 2,$$

where $\mu_{ij} \geq 0$ (with $\mu_{ii} = 0$) and $\sigma_{ij}^{(k)} (\sigma_{ji}^{(k)})$ is real for $i, j = 1, \dots, r$, then the transition function of Z^N , starting at $(P_{ij}) \in E_N$, is specified by (3.44)–(3.49).

To evaluate the appropriate moments, observe that

$$(3.50) \quad E[P'_{ij} - P_{ij}] = N^{-1}[-\gamma_{ij}(1 - \beta_{ij}) + (1 - \gamma_{ij})\beta_{ij}] = N^{-1}(\beta_{ij} - \gamma_{ij})$$

and

$$(3.51) \quad E[(P'_{ik} - P_{ik})(P'_{jl} - P_{jl})] = N^{-2}[-(\gamma_{ik}\beta_{jl} + \gamma_{jl}\beta_{ik}) + \delta_{ij}\delta_{kl}(\gamma_{ik} + \beta_{ik})].$$

Noting that $\tilde{\gamma}_{ij} = w_{ij}^{(2)} \tilde{P}_{ij} / \sum_{k,l} w_{kl}^{(2)} \tilde{P}_{kl}$ and $\tilde{\beta}_{ij} = p_i^{**} p_j^{**}$, where $w_{ji}^{(2)} \equiv w_{ij}^{(2)}$ for $1 \leq i < j \leq r$, and letting $\gamma_i = \sum_j \tilde{\gamma}_{ij}$, we have

$$\begin{aligned} (3.52) \quad N^2 E[p'_i - p_i] &= N^2 \sum_j E[\tilde{P}'_{ij} - \tilde{P}_{ij}] \\ &= N \sum_j (\tilde{\beta}_{ij} - \tilde{\gamma}_{ij}) \\ &= N(p_i^{**} - \gamma_i) \\ &= -\sum_j \mu_{ij} p_i + \sum_j \mu_{ji} p_j + \sum_j \sigma_{ij} \tilde{P}_{ij} \\ &\quad - p_i \sum_{k,l} \sigma_{kl} \tilde{P}_{kl} + O(N^{-1}), \end{aligned}$$

where $\sigma_{ij} = \sigma_{ij}^{(1)} - \sigma_{ij}^{(2)}$ is the difference between the scaled fertility and mortality, and

$$\begin{aligned}
 (3.53) \quad N^2 E[(p'_i - p_i)(p'_j - p_j)] &= N^2 \sum_{k,l} E[(\tilde{P}'_{ik} - \tilde{P}_{ik})(\tilde{P}'_{jl} - \tilde{P}_{jl})] \\
 &= \sum_{k,l} [-(\tilde{\gamma}_{ik}\tilde{\beta}_{jl} + \tilde{\gamma}_{jl}\tilde{\beta}_{ik}) \\
 &\quad + \frac{1}{2}(\delta_{ij}\delta_{kl} \vee \delta_{il}\delta_{jk})(1 + \delta_{jl})(\tilde{\gamma}_{ik} + \tilde{\beta}_{ik})] \\
 &= -(\gamma_i p_j^{**} + \gamma_j p_i^{**}) + \frac{1}{2}[\delta_{ij}(\gamma_i + p_i^{**}) + \tilde{\gamma}_{ij} + \tilde{\beta}_{ij}] \\
 &= -2p_i p_j + \frac{1}{2}(2\delta_{ij}p_i + \tilde{P}_{ij} + p_i p_j) + O(N^{-1}) \\
 &= p_i(\delta_{ij} - p_j) + \frac{1}{2}\tilde{Q}_{ij} + O(N^{-1}),
 \end{aligned}$$

where the third equality uses (3.35).

We also have

$$\begin{aligned}
 (3.54) \quad NE[Q'_{ij} - Q_{ij}] &= NE[P'_{ij} - P_{ij}] + O(N^{-1}) \\
 &= \beta_{ij} - \gamma_{ij} + O(N^{-1}) \\
 &= -Q_{ij} + O(N^{-1}),
 \end{aligned}$$

so since $|P'_{ij} - P_{ij}| \leq N^{-1}$ with probability one, the conditions (3.12)–(3.16) of Theorem 3.5 are satisfied with $\varepsilon_N = N^{-2}$ and $\delta_N = N^{-1}$ (recall (3.27)). Thus, the theorem is applicable to this model as well. \square

4. INFINITELY-MANY-ALLEL MODELS

In the absence of selection, the Wright–Fisher model (defined by (1.12), (1.6) with $p_k^* = p_k$, and (1.5)) can be described as follows. Each of the $2N$ genes in generation $k+1$ selects a “parent” gene at random (with replacement) from generation k . If the “parent” gene is of allelic type A_i , then its “offspring” gene is of allelic type A_j with probability u_{ij}^* .

In this section we consider a generalization of this model, as well as its diffusion limit. Let E be a compact metric space. E is the space of “types.” For each positive integer M , let $P_M(x, \Gamma)$ be a transition function on $E \times \mathcal{B}(E)$, and define a Markov chain $\{Y^M(k), k = 0, 1, \dots\}$ in $E^M = E \times \dots \times E$ (M factors) as follows, where $Y_i^M(k)$ represents the type of the i th individual in generation k . Each of the M individuals in generation $k+1$ selects a parent at random (with replacement) from generation k . If the parent is of type x , then its offspring’s type belongs to Γ with probability $P_M(x, \Gamma)$. In particular,

$$\begin{aligned}
 (4.1) \quad E\left[\prod_{i=1}^M f_i(Y_{j_i}^M(k+1)) \middle| Y^M(k) = (x_1, \dots, x_M)\right] \\
 = \prod_{i=1}^M \left\{ \frac{1}{M} \sum_{j=1}^M \int f_i(y) P_M(x_j, dy) \right\}
 \end{aligned}$$

if $j_1, \dots, j_l \in \{1, \dots, M\}$ are distinct and $f_1, \dots, f_l \in B(E)$, since the components of $Y^M(k+1)$ are conditionally independent given $Y^M(k)$.

Observe that the process

$$(4.2) \quad X^M(t) = \sum_{i=1}^M \frac{1}{M} \delta_{Y_i^M([Mt])}, \quad t \geq 0,$$

has sample paths in $D_{\mathcal{P}(E)}[0, \infty)$. Our first result gives conditions under which there exists a Markov process X with sample paths in $D_{\mathcal{P}(E)}[0, \infty)$ such that $X^M \Rightarrow X$.

Suppose that B is a linear operator on $C(E)$, and let

$$(4.3) \quad \mathcal{D} = \left\{ \varphi = \prod_{i=1}^l \langle f_i, \cdot \rangle : l \geq 1, f_1, \dots, f_l \in \mathcal{D}(B) \right\} \subset C(\mathcal{P}(E)),$$

where $\langle f, \mu \rangle$ denotes $\int f d\mu$ for $f \in B(E)$ and $\mu \in \mathcal{P}(E)$. Given $\varphi = \prod_{i=1}^l \langle f_i, \cdot \rangle \in \mathcal{D}$, define

$$(4.4) \quad G\varphi(\mu) = \frac{1}{2} \sum_{i \neq j} (\langle f_i f_j, \mu \rangle - \langle f_i, \mu \rangle \langle f_j, \mu \rangle) \prod_{m: m \neq i, j} \langle f_m, \mu \rangle \\ + \sum_{i=1}^l \langle Bf_i, \mu \rangle \prod_{j: j \neq i} \langle f_j, \mu \rangle,$$

and let

$$(4.5) \quad A = \{(\varphi, G\varphi) : \varphi \in \mathcal{D}\}.$$

4.1 Theorem Suppose that the linear operator B on $C(E)$ is dissipative, $\mathcal{D}(B)$ is dense in $C(E)$, and the closure of B (which is single-valued) contains $(1, 0)$. Then, for each $\tilde{v} \in \mathcal{P}(\mathcal{P}(E))$, there exists a solution of the $D_{\mathcal{P}(E)}[0, \infty)$ martingale problem for (A, \tilde{v}) . If the closure of B generates a Feller semigroup on $C(E)$, then the martingale problem for A is well-posed.

For $M = 1, 2, \dots$, define X^M in terms of $P_M(x, \Gamma)$ as in (4.2), and define Q_M and B_M on $B(E)$ by

$$(4.6) \quad Q_M f(x) = \int f(y) P_M(x, dy), \quad B_M = M(Q_M - I).$$

If the closure of B generates a Feller semigroup on $C(E)$, if

$$(4.7) \quad B \subset \text{ex-lim}_{M \rightarrow \infty} B_M,$$

and if X is a solution of the $D_{\mathcal{P}(E)}[0, \infty)$ martingale problem for A , then $X^M(0) \Rightarrow X(0)$ in $\mathcal{P}(E)$ implies $X^M \Rightarrow X$ in $D_{\mathcal{P}(E)}[0, \infty)$.

Proof. Under the first set of hypotheses on B , Lemma 5.3 of Chapter 4 implies the existence of a sequence of transition functions $P_M(x, \Gamma)$ on $E \times \mathcal{D}(E)$ such that the operators B_M , defined by (4.6), satisfy $\lim_{M \rightarrow \infty} B_M f = Bf$ for all $f \in \mathcal{D}(B)$. In particular, (4.7) holds.

Hence it suffices to prove existence of solutions of the martingale problem for A assuming that $\mathcal{D}(B)$ is dense in $C(E)$ and (4.7) holds. Let $\varphi = \prod_{i=1}^k \langle f_i, \cdot \rangle \in \mathcal{D}$, choose $f_1^M, \dots, f_k^M \in B(E)$ such that $f_i^M \rightarrow f_i$ and $B_M f_i^M \rightarrow Bf_i$ for $i = 1, \dots, k$, and put $\varphi_M = \prod_{i=1}^k \langle f_i^M, \cdot \rangle$. Given $\mu = M^{-1} \sum_{j=1}^M \delta_{x_j}$, where $(x_1, \dots, x_M) \in E^M$, we have

$$\begin{aligned}
 (4.8) \quad & E\left[\varphi_M\left(X^M\left(\frac{1}{M}\right)\right) \middle| X^M(0) = \mu\right] \\
 &= M^{-k} E\left[\prod_{i=1}^k \sum_{j=1}^M f_i^M(Y_j^M(1)) \middle| Y^M(0) = (x_1, \dots, x_M)\right] \\
 &= M^{-k} E\left[\sum_{j_1=1}^M \cdots \sum_{j_k=1}^M \prod_{i=1}^k f_i^M(Y_{j_i}^M(1)) \middle| Y^M(0) = (x_1, \dots, x_M)\right] \\
 &= M^{-k} \frac{M!}{(M-k)!} \prod_{i=1}^k \langle Q_M f_i^M, \mu \rangle \\
 &\quad + M^{-k} \frac{M!}{(M-k+1)!} \sum_{l < m} \langle Q_M f_l^M f_m^M, \mu \rangle \prod_{n: n \neq l, m} \langle Q_M f_n^M, \mu \rangle \\
 &\quad + O(M^{-2}),
 \end{aligned}$$

where the factor $M!/(M-k)!$ is the number of ways of selecting j_1, \dots, j_k so that they are all distinct, and $M!/(M-k+1)!$ is the number of ways of selecting j_1, \dots, j_k so that $j_l = j_m$ (l, m fixed) but they are otherwise distinct. Hence

$$\begin{aligned}
 (4.9) \quad & M E\left[\varphi_M\left(X^M\left(\frac{1}{M}\right)\right) - \varphi_M(\mu) \middle| X^M(0) = \mu\right] \\
 &= M^{-k} \frac{M!}{(M-k)!} \sum_{i=1}^k \left(\prod_{j: j < i} \langle f_j^M, \mu \rangle \right) \langle B_M f_i^M, \mu \rangle \left(\prod_{j: j > i} \langle Q_M f_j^M, \mu \rangle \right) \\
 &\quad + M^{-k+1} \frac{M!}{(M-k+1)!} \\
 &\quad \times \frac{1}{2} \sum_{l \neq m} \left\{ \langle Q_M f_l^M f_m^M, \mu \rangle \prod_{n: n \neq l, m} \langle Q_M f_n^M, \mu \rangle - \prod_{i=1}^k \langle f_i^M, \mu \rangle \right\} \\
 &\quad + O(M^{-1}) \\
 &= G\varphi(\mu) + o(1),
 \end{aligned}$$

and the convergence is uniform in μ of the form $\mu = M^{-1} \sum_{j=1}^M \delta_{x_j}$, where $(x_1, \dots, x_M) \in E^M$. Here we are using the fact that, since $\mathcal{D}(B)$ is dense in $C(E)$, $Q_M f \rightarrow f$ for every $f \in C(E)$. As in Remark 5.2 of Chapter 4, it follows from Theorems 9.1 and 9.4, both of Chapter 3, that $\{X^M\}$ is relatively compact. As in Lemma 5.1 of Chapter 4 we conclude that for each $\tilde{v} \in \mathcal{P}(\mathcal{P}(E))$, there exists a solution of the martingale problem for (A, \tilde{v}) .

Fix $\tilde{v} \in \mathcal{P}(\mathcal{P}(E))$. To complete the proof, it will suffice by Corollary 8.17 of Chapter 4 to show that the martingale problem for (A, \tilde{v}) has a unique solution, assuming that \bar{B} generates a Feller semigroup $\{S(t)\}$ on $C(E)$. Let X be a solution of the martingale problem for (A, \tilde{v}) . By Corollary 3.7 of Chapter 4, X has a modification X^* with sample paths in $D_{\mathcal{P}(E)}[0, \infty)$. Let $f_1, \dots, f_k \in \mathcal{D}(\bar{B})$. Then

$$(4.10) \quad \frac{\partial}{\partial t} \prod_{i=1}^k \langle S(u-t)f_i, \mu \rangle = - \sum_{i=1}^k \langle \bar{B}S(u-t)f_i, \mu \rangle \prod_{j: j \neq i} \langle S(u-t)f_j, \mu \rangle$$

for $0 \leq t \leq u$ and all $\mu \in \mathcal{P}(E)$, so by Lemma 3.4 of Chapter 4,

$$(4.11) \quad \begin{aligned} & \prod_{i=1}^k \langle S(u-t \wedge u)f_i, X^*(t \wedge u) \rangle \\ & - \int_0^{t \wedge u} \frac{1}{2} \sum_{i \neq m} \{ \langle S(u-s)f_i S(u-s)f_m, X^*(s) \rangle - \langle S(u-s)f_i, X^*(s) \rangle \\ & \quad \cdot \langle S(u-s)f_m, X^*(s) \rangle \} \prod_{n: n \neq i, m} \langle S(u-s)f_n, X^*(s) \rangle ds \end{aligned}$$

is a martingale in t for each $u \geq 0$. Hence

$$(4.12) \quad \begin{aligned} E \left[\prod_{i=1}^k \langle f_i, X(u) \rangle \right] &= \int \prod_{i=1}^k \langle S(u)f_i, \mu \rangle \tilde{v}(d\mu) \\ &+ \frac{1}{2} \int_0^u \sum_{i \neq m} E \left[\langle S(u-s)f_i S(u-s)f_m, X(s) \rangle \right. \\ & \quad \times \left. \prod_{n: n \neq i, m} \langle S(u-s)f_n, X(s) \rangle \right] ds \\ &- \binom{k}{2} \int_0^u E \left[\prod_{i=1}^k \langle S(u-s)f_i, X(s) \rangle \right] ds. \end{aligned}$$

Moreover, (4.12) holds for all $f_1, \dots, f_k \in C(E)$ since $\mathcal{D}(B)$ is dense in $C(E)$.

Let Y be another solution of the martingale problem for (A, \tilde{v}) , and put

$$(4.13) \quad \beta_k(u) = \sup_{f_1, \dots, f_k \in C(E), \|f_i\| \leq 1} \left| E \left[\prod_{i=1}^k \langle f_i, X(u) \rangle - \prod_{i=1}^k \langle f_i, Y(u) \rangle \right] \right|.$$

Then, since (4.12) holds with X replaced by Y ,

$$(4.14) \quad \beta_k(u) \leq k(k-1) \int_0^u \beta_k(s) \, ds, \quad u \geq 0.$$

We conclude that $\beta_k(u) = 0$ for all $k \geq 1$ and $u \geq 0$, and hence that X and Y have the same one-dimensional distributions. Uniqueness then follows from Theorem 4.2 of Chapter 4. \square

The process X of Theorem 4.1 is therefore characterized by its type space E and the linear operator B on $C(E)$. Let E_1, E_2, \dots and E be compact metric spaces. For $n = 1, 2, \dots$, let $\eta_n: E_n \rightarrow E$ be continuous, and define $\pi_n: C(E) \rightarrow C(E_n)$ by $\pi_n f = f \circ \eta_n$, $\hat{\eta}_n: \mathcal{P}(E_n) \rightarrow \mathcal{P}(E)$ by $\hat{\eta}_n \mu = \mu \eta_n^{-1}$, and $\hat{\pi}_n: C(\mathcal{P}(E)) \rightarrow C(\mathcal{P}(E_n))$ by $\hat{\pi}_n f = f \circ \hat{\eta}_n$.

4.2 Proposition Let B_1, B_2, \dots and B be linear operators on $C(E_1), C(E_2), \dots$ and $C(E)$, respectively, satisfying the conditions in the first sentence of Theorem 4.1. Define A_1, A_2, \dots and A in terms of E_1, E_2, \dots and E and B_1, B_2, \dots and B as in (4.3)–(4.5). For $n = 1, 2, \dots$, let X_n be a solution of the $D_{\mathcal{P}(E_n)}[0, \infty)$ martingale problem for A_n . If the closure of B generates a Feller semigroup on $C(E)$, if

$$(4.15) \quad B \subset \text{ex-lim}_{n \rightarrow \infty} B_n \text{ (with respect to } \{\pi_n\}),$$

and if X is a solution of the $D_{\mathcal{P}(E)}[0, \infty)$ martingale problem for A , then $\hat{\eta}_n(X_n(0)) \Rightarrow X(0)$ in $\mathcal{P}(E)$ implies $\hat{\eta}_n \circ X_n \Rightarrow X$ in $D_{\mathcal{P}(E)}[0, \infty)$.

Proof. By (4.15),

$$(4.16) \quad A \subset \text{ex-lim}_{n \rightarrow \infty} A_n \text{ (with respect to } \{\hat{\pi}_n\}),$$

so the result follows from Corollary 8.16 of Chapter 4. \square

We give two examples of Proposition 4.2. In both, E_n is a subset of E , η_n is an inclusion map, and hence $\hat{\eta}_n$ can be regarded as an inclusion map (that is, elements of $\mathcal{P}(E_n)$ can be regarded as belonging to $\mathcal{P}(E)$). With this understanding, we can suppress the notation η_n and $\hat{\eta}_n$.

4.3 Example For $n = 1, 2, \dots$ let $E_n = \{k/\sqrt{n} : k \in \mathbb{Z}\} \cup \{\Delta\}$ (the one-point compactification), define B_n to be the bounded linear operator on $C(E_n)$ given by

$$(4.17) \quad B_n f\left(\frac{k}{\sqrt{n}}\right) = \sigma^2 n \left\{ \frac{1}{2} f\left(\frac{k-1}{\sqrt{n}}\right) + \frac{1}{2} f\left(\frac{k+1}{\sqrt{n}}\right) - f\left(\frac{k}{\sqrt{n}}\right) \right\}$$

and $B_n f(\Delta) = 0$, where $\sigma^2 \geq 0$, and let X_n be a solution of the $D_{\mathcal{P}(E_n)}[0, \infty)$ martingale problem for A_n (defined as in (4.3)–(4.5)). X_1 is known as the Ohta–Kimura model.

Let $E = \mathbb{R} \cup \{\Delta\}$ (the one-point compactification), define B to be the linear operator on $C(E)$ given by

$$(4.18) \quad Bf(x) = \frac{1}{2} \sigma^2 f''(x)$$

and $Bf(\Delta) = 0$, where $\mathcal{D}(B) = \{f \in C(E) : (f - f(\Delta))|_{\mathbb{R}} \in C_c^2(\mathbb{R})\}$, and let X be a solution of the $D_{\mathcal{P}(E)}[0, \infty)$ martingale problem for A (defined by (4.3)–(4.5)). X is known as the Fleming–Viot model.

By Proposition 1.1 of Chapter 5 and Proposition 4.2 of this chapter, $X_n(0) \Rightarrow X(0)$ in $\mathcal{P}(E)$ implies $X_n \Rightarrow X$ in $D_{\mathcal{P}(E)}[0, \infty)$. (Recall that we are regarding $\mathcal{P}(E_n)$ as a subset of $\mathcal{P}(E)$.) The use of one-point compactifications here is only so that Theorem 4.1 and Proposition 4.2 will apply. It is easy to see that, for example, $P\{X(0)(\mathbb{R}) = 1\} = 1$ implies $P\{X(t)(\mathbb{R}) = 1 \text{ for all } t > 0\} = 1$. \square

4.4 Example For $n = 2, 3, \dots$, let $E_n = \{1/n, 2/n, \dots, 1\}$, define B_n to be the bounded linear operator on $C(E_n)$ given by

$$(4.19) \quad B_n f\left(\frac{i}{n}\right) = \frac{1}{2} \frac{\theta}{n-1} \sum_{j=1}^n \left\{ f\left(\frac{j}{n}\right) - f\left(\frac{i}{n}\right) \right\},$$

where $\theta > 0$, and let X_n be a solution of the martingale problem for A_n (defined as in (4.3)–(4.5)). Observe that $X_n(t) = \sum_{i=1}^r p_i(t) \delta_{i/r}$, where $(p_1(t), \dots, p_{r-1}(t))$ is the diffusion process of Section 1 with $\sum_{j:j \neq i} \mu_{ij} = \theta/2$ for $i = 1, \dots, r$, $\mu_{ij} (i \neq j)$ independent of i and j , and $\sigma_{ij} = 0$ for $i, j = 1, \dots, r$. Thus, X , could be called the neutral r -allele model with uniform mutation. (The term “neutral” refers to the lack of selection.)

Let $E = [0, 1]$, define B to be the bounded linear operator on $C[0, 1]$ given by

$$(4.20) \quad Bf(x) = \frac{1}{2} \theta \int_0^1 (f(y) - f(x)) dy = \frac{1}{2} \theta (\langle f, \lambda \rangle - f(x)),$$

where λ denotes Lebesgue measure on $[0, 1]$, and let X be a solution of the $D_{\mathcal{P}[0, 1]}[0, \infty)$ martingale problem for A (defined by (4.3)–(4.5)). We call X the infinitely-many-neutral-alleles model with uniform mutation.

By Proposition 4.2, $X_n(0) \Rightarrow X(0)$ in $\mathcal{P}[0, 1]$ implies $X_n \Rightarrow X$ in $D_{\mathcal{P}[0, 1]}[0, \infty)$. (Again, we are regarding $\mathcal{P}(E_n)$ as a subset of $\mathcal{P}[0, 1]$.) Of course, with $E = [0, 1]$ and

$$(4.21) \quad P_M(x, dy) = \left(1 - \frac{\theta}{2M}\right)\delta_x(dy) + \frac{\theta}{2M} \lambda(dy), \quad M \geq \theta,$$

in Theorem 4.1, we have $X^M(0) \Rightarrow X(0)$ in $\mathcal{P}[0, 1]$ implies $X^M \Rightarrow X$ in $D_{\mathcal{P}[0, 1]}[0, \infty)$. Thus, X can be thought of as either a limit in distribution of certain $(n - 1)$ -dimensional diffusions as $n \rightarrow \infty$, or as a limit in distribution of a certain sequence of infinite-dimensional Markov chains. \square

The remainder of this section is devoted to a more detailed examination of the infinitely-many-neutral-alleles model with uniform mutation.

4.5 Theorem Given $\tilde{v} \in \mathcal{P}(\mathcal{P}[0, 1])$, let X be as in Example 4.4 with initial distribution \tilde{v} . (In other words, X is the process of Theorem 4.1 with $E = [0, 1]$, with B defined on $C[0, 1]$ by $Bf = \frac{1}{2}\theta(\langle f, \lambda \rangle - f)$, where $\theta > 0$ and λ is Lebesgue measure on $[0, 1]$, and with initial distribution \tilde{v} .) Then almost all sample paths of X belong to $C_{\mathcal{P}[0, 1]}[0, \infty)$, and

$$(4.22) \quad P\{X(t) \in \mathcal{P}_a[0, 1] \text{ for all } t > 0\} = 1,$$

where $\mathcal{P}_a[0, 1]$ denotes the set of purely atomic Borel probability measures on $[0, 1]$.

Proof. Using the notation of Example 4.4, let X_n have initial distribution $\tilde{v}_n \in \mathcal{P}(\mathcal{P}(E_n))$, where the sequence $\{\tilde{v}_n\}$ is chosen so that $\tilde{v}_n \Rightarrow \tilde{v}$ on $\mathcal{P}[0, 1]$. Then $X_n \Rightarrow X$ in $D_{\mathcal{P}[0, 1]}[0, \infty)$, and since $C_{\mathcal{P}[0, 1]}[0, \infty)$ is a closed subset of $D_{\mathcal{P}[0, 1]}[0, \infty)$, we have

$$(4.23) \quad 1 = \overline{\lim}_{n \rightarrow \infty} P\{X_n \in C_{\mathcal{P}[0, 1]}[0, \infty)\} \leq P\{X \in C_{\mathcal{P}[0, 1]}[0, \infty)\}$$

by Theorem 3.1 of Chapter 3.

The proof of the second assertion is more complicated. Observe first that for $f \in C[0, 1]$,

$$(4.24) \quad M_f(t) \equiv \langle f, X(t) \rangle - \langle f, X(0) \rangle - \int_0^t \frac{1}{2} \theta(\langle f, \lambda \rangle - \langle f, X(s) \rangle) ds$$

is a continuous, square-integrable martingale, and (see Problem 29 of Chapter 2) its increasing process has the form

$$(4.25) \quad \langle M_f \rangle_t = \int_0^t (\langle f^2, X(s) \rangle - \langle f, X(s) \rangle^2) ds.$$

Consequently, if $\gamma \geq 2$ and $f, g \in C[0, 1]$ with $f, g \geq 0$, Itô's formula (Theorem 2.9 of Chapter 5) implies that

$$(4.26) \quad M_f^\gamma(t) \equiv \langle f, X(t) \rangle^\gamma - \langle f, X(0) \rangle^\gamma$$

$$- \int_0^t \left\{ \binom{\gamma}{2} (\langle f^2, X(s) \rangle - \langle f, X(s) \rangle^2) \langle f, X(s) \rangle^{\gamma-2} \right.$$

$$\left. + \frac{\gamma}{2} \theta (\langle f, \lambda \rangle - \langle f, X(s) \rangle) \langle f, X(s) \rangle^{\gamma-1} \right\} ds$$

is a continuous, square-integrable martingale, and

$$(4.27) \quad \langle M_f^\gamma, M_g^\gamma \rangle_t = \gamma^2 \int_0^t \langle f, X(s) \rangle^{\gamma-1} (\langle fg, X(s) \rangle$$

$$- \langle f, X(s) \rangle \langle g, X(s) \rangle) \langle g, X(s) \rangle^{\gamma-1} ds.$$

(Note that $\langle \cdot, \cdot \rangle$ is used in two different ways here.) Let us define $\varphi_{n,\gamma}: \mathcal{P}[0, 1] \rightarrow [0, \infty)$ for each $\gamma > 0$ and $n = 1, 2, \dots$ by

$$(4.28) \quad \varphi_{n,\gamma}(\mu) = \mu \left(\left[0, \frac{1}{2^n} \right] \right)^\gamma + \mu \left(\left[\frac{1}{2^n}, \frac{2}{2^n} \right] \right)^\gamma + \cdots + \mu \left(\left[1 - \frac{1}{2^n}, 1 \right] \right)^\gamma.$$

It follows that, for each $\gamma \geq 2$ and $n = 1, 2, \dots$,

$$(4.29) \quad Z_{n,\gamma}(t) \equiv \varphi_{n,\gamma}(X(t)) - \varphi_{n,\gamma}(X(0))$$

$$+ \int_0^t \left\{ \binom{\gamma}{2} (\varphi_{n,\gamma-1} - \varphi_{n,\gamma})(X(s)) \right.$$

$$\left. + \frac{\gamma}{2} \theta (2^{-n} \varphi_{n,\gamma-1} - \varphi_{n,\gamma})(X(s)) \right\} ds$$

is a continuous, square-integrable martingale with increasing process

$$(4.30) \quad I_{n,\gamma}(t) = \gamma^2 \int_0^t (\varphi_{n,2\gamma-1} - \varphi_{n,\gamma}^2)(X(s)) ds;$$

in fact, this holds for each $\gamma > 1$ as can be seen by approximating the function x^γ by the $C^2[0, \infty)$ function $(x + \varepsilon)^\gamma$. Defining $\varphi_\gamma: \mathcal{P}[0, 1] \rightarrow [0, \infty]$ for each $\gamma > 0$ by

$$(4.31) \quad \varphi_\gamma(\mu) = \begin{cases} \sum_{0 \leq x \leq 1} \mu(\{x\})^\gamma, & \text{if } \gamma \neq 1, \\ 1, & \text{if } \gamma = 1, \end{cases}$$

we have $\text{bp-lim}_{n \rightarrow \infty} \varphi_{n,\gamma} = \varphi_\gamma$ for each $\gamma \geq 1$, while $\varphi_{n,\gamma} \nearrow \varphi_\gamma$ as $n \nearrow \infty$ for $0 < \gamma < 1$. We conclude that, for each $\gamma > 1$,

$$(4.32) \quad Z_\gamma(t) \equiv \varphi_\gamma(X(t)) - \varphi_\gamma(X(0))$$

$$- \int_0^t \left\{ \binom{\gamma}{2} \varphi_{\gamma-1} - \left[\binom{\gamma}{2} + \frac{\gamma}{2} \theta \right] \varphi_\gamma \right\} (X(s)) \, ds$$

is a continuous, square-integrable martingale with increasing process

$$(4.33) \quad I_\gamma(t) = \gamma^2 \int_0^t (\varphi_{2\gamma-1} - \varphi_\gamma^2)(X(s)) \, ds;$$

here we are using $\lim_{n \rightarrow \infty} E[Z_{n,\gamma}(t)^2] = \lim_{n \rightarrow \infty} E[I_{n,\gamma}(t)] = E[I_\gamma(t)]$ and the monotone convergence theorem to show that, when $1 < \gamma < 2$,

$$(4.34) \quad E \left[\left(\int_0^t \varphi_{\gamma-1}(X(s)) \, ds \right)^2 \right] < \infty, \quad t \geq 0.$$

Letting $\varphi_{1+}(\mu) = \text{bp-lim}_{\gamma \rightarrow 1+} \varphi_\gamma(\mu) = \sum_{0 \leq x \leq 1} \mu(\{x\})$, we have

$$(4.35) \quad 0 = \lim_{\gamma \rightarrow 2+} E[Z_\gamma(t) - Z_2(t)] = E \left[\int_0^t (1 - \varphi_{1+})(X(s)) \, ds \right]$$

for all $t \geq 0$, so $P\{X(t) \in \mathcal{P}_a[0, 1] \text{ for almost every } t > 0\} = 1$. To remove the word "almost," observe that

$$\begin{aligned} (4.36) \quad \lim_{\gamma \rightarrow 1+} E[Z_\gamma(t)^2] &= \lim_{\gamma \rightarrow 1+} E[I_\gamma(t)] \\ &= E \left[\int_0^t \varphi_{1+}(1 - \varphi_{1+})(X(s)) \, ds \right] = 0 \end{aligned}$$

for each $t \geq 0$ by (4.33) and (4.35). Fix $t_0 > 0$. By Doob's martingale inequality, $\sup_{0 \leq t \leq t_0} |Z_\gamma(t)| \rightarrow 0$ in probability as $\gamma \rightarrow 1+$, so there exists a sequence $\gamma_n \rightarrow 1+$ such that $\sup_{0 \leq t \leq t_0} |Z_{\gamma_n}(t)| \rightarrow 0$ a.s. Letting

$$(4.37) \quad \eta(t) = \overline{\lim}_{n \rightarrow \infty} \binom{\gamma_n}{2} \int_0^t \varphi_{\gamma_n-1}(X(s)) \, ds,$$

we obtain from (4.32) and (4.35) that, almost surely,

$$(4.38) \quad \varphi_{1+}(X(t)) - \varphi_{1+}(X(0)) - \eta(t) + \frac{1}{2}\theta t = 0, \quad 0 \leq t \leq t_0.$$

Since $\eta(t)$ is nondecreasing in t , we conclude that $P\{X(t) \in \mathcal{P}_a[0, 1] \text{ for all } t > 0\} = 1$, as required. \square

4.6 Theorem The measure-valued diffusions X_n , $n = 2, 3, \dots$, and X defined as in Example 4.4, have unique stationary distributions $\tilde{\mu}_n$, $n = 2, 3, \dots$, and $\tilde{\mu}$, respectively. In fact, $\tilde{\mu}_n$ is the distribution of $\sum_{i=1}^n \xi_i^n \delta_{i/n}$, where $(\xi_1^n, \dots, \xi_n^n)$ has a symmetric Dirichlet distribution with parameter $\theta/(n-1)$ (defined below).

Moreover, there exist random variables $\xi_1 \geq \xi_2 \geq \dots \geq 0$ with $\sum_{i=1}^{\infty} \xi_i = 1$ such that

$$(4.39) \quad (\xi_{(1)}^n, \dots, \xi_{(k)}^n) \Rightarrow (\xi_1, \dots, \xi_k)$$

as $n \rightarrow \infty$ for each $k \geq 1$, where $\xi_{(1)}^n, \dots, \xi_{(n)}^n$ denote the descending order statistics of ξ_1^n, \dots, ξ_n^n . Finally, $\tilde{\mu}$ is the distribution of $\sum_{i=1}^{\infty} \xi_i \delta_{u_i}$, where u_1, u_2, \dots is a sequence of independent, uniformly distributed random variables on $[0, 1]$, independent of ξ_1, ξ_2, \dots .

Proof. Fix $n \geq 2$. Let $E_n = \{1/n, 2/n, \dots, 1\}$ and define $f_1, \dots, f_n \in C(E_n)$ by $f_j(j/n) = \delta_{ij}$. Let $(\xi_1^n, \dots, \xi_n^n)$ have a symmetric Dirichlet distribution with parameter $\varepsilon_n \equiv \theta/(n-1)$, that is, $(\xi_1^n, \dots, \xi_{n-1}^n)$ is a K -valued random variable (recall (1.16) with $r = n$) with Lebesgue density $\Gamma(n\varepsilon_n)\Gamma(\varepsilon_n)^{-n}(p_1 \cdots p_n)^{\varepsilon_n-1}$, and $\xi_n^n = 1 - \sum_{i=1}^{n-1} \xi_i^n$. Let $\tilde{v}_n \in \mathcal{P}(\mathcal{P}(E_n))$ be the distribution of $\sum_{i=1}^n \xi_i^n \delta_{i/n}$. To show that \tilde{v}_n is a stationary distribution for X_n , it suffices by Theorem 9.17 of Chapter 4 to show that

$$(4.40) \quad \int_{\mathcal{P}(E_n)} G_n \left(\prod_{i=1}^n \langle f_i, \mu \rangle^{m_i} \right) \tilde{v}_n(d\mu) = 0$$

for all integers $m_1, \dots, m_n \geq 0$, where G_n is as in (4.4) with B_n given by (4.19). (Actually, this can be proved without the aid of Theorem 9.17 of Chapter 4 by checking that the span of the functions within parentheses in (4.40) forms a core.)

But with $|m| = m_1 + \dots + m_n$, the left side of (4.40) becomes

$$(4.41)$$

$$\begin{aligned} & \int_{\mathcal{P}(E_n)} \left\{ \frac{1}{2} \sum_{i,j=1}^n (\langle f_i f_j, \mu \rangle - \langle f_i, \mu \rangle \langle f_j, \mu \rangle) m_i (m_j - \delta_{ij}) \prod_{l=1}^n \langle f_l, \mu \rangle^{m_l - \delta_{il}} \right. \\ & \quad \left. + \frac{1}{2} \theta \sum_{i=1}^n \left(\frac{1}{n-1} - \frac{n}{n-1} \langle f_i, \mu \rangle \right) m_i \prod_{l=1}^n \langle f_l, \mu \rangle^{m_l - \delta_{il}} \right\} \tilde{v}_n(d\mu) \\ &= E \left[\frac{1}{2} \sum_{i,j=1}^n \xi_i^n (\delta_{ij} - \xi_j^n) m_i (m_j - \delta_{ij}) \prod_{l=1}^n (\xi_l^n)^{m_l - \delta_{il}} \right. \\ & \quad \left. + \frac{1}{2} \theta \sum_{i=1}^n \left(\frac{1}{n-1} - \frac{n}{n-1} \xi_i^n \right) m_i \prod_{l=1}^n (\xi_l^n)^{m_l - \delta_{il}} \right] \\ &= E \left[\frac{1}{2} \sum_{i=1}^n m_i (m_i - 1 + n\varepsilon_n) \prod_{l=1}^n (\xi_l^n)^{m_l - \delta_{il}} \right. \\ & \quad \left. - \frac{1}{2} |m|(|m| - 1 + n\varepsilon_n) \prod_{l=1}^n (\xi_l^n)^{m_l} \right], \end{aligned}$$

and this is zero because

$$(4.42) \quad E[(\xi_1^n)^{m_1} \cdots (\xi_n^n)^{m_n}] = \frac{\Gamma(m_1 + \varepsilon_n) \cdots \Gamma(m_n + \varepsilon_n)}{\Gamma(|m| + n\varepsilon_n)}$$

and $\Gamma(u) = (u-1)\Gamma(u-1)$ if $u > 1$. As for uniqueness, suppose $\tilde{\mu}_n$ is a stationary distribution for X_n , and note that the left side of (4.41) with \tilde{v}_n replaced by $\tilde{\mu}_n$ is zero. Inducting on the degree of $\prod_{i=1}^n \langle f_i, \mu \rangle^{m_i}$ (namely, $|m|$), we find that $\int \prod_{i=1}^n \langle f_i, \mu \rangle^{m_i} \tilde{\mu}_n(d\mu)$ is uniquely determined for all $m_1, \dots, m_n \geq 0$. Hence $\tilde{\mu}_n$ is uniquely determined.

By Theorem 9.3 of Chapter 4, X has a stationary distribution $\tilde{\mu}$. Noting that

$$(4.43) \quad \int G\left(\prod_{i=1}^k \langle f_i, \mu \rangle\right) \tilde{\mu}(d\mu) = 0$$

for all $k \geq 1$ and $f_1, \dots, f_k \in C[0, 1]$, we obtain the uniqueness of $\tilde{\mu}$ as above, except here we induct on k . It follows from Theorem 9.12 of Chapter 4 that $\tilde{\mu}_n \Rightarrow \tilde{\mu}$ on $\mathcal{P}[0, 1]$.

Theorem 4.5 immediately implies that $\tilde{\mu}(\mathcal{P}_n[0, 1]) = 1$. We leave it to the reader to check that therefore there exist random variables $\xi_1 \geq \xi_2 \geq \cdots \geq 0$ with $\sum_{i=1}^\infty \xi_i = 1$ and u_1, u_2, \dots with distinct values in $[0, 1]$ such that $\tilde{\mu}$ is the distribution of $\sum_{i=1}^\infty \xi_i \delta_{u_i}$. The assertion that (4.39) holds says simply that the joint $\tilde{\mu}_n$ -distribution of the sizes of the k largest atoms converges to the joint $\tilde{\mu}$ -distribution. Unfortunately, this cannot be deduced merely from the fact that $\tilde{\mu}_n \Rightarrow \tilde{\mu}$ on $\mathcal{P}[0, 1]$. However, by giving a stronger topology to $\mathcal{P}[0, 1]$, we can obtain the desired conclusion.

We define the metric ρ^* on $\mathcal{P}[0, 1]$ as follows: given $\mu, v \in \mathcal{P}[0, 1]$, let F_μ, F_v be the corresponding (right-continuous) cumulative distribution functions, and put

$$(4.44) \quad \rho^*(\mu, v) = d_{D[0, 1]}(F_\mu, F_v),$$

where $d_{D[0, 1]}$ denotes a metric that induces the Skorohod topology on $D[0, 1]$ (see Billingsley (1968)). The separability of $(\mathcal{P}[0, 1], \rho^*)$ follows as does the separability of $D[0, 1]$. We note that $\sum_{i=1}^n \xi_i^n \delta_{u_i}$, $n = 2, 3, \dots$, and $\sum_{i=1}^\infty \xi_i \delta_{u_i}$ are $(\mathcal{P}[0, 1], \rho^*)$ -valued random variables, so we regard their distributions $\tilde{\mu}_n$, $n = 2, 3, \dots$, and $\tilde{\mu}$ as belonging to $\mathcal{P}(\mathcal{P}[0, 1], \rho^*)$. We claim that

$$(4.45) \quad \tilde{\mu}_n \Rightarrow \tilde{\mu} \text{ on } (\mathcal{P}[0, 1], \rho^*).$$

Letting

$$(4.46) \quad F_n(t) = \sum_{i/n \leq t} \xi_i^n, \quad F(t) = \sum_{u_i \leq t} \xi_i, \quad 0 \leq t \leq 1,$$

we see from the definition of ρ^* that it suffices to show that $F_n \Rightarrow F$ in $D[0, 1]$. We verify this using Theorem 15.6 of Billingsley (1968), which is the analogue for $D[0, 1]$ of Theorem 8.8 of Chapter 3.

Let $C = \bigcup_{i=1}^{\infty} \{t \in [0, 1] : P\{u_i = t\} > 0\}$. Then C is at most countable, and for $t_1, \dots, t_k \in [0, 1] - C$, the function $\mu \rightarrow (\mu([0, t_1]), \dots, \mu([0, t_k]))$ is $\tilde{\mu}$ -a.s. continuous on $(\mathcal{P}[0, 1], \rho)$ (ρ being the Prohorov metric), so

$$(4.47) \quad (F_n(t_1), \dots, F_n(t_k)) \Rightarrow (F(t_1), \dots, F(t_k))$$

by Corollary 1.9 of Chapter 3. In particular, if $t \in [0, 1] - C$, then

$$\begin{aligned} (4.48) \quad E[F(t)] &= \lim_{n \rightarrow \infty} E[F_n(t)] \\ &= \lim_{n \rightarrow \infty} E[\xi_1^n + \dots + \xi_{[nt]}^n] \\ &= \lim_{n \rightarrow \infty} \frac{[nt]}{n} = t, \end{aligned}$$

so $E[F(t) - F(t-)] = 0$, and hence $C = \emptyset$. It follows that (4.47) holds for all $t_1, \dots, t_k \in [0, 1]$. Finally, let $0 \leq t_1 \leq t \leq t_2 \leq 1$. Then for $n = 1, 2, \dots$,

$$\begin{aligned} (4.49) \quad E[(F_n(t) - F_n(t_1))^2(F_n(t_2) - F_n(t))^2] \\ &= E[(\xi_{[nt_1]+1}^n + \dots + \xi_{[nt]}^n)^2(\xi_{[nt_1]+1}^n + \dots + \xi_{[nt_2]}^n)^2] \\ &= E[(Z_1^n)^2(Z_2^n)^2], \end{aligned}$$

where (Z_1^n, Z_2^n) is a K -valued random variable (recall (1.16) with $r = 3$) with Lebesgue density $\Gamma(\alpha_n + \beta_n + \gamma_n)\{\Gamma(\alpha_n)\Gamma(\beta_n)\Gamma(\gamma_n)\}^{-1}p_1^{\alpha_n-1}p_2^{\beta_n-1}p_3^{\gamma_n-1}$ and $(\alpha_n, \beta_n, \gamma_n) = (([nt] - [nt_1])\varepsilon_n, ([nt_2] - [nt])\varepsilon_n, (n - [nt_2] + [nt_1])\varepsilon_n)$. Hence (4.49) becomes

$$\begin{aligned} (4.50) \quad \frac{\alpha_n(\alpha_n + 1)\beta_n(\beta_n + 1)}{(\alpha_n + \beta_n + \gamma_n) \cdots (\alpha_n + \beta_n + \gamma_n + 3)} &\leq \left(\frac{[nt] - [nt_1]}{n} \right) \left(\frac{[nt_2] - [nt]}{n} \right) \\ &\leq (t_2 - t_1)^2. \end{aligned}$$

We conclude that $F_n \Rightarrow F$ in $D[0, 1]$, and hence (4.45) holds.

Let us say that $x \in D[0, 1]$ has a jump of size $\delta > 0$ at a location $t \in [0, 1]$ if $|x(t) - x(t-)| = \delta$. For each $x \in D[0, 1]$ and $i \geq 1$, we define $s_i(x)$ and $l_i(x)$ to be the size and location of the i th largest jump of x . If $s_i(x) = s_{i+1}(x)$, we adopt the convention that $l_i(x) < l_{i+1}(x)$ (that is, ties are labeled from left to right). If x has only k jumps, we define $s_i(x) = l_i(x) = 0$ for each $i > k$. We leave it to the reader to check that s_1, s_2, \dots and l_1, l_2, \dots are Borel measurable on $D[0, 1]$. Suppose $\{x_n\} \subset D[0, 1]$, $x \in D[0, 1]$, and $d_{D[0, 1]}(x_n, x) \rightarrow 0$. Then

$$(4.51) \quad (s_1(x_n), \dots, s_k(x_n)) \rightarrow (s_1(x), \dots, s_k(x))$$

and, if $s_1(x) > s_2(x) > \dots$, then

$$(4.52) \quad (s_1(x_n), \dots, s_k(x_n), l_1(x_n), \dots, l_k(x_n)) \rightarrow (s_1(x), \dots, s_k(x), l_1(x), \dots, l_k(x)).$$

It follows from the definition of ρ^* that

$$(4.53) \quad (s_1(F_\mu), \dots, s_k(F_\mu))$$

is a ρ^* -continuous function of $\mu \in \mathcal{P}[0, 1]$, where F_μ is as in (4.44). Now since the $\tilde{\mu}_n$ -distribution of (4.53) is the distribution of $(\xi_{(1)}^n, \dots, \xi_{(k)}^n)$ for each $n \geq k$, and since the $\tilde{\mu}$ -distribution of (4.53) is the distribution of (ξ_1, \dots, ξ_k) , we obtain (4.39) from (4.45).

We leave it as a problem to show that

$$(4.54) \quad P\{\xi_1 > \xi_2 > \dots\} = 1$$

(Problem 12). It follows that

$$(4.55) \quad (s_1(F_\mu), \dots, s_k(F_\mu), l_1(F_\mu), \dots, l_k(F_\mu))$$

is a $\tilde{\mu}$ -a.s. ρ^* -continuous function of $\mu \in \mathcal{P}[0, 1]$. Now the $\tilde{\mu}_n$ -distribution of (4.55) is the distribution of

$$(4.56) \quad (\xi_{(1)}^n, \dots, \xi_{(k)}^n, u_1^n, \dots, u_k^n)$$

for each $n \geq k$, where (nu_1^n, \dots, nu_k^n) is independent of $(\xi_1^n, \dots, \xi_n^n)$ and takes on each of the permutations of $(1, 2, \dots, n)$ with probability $1/n!$. By Corollary 1.9 of Chapter 3, (4.56) converges in distribution as $n \rightarrow \infty$ to the $\tilde{\mu}$ -distribution of (4.55), that is, to the distribution of $(\xi_1, \dots, \xi_k, u_1, \dots, u_k)$. This allows us to conclude that u_1, u_2, \dots is a sequence of independent, uniformly distributed random variables on $[0, 1]$, independent of ξ_1, ξ_2, \dots . \square

We close this section with a derivation of Ewens' sampling formula. Given a positive integer r , a vector $\beta = (\beta_1, \dots, \beta_r)$ belonging to the finite set

$$(4.57) \quad \Gamma_r = \left\{ \alpha \in (\mathbb{Z}_+)^r : \sum_{j=1}^r j\alpha_j = r \right\},$$

and $\tilde{v} \in \mathcal{P}(\mathcal{P}_a[0, 1])$, let $P(\beta, \tilde{v})$ denote the probability that in a random sample of size r from a population whose "type" frequencies are random and distributed according to \tilde{v} , β_j "types" are represented j times ($j = 1, \dots, r$).

4.7 Theorem Let $\tilde{\mu}$ be as in Theorem 4.6, let $r \geq 1$, and let $\beta \in \Gamma_r$. Then

$$(4.58) \quad P(\beta, \tilde{\mu}) = \prod_{j=1}^r \frac{j}{j + \theta - 1} \frac{\theta^{\beta_j}}{j^{\beta_j} \beta_j!}.$$

Proof. Observe that for each $\tilde{v} \in \mathcal{P}(\mathcal{P}_a[0, 1])$,

$$(4.59) \quad P(\beta, \tilde{v}) = \int_{\mathcal{P}_a[0, 1]} P(\beta, \delta_\mu) \tilde{v}(d\mu)$$

and

$$(4.60) \quad P(\beta, \delta_\mu) = \sum \frac{r!}{m_1! m_2! \dots} s_1(F_\mu)^{m_1} s_2(F_\mu)^{m_2} \dots,$$

where the sum ranges over all sequences (m_1, m_2, \dots) of nonnegative integers for which $\sum_{i=1}^{\infty} m_i = r$ and β_j is the cardinality of $\{i \geq 1 : m_i = j\}$ ($j = 1, \dots, r$). Denote (4.60) by $\varphi_{\beta}(\mu)$. Then φ_{β} is lower semicontinuous with respect to ρ^* and

$$(4.61) \quad \sum_{\alpha \in \Gamma_r} \varphi_{\alpha}(\mu) = (s_1(F_{\mu}) + s_2(F_{\mu}) + \cdots)^r = 1,$$

implying that $\varphi_{\beta} = 1 - \sum_{\alpha \in \Gamma_r - \{\beta\}} \varphi_{\alpha}$ is also upper semicontinuous with respect to ρ^* , hence ρ^* -continuous. We conclude that

$$(4.62) \quad \lim_{n \rightarrow \infty} \int \varphi_{\beta} d\tilde{\mu}_n = \int \varphi_{\beta} d\tilde{\mu}.$$

The proof is completed by showing that the left side of (4.62) equals the right side of (4.58). That is,

$$\begin{aligned} (4.63) \quad & \int \sum \frac{r!}{m_1! m_2! \cdots} s_1(F_{\mu})^{m_1} s_2(F_{\mu})^{m_2} \cdots d\tilde{\mu}_n \\ &= \frac{r!}{(1!)^{\beta_1} (2!)^{\beta_2} \cdots (r!)^{\beta_r}} \sum E[(\zeta_{(1)}^n)^{m_1} \cdots (\zeta_{(n)}^n)^{m_n}] \\ &= \prod_{j=1}^r \frac{j}{(j!)^{\beta_j}} \sum E[(\zeta_1^n)^{m_1} \cdots (\zeta_n^n)^{m_n}] \\ &= \prod_{j=1}^r \frac{j}{(j!)^{\beta_j}} \sum \frac{\Gamma(m_1 + \varepsilon_n) \cdots \Gamma(m_n + \varepsilon_n)}{\Gamma(r + n\varepsilon_n)} \frac{\Gamma(n\varepsilon_n)}{\Gamma(\varepsilon_n)^n} \\ &= \left(\prod_{j=1}^r \frac{j}{(j!)^{\beta_j}} \right) \frac{n! \Gamma(\varepsilon_n)^{n-\Sigma \beta_j} \Gamma(1 + \varepsilon_n)^{\beta_1} \cdots \Gamma(r + \varepsilon_n)^{\beta_r}}{\beta_1! \cdots \beta_r! (n - \Sigma \beta_j)! \Gamma(r + n\varepsilon_n)} \frac{\Gamma(n\varepsilon_n)}{\Gamma(\varepsilon_n)^n} \\ &\rightarrow \prod_{j=1}^r \frac{j}{j + \theta - 1} \frac{\theta^{\theta_j}}{j^{\theta_j} \beta_j!}, \end{aligned}$$

where the sums are as in (4.60); here we are using $\Gamma(\varepsilon_n) = \Gamma(1 + \varepsilon_n)/\varepsilon_n$. \square

5. PROBLEMS

1. Show that (1.23), (1.26), and (1.27) imply (1.22).
2. Let X be the diffusion process in K of Section 1 in the special case in which $\mu_{ij} = \gamma_j > 0$ for $i, j = 1, \dots, r$. Show that the measure $\mu \in \mathcal{P}(K)$, defined by

$$(5.1) \quad \mu(dp) = \beta p_1^{2\gamma_1-1} \cdots p_r^{2\gamma_r-1} \exp \left(\sum_{i,j=1}^r \sigma_{ij} p_i p_j \right) dp_1 \cdots dp_{r-1}$$

for some constant $\beta > 0$, is a stationary distribution for X . (This generalizes parts of Theorems 2.1 and 4.6.)

3. Let X be as in Theorem 2.4. Show that ζ^ϵ , defined by (2.56), is continuous a.s. with respect to the distribution of X for $0 < \epsilon < \frac{1}{2}$.
4. Let X be the diffusion process in K of Section 1 in the special case in which $\mu_{ii} = \gamma > 0$ for $i = 1, \dots, r - 1$, $\mu_{ij} = 0$ otherwise, and $\sigma_{ij} = 0$ for $i, j = 1, \dots, r$. Let $\tau = \inf\{t \geq 0: \min_{1 \leq i \leq r-1} X_i(t) = 0\}$. If $p \in K$ and $P\{X(0) = p\} = 1$, show that, for $i = 1, \dots, r - 1$,

$$(5.2) \quad P\{X_i(\tau) = 0\} = 1 - p_i \left\{ \sum_{\substack{1 \leq j \leq r-1 \\ j \neq i}} (p_i + p_j)^{-1} - \sum_{\substack{1 \leq j < k \leq r-1 \\ j, k \neq i}} (p_i + p_j + p_k)^{-1} + \cdots + (-1)^{r-1} (1 - p_r)^{-1} \right\}.$$

5. Put $I = \{1, \dots, r\}$, and let X be the diffusion process in $E = [0, 1]^I$ with generator $A \equiv \{(f, Gf): f \in C^2(E)\}$, where

$$(5.3) \quad G = \frac{1}{2} \sum_{i \in I} \kappa_i^{-1} p_i (1 - p_i) \partial_i^2 + \sum_{i \in I} \left\{ \sigma p_i (1 - p_i) + \sum_{j \in I} \mu_{ij} p_j \right\} \partial_i,$$

$(\mu_{ij})_{i, j \in I}$ is the infinitesimal matrix of an irreducible jump Markov process in I with stationary distribution $(\kappa_i)_{i \in I}$, and σ is real. (Specifically, $\mu_{ij} \geq 0$ for $i \neq j$, there does not exist a nonempty proper subset J of I with $\mu_{ij} = 0$ for all $i \in J$ and $j \notin J$, $\kappa_i > 0$ for each $i \in I$, $\sum_{j \in I} \kappa_j = 1$, and $\sum_{i \in I} \kappa_i \mu_{ij} = 0$ for each $j \in I$.)

- (a) Formulate a geographically structured Wright–Fisher model of which X is the diffusion limit, proving an appropriate limit theorem (cf. Problem 7).
- (b) Let $\tau = \inf\{t \geq 0: X(t) = (0, \dots, 0) \text{ or } X(t) = (1, \dots, 1)\}$. Show that $E[\tau] < \infty$, regardless of what the initial distribution of X may be.
- (c) If $p \in E$ and $P\{X(0) = p\} = 1$, show that

$$(5.4) \quad P\{X(\tau) = (1, \dots, 1)\} = \frac{1 - \exp\{-2\sigma \sum_{i \in I} \kappa_i p_i\}}{1 - \exp\{-2\sigma\}}.$$

6. Construct a sequence of diffusion processes X^N in $[0, 1]$, and a diffusion X in $[0, 1]$ with the following properties: 0 and 1 are absorbing boundaries for X^N and X , and $X^N \Rightarrow X$ in $D_{[0, 1]}[0, \infty)$; but, defining ζ by (2.20) with $F = \{0, 1\}$ and τ_N and τ by (2.21), τ_N fails to converge in distribution to τ .
7. Apply Theorem 3.5 to Nagylaki's (1980) geographically structured Wright–Fisher model. In fact, this is already done in the given reference,

so the problem consists merely of verifying the analysis appearing there and checking the technical conditions.

8. Apply Theorem 3.5 to the dioecious multinomial-sampling model described by Watterson (1964).

Hint: Rather than numbering the genotypes arbitrarily, it simplifies matters to let $P_{ij}^{(s)}$ be the frequency of $A_i A_j$ ($i \leq j$) in sex s ($s = 1, 2$). This suggests, incidentally, that the general case of r alleles is no more difficult than the special case $r = 2$. Finally, we remark that the assumption that mutation rates and selection intensities are equal in the two sexes is unnecessarily restrictive.

9. Apply Theorem 3.5 to the dioecious overlapping-generation model described by Watterson (1964).

10. Karlin and Levikson (1974) state some results concerning diffusion limits of genetic models with random selection intensities. Prove these results.

Hint: As a first step, one must specify the discrete models precisely.

11. Let X_1 (respectively, X) be the Ohta–Kimura model (respectively, the Fleming–Viot model) of Example 4.3, regarded as taking values in $\mathcal{P}(\mathbb{Z})$ (respectively, $\mathcal{P}(\mathbb{R})$). Show that X_1 (respectively, X) has no stationary distribution.

12. Let $\xi_1 \geq \xi_2 \geq \dots$ be as in Theorem 4.6. Show that $P\{\xi_1 > \xi_2 > \dots\} = 1$.

13. Let $\xi_1 \geq \xi_2 \geq \dots$ be as in Theorem 4.6. Consider an inhomogeneous Poisson process on $(0, \infty)$ with rate function $\rho(x) = \theta x^{-1} e^{-x}$. In particular, the number of points in the interval (a, b) is Poisson distributed with parameter $\int_a^b \rho(x) dx$. Because $\int_a^\infty \rho(x) dx < \infty$ ($= \infty$) if $a > 0$ ($= 0$), the points of the Poisson process can be labeled as $\eta_1 > \eta_2 > \dots$. Moreover, $\sum_{i=1}^\infty \eta_i$ has expectation $\int_0^\infty x \rho(x) dx = \theta$, and is therefore finite a.s. Show that (ξ_1, ξ_2, \dots) has the same distribution as

$$(5.5) \quad \left(\frac{\eta_1}{\sum_{i=1}^\infty \eta_i}, \frac{\eta_2}{\sum_{i=1}^\infty \eta_i}, \dots \right).$$

14. Let X be the stationary infinitely-many-neutral-alleles model with uniform mutation (see Theorem 4.6), with its time-parameter set extended to $(-\infty, \infty)$.
- Show that $\{X(t), -\infty < t < \infty\}$ and $\{X(-t), -\infty < t < \infty\}$ induce the same distribution on $C_{\mathcal{P}(0, 1)}(-\infty, \infty)$. (Because of this, X is said to be *reversible*.)
 - Using (a), show that the probability that the most frequent allele (or “type”) at time 0, say, is oldest equals the probability that the most frequent allele at time 0 will survive the longest.

- (c) Show that the second probability in (b) is just $E[\xi_1]$, where ξ_1 is as in Theorem 4.6. (See Watterson and Guess (1977) for an evaluation of this expectation.)

6. NOTES

The best general reference on mathematical population genetics is Ewens (1979). Also useful is Kingman (1980).

The genetic model described in Section 1 is a variation of a model of Moran (1958c) due to Ethier and Nagylaki (1980). The Wright–Fisher model was formulated implicitly by Fisher (1922) and explicitly by Wright (1931). Various versions of Theorem 1.1 have been obtained by various authors. Trotter (1958) treated the neutral diallelic case, Norman (1972) the general diallelic case, Littler (1972) the neutral multi-allelic case, and Sato (1976) the general multi-allelic case.

The proof of Lemma 2.1 follows Norman (1975b). Theorem 2.4 is essentially from Ethier (1979), but a special case had earlier been obtained by Guess (1973). Corollary 2.7 is due to Norman (1972).

Section 3 comes from Ethier and Nagylaki (1980). Earlier work on diffusion approximations of non-Markovian models includes that of Watterson (1962) and Norman (1975a). Example 3.8, as noted above, is similar to a model of Moran (1958c). Example 3.9 is essentially due to Moran (1958a, b).

Theorem 4.1 is due to Kurtz (1981a). The characterization of X had earlier been obtained in certain cases by Fleming and Viot (1979). The processes of Example 4.3 are those of Ohta and Kimura (1973) and Fleming and Viot (1979). Example 4.4 was motivated by Watterson (1976), but the model goes back to Kimura and Crow (1964). Theorem 4.5 is analogous to a result of Ethier and Kurtz (1981). The main conclusion of Theorem 4.6, namely (4.39), is due to Kingman (1975). Finally, Theorem 4.7 is Ewens' (1972) sampling formula; our proof is based on Watterson (1976) and Kingman (1977).

Problem 2 is essentially Wright's (1949) formula. The reader is referred to Shiga (1981) for uniqueness. Problem 4 comes from Littler and Good (1978). See Nagylaki (1982) for Problem 5(a)(c). Problem 11 (for X_1) is due to Shiga (1982), who obtains much more general results. Problem 13 is a theorem of Kingman (1975), while Problem 14 is adapted from Watterson and Guess (1977).

11

DENSITY DEPENDENT POPULATION PROCESSES

By a population process we mean a stochastic model for a system involving a number of similar particles. We use the term "particles" broadly to include molecules in a chemical reaction model and infected individuals in an epidemic model. The branching and genetic models of the previous two chapters are examples of what we have in mind.

In this chapter we consider certain one-parameter families of processes that arise in a variety of applications. Section 1 gives examples that motivate the general formulation. Section 2 gives the basic law of large numbers and central limit theorem and Section 3 examines the corresponding diffusion approximation. Asymptotics for hitting distributions are considered in Section 4.

1. EXAMPLES

We are interested in certain families of jump Markov processes depending on a parameter that has different interpretations in different contexts, for example, total population size, area, or volume. We always denote this parameter by n . To motivate and identify the structure of these particular families, we give some examples:

A. Logistic Growth

In this context we interpret n as the area of a region occupied by a certain population. If the population size is k , then the population density is k/n . For simplicity we assume births and deaths occur singly. The intensities for births and deaths should be approximately proportional to the population size. We assume, however, that crowding affects the birth and death rates, which therefore depend on the population density. Hence the intensities can be written

$$(1.1) \quad \begin{aligned} q_{k,k+1}^{(n)} &= \lambda \left(\frac{k}{n} \right) k = n\lambda \left(\frac{k}{n} \right) \frac{k}{n} \\ q_{k,k-1}^{(n)} &= \mu \left(\frac{k}{n} \right) k = n\mu \left(\frac{k}{n} \right) \frac{k}{n}. \end{aligned}$$

If we take $\lambda(x) \equiv a$ and $\mu(x) = b + cx$, we have a stochastic model analogous to the deterministic logistic model given by

$$(1.2) \quad \dot{X} = (a - b)X - cX^2. \quad \square$$

B. Epidemics

Here we interpret n as the total population size, which remains constant. In the population at any given time there are a number of individuals i that are susceptible to a particular disease and a number of individuals j who have the disease and can pass it on. A susceptible individual encounters diseased individuals at a rate proportional to the fraction of the total population that is diseased. Consequently, the intensity for a new infection is

$$(1.3) \quad q_{(i,j)(i-1,j+1)} = \lambda i \frac{j}{n} = n\lambda \frac{i}{n} \frac{j}{n}.$$

We assume diseased individuals recover and become immune independently of each other, which leads to the assumption

$$(1.4) \quad q_{(i,j)(i,j-1)} = \mu j = n\mu \frac{j}{n}.$$

The analogous deterministic model in this case is

$$(1.5) \quad \begin{aligned} \dot{X}_1 &= -\lambda X_1 X_2, \\ \dot{X}_2 &= \lambda X_1 X_2 - \mu X_2. \end{aligned} \quad \square$$

C. Chemical Reactions

We now interpret n as the volume of a chemical system containing d chemical reactants R_1, R_2, \dots, R_d undergoing r chemical reactions



$$j = 1, 2, \dots, r,$$

that is, for example, when the j th reaction occurs in the forward direction, b_{1j} molecules of reactant R_1 , b_{2j} molecules of reactant R_2 , and so on, react to form c_{1j} molecules of R_1 , c_{2j} molecules of R_2 , and so on. Let $b_j = (b_{1j}, b_{2j}, \dots, b_{dj})$, $c_j = (c_{1j}, c_{2j}, \dots, c_{dj})$, and define

$$(1.7) \quad |b_j| = b_{1j} + b_{2j} + b_{3j} + \cdots + b_{dj}, \quad x^{b_j} = \prod_{i=1}^d x_i^{b_{ij}}.$$

The stochastic analogue of the "law of mass action" suggests that the intensity for the occurrence of the forward reaction should be

$$(1.8) \quad n^{-|b_j|+1} \lambda_j \prod_{i=1}^d \binom{k_i}{b_{ij}} = n \left[\frac{\lambda_j}{\prod_{i=1}^d b_{ij}!} \left(\frac{k}{n} \right)^{b_j} + O\left(\frac{1}{n}\right) \right],$$

where $k = (k_1, k_2, \dots, k_d)$ are the numbers of molecules of the reactants. The intensity for the reverse reaction is

$$(1.9) \quad n^{-|c_j|+1} \mu_j \prod_{i=1}^d \binom{k_i}{c_{ij}} = n \left[\frac{\mu_j}{\prod_{i=1}^d c_{ij}!} \left(\frac{k}{n} \right)^{c_j} + O\left(\frac{1}{n}\right) \right].$$

If we take as the state the numbers of molecules of the reactants, then the transition intensities become

$$(1.10) \quad q_{k, k+i}^{(n)} = \sum_{c_j - b_j = i} n^{-|b_j|+1} \lambda_j \prod_{i=1}^d \binom{k_i}{b_{ij}} + \sum_{b_j - c_j = i} n^{-|c_j|+1} \mu_j \prod_{i=1}^d \binom{k_i}{c_{ij}} \\ = n \left[\sum_{c_j - b_j = i} \frac{\lambda_j}{\prod_{i=1}^d b_{ij}!} \left(\frac{k}{n} \right)^{b_j} + \sum_{b_j - c_j = i} \frac{\mu_j}{\prod_{i=1}^d c_{ij}!} \left(\frac{k}{n} \right)^{c_j} + O\left(\frac{1}{n}\right) \right]$$

and the analogous deterministic model is

$$(1.11) \quad \dot{X}_i = \sum_{j=1}^r ((c_{ij} - b_{ij}) \tilde{\lambda}_j X^{b_j} + (b_{ij} - c_{ij}) \tilde{\mu}_j X^{c_j})$$

where

$$\tilde{\lambda}_j = \frac{\lambda_j}{\prod_{i=1}^d b_{ij}!} \quad \text{and} \quad \tilde{\mu}_j = \frac{\mu_j}{\prod_{i=1}^d c_{ij}!}. \quad \square$$

In the first two examples the transition intensities are of the form

$$(1.12) \quad q_{k,k+1}^{(n)} = n\beta_l\left(\frac{k}{n}\right).$$

In the last example, (1.12) is the correct form if c_{ij} and b_{ij} only assume the values 0 and 1, while in general we have

$$(1.13) \quad q_{k,k+1}^{(n)} = n\left[\beta_l\left(\frac{k}{n}\right) + O\left(\frac{1}{n}\right)\right].$$

We consider families with transition intensities of the form (1.12) and observe that the results usually carry over to the more general form with little additional effort.

To be precise, we assume we are given a collection of nonnegative functions β_l , $l \in \mathbb{Z}^d$, defined on a subset $E \subset \mathbb{R}^d$. Setting

$$(1.14) \quad E_n = E \cap \{n^{-1}k : k \in \mathbb{Z}^d\},$$

we require that $x \in E_n$ and $\beta_l(x) > 0$ imply $x + n^{-1}l \in E_n$. By a *density dependent family* corresponding to the β_l we mean a sequence $\{X_n\}$ of jump Markov processes such that X_n has state space E_n and transition intensities

$$(1.15) \quad q_{x,y}^{(n)} = n\beta_{n(y-x)}(x), \quad x, y \in E_n.$$

2. LAW OF LARGE NUMBERS AND CENTRAL LIMIT THEOREM

By Theorem 4.1 of Chapter 6 we see that the Markov process \hat{X}_n with intensities $q_{k,k+1}^{(n)} = n\beta_l(k/n)$, satisfies, for t less than the first infinity of jumps,

$$(2.1) \quad \hat{X}_n(t) = \hat{X}_n(0) + \sum_l lY_l\left(n \int_0^t \beta_l\left(\frac{\hat{X}_n(s)}{n}\right) ds\right),$$

where the Y_l are independent standard Poisson processes.

Setting

$$(2.2) \quad F(x) = \sum_l l\beta_l(x)$$

and $X_n = n^{-1}\hat{X}_n$, we have

$$(2.3) \quad X_n(t) = X_n(0) + \sum_l ln^{-1} Y_l\left(n \int_0^t \beta_l(X_n(s)) ds\right) + \int_0^t F(X_n(s)) ds$$

where $\tilde{Y}_l(u) = Y_l(u) - u$ is the Poisson process centered at its expectation. The state space for X_n is E_n given by (1.14), and the form of the generator for X_n is

$$(2.4) \quad \begin{aligned} A_n f(x) &= \sum_l n\beta_l(x)(f(x + n^{-1}l) - f(x)) \\ &= \sum_l n\beta_l(x)(f(x + n^{-1}l) - f(x) - n^{-1}l \cdot \nabla f(x)) \\ &\quad + F(x) \cdot \nabla f(x), \quad x \in E_n. \end{aligned}$$

Observing that

$$(2.5) \quad \lim_{n \rightarrow \infty} \sup_{u \leq v} |n^{-1} \tilde{Y}_l(nu)| = 0 \quad \text{a.s.,} \quad v \geq 0,$$

we have the following theorem.

2.1 Theorem Suppose that for each compact $K \subset E$,

$$(2.6) \quad \sum_l l \sup_{x \in K} \beta_l(x) < \infty$$

and there exists $M_K > 0$ such that

$$(2.7) \quad |F(x) - F(y)| \leq M_K |x - y|, \quad x, y \in K.$$

Suppose X_n satisfies (2.3), $\lim_{n \rightarrow \infty} X_n(0) = x_0$, and X satisfies

$$(2.8) \quad X(t) = x_0 + \int_0^t F(X(s)) ds, \quad t \geq 0.$$

Then for every $t \geq 0$,

$$(2.9) \quad \lim_{n \rightarrow \infty} \sup_{s \leq t} |X_n(s) - X(s)| = 0 \quad \text{a.s.}$$

2.2 Remark Implicitly we are assuming global existence for $\dot{X} = F(X)$, $X(0) = x_0$. Of course (2.7) guarantees uniqueness. \square

Proof. Since for fixed $t \geq 0$ the validity of (2.9) depends only on the values of the β_l in some small neighborhood of $\{X(s): s \leq t\}$, we may as well assume that $\tilde{\beta}_l = \sup_{x \in E} \beta_l(x)$ satisfies

$$(2.10) \quad \sum_l l \tilde{\beta}_l < \infty$$

and that there exists a fixed $M > 0$ such that

$$(2.11) \quad |F(x) - F(y)| \leq M |x - y|, \quad x, y \in E.$$

Then

$$(2.12) \quad \begin{aligned} \varepsilon_n(t) &\equiv \sup_{u \leq t} \left| X_n(u) - X_n(0) - \int_0^u F(X_n(s)) \, ds \right| \\ &\leq \sum_l |l| n^{-1} \sup_{u \leq t} |\tilde{Y}_l(n\bar{\beta}_l u)| \\ &\leq \sum_l |l| n^{-1} (Y_l(n\bar{\beta}_l t) + n\bar{\beta}_l t), \end{aligned}$$

where the last inequality is term by term. Note that the process on the right is a process with independent increments, and the law of large numbers implies

$$(2.13) \quad \begin{aligned} \lim_{n \rightarrow \infty} \sum_l |l| n^{-1} (Y_l(n\bar{\beta}_l t) + n\bar{\beta}_l t) \\ &= \sum_l 2|l|\bar{\beta}_l t \\ &= \sum_l \lim_{n \rightarrow \infty} |l| n^{-1} (Y_l(n\bar{\beta}_l t) + n\bar{\beta}_l t), \end{aligned}$$

that is, we can interchange the limit and the summation. But the term by term inequality in (2.12) implies we can interchange the limit and summation for the middle expression as well, and we conclude from (2.5) that

$$(2.14) \quad \lim_{n \rightarrow \infty} \varepsilon_n(t) = 0 \quad \text{a.s.}$$

Now (2.11) implies

$$(2.15) \quad |X_n(t) - X(t)| \leq |X_n(0) - x_0| + \varepsilon_n(t) + \int_0^t M |X_n(s) - X(s)| \, ds,$$

and hence by Gronwall's inequality (Appendix 5),

$$(2.16) \quad |X_n(t) - X(t)| \leq (|X_n(0) - x_0| + \varepsilon_n(t)) e^{Mt}$$

and (2.9) follows. \square

Set $W_l^{(n)}(u) = n^{-1/2} \tilde{Y}_l(nu)$. The fact that $W_l^{(n)} \Rightarrow W_l$, standard Brownian motion, immediately suggests a central limit theorem for the deviation of X_n from X . Let $V_n(t) = \sqrt{n}(X_n(t) - X(t))$. Then

$$(2.17) \quad V_n(t) = V_n(0) + \sum_l l W_l^{(n)} \left(\int_0^t \beta_l(X_n(s)) \, ds \right) + \int_0^t \sqrt{n} (F(X_n(s)) - F(X(s))) \, ds.$$

Observing that $X_n = X + n^{-1/2}V_n$, (2.17) suggests the following limiting equation:

$$(2.18) \quad V(t) = V(0) + \sum_l l W_l \left(\int_0^t \beta_l(X(s)) ds \right) + \int_0^t \partial F(X(s)) V(s) ds \\ \equiv V(0) + U(t) + \int_0^t \partial F(X(s)) V(s) ds,$$

where $\partial F(x) = ((\partial_j F_i(x)))$.

Let Φ be the solution of the matrix equation

$$(2.19) \quad \frac{\partial}{\partial t} \Phi(t, s) = \partial F(X(t)) \Phi(t, s), \quad \Phi(s, s) = I.$$

Then

$$(2.20) \quad V(t) = \Phi(t, 0)V(0) + \int_0^t \Phi(t, s) dU(s) \\ = \Phi(t, 0)V(0) + U(t) + \int_0^t \Phi(t, s) \partial F(X(s)) U(s) ds \\ = \Phi(t, 0)(V(0) + U(t)) + \int_0^t \Phi(t, s) \partial F(X(s))(U(s) - U(t)) ds.$$

Since U is Gaussian (in fact, a time-inhomogeneous Brownian motion), V is Gaussian with mean $\Phi(t, 0)V(0)$ (we assume $V(0)$ is nonrandom) and covariance matrix

$$(2.21) \quad \text{cov}(V(t), V(r)) = \int_0^{t \wedge r} \Phi(t, s) G(X(s)) [\Phi(r, s)]^T ds,$$

where

$$(2.22) \quad G(x) = \sum_l l l^T \beta_l(x).$$

2.3 Theorem Suppose for each compact $K \subset E$,

$$(2.23) \quad \sum_l \|l\|^2 \sup_{x \in K} \beta_l(x) < \infty,$$

and that the β_l and ∂F are continuous. Suppose X_n satisfies (2.3), X satisfies (2.8), $V_n = \sqrt{n}(X_n - X)$, and $\lim_{n \rightarrow \infty} V_n(0) = V(0)$ ($V(0)$ constant). Then $V_n \Rightarrow V$ where V is the solution of (2.18).

Proof. Comparing (2.17) to (2.18), set

$$(2.24) \quad U_n(t) = \sum_l l W_l^{(n)} \left(\int_0^t \beta_l(X_n(s)) ds \right)$$

and

$$(2.25) \quad \varepsilon_n(t) = \int_0^t \sqrt{n} (F(X_n(s)) - F(X(s)) - n^{-1/2} \partial F(X(s)) V_n(s)) ds.$$

Theorem 2.1 implies $\sup_{s \leq t} |\varepsilon_n(s)| \rightarrow 0$ a.s. and $U_n \Rightarrow U$ in $D_{\mathbb{R}^d}[0, \infty)$. But, as in (2.20),

$$(2.26) \quad V_n(t) = \Phi(t, 0) V_n(0) + U_n(t) + \varepsilon_n(t) + \int_0^t \Phi(t, s) \partial F(X(s)) (U_n(s) + \varepsilon_n(s)) ds,$$

and $V_n \Rightarrow V$ by the continuous mapping theorem, Corollary 1.9 of Chapter 3. \square

3. DIFFUSION APPROXIMATIONS

The basic implication of Theorem 2.3 is that X_n can be approximated (in distribution) by $\tilde{Z}_n = X + n^{-1/2} V$. An alternative approximation is the "diffusion approximation" Z_n , whose generator is obtained heuristically by expanding f in (2.4) in a Taylor series and dropping terms beyond the second order. This gives

$$(3.1) \quad B_n f(x) = \frac{1}{2n} \sum_{i,j} G_{ij}(x) \partial_i \partial_j f(x) + \sum_i F_i(x) \partial_i f(x).$$

The statement that \tilde{Z}_n approximates X_n is, of course, justified by the central limit theorem. No similar limit theorem can justify the statement that Z_n approximates X_n , since the Z_n are not expressible in terms of any sort of limiting process. To overcome this problem we use the coupling theorem of Komlos, Major and Tusnady, Corollary 5.5 of Chapter 7, to obtain a direct comparison between X_n and Z_n .

Suppose the β_i are continuous and the solution of the martingale problem for $(B_n, \delta_{X_n(0)})$ is unique. Then it follows from Theorem 5.1 of Chapter 6 that the solution can be obtained as a solution of

$$(3.2) \quad Z_n(t) = X_n(0) + \sum_i \ln^{-1} W_i \left(n \int_0^t \beta_i(Z_n(s)) ds \right) + \int_0^t F(Z_n(s)) ds$$

where the W_i are independent standard Brownian motions. By Corollary 5.5 and Remark 5.4 both of Chapter 7, we can assume the existence of centered Poisson processes \tilde{Y}_i such that

$$(3.3) \quad \sup_t \frac{|\tilde{Y}_i(t) - W_i(t)|}{\log(2 \vee t)} < \infty$$

and for $\beta, \gamma > 0$ there exist $\lambda, K, C > 0$ such that

$$(3.4) \quad P \left\{ \sup_{t \leq \beta n} |\tilde{Y}_l(t) - W_l(t)| > C \log n + x \right\} \leq Kn^{-\gamma} e^{-\lambda x}.$$

Since the W_l are independent, we can take the \tilde{Y}_l to be independent as well. Let X_n satisfy (2.3) using the \tilde{Y}_l constructed from the W_l . Then X_n and Z_n are defined on the same sample space and we can consider the difference $|X_n(t) - Z_n(t)|$. (Note that the pair (X_n, Z_n) is *not* a Markov process even though each component is.)

3.1 Theorem Let X, X_n , and Z_n be as above, and assume $\lim_{n \rightarrow \infty} X_n(0) = X(0)$. Fix $\varepsilon, T > 0$, and set $N_\varepsilon = \{y \in E : \inf_{t \leq T} |X(t) - y| \leq \varepsilon\}$. Let $\tilde{\beta}_l \equiv \sup_{x \in N_\varepsilon} \beta_l(x) < \infty$ and suppose $\tilde{\beta}_l = 0$ except for finitely many l . Suppose $M > 0$ satisfies

$$(3.5) \quad |\beta_l(x) - \beta_l(y)| \leq M|x - y|, \quad x, y \in N_\varepsilon,$$

and

$$(3.6) \quad |F(x) - F(y)| \leq M|x - y|, \quad x, y \in N_\varepsilon.$$

Let $\tau_n = \inf \{t : X_n(t) \notin N_\varepsilon \text{ or } Z_n(t) \notin N_\varepsilon\}$. (Note $P\{\tau_n > T\} \rightarrow 1$.) Then for $n \geq 2$ there is a random variable Γ_n^T and positive constants λ_T, C_T , and K_T depending on T , on M , and on the $\tilde{\beta}_l$, but *not* on n , such that

$$(3.7) \quad \sup_{t \leq T \wedge \tau_n} |X_n(t) - Z_n(t)| \leq \Gamma_n^T \frac{\log n}{n},$$

and

$$(3.8) \quad P\{\Gamma_n^T > C_T + x\} \leq K_T n^{-2} \exp \left\{ -\lambda_T x^{1/2} - \frac{\lambda_T x}{\log n} \right\}.$$

Proof. Again we can assume $\tilde{\beta}_l = \sup_{x \in E} \beta_l(x) < \infty$ and (3.5) and (3.6) are satisfied for all $x, y \in E$. Under these hypotheses we can drop the τ_n in (3.7).

By (3.4) there exist constants C_l^1, K_l^1 independent of n and nonnegative random variables $L_{l,n}^1$ such that

$$(3.9) \quad \sup_{t \leq \min\{\beta_l T, \Gamma_n^T\}} |\tilde{Y}_l(t) - W_l(t)| \leq C_l^1 \log n + L_{l,n}^1$$

and

$$(3.10) \quad P\{L_{l,n}^1 > x\} \leq K_l^1 n^{-2} e^{-4x}.$$

Define

$$(3.11) \quad \Lambda_{l,n}(z) = \sup_{u \leq n\beta_l T} \sup_{v \leq z \log n} |W_l(u+v) - W_l(u)|.$$

Let $k_n = [n\beta_l T/z \log n] + 1$. Then

$$(3.12) \quad \Lambda_{l,n}(z) \leq 3 \sup_{k \leq k_n} \sup_{v \leq z \log n} |W_l(kz \log n + v) - W_l(kz \log n)|,$$

and hence for $z, c, x \geq 0$

$$\begin{aligned} (3.13) \quad P\{\Lambda_{l,n}(z) \geq z^{1/2}(c \log n + x)\} &\leq k_n P\left\{3 \sup_{v \leq z \log n} |W_l(v)| \geq z^{1/2}(c \log n + x)\right\} \\ &\leq 2k_n P\{|W_l(1)| \geq \frac{1}{3}(c \log n + x)(\log n)^{-1/2}\} \\ &\leq ak_n \exp\left\{-\frac{(c^2 \log n + 2cx + x^2/\log n)}{18}\right\} \end{aligned}$$

where $a = \sup_{x \geq 0} e^{x^2/2} P\{|W_l(1)| \geq x\}$. Since $\Lambda_{l,n}(z)$ is monotone in z ,

$$\begin{aligned} (3.14) \quad \bar{\Lambda}_{l,n} &\equiv \sup_{0 \leq z \leq n\beta_l T} (z+1)^{-1/2} \Lambda_{l,n}(z) \\ &\leq \sup_{1 \leq m \leq n\beta_l T + 1} m^{-1/2} \Lambda_{l,n}(m) \end{aligned}$$

for integer m . Therefore

$$\begin{aligned} (3.15) \quad P\{\bar{\Lambda}_{l,n} > c \log n + x\} &\leq (1 + n\beta_l T)ak_n \exp\left\{-\frac{(c^2 \log n + 2cx + x^2/\log n)}{18}\right\}. \end{aligned}$$

Since $k_n = O(n)$ there exist constants C_l^2 and K_l^2 , independent of n , such that

$$(3.16) \quad P\{\bar{\Lambda}_{l,n} > C_l^2 \log n + x\} \leq K_l^2 n^{-2} \exp\left\{-\lambda x - \frac{x^2}{18 \log n}\right\}.$$

Setting $L_{l,n}^2 = (\bar{\Lambda}_{l,n} - C_l^2 \log n) \vee 0$, we have

$$(3.17) \quad P\{L_{l,n}^2 > x\} \leq K_l^2 n^{-2} \exp\left\{-\lambda x - \frac{x^2}{18 \log n}\right\}.$$

Taking the difference between (2.3) and (3.2) (recall only finitely many β_i are nonzero),

$$\begin{aligned}
(3.18) \quad & |X_n(t) - Z_n(t)| \\
& \leq n^{-1} \sum_l |l| \left| \tilde{Y}_l \left(n \int_0^t \beta_l(X_n(s)) ds \right) - W_l \left(n \int_0^t \beta_l(X_n(s)) ds \right) \right| \\
& + n^{-1} \sum_l |l| \left| W_l \left(n \int_0^t \beta_l(X_n(s)) ds \right) - W_l \left(n \int_0^t \beta_l(Z_n(s)) ds \right) \right| \\
& + \int_0^t |F(X_n(s)) - F(Z_n(s))| ds \\
& \leq n^{-1} \sum_l |l| (C_l^1 \log n + L_{l,n}^1) \\
& + n^{-1} \sum_l |l| \Lambda_{l,n} \left(\frac{n}{\log n} \left| \int_0^t (\beta_l(X_n(s)) - \beta_l(Z_n(s))) ds \right| \right) \\
& + \int_0^t M |X_n(s) - Z_n(s)| ds.
\end{aligned}$$

Let $\gamma_n(t) = n |X_n(t) - Z_n(t)| / \log n$. Then

$$\begin{aligned}
(3.19) \quad \gamma_n(t) & \leq e^{Mt} \left[\sum_l |l| \left(C_l^1 + \frac{L_{l,n}^1}{\log n} \right) \right. \\
& \quad \left. + \left(1 + M \int_0^t \gamma_n(s) ds \right)^{1/2} \sum_l |l| \left(C_l^2 + \frac{L_{l,n}^2}{\log n} \right) \right],
\end{aligned}$$

and setting $\bar{\gamma}_n = \sup_{t \leq T} \gamma_n(t)$ we have

$$\begin{aligned}
(3.20) \quad \bar{\gamma}_n & \leq e^{MT} \sum_l |l| \left(C_l^1 + \frac{L_{l,n}^1}{\log n} \right) \\
& \quad + \left(\frac{1}{MT} + \bar{\gamma}_n \right)^{1/2} e^{MT} (MT)^{1/2} \sum_l |l| \left(C_l^2 + \frac{L_{l,n}^2}{\log n} \right).
\end{aligned}$$

The inequality $\gamma \leq a + \gamma^{1/2}b$ implies $\gamma \leq 2a + b^2$. Hence there is a constant C_T such that

$$\begin{aligned}
(3.21) \quad \bar{\gamma}_n & \leq \frac{1}{MT} + 2e^{MT} \sum_l |l| \left(C_l^1 + \frac{L_{l,n}^1}{\log n} \right) + MT e^{2MT} \left(\sum_l |l| \left(C_l^2 + \frac{L_{l,n}^2}{\log n} \right) \right)^2 \\
& \leq C_T + \frac{2e^{MT}}{\log n} \sum_l |l| L_{l,n}^1 + \frac{MT e^{2MT}}{(\log n)^2} \left(\sum_l |l| L_{l,n}^2 \right)^2 \\
& \equiv C_T + L_n \equiv \Gamma_n^T.
\end{aligned}$$

Since the sums in (3.21) are finite, (3.10) and (3.17) imply there exist constants $K_T, \lambda_T > 0$ such that

$$(3.22) \quad P\{L_n > x\} \leq K_T n^{-1} \exp \left\{ -\lambda_T x^{1/2} - \frac{\lambda_T x}{\log n} \right\},$$

and (3.8) follows. \square

We now have two approximations for X_n , namely Z_n and \tilde{Z}_n . The question arises as to which is "best." The answer is that, at least asymptotically, for bounded time intervals they are essentially equivalent. In order to make this precise, we consider Z_n and \tilde{Z}_n as solutions of stochastic integral equations:

$$(3.23) \quad Z_n(t) = Z_n(0) + n^{-1/2} \sum_l \int_0^t l \beta_l^{1/2}(Z_n(s)) dW_l(s) + \int_0^t F(Z_n(s)) ds$$

and

$$(3.24) \quad \begin{aligned} \tilde{Z}_n(t) = X(0) + n^{-1/2} \sum_l \int_0^t l \beta_l^{1/2}(X(s)) dW_l(s) \\ + \int_0^t (F(X(s)) + \partial F(X(s))(\tilde{Z}_n(s) - X(s))) ds. \end{aligned}$$

3.2 Theorem In addition to the assumptions of Theorem 3.1, suppose that the $\beta_l^{1/2}$ are continuously differentiable and that F is twice continuously differentiable in N_t . Let Z_n and \tilde{Z}_n satisfy (3.23) and (3.24). Let $n(X_n(0) - X(0)) \rightarrow 0$. Then

$$(3.25) \quad \sup_{t \leq T} |n(Z_n(t) - \tilde{Z}_n(t)) - \hat{V}(t)| \xrightarrow{P} 0,$$

where \hat{V} satisfies

$$(3.26) \quad \begin{aligned} \hat{V}(t) = \sum_l l \int_0^t \nabla \beta_l^{1/2}(X(s)) \cdot V(s) dW_l(s) + \int_0^t \partial F(X(s)) \hat{V}(s) ds \\ + \int_0^t \sum_{i,j} \partial_i \partial_j F(X(s)) V_i(s) V_j(s) ds \end{aligned}$$

and V satisfies

$$(3.27) \quad V(t) = \sum_l l \int_0^t \beta_l^{1/2}(X(s)) dW_l(s) + \int_0^t \partial F(X(s)) V(s) ds.$$

3.3 Remark The assumption that $\beta_l > 0$ for only finitely many l can be replaced by

$$\sum_l l^2 \beta_l < \infty \quad \text{and} \quad \sum_l l^2 \sup_{x \in N_t} |\nabla \beta_l^{1/2}(x)|^2 < \infty. \quad \square$$

Proof. Let $U_n = \sqrt{n}(Z_n - X)$. A simple argument gives

$$(3.28) \quad E \left[\sup_{t \leq T} |U_n(t) - V(t)|^2 \right] \rightarrow 0.$$

We have

$$\begin{aligned}
 (3.29) \quad & \hat{V}_n(t) \equiv n(Z_n(t) - \bar{Z}_n(t)) \\
 & = n(X_n(0) - X(0)) + \sum_l l \int_0^t \sqrt{n}(\beta_l^{1/2}(Z_n(s)) - \beta_l^{1/2}(X(s))) dW_l(s) \\
 & \quad + \int_0^t \partial F(X(s)) \hat{V}_n(s) ds \\
 & \quad + \int_0^t \sqrt{n}(\sqrt{n}(F(Z_n(s)) - F(X(s))) - n^{-1/2} \partial F(X(s)) U_n(s)) ds.
 \end{aligned}$$

By hypothesis $n(X_n(0) - X(0)) \rightarrow 0$. The second term on the right converges to the first term on the right of (3.26) by (3.28) of this chapter and (2.19) of Chapter 5. The limit in (3.28) also implies the last term in (3.29) converges to the last term in (3.26), and the theorem follows. \square

By the construction in Theorem 3.1, for each $T > 0$, there is a constant $C_T > 0$ such that

$$(3.30) \quad \lim_{n \rightarrow \infty} P \left\{ \sup_{t \leq T} |X_n(t) - Z_n(t)| > \frac{C_T \log n}{n} \right\} = 0,$$

whereas by Theorem 3.2, for any sequence $a_n \rightarrow \infty$,

$$(3.31) \quad \lim_{n \rightarrow \infty} P \left\{ \sup_{t \leq T} |Z_n(t) - \bar{Z}_n(t)| > \frac{a_n}{n} \right\} = 0.$$

Since Bartfai's theorem, Theorem 5.6 of Chapter 7, implies that (3.30) is best possible, at least in some cases, we see that asymptotically the two approximations are essentially the same.

4. HITTING DISTRIBUTIONS

The time and location of the first exit of a process from a region are frequently of interest. The results of Section 2 give the asymptotic behavior for these quantities as well. We characterize the region of interest as the set on which a given function φ is positive.

4.1 Theorem Let φ be continuously differentiable on \mathbb{R}^d . Let X_n and X satisfy (2.3) and (2.8), respectively, with $\varphi(X(0)) > 0$, and suppose the conditions of Theorem 2.3 hold. Let

$$(4.1) \quad \tau_n = \inf \{t: \varphi(X_n(t)) \leq 0\}$$

and

$$(4.2) \quad \tau = \inf \{t: \varphi(X(t)) \leq 0\}.$$

Suppose $\tau < \infty$ and

$$(4.3) \quad \nabla \varphi(X(\tau)) \cdot F(X(\tau)) < 0.$$

Then

$$(4.4) \quad \sqrt{n}(\tau_n - \tau) \Rightarrow - \frac{\nabla \varphi(X(\tau)) \cdot V(\tau)}{\nabla \varphi(X(\tau)) \cdot F(X(\tau))}$$

and

$$(4.5) \quad \sqrt{n}(X_n(\tau_n) - X(\tau)) \Rightarrow V(\tau) - \frac{\nabla \varphi(X(\tau)) \cdot V(\tau)}{\nabla \varphi(X(\tau)) \cdot F(X(\tau))} F(X(\tau)).$$

4.2 Remark One example of interest is the number of susceptibles remaining in the population when the last infective is removed in the epidemic model described in Section 1. This situation is not covered directly by the theorem. (In particular, $\tau = \infty$). However see Problem 5. \square

Proof. Note that

$$(4.6) \quad \frac{\partial}{\partial t} \varphi(X(t)) = \nabla \varphi(X(t)) \cdot F(X(t))$$

so that (4.3) implies $\varphi(X(\tau - \varepsilon)) > 0$ and $\varphi(X(\tau + \varepsilon)) < 0$ for $0 < \varepsilon < \tau$.

Since $X_n \rightarrow X$ a.s. uniformly on bounded time intervals, it follows that $\tau_n \rightarrow \tau$ a.s. Since $\varphi(X_n(\tau_n)) \leq 0$ and $\varphi(X_n(\tau_n -)) \geq 0$,

$$(4.7) \quad \begin{aligned} |\sqrt{n}\varphi(X_n(\tau_n))| &\leq |\sqrt{n}(\varphi(X_n(\tau_n)) - \varphi(X_n(\tau_n -)))| \\ &= |\nabla \varphi(\theta_n) \cdot (V_n(\tau_n) - V_n(\tau_n -))| \end{aligned}$$

for some θ_n on the line between $X_n(\tau_n)$ and $X_n(\tau_n -)$, and since $V_n \Rightarrow V$ and V is continuous, the right side of (4.7) converges in distribution to zero. By the continuity of X , $\varphi(X(\tau)) = 0$ and

$$(4.8) \quad \begin{aligned} \sqrt{n}(\varphi(X(\tau)) - \varphi(X(\tau_n))) \\ &= \sqrt{n}(\varphi(X_n(\tau_n)) - \varphi(X(\tau_n))) - \sqrt{n}\varphi(X_n(\tau_n)) \\ &= \sqrt{n}(\varphi(X(\tau_n) + n^{-1/2}V_n(\tau_n)) - \varphi(X(\tau_n))) - \sqrt{n}\varphi(X_n(\tau_n)) \\ &\Rightarrow \nabla \varphi(X(\tau)) \cdot V(\tau). \end{aligned}$$

But the left side of (4.8) is asymptotic to

$$(4.9) \quad -\nabla \varphi(X(\tau)) \cdot F(X(\tau))\sqrt{n}(\tau_n - \tau),$$

and (4.4) follows.

Finally

$$\begin{aligned}
 (4.10) \quad & \sqrt{n}(X_n(\tau_n) - X(\tau)) \\
 &= V_n(\tau_n) + \sqrt{n}(X(\tau_n) - X(\tau)) \\
 &\Rightarrow V(\tau) - \frac{\nabla\varphi(X(\tau)) \cdot V(\tau)}{\nabla\varphi(X(\tau)) \cdot F(X(\tau))} F(X(\tau)). \quad \square
 \end{aligned}$$

5. PROBLEMS

1. Let X_n be the logistic growth model described in Section 1.
 - (a) Compute the parameters of the limiting Gaussian process V given by Theorem 2.3.
 - (b) Let Z_n and \tilde{Z}_n be the approximations of X_n discussed in Section 3. Assuming $Z_n(0) = \tilde{Z}_n(0) \neq 0$, show that Z_n eventually absorbs at zero, but that \tilde{Z}_n is asymptotically stationary (and nondegenerate).
2. Consider the chemical reaction model for $R_1 + R_2 \rightleftharpoons R_3$ with parameters given by (1.10).
 - (a) Compute the parameters of the limiting Gaussian process V given by Theorem 2.3.
 - (b) Let $X(0)$ be the fixed point of the limiting deterministic model (so $X(t) = X(0)$ for all $t \geq 0$). Then V_n is a Markov process with stationary transition probabilities. Apply Theorem 9.14 of Chapter 4 to show that the stationary distribution for V_n converges to the stationary distribution for V .
3. Use the fact that, under the assumptions of Theorem 2.1,

$$(5.1) \quad X_n(t) - X_n(0) - \int_0^t F(X_n(s)) ds$$

is a local martingale and Gronwall's inequality to estimate $P\{\sup_{s \leq t} |X_n(s) - X(s)| \geq \varepsilon\}$.

4. Under the hypotheses of Theorems 3.1 and 3.2, show that for any bounded $U \subset \mathbb{R}^d$ with smooth boundary,
- $$(5.2) \quad |P\{V_n(t) \in U\} - P\{V(t) \in U\}| = O\left(\frac{\log n}{\sqrt{n}}\right).$$
5. Let $X_n = (S_n, I_n)$ be the epidemic model described in Section 1 and let $X = (S, I)$ denote the limiting deterministic model (S for susceptible, I for infectious). Let $\tau_n = \inf\{t: I_n(t) = 0\}$.

- (a) Show that if $I(0) > 0$, then $I(t) > 0$ for all $t \geq 0$, but that $\lim_{t \rightarrow \infty} I(t) = 0$ and $S(\infty) \equiv \lim_{t \rightarrow \infty} S(t)$ exists.
- (b) Show that if $\sqrt{n}(X_n(0) - X(0))$ converges, then $\sqrt{n}(S_n(\tau_n) - S(\infty))$ converges in distribution.

Hint: Let γ_n satisfy

$$(5.3) \quad \int_0^{\gamma_n(t)} I_n(s) \, ds = t, \quad t < \int_0^\infty I_n(s) \, ds,$$

and show that $X_n(\gamma_n(\cdot))$ extends to a process satisfying the conditions of Theorem 4.1 with $\varphi(x_1, x_2) = x_2$.

6. NOTES

Most of the material in this chapter is from Kurtz (1970b, 1971, 1978a). Norman (1974) gives closely related results including conditions under which the convergence of $V_n(t)$ to $V(t)$ is uniform for all t (see Problem 2). Barbour (1974, 1980) studies the same class of processes giving rates of convergence for the distributions of certain functionals of V_n in the first paper and for the stationary distributions in the second. Berry-Esseen type results have been given by Allain (1976) and Alm (1978). Analogous results for models with age dependence have been given by Wang (1977).

Darden and Kurtz (1985) study the situation in which the limiting deterministic model has a stable fixed point, extending the uniformity results of Norman and obtaining asymptotic exponentiality for the distribution of the exit time from a neighborhood of the stable fixed point.

12 | RANDOM EVOLUTIONS

The aim in this chapter is to study the asymptotic behavior of certain random evolutions as a small parameter tends to zero. We do not attempt to achieve the greatest generality, but only enough to be able to treat a variety of examples. Section 1 introduces the basic ideas and terminology in terms of perhaps the simplest example. Sections 2 and 3 consider the case in which the underlying process (or driving process) is Markovian and ergodic, while Section 4 requires it to be stationary and uniform mixing.

1. INTRODUCTION

One of the simplest examples of a random evolution can be described as follows. Let N be a Poisson process with parameter λ , and fix $\alpha > 0$. Given $(x, y) \in \mathbb{R} \times \{-1, 1\}$, define the pair of processes (X, Y) by

$$(1.1) \quad X(t) = x + \alpha \int_0^t Y(s) \, ds, \quad Y(t) = (-1)^{N(t)} y.$$

$X(t)$ and $\alpha Y(t)$ represent the position and velocity at time t of a particle moving in one dimension at constant speed α , but subject to reversals of direction at the jump times of N , given initial position x and velocity αy .

Let us first observe that (X, Y) is a Markov process in $\mathbb{R} \times \{-1, 1\}$. For if we define for each $t \geq 0$ the linear operator $T(t)$ on $\hat{C}(\mathbb{R} \times \{-1, 1\})$ by

$$(1.2) \quad T(t)f(x, y) = E[f(X(t), Y(t))],$$

where (X, Y) is given by (1.1), then the Markov property of Y implies that

$$(1.3) \quad E[f(X(t), Y(t)) | \mathcal{F}_s^Y] = T(t-s)f(X(s), Y(s))$$

for all $f \in \hat{C}(\mathbb{R} \times \{-1, 1\})$, $(x, y) \in \mathbb{R} \times \{-1, 1\}$, and $t > s \geq 0$. It follows easily that $\{T(t)\}$ is a Feller semigroup on $\hat{C}(\mathbb{R} \times \{-1, 1\})$ and (X, Y) is a Markov process in $\mathbb{R} \times \{-1, 1\}$ corresponding to $\{T(t)\}$.

Clearly, however, X itself is non-Markovian. Nevertheless, while Y visits $y \in \{-1, 1\}$, X evolves according to the Feller semigroup $\{T_y(t)\}$ on $\hat{C}(\mathbb{R})$ defined by

$$(1.4) \quad T_y(t)f(x) = f(x + \alpha t y).$$

Consequently, letting τ_1, τ_2, \dots denote the jump times of Y , the evolution of X over the time interval $[s, t]$ is described by the operator-valued random variable

$$(1.5) \quad \mathcal{F}(s, t) = T_{Y(s)}((\tau_1 \vee s) \wedge t - s) T_{Y(\tau_1)}((\tau_2 \vee s) \wedge t - (\tau_1 \vee s) \wedge t) \cdots$$

in the sense that

$$(1.6) \quad E[f(X(t), Y(t)) | \mathcal{F}_s^X \vee \mathcal{F}_t^Y] = \mathcal{F}(s, t)\{f(\cdot, Y(t))\}(X(s))$$

for all $f \in \hat{C}(\mathbb{R} \times \{-1, 1\})$. The family $\{\mathcal{F}(s, t), t \geq s \geq 0\}$ satisfies

$$(1.7) \quad \mathcal{F}(s, t)\mathcal{F}(t, u) = \mathcal{F}(s, u), \quad s \leq t \leq u,$$

and is therefore called a *random evolution*. Because Y "controls" the development of X , we occasionally refer to Y as the *driving process* and to X as the *driven process*. Observe that (1.3) and (1.6) specify the relationship between $\{T(t)\}$ and $\{\mathcal{F}(t)\}$. (Of course, in the special case of (1.1), the left side of (1.6) can be replaced by $f(X(t), Y(t))$ because $\mathcal{F}_s^X \subset \mathcal{F}_t^Y$. In general, however, X need not evolve deterministically while Y visits y .)

To determine the generator of the semigroup $\{T(t)\}$, let $f \in C^{1,0}(\mathbb{R} \times \{-1, 1\})$ and $t_2 > t_1 \geq 0$. Then

$$(1.8) \quad f(X(t_2), Y(t_2)) - f(X(t_1), Y(t_2)) = \int_{t_1}^{t_2} \alpha Y(s) f_x(X(s), Y(t_2)) ds$$

and

$$(1.9) \quad E[f(X(t_1), Y(t_2)) - f(X(t_1), Y(t_1)) | \mathcal{F}_{t_1}^Y]$$

$$= E \left[\int_{t_1}^{t_2} \lambda \{f(X(t_1), -Y(s)) - f(X(t_1), Y(s))\} ds \middle| \mathcal{F}_{t_1}^Y \right],$$

so by Lemma 3.4 of Chapter 4,

$$(1.10) \quad f(X(t), Y(t)) - \int_0^t Af(X(s), Y(s)) \, ds$$

is an $\{\mathcal{F}_t^\gamma\}$ -martingale, where

$$(1.11) \quad Af(x, y) = \alpha y f_x(x, y) + \lambda \{f(x, -y) - f(x, y)\}.$$

Identifying $\hat{C}(\mathbb{R} \times \{-1, 1\})$ with $\hat{C}(\mathbb{R}) \times \hat{C}(\mathbb{R})$, we can rewrite A as

$$(1.12) \quad A = \alpha \begin{pmatrix} \frac{d}{dx} & 0 \\ 0 & -\frac{d}{dx} \end{pmatrix} + \lambda \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$$

with $\mathcal{D}(A) = \hat{C}^1(\mathbb{R}) \times \hat{C}^1(\mathbb{R})$. Since $\mathcal{D}(A) \subset \hat{C}(\mathbb{R}) \times \hat{C}(\mathbb{R})$, it follows from the martingale property and the strong continuity of $\{T(t)\}$ that the generator of $\{T(t)\}$ extends A . But by Problem 1 and Corollary 7.2, both of Chapter 1, A generates a strongly continuous semigroup on $\hat{C}(\mathbb{R}) \times \hat{C}(\mathbb{R})$. We conclude from Proposition 4.1 of Chapter 1 that A is precisely the generator of $\{T(t)\}$.

This has an interesting consequence. Let $f \in \hat{C}^2(\mathbb{R})$ and define

$$(1.13) \quad \begin{pmatrix} g(t, x) \\ h(t, x) \end{pmatrix} = T(t) \begin{pmatrix} f \\ f \end{pmatrix}(x)$$

for all $(t, x) \in [0, \infty) \times \mathbb{R}$. Then g and h belong to $\hat{C}^2([0, \infty) \times \mathbb{R})$ and satisfy the system of partial differential equations

$$(1.14) \quad \begin{aligned} g_t &= \alpha g_x - \lambda(g - h), \\ h_t &= -\alpha h_x + \lambda(g - h). \end{aligned}$$

Letting $u = \frac{1}{2}(g + h)$ and $v = \frac{1}{2}(g - h)$, we have

$$(1.15) \quad \begin{aligned} u_t &= \alpha v_x, \\ v_t &= \alpha u_x - 2\lambda v. \end{aligned}$$

Hence $u_{tt} = \alpha v_{tx} = \alpha u_{xx} - 2\lambda v_x = \alpha u_{xx} - (2\lambda/\alpha)u_t$, or

$$(1.16) \quad u_{tt} = \frac{\alpha^2}{2\lambda} u_{xx} - \frac{\alpha}{2\lambda} u_{tt}.$$

This is a hyperbolic equation known as the telegrapher's equation. Random evolutions can be used to represent the solutions of certain of these equations probabilistically, though that is not our concern here.

In the context of the present example, we are interested instead in the asymptotic distribution of X as $\alpha \rightarrow \infty$ and $\lambda \rightarrow \infty$ with $\alpha^2 = \lambda$. Let $0 < \varepsilon < 1$ and observe that with $\alpha = 1/\varepsilon$ and $\lambda = 1/\varepsilon^2$, (1.16) becomes

$$(1.17) \quad u_t = \frac{1}{2} u_{xx} - \frac{\varepsilon}{2} u_{tt}.$$

This suggests that as $\varepsilon \rightarrow 0$, we should have $X \Rightarrow x + W$ in $C_{\mathbb{R}}[0, \infty)$, where W is a standard one-dimensional Brownian motion. To make this precise, let N be a Poisson process with parameter 1. Given $(x, y) \in \mathbb{R} \times \{-1, 1\}$, define $(X^\varepsilon, Y^\varepsilon)$ for $0 < \varepsilon < 1$ by

$$(1.18) \quad X^\varepsilon(t) = x + \frac{1}{\varepsilon} \int_0^t Y^\varepsilon(s) \, ds, \quad Y^\varepsilon(t) = (-1)^{N(t/\varepsilon^2)} y.$$

By the Markov property of Y^ε ,

$$(1.19) \quad \begin{aligned} M^\varepsilon(t) &\equiv \frac{\varepsilon}{2} Y^\varepsilon(t) - \frac{\varepsilon}{2} y + \frac{1}{\varepsilon} \int_0^t Y^\varepsilon(s) \, ds \\ &= X^\varepsilon(t) - x + \frac{\varepsilon}{2} Y^\varepsilon(t) - \frac{\varepsilon}{2} y \end{aligned}$$

is a zero-mean $\{\mathcal{F}_t^{Y^\varepsilon}\}$ -martingale, and $M^\varepsilon(t)^2 - t$ is also an $\{\mathcal{F}_t^{Y^\varepsilon}\}$ -martingale. It follows immediately from the martingale central limit theorem (Theorem 1.4 of Chapter 7) that $M^\varepsilon \Rightarrow W$ in $D_{\mathbb{R}}[0, \infty)$, hence by (1.19) and Problem 25 of Chapter 3, that $X^\varepsilon \Rightarrow x + W$ in $C_{\mathbb{R}}[0, \infty)$.

There is an alternative proof of this result that generalizes much more readily. Let $f \in C^2(\mathbb{R})$, and define $f_\varepsilon \in C^{1,0}(\mathbb{R} \times \{-1, 1\})$ for $0 < \varepsilon < 1$ by

$$(1.20) \quad f_\varepsilon(x, y) = f(x) + \frac{\varepsilon}{2} y f'(x).$$

Then, defining A_ε by (1.11) with $\alpha = 1/\varepsilon$ and $\lambda = 1/\varepsilon^2$, we have

$$(1.21) \quad A_\varepsilon f_\varepsilon(x, y) = \frac{1}{\varepsilon} y f'(x) + \frac{1}{2} y^2 f''(x) - \frac{1}{\varepsilon} y f'(x) = \frac{1}{2} f''(x)$$

for all $(x, y) \in \mathbb{R} \times \{-1, 1\}$. The desired conclusion now follows from Proposition 1.1 of Chapter 5 and Corollary 8.7 of Chapter 4.

It is the purpose of this chapter to obtain limit theorems of the above type under a variety of assumptions. More specifically, given $F, G: \mathbb{R}^d \times E \rightarrow \mathbb{R}^d$, a process Y with sample paths in $D_E[0, \infty)$, and $x \in \mathbb{R}^d$, we consider the solution X^ε , where $0 < \varepsilon < 1$, of the differential equation

$$(1.22) \quad \frac{d}{dt} X^\varepsilon(t) = F\left(X^\varepsilon(t), Y\left(\frac{t}{\varepsilon^2}\right)\right) + \frac{1}{\varepsilon} G\left(X^\varepsilon(t), Y\left(\frac{t}{\varepsilon^2}\right)\right)$$

with initial condition $X^\varepsilon(0) = x$. Of course, F and G must satisfy certain smoothness and growth assumptions in order for X^ε to be well defined. In

Section 2, we consider the case in which E is a compact metric space and the driving process Y is Markovian and ergodic. (This clearly generalizes (1.18).) In Section 3, we allow E to be a locally compact, separable metric space. In Section 4, we again require that E be compact but allow Y to be stationary and uniform mixing instead of Markovian.

2. DRIVING PROCESS IN A COMPACT STATE SPACE

The argument following (1.20) provided the motivation for Corollary 7.8 of Chapter 1. We include essentially a restatement of that result in a form that is suitable for application to random evolutions with driving process in a compact state space.

First, however, we need to generalize the Riemann integral of Chapter 1, Section 1.

2.1 Lemma Let E be a metric space, let L be a separable Banach space, and let $\mu \in \mathcal{P}(E)$. If $f: E \rightarrow L$ is Borel measurable and

$$(2.1) \quad \int \|f(y)\| \mu(dy) < \infty,$$

then there exists a sequence $\{f_n\}$ of Borel measurable simple functions from E into L such that

$$(2.2) \quad \lim_{n \rightarrow \infty} \int \|f_n(y) - f(y)\| \mu(dy) = 0.$$

The separability assumption on L is unnecessary if E is σ -compact and f is continuous.

Proof. If L is separable, let $\{g_n\}$ be dense in L ; if E is σ -compact and f is continuous, then $f(E)$ is σ -compact, hence separable, so let $\{g_n\}$ be dense in $f(E)$. For $m, n = 1, 2, \dots$ define $A_{n,m} = \{g \in L: \|g - g_n\| < 1/m\} = \bigcup_{k=1}^{m-1} A_{k,m}$ and

$$(2.3) \quad h_{n,m}(y) = \sum_{k=1}^m \chi_{f^{-1}(A_{k,m})}(y) g_k.$$

Then, letting $B_{n,m} = \bigcup_{k=1}^m f^{-1}(A_{k,m})$, we have

$$(2.4) \quad \int \|h_{n,m}(y) - f(y)\| \mu(dy) \leq \frac{1}{m} \mu(B_{n,m}) + \int_{B_{n,m}} \|f(y)\| \mu(dy)$$

for $m, n = 1, 2, \dots$, and hence

$$(2.5) \quad \lim_{m \rightarrow \infty} \lim_{n \rightarrow \infty} \int \|h_{n,m}(y) - f(y)\| \mu(dy) = 0.$$

We conclude that there exists $\{m_n\}$ such that (2.2) holds with $f_n = h_{n, m_n}$. \square

Let E , L , and μ be as in Lemma 2.1, and let $f: E \rightarrow L$ be a Borel measurable simple function, that is,

$$(2.6) \quad f(y) = \sum_{k=1}^n \chi_{B_k}(y) g_k,$$

where $B_1, \dots, B_n \in \mathcal{B}(E)$ are disjoint, $g_1, \dots, g_n \in L$, and $n \geq 1$. Then we define

$$(2.7) \quad \int f d\mu = \sum_{k=1}^n \mu(B_k) g_k.$$

More generally, suppose $f: E \rightarrow L$ is Borel measurable and (2.1) holds. Let $\{f_n\}$ be as in Lemma 2.1. Then we define the (Bochner) integral of f with respect to μ by

$$(2.8) \quad \int f d\mu = \lim_{n \rightarrow \infty} \int f_n d\mu.$$

It is easily checked that this limit exists and is independent of the choice of the approximating sequence $\{f_n\}$.

In particular, if E is compact, L is arbitrary, and $\mu \in \mathcal{P}(E)$, then $\int f d\mu$ exists for all f belonging to $C_L(E)$, the space of continuous functions from E into L . We note that $C_L(E)$ is a Banach space with norm $\|f\| = \sup_{y \in E} \|f(y)\|$.

2.2 Proposition Let E be a compact metric space, L a Banach space, $P(t, y, \Gamma)$ a transition function on $[0, \infty) \times E \times \mathcal{B}(E)$, and $\mu \in \mathcal{P}(E)$. Assume that the formula

$$(2.9) \quad S(t)g(y) = \int g(z)P(t, y, dz)$$

defines a Feller semigroup $\{S(t)\}$ on $C(E)$ satisfying

$$(2.10) \quad \lim_{\lambda \rightarrow 0+} \lambda \int_0^\infty e^{-\lambda t} S(t)g dt = \int g d\mu$$

for all $g \in C(E)$, and let B_0 denote its generator. Observe that (2.9) also defines a strongly continuous contraction semigroup $\{S(t)\}$ on $C_L(E)$, and let B denote its generator.

Let D be a dense subspace of L , and for each $y \in E$, let Π_y and A_y be linear operators on L with domains containing D such that the functions $y \mapsto \Pi_y f$

and $y \rightarrow A_y f$ belong to $C_L(E)$ for each $f \in D$. Define linear operators Π and A on $C_L(E)$ with

$$(2.11) \quad \mathcal{D}(\Pi) = \{f \in C_L(E) : f(y) \in \mathcal{D}(\Pi_y) \text{ for every } y \in E,$$

and $y \rightarrow \Pi_y(f(y))$ belongs to $C_L(E)\}$

and

$$(2.12) \quad \mathcal{D}(A) = \{f \in C_L(E) : f(y) \in \mathcal{D}(A_y) \text{ for every } y \in E$$

and $y \rightarrow A_y(f(y))$ belongs to $C_L(E)\}$

by $(\Pi f)(y) = \Pi_y(f(y))$ and $(Af)(y) = A_y(f(y))$. Let \mathcal{D} be a subspace of $C_L(E)$ such that

$$(2.13) \quad \mathcal{D}_0 \equiv \left\{ \sum_{i=1}^n g_i(\cdot) f_i : g_1, \dots, g_n \in \mathcal{D}(B_0), f_1, \dots, f_n \in D, n \geq 1 \right\}$$

$$\subset \mathcal{D} \subset \mathcal{D}(\Pi) \cap \mathcal{D}(A) \cap \mathcal{D}(B),$$

and assume that, for $0 < \varepsilon < 1$, an extension of $\{(f, \Pi f + \varepsilon^{-1} Af + \varepsilon^{-2} Bf) : f \in \mathcal{D}\}$ generates a strongly continuous contraction semigroup $\{T_\varepsilon(t)\}$ on $C_L(E)$.

Suppose there is a linear operator V on $C_L(E)$ such that $Af \in \mathcal{D}(V)$ and $VAf \in \mathcal{D}$ for all $f \in D$ and $BVg = -g$ for all $g \in \mathcal{D}(V)$. (Here and below, we identify elements of L with constant functions in $C_L(E)$.) Put

$$(2.14) \quad C = \left\{ \left(f, \int (\Pi f + AV Af)(y) \mu(dy) \right) : f \in D \right\}.$$

Then C is dissipative, and if \tilde{C} , which is single-valued, generates a strongly continuous contraction semigroup $\{T(t)\}$ on L , then, for each $f \in L$, $\lim_{t \rightarrow 0} T_t(t)f = T(t)f$ for all $t \geq 0$, uniformly on bounded intervals.

2.3 Remark Suppose that (2.10) can be strengthened to

$$(2.15) \quad \int_0^\infty \left\| S(t)g - \int g \, d\mu \right\| dt < \infty, \quad g \in C_L(E).$$

By the uniform boundedness principle, there exists a constant M such that

$$(2.16) \quad \left\| \int_0^\infty \left(S(t)g - \int g \, d\mu \right) dt \right\| \leq M \|g\|, \quad g \in C_L(E),$$

and hence for each $y \in E$ there exists a finite signed Borel measure $v(y, \cdot)$ such that

$$(2.17) \quad \int_0^\infty \left(S(t)g(y) - \int g \, d\mu \right) dt = \int g(z) v(y, dz), \quad g \in C_L(E),$$

where the right side is defined using (2.8). If $\int (Af)(z)v(\cdot, dz) \in \mathcal{D}$ for all $f \in D$, then, by Remark 7.9(b) of Chapter 1, $V: \{Af: f \in D\} \rightarrow \mathcal{D}$ is given by

$$(2.18) \quad (Vg)(y) = \int g(z)v(y, dz). \quad \square$$

Proof. We claim first that

$$(2.19) \quad \left\{ \sum_{i=1}^n g_i(\cdot) f_i : g_1, \dots, g_n \in C(E), f_1, \dots, f_n \in L, n \geq 1 \right\}$$

is dense in $C_L(E)$. To see this, let $\varepsilon > 0$ and choose $y_1, \dots, y_n \in E$ such that $E = \bigcup_{i=1}^n B(y_i, \varepsilon)$. Let $g_1, \dots, g_n \in C(E)$ be a partition of unity, that is, for $i = 1, \dots, n$, $g_i \geq 0$, $\text{supp } g_i \subset B(y_i, \varepsilon)$, and $\sum_{i=1}^n g_i \equiv 1$ (Rudin (1974), Theorem 2.13). Given $f \in C_L(E)$, let $f_\varepsilon = \sum_{i=1}^n g_i(\cdot) f(y_i)$. Then f_ε belongs to (2.19) and

$$(2.20) \quad \|f_\varepsilon - f\| \leq \sup \{\|f(x) - f(y)\| : r(x, y) < \varepsilon\},$$

where r is the metric for E . But the right side of (2.20) tends to zero as $\varepsilon \rightarrow 0$, so the claim is proved. Since $\mathcal{D}(B_0)$ is dense in $C(E)$ and D is dense in L , we also have \mathcal{D}_0 dense in $C_L(E)$.

It follows that $\{S(t)\}$ is strongly continuous on $C_L(E)$, (2.10) holds for all $f \in C_L(E)$, and (2.16) implies (2.17). Note also that $S(t): \mathcal{D}_0 \rightarrow \mathcal{D}_0$, so \mathcal{D}_0 is a core for B by Proposition 3.3 of Chapter 1.

We apply Corollary 7.8 of Chapter 1 (see Remark 7.9(c) in that chapter) with the roles of Π and A played by the restrictions of Π and A to \mathcal{D} . \square

2.4 Theorem Let E be a compact metric space, let $F, G \in C_{\mathbb{R}^d}(\mathbb{R}^d \times E)$, and suppose that for each $n \geq 1$ there exists a constant M_n for which

$$(2.21) \quad |F(x, y) - F(x', y)| \leq M_n |x - x'|, \quad |x| \vee |x'| \leq n, y \in E,$$

that $G_1, \dots, G_d \in C^{1,0}(\mathbb{R}^d \times E)$, and that

$$(2.22) \quad \sup_{(x, y) \in \mathbb{R}^d \times E} \frac{|F(x, y)| \vee |G(x, y)|}{1 + |x|} < \infty.$$

Let $\{S(t)\}$ be a Feller semigroup on $C(E)$, let $\mu \in \mathcal{P}(E)$, and assume that

$$(2.23) \quad \lim_{\lambda \rightarrow 0+} \lambda \int_0^\infty e^{-\lambda t} S(t)g dt = \int g d\mu, \quad g \in C(E).$$

Let B_0 denote the generator of $\{S(t)\}$.

Suppose that $\int G(x, y)\mu(dy) = 0$ for all $x \in \mathbb{R}^d$ and that there exists for each $y \in E$ a finite, signed, Borel measure $v(y, \cdot)$ on E such that the function $H: \mathbb{R}^d \times E \rightarrow \mathbb{R}^d$, defined by

$$(2.24) \quad H(x, y) = \int G(x, z)v(y, dz),$$

satisfies, for $i = 1, \dots, d$, $H_i \in C^{1,0}(\mathbb{R}^d \times E)$, $H_i(x, \cdot) \in \mathcal{D}(B_0)$ for each $x \in \mathbb{R}^d$, and $B_0[H_i(x, \cdot)](y) = -G_i(x, y)$ for all $(x, y) \in \mathbb{R}^d \times E$. Fix $\mu_0 \in \mathcal{P}(E)$, and let Y be a Markov process corresponding to $\{S(t)\}$ with sample paths in $D_E[0, \infty)$ and initial distribution μ_0 . Fix $x_0 \in \mathbb{R}^d$, and define X^ϵ for $0 < \epsilon < 1$ to be the solution of the differential equation

$$(2.25) \quad \frac{d}{dt} X^\epsilon(t) = F\left(X^\epsilon(t), Y\left(\frac{t}{\epsilon^2}\right)\right) + \frac{1}{\epsilon} G\left(X^\epsilon(t), Y\left(\frac{t}{\epsilon^2}\right)\right)$$

with initial condition $X^\epsilon(0) = x_0$. Put

$$(2.26) \quad C = \left\{ \left(f, \frac{1}{2} \sum_{i,j=1}^d a_{ij} \partial_i \partial_j f + \sum_{i=1}^d b_i \partial_i f \right) : f \in C_c^2(\mathbb{R}^d) \right\},$$

where

$$(2.27) \quad a_{ij}(x) = \int G_i(x, y) H_j(x, y) \mu(dy) + \int G_j(x, y) H_i(x, y) \mu(dy)$$

and

$$(2.28) \quad b_i(x) = \int F_i(x, y) \mu(dy) + \int G(x, y) \cdot \nabla_x H_i(x, y) \mu(dy).$$

Then C is dissipative. Assume that \hat{C} generates a Feller semigroup $\{T(t)\}$ on $\hat{C}(\mathbb{R}^d)$, and let X be a Markov process corresponding to $\{T(t)\}$ with sample paths in $C_{\mathbb{R}^d}[0, \infty)$ and initial distribution δ_{x_0} . Then $X^\epsilon \Rightarrow X$ in $C_{\mathbb{R}^d}[0, \infty)$ as $\epsilon \rightarrow 0$.

2.5 Remark Suppose that (2.23) can be strengthened to

$$(2.29) \quad \int_0^\infty \sup_{(x, y) \in \mathbb{R}^d \times E} \left| S(t)[g(x, \cdot)](y) - \int g(x, z) \mu(dz) \right| dt < \infty,$$

$$g \in \hat{C}(\mathbb{R}^d \times E).$$

Then $v(y, dz)$ is as in (2.17) with $L = \hat{C}(\mathbb{R}^d)$. \square

Proof. We identify $C_{\hat{C}(\mathbb{R}^d)}(E)$ with $\hat{C}(\mathbb{R}^d \times E)$ and apply Proposition 2.2 with $L = \hat{C}(\mathbb{R}^d)$, $D = C_c^2(\mathbb{R}^d)$, $\Pi_y = F(\cdot, y) \cdot \nabla$, $A_y = G(\cdot, y) \cdot \nabla$, $\mathcal{D}(\Pi_y) = \mathcal{D}(A_y) = C_c^1(\mathbb{R}^d)$, and

$$(2.30) \quad \mathcal{D} = \{f \in C_c^{1,0}(\mathbb{R}^d \times E) : f(x, \cdot) \in \mathcal{D}(B_0) \text{ for all } x \in \mathbb{R}^d,$$

$$\text{and } (x, y) \mapsto B_0[f(x, \cdot)](y) \text{ belongs to } \hat{C}(\mathbb{R}^d \times E)\}.$$

Clearly, $\mathcal{D} \subset D(\Pi) \cap \mathcal{D}(A)$. We claim that $\mathcal{D} \subset \mathcal{D}(B)$. To see this, let

$$(2.31) \quad \hat{B} = \{(f, g) \in \hat{C}(\mathbb{R}^d \times E) \times \hat{C}(\mathbb{R}^d \times E) :$$

$$f(x, \cdot) \in \mathcal{D}(B_0) \text{ for all } x \in \mathbb{R}^d,$$

$$\text{and } g(x, y) = B_0[f(x, \cdot)](y) \text{ for all } (x, y) \in \mathbb{R}^d \times E\},$$

and observe that \hat{B} is a dissipative linear extension of the generator B of $\{S(t)\}$ on $\hat{C}(\mathbb{R}^d \times E)$, and hence $\hat{B} = B$ by Proposition 4.1 of Chapter 1.

Next, fix $\varepsilon \in (0, 1)$, and define the contraction semigroup $\{T_\varepsilon(t)\}$ on $B(\mathbb{R}^d \times E)$ by

$$(2.32) \quad T_\varepsilon(t)f(x, y) = E \left[f \left(X^\varepsilon(t), Y \left(\frac{t}{\varepsilon^2} \right) \right) \right],$$

where Y is a Markov process corresponding to $\{S(t)\}$ with sample paths in $D_E[0, \infty)$ and initial distribution δ_y , and X^ε satisfies the differential equation (2.25) with initial condition $X^\varepsilon(0) = x$. The semigroup property follows from the identity

$$\begin{aligned} (2.33) \quad & E \left[f \left(X^\varepsilon(t), Y \left(\frac{t}{\varepsilon^2} \right) \right) \middle| \mathcal{F}_{s/\varepsilon^2}^Y \right] \\ &= E \left[f \left(X^\varepsilon(s) + \int_s^t \left(F + \frac{1}{\varepsilon} G \right) \left(X^\varepsilon(u), Y \left(\frac{u}{\varepsilon^2} \right) \right) du, Y \left(\frac{t}{\varepsilon^2} \right) \right) \middle| \mathcal{F}_{s/\varepsilon^2}^Y \right] \\ &= T_\varepsilon(t-s)f \left(X^\varepsilon(s), Y \left(\frac{s}{\varepsilon^2} \right) \right), \end{aligned}$$

valid for all $t > s \geq 0$ and $f \in B(\mathbb{R}^d \times E)$ by the Markov property of Y and the fact that $X^\varepsilon(s + \cdot)$ solves the differential equation (2.25) with $(X^\varepsilon(0), Y(\cdot/\varepsilon^2))$ replaced by $(X^\varepsilon(s), Y((s + \cdot)/\varepsilon^2))$. We leave it to the reader to check that $T_\varepsilon(t): \hat{C}(\mathbb{R}^d \times E) \rightarrow \hat{C}(\mathbb{R}^d \times E)$ for all $t \geq 0$. Using (2.22) we conclude that $T_\varepsilon(t): \hat{C}(\mathbb{R}^d \times E) \rightarrow \hat{C}(\mathbb{R}^d \times E)$ for every $t \geq 0$. Let $f \in \mathcal{D}$ and $t_2 > t_1 \geq 0$. Then

$$\begin{aligned} (2.34) \quad & f \left(X^\varepsilon(t_2), Y \left(\frac{t_2}{\varepsilon^2} \right) \right) - f \left(X^\varepsilon(t_1), Y \left(\frac{t_2}{\varepsilon^2} \right) \right) \\ &= \int_{t_1}^{t_2} \left(F + \frac{1}{\varepsilon} G \right) \left(X^\varepsilon(s), Y \left(\frac{t_2}{\varepsilon^2} \right) \right) \cdot \nabla_x f \left(X^\varepsilon(s), Y \left(\frac{t_2}{\varepsilon^2} \right) \right) ds \end{aligned}$$

and

$$\begin{aligned} (2.35) \quad & E \left[f \left(X^\varepsilon(t_1), Y \left(\frac{t_2}{\varepsilon^2} \right) \right) - f \left(X^\varepsilon(t_1), Y \left(\frac{t_1}{\varepsilon^2} \right) \right) \middle| \mathcal{F}_{t_1/\varepsilon^2}^Y \right] \\ &= E \left[\int_{t_1}^{t_2} \frac{1}{\varepsilon^2} B_0[f(X^\varepsilon(t_1), \cdot)] \left(Y \left(\frac{s}{\varepsilon^2} \right) \right) ds \middle| \mathcal{F}_{t_1/\varepsilon^2}^Y \right], \end{aligned}$$

so by Lemma 3.4 of Chapter 4,

$$(2.36) \quad f \left(X^\varepsilon(t), Y \left(\frac{t}{\varepsilon^2} \right) \right) - \int_0^t \left(\Pi f + \frac{1}{\varepsilon} Af + \frac{1}{\varepsilon^2} Bf \right) \left(X^\varepsilon(s), Y \left(\frac{s}{\varepsilon^2} \right) \right) ds$$

is a martingale. It follows that $\{T(t)\}$ is strongly continuous on \mathcal{D} , hence on $\hat{\mathcal{C}}(\mathbb{R}^d \times E)$. We conclude therefore that the generator of $\{T_\epsilon(t)\}$ extends $\{(f, \Pi f + \epsilon^{-1} Af + \epsilon^{-2} Bf) : f \in \mathcal{D}\}$.

We define V on $\mathcal{D}(V) \equiv \{Af : f \in C_c^2(\mathbb{R}^d)\}$ by $(Vg)(x, y) = \int g(x, z)v(y, dz)$ and note that $V : \mathcal{D}(V) \rightarrow \mathcal{D}$ and

$$(2.37) \quad \begin{aligned} (BV Af)(x, y) &= B_0[H(x, \cdot) \cdot \nabla f(x)](y) \\ &= -G(x, y) \cdot \nabla f(x) = -Af(x, y) \end{aligned}$$

for all $f \in C_c^2(\mathbb{R}^d)$ and $(x, y) \in \mathbb{R}^d \times E$. It is immediate that C , defined by (2.14), has the form (2.26)–(2.28). Under the assumptions of the theorem, we infer from Proposition 2.2 that, for each $f \in \hat{\mathcal{C}}(\mathbb{R}^d)$, $T_\epsilon(t)f \rightarrow T(t)f$ as $\epsilon \rightarrow 0$ for all $t \geq 0$, uniformly on bounded intervals. By (2.33), $(X^\epsilon(\cdot), Y(\cdot/\epsilon^2))$ is a Markov process corresponding to $\{T_\epsilon(t)\}$ with sample paths in $D_{\mathbb{R}^d \times E}[0, \infty)$ and initial distribution $\delta_{x_0} \times \mu_0$ and therefore, by Corollary 8.7 of Chapter 4 and Problem 25 of Chapter 3, $X^\epsilon \Rightarrow X$ in $C_{\mathbb{R}^d}[0, \infty)$ as $\epsilon \rightarrow 0$. \square

2.6 Example Let E be finite, and define

$$(2.38) \quad B_0 g(i) = \sum_{j \in E} q_{ij} g(j),$$

where $Q = (q_{ij})_{i, j \in E}$ is an irreducible, infinitesimal matrix (i.e., $q_{ij} \geq 0$ for all $i \neq j$, $\sum_{j \in E} q_{ij} = 0$ for all $i \in E$, and there does not exist a nonempty, proper subset J of E such that $q_{ij} = 0$ for all $i \in J$ and $j \notin J$). Let $\mu = (\mu_i)_{i \in E}$ denote the unique stationary distribution. It is well known that

$$(2.39) \quad \lim_{t \rightarrow \infty} P(t, i, \{j\}) = \mu_j, \quad i, j \in E,$$

and (2.23) follows from this. By the existence of generalized inverses (Rao (1973), p. 25) and Lemma 7.3(d) of Chapter 1, there exists a real matrix $v = (v_{ij})_{i, j \in E}$ such that $Qv\lambda = -\lambda$ for all real column vectors $\lambda = (\lambda_i)_{i \in E}$ for which $\lambda \cdot \mu = 0$. It follows that the function H of Theorem 2.4 is given by

$$(2.40) \quad H(x, i) = \sum_{j \in E} v_{ij} G(x, j).$$

Alternatively, using the fact that the convergence in (2.39) is exponentially fast (Doob (1953), Theorem VI.1.1), Remark 2.5 gives (2.40) with

$$(2.41) \quad v_{ij} = \int_0^\infty (P(t, i, \{j\}) - \mu_j) dt.$$

This generalizes the example of Section 1. \square

2.7 Example Let $E = [0, 1]$, and define

$$(2.42) \quad B_0 = \{(g, \frac{1}{2}g'') : g \in C^2[0, 1], g'(0) = g'(1) = 0\}.$$

We claim that the Feller semigroup $\{S(t)\}$ on $C[0, 1]$ generated by B_0 (see Problem 6(a) in Chapter 1) satisfies

$$(2.43) \quad \lim_{t \rightarrow \infty} \sup_{0 \leq y \leq 1} \left| S(t)g(y) - \int_0^1 g(z) dz \right| = 0, \quad g \in C[0, 1].$$

This follows from the fact that $\{S(t)\}$ has the form

$$(2.44) \quad S(t)g(y) = \int_0^1 g(z) \tilde{p}(t, y, z) dz,$$

where

$$(2.45) \quad \tilde{p}(t, y, z) = \sum_{n=-\infty}^{\infty} p(t, y, 2n + z) + \sum_{n=-\infty}^{\infty} p(t, y, 2n - z)$$

and $p(t, y, z) = (2\pi t)^{-1/2} \exp \{-(z-y)^2/2t\}$, together with the crude inequality

$$(2.46) \quad \sup_{0 \leq z \leq 1} \tilde{p}(t, y, z) - \inf_{0 \leq z \leq 1} \tilde{p}(t, y, z) \leq \frac{2}{\sqrt{2\pi t}},$$

valid for $0 \leq y \leq 1$ and $t > 0$. In particular, μ is Lebesgue measure on $[0, 1]$. The function H of Theorem 2.4 can be defined by

$$(2.47) \quad H(x, y) = -2 \int_0^y \int_0^z G(x, w) dw dz.$$

Note that $H_i(x, \cdot) \in \mathcal{D}(B_0)$ for each $x \in \mathbb{R}^d$ and $i = 1, \dots, d$ since $\int_0^1 G(x, w) dw = 0$ for all $x \in \mathbb{R}^d$ by assumption. \square

3. DRIVING PROCESS IN A NONCOMPACT STATE SPACE

Let Y be an Ornstein–Uhlenbeck process, that is, a diffusion process in \mathbb{R} with generator

$$(3.1) \quad B_0 = \{(g, \mathcal{G}g): g \in \hat{C}(\mathbb{R}) \cap C^2(\mathbb{R}), \mathcal{G}g \in \hat{C}(\mathbb{R})\},$$

where $\mathcal{G}g(y) = g''(y) - yg'(y)$. The analogue of (1.18) is

$$(3.2) \quad X^\epsilon(t) = x + \frac{1}{\epsilon} \int_0^t Y^\epsilon(s) ds, \quad Y^\epsilon(t) = Y\left(\frac{t}{\epsilon^2}\right),$$

and one might ask whether the analogous conclusion holds. Even if Theorem 2.4 could be extended to the case of E locally compact (with $C(E)$ replaced by $\hat{C}(E)$), it would still be inadequate for at least three reasons. First, (2.22) is not satisfied. Second, convergence in (2.23) cannot be uniform if the right side is nonzero. Third, with $G(x, y) = y$, we have $H(x, y) = y$, which is not even bounded in $y \in \mathbb{R}$, much less an element of $\mathcal{D}(B_0)$. This last problem causes the

most difficulty. We may be able to find an operator V that formally satisfies $B_0 Vg = -g$ (e.g., if B_0 is given by a differential operator \mathcal{G} , we may be able to solve $\mathcal{G}h = -g$ for a large class of g), but $Vg \notin \mathcal{D}(B_0)$ for the functions g in which we are interested.

There are several ways to proceed. One is to prove an analogue of Proposition 2.2 with the role of $C_c(E)$ played by the space of (equivalence classes of) Borel measurable functions $f: E \rightarrow L$ with $\|f(\cdot)\| \in L^1(\mu)$, where μ is the stationary distribution of Y . However, this approach seems to require that Y have initial distribution μ .

Instead, we apply Corollary 8.7 (or 8.16) of Chapter 4, which was formulated with problems such as this in mind. The basic idea in the theorem is to "cut off" unbounded Vg by multiplying by a function $\phi_\varepsilon \in C_c(E)$, with $\phi_\varepsilon = 1$ on a large compact set, selected so that $\phi_\varepsilon Vg \in \mathcal{D}(B_0)$ and $B_0(\phi_\varepsilon Vg)$ is approximately $-g$. We show, in the case of (3.2), that $X^\varepsilon \Rightarrow x + \sqrt{2W}$ in $C_{\mathbb{R}}[0, \infty)$ as $\varepsilon \rightarrow 0+$.

3.1 Theorem Let E be a locally compact, separable metric space, let $F, G \in C_{\mathbb{R}}(\mathbb{R}^d \times E)$, and suppose that $F_1, \dots, F_d \in C^{1,0}(\mathbb{R}^d \times E)$, that $G_1, \dots, G_d \in C^{2,0}(\mathbb{R}^d \times E)$, and that

$$(3.3) \quad \sup_{(x, y) \in \mathbb{R}^d \times K} \frac{|F(x, y)| \vee |G(x, y)|}{1 + |x|} < \infty$$

for every compact set $K \subset E$. Let $\{S(t)\}$ be a Feller semigroup on $\hat{C}(E)$ with generator B_0 and let $\mu \in \mathcal{P}(E)$. Let $\rho: E \rightarrow (0, \infty)$ satisfy $1/\rho \in \hat{C}(E)$, let $\phi \in C_c^2[0, \infty)$ satisfy $\chi_{[0,1]} \leq \phi \leq \chi_{[0,2]}$, fix $0 < \theta < 1$, and define $\phi_\varepsilon \in C_c(E)$ and $K_\varepsilon \subset E$ by

$$(3.4) \quad \phi_\varepsilon(y) = \phi(\varepsilon\rho(y)) \quad \text{and} \quad K_\varepsilon = \{y \in E: \varepsilon^\theta \rho(y) \leq 1\}.$$

Assume that $\phi_\varepsilon \in \mathcal{D}(B_0)$ for each $\varepsilon \in (0, 1)$ and

$$(3.5) \quad \sup_{y \in K_\varepsilon} |B_0 \phi_\varepsilon(y)| = o(\varepsilon^2) \quad \text{as} \quad \varepsilon \rightarrow 0.$$

Define

$$(3.6) \quad \mathcal{M} = \left\{ g \in C(\mathbb{R}^d \times E): \int \sup_{|x| \leq l} |g(x, y)| \mu(dy) < \infty \quad \text{for } l = 1, 2, \dots \right\},$$

and let V be a linear operator on \mathcal{M} with $\mathcal{D}(V) \subset \{g \in \mathcal{M}: \int g(x, y) \mu(dy) = 0 \text{ for all } x \in \mathbb{R}^d\}$ such that if $g \in \mathcal{D}(V)$, then $(Vg)(x, \cdot) \phi_\varepsilon(\cdot) \in \mathcal{D}(B_0)$ for every $x \in \mathbb{R}^d$ and $0 < \varepsilon < 1$ and

$$(3.7) \quad \sup_{|x| \leq l, y \in K_\varepsilon} |B_0[(Vg)(x, \cdot) \phi_\varepsilon(\cdot)](y) + g(x, y)| = o(1) \quad \text{as} \quad \varepsilon \rightarrow 0$$

for $l = 1, 2, \dots$. Assume that $f \in C(\mathbb{R}^d)$ and $g \in \mathcal{D}(V)$ imply $fg \in \mathcal{D}(V)$ and $V(fg) = fVg$.

The following assumptions and definitions are made for $i, j, k = 1, \dots, d$. Suppose $G_i \in \mathcal{D}(V)$ and the left side of (3.7) with $g = G_i$ is $o(\varepsilon)$ as $\varepsilon \rightarrow 0$ for $i = 1, 2, \dots$. Assume $H_i \equiv VG_i \in C^{2,0}(\mathbb{R}^d \times E)$ and $F_i, G_i H_j, G_i \partial H_j / \partial x_i \in \mathcal{M}$. Suppose a_{ij} and b_i , defined by (2.27) and (2.28), belong to $C^1(\mathbb{R}^d)$. Suppose $G_i H_j + G_j H_i - a_{ij} \in \mathcal{D}(V)$, $F_i + G \cdot \nabla_x H_i - b_i \in \mathcal{D}(V)$, $\hat{a}_{ij} \equiv V(G_i H_j + G_j H_i - a_{ij}) \in C^{1,0}(\mathbb{R}^d \times E)$, and $\hat{b}_i \equiv V(F_i + G \cdot \nabla_x H_i - b_i) \in C^{1,0}(\mathbb{R}^d \times E)$. Assume that $H_i, F_i H_j, F_i \partial H_j / \partial x_i, G_i \hat{a}_{jk}, G_i \partial \hat{a}_{jk} / \partial x_i, G_i \hat{b}_j, G_i \partial \hat{b}_j / \partial x_i$, when multiplied by the function $(x, y) \mapsto \chi_{[0,1]}(|x|)/\rho(y)$, are bounded on $\mathbb{R}^d \times E$ for $i = 1, 2, \dots$. Assume further that $\hat{a}_{ij}, \hat{b}_i, F_i \hat{a}_{jk}, F_i \partial \hat{a}_{jk} / \partial x_i, F_i \hat{b}_j, F_i \partial \hat{b}_j / \partial x_i$, when multiplied by the function $(x, y) \mapsto \chi_{[0,1]}(|x|)/\rho(y)^2$, are bounded on $\mathbb{R}^d \times E$ for $i = 1, 2, \dots$.

Fix $\mu_0 \in \mathcal{P}(E)$, and let Y be a Markov process corresponding to $\{S(t)\}$ with sample paths in $D_E[0, \infty)$ and initial distribution μ_0 . Assume that

$$(3.8) \quad \lim_{\varepsilon \rightarrow 0} P \left\{ Y \left(\frac{t}{\varepsilon^2} \right) \in K_\varepsilon \text{ for } 0 \leq t \leq T \right\} = 1$$

for each $T > 0$. Fix $x_0 \in \mathbb{R}^d$ and define X^ε for $0 < \varepsilon < 1$ to be the solution of the differential equation (2.25) with initial condition $X^\varepsilon(0) = x_0$. Put

$$(3.9) \quad C = \left\{ \left(f, \frac{1}{2} \sum_{i,j=1}^d a_{ij} \partial_i \partial_j f + \sum_{i=1}^d b_i \partial_i f \right) : f \in C_c^3(\mathbb{R}^d) \right\}.$$

Then C is dissipative. Assume that \bar{C} , which is single-valued, generates a Feller semigroup $\{\bar{T}(t)\}$ on $\bar{C}(\mathbb{R}^d)$, and let X be a Markov process corresponding to $\{\bar{T}(t)\}$ with sample paths in $C_{\mathbb{R}^d}[0, \infty)$ and initial distribution δ_{x_0} . Then $X^\varepsilon \Rightarrow X$ in $C_{\mathbb{R}^d}[0, \infty)$ as $\varepsilon \rightarrow 0$.

3.2 Remark Instead of assuming an ergodicity condition such as (2.29), which would be rather difficult to exploit here (and may be rather difficult to verify), we assume the existence of a linear operator V such that (essentially) $B_0 V = -I$. \square

Proof. For each $\varepsilon \in (0, 1)$, exactly the same argument as used in the proof of Theorem 2.4 shows that $\{(X^\varepsilon(t), Y(t/\varepsilon^2)), t \geq 0\}$ is a progressive Markov process in $\mathbb{R}^d \times E$ corresponding to a measurable contraction semigroup with full generator that extends $\{(f, A_\varepsilon f) : f \in \mathcal{D}\}$, where

$$(3.10) \quad \mathcal{D} = \{f \in C_c^{1,0}(\mathbb{R}^d \times E) : f(x, \cdot) \in \mathcal{D}(B_0) \text{ for all } x \in \mathbb{R}^d\}$$

and

$$(3.11) \quad A_\varepsilon f(x, y) = \{F(x, y) + \varepsilon^{-1}G(x, y)\} \cdot \nabla_x f(x, y) + \varepsilon^{-2}B_0[f(x, \cdot)](y).$$

(Note that if $f \in \mathcal{D}$, the function $(x, y) \mapsto B_0[f(x, \cdot)](y)$ is automatically jointly measurable.) Let $(f, g) \in C$ and define

$$(3.12) \quad h_1 = V(G \cdot \nabla f) = H \cdot \nabla f$$

and

$$(3.13) \quad h_2 = V(F \cdot \nabla f + G \cdot \nabla_x h_1 - g)$$

$$= \frac{1}{2} \sum_{i,j=1}^d \hat{a}_{ij} \partial_i \partial_j f + \sum_{i=1}^d b_i \partial_i f.$$

For each $\varepsilon \in (0, 1)$, define $f_\varepsilon \in \mathcal{D}$ by

$$(3.14) \quad f_\varepsilon(x, y) = (f(x) + \varepsilon h_1(x, y) + \varepsilon^2 h_2(x, y))\phi_\varepsilon(y),$$

and observe that

$$\begin{aligned} (3.15) \quad A_\varepsilon f_\varepsilon(x, y) &= \varepsilon^{-2} f(x) B_0 \phi_\varepsilon(y) \\ &\quad + \varepsilon^{-1} \{G(x, y) \cdot \nabla f(x)\phi_\varepsilon(y) + B_0[h_1(x, \cdot)\phi_\varepsilon(\cdot)](y)\} \\ &\quad + \{F(x, y) \cdot \nabla f(x) + G(x, y) \cdot \nabla_x h_1(x, y)\}\phi_\varepsilon(y) \\ &\quad + B_0[h_2(x, \cdot)\phi_\varepsilon(\cdot)](y) \\ &\quad + \varepsilon \{F(x, y) \cdot \nabla_x h_1(x, y) + G(x, y) \cdot \nabla_x h_2(x, y)\}\phi_\varepsilon(y) \\ &\quad + \varepsilon^2 F(x, y) \cdot \nabla_x h_2(x, y)\phi_\varepsilon(y) \end{aligned}$$

for all $(x, y) \in \mathbb{R}^d \times E$. By (3.5) and the other assumptions,

$$(3.16) \quad \sup_{0 < \varepsilon < 1} \sup_{(x, y) \in \mathbb{R}^d \times E} |f_\varepsilon(x, y)| < \infty,$$

$$(3.17) \quad \lim_{\varepsilon \rightarrow 0} \sup_{(x, y) \in \mathbb{R}^d \times K_\varepsilon} |f_\varepsilon(x, y) - f(x)| = 0,$$

$$(3.18) \quad \lim_{\varepsilon \rightarrow 0} \sup_{(x, y) \in \mathbb{R}^d \times K_\varepsilon} |A_\varepsilon f_\varepsilon(x, y) - g(x)| = 0.$$

In view of (3.8), the result follows from Corollary 8.7 of Chapter 4. \square

3.3 Example Let $E = \mathbb{R}$ and define B_0 by (3.1), where $\mathcal{G}g(y) = g''(y) - yg'(y)$. It is quite easy to show that the Feller semigroup $\{S(t)\}$ on $\hat{C}(\mathbb{R})$ generated by B_0 has a unique stationary distribution μ , that μ is $N(0, 1)$, the standard normal distribution, and that

$$(3.19) \quad \text{bp-lim}_{t \rightarrow \infty} \left| \frac{1}{t} \int_0^t S(s)g \, ds - \int g(z)\mu(dz) \right| = 0, \quad g \in \hat{C}(\mathbb{R}).$$

However, these results are not explicitly needed.

For each $n \geq 1$, define $\phi_n: \mathbb{R} \rightarrow (0, \infty)$ by $\phi_n(y) = (1 + y^2)^{n/2}$,

$$(3.20) \quad \mathcal{M}_n = \left\{ g \in C(\mathbb{R}^d \times \mathbb{R}): \sup_{|x| \leq l, y \in \mathbb{R}} \frac{|g(x, y)|}{\phi_n(y)} < \infty \quad \text{for } l = 1, 2, \dots \right\},$$

and $\mathcal{M}_\infty = \bigcup_{n=1}^\infty \mathcal{M}_n$. Define V on

$$(3.21) \quad \mathcal{D}(V) = \left\{ g \in \mathcal{M}_\infty: \int_{-\infty}^\infty g(x, y)e^{-y^2/2} \, dy = 0 \quad \text{for all } x \in \mathbb{R}^d \right\}$$

by

$$(3.22) \quad Vg(x, y) = \int_0^y e^{z^2/2} \int_z^\infty g(x, w) e^{-w^2/2} dw dz,$$

and note that $V: \mathcal{D}(V) \cap \mathcal{M}_n \rightarrow \mathcal{M}_n$ for each $n \geq 1$. Also, if $g \in \mathcal{D}(V) \cap C^{1,0}(\mathbb{R}^d \times \mathbb{R})$ and $|\nabla_x g| \in \mathcal{M}_\infty$, then $Vg \in C^{1,0}(\mathbb{R}^d \times \mathbb{R})$ and, for $i = 1, \dots, d$, $\partial g/\partial x_i \in \mathcal{D}(V)$ and $\partial(Vg)/\partial x_i = V(\partial g/\partial x_i)$.

Fix $m \geq 1$ and let $\rho = \phi_{3m}$ and $\theta = \frac{1}{2}$. Observe that V satisfies the required conditions (in fact (3.7) is zero). Assume, in addition to the assumptions on F and G in the first sentence of the theorem, that $G_i \in \mathcal{D}(V) \cap \mathcal{M}_m$ and $F_i, \partial F_i/\partial x_j, G_i, \partial G_i/\partial x_j, \partial^2 G_i/\partial x_j \partial x_k \in \mathcal{M}_m$ for $i, j, k = 1, \dots, d$. If \bar{C} satisfies the condition of the theorem, the only condition that remains to be verified is (3.8). For this it suffices to show that

$$(3.23) \quad \lim_{\varepsilon \rightarrow 0} \sup_{0 \leq s \leq t} \varepsilon^{1/2} \left(1 + Y \left(\frac{s}{\varepsilon^2} \right)^2 \right)^{3m/2} = 0 \quad \text{a.s.,} \quad t \geq 0.$$

For the latter it is enough to show that for each $\lambda > 0$ there exists a random variable η such that

$$(3.24) \quad P\{|Y(t)| \leq \eta + t^\lambda \text{ for all } t \geq 0\} = 1.$$

To verify (3.24), we need only show that $\lim_{t \rightarrow \infty} |Y(t)|/t^\lambda = 0$ a.s. for every $\lambda > 0$, which follows from the representation

$$(3.25) \quad Y(t) = e^{-t} Y(0) + e^{-t} W(e^{2t} - 1),$$

where W is a standard one-dimensional Brownian motion, and the law of the iterated logarithm for W . \square

4. NON-MARKOVIAN DRIVING PROCESS

We again consider the limit in distribution as $\varepsilon \rightarrow 0+$ of the solution X^ε of the differential equation

$$(4.1) \quad \frac{d}{dt} X^\varepsilon(t) = F\left(X^\varepsilon(t), Y\left(\frac{t}{\varepsilon^2}\right)\right) + \frac{1}{\varepsilon} G\left(X^\varepsilon(t), Y\left(\frac{t}{\varepsilon^2}\right)\right)$$

driven by $Y(\cdot/\varepsilon^2)$, where Y is a process in a compact state space. However, instead of assuming that Y is Markovian and ergodic as in Section 2, we require that Y be stationary and uniform mixing.

4.1 Theorem Let E be a compact metric space, let $F, G: \mathbb{R}^d \times E \rightarrow \mathbb{R}^d$, and suppose that $F_1, \dots, F_d \in C^{1,0}(\mathbb{R}^d \times E)$, $G_1, \dots, G_d \in C^{2,0}(\mathbb{R}^d \times E)$, and

$$(4.2) \quad \sup_{(x,y) \in \mathbb{R}^d \times E} \frac{|F(x,y)| \vee |G(x,y)|}{1 + |x|} < \infty.$$

Let Y be a stationary process with sample paths in $D_E[0, \infty)$, and for each $t \geq 0$, let \mathcal{F}_t and \mathcal{F}' denote the completions of the σ -algebras \mathcal{F}_t' and $\sigma\{Y(s): s \geq t\}$, respectively. Assume that the filtration $\{\mathcal{F}_t\}$ is right continuous, and that

$$(4.3) \quad \varphi(u) \equiv \sup_{t \geq 0} \sup_{A \in \mathcal{F}_t, B \in \mathcal{F}'^{t+u}} |P(B|A) - P(B)|$$

satisfies

$$(4.4) \quad \int_0^\infty u \varphi(u) du < \infty.$$

Suppose that

$$(4.5) \quad E[G(x, Y(0))] = 0, \quad x \in \mathbb{R}^d.$$

Fix $x_0 \in \mathbb{R}^d$, and define X^ε for $0 < \varepsilon < 1$ to be the solution of the differential equation (4.1) with initial condition $X^\varepsilon(0) = x_0$. Put

$$(4.6) \quad C = \left\{ \left(f, \frac{1}{2} \sum_{i,j=1}^d a_{ij} \partial_i \partial_j f + \sum_{i=1}^d b_i \partial_i f \right) : f \in C_c^3(\mathbb{R}^d) \right\},$$

where

$$(4.7) \quad a_{ij}(x) = \int_0^\infty E[G_i(x, Y(0)) G_j(x, Y(t))] dt + \int_0^\infty E[G_i(x, Y(0)) \cdot \nabla_x G_j(x, Y(t))] dt$$

and

$$(4.8) \quad b_i(x) = E[F_i(x, Y(0))] + \int_0^\infty E[G(x, Y(0)) \cdot \nabla_x G_i(x, Y(t))] dt.$$

Then C is dissipative. Assume that \bar{C} , which is single-valued, generates a Feller semigroup $\{T(t)\}$ on $\bar{C}(\mathbb{R}^d)$, and let X be a Markov process corresponding to $\{T(t)\}$ with sample paths in $C_{\mathbb{R}^d}[0, \infty)$ and initial distribution δ_{x_0} . Then $X^\varepsilon \Rightarrow X$ in $C_{\mathbb{R}^d}[0, \infty)$ as $\varepsilon \rightarrow 0$.

Proof. Let $t, u \geq 0$ and let X be essentially bounded and \mathcal{F}'^{t+u} -measurable. Then by Proposition 2.6 of Chapter 7 ($r = 1, p = \infty$),

$$(4.9) \quad \|E[X|\mathcal{F}_t] - E[X]\|_\infty \leq 2\varphi(u)\|X\|_\infty,$$

where $\varphi(u)$ is defined by (4.3). For example, conditioning on \mathcal{F}_0 and using (4.5), we find that

$$(4.10) \quad |E[G_i(x, Y(0))G_j(x, Y(t))]| \leq \sup_y |G_i(x, y)| 2\varphi(t) \sup_y |G_j(x, y)|$$

for all $x \in \mathbb{R}^d$, $t \geq 0$, and $i, j = 1, \dots, d$. The same inequality holds when G_i and/or G_j are replaced by any of their first- or second-order partial x -derivatives, and therefore the coefficients (4.7) and (4.8) are continuously differentiable on \mathbb{R}^d .

We also observe that the diffusion matrix $(a_{ij}(x))$ is nonnegative definite for each $x \in \mathbb{R}^d$. For if $x, \xi \in \mathbb{R}^d$ and $T > 0$,

$$\begin{aligned} (4.11) \quad & \frac{1}{T} E \left[\left(\xi \cdot \int_0^T G(x, Y(t)) dt \right)^2 \right] \\ &= \sum_{i, j=1}^d \xi_i \xi_j \frac{1}{T} E \int_0^T \int_0^t [G_i(x, Y(s))G_j(x, Y(t)) \\ &\quad + G_j(x, Y(s))G_i(x, Y(t))] ds dt \\ &= \sum_{i, j=1}^d \xi_i \xi_j \frac{1}{T} \int_0^T \int_0^t E[G_i(x, Y(0))G_j(x, Y(t-s)) \\ &\quad + G_j(x, Y(0))G_i(x, Y(t-s))] ds dt \\ &= \sum_{i, j=1}^d \xi_i \xi_j \frac{1}{T} \int_0^T \left\{ E \left[G_i(x, Y(0)) \int_0^t G_j(x, Y(s)) ds \right] \right. \\ &\quad \left. + E \left[G_j(x, Y(0)) \int_0^t G_i(x, Y(s)) ds \right] \right\} dt. \end{aligned}$$

As $T \rightarrow \infty$, (4.11), which is nonnegative, converges to

$$(4.12) \quad \sum_{i, j=1}^d \xi_i \xi_j a_{ij}(x).$$

Thus, C satisfies the positive maximum principle, hence C is dissipative (Lemma 2.1 of Chapter 4) and \bar{C} is single-valued (Lemma 4.2 of Chapter 1).

The growth condition (4.2) guarantees the global existence of the solution X^t of (4.1). Denote by $\{\mathcal{F}_t^t\}$ the filtration given by $\mathcal{F}_t^t = \mathcal{F}_{t/t^2}$, and let $\hat{\mathcal{A}}^t$ be the full generator of the associated semigroup of conditioned shifts (Chapter 2, Section 7). By Theorem 8.2 of Chapter 4, the finite-dimensional distributions

of X^ε will converge weakly to those of X if for each $(f, g) \in C$, we can find $(f^\varepsilon, g^\varepsilon) \in \hat{\mathcal{A}}^\varepsilon$ for every $\varepsilon \in (0, 1)$ such that

$$(4.13) \quad \sup_{\varepsilon} \sup_{t \leq T} E[|f^\varepsilon(t)|] < \infty, \quad T > 0,$$

$$(4.14) \quad \sup_{\varepsilon} \sup_{t \leq T} E[|g^\varepsilon(t)|] < \infty, \quad T > 0,$$

$$(4.15) \quad \lim_{\varepsilon \rightarrow 0} E[|f^\varepsilon(t) - f(X^\varepsilon(t))|] = 0, \quad t \geq 0,$$

and

$$(4.16) \quad \lim_{\varepsilon \rightarrow 0} E[|g^\varepsilon(t) - g(X^\varepsilon(t))|] = 0, \quad t \geq 0.$$

By Corollary 8.6 of Chapter 4 and Problem 25 of Chapter 3, we have $X^\varepsilon \Rightarrow X$ in $C_{\mathbb{R}^d}[0, \infty)$ as $\varepsilon \rightarrow 0$ if (4.14) and (4.15) can be replaced by the stronger conditions

$$(4.17) \quad \sup_{\varepsilon} E \left[\text{ess sup}_{t \leq T} |g^\varepsilon(t)| \right] < \infty, \quad T > 0,$$

and

$$(4.18) \quad \lim_{\varepsilon \rightarrow 0} E \left[\sup_{t \in \mathbb{Q} \cap [0, T]} |f^\varepsilon(t) - f(X^\varepsilon(t))| \right] = 0, \quad T > 0.$$

Fix $(f, g) \in C$, and let $\varepsilon \in (0, 1)$ be arbitrary. We let

$$(4.19) \quad f^\varepsilon(t) = f(X^\varepsilon(t)) + \varepsilon \hat{h}_1^\varepsilon(t) + \varepsilon^2 \hat{h}_2^\varepsilon(t),$$

where the correction terms $\hat{h}_1^\varepsilon, \hat{h}_2^\varepsilon \in \mathcal{D}(\hat{\mathcal{A}}^\varepsilon)$ are chosen by analogy with (3.12) and (3.13). Let us first consider \hat{h}_1^ε . We define $f_1^\varepsilon: \mathbb{R}^d \times [0, \infty) \times \Omega \rightarrow \mathbb{R}$ by

$$(4.20) \quad f_1^\varepsilon(x, t, \omega) = G \left(x, Y \left(\frac{t}{\varepsilon^2}, \omega \right) \right) \cdot \nabla f(x).$$

Clearly, f_1^ε is $\mathcal{B}(\mathbb{R}^d) \times \mathcal{B}[0, \infty) \times \mathcal{F}$ -measurable and is C_c^2 in x for fixed (t, ω) . In fact, there is a constant k_1 such that $f_1^\varepsilon(x, t, \omega) = 0$ for all $|x| \geq k_1$, $t \geq 0$, and $\omega \in \Omega$, and

$$(4.21) \quad \|f_1^\varepsilon(\cdot, t, \omega)\|_{C^2} \leq \sup_y \|G(\cdot, y) \cdot \nabla f(\cdot)\|_{C^2} \equiv \gamma < \infty.$$

for all $t \geq 0$ and $\omega \in \Omega$, where $\|f\|_{C^m} \equiv \sum_{|\alpha| \leq m} \|D^\alpha f\|$. By Corollary 4.5 of Chapter 2 there exists $g_1^\epsilon: \mathbb{R}^d \times [0, \infty) \times [0, \infty) \times \Omega \rightarrow \mathbb{R}$, $\mathcal{B}(\mathbb{R}^d) \times \mathcal{B}([0, \infty)) \times \mathcal{O}$ -measurable, C_c^2 in x for fixed (s, t, ω) , such that

$$(4.22) \quad g_1^\epsilon(x, s, t, \omega) = E_t^\epsilon[f_1^\epsilon(x, t + s, \cdot)](\omega)$$

for all $x \in \mathbb{R}^d$ and $s, t \geq 0$, where E_t^ϵ denotes conditional expectation given \mathcal{F}_t^ϵ here and below. Moreover, g_1^ϵ may be chosen so that $g_1^\epsilon(x, s, t, \omega) = 0$ for all $|x| \geq k_1$, $s, t \geq 0$, and $\omega \in \Omega$, and

$$(4.23) \quad \|g_1^\epsilon(\cdot, s, t, \omega)\|_{C^2} \leq 2\gamma\phi\left(\frac{s}{\epsilon^2}\right)$$

for all $s, t \geq 0$ and $\omega \in \Omega$. The latter can be deduced from (4.5), (4.9), and (4.21). We now define $h_1^\epsilon: \mathbb{R}^d \times [0, \infty) \times \Omega \rightarrow \mathbb{R}$ by

$$(4.24) \quad h_1^\epsilon(x, t, \omega) = \epsilon^{-2} \int_0^\infty g_1^\epsilon(x, s, t, \omega) ds.$$

Clearly, h_1^ϵ is $\mathcal{B}(\mathbb{R}^d) \times \mathcal{O}$ -measurable and is C_c^2 in x for fixed (t, ω) . In fact, $h_1^\epsilon(x, t, \omega) = 0$ for all $|x| \geq k_1$, $t \geq 0$, and $\omega \in \Omega$, and

$$(4.25) \quad \|h_1^\epsilon(\cdot, t, \omega)\|_{C^2} \leq 2\gamma \int_0^\infty \varphi(s) ds$$

for all $t \geq 0$ and $\omega \in \Omega$. Finally, we define $\hat{h}_1^\epsilon: [0, \infty) \times \Omega \rightarrow \mathbb{R}$ by

$$(4.26) \quad \hat{h}_1^\epsilon(t, \omega) = h_1^\epsilon(X^\epsilon(t, \omega), t, \omega).$$

It follows that \hat{h}_1^ϵ is optional (hence progressive).

To show that $\hat{h}_1^\epsilon \in \mathcal{D}(\hat{\mathcal{A}}^\epsilon)$, we apply Lemma 3.4 of Chapter 4. Fix $t_2 > t_1 \geq 0$. Clearly,

$$\begin{aligned} (4.27) \quad & h_1^\epsilon(X^\epsilon(t_2), t_2) - h_1^\epsilon(X^\epsilon(t_1), t_1) \\ &= \int_{t_1}^{t_2} \nabla_x h_1^\epsilon(X^\epsilon(s), t_2) \cdot \frac{d}{ds} X^\epsilon(s) ds \\ &= \int_{t_1}^{t_2} \left\{ F\left(X^\epsilon(s), Y\left(\frac{s}{\epsilon^2}\right)\right) + \frac{1}{\epsilon} G\left(X^\epsilon(s), Y\left(\frac{s}{\epsilon^2}\right)\right) \right\} \\ &\quad \cdot \nabla_x h_1^\epsilon(X^\epsilon(s), t_2) ds. \end{aligned}$$

For each $x \in \mathbb{R}^d$ and $s \geq 0$, we have

$$\begin{aligned} (4.28) \quad E_{t_1}^\epsilon[g_1^\epsilon(x, s, t_2, \cdot)](\omega) &= E_{t_1}^\epsilon[E_{t_2}^\epsilon[f_1^\epsilon(x, t_2 + s, \cdot)]](\omega) \\ &= E_{t_1}^\epsilon[f_1^\epsilon(x, t_2 + s, \cdot)](\omega) \\ &= g_1^\epsilon(x, s + t_2 - t_1, t_1, \omega), \end{aligned}$$

and therefore

$$\begin{aligned}
 (4.29) \quad & E_{t_1}^{\epsilon}[h_1^{\epsilon}(X^{\epsilon}(t_1), t_2) - h_1^{\epsilon}(X^{\epsilon}(t_1), t_1)] \\
 &= \epsilon^{-2} \left\{ E_{t_1}^{\epsilon} \left[\int_0^{\infty} g_1^{\epsilon}(X^{\epsilon}(t_1), s, t_2) ds \right] - \int_0^{\infty} g_1^{\epsilon}(X^{\epsilon}(t_1), s, t_1) ds \right\} \\
 &= \epsilon^{-2} \left\{ \int_0^{\infty} g_1^{\epsilon}(X^{\epsilon}(t_1), s + t_2 - t_1, t_1) ds - \int_0^{\infty} g_1^{\epsilon}(X^{\epsilon}(t_1), s, t_1) ds \right\} \\
 &= -\epsilon^{-2} \int_0^{t_2 - t_1} g_1^{\epsilon}(X^{\epsilon}(t_1), s, t_1) ds \\
 &= -\epsilon^{-2} E_{t_1}^{\epsilon} \left[\int_{t_1}^{t_2} f_1^{\epsilon}(X^{\epsilon}(t_1), s) ds \right].
 \end{aligned}$$

Finally, we must verify condition (3.15) of Chapter 5, which amounts to showing that, for each $t \geq 0$,

$$(4.30) \quad \lim_{\delta \rightarrow 0^+} E[|\nabla_x h_1^{\epsilon}(X^{\epsilon}(t), t + \delta) - \nabla_x h_1^{\epsilon}(X^{\epsilon}(t), t)|] = 0.$$

(We can ignore the factor $F + \epsilon^{-1}G$ because $\nabla_x h_1^{\epsilon}(x, t, \omega)$ has compact support in x , uniformly in (t, ω) .) Using the bound (4.23), the dominated convergence theorem reduces the problem to one of showing that, for each $s, t \geq 0$,

$$(4.31) \quad \lim_{\delta \rightarrow 0^+} E[|\nabla_x g_1^{\epsilon}(X^{\epsilon}(t), s, t + \delta, \cdot) - \nabla_x g_1^{\epsilon}(X^{\epsilon}(t), s, t, \cdot)|] = 0$$

or

$$(4.32) \quad \lim_{\delta \rightarrow 0^+} E[|E_{t+\delta}^{\epsilon}[\nabla_x f_1^{\epsilon}(X^{\epsilon}(t), t + \delta + s)] - E_t^{\epsilon}[\nabla_x f_1^{\epsilon}(X^{\epsilon}(t), t + s)]|] = 0.$$

But (4.32) follows easily from the right continuity of Y and the right continuity of the filtration $\{\mathcal{F}_t^{\epsilon}\}$. We conclude from Lemma 3.4 of Chapter 4 that

$$(4.33) \quad (\hat{h}_1^{\epsilon}(t), \{F(x, y) + \epsilon^{-1}G(x, y)\} \cdot \nabla_x h_1^{\epsilon}(x, t) - \epsilon^{-2}G(x, y) \cdot \nabla f(x)) \in \hat{\mathcal{A}}^{\epsilon},$$

where $x = X^{\epsilon}(t)$ and $y = Y(t/\epsilon^2)$.

We turn now to the definition of $\hat{h}_2^{\epsilon}(t)$. We define $f_2^{\epsilon}: \mathbb{R}^d \times [0, \infty) \times \Omega \rightarrow \mathbb{R}$ by

$$\begin{aligned}
 (4.34) \quad f_2^{\epsilon}(x, t, \omega) &= F\left(x, Y\left(\frac{t}{\epsilon^2}, \omega\right)\right) \cdot \nabla f(x) + G\left(x, Y\left(\frac{t}{\epsilon^2}, \omega\right)\right) \\
 &\quad \cdot \nabla_x h_1^{\epsilon}(x, t, \omega) - g(x).
 \end{aligned}$$

Observe that f_2^{ϵ} is $\mathcal{B}(\mathbb{R}^d) \times \mathcal{O}$ -measurable and is C_c^1 in x for fixed (t, ω) . In fact, there is a constant k_2 such that $f_2(x, t, \omega) = 0$ for all $|x| \geq k_2$, $t \geq 0$, and $\omega \in \Omega$, and

$$(4.35) \quad \|f_2^\varepsilon(\cdot, t, \omega)\|_{C^1} \leq \sup_y \|F(\cdot, y) \cdot \nabla f(x)\|_{C^1} \\ + \sup_{y, t, \omega} \|G(\cdot, y) \cdot \nabla_x h_1^\varepsilon(\cdot, t, \omega)\|_{C^1} + \|g\|_{C^1} \\ \equiv \eta < \infty$$

by (4.25). We now define g_2^ε , h_2^ε , and \tilde{h}_2^ε by analogy with g_1^ε , h_1^ε , and \tilde{h}_1^ε . The only thing that needs to be checked is the analogue of (4.23), which is that, for appropriate constants $c_1, c_2 > 0$,

$$(4.36) \quad \|g_2^\varepsilon(\cdot, s, t, \omega)\|_{C^1} \leq c_1 \varphi\left(\frac{s}{\varepsilon^2}\right) + c_2 \int_0^\infty \varphi\left(\left(\frac{s}{\varepsilon^2}\right) \vee s'\right) ds'$$

for all $s, t \geq 0$ and $\omega \in \Omega$. Observe that the right side of (4.36) is Lebesgue integrable on $[0, \infty)$ by (4.4).

To justify (4.36), fix $x \in \mathbb{R}^d$ and $s, t \geq 0$. Then

$$(4.37) \quad g_2^\varepsilon(x, s, t) = E_t^\varepsilon[f_2^\varepsilon(x, t + s)] \\ = E_t^\varepsilon \left[F\left(x, Y\left(\frac{t+s}{\varepsilon^2}\right)\right) \cdot \nabla f(x) \right] \\ - E \left[F\left(x, Y\left(\frac{t+s}{\varepsilon^2}\right)\right) \cdot \nabla f(x) \right] \\ + E_t^\varepsilon \left[G\left(x, Y\left(\frac{t+s}{\varepsilon^2}\right)\right) \cdot \nabla_x h_1^\varepsilon(x, t + s) \right] \\ - E \left[G\left(x, Y\left(\frac{t+s}{\varepsilon^2}\right)\right) \cdot \nabla_x h_1^\varepsilon(x, t + s) \right]$$

by the definition of C . Consequently, a similar equation holds for $\nabla_x g_2^\varepsilon(x, s, t, \cdot)$ with each integrand replaced by its x -gradient. By (4.9),

$$(4.38) \quad \left\| E_t^\varepsilon \left[F\left(x, Y\left(\frac{t+s}{\varepsilon^2}\right)\right) \cdot \nabla f(x) \right] - E \left[F\left(x, Y\left(\frac{t+s}{\varepsilon^2}\right)\right) \cdot \nabla f(x) \right] \right\|_\infty \\ \leq 2\varphi\left(\frac{s}{\varepsilon^2}\right) \sup_y |F(x, y) \cdot \nabla f(x)|.$$

Since $\nabla_x h_1^\varepsilon(x, t + s) = \varepsilon^{-2} \int_0^\infty \nabla_x g_1^\varepsilon(x, s', t + s) ds'$, we need to consider

$$(4.39) \quad \left\| E_t^\varepsilon \left[G\left(x, Y\left(\frac{t+s}{\varepsilon^2}\right)\right) \cdot \nabla_x \left\{ G\left(x, Y\left(\frac{t+s+s'}{\varepsilon^2}\right)\right) \cdot \nabla f(x) \right\} \right] \right. \\ \left. - E \left[G\left(x, Y\left(\frac{t+s}{\varepsilon^2}\right)\right) \cdot \nabla_x \left\{ G\left(x, Y\left(\frac{t+s+s'}{\varepsilon^2}\right)\right) \cdot \nabla f(x) \right\} \right] \right\|_\infty$$

for fixed $s' \geq 0$. By (4.9), this is bounded by

$$(4.40) \quad 2\varphi\left(\frac{s}{\varepsilon^2}\right) \sup_{y, z} |G(x, y) \cdot \nabla_x\{G(x, z) \cdot \nabla f(x)\}|.$$

Moreover, conditioning on \mathcal{F}_{t+s}^t and applying (4.9) and (4.5), each of the two expectations in (4.39) is bounded a.s. by

$$(4.41) \quad \sup_y |G(x, y)| 2\varphi\left(\frac{s'}{\varepsilon^2}\right) \sup_y |\nabla_x\{G(x, y) \cdot \nabla f(x)\}|.$$

Thus, (4.39) is bounded by $c_3 \varphi((s \vee s')/\varepsilon^2)$ for an appropriate constant c_3 . Similar bounds hold when all integrands are replaced by their x -gradients, and thus (4.36) can be assumed to hold. It follows from (4.36) that

$$(4.42) \quad \|h_2^\varepsilon(\cdot, t, \omega)\|_{C^1} \leq c_1 \int_0^\infty \varphi(s) ds + 2c_2 \int_0^\infty s\varphi(s) ds$$

for all $t \geq 0$ and $\omega \in \Omega$.

The argument used to show that $h_1^\varepsilon \in \mathcal{D}(\hat{\mathcal{A}}^\varepsilon)$ now applies, almost word-for-word, to show that

$$(4.43) \quad \begin{aligned} (\hat{h}_2^\varepsilon(t), \{F(x, y) + \varepsilon^{-1}G(x, y)\} \cdot \nabla_x h_2^\varepsilon(x, t) \\ - \varepsilon^{-2}\{F(x, y) \cdot \nabla f(x) + G(x, y) \cdot \nabla_x h_1^\varepsilon(x, t) - g(x)\}) \in \hat{\mathcal{A}}^\varepsilon, \end{aligned}$$

where $x = X^\varepsilon(t)$ and $y = Y(t/\varepsilon^2)$. The only point that should be made is that, in proving the analogue of (4.32), $\nabla_x f_2^\varepsilon(X^\varepsilon(t), t + \delta + s)$ no longer converges pointwise in ω , but only in L^1 . However, this suffices.

Clearly,

$$(4.44) \quad (f(x), \{F(x, y) + \varepsilon^{-1}G(x, y)\} \cdot \nabla f(x)) \in \hat{\mathcal{A}}^\varepsilon$$

where $x = X^\varepsilon(t)$ and $y = Y(t/\varepsilon^2)$. Recalling (4.19), we obtain from (4.44), (4.33), and (4.43) that

$$(4.45) \quad \begin{aligned} (f^\varepsilon(t), g(x) + \varepsilon F(x, y) \cdot \nabla_x h_1^\varepsilon(x, t) + \varepsilon G(x, y) \cdot \nabla_x h_2^\varepsilon(x, t) \\ + \varepsilon^2 F(x, y) \cdot \nabla_x h_2^\varepsilon(x, t)) \in \hat{\mathcal{A}}^\varepsilon, \end{aligned}$$

where $x = X^\varepsilon(t)$ and $y = Y(t/\varepsilon^2)$. By (4.25) and (4.42), together with the fact that $\nabla_x h_1^\varepsilon(x, t, \omega)$ and $\nabla_x h_2^\varepsilon(x, t, \omega)$ have compact support in x , uniformly in (t, ω) , we see that (4.13)–(4.18) are satisfied, and hence the proof is complete. \square

5. PROBLEMS

1. Formulate and prove a discrete-parameter analogue of Theorem 2.4. (Both X and Y are discrete-parameter processes, and the differential equation (2.25) is a difference equation.)
2. Give a simpler proof that $X^\varepsilon \Rightarrow x + \sqrt{2}W$ in (3.2) by using the representation

$$(5.1) \quad X^\varepsilon(t) = x + \varepsilon Y(0) + \sqrt{2\varepsilon} W\left(\frac{t}{\varepsilon^2}\right) - \varepsilon Y\left(\frac{t}{\varepsilon^2}\right), \quad t \geq 0,$$

where W is a one-dimensional Brownian motion.

3. Generalize Example 3.3 to the case in which $Y(t) \equiv UZ(t)$, where

$$(5.2) \quad dZ(t) = S \, dW(t) + NZ(t) \, dt$$

and U , S , and N are (constant) $d \times d$ matrices with the eigenvalues of N having negative real parts, and W is a d -dimensional Brownian motion.

4. Extend Theorem 4.1 to noncompact E . The extension should include the case in which Y is a stationary Gaussian process.

6. NOTES

Random evolutions, introduced by Griego and Hersh (1969), are surveyed by Hersh (1974) and Pinsky (1974).

The derivation of the telegrapher's equation in Section 1 is due to Goldstein (1951) and Kac (1956).

The results of Section 2 were motivated by work of Pinsky (1968), Griego and Hersh (1971), Hersh and Papanicolaou (1972), and Kurtz (1973).

Theorem 4.1 is due essentially to Kushner (1979) (see also Kushner (1984)), though the problem had earlier been treated by Stratonovich (1963, 1967), Khas'minskii (1966), Papanicolaou and Varadhan (1973), and Papanicolaou and Kohler (1974).

APPENDIXES

1. CONVERGENCE OF EXPECTATIONS

Recall that $X_n \xrightarrow{a.s.} X$ implies $X_n \xrightarrow{P} X$ implies $X_n \Rightarrow X$, so the following results, which are stated in terms of convergence in distribution, apply to the other types of convergence as well.

1.1 Proposition (Fatou's Lemma) Let $X_n \geq 0$, $n = 1, 2, \dots$, and $X_n \Rightarrow X$. Then

$$(1.1) \quad \liminf_{n \rightarrow \infty} E[X_n] \geq E[X].$$

Proof. For $M > 0$

$$(1.2) \quad \liminf_{n \rightarrow \infty} E[X_n] \geq \lim_{n \rightarrow \infty} E[X_n \wedge M] = E[X \wedge M],$$

where the equality holds by definition. Letting $M \rightarrow \infty$ we have (1.1). \square

1.2 Theorem (Dominated Convergence Theorem) Suppose

$$|X_n| \leq Y_n, \quad n = 1, 2, \dots, \quad X_n \Rightarrow X, \quad Y_n \Rightarrow Y$$

and $\lim_{n \rightarrow \infty} E[Y_n] = E[Y]$. Then

$$(1.3) \quad \lim_{n \rightarrow \infty} E[X_n] = E[X].$$

Proof. It is not necessarily the case that $(X_n, Y_n) \Rightarrow (X, Y)$. However, by Proposition 2.4 of Chapter 3, every subsequence of $\{(X_n, Y_n)\}$ has a further subsequence such that $(X_{n_k}, Y_{n_k}) \Rightarrow (\tilde{X}, \tilde{Y})$, where \tilde{X} and X have the same distribution, as do \tilde{Y} and Y . Consequently, $Y_{n_k} + X_{n_k} \Rightarrow \tilde{Y} + \tilde{X}$ and $Y_{n_k} - X_{n_k} \Rightarrow \tilde{Y} - \tilde{X}$, so by Fatou's lemma,

$$(1.4) \quad \liminf_{k \rightarrow \infty} (E[Y_{n_k}] + E[X_{n_k}]) \geq E[Y] + E[X]$$

and

$$(1.5) \quad \liminf_{k \rightarrow \infty} (E[Y_{n_k}] - E[X_{n_k}]) \geq E[Y] - E[X].$$

Therefore $\lim_{k \rightarrow \infty} E[X_{n_k}] = E[X]$, and (1.3) follows. \square

2. UNIFORM INTEGRABILITY

A collection of real-valued random variables $\{X_\alpha\}$ is *uniformly integrable* if $\sup_\alpha E[|X_\alpha|] < \infty$, and for every $\varepsilon > 0$ there exists a $\delta > 0$ such that for every α , $P(A_\alpha) < \delta$ implies $|E[X_\alpha \chi_{A_\alpha}]| < \varepsilon$.

2.1 Proposition The following are equivalent:

- (a) $\{X_\alpha\}$ is uniformly integrable.
- (b) $\lim_{N \rightarrow \infty} \sup_\alpha E[\chi_{\{|X_\alpha| > N\}} | X_\alpha |] = 0$.
- (c) $\lim_{N \rightarrow \infty} \sup_\alpha E[|X_\alpha| - N \wedge |X_\alpha|] = 0$.

Proof. Since $P\{|X_\alpha| > N\} \leq N^{-1}E[|X_\alpha|]$, it is immediate that (a) implies (b). More precisely,

$$NP\{|X_\alpha| > N\} \leq E[\chi_{\{|X_\alpha| > N\}} | X_\alpha |],$$

and since

$$(2.1) \quad \begin{aligned} E[|X_\alpha| - N \wedge |X_\alpha|] &= E[\chi_{\{|X_\alpha| > N\}} (|X_\alpha| - N)] \\ &= E[\chi_{\{|X_\alpha| > N\}} | X_\alpha |] - NP\{|X_\alpha| > N\}, \end{aligned}$$

(b) implies (c). Finally note that if $P(A_\alpha) \leq N^{-2}$, then

$$(2.2) \quad \begin{aligned} E[\chi_{A_\alpha} | X_\alpha |] &\leq E[|X_\alpha| - N \wedge |X_\alpha|] + NP(A_\alpha) \\ &\leq E[|X_\alpha| - N \wedge |X_\alpha|] + N^{-1}, \end{aligned}$$

and (c) implies (a). \square

2.2 Proposition A collection of real-valued random variables $\{X_n\}$ is uniformly integrable if and only if there exists an increasing convex function φ on $[0, \infty)$ such that $\lim_{x \rightarrow \infty} \varphi(x)/x = \infty$ and $\sup_n E[\varphi(|X_n|)] < \infty$.

Proof. We can assume $\varphi(0) = 0$. Then $\varphi(x)/x$ is increasing and

$$(2.3) \quad E[\chi_{\{|X_n| > N\}} |X_n|] \leq \frac{N E[\varphi(|X_n|)]}{\varphi(N)}.$$

Therefore sufficiency follows from Proposition 2.1(b).

By (b) there exists an increasing sequence $\{N_k\}$ such that

$$(2.4) \quad \sup_n \sum_{k=1}^{\infty} k E[\chi_{\{|X_n| > N_k\}} |X_n|] < \infty.$$

Assume $N_0 = 0$ and define $\varphi(0) = 0$ and

$$(2.5) \quad \varphi'(x) = \left(k - \frac{N_{k+1} - x}{N_{k+1} - N_k} \right), \quad N_k \leq x < N_{k+1}. \quad \square$$

2.3 Proposition If $X_n \Rightarrow X$ and $\{X_n\}$ is uniformly integrable, then $\lim_{n \rightarrow \infty} E[X_n] = E[X]$. Conversely, if the X_n are integrable, $X_n \Rightarrow X$, and $\lim_{n \rightarrow \infty} E[|X_n|] = E[|X|]$, then $\{X_n\}$ is uniformly integrable.

Proof. If $\{X_n\}$ is uniformly integrable, the first term on the right of

$$(2.6) \quad E[|X_n|] = E[|X_n| - N \wedge |X_n|] + E[N \wedge |X_n|]$$

can be made small uniformly in n . Consequently $\lim_{n \rightarrow \infty} E[|X_n|] = E[|X|]$, and hence $\lim_{n \rightarrow \infty} E[X_n] = E[X]$ by Theorem 1.2.

Conversely, since

$$(2.7) \quad \lim_{n \rightarrow \infty} E[|X_n| - N \wedge |X_n|] = E[|X|] - E[N \wedge |X|]$$

and the right side of (2.7) can be made arbitrarily small by taking N large, (b) of Proposition 2.1 follows. \square

2.4 Proposition Let $\{X_n\}$ be a uniformly integrable sequence of random variables defined on (Ω, \mathcal{F}, P) . Then there exists a subsequence $\{X_{n_k}\}$ and a random variable X such that

$$(2.8) \quad \lim_{k \rightarrow \infty} E[X_{n_k} Z] = E[XZ]$$

for every bounded random variable Z defined on (Ω, \mathcal{F}, P) .

2.5 Remark The converse also holds. See Dunford and Schwartz (1957), page 294. \square

Proof. Since we can consider $\{X_n \vee 0\}$ and $\{X_n \wedge 0\}$ separately, we may as well assume $X_n \geq 0$. Let \mathcal{A} be the algebra generated by $\{\{X_n < a\} : n = 1, 2, \dots, a \in \mathbb{Q}\}$. Note that \mathcal{A} is countable so there exists a subsequence $\{X_{n_k}\}$ such that

$$(2.9) \quad \mu(A) \equiv \lim_{k \rightarrow \infty} E[X_{n_k} \chi_A]$$

exists for every $A \in \mathcal{A}$. Let \mathcal{G} be the collection of sets $A \in \mathcal{F}$ for which the limit in (2.9) exists. Then $\mathcal{A} \subset \mathcal{G}$. If $A, B \in \mathcal{G}$ and $A \subset B$, then $B - A \in \mathcal{G}$. The uniform integrability implies that, if $\{A_k\} \subset \mathcal{G}$ and $A_1 \subset A_2 \subset \dots$, then $\bigcup_k A_k \in \mathcal{G}$. ($P(\bigcup_k A_k - A_m)$ can be made arbitrarily small by choosing m large.) Therefore the Dynkin class theorem (Appendix 4) implies $\mathcal{G} \supset \sigma(\mathcal{A})$.

Clearly μ is finitely additive on $\sigma(\mathcal{A})$, and the uniform integrability implies μ is countably additive.

Clearly $\mu \ll P$ on $\sigma(\mathcal{A})$, so there exists a $\sigma(\mathcal{A})$ -measurable random variable X such that $\mu(A) = E[X \chi_A]$, $A \in \sigma(\mathcal{A})$.

By (2.9),

$$(2.10) \quad \lim_{k \rightarrow \infty} E[X_{n_k} Z] = E[XZ]$$

for all simple $\sigma(\mathcal{A})$ -measurable random variables and the uniform integrability allows the extension of that conclusion to all bounded, $\sigma(\mathcal{A})$ -measurable random variables. Finally, for any bounded random variable Z ,

$$(2.11) \quad \begin{aligned} \lim_{k \rightarrow \infty} E[X_{n_k} Z] &= \lim_{k \rightarrow \infty} E[X_{n_k} E[Z | \sigma(\mathcal{A})]] \\ &= E[XE[Z | \sigma(\mathcal{A})]] = E[XZ]. \end{aligned}$$

□

3. BOUNDED POINTWISE CONVERGENCE

Let E be a metric space and let $V(E)$ denote the space of finite signed Borel measures on E with total variation norm

$$(3.1) \quad \|v\| = \sup_{A \in \mathcal{B}(E)} (|v(A)| + |v(E - A)|).$$

A sequence $\{f_n\} \subset B(E)$ converges in the weak* topology to f (denoted by $w^*\text{-lim}_{n \rightarrow \infty} f_n = f$) if $\lim_{n \rightarrow \infty} \int f_n dv = \int f dv$ for each $v \in V(E)$. A sequence $\{f_n\}$ converges boundedly and pointwise to f (denoted by $bp\text{-lim}_{n \rightarrow \infty} f_n = f$) if $\sup_n \|f_n\| < \infty$ and $\lim_{n \rightarrow \infty} f_n(x) = f(x)$ for each $x \in E$.

3.1 Proposition Let f_n , $n = 1, 2, \dots$, and f belong to $B(E)$. $w^*\text{-lim}_{n \rightarrow \infty} f_n = f$ if and only if $bp\text{-lim}_{n \rightarrow \infty} f_n = f$.

3.2 Remark This result holds only for sequences, not for nets. □

Proof. If $w^*\text{-}\lim_{n \rightarrow \infty} f_n = f$, then $\sup_v |\int f_n dv| < \infty$ for each $v \in V(E)$ and the uniform boundedness theorem (see e.g., Rudin (1974), page 104) implies $\sup_n \|f_n\| < \infty$. Of course, taking $v = \delta_x$ implies $\lim_{n \rightarrow \infty} f_n(x) = f(x)$.

The converse follows by the dominated convergence theorem. \square

Let $H \subset B(E)$. The *bp-closure* of H is the smallest subset \tilde{H} of $B(E)$ containing H such that $\{f_n\} \subset \tilde{H}$ and $\text{bp-lim}_{n \rightarrow \infty} f_n = f$ imply $f \in \tilde{H}$. Note the bp-closure of H is not necessarily the same as the weak* closure. For example, let $E = [0, 1]$ and $H = \{n\chi_{[k/n^2, (k+1)/n^2]} : 0 \leq k < n^2, n = 1, 2, \dots\}$. Then H is bp-closed, but it is not closed in the weak* topology.

4. MONOTONE CLASS THEOREMS

Let Ω be a set. A collection \mathcal{M} of subsets of Ω is a *monotone class* if

$$(M1) \quad \{A_n\} \subset \mathcal{M} \quad \text{and} \quad A_1 \subset A_2 \subset \dots \quad \text{imply} \quad \cup A_n \in \mathcal{M}$$

and

$$(M2) \quad \{A_n\} \subset \mathcal{M} \quad \text{and} \quad A_1 \supset A_2 \supset \dots \quad \text{imply} \quad \cap A_n \in \mathcal{M}.$$

A collection \mathcal{D} of subsets of Ω is a *Dynkin class* if

$$(D1) \quad \Omega \in \mathcal{D},$$

$$(D2) \quad A, B \in \mathcal{D} \quad \text{and} \quad A \subset B \quad \text{imply} \quad B - A \in \mathcal{D},$$

$$(D3) \quad \{A_n\} \subset \mathcal{D} \quad \text{and} \quad A_1 \subset A_2 \subset \dots \quad \text{imply} \quad \cup A_n \in \mathcal{D}.$$

4.1 Theorem (Monotone Class Theorem) If \mathcal{A} is an algebra and \mathcal{M} is a monotone class with $\mathcal{A} \subset \mathcal{M}$, then $\sigma(\mathcal{A}) \subset \mathcal{M}$.

Proof. Let $M(\mathcal{A})$ be the smallest monotone class containing \mathcal{A} . We want to show $M(\mathcal{A}) = \sigma(\mathcal{A})$. Clearly it will be sufficient to show that $M(\mathcal{A})$ is a σ -algebra.

First note that $\{A \in M(\mathcal{A}) : A^c \in M(\mathcal{A})\}$ is a monotone class that contains \mathcal{A} and hence $M(\mathcal{A})$, that is, $A \in M(\mathcal{A})$ implies $A^c \in M(\mathcal{A})$. Next note that for $A \in \mathcal{A}$, $\{B : A \cup B \in M(\mathcal{A})\}$ is a monotone class containing \mathcal{A} and hence $M(\mathcal{A})$, that is, $A \in \mathcal{A}$ and $B \in M(\mathcal{A})$ imply $A \cup B \in M(\mathcal{A})$. Finally, by this last observation, if $A \in M(\mathcal{A})$ then $\{B : A \cup B \in M(\mathcal{A})\}$ is a monotone class containing \mathcal{A} and hence $M(\mathcal{A})$, that is, $M(\mathcal{A})$ is closed under finite unions. Since $M(\mathcal{A})$ is closed under finite unions, by (M1) it is closed under countable unions, and hence is a σ -algebra. \square

4.2 Theorem (Dynkin Class Theorem) Let \mathcal{S} be a collection of subsets of Ω such that $A, B \in \mathcal{S}$ implies $A \cap B \in \mathcal{S}$. If \mathcal{D} is a Dynkin class with $\mathcal{S} \subset \mathcal{D}$, then $\sigma(\mathcal{S}) \subset \mathcal{D}$.

Proof. Let $D(\mathcal{S})$ be the smallest Dynkin class that contains \mathcal{S} . It is sufficient to show that $D(\mathcal{S})$ is a σ -algebra. This will follow from (D3) if we show $D(\mathcal{S})$ is closed under finite unions.

If $A, B \in \mathcal{S}$, then A^c, B^c , and $A^c \cup B^c = \Omega - A \cap B$ are in $D(\mathcal{S})$. Consequently $A^c \cup B^c - A^c = A \cap B^c$, $A^c \cup B = \Omega - A \cap B^c$, $A^c \cap B^c = A^c \cup B - B$, and $A \cup B = \Omega - A^c \cap B^c$ are in $D(\mathcal{S})$.

For $A \in \mathcal{S}$, $\{B: A \cup B \in D(\mathcal{S})\}$ is a Dynkin class containing \mathcal{S} , and hence $A \in \mathcal{S}$ and $B \in D(\mathcal{S})$ imply $A \cup B \in D(\mathcal{S})$. Consequently, for $A \in D(\mathcal{S})$, $\{B: A \cup B \in D(\mathcal{S})\}$ is a Dynkin class containing \mathcal{S} , and hence $A, B \in D(\mathcal{S})$ implies $A \cup B \in D(\mathcal{S})$. \square

4.3 Theorem Let H be a linear space of bounded functions on Ω that contains constants, and let \mathcal{S} be a collection of subsets of Ω such that $A, B \in \mathcal{S}$ implies $A \cap B \in \mathcal{S}$. Suppose $\chi_A \in H$ for all $A \in \mathcal{S}$, and $\{f_n\} \subset H, f_1 \leq f_2 \leq \dots$, and $\sup_n f_n \leq c$ for some constant c imply $f \equiv \text{bp-lim}_{n \rightarrow \infty} f_n \in H$. Then H contains all bounded $\sigma(\mathcal{S})$ -measurable functions.

Proof. Note that $\{A: \chi_A \in H\}$ is a Dynkin class containing \mathcal{S} and hence $\sigma(\mathcal{S})$. Since H is linear, H contains all simple $\sigma(\mathcal{S})$ -measurable functions. Since any bounded $\sigma(\mathcal{S})$ -measurable function is the pointwise limit of an increasing sequence of simple functions, the theorem follows. \square

4.4 Corollary Let H be a linear space of bounded functions on Ω containing constants that is closed under uniform convergence and under bounded pointwise convergence of nondecreasing sequences (as in Theorem 4.3). Suppose $H_0 \subset H$ is closed under multiplication ($f, g \in H_0$ implies $fg \in H_0$). Then H contains all bounded $\sigma(H_0)$ -measurable functions.

Proof. Let $F \in C(\mathbb{R})$. Then on any bounded interval, F is the uniform limit of polynomials, and hence $f \in H_0$ implies $F(f) \in H$. In particular, $f_n = [1 \wedge (f - a) \vee 0]^{1/n}$ is in H . Note that $f_1 \leq f_2 \leq \dots \leq 1$, and hence

$$(4.1) \quad \chi_{\{f > a\}} = \lim_{n \rightarrow \infty} f_n \in H.$$

Similarly, for $f_1, \dots, f_m \in H_0$,

$$(4.2) \quad \chi_{\{f_1 > a_1, \dots, f_m > a_m\}} \in H$$

and, since $\sigma(H_0) = \sigma(\{\{f_i > a_1, \dots, f_m > a_m\}: f_i \in H_0, a_i \in \mathbb{R}\})$, the corollary follows. \square

We give an additional application of Theorem 4.3.

4.5 Proposition Let E_1 and E_2 be separable metric spaces. Let X be an E_1 -valued random variable defined on (Ω, \mathcal{F}, P) , and let \mathcal{H} be a sub- σ -algebra of \mathcal{F} . Then for each $\psi \in B(E_1 \times E_2)$ there is a bounded $\mathcal{B}(E_2) \times \mathcal{H}$ -measurable function φ such that

$$(4.3) \quad E[\psi(X, Y) | \mathcal{H}](\omega) = \varphi(Y(\omega), \omega),$$

for every \mathcal{H} -measurable, E_2 -valued random variable Y . If X is independent of \mathcal{H} , then φ does not depend on ω . Specifically,

$$(4.4) \quad \varphi(y, \omega) = \varphi(y) = E[\psi(X, y)].$$

Proof. If $\psi(x, y) = g(x)h(y)$, then $\varphi(y, \cdot) = h(y)E[g(X) | \mathcal{H}]$. Let H be the collection of $\psi \in B(E_1 \times E_2)$ for which the conclusion of the proposition is valid, and let $\mathcal{S} = \{A \times B : A \in \mathcal{B}(E_1), B \in \mathcal{B}(E_2)\}$. Since $\chi_{A \times B}(x, y) = \chi_A(x)\chi_B(y)$, the proposition follows by Theorem 4.3. \square

5. GRONWALL'S INEQUALITY

5.1 Theorem Let μ be a Borel measure on $[0, \infty)$, let $\varepsilon \geq 0$, and let f be a Borel measurable function that is bounded on bounded intervals and satisfies

$$(5.1) \quad 0 \leq f(t) \leq \varepsilon + \int_{[0, t]} f(s) \mu(ds), \quad t \geq 0.$$

Then

$$(5.2) \quad f(t) \leq \varepsilon e^{\mu[0, t]}, \quad t \geq 0.$$

In particular, if $M > 0$ and

$$(5.3) \quad 0 \leq f(t) \leq \varepsilon + M \int_0^t f(s) ds, \quad t \geq 0,$$

then

$$(5.4) \quad f(t) \leq \varepsilon e^{Mt}, \quad t \geq 0.$$

Proof. Iterating (5.1) gives

$$\begin{aligned} (5.5) \quad f(t) &\leq \varepsilon + \varepsilon \sum_{k=1}^{\infty} \int_{[0, t]} \int_{[0, s_1]} \cdots \int_{[0, s_{k-1}]} \mu(ds_k) \cdots \mu(ds_1) \\ &\leq \varepsilon + \varepsilon \sum_{k=1}^{\infty} \frac{1}{k!} (\mu[0, t])^k \\ &= \varepsilon e^{\mu[0, t]}. \end{aligned}$$

\square

6. THE WHITNEY EXTENSION THEOREM

For $x \in \mathbb{R}^d$ and $\alpha \in \mathbb{Z}_+^d$, let $x^\alpha = \prod_{k=1}^d x_k^{\alpha_k}$, $|\alpha| = \sum_{k=1}^d \alpha_k$, and $\alpha! = \prod_{k=1}^d \alpha_k!$. Similarly, if D_k denotes differentiation in the k th variable,

$$(6.1) \quad D^\alpha f = \prod_{k=1}^d D_k^{\alpha_k} f,$$

and if f is r times continuously differentiable on a convex, open set in \mathbb{R}^d , then by Taylor's theorem

$$(6.2) \quad \begin{aligned} D^\alpha f(y) &= \sum_{|\beta| \leq r - |\alpha|} \frac{1}{\beta!} D^{\alpha+\beta} f(x)(y-x)^\beta \\ &\quad + \sum_{|\beta|=r-|\alpha|} \frac{r-|\alpha|}{\beta!} \int_0^1 (1-u)^{r-|\alpha|-1} [D^{\alpha+\beta} f(x+u(y-x)) \\ &\quad \quad \quad - D^{\alpha+\beta} f(x)] du (y-x)^\beta. \end{aligned}$$

6.1 Theorem Let $E \subset \mathbb{R}^d$ be closed. Suppose a collection of functions $\{f_\alpha : \alpha \in \mathbb{Z}_+^d, |\alpha| \leq r\}$ satisfies $f_\alpha : E \rightarrow \mathbb{R}$ for each α ,

$$(6.3) \quad f_\alpha(y) = \sum_{|\beta| \leq r - |\alpha|} \frac{1}{\beta!} f_{\alpha+\beta}(x)(y-x)^\beta + R_\alpha(x, y)$$

for all $x, y \in E$, and for each compact set $K \subset \mathbb{R}^d$,

$$(6.4) \quad \lim_{\delta \rightarrow 0} \sup \left\{ \frac{|R_\alpha(x, y)|}{|x-y|^{r-|\alpha|}} : x, y \in E \cap K, |x-y| < \delta \right\} = 0.$$

Then there exists $f \in C^r(\mathbb{R}^d)$ such that $f|_E = f_0$.

6.2 Remark Essentially the theorem states that a function f_0 that is r times continuously differentiable on E can be extended to a function that is r times continuously differentiable on \mathbb{R}^d . \square

Proof. The theorem is due to Whitney (1934). For a more recent exposition see Abraham and Robbin (1967), Appendix A. \square

6.3 Corollary Let E be convex, and suppose E is the closure of its interior E° . Suppose f_0 is r times continuously differentiable on E° and that the derivatives $D^\alpha f_0$ are uniformly continuous on E° . Then there exists $f \in C^r(\mathbb{R}^d)$ such that $f|_{E^\circ} = f_0$.

Proof. Let $R_\alpha(x, y)$ be the remainder (second) term on the right of (6.2). There exists a constant C such that for $x, y \in E^\circ$,

$$(6.5) \quad |R_\alpha(x, y)| \leq C|x-y|^{r-|\alpha|} w(|x-y|),$$

where

$$(6.6) \quad w(\delta) = \max_{|x|=r} \sup_{\substack{x, y \in E \\ |x-y|<\delta}} |D^q f(y) - D^q f(x)|.$$

By continuity (6.5) extends to all $x, y \in E$. Since $\lim_{\delta \rightarrow 0} w(\delta) = 0$, the corollary follows. \square

7. APPROXIMATION BY POLYNOMIALS

In a variety of contexts it is useful to approximate a function $f \in C_c(\mathbb{R}^d)$ by polynomials in such a way that not only f but all of its derivatives of order less than or equal to r are approximated uniformly on compact sets. To obtain such approximations, one need only construct a sequence of polynomials $\{\rho_n\}$ that are approximate delta functions in the sense that for every $f \in C_c(\mathbb{R}^d)$,

$$(7.1) \quad \lim_{n \rightarrow \infty} \int f(y) \rho_n(x-y) dy = f(x),$$

and the convergence is uniform for x in compact sets. Such a sequence can be constructed in a variety of ways. A simple example is

$$(7.2) \quad \rho_n(z) = n^d \left(1 - \frac{|z|^2}{n^2} \right)^{n^d} \pi^{-d/2}.$$

To see that this sequence has the desired property, first note that

$$(7.3) \quad \sup_{|u| \leq n^2} e^{|u|^2} \left(1 - \frac{|u|^2}{n^4} \right)^{n^d} = 1.$$

For x in a fixed compact set and n sufficiently large

$$\begin{aligned} (7.4) \quad & \int_{\mathbb{R}^d} f(y) \rho_n(x-y) dy \\ &= \int_{\mathbb{R}^d} f\left(x - \frac{1}{n} u\right) \left(1 - \frac{|u|^2}{n^4} \right)^{n^d} \pi^{-d/2} du \\ &= \int_{|u| \leq n^2} f\left(x - \frac{1}{n} u\right) e^{-|u|^2} e^{|u|^2} \left(1 - \frac{|u|^2}{n^4} \right)^{n^d} \pi^{-d/2} du. \end{aligned}$$

The second equality follows from the fact that f has compact support, and (7.1) follows by the dominated convergence theorem.

7.1 Proposition Let $f \in C^r(\mathbb{R}^d)$. Then for each compact set K and $\varepsilon > 0$, there exists a polynomial p such that for every $|\alpha| \leq r$,

$$(7.5) \quad \sup_{x \in K} |D^\alpha f(x) - D^\alpha p(x)| \leq \varepsilon.$$

Proof. Without loss of generality we can assume f has compact support. (Replace f by $\zeta \cdot f$, where $\zeta \in C_c^\infty(\mathbb{R}^d)$ and $\zeta = 1$ on K .) Take

$$(7.6) \quad p_n(x) = \int_{\mathbb{R}^d} f(y) \rho_n(x-y) dy = \int_{\mathbb{R}^d} f(x-y) \rho_n(y) dy$$

and note

$$(7.7) \quad D^\alpha p_n(x) = \int_{\mathbb{R}^d} D^\alpha f(x-y) \rho_n(y) dy = \int_{\mathbb{R}^d} D^\alpha f(y) \rho_n(x-y) dy.$$

For n sufficiently large, p_n will have the desired properties. \square

As an application of the previous result we have the following.

7.2 Proposition Let φ be convex on \mathbb{R}^d . Then for each compact, convex set K and $\varepsilon > 0$, there exists a polynomial p such that p is convex on K and

$$(7.8) \quad \sup_{x \in K} |\varphi(x) - p(x)| \leq \varepsilon.$$

Proof. Let $\rho \in C_c^\infty(\mathbb{R}^d)$ be nonnegative and

$$(7.9) \quad \int_{\mathbb{R}^d} \rho(y) dy = 1.$$

Then for n sufficiently large,

$$(7.10) \quad \varphi_1(x) \equiv \int_{\mathbb{R}^d} \varphi(y) n^d \rho(n(x-y)) dy$$

is infinitely differentiable, convex, and satisfies

$$(7.11) \quad \sup_{x \in K} |\varphi(x) - \varphi_1(x)| \leq \frac{\varepsilon}{3}.$$

For δ sufficiently small, $\varphi_2(x) \equiv \varphi_1(x) + \delta |x|^2$ satisfies

$$(7.12) \quad \sup_{x \in K} |\varphi(x) - \varphi_2(x)| \leq \frac{2\varepsilon}{3}.$$

Recall that a function $\psi \in C^2(\mathbb{R}^d)$ is convex on K if and only if the Hessian matrix $((D_i D_j \psi))$ is nonnegative definite. Note that $((D_i D_j \varphi_2))$ is positive definite. In particular

$$(7.13) \quad \sum z_i z_j D_i D_j \varphi_2(x) \geq \delta |z|^2.$$

By Proposition 7.1 there exists a polynomial p such that

$$(7.14) \quad \sup_{x \in K} |\varphi_2(x) - p(x)| \leq \frac{\varepsilon}{3},$$

and $D_i D_j p$ approximates $D_i D_j \varphi_2$ closely enough so that

$$(7.15) \quad \sum z_i z_j D_i D_j p(x) \geq \frac{\delta |z|^2}{2}, \quad x \in K.$$

Consequently, p is convex on K , and (7.12) and (7.14) imply p satisfies (7.8). \square

8. BIMEASURES AND TRANSITION FUNCTIONS

Let (M, \mathcal{M}) be a measurable space and (E, r) a complete, separable metric space. A function $v_0(A, B)$ defined for $A \in \mathcal{M}$ and $B \in \mathcal{B}(E)$ is a *bimeasure* if for each $A \in \mathcal{M}$, $v_0(A, \cdot)$ is a measure on $\mathcal{B}(E)$ and for each $B \in \mathcal{B}(E)$, $v_0(\cdot, B)$ is a measure on \mathcal{M} .

8.1 Theorem Let v_0 be a bimeasure on $\mathcal{M} \times \mathcal{B}(E)$ such that $0 < v_0(M, E) < \infty$, and define $\mu = v_0(\cdot, E)$. Then there exists $\eta: M \times \mathcal{B}(E) \rightarrow [0, \infty)$ such that for each $x \in M$, $\eta(x, \cdot)$ is a measure on $\mathcal{B}(E)$, for each $B \in \mathcal{B}(E)$, $\eta(\cdot, B)$ is \mathcal{M} -measurable, and

$$(8.1) \quad v_0(A, B) = \int_A \eta(x, B) \mu(dx), \quad A \in \mathcal{M}, B \in \mathcal{B}(E).$$

Furthermore,

$$(8.2) \quad v(C) \equiv \iint \chi_C(x, y) \eta(x, dy) \mu(dx)$$

defines a measure on the product σ -algebra $\mathcal{M} \times \mathcal{B}(E)$ satisfying $v(A \times B) = v_0(A, B)$ for all $A \in \mathcal{M}$, $B \in \mathcal{B}(E)$.

8.2 Remark The first part of the theorem is essentially just the existence of a regular conditional distribution. The observation that a bimeasure (as defined by Kingman (1967)) determines a measure on the product σ -algebra is due to Morando (1969), page 224.

Proof. Without loss of generality, we can assume $v_0(M, E) = 1$ (otherwise replace $v_0(A, B)$ by $v_0(A, B)/v_0(M, E)$). Let $\{x_i\}$ be a countable dense subset of E , and let B_1, B_2, \dots be an ordering of $\{B(x_i, k^{-1}) : i = 1, 2, \dots, k = 1, 2, \dots\}$.

For each $B \in \mathcal{B}(E)$, $v_0(\cdot, B) \ll \mu$, so there exists $\eta_0(\cdot, B)$, \mathcal{M} -measurable, such that

$$(8.3) \quad v_0(A, B) = \int_A \eta_0(x, B) \mu(dx), \quad A \in \mathcal{M}.$$

We can always assume $\eta_0(x, B) \leq 1$, and for fixed B, C , with $B \subset C$, we can assume $\eta_0(x, B) \leq \eta_0(x, C)$ for all x . Therefore we may define $\eta_0(x, E) = 1$, select $\eta_0(x, B_1)$ satisfying (8.3) (with $B = B_1$), and define $\eta_0(x, B_1^c) = 1 - \eta_0(x, B_1)$, which satisfies (8.3) with $B = B_1^c$. For any sequence C_1, C_2, \dots where C_i is B_i or B_i^c , working recursively we can select $\eta_0(x, C_1 \cap C_2 \cap \dots \cap C_k \cap B_{k+1})$ satisfying (8.3) with $B = C_1 \cap C_2 \cap \dots \cap C_k \cap B_{k+1}$ and $\eta_0(x, C_1 \cap C_2 \cap \dots \cap C_k \cap B_{k+1}) \leq \eta_0(x, C_1 \cap C_2 \cap \dots \cap C_k)$, and define

$$\eta_0(x, C_1 \cap C_2 \cap \dots \cap C_k \cap B_{k+1}) = \eta_0(x, C_1 \cap C_2 \cap \dots \cap C_k) - \eta_0(x, C_1 \cap C_2 \cap \dots \cap C_k \cap B_{k+1}),$$

which satisfies (8.3) with $B = C_1 \cap C_2 \cap \dots \cap C_k \cap B_{k+1}^c$. For $B \in \mathcal{F}_n = \sigma(B_1, \dots, B_n)$, define $\eta_0(x, B) = \sum \eta_0(x, C_1 \cap C_2 \cap \dots \cap C_n)$ where the sum is over $\{C_1 \cap C_2 \cap \dots \cap C_n : C_i \text{ is } B_i \text{ or } B_i^c, C_1 \cap C_2 \cap \dots \cap C_n \subset B\}$. Then $\eta_0(x, B)$ satisfies (8.3) and $\eta_0(x, \cdot)$ is finitely additive on $\bigcup_n \mathcal{F}_n$.

Let $\Gamma_n = \{C_1 \cap C_2 \cap \dots \cap C_n : C_i \text{ is } B_i \text{ or } B_i^c\}$, and for $C \in \Gamma_n$ such that $C \neq \emptyset$, let $z_C \in C$. Define $\eta_n(x, \cdot) \in \mathcal{P}(E)$ by

$$(8.4) \quad \eta_n(x, B) = \sum_{C \in \Gamma_n} \delta_{z_C}(B) \eta_0(x, C).$$

Note that for $B \in \mathcal{F}_n$, $\eta_n(x, B) = \eta_0(x, B)$. For $m = 1, 2, \dots$ let K_m be compact and satisfy $v_0(M, K_m) \geq 1 - 2^{-m}$. For each m , there exists N_m such that for $n \geq N_m$ there is a $B \in \mathcal{F}_n$ satisfying $K_m \subset B \subset K_m^{1/m}$. Hence

$$(8.5) \quad \int \inf_{n \geq N_m} \eta_n(x, K_m^{1/m}) \mu(dx) \geq \int \eta_0(x, B) \mu(dx) \geq v_0(M, K_m) \geq 1 - 2^{-m}.$$

Therefore

$$(8.6) \quad \mu \left\{ x : \inf_{n \geq N_m} \eta_n(x, K_m^{1/m}) < 1 - m^{-1} \right\} \leq m 2^{-m},$$

and hence by Borel-Cantelli

$$(8.7) \quad G = \left\{ x : \limsup_{n \rightarrow \infty} \eta_n(x, K_m^{1/m}) \geq 1 - m^{-1} \text{ for all but finitely many } m \right\}$$

satisfies $\mu(G) = 1$. It follows that for each $x \in G$, $\{\eta_n(x, \cdot)\}$ is relatively compact. Since $\lim_{n \rightarrow \infty} \eta_n(x, B) = \eta_0(x, B)$ for every $B \in \bigcup_n \mathcal{F}_n$, for $x \in G$ there exists

$\eta(x, \cdot)$ such that $\eta_n(x, \cdot) \Rightarrow \eta(x, \cdot)$. (See Problem 27 in Chapter 3.) By Theorem 3.1 of Chapter 3

$$(8.8) \quad \lim_{n \rightarrow \infty} \int_A \eta_n(x, B) \mu(dx) = \int_A \eta(x, B) \mu(dx)$$

for all $B \in \mathcal{B}(E)$ such that

$$(8.9) \quad \int_A \eta(x, \partial B) \mu(dx) = 0.$$

Since for $B \in \bigcup_n \mathcal{F}_n$

$$(8.10) \quad \lim_{n \rightarrow \infty} \int_A \eta_n(x, B) \mu(dx) = v_0(A, B),$$

it follows from Problem 27 of Chapter 3 that (8.1) holds. \square

9. TULCEA'S THEOREM

9.1 Theorem Let $(\Omega_k, \mathcal{F}_k)$, $k = 1, 2, \dots$, be measurable spaces, $\Omega = \Omega_1 \times \Omega_2 \times \dots$ and $\mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2 \times \dots$. Let P_1 be a probability measure on \mathcal{F}_1 , and for $k = 2, 3, \dots$ let $P_k: \Omega_1 \times \dots \times \Omega_{k-1} \times \mathcal{F}_k \rightarrow [0, 1]$ be such that for each $(\omega_1, \dots, \omega_{k-1}) \in \Omega_1 \times \dots \times \Omega_{k-1}$, $P_k(\omega_1, \dots, \omega_{k-1}, \cdot)$ is a probability measure on \mathcal{F}_k , and for each $A \in \mathcal{F}_k$, $P_k(\cdot, A)$ is $\mathcal{F}_1 \times \dots \times \mathcal{F}_{k-1}$ -measurable. Then there is a probability measure P on \mathcal{F} such that for $A \in \mathcal{F}_1 \times \dots \times \mathcal{F}_k$,

$$(9.1) \quad P(A \times \Omega_{k+1} \times \Omega_{k+2} \times \dots)$$

$$= \int_{\Omega_1} \dots \int_{\Omega_k} \chi_A(\omega_1, \dots, \omega_k) P_k(\omega_1, \dots, \omega_{k-1}, d\omega_k) \dots P_1(d\omega_1).$$

Proof. The collection of sets

$$\mathcal{A} = \{A \times \Omega_{k+1} \times \Omega_{k+2} \times \dots : A \in \mathcal{F}_1 \times \dots \times \mathcal{F}_k, k = 1, 2, \dots\}$$

is an algebra. Clearly P defined by (9.1) is finitely additive on \mathcal{A} . To apply the Caratheodory extension theorem to extend P to a measure on $\sigma(\mathcal{A}) = \mathcal{F}$, we must show that P is countably additive on \mathcal{A} . (See Billingsley (1979), Theorem 3.1.)

To verify countable additivity it is enough to show that $\{B_n\} \subset \mathcal{A}$, $B_1 \supset B_2 \supset \dots$ and $\lim_{n \rightarrow \infty} P(B_n) > 0$ imply $\bigcap_n B_n \neq \emptyset$. Let $B_n = A_n \times \Omega_{n+1}$.

$\times \Omega_{k_n+2} \times \cdots$ for $A_n \in \mathcal{F}_1 \times \cdots \times \mathcal{F}_{k_n}$, $B_1 \supset B_2 \supset \cdots$, and $\lim_{n \rightarrow \infty} P(B_n) > 0$. (We can assume $k_n \rightarrow \infty$.) For $n = 1, 2, \dots$ and $k < k_n$, define

$$(9.2) \quad f_{k,n}(\omega_1, \dots, \omega_k) = \int_{\Omega_{k+1}} \cdots \int_{\Omega_{k_n}} \chi_{A_n}(\omega_1, \dots, \omega_{k_n}) \\ \times P_{k_n}(\omega_1, \dots, \omega_{k_n-1}, d\omega_{k_n}) \cdots P_{k+1}(\omega_1, \dots, \omega_k, d\omega_{k+1}),$$

and for $k \geq k_n$, $f_{k,n}(\omega_1, \dots, \omega_k) = \chi_{A_n}(\omega_1, \dots, \omega_{k_n})$. Note that

$$(9.3) \quad f_{k,n}(\omega_1, \dots, \omega_k) = \int f_{k+1,n}(\omega_1, \dots, \omega_{k+1}) P_{k+1}(\omega_1, \dots, \omega_k, d\omega_{k+1})$$

and

$$(9.4) \quad P(B_n) = \int f_{1,n}(\omega_1) P_1(d\omega_1).$$

Furthermore note that $f_{k,n} \geq f_{k,n+1}$ so $g_k \equiv \text{bp-lim}_{n \rightarrow \infty} f_{k,n}$ exists, and by the monotone convergence theorem,

$$(9.5) \quad \lim_{n \rightarrow \infty} P(B_n) = \int g_k(\omega_1) P_1(d\omega_1)$$

and

$$(9.6) \quad g_k(\omega_1, \dots, \omega_k) = \int g_{k+1}(\omega_1, \dots, \omega_{k+1}) P_{k+1}(\omega_1, \dots, \omega_k, d\omega_{k+1}).$$

If $\lim_{n \rightarrow \infty} P(B_n) > 0$, there must be $\tilde{\omega}_1 \in \Omega_1$ such that $g_1(\tilde{\omega}_1) > 0$ and by induction a sequence $\tilde{\omega}_1, \tilde{\omega}_2, \dots$ such that $g_k(\tilde{\omega}_1, \dots, \tilde{\omega}_k) > 0$, $k = 1, 2, \dots$. Finally, since

$$(9.7) \quad g_{k_n}(\tilde{\omega}_1, \dots, \tilde{\omega}_{k_n}) \leq f_{k_n,n}(\tilde{\omega}_1, \dots, \tilde{\omega}_{k_n}) \\ = \chi_{A_n}(\tilde{\omega}_1, \dots, \tilde{\omega}_{k_n}),$$

$(\tilde{\omega}_1, \tilde{\omega}_2, \dots) \in B_n$ for every n , and hence $(\tilde{\omega}_1, \tilde{\omega}_2, \dots) \in \bigcap_n B_n$. \square

10. MEASURABLE SELECTIONS AND MEASURABILITY OF INVERSES

Let (M, \mathcal{M}) be a measurable space and (S, ρ) a complete, separable metric space. Suppose for each $x \in M$, $\Gamma_x \subset S$. A measurable selection of $\{\Gamma_x\}$ is an \mathcal{M} -measurable function $f: M \rightarrow S$ such that $f(x) \in \Gamma_x$ for every $x \in M$.

10.1 Theorem Suppose for each $x \in M$, Γ_x is a closed subset of S and that for every open set $U \subset S$, $\{x \in M: \Gamma_x \cap U \neq \emptyset\} \in \mathcal{M}$. Then there exist $f_n: M \rightarrow S$, $n = 1, 2, \dots$, such that f_n is \mathcal{M} -measurable, $f_n(x) \in \Gamma_x$ for every $x \in M$, and Γ_x is the closure of $\{f_1(x), f_2(x), \dots\}$.

10.2 Remark Regarding $x \rightarrow \Gamma_x$ as a set-valued function, if $\{x \in M : \Gamma_x \cap U \neq \emptyset\} \in \mathcal{M}$ for every open U , the function is said to be *weakly measurable*. The function is *measurable* if “open” can be replaced by “closed.” The theorem not only gives the existence of a measurable selection, but also shows that any closed-set-valued, weakly measurable function has the representation (known as the Castaing representation) $\Gamma_x = \text{closure } \{f_1(x), f_2(x), \dots\}$ for some countable collection of \mathcal{M} -measurable functions. \square

Proof. See Himmelberg (1975), Theorem 5.6. Earlier versions of the result are in Castaing (1967) and Kuratowski and Ryll-Nardzewski (1965). \square

10.3 Corollary Suppose $(M, \mathcal{M}) = (E, \mathcal{B}(E))$ for a metric space E . If $y_n \in \Gamma_{x_n}$, $n = 1, 2, \dots$, and $\lim_{n \rightarrow \infty} x_n = x$ imply that $\{y_n\}$ has a limit point in Γ_x , then there is a measurable selection of $\{\Gamma_x\}$.

10.4 Remark The assumptions of the corollary imply that for $K \subset E$ compact, $\bigcup_{x \in K} \Gamma_x$ is compact. \square

Proof. Note that for a closed set F , $\{x : \Gamma_x \cap F \neq \emptyset\}$ is closed, hence measurable. If U is open, then $U = \bigcup_n F_n$ for some sequence of closed sets $\{F_n\}$, and hence $\{x : \Gamma_x \cap U \neq \emptyset\} = \bigcup_n \{x : \Gamma_x \cap F_n \neq \emptyset\}$ is measurable. \square

For a review of results on measurable selections, see Wagner (1977).

One source of set-valued functions is the inverse mapping of a given function $\varphi : E_1 \rightarrow E_2$, that is, for $x \in E_2$ take $\Gamma_x = \varphi^{-1}(x) = \{y \in E_1 : \varphi(y) = x\}$. If φ is one-to-one, then the existence of a measurable selection is precisely the measurability of the inverse function. The following theorem of Kuratowski gives conditions for this measurability.

10.5 Theorem Let (S_1, ρ_1) and (S_2, ρ_2) be complete, separable metric spaces. Let $E_1 \in \mathcal{B}(S_1)$, and let $\varphi : E_1 \rightarrow S_2$ be Borel measurable and one-to-one. Then $E_2 \equiv \varphi(E_1) \equiv \{\varphi(x) : x \in E_1\}$ is a Borel subset of S_2 and φ^{-1} is a Borel measurable function from E_2 onto E_1 .

Proof. See Theorem 3.9 and Corollary 3.3 of Chapter I of Parthasarathy (1967). \square

11. ANALYTIC SETS

Let \mathbb{N} denote the set of positive integers and $\mathcal{N} = \mathbb{N}^\omega$. We give \mathbb{N} the discrete topology and \mathcal{N} the corresponding product topology. Let (S, ρ) be a complete, separable metric space. A subset $A \subset S$ is *analytic* if there exists a continuous function φ mapping \mathcal{N} onto A .

11.1 Proposition Every Borel subset of a complete, separable metric space is analytic.

Proof. See Theorem 2.5 of Parthasarathy (1967). \square

Analytic sets arise most naturally as images of Borel sets.

11.2 Proposition Let (S_1, ρ_1) and (S_2, ρ_2) be complete, separable metric spaces and let $\varphi: S_1 \rightarrow S_2$ be Borel measurable. If $A \in \mathcal{B}(S_1)$, then $\varphi(A) = \{\varphi(x): x \in A\}$ is an analytic subset of S_2 .

Proof. See Theorem 3.4 of Parthasarathy (1967). \square

11.3 Theorem Let (S, ρ) be a complete, separable metric space and let (Ω, \mathcal{F}, P) be a complete probability space. If Y is an S -valued random variable defined on (Ω, \mathcal{F}, P) and A is an analytic subset of S , then $\{Y \in A\} \in \mathcal{F}$.

Proof. See Dellacherie and Meyer (1978), page 58. The definition of analytic set used there is more general than that given above. The role of the paved set (F, \mathcal{F}) in the definition in Dellacherie and Meyer (page 41) is taken by $(S, \mathcal{B}(S))$, and the auxiliary compact space E is $(\mathbb{N}^A)^\omega$, where \mathbb{N}^A is the one-point compactification of \mathbb{N} . Let $B \subset E \times S$ be given by $B = \{(x, \varphi(x)): x \in \mathbb{N}^\omega\}$, where φ is continuous on \mathbb{N}^ω . Then for $\{z_i\}$ dense in S , $B = \bigcap_n \bigcup_i \bigcup_m \text{cl}\{x \in \mathbb{N}^\omega: |x_j| \leq m, j = 1, \dots, n, \varphi(x) \in B(z_i, n^{-1})\} \times B(z_i, n^{-1})$, where cl denotes the closure in $(\mathbb{N}^A)^\omega$. Consequently $B \in (\mathcal{K}(E) \times \mathcal{B}(S))_{ss}$ ($\mathcal{K}(E)$ is the class of compact subsets of E) and A is the projection onto S of B , so A is $\mathcal{B}(S)$ -analytic in the terminology of Dellacherie and Meyer. \square

REFERENCES

- Abraham, Ralph and Robbin, Joel (1967). *Transversal Mappings and Flows*. Benjamin, New York.
- Aldous, David (1978). Stopping times and tightness. *Ann. Probab.* **6**, 335–340.
- Alexandroff, A. D. (1940–1943). Additive set functions in abstract spaces. *Mat. Sb.* **8**, 307–348; **9**, 563–628; **13**, 169–238.
- Allain, Marie-France (1976). Étude de la vitesse de convergence d'une suite de processus de Markov de saut pur. *C. R. Acad. Sci. Paris* **282**, 1015–1018.
- Aim, Sven Erick (1978). On the rate of convergence in diffusion approximation of jump Markov processes. Technical Report, Department of Mathematics, Uppsala University, Sweden.
- Anderson, Robert F. (1976). Diffusion with second order boundary conditions, I, II. *Indiana Univ. Math. J.* **25**, 367–395, 403–441.
- Artstein, Zvi (1983). Distributions of random sets and random selections. *Israel J. Math.* **46**, 313–324.
- Athreya, Krishna B. and Ney, Peter E. (1972). *Branching Processes*. Springer-Verlag, Berlin.
- Barbour, Andrew D. (1974). On a functional central limit theorem for Markov population processes. *Adv. Appl. Probab.* **6**, 21–39.
- Barbour, Andrew D. (1976). Second order limit theorems for the Markov branching process in random environments. *Stochastic Process. Appl.* **4**, 33–40.
- Barbour, Andrew D. (1980). Equilibrium distributions for Markov population processes. *Adv. Appl. Probab.* **12**, 591–614.
- Bártfai, Pál (1966). Die Bestimmung des zu einem wiederkehrenden Prozess gehörenden Verteilungsfunktion aus den mit Fehlern behafteten Daten einer Einziger Relation. *Stud. Sci. Math. Hung.* **1**, 161–168.

- Beneš, Václav Edvard (1968). Finite regular invariant measures for Feller processes. *J. Appl. Probab.* **5**, 203–209.
- Bhattacharya, Rabi N. (1982). On the functional central limit theorem and the law of the iterated logarithm for Markov processes. *Z. Wahrsch. verw. Gebiete* **60**, 185–201.
- Billingsley, Patrick (1968). *Convergence of Probability Measures*. Wiley, New York.
- Billingsley, Patrick (1979). *Probability and Measure*. Wiley, New York.
- Blackwell, David and Dubins, Lester E. (1983). An extension of Skorohod's almost sure representation theorem. *Proc. Amer. Math. Soc.* **89**, 691–692.
- Blankenship, Gilmer L. and Papanicolaou, George C. (1978). Stability and control of stochastic systems with wide band noise disturbances I. *SIAM J. Appl. Math.* **34**, 437–476.
- Borovkov, A. A. (1970). Theorems on the convergence to Markov diffusion processes. *Z. Wahrsch. verw. Gebiete* **16**, 47–76.
- Breiman, Leo (1968). *Probability*. Addison-Wesley, Reading, Mass.
- Brown, Bruce M. (1971). Martingale central limit theorems. *Ann. Math. Statist.* **42**, 59–66.
- Brown, Timothy C. (1978). A martingale approach to the Poisson convergence of simple point processes. *Ann. Probab.* **6**, 615–628.
- Castaing, Charles (1967). Sur les multi-applications mesurables. *Rev. Francaise Inf. Rech. Opéra.* **1**, 91–126.
- Chenčov, N. N. (1956). Weak convergence of stochastic processes whose trajectories have no discontinuities of the second kind and the heuristic approach to the Kolmogorov-Smirnov tests. *Theory Probab. Appl.* **1**, 140–149.
- Chernoff, Herman (1956). Large sample theory: parametric case. *Ann. Math. Statist.* **27**, 1–22.
- Chernoff, Paul R. (1968). Note on product formulas for operator semigroups. *J. Funct. Anal.* **2**, 238–242.
- Chow, Yuah Shih (1960). Martingales in a σ -finite measure space indexed by directed sets. *Trans. Amer. Math. Soc.* **97**, 254–285.
- Chung, Kai Lai and Williams, Ruth J. (1983). *Introduction to Stochastic Integration*. Birkhauser, Boston.
- Cohn, Donald L. (1980). *Measure Theory*. Birkhauser, Boston.
- Costantini, Cristina, Gerardi, Anna and Nappo, Giovanna (1982). On the convergence of sequences of stationary jump Markov processes. *Statist. Probab. Lett.* **1**, 155–160.
- Costantini, Cristina and Nappo, Giovanna (1982). Some results on weak convergence of jump Markov processes and their stability properties. *Systems Control Lett.* **2**, 175–183.
- Courrège, Philippe (1963). Intégrales stochastiques et martingales de carré intégrable. Séminaire Brelot-Choquet-Deny, 7th year.
- Cox, E. Theodore, and Rösler, Uwe (1982). A duality relation for entrance and exit laws for Markov processes. *Stochastic Process. Appl.* **16**, 141–151.
- Crandall, Michael G. and Liggett, Thomas M. (1971). Generation of semi-groups of nonlinear transformations on general Banach spaces. *Amer. J. Math.* **93**, 265–298.
- Csörgő, Miklós and Révész, Pál (1981). *Strong Approximations in Probability and Statistics*. Academic, New York.
- Darden, Thomas and Kurtz, Thomas G. (1986). Nearly deterministic Markov processes near a stable point. To appear.
- Davies, Edward Brian (1980). *One-Parameter Semigroups*. Academic, London.
- Davydov, Yu. A. (1968). Convergence of distributions generated by stationary stochastic processes. *Theory Probab. Appl.* **13**, 691–696.
- Dawson, Donald A. (1975). Stochastic evolution equations and related measure processes. *J. Multivar. Anal.* **5**, 1–52.

- Dawson, Donald A. (1977). The critical measure diffusion process. *Z. Wahrsch. verw. Gebiete* **40**, 125–145.
- Dawson, Donald A. (1979). Stochastic measure diffusion processes. *Canad. Math. Bull.* **22**, 129–138.
- Dawson, Donald A. and Hochberg, Kenneth J. (1979). The carrying dimension of a measure diffusion process. *Ann. Probab.* **7**, 693–703.
- Dawson, Donald A. and Kurtz, Thomas G. (1982). Applications of duality to measure-valued processes. *Advances in Filtering and Optimal Stochastic Control. Lect. Notes Cont. Inf. Sci.* **42**, Springer-Verlag, Berlin, pp. 177–191.
- Dellacherie, Claude and Meyer, Paul-André (1978). *Probabilities and Potential*. North-Holland, Amsterdam.
- Dellacherie, Claude and Meyer, Paul-André (1982). *Probabilities and Potential B*. North-Holland, Amsterdam.
- Doleans-Dade, Catherine (1969). Variation quadratique des martingales continues à droite. *Ann. Math. Statist.* **40**, 284–289.
- Doleans-Dade, Catherine and Meyer, Paul-André (1970). Intégrales stochastiques par rapport aux martingales locales. *Séminaire de Probabilités IV, Lect. Notes Math.*, **124**, Springer-Verlag, Berlin.
- Donsker, Monroe D. (1951). An invariance principle for certain probability limit theorems. *Mem. Amer. Math. Soc.* **6**.
- Doob, Joseph L. (1953). *Stochastic Processes*. Wiley, New York.
- Dudley, Richard M. (1968). Distances of probability measures and random variables. *Ann. Math. Statist.* **39**, 1563–1572.
- Dunford, Nelson and Schwartz, Jacob T. (1957). *Linear Operators Part I: General Theory*. Wiley-Interscience, New York.
- Dvoretzky, Aryeh (1972). Asymptotic normality for sums of dependent random variables. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **2**, University of California Press, pp. 513–535.
- Dynkin, E. B. (1961). *Theory of Markov Processes*. Prentice-Hall, Englewood Cliffs, N.J.
- Dynkin, E. B. (1965). *Markov Processes I, II*. Springer-Verlag, Berlin.
- Echeverria, Pedro E. (1982). A criterion for invariant measures of Markov processes. *Z. Wahrsch. verw. Gebiete* **61**, 1–16.
- Elliot, Robert J. (1982). *Stochastic Calculus and Applications*. Springer-Verlag, New York.
- Ethier, Stewart N. (1976). A class of degenerate diffusion processes occurring in population genetics. *Comm. Pure Appl. Math.* **29**, 483–493.
- Ethier, Stewart N. (1978). Differentiability preserving properties of Markov semigroups associated with one-dimensional diffusions. *Z. Wahrsch. verw. Gebiete* **45**, 225–238.
- Ethier, Stewart N. (1979). Limit theorems for absorption times of genetic models. *Ann. Probab.* **7**, 622–738.
- Ethier, Stewart N. (1981). A class of infinite-dimensional diffusions occurring in population genetics. *Indiana Univ. Math. J.* **30**, 925–935.
- Ethier, Stewart N. and Kurtz, Thomas G. (1981). The infinitely-many-neutral-alleles diffusion model. *Adv. Appl. Probab.* **13**, 429–452.
- Ethier, Stewart N. and Kurtz, Thomas G. (1986). The infinitely-many-alleles model with selection as a measure-valued diffusion. To appear.
- Ethier, Stewart N. and Nagylaki, Thomas (1980). Diffusion approximation of Markov chains with two time scales and applications to population genetics. *Adv. Appl. Probab.* **12**, 14–49.
- Ewens, Warren J. (1972). The sampling theory of selectively neutral alleles. *Theor. Pop. Biol.* **3**, 87–112.
- Ewens, Warren J. (1979). *Mathematical Population Genetics*. Springer-Verlag, Berlin.

- Feller, William (1951). Diffusion processes in genetics. *Proc. Second Berkeley Symp. Math. Statist. Prob.*, University of California Press, Berkeley, pp. 227–246.
- Feller, William (1952). The parabolic differential equations and the associated semi-groups of transformations. *Ann. Math.* **55**, 468–519.
- Feller, William (1953). On the generation of unbounded semi-groups of bounded linear operators. *Ann. Math.* **58**, 166–174.
- Feller, William (1971). *An Introduction to Probability Theory and Its Applications II*, 2nd ed., Wiley, New York.
- Fiorenza, Renato (1959). Sui problemi di derivata obliqua per le equazioni ellittiche. *Ric. Mat.* **8**, 83–110.
- Fisher, R. A. (1922). On the dominance ratio. *Proc. Roy. Soc. Edin.* **42**, 321–341.
- Fleming, Wendall H. and Viot, Michel (1979). Some measure-valued Markov processes in population genetics theory. *Indiana Univ. Math. J.* **28**, 817–843.
- Freidlin, M. I. (1968). On the factorization of non-negative definite matrices. *Theory Probab. Appl.* **13**, 354–356.
- Friedman, Avner (1975). *Stochastic Differential Equations I*. Academic, New York.
- Gänsler, Peter and Häusler, Erich (1979). Remarks on the functional central limit theorem for martingales. *Z. Wahrsch. verw. Gebiete* **50**, 237–243.
- Gihman, I. I. and Skorohod, A. V. (1969). *Introduction to the Theory of Random Processes*. W. B. Saunders Co., Philadelphia.
- Gihman, I. I. and Skorohod, A. V. (1972). *Stochastic Differential Equations*. Springer-Verlag, Berlin.
- Gihman, I. I. and Skorohod, A. V. (1974). *The Theory of Stochastic Processes I*. Springer-Verlag, Berlin.
- Goldstein, Jerome A. (1976). Semigroup-theoretic proofs of the central limit theorem and other theorems of analysis. *Semigroup Forum* **12**, 189–206.
- Goldstein, Sydney (1951). On diffusion by discontinuous movements, and on the telegraph equation. *Quart. J. Mech. Appl. Math.* **4**, 129–156.
- Gray, Lawrence and Griffeath, David (1977a). On the uniqueness and nonuniqueness of proximity processes. *Ann. Probab.* **5**, 678–692.
- Gray, Lawrence and Griffeath, David (1977b). Unpublished manuscript.
- Griego, Richard J. and Hersh, Reuben (1969). Random evolutions, Markov chains, and systems of partial differential equations. *Proc. Nat. Acad. Sci. USA* **62**, 305–308.
- Griego, Richard J. and Hersh, Reuben (1971). Theory of random evolutions with applications to partial differential equations. *Trans. Amer. Math. Soc.* **156**, 405–418.
- Grigelionis, Bronius and Mikulevičius, R. (1981). On the weak convergence of random point processes. *Lithuanian Math. Trans.* **21**, 49–55.
- Grimvall, Anders (1974). On the convergence of sequences of branching processes. *Ann. Probab.* **2**, 1027–1045.
- Guess, Harry A. (1973). On the weak convergence of Wright–Fisher models. *Stochastic Process. Appl.* **1**, 287–306.
- Gustafson, Karl (1966). A perturbation lemma. *Bull. Amer. Math. Soc.* **72**, 334–338.
- Hall, P. (1935). On representatives of subsets. *J. London Math. Soc.* **10**, 26–30.
- Hall, Peter and Heyde, C. C. (1980). *Martingale Limit Theory and Its Applications*. Academic, New York.
- Hardin, Clyde (1985). A spurious Brownian motion. *Proc. Amer. Math. Soc.* **93**, 350.
- Harris, Theodore E. (1976). On a class of set-valued Markov processes. *Ann. Probab.* **4**, 175–194.

- Hasegawa, Minoru (1964). A note on the convergence of semi-groups of operators. *Proc. Japan Acad.* **40**, 262–266.
- Helland, Inge S. (1978). Continuity of a class of random time transformations. *Stochastic Process. Appl.* **7**, 79–99.
- Helland, Inge S. (1981). Minimal conditions for weak convergence to a diffusion process on the line. *Ann. Probab.* **9**, 429–452.
- Helland, Inge S. (1982). Central limit theorems for martingales with discrete or continuous time. *Scand. J. Statist.* **9**, 79–94.
- Helms, Lester L. (1974). Ergodic properties of several interacting Poisson particles. *Adv. Math.* **12**, 32–57.
- Hersh, Reuben (1974). Random evolutions: a survey of results and problems. *Rocky Mt. J. Math.* **4**, 443–477.
- Hersh, Reuben and Papanicolaou, George C. (1972). Non-commuting random evolutions, and an operator-valued Feynman–Kac formula. *Comm. Pure Appl. Math.* **25**, 337–367.
- Heyde, C. C. (1974). On the central limit theorem for stationary processes. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **30**, 315–320.
- Hille, Einar (1948). *Functional Analysis and Semi-groups*, Am. Math. Soc. Colloq. Publ. **31**, New York.
- Hille, Einar and Phillips, Ralph (1957). *Functional Analysis and Semi-groups*, rev. ed., Amer. Math. Soc. Colloq. Publ. **31**, Providence, R.I.
- Himmelberg, C. J. (1975). Measurable relations. *Fund. Math.* **87**, 53–72.
- Holley, Richard A. and Liggett, Thomas M. (1975). Ergodic theorems for weakly interacting systems and the voter model. *Ann. Probab.* **3**, 643–663.
- Holley, Richard A. and Stroock, Daniel W. (1976). A martingale approach to infinite systems of interacting processes. *Ann. Probab.* **4**, 195–228.
- Holley, Richard A. and Stroock, Daniel W. (1978). Generalized Ornstein–Uhlenbeck processes and infinite particle branching Brownian motions. *Publ. RIMS, Kyoto Univ.* **14**, 741–788.
- Holley, Richard A. and Stroock, Daniel W. (1979). Dual processes and their applications to infinite interacting systems. *Adv. Math.* **32**, 149–174.
- Holley, Richard A., Stroock, Daniel W., and Williams, David (1977). Applications of dual processes to diffusion theory. *Proc. Symp. Pure Math.* **31**, AMS, Providence, R.I., pp. 23–36.
- Ibragimov, I. A. (1959). Some limit theorems for strict-sense stationary stochastic processes. *Dokl. Akad. Nauk SSSR* **125**, 711–714.
- Ibragimov, I. A. (1962). Some limit theorems for stationary processes. *Theory Probab. Appl.* **7**, 349–382.
- Ibragimov, I. A. and Linnik, Yu. V. (1971). *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff, Groningen.
- Ikeda, Nobuyuki, Nagasawa, Masao, and Watanabe, Shinzo (1968, 1969). Branching Markov processes I, II, III. *J. Math. Kyoto* **8**, 233–278, 365–410; **9**, 95–160.
- Ikeda, Nobuyuki and Watanabe, Shinzo (1981). *Stochastic Differential Equations and Diffusion Processes*. North Holland, Amsterdam.
- Il'in, A. M., Kalashnikov, A. S., and Oleinik, O. A. (1962). Linear equations of the second order of parabolic type. *Russ. Math. Surveys* **17**, 1–143.
- Itô, Kiyosi (1951). On stochastic differential equations. *Mem. Amer. Math. Soc.* **4**.
- Itô, Kiyosi and Watanabe, Shinzo (1965). Transformations of Markov processes by multiplicative functionals. *Ann. Inst. Fourier* **15**, 15–30.
- Jacob, Jean, Mémin, Jean, and Métivier, Michel (1983). On tightness and stopping times. *Stochastic Process. Appl.* **14**, 109–146.

- Jagers, Peter (1971). Diffusion approximations of branching processes. *Ann. Math. Statist.* **42**, 2074-2078.
- Jitina, Miloslav (1969). On Feller's branching diffusion processes. *Časopis Pěst. Mat.* **94**, 84-90.
- Joffe, Anatole and Metivier, Michel (1984). Weak convergence of sequences of semimartingales with applications to multitype branching processes. Tech. Rep., Université de Montréal.
- Kabanov, Yu. M., Lipster, R. Sh., and Shiryaev, A. N. (1980). Some limit theorems for simple point processes (a martingale approach). *Stochastics* **3**, 203-216.
- Kac, Mark (1956). Some stochastic problems in physics and mathematics. Magnolia Petroleum Co. Colloq. Lect. **2**.
- Kac, Mark (1974). A stochastic model related to the telegrapher's equation. *Rocky Mt. J. Math.* **4**, 497-509.
- Kallman, Robert R. and Rota, Gian-Carlo (1970). On the inequality $\|f'\|^2 \leq 4\|f\| \cdot \|f''\|$. *Inequalities, Vol. II*, Oved Shisha, Ed. Academic, New York, pp. 187-192.
- Karlin, Samuel and Levikson, Benny (1974). Temporal fluctuations in selection intensities: Case of small population size. *Theor. Pop. Biol.* **6**, 383-412.
- Kato, Tosio (1966). *Perturbation Theory for Linear Operators*. Springer-Verlag, New York.
- Keiding, Niels (1975). Extinction and exponential growth in random environments. *Theor. Pop. Biol.* **8**, 49-63.
- Kertz, Robert P. (1974). Perturbed semigroup limit theorems with applications to discontinuous random evolutions. *Trans. Amer. Math. Soc.* **199**, 29-53.
- Kertz, Robert P. (1978). Limit theorems for semigroups with perturbed generators, with applications to multiscaled random evolutions. *J. Funct. Anal.* **27**, 215-233.
- Khas'minskii, R. Z. (1960). Ergodic properties of recurrent diffusion processes and stabilization of the solution of the Cauchy problem for parabolic equations. *Theory Probab. Appl.* **5**, 179-196.
- Khas'minskii, R. Z. (1966). A limit theorem for the solutions of differential equations with random right-hand sides. *Theory Probab. Appl.* **11**, 390-406.
- Khas'minskii, R. Z. (1980). *Stochastic Stability of Differential Equations*. Sijthoff and Nordhoff.
- Khintchine, A. (1933). *Asymptotische Gesetze der Wahrscheinlichkeitsrechnung*. Springer-Verlag, Berlin.
- Kimura, Motoo and Crow, James F. (1964). The number of alleles that can be maintained in a finite population. *Genetics* **49**, 725-738.
- Kingman, J. F. C. (1967). Completely random measures. *Pacific J. Math.* **21**, 59-78.
- Kingman, J. F. C. (1975). Random discrete distributions. *J. R. Statist. Soc. B* **37**, 1-22.
- Kingman, J. F. C. (1977). The population structure associated with the Ewens sampling formula. *Theor. Pop. Biol.* **11**, 274-283.
- Kingman, J. F. C. (1980). *Mathematics of Genetic Diversity*. CBMS-NSF Regional Conf. Series in Appl. Math., **34**. SIAM, Philadelphia.
- Kolmogorov, A. N. (1956). On Skorohod convergence. *Theory Probab. Appl.* **1**, 213-222.
- Komlós, János, Major, Peter, and Tusnády, Gábor (1975, 1976). An approximation of partial sums of independent RV's and the sample DF I, II. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **32**, 111-131; **34**, 33-58.
- Krylov, N. V. (1973). The selection of a Markov process from a system of processes and the construction of quasi-diffusion processes. *Math. USSR Izvestia* **7**, 691-709.
- Kunita, Hiroshi and Watanabe, Shinzo (1967). On square integrable martingales. *Nagoya Math. J.* **30**, 209-245.
- Kuratowski, K. and Ryll-Nardzewski, C. (1965). A general theorem on selectors. *Bull. Acad. Polon. Sci. Ser. Sci. Math. Astronom. Phys.* **13**, 397-403.

- Kurtz, Thomas G. (1969). Extensions of Trotter's operator semigroup approximation theorems. *J. Funct. Anal.* 3, 354–375.
- Kurtz, Thomas G. (1970a). A general theorem on the convergence of operator semigroups. *Trans. Amer. Math. Soc.* 148, 23–32.
- Kurtz, Thomas G. (1970b). Solutions of ordinary differential equations as limits of pure jump Markov processes. *J. Appl. Probab.* 7, 49–58.
- Kurtz, Thomas G. (1971). Limit theorems for sequences of jump Markov processes approximating ordinary differential processes. *J. Appl. Probab.* 8, 344–356.
- Kurtz, Thomas G. (1973). A limit theorem for perturbed operator semigroups with applications to random evolutions. *J. Funct. Anal.* 12, 55–67.
- Kurtz, Thomas G. (1975). Semigroups of conditioned shifts and approximation of Markov processes. *Ann. Probab.* 3, 618–642.
- Kurtz, Thomas G. (1977). Applications of an abstract perturbation theorem to ordinary differential equations. *Houston J. Math.* 3, 67–82.
- Kurtz, Thomas G. (1978a). Strong approximation theorems for density dependent Markov chains. *Stochastic Process. Appl.* 6, 223–240.
- Kurtz, Thomas G. (1978b). Diffusion approximations for branching processes. *Branching Processes, Adv. Prob.* 5, Anatole Joffe and Peter Ney, Eds., Marcel Dekker, New York, pp. 262–292.
- Kurtz, Thomas G. (1980a). Representation of Markov processes as multiparameter time changes. *Ann. Probab.* 8, 682–715.
- Kurtz, Thomas G. (1980b). The optional sampling theorem for martingales indexed by directed sets. *Ann. Probab.* 8, 675–681.
- Kurtz, Thomas G. (1981a). *Approximation of Population Processes. CBMS-NSF Regional Conf. Series in Appl. Math.* 36, SIAM, Philadelphia.
- Kurtz, Thomas G. (1981b). The central limit theorem for Markov chains. *Ann. Probab.* 9, 557–560.
- Kurtz, Thomas G. (1981c). Approximation of discontinuous processes by continuous processes. *Stochastic Nonlinear Systems*, L. Arnold and R. Lefever, Eds. Springer-Verlag, Berlin, pp. 22–35.
- Kurtz, Thomas G. (1982). Representation and approximation of counting processes. *Advances in Filtering and Optimal Stochastic Control, Lect. Notes Cont. Inf. Sci.* 42, W. H. Fleming and L. G. Gorostiza, Eds., Springer-Verlag, Berlin.
- Kushner, Harold J. (1974). On the weak convergence of interpolated Markov chains to a diffusion. *Ann. Probab.* 2, 40–50.
- Kushner, Harold J. (1979). Jump-diffusion approximation for ordinary differential equations with wide-band random right hand sides. *SIAM J. Control Optim.* 17, 729–744.
- Kushner, Harold J. (1980). A martingale method for the convergence of a sequence of processes to a jump-diffusion process on the line. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 53, 207–219.
- Kushner, Harold J. (1982). Asymptotic distributions of solutions of ordinary differential equations with wide band noise inputs; approximate invariant measures. *Stochastics* 6, 259–277.
- Kushner, Harold J. (1984). *Approximation and Weak Convergence Methods for Random Processes*. MIT Press, Cambridge, Massachusetts.
- Ladyzhenskaya, O. A. and Ural'tseva, N. N. (1968). *Linear and Quasilinear Elliptic Partial Differential Equations*. Academic, New York.
- Lamperti, John (1967a). The limit of a sequence of branching processes. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 7, 271–288.
- Lamperti, John (1967b). On random time substitutions and the Feller property. *Markov Processes and Potential Theory*, Joshua Chover, Ed., Wiley, New York, pp. 87–101.
- Lamperti, John (1977). *Stochastic Processes*. Springer-Verlag, New York.

- Lévy, Paul (1948). *Processus Stochastique et Mouvement Brownien*. Gauthier-Villars, Paris.
- Liggett, Thomas M. (1972). Existence theorems for infinite particle systems. *Trans. Amer. Math. Soc.* **165**, 471–481.
- Liggett, Thomas M. (1977). The stochastic evolution of infinite systems of interacting particles. *Lect. Notes Math.* **598**, 187–248. Springer-Verlag, New York.
- Liggett, Thomas M. (1985). *Interacting Particle Systems*. Springer-Verlag, New York.
- Lindvall, Torgny (1972). Convergence of critical Galton-Watson branching processes. *J. Appl. Probab.* **9**, 445–450.
- Lindvall, Torgny (1974). Limit theorems for some functionals of certain Galton-Watson branching processes. *Adv. Appl. Probab.* **6**, 309–321.
- Littler, Raymond A. (1972). Multidimensional stochastic models in genetics. Ph.D thesis, Monash Univ.
- Littler, Raymond A. and Good, A. J. (1978). Ages, extinction times, and first passage probabilities for a multiallele diffusion model with irreversible mutation. *Theor. Pop. Biol.* **13**, 214–225.
- Mackevičius, V. (1974). On the question of the weak convergence of random processes in the space $D[0, \infty)$. *Lithuanian Math. Trans.* **14**, 620–623.
- Maigret, Nelly (1978). Théorème de limite centrale fonctionnel pour une chaîne de Markov récurrente au sens de Harris et positive. *Ann. Inst. Henri Poincaré* **14**, 425–440.
- Major, Peter (1976). The approximation of partial sums of independent RV's. *Z. Wahrsch. verw. Gebiete* **35**, 213–220.
- Malek-Mansour, M., Van Den Broeck, C., Nicolis, G., and Turner, J. W. (1981). Asymptotic properties of Markovian master equations. *Ann. Phys.* **131**, 283–313.
- Mann, Henry B. and Wald, Abraham (1943). On stochastic limit and order relations. *Ann. Math. Statist.* **14**, 217–226.
- Mandl, Petr (1968). *Analytical Treatment of One-Dimensional Markov Processes*. Springer-Verlag, Berlin.
- McKean, Henry P., Jr. (1969). *Stochastic Integrals*. Academic, New York.
- McLeish, D. L. (1974). Dependent central limit theorems and invariance principles. *Ann. Probab.* **2**, 608–619.
- Métivier, Michel (1982). *Semimartingales: A Course on Stochastic Processes*. Walter de Gruyter, Berlin.
- Meyer, Paul-André (1966). *Probability and Potentials*. Blaisdell, Waltham, Mass.
- Meyer, Paul-André (1967). Intégrales stochastiques I, II, III, IV. *Séminaire de Probabilités I. Lect. Notes Math.* **39**, 72–162.
- Meyer, Paul-André (1968). Guide détaillé de la théorie générale des processus. *Séminaire de Probabilités II. Lect. Notes Math.* **51**, 140–165.
- Mirando, Carlo (1970). *Partial Differential Equations of Elliptic Type*. Springer-Verlag, Berlin.
- Moran, P. A. P. (1958a). A general theory of the distribution of gene frequencies I. Overlapping generations. *Proc. Roy. Soc. London B* **149**, 102–111.
- Moran, P. A. P. (1958b). A general theory of the distribution of gene frequencies. II. Non-overlapping generations. *Proc. Roy. Soc. London B* **149**, 113–116.
- Moran, P. A. P. (1958c). Random processes in genetics. *Proc. Camb. Phil. Soc.* **54**, 60–71.
- Morando, Philippe (1969). Mesures aléatoires. *Séminaire de Probabilités III, Lect. Notes Math.* **88**, 190–229. Springer-Verlag, Berlin.
- Morkvenas, R. (1974). Convergence of Markov chains to solution of martingale problem. *Lithuanian Math. Trans.* **14**, 460–466.
- Nagylaki, Thomas (1980). The strong-migration limit in geographically structured populations. *J. Math. Biol.* **9**, 101–114.

- Nagylaki, Thomas (1982). Geographical invariance in population genetics. *J. Theor. Biol.* **99**, 159–172.
- Neveu, J. (1958). Théorie des semi-groups de Markov. *Univ. California Publ. Statist.* **2**, 319–394.
- Norman, M. Frank (1971). Slow learning with small drift in two-absorbing-barrier models. *J. Math. Psych.* **8**, 1–21.
- Norman, M. Frank (1972). *Markov Processes and Learning Models*. Academic, New York.
- Norman, M. Frank (1974). A central limit theorem for Markov processes that move by small steps. *Ann. Probab.* **2**, 1065–1074.
- Norman, M. Frank (1975a). Diffusion approximation of non-Markovian processes. *Ann. Probab.* **3**, 358–364.
- Norman, M. Frank (1975b). Limit theorems for stationary distributions. *Adv. Appl. Probab.* **7**, 561–575.
- Norman, M. Frank (1977). Ergodicity of diffusion and temporal uniformity of diffusion approximation. *J. Appl. Probab.* **14**, 399–404.
- Ohta, Tomoko and Kimura, Motoo (1969). Linkage disequilibrium due to random genetic drift. *Genet. Res. Camb.* **13**, 47–55.
- Ohta, Tomoko and Kimura, Motoo (1973). A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genet. Res. Camb.* **22**, 201–204.
- Oleinik, Olga A. (1966). Alcuni risultati sulle equazioni lineari e quasi lineari ellittico-paraboliche a derivate parziali del secondo ordine. *Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur.* (8) **40**, 775–784.
- Papanicolaou, George C. and Kohler, W. (1974). Asymptotic theory of mixing stochastic ordinary differential equations. *Comm. Pure Appl. Math.* **27**, 641–668.
- Papanicolaou, George C., Stroock, Daniel W., and Varadhan, S. R. S. (1977). Martingale approach to some limit theorems. *Conference on Statistical Mechanics, Dynamical Systems, and Turbulence*, Duke University, M. Reed, Ed., Duke Univ. Math. Series 3.
- Papanicolaou, George C. and Varadhan, S. R. S. (1973). A limit theorem with strong mixing in Banach space and two applications to stochastic differential equations. *Comm. Pure Appl. Math.* **26**, 497–524.
- Parthasarathy, K. R. (1967). *Probability Measures on Metric Spaces*. Academic, New York.
- Pazy, Amnon (1983). *Semigroups of Linear Operators and Applications to Partial Differential Equations*. Springer-Verlag, New York.
- Peligrad, Magda (1982). Invariance principles for mixing sequences of random variables. *Ann. Probab.* **10**, 968–981.
- Phillips, Ralph S. and Sarason, Leonard (1968). Elliptic-parabolic equations of the second order. *J. Math. Mech.* **17**, 891–917.
- Pinsky, Mark A. (1968). Differential equations with a small parameter and the central limit theorem for functions defined on a finite Markov chain. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **9**, 101–111.
- Pinsky, Mark A. (1974). Multiplicative operator functionals and their asymptotic properties. *Advances in Probability* **3**, Dekker, New York.
- Priouret, P. (1974). Processus de diffusion et équations différentielles stochastiques. *Lect. Notes Math.* **390**. Springer-Verlag, Berlin.
- Prohorov, Yu. V. (1956). Convergence of random processes and limit theorems in probability theory. *Theory Probab. Appl.* **1**, 157–214.
- Rao, C. Radhakrishna (1973). *Linear Statistical Inference and Its Applications*, 2nd ed. Wiley, New York.
- Rebolledo, Rolando (1979). La méthode des martingales appliquée à l'étude de la convergence en loi de processus. *Bull. Soc. Math. France Mem.*, **62**.

- Rebolledo, Rolando (1980). Central limit theorems for local martingales. *Z. Wahrsch. verw. Gebiete* **51**, 269–286.
- Rishel, Raymond (1970). Necessary and sufficient dynamic programming conditions for continuous time stochastic control. *SIAM J. Control* **8**, 559–571.
- Rootzén, Holger (1977). On the functional central limit theorem for martingales. *Z. Wahrsch. verw. Gebiete* **38**, 199–210.
- Rootzén, Holger (1980). On the functional central limit theorem for martingales II. *Z. Wahrsch. verw. Gebiete* **51**, 79–93.
- Rosenblatt, Murray (1956). A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci. U.S.A.* **42**, 43–47.
- Rosenkrantz, Walter A. (1974). Convergent family of diffusion processes whose diffusion coefficients diverge. *Bull. Amer. Math. Soc.* **80**, 973–976.
- Rosenkrantz, Walter A. (1975). Limit theorems for solutions to a class of stochastic differential equations. *Indiana Univ. Math. J.* **24**, 613–625.
- Rosenkrantz, Walter A. and Dorea, C. C. Y. (1980). Limit theorems for Markov processes via a variant of the Trotter-Kato theorem. *J. Appl. Probab.* **17**, 704–715.
- Roth, Jean-Pierre (1976). Opérateurs dissipatifs et semigroups dans les espaces de fonctions continues. *Ann. Inst. Fourier, Grenoble* **26**, 1–97.
- Roth, Jean-Pierre (1977). Opérateurs elliptiques comme générateurs infinitésimaux de semi-groupes de Feller. *C. R. Acad. Sci. Paris* **284**, 755–757.
- Rudin, Walter (1973). *Functional Analysis*. McGraw-Hill, New York.
- Rudin, Walter (1974). *Real and Complex Analysis*, 2nd Ed. McGraw-Hill, New York.
- Sato, Ken-iti (1976). Diffusion processes and a class of Markov chains related to population genetics. *Osaka J. Math.* **13**, 631–659.
- Sato, Ken-iti (1977). A note on convergence of probability measures on C and D. *Ann. Sci. Kanazawa Univ.* **14**, 1–5.
- Schauder, J. (1934). Über lineare elliptische Differentialgleichungen zweiter Ordnung. *Math. Z.* **38**, 257–282.
- Schlögl, F. (1972). Chemical reaction models for non-equilibrium phase transitions. *Z. Physik* **253**, 147–161.
- Serant, Daniel and Villard, Michel (1972). Linearization of crossing-over and mutation in a finite random-mating population. *Theor. Pop. Biol.* **3**, 249–257.
- Shiga, Tokuzo (1980). An interacting system in population genetics. *J. Math. Kyoto Univ.* **20**, 213–242.
- Shiga, Tokuzo (1981). Diffusion processes in population genetics. *J. Math. Kyoto Univ.* **21**, 133–151.
- Shiga, Tokuzo (1982). Wandering phenomena in infinite allelic diffusion models. *Adv. Appl. Probab.* **14**, 457–483.
- Siegmund, David (1976). The equivalence of absorbing and reflecting barrier problems for stochastically monotone Markov processes. *Ann. Probab.* **4**, 914–924.
- Skorohod, A. V. (1956). Limit theorems for stochastic processes. *Theory Probab. Appl.* **1**, 261–290.
- Skorohod, A. V. (1958). Limit theorems for Markov processes. *Theory Probab. Appl.* **3**, 202–246.
- Skorohod, A. V. (1965). *Studies in the Theory of Random Processes*. Addison-Wesley, Reading, Mass.
- Sova, Miroslav (1967). Convergence d'opérations linéaires non bornées. *Rev. Roumaine Math. Pures Appliq.* **12**, 373–389.
- Spitzer, Frank (1970). Interaction of Markov processes. *Adv. Math.* **5**, 246–290.

- Stone, Charles (1963). Weak convergence of stochastic processes defined on a semi-infinite time interval. *Proc. Amer. Math. Soc.* **14**, 694–696.
- Stratonovich, R. L. (1963, 1967). *Topics in the Theory of Random Noise I, II*. Gordon and Breach, New York.
- Strassen, Volker (1965). The existence of probability measures with given marginals. *Ann. Math. Statist.* **36**, 423–439.
- Stroock, Daniel W. (1975). Diffusion processes associated with Levy generators. *Z. Wahrsch. verw. Gebiete* **32**, 209–244.
- Stroock, Daniel W. and Varadhan, S. R. S. (1969). Diffusion processes with continuous coefficients I, II. *Comm. Pure Appl. Math.* **22**, 345–400, 479–530.
- Stroock, Daniel W. and Varadhan, S. R. S. (1971). Diffusion processes with boundary conditions. *Comm. Pure Appl. Math.* **24**, 147–225.
- Stroock, Daniel W. and Varadhan, S. R. S. (1972). On the support of diffusion processes with applications to the strong maximum principle. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **3**, 333–359.
- Stroock, Daniel W. and Varadhan, S. R. S. (1979). *Multidimensional Diffusion Processes*. Springer-Verlag, Berlin.
- Trotter, Hale F. (1958). Approximation of semi-groups of operators. *Pacific J. Math.* **8**, 887–919.
- Trotter, Hale F. (1959). On the product of semi-groups of operators. *Proc. Amer. Math. Soc.* **10**, 545–551.
- Vasershtain, L. N. (1969). Markov processes over denumerable products of spaces describing large systems of automata. *Probl. Peredachi Inform.* **5**(3), 64–73.
- Vasershtain, L. N. and Leontovich, A. M. (1970). Invariant measures of certain Markov operators describing a homogeneous random medium. *Probl. Peredachi Inform.* **6**(1), 71–80.
- Volkonskii, V. A. (1958). Random substitution of time in strong Markov processes. *Theory Probab. Appl.* **3**, 310–326.
- Volkonskii, V. A. and Rozanov, Yu. A. (1959). Some limit theorems for random functions I. *Theory Probab. Appl.* **4**, 178–197.
- Wagner, Daniel H. (1977). Survey of measurable selection theorems. *SIAM J. Control Optim.* **15**, 859–903.
- Wang, Frank J. S. (1977). A central limit theorem for age and density dependent population processes. *Stochastic Process. Appl.* **5**, 173–193.
- Wang, Frank J. S. (1982a). Probabilities of extinction of multiplicative measure diffusion processes with absorbing boundary. *Indiana Univ. Math. J.* **31**, 97–107.
- Wang, Frank J. S. (1982b). Diffusion approximations of age-and-position dependent branching processes. *Stochastic Process. Appl.* **13**, 59–74.
- Watanabe, Shinzo (1964). Additive functionals of Markov processes and Lévy systems. *Japanese J. Math.* **34**, 53–70.
- Watanabe, Shinzo (1968). A limit theorem of branching processes and continuous state branching processes. *J. Math. Kyoto Univ.* **8**, 141–167.
- Watanabe, Shinzo (1971). On stochastic differential equations for multi-dimensional diffusion processes with boundary conditions. *J. Math. Kyoto Univ.* **11**, 169–180.
- Waterson, G. A. (1962). Some theoretical aspects of diffusion theory in population genetics. *Ann. Math. Statist.* **33**, 939–957.
- Waterson, G. A. (1964). The application of diffusion theory to two population genetic models of Moran. *J. Appl. Probab.* **1**, 233–246.
- Waterson, G. A. (1970). On the equivalence of random mating and random union of gametes in finite, monoecious populations. *Theor. Pop. Biol.* **1**, 233–250.

- Watterson, G. A. (1976). The stationary distribution of the infinitely-many neutral alleles diffusion model. *J. Appl. Probab.* **13**, 639–651.
- Watterson, G. A. and Guess, Harry A. (1977). Is the most frequent allele the oldest? *Theor. Pop. Biol.* **11**, 141–160.
- Weiss, Alan (1981). Invariant measures of diffusions in bounded domains. Ph.D. dissertation, New York University.
- Whitney, Hassler (1934). Analytic extensions of differentiable functions defined on closed sets. *Trans. Amer. Math. Soc.* **36**, 369–387.
- Williams, David (1979). *Diffusions, Markov Processes, and Martingales*. Wiley, New York.
- Withers, C. S. (1981). Central limit theorems for dependent variables. I. *Z. Wahrsch. verw. Gebiete* **57**, 509–534.
- Wonham, W. M. (1966). Lyapunov criteria for weak stochastic stability. *J. Differential Equations* **2**, 195–207.
- Wright, Sewall (1931). Evolution in Mendelian populations. *Genetics* **16**, 97–159.
- Wright, Sewall (1949). Adaptation and selection. *Genetics, Paleontology, and Evolution*, G. L. Jepson, E. Mayr, and G. G. Simpson, Eds. Princeton University Press, Princeton, pp. 365–389.
- Yamada, Toshio and Watanabe, Shinzo (1971). On the uniqueness of solutions of stochastic differential equations. *J. Math. Kyoto Univ.* **11**, 155–167.
- Yosida, Kosaku (1948). On the differentiability and the representation of one-parameter semi-groups of linear operators. *J. Math. Soc. Japan* **1**, 15–21.
- Yosida, Kosaku (1980). *Functional Analysis*, 6th ed. Springer-Verlag, Berlin.
- Zakai, Moshe (1969). A Lyapunov criterion for the existence of stationary probability distributions for systems perturbed by noise. *SIAM J. Control* **7**, 390–397.

INDEX

Note: * indicates definition

- À, 23*
- ./, 145*
- Abraham, 499
- Absorption probabilities, convergence, 420
- Absorption time, convergence in distribution, 419
- Adapted:
- to directed filtration, 86*
 - stochastic process, 50*
- Aldous, 154
- Alexandroff, 154
- Allain, 467
- Alm, 467
- Analytic sets, 506
- Anderson, 273
- Artstein, 154
- Athreya, 409
- B(E), 155*
- .n, 96*
- Barbour, 409, 467
- Bárfai, 359, 364, 464
- Beneš, 274
- Bhattacharya, 364
- Billingsley, 154, 364, 445, 504
- Bimeasure, 502
- Blackwell, 154
- Blankenship, 274
- Bochner integral, 473*
- Borovkov, 364
- Boundary classification, one-dimensional diffusion, 366, 382
- Boundary of set, 108*
- Bounded generator, 162, 222
- Bounded pointwise convergence, 111*, 495
- bp-closure, 111*, 496*
- bp-convergence, 111*
- bp-dense, 111*
- bp-lim, 495*
- Branching Markov process, 400
- generator, 402
- Branching process:
- Galton-Watson, 386*
 - in random environments, 396
 - two-type Markov, 392
- Brown, 364
- Brownian motion, 276*, 302, 359
- martingale characterization, 290, 338
 - strong approximation, 356
- C_{m,u}(Ω), 368
- Castaing, 506
- Č(E), 164*
- Č(E), 155*
- C_{t,0,∞} martingale problem, 186*
- Chapman-Kolmogorov property, 156*
- Chemical reaction model, 2, 454, 466
- Chenčov, 154
- Chernoff, 48, 154
- Chernoff inequality, 30
- Chernoff product formula, 32
- Chow, 94
- Chung, 305
- Closed linear operator, 8*
- Closure of linear operator, 16*

- Compact containment condition**, 129*
 for Schlögl model, 4
 sufficient conditions, 203
- Compact sets**:
 in $D_E[0, \infty)$, 123
 in $\mathcal{A}(S)$, 104
- Conservative operator**, 166*
- Contact time**, 54*
- Continuous mapping theorem**, 103
 generalization, 151
- Convergence determining set**, 112*
 conditions for, 151
 counterexample, 151
 on product space, 115
- Convergence in distribution**, 108*
 for branching Markov processes, 406
 via convergence of generators, 2, 388, 393,
 406, 415, 428, 436, 475, 480, 484
 in $D_E[0, \infty)$, 127, 131, 143, 144
 for Feller processes, 167
 for Markov chains, 2, 3, 168, 173, 230, 233,
 236
 for Markov processes, 172, 230, 232, 236
 for measure-valued genetic model, 439
 to process in $C_E[0, \infty)$, 148
 using random time change equation, 3, 310,
 322, 323, 390, 397, 458
- Convergence in probability**, metric for, 60*, 90
- Core**, 17*
 conditions for, 17, 19
 examples, 3, 43, 365
 of generator, 17*
 for Laplacian, 276
- Costantini, 5, 274
- Courrège, 305
- Cox, 274
- Crandall, 47
- Crow, 451
- Csörgő, 364
- $D_E[0, \infty)$, 116*
 Borel sets in, 127
 compact sets in, 122
 completeness, 121
 modulus of continuity, 122*, 134
 separability, 121
- Darden, 467
- Davies, 47, 48
- Davydov, 364
- Dawson, 274, 409
- Dellacherie, 73, 75, 93, 94, 507
- $D_E[0, \infty)$ martingale problem, 186*
 uniqueness, 187
- Density dependent family**, 455*
 diffusion approximation, 460
 Gaussian approximation, 458
 law of large numbers, 456
- Diffusion approximation**:
 density dependent family, 460
 examples, 1, 360, 361
 Galton-Watson process, 388
 genotypic-frequency model, 426
 for Markov chain, 355, 428
 martingale problem, 354
 random evolution, 475, 480, 484
 random time change equation, 330
 strong approximation, 460
 for Wright-Fisher model, 363, 415
- Diffusion process**:
 absorbing boundary, 368
 boundary classification, 366, 382
 degenerate, 371
 generator, 366
 one-dimensional, 367
 in \mathbb{R}^4 , 370
 random time change equation, 328
 reflecting boundary, 369
 stochastic integral equation for, 290
- Discontinuities of functions in $D_E[0, \infty)$** :
 convergence of, 147
 countability of, 116
- Discrete stopping time**:
 approximation by, 86
 strong Markov property, 159
- Dissipativity**:
 of linear operator, 11*
 martingale problem, 178
 of multivalued linear operator, 21
 positive maximum principle, 165*
- Distribution**:
 convergence in, 108*
 of random variable, 107*
- Doob-Dade, 94, 305
- Dominated convergence theorem**, 492
- Donsker, 364
- Doob, 93, 94, 273, 305, 364, 478
- Doob inequality, 63, 64
- Doob-Meyer decomposition, 74*
- Driving process**, 469
- Duality**, 188*, 266
- Dubins, 154
- Dudley, 154
- Dunford, 494
- Dvoretzky, 364
- Dynkin, 47, 93, 273, 370, 385
- Dynkin class, 496*
- Dynkin class theorem, 497

- E¹, 165*
- Echeverria, 274
- Elliot, 75, 305
- Entrance time, 54*
- Epidemic model, 453, 466
- Equivalence of stochastic processes, 50*
- Ergodic theorem, 352
- Ethier, 44, 48, 274, 371, 372, 375, 385, 451
- Ewens, 451
- Ewens sampling formula, 447
- Exit time, 54*
- convergence in distribution, 419, 464
- ex-LIM, 34*
- ex-lim, 22*
- Extended limit, 22*
- generalized, 34
- $\varphi_n(\cdot | \mathcal{F}_t)$, 346
- \mathcal{F}_t^X , 50*
- Fatou's lemma, 492
- Feller, 116, 385, 409
- Feller process:
- continuity of sample paths, 171
 - convergence in distribution, 167
 - version in $D_{\mathbb{R}}[0, \infty)$, 169
- Feller semigroup, 166*
- for Brownian motion, 276
- Filtration, 50*
- complete, 50*
 - directed index set, 85
 - Markov with respect to, 156
 - right continuous, 50*
- Finite-dimensional distributions
- convergence, 225, 226
 - determined by semigroup, 161
 - for Markov process, 157
 - of stochastic process, 50*
- Fiorenza, 385
- Fisher, 451
- Fleming, 451
- Fleming-Viot model, 440, 450
- martingale problem, 436
- Forward equation, uniqueness, 251, 252
- Freidlin, 385
- Friedman, 305, 369
- Full generator, 23*, 261
- related martingales, 162
- Galton-Watson branching process, 386
- Gänsler, 364
- Generator:
- absorbing diffusion, 368
 - bounded, 162, 222
 - bounded perturbation, 256, 257, 261
- branching Markov process, 402
- core of, 17*
- d-dimensional diffusion, 370, 372, 373, 374, 375
- degenerate diffusion, 371, 372, 373, 374, 375, 408
- examples, 3, 43, 365
- extended limit, 22*, 34
- full, 23*
- Hille-Yosida theorem, 13
- independent increments process, 380
- infinite particle system, 381
- jump Markov process, 376
- Lévy process, 379
- nondegenerate diffusion, 366, 370
- one-dimensional diffusion, 366, 371
- perturbation, 37, 44
- properties, 9
- reflecting diffusion, 369
- resolvent of, 10*
- of semigroup, 8*
- uniqueness of semigroup, 15
- Yosida approximation of, 12*
- Genotypic-frequency model, 426
- Gerardi, 274
- Gihman, 305, 380
- Goldstein, 48, 491
- Good, 451
- Gray, 274
- Griego, 491
- Griffeath, 274
- Grigelionis, 364
- Grimvall, 409
- Gronwall's inequality, 498
- Guess, 451
- Gustafson, 48
- Hall, 154, 364
- Hardin, 364
- Hardy-Weinberg proportions, 412*
- deviation from, 431
- Harris, 273
- Hasagawa, 48
- Häusler, 364
- Helland, 274, 335, 364, 409
- Helms, 336
- Hersh, 491
- Heyde, 364
- Hille, 47, 48
- Hille-Yosida theorem, 13, 16
- for multivalued generators, 21
 - for positive semigroups on \mathbb{C} , 165
- Himmelberg, 506
- Hitting distribution, convergence, 420, 464

- Hochberg, 409
 Holley, 273, 274, 336
- Ibragimov, 364
 Ikeda, 305, 409
 Il'in, 385
Increasing process, 74*
Independent increments process, generator, 380
Indistinguishability of stochastic processes, 50*
Infinitely-many-allele model, 435
 convergence in distribution, 436
Infinite particle system, generator, 381
Infinitesimal generator (see *generator*)
Initial distribution, 157*
Integral, Banach space-valued, 8, 473
Integration by parts, 65
Invariance principle, 278
 for stationary sequence, 350
 strong, 356
Inverse function, measurability of, 506
Itô, 93, 305
Itô's formula, 287
- Jacod, 154
 Jagers, 409
Jensen's inequality, 55
 Jitina, 409
 Joffe, 409
Jump Markov process:
 construction, 162, 263
 examples, 452
 generator, 376
 random time change equation, 326, 455
- Kabanov, 364
 Kac, 491
 Kalashnikov, 385
 Kallman, 48
 Karlin, 450
 Kato, 48
 Keiding, 409
 Kertz, 48
 Khas'minskii, 274, 491
 Khintchine, 274
 Kimura, 451
 Kingman, 451, 502
 Kohler, 491
 Kolmogorov, 154
 Komlós, 356, 364, 459
 Krylov, 274
Krylov's theorem, 210
 Kunita, 305
 Kuratowski, 506
- Kurtz, 5, 47, 48, 94, 154, 273, 274, 336, 364,
 409, 451, 467, 491
 Kushner, 4, 274, 364, 491
- Ladyzhenskaya, 369, 385
 Lamperti, 335, 352, 409
 Leontovich, 273
 Levikson, 450
 Lévy, 273, 364
Lévy process, generator, 379
 Liggett, 47, 48, 273, 381, 385
 Lindvall, 154, 409
Linear operator, 8*
 closable, 16*
 closed, 8*
 closure of, 16*
 dissipative, 11*
 graph of, 8*
 multivalued, 20
 single-valued, 20
 Linnik, 364
 Lipster, 364
 Littler, 385, 451
 $L^2(\{M\})$, 280
Local martingale, 64*
 example, 90
see also Martingale
Local-martingale problem, 223*
Logistic growth model, 453, 466
Lyapunov function, 240
- McKean, 305
 Mackevičius, 273
 McLeish, 364
 Maigret, 364
 Major, 356, 364, 459
 Malek-Mansour, 5
 Mandl, 367, 385
 Mann, 154
Markov chain, 158*
 diffusion approximation, 355, 428
Markov process, 156*
 convergence in distribution, 172, 230, 232,
 236
 corresponding semigroup, 161
 sample paths in $C_{\epsilon}[0, \infty)$, 264, 265
 sample paths in $D_{\epsilon}[0, \infty)$, 264
- Marriage lemma, 97
Martingale, 55*
 central limit theorem, 339, 471
 characterization using stopping times, 93
 class DL, 74*
 continuous, 79
 convergence in distribution, 362

- cross variation, 79
 directed index set, 87*
 Doob inequality, 63, 64
 local, 64
 multiparameter, 317
 optional sampling theorem, 61, 88, 92, 93
 orthogonal, 80*
 (\cdot) process, 79*, 280, 282
 quadratic variation, 67*, 71
 relative compactness, 343
 right continuity of, 61
 sample paths of, 59, 61
 square integrable, 78*, 279
 upcrossing inequality for, 57
Martingale problem, 173*
 bounded perturbation, 256, 257
 branching Markov process, 404
 $C_E[0, \infty)$, 186*
 collection of solutions, 202
 continuation of solutions, 206
 convergence in distribution, 234
 $D_E[0, \infty)$, 186*
 diffusion approximation, 354
 discrete time, 263
 for distributions, 174*
 equivalent formulations, 176
 existence, 199, 219, 220
 existence of Markovian solution, 210
 independent components, 253
 local, 223*
 localization, 216
 Markov property, 184
 measure-valued process, 436
 for processes, 173*
 for random time change, 308
 sample path properties, 178
 sample paths in $C_E[0, \infty)$, 295
 stopped, 216*
 Schlögl model, 3
 time-dependent, 221*
 uniqueness, 182*, 184, 187, 217, 219
 well-posed, 182*
Matrix square root, 374
M(E), 155*
Measurability of P_x , 158, 188, 210
Measurable selection, 505
Measurable semigroup, 23*
Measurable stochastic process, 50*
Measure-valued process, 401, 436
 Mémin, 154
 Métivier, 154, 305, 409
Metric lattice, 85*
 separable from above, 85
 Meyer, 73, 75, 93, 94, 305, 507
Mikulevičius, 364
Mixing, 345, 362
Modification:
 progressive, 89
 of stochastic process, 50*
Modulus of continuity, 277*
 in $D_E[0, \infty)$, 122*, 134, 310, 321, 334
Monotone class, 496*
Monotone class theorem, 496
 for functions, 497
Moran, 451
Moran model, 433
Morando, 502
Morkvenas, 274
Multiparameter random time change, *see*
 Random time change equation
Multiparameter stopping time, 312
Multivalued operator, 20*
 domain and range of, 20*
 Nagasawa, 409
 Nagylaki, 48, 449, 451
 Nappo, 5, 274
 Neveu, 47
 Ney, 409
 Norman, 48, 274, 385, 451, 467
Offspring distribution, 386
 Ohta, 451
 Ohta-Kimura model, 440, 450
 Oleinik, 374, 385
Optional:
 modification, 72
 process, 71*
 sets, 71*
Optional projection, 73*
 in Banach space, 91
Optional projection theorem, 72
Optional sampling theorem, 61, 92, 93
 directed index set, 88
 multiparameter, 317
Ornstein-Uhlenbeck process, 191
 Papanicolaou, 274, 491
 Parthasarathy, 506, 507
 Pazy, 47
P-continuity set, 108*
 $\mathcal{P}(E)$, 96
 Peligrad, 364
Perturbation by bounded operator, 38
Perturbation of generator, 37, 44
 Phillips, 47, 385
Picard iteration, 299
 Pinsky, 491

- Poisson process, martingale characterization, 360
 Portmanteau theorem, 108
 Positive maximum principle, 165*
 Positive operator, 165*
 Positive semigroup, 165*
 Predictable process, 75
 Priouret, 305
 Process, *see* Stochastic process
 Product space:
 separating and convergence determining sets
 in, 115
 tightness in, 107
 Progressive modification, 89
 Progressive sets, 71*
 with directed index set, 86
 Progressive stochastic process, 50*
 Prohorov, 154
 Prohorov metric, 96, 357, 408
 completeness of, 101
 separability of, 101
 Prohorov theorem, 104

 Quadratic variation of local martingale, 67
 Quasi-left continuity, 181

 Random evolution, 469*
 diffusion approximation, 475, 480, 484
 Random time change, multiparameter, 311
 Random time change equation, 306*
 convergence in distribution, 310, 322, 323,
 390, 397, 458
 corresponding martingale problem, 308, 309,
 316
 corresponding stochastic integral equation, 329
 diffusion approximation, 330
 for diffusion process, 328
 for jump Markov process, 326, 455
 multiparameter, 312*
 nonanticipating solution, 314, 315
 nonuniqueness, 332
 relative compactness, 321
 for Schlägl model, 3
 strong uniqueness, 314
 uniqueness, 307
 weak solution, 313
 weak uniqueness, 314
 Rao, 478
 Rebolledo, 274, 364
 Relative compactness:
 in $D_E[0, \infty)$, 197, 343
 in $\mathcal{A}(D_E[0, \infty))$, 128, 137, 139, 142, 152
 Resolvent identity, 11
 Resolvent for semigroup, 10*
 Resolvent set, 10*
- Reversibility, 450*
 Révész, 364
 Rishel, 94
 Robbin, 499
 Rootzén, 364
 Rosenblatt, 364
 Rosenkrantz, 274
 Rösler, 274
 Rota, 48
 Roth, 274, 373, 385
 Rozanov, 364
 Rudin, 496
 Ryll-Nardzewski, 506

 Sample paths:
 continuity of, 171
 for Feller process, 167
 for solution of martingale problem, 178
 of stochastic process, 50*
 Sarason, 385
 Sato, 154, 451
 Schauder, 385
 Schlägl, 5
 Schlägl model, 2
 Schwartz, 494
 Semigroup, 6*
 approximation theorem, 28, 31, 34, 36, 39,
 45, 46
 with bounded generator, 7
 of conditioned shifts, 80*, 92, 226, 229, 485
 contraction, 6*
 convergence, 225, 388
 convergence of resolvents, 44
 corresponding to a Markov process, 161
 ergodic properties, 39
 Feller, 166*
 generator of, 8*
 Hille-Yosida theorem, 13
 for jump Markov process, 163
 limit of perturbed, 41, 45, 473
 measurable, 23*, 80
 perturbation, 37
 positive, 165*
 strongly continuous, 6*
 unique determination, 15
 Separable from above, 85*
 Separating set (of functions), 112*
 on product space, 115
 on subset of $\mathcal{A}(S)$, 116
 Separation (of points), 112*
 Serant, 385
 Set-valued functions, measurability, 506
 Shiga, 274, 451
 Shiryaev, 364

- Siegmund, 273
 Single-valued operator, 20*
 Skorohod, 47, 154, 274, 305, 364, 380
 Skorohod representation, 102
 in R, 150
 Skorohod topology:
 compact sets in, 122
 completeness, 121
 metric for, 117*, 120*
 separability, 121
 Slutsky, 154
 Slutsky theorem, 110
 Sova, 47, 48
 Space-time process, 221, 295
 Spitzer, 273
 Stationary distribution, 238*, 239
 characterization, 248
 convergence, 244, 245, 418
 existence, 240, 243
 for genetic diffusion, 417, 448
 infinitely-many-allele model, 443
 relative compactness, 246
 uniqueness, 270
 Stationary process, 238*
 Stationary sequence:
 invariance principle for, 350
 Poisson approximation, 362
 Stieltjes integral, 280
 Stochastic integral:
 iterated, 286, 287
 with respect to local martingale, 286*
 with respect to martingale, 282*
 for simple functions, 280
 Stochastic integral equation, 290
 corresponding martingale problem, 292, 293
 corresponding random time change equation,
 329
 existence, 299, 300
 pathwise uniqueness, 291, 296, 297, 298
 uniqueness in distribution, 291, 295, 296
 Stochastic process, 49*
 adapted, 50*
 (right, left) continuous, 50*
 equivalence of, 50*
 finite-dimensional distributions, 50*
 increasing, 74*
 index set of, 49*
 indistinguishability, 50*
 measurable, 50*
 modification of, 50*
 progressive, 50*
 sample paths of, 50*
 state space of, 49*
 version of, 50*
- Stone, 154
 Stopped martingale problem, 216*
 Stopped process, X', 64, 68, 285
 Stopping time, 51*
 approximation by discrete, 51, 86
 bounded, 51*
 closure properties of collection, 51
 contact time, 54*
 corresponding σ -algebra, 52*, 89
 directed index set, 85
 discrete, 51*
 entrance time, 54*
 exit time, 54*
 finite, 51*
 truncation of, 51, 86
 Strassen, 154
 Stratonovich, 491
 Strong approximation, 356, 460
 Strong Markov process, 158*
 Strong Markov property, 158*
 for Brownian motion, 278
 Strong mixing, 345*
 Strong separation (of points), 113*, 143
 Stroock, 273, 274, 305, 336, 364, 369, 371,
 374, 375, 380, 385
 Submartingale, 55*
 of class DL, 74*
 Supermartingale, 55*
 nonnegative, 62
 Telegrapher's equation, 470
 Tightness, 103*
 Time homogeneity, 156*
 Total variation norm, 495*
 Transition function:
 continuous time, 156*
 discrete time, 158*
 existence, 502
 Trotter, 47, 48, 274, 451
 Trotter product formula, 33
 alternative proof, 45
 Tulcea's theorem, 504
 Tusnády, 356, 364, 459
 Uniform integrability, 493*
 of class DL submartingale, 74
 of conditional expectations, 90
 of submartingales, 60, 90
 weak compactness, 76
 Uniform mixing, 345*, 348, 484
 Uniqueness:
 for forward equation, 251, 252
 for martingale problem, 182
 for random time change equation, 307, 314

- Uniqueness (*Continued*)**
for stochastic integral equation, 291, 295, 296,
297, 298
for $u' = Au$, 18, 26
- Upcrossing inequality**, 57
- Ural'tseva**, 369, 385
- Varadhan**, 273, 274, 305, 364, 369, 371, 374,
375, 385, 491
- Vasershtein**, 273
- Version of stochastic process**, 50*
- Villard**, 385
- Viot**, 451
- Volkonski**, 335, 364
- Wagner**, 506
- Wald**, 154
- Wang**, 409, 467
- Watanabe**, 47, 93, 273, 305, 364, 385, 409
- Watterson**, 450, 451
- Weak convergence**, 107*. *See also* Convergence
in distribution
- Weak topology**, metric for, 96, 150*
- Weiss**, 274
- Whitney**, 499
- Whitney extension theorem**, 499
- Williams**, 273, 274, 305
- Withers**, 364
- Wonham**, 274
- Wright**, 451
- Wright-Fisher model**, 414
- X'**, 64*
- Yamada**, 305, 385
- Yosida**, 47
- Yosida approximation of generator**, 12*, 261
- Zakai**, 274

FLOWCHART

This table indicates the relationships between theorems, corollaries, and so on. For example, the entry **C2.8 P2.1 P2.7 T6.9 T4.2.2** under Chapter 1 means that Corollary 2.8 of Chapter 1 requires Propositions 2.1 and 2.7 of that chapter for its proof and is used in the proofs of Theorem 6.9 of Chapter 1 and Theorem 2.2 of Chapter 4.

Chapter 1

P1.1 C1.2 P1.5b C1.6 C1.2 P1.1 R1.3 L1.4a Prob.3 L1.4b Prob.3 P2.1 P3.4 L1.4c Prob.3 P1.5c L2.5 P5.4 L6.2 P1.5a C1.6 P1.5b P1.1 P2.1 P3.3 P3.7 T10.4.1 P1.5c L1.4c C1.6 T2.6 P2.7 P3.4 L6.2 T6.11 R4.2.10 P4.9.2 T8.3.1 C1.6 P1.1 P1.5ac P2.1 T2.6 P2.7 P4.9.2 T10.4.1 P2.1 L1.4b L1.5b C1.6 T2.6 P2.7 C2.8 P3.3 P3.7 P4.1 T6.1 L6.3 T6.5 T6.9 T7.1 R7.9b T2.7.1 C4.2.8 L2.2 L2.3 T2.6 L2.11 L2.3 L2.2 T2.6 T4.3 T6.9 T6.11 T4.5.19a L2.4a T2.6 P2.7 T6.1 L2.4b T2.6 P2.7 L2.4c T2.6 P2.7 T6.1 C6.8 T7.1 L2.5 L1.4c T2.6 P2.7 T2.6 P1.5c C1.6 P2.1 L2.2 L2.3 L2.4abc L2.5 T2.12 P3.4 T4.3 T7.1 T4.4.1 P2.7 P1.5c C1.6 P2.1 L2.4abc L2.5 P2.9 C2.8 T6.1 T4.2.7 C2.8 P2.1 P2.7 T6.9 T4.2.2 P2.9 P2.10 P2.7 P2.10 P2.9 P3.4 L2.11 L2.2 T2.12 P3.1 P3.4 T2.12 T2.6 L2.11 P3.1 P3.5 P3.7 T4.2.2 P3.1 L2.11 T2.12 P3.3 T6.1 L6.3 T6.5 R3.2 T8.1.5 T8.3.1 P3.3 P1.5b P2.1 P3.1 P5.1.1 T8.1.6 T8.2.1 T8.3.1 T8.3.4 L10.3.1 P12.2.2 P3.4 L1.4b P1.5c T2.6 P2.10 L2.11 P3.5 T2.12 L3.6 P3.7 T8.3.1 P3.7 P1.5b P2.1 T2.12 L3.6 C3.8 T8.2.1 T8.2.5 T8.2.8 C3.8 P3.7 P4.1 P2.1 T7.1 R7.9c T12.2.4 L4.2 T4.3 T4.4.1 T12.4.1 T4.3 L2.3 T2.6 L4.2 T8.3.1 P5.1 C4.8.7 C4.8.16 P5.2 P2.7.5 P5.3 P5.4 L1.4c T9.4.3 R5.5 T6.1 P2.1 L2.4ac P2.7 P3.1 L6.2 L6.3 T6.5 T7.6a C7.7a T4.2.5 T4.2.11 R4.8.8a T8.3.1 L6.2 P1.4c P1.5c T6.1 L6.3 P2.1 P3.1 T6.1 L6.4 T6.5 T6.11 T6.5 P2.1 P3.1 L6.4 T6.1 C6.6 C6.7 C6.8 T7.6b C7.7b T4.2.6 T4.2.12 T5.1.2c T9.1.3 T10.1.1 C6.6 T6.5 C6.7 C6.8 C6.7 T6.5 C6.6 C6.8 L2.4c T6.5 C6.6 T2.7.1 T4.4.1 P4.9.2 T4.9.3 T6.9 P2.1 L2.3 C2.8 T6.11 T4.8.2 R6.10 Prob.16 T6.11 P1.5c L2.3 L6.4 T6.9 T7.1 P2.1 L2.4c T2.6 P4.1 C7.2 C7.2 T7.1 L7.3a Prob.18 C7.7ab L7.3b Prob.18 L7.3c Prob.18 L7.3d Prob.18 T7.6ab R7.9a E12.2.6 R7.4 T10.3.5ab R7.5 T10.3.5ab T7.6a T6.1 L7.3d C7.7a C7.8 T7.6b T6.5 L7.3d C7.7b T10.3.5ab C7.7a T6.1 L7.3a T7.6a C7.7b T6.5 L7.3a T7.6b T10.3.5b C7.8 T7.6a P12.2.2 R7.9a L7.3d R7.9b P2.1 R12.2.3 R7.9c P4.1 P12.2.2 R7.9d

Chapter 2

L1.1 P1.2d P1.5b L4.1 P1.2a P1.2b P1.4g P1.2c P1.2d L1.1 P1.3 T2.13
 R3.8.5a R4.1.4 T4.2.7 T5.1.2a P1.4a P1.4b P1.4def P2.15 T4.2 C4.4 C4.5
 P1.4c P1.4ef L2.2 T2.13 P3.2 L4.1 P1.4d P1.4b L2.2 T2.13 P3.2 L4.1 R4.3 P4.1.5
 T4.3.12 T4.4.2bc L5.2.4 P1.4e P1.4bc P1.4f P1.4f P1.4bce P1.4g P1.2b L3.8.4
 P1.5a P3.2 P3.4 L3.5 P3.6 T5.1 T4.6.1 T4.6.2 T4.6.3 C4.6.4 L4.6.5 L4.10.6 T5.2.9
 T5.3.7 T5.3.11 P1.5b L1.1 P2.15 P2.16b C5.3 C5.4 T4.3.8 C4.3.13 T7.1.4ab T7.4.1
 T8.3.1 T1.6 P2.1a C2.17 P3.4 L3.5 P3.6 P6.2 T7.1.4a T9.2.1a P2.1b P2.9 T2.13
 L7.2 L2.2 P1.4cd L2.3 L2.5 T2.13 L2.3 L2.2 C2.4 C2.4 L2.3 P2.9 C2.11 R2.12
 P2.16a L2.5 L2.2 C2.6 C2.6 L2.5 P2.9 C2.11 R2.12 L2.7 P2.9 L2.8 P2.9
 T4.3.6 P2.9 P2.1b C2.4 C2.6 L2.7 L2.8 Prob.8 Prob.9 Prob.10a C2.10 L4.1 R7.3
 T4.3.6 C2.10 P2.9 Prob.10a C2.11 R2.14 T12.4.1 C2.11 C2.4 C2.6 C2.10 Prob.9
 T5.1 R5.2a R2.12 C2.4 C2.6 Prob.9 P3.4 L3.5 P3.6 T5.1 L4.6.5 T2.13 P1.3
 P1.4cd P2.1b L2.2 Prob.10a R2.14 P2.13 P2.16b P3.1 P3.2 P3.4 L3.5 P3.6 T4.2 C4.4
 C4.5 T5.1 R5.2b C5.4 P4.2.9 T4.3.8 P4.3.9 P4.3.10 T4.3.12 C4.4.14 T4.5.11b L4.5.13
 T4.6.1 T4.6.2 T4.6.3 C4.6.4 L4.6.5 T4.10.1 L4.10.6 L5.2.4 T5.3.7 T6.1.3 T6.1.4
 T6.2.8b T6.5.3b T7.1.4ab T7.4.1 T8.3.1 L9.1.6 P10.2.8 L10.2.10 R2.14 C2.10 T2.13
 P2.15 T4.3.8 P4.3.9 P4.3.10 C4.3.13 L4.5.13 T6.1.3 T6.1.4 T6.2.8b T6.5.3b T7.1.4ab
 T7.4.1 T8.3.1 P2.15 P1.4b P1.5b T2.13 R2.14 P4.2.4 P2.16a C2.4 C2.17 P3.4
 P3.6 T7.1.4a T9.3.1 L9.4.1 P2.16b P1.5b T2.13 C2.17 T5.2.3 T5.3.11 T9.2.1a
 C2.17 P2.1a P2.16ab P3.4 L3.5 P3.6 T10.4.5 P3.1 T2.13 P3.2 P1.4cd P1.5a
 T2.13 C3.3 L4.3.2 C3.3 P3.2 P3.4 P1.5a P2.1a R2.12 T2.13 P2.16a C2.17 L3.5
 Prob.8 Prob.10a P6.1 T5.2.9 T7.1.4a L3.5 P1.5a P2.1a R2.12 T2.13 C2.17 Prob.10a
 P3.4 P3.6 P1.5a P2.1a R2.12 T2.13 P2.16a C2.17 Prob.10a L4.1 L1.1 P1.4cd
 P2.9 T4.2 C4.4 C4.5 T4.2 P1.4b T2.13 L4.1 TA.4.2 P7.5 R4.3 P1.4d C4.4 P1.4b
 T2.13 L4.1 TA.4.2 C4.5 P1.4b T2.13 L4.1 TA.4.2 T12.4.1 P4.6 TA.4.3 R4.7 R4.7
 P4.6 T7.1 T4.8.2 R4.8.3b C4.8.4 C4.8.5 C4.8.12 C4.8.13 T5.1 P1.5a C2.11 R2.12
 T2.13 Prob.15 PA.2.1 PA.2.4 C5.3 C5.4 P6.2 L7.2 T5.2.3 L5.2.4 R5.2a C2.11
 R5.2b T2.13 R5.2c C5.3 P1.5b T5.1 C5.4 T2.13 T5.1 P6.2 T5.2.9 P6.1 P3.4
 P6.2 T5.2.9 P6.2 P2.1a T5.1 C5.4 P6.1 Prob.10c T5.2.3 L5.2.4 T7.1 P1.2.1 C1.6.8
 R4.7 P7.6 T4.8.2 R4.8.3a T4.8.10 L7.2 P2.1b T5.1 Prob.10b C7.4 R7.3 P2.9
 C7.4 L7.2 P7.5 P1.5.2 T4.2 P7.6 R4.8.3b P7.6 T7.1 P7.5 P8.1a P8.1b P8.6
 P8.2 P8.5c R8.3 P8.4 T8.7 P8.5a P8.5b P8.5c P8.2 P8.6 P8.1b
 P6.2.10 T8.7 P8.4 T6.2.8a

Chapter 3

L1.1 T1.2 L1.3 C1.6 L1.3 C1.5 T1.2 T1.7 T1.8 L1.4 C1.5 C1.5 L1.4 L1.3
 C1.6 T1.2 T1.7 C1.9 T1.7 L1.3 C1.6 Prob.3 T2.2 T4.4.6 T4.6.3 T1.8 L1.3 C1.9
 T7.8a T6.1.5 T6.3.3 C1.9 C1.6 T1.8 T4.5b T6.3.4a T6.5.4 T7.4.1 T9.2.1b T9.3.1
 C10.2.6 C10.2.7 T10.4.6 T11.2.3 TA.1.2 L2.1 T2.2 T4.5ab C7.4 C8.10 C9.2 T4.1.1
 P4.4.7 TA.8.1 T2.2 T1.7 L2.1 C2.3 T4.5b P4.6b T7.2 R7.3 L7.5 L4.5.3 T4.5.11b
 L4.5.15 R4.9.4 T4.9.9 T4.9.10 T10.2.2 TA.1.2 TA.8.1 C2.3 T2.2 L4.9.13 P2.4
 P4.6b T4.1.1 T3.3.4b TA.1.2 T3.1 C3.2 C3.3 P4.4 T4.5b C8.10 C9.2 T10.2b P10.4
 T4.5.11c L4.5.17 T10.4.5 TA.8.1 C3.2 T3.1 C9.3 R9.1.5 C3.3

T3.1 P10.4 T7.3.1 T7.3.3 T7.4.1 T11.2.3 R3.4 L8.1 L4.1 P4.2 P4.2 L4.1 TA.4.3
 T4.5a T4.4.6 T4.5.19a C7.2.8 L4.3 T4.5b P4.6b L4.8.1 P4.4 T3.1 T4.5a L2.1
 P4.2 T4.5b T4.5b C1.9 L2.1 T2.2 T3.1 L4.3 T4.5a P4.8a P4.1.6 T4.4.2a C4.4.3
 L4.8.1 P4.6b T2.2 P2.4 L4.3 L5.1 P5.2 P6.5 P7.1 T6.1.1b P5.2 L5.1 P6.5 P7.1
 T7.8a T4.5.11c P5.3 C5.5 T5.6 L6.2b P6.5 C9.2 L10.1 L4.5.10 T4.5.19a R5.4 T5.6
 P6.5 C5.5 P5.3 T5.6 P5.3 R5.4 Prob.14 P7.1 T7.2 R7.3 T7.8a C9.2 T4.4.6 T4.6.3
 L6.1 T6.3 T7.2 L7.5 L6.2a L6.2b T6.3 P6.5 C7.4 L6.2b P5.3 L6.2a L6.2c T6.3
 L6.2c L6.2b T6.3 L6.1 L6.2ab Prob.16 T7.2 T6.3.3 R6.4 Prob.16 R7.3 C9.2
 P6.5 L5.1 P5.2 P5.3 R5.4 L6.2a P6.3.2 P7.1 L5.1 P5.2 T5.6 T7.8b C4.4.3 T7.2
 T2.2 T5.6 L6.1 T6.3 C7.4 T8.6 T9.1 R7.3 T2.2 T5.6 R6.4 C7.4 L2.1 L6.2a T7.2
 T6.1.5 P6.3.1 T6.3.4b C6.3.6 T6.5.4 T9.3.1 L7.5 T2.2 L6.1 Prob.15 T7.6 T7.8 L7.5
 L7.7 T7.8ab L4.5.1 T4.8.10 T7.8a T1.8 P5.2 T5.6 L7.7 T7.8b C8.10 C9.2 L4.5.1
 T4.8.10 T7.8b P7.1 L7.7 T7.8a TA.4.2 C9.3 C4.8.6 C4.8.15 L8.1 R3.4 P8.3 L8.2
 P8.3 P8.3 L8.1 L8.2 T8.6 L8.4 P2.1.4g R8.5b T8.6 R8.5a P2.1.3 R8.5b L8.4
 T8.6 T8.6 Prob.2.25 T7.2 P8.3 L8.4 R8.5b R8.7a C8.10 T9.1 T9.4 T7.1.4ab T7.4.1
 T9.1.4 R8.7a T8.6 T9.4 R8.7b T8.8 C8.10 R8.9a R8.9b C8.10 L2.1
 T3.1 T7.8a T8.6 T8.8 T9.1 T7.2 T8.6 Prob.13 C9.2 C9.3 R4.5.2 L4.5.17 C4.8.6
 C4.8.15 T4.9.17 T7.1.4a T10.4.1 C9.2 L2.1 T3.1 P5.3 T5.6 R6.4 T7.8a T9.1 C4.8.6
 C9.3 C3.2 T7.8b T9.1 T4.2.5 T4.2.6 T4.2.11 T4.2.12 T9.4 T8.6 R8.7a Prob.23
 T4.2.5 T4.2.6 T4.2.11 T4.2.12 R4.5.2 L4.5.17 C4.8.6 C4.8.15 T4.9.17 T10.4.1 R9.5a
 R8.5b L10.1 P5.3 T10.2b T10.2a P10.4 T7.1.4b T9.1.4 T11.4.1 T10.2b T3.1
 L10.1 P10.3 P10.4 T3.1 C3.3 T10.2a

Chapter 4

T1.1 Prob.2.27 L3.2.1 P3.2.4 TA.9.1 P1.2 TA.4.2 P1.5 T5.19c P1.3 T2.7 R1.4
 P2.1.3 P1.5 P2.1.4d P1.2 TA.4.2 P1.6 P3.4.6a T4.1 P1.7 P10.2.8 L2.1 T2.2
 T12.4.1 T2.2 C1.2.8 T1.2.12 L2.1 T8.1.4 T8.1.5 T8.3.1 L2.3 P2.4 T2.5 T2.6 T2.11
 T2.12 T8.3.1 P2.4 P2.2.15 L2.3 T2.5 T2.6 T2.11 T2.12 T2.5 T1.6.1 C3.9.3 T3.9.4
 L2.3 P2.4 T2.7 T11.2.3 T2.6 T1.6.5 C3.9.3 T3.9.4 L2.3 P2.4 T5.1.2c T2.7 P1.2.7
 P2.1.3 P1.3 T2.5 C2.8 P2.9 T5.1.2 T8.3.1 T10.1.1 T12.2.4 T12.3.1 T12.4.1 C2.8
 P1.2.1 T2.7 T8.3.1 P2.9 T2.2.13 T2.7 T5.1.2 T10.2.4 T12.2.4 T12.3.1 T12.4.1
 R2.10 P1.1.5c T2.11 T1.6.1 C3.9.3 T3.9.4 L2.3 P2.4 T2.12 T1.6.5 C3.9.3 T3.9.4
 L2.3 P2.4 P3.1 T4.1 L3.2 P2.3.2 Prob.2.22 Prob.14 P3.3 P3.5 T4.1 C4.4 L5.6
 L5.18 T5.19a P9.2 T9.3 T10.1 L6.2.6 T6.2.8ab L9.4.1 P3.3 L3.2 L3.4 T10.3
 L6.2.8b T9.4.3 T10.4.1 T12.2.4 T12.3.1 T12.4.1 P3.5 L3.2 T5.19a T3.6 L2.2.8
 P2.2.9 Prob.3.7 C3.7 C3.7 T3.6 P5.3.5 T8.3.3 T10.4.1 T3.8 P2.1.5b T2.2.13
 R2.2.14 R5.5 T8.3.3 P3.9 P2.1.5b T2.2.13 R2.2.14 P5.3.5 P5.3.10 P3.10 P2.1.5b
 T2.2.13 R2.2.14 R3.11 T3.12 P2.1.4d T2.2.13 Prob.3.7 C3.13 T5.11c R6.1.6
 T6.3.4a C3.13 P2.1.5b R2.2.14 T3.12 T4.1 T1.2.6 L1.4.2 C1.6.8 P1.6 P3.1 L3.2
 T5.19c T4.2a P3.4.6a P4.7 T10.1 T10.3 T9.4.3 T10.4.1 T4.2b P2.1.4d Prob.2.11b
 T4.2c P2.1.4d Prob.2.11b T5.11c P9.19 T9.14 C4.3 P3.4.6a P3.7.1 C4.4 T6.2 P9.19
 C4.4 L3.2 C4.3 R4.5 Prob.22 T4.6 T3.1.7 P3.4.2 T3.5.6 TA.10.5 P4.7 L3.2.1
 T4.2a R4.8a R4.8b E4.9 Prob.23 L4.10 R4.12 T4.11 C4.13 R4.12 L4.10
 C4.13 T4.11 C4.14 T2.2.13 C4.15 C4.15 C4.14 R4.16 L5.1 L3.7.7 T3.7.8a
 T5.4 L5.17 T9.17 R5.2 T3.9.1 T3.9.4 T5.4 L5.3 T3.2.2 TA.10.1 T5.4 T10.4.1 T5.4
 L5.1 R5.2

L5.3 T5.3.10 R5.5 T3.8 L5.6 L3.2 L5.8 L5.9 L5.10 L5.10 P3.5.3 L5.9 T5.11b
 T5.11a L5.16 T5.11b T5.11b T2.2.13 T3.2.2 L5.10 T5.11a L5.17 T5.11c T5.11c
 T3.3.1 P3.5.2 T3.12 T4.2c T5.11b R5.12 L5.13 T2.2.13 R2.2.14 E5.14 L5.15
 Prob.2.26 T3.2.2 L5.16 T5.3.6 C7.5.3 L5.16 L5.15 T5.11a T6.1 L6.5 T6.1.4 T6.2.8b
 L5.17 T3.3.1 T3.9.1 T3.9.4 L5.1 T5.11b L5.18 L3.2 T5.19a T5.19a L1.2.3 P3.4.2
 P3.5.3 L3.2 P3.5 L5.18 L5.20 T5.19b T5.19c T4.1 P1.2 T5.19d L5.20 L5.20
 P3.5.3 L5.10 T5.19ad T6.1 P2.1.5a T2.2.13 L5.16 C6.4 T7.4.1 T8.2 P2.1.5a
 T2.2.13 C4.3 Prob.27 T6.3 P2.1.5a T2.2.13 T3.1.7 T3.5.6 C6.4 C6.4 P2.1.5a
 T2.2.13 T6.1 T6.3 L6.5 P2.1.5a R2.2.12 T2.2.13 L5.16 T8.6 T8.8 L6.5 T7.1
 P5.3.5 T5.3.10 L7.2 T7.3 Prob.29 L8.1 L3.4.3 P3.4.6a T8.2 R6.3.5a T8.2 T1.6.9
 R2.4.7 T2.7.1 L8.1 C8.4 C8.5 C8.6 R8.3a T2.7.1 R8.3b R2.4.7 P2.7.5 R8.3c
 C8.7 C8.9 R8.3d C8.4 R2.4.7 T8.2 C8.5 R2.4.7 T8.2 C8.6 T3.7.8b T3.9.1
 C3.9.2 T3.9.4 T8.2 C8.7 C8.9 T9.2.1a T12.4.1 C8.7 P1.5.1 R8.3c C8.6 T12.2.4
 T12.3.1 R8.8a T1.6.1 R8.8b C8.9 R8.3c C8.6 T9.1.3 T10.1.1 T10.3.5a T8.10
 T2.7.1 L3.7.7 T3.7.8a C8.12 C8.13 C8.15 R8.11 C8.16 C8.17 C8.12 R2.4.7 T8.10
 C8.13 R2.4.7 T8.10 R8.14 C8.15 T3.7.8b T3.9.1 T3.9.4 T8.10 C8.16 C8.17
 C8.16 P1.5.1 R8.11 C8.15 T9.4.3 P10.4.2 C8.17 R8.11 C8.15 T10.4.1 P8.18
 L9.1 P9.2 P9.2 P1.1.5c C1.1.6 C1.6.8 L3.2 L9.1 T9.3 L10.2.1 T9.3 C1.6.8 L3.2
 P9.2 T10.4.6 R9.4 T3.2.2 L9.5 R9.6 L9.7 T3.2.2 C9.8 T9.9 T3.2.2
 T9.10 T3.2.2 R9.11 T9.12 T10.4.6 L9.13 C3.2.3 T9.14 R9.15 L9.16
 T9.17 T9.17 T3.9.1 T3.9.4 L5.1 L9.16 Prob.41 TA.8.1 P9.19 T10.4.6 P9.18
 P9.19 T4.2c C4.3 T9.17 T10.3 T10.1 T2.2.13 Prob.2.23 L3.2 L4.2a T10.3 P10.2
 PA.4.5 T10.3 T10.3 L3.4 T4.2a T10.1 P10.2 L10.5 L10.6 P9.19 R10.4 L10.5
 T10.3 L10.6 P2.1.5a T2.2.13 T10.3

Chapter 5

P1.1 P1.3.3 T1.2 T1.2c T9.3.1 E10.4.3 T11.2.3 T1.2 T4.2.7 P4.2.9 P1.1 P3.1
 T9.3.1 T1.2a P2.1.3 T1.2bc T11.2.3 T1.2b T1.2a C3.4 T3.7 T3.8 T3.11 E9.3.2
 T1.2c T1.6.5 T4.2.6 P1.1 T1.2a L2.1 T2.3 L2.2 T2.3 T2.3 P2.2.16b T2.5.1
 P2.6.2 L2.1 L2.2 L2.4 T2.6 T3.11 L2.4 P2.1.4d T2.2.13 T2.5.1 P2.6.2 T2.3 T2.6
 L2.7 L2.5 Prob.11 L2.8 T2.6 T2.3 L2.4 L2.7 L2.8 T2.9 T3.1 T3.3 T3.7 T7.1.1
 L2.7 L2.4 T2.6 T2.9 L2.8 L2.5 T2.6 P3.1 T9.3 T2.9 P2.1.5a P2.3.4 C2.5.4 P2.6.1
 T2.6 L2.7 L2.11 Prob.12 T2.12 P3.1 C3.4 T3.8 T7.1.1 T7.1.2 T7.4.1 E9.3.2 T10.4.5
 R2.10 L2.11 T2.9 T2.12 T2.9 P3.1 T3.3 T6.5.3b P3.1 T1.2b T2.6 L2.8 T2.9
 T2.12 L3.2 TA.10.1 T3.3 T3.3 T2.6 L2.8 T2.12 L3.2 C3.4 T6.5.3a C3.4 T1.2b
 T2.9 T3.3 T3.10 T8.2.3 T8.2.6 P3.5 C4.3.7 P4.3.9 T4.7.1 Prob.4.19 T3.10 T8.1.7
 T8.2.3 T8.2.6 T3.6 L4.5.15 T8.2.6 T3.7 P2.1.5a P2.2.13 T1.2b T2.6 TA.5.1 T3.11
 T3.8 T1.2b T2.9 TA.5.1 T8.2.3 R3.9 T3.10 P4.3.9 T4.5.4 T4.7.1 Prob.4.19 C3.4
 P3.5 T8.2.3 T8.2.6 T3.11 P2.1.5a P2.2.16b T1.2b T2.3 T3.7 T11.3.2

Chapter 6

T1.1a T1.1b T1.1b L3.5.1 T1.1a T1.5 R1.2 T1.3 T2.2.13 R2.2.14 T9.3.1 T1.4
 T2.2.13 R2.2.14 L4.5.16 Prob.4.45 Prob.12 T1.5 T3.1.8 C3.7.4 T1.1.b T9.1.4 R1.6
 T4.3.12 L2.1 TA.11.3 T2.2a TA.11.3 T2.2b T2.2b T2.2a T4.1b T5.1 R2.3 Prob.1
 P2.4 TA.11.3 R2.5 L2.6 L4.3.2 L2.7 T2.8ab P2.10 L2.7 L2.6 T3.4a T5.4
 T2.8a T2.8.7 L4.3.2 L2.6 T4.1b T5.1 T2.8b Prob.1.23 T2.2.13 R2.2.14 Prob.2.24
 L4.3.2 L4.3.4 L4.5.16 Prob.4.45 L2.6 R2.9a R2.9b P2.10

P2.8.6 L2.6 T4.1b T5.1 P3.1 C3.7.4 P3.2 P3.6.5 Prob.5 T3.3 T3.1.8 T3.6.3
 T3.4a C3.1.9 T4.3.12 L2.7 T3.4b C3.6 T3.4b P3.2.4 C3.7.4 T3.4a Prob.7 R3.5a
 L4.8.1 R3.5b C3.6 C3.7.4 T3.4a T4.1a T4.1b T2.2b T2.8a P2.10 T4.1c
 T4.1c T4.1b T11.2.1 T5.1 T2.2b T2.8a P2.10 R5.2a T5.3a T11.3.1 R5.2a T5.1
 R5.2b T5.3a T5.3.3 T5.1 T5.3b T2.2.13 R2.2.14 T5.2.12 T5.4 C3.1.9 C3.7.4
 L2.7 R5.5

Chapter 7

T1.1 T5.2.6 T5.2.9 T1.4b T1.2 T5.2.9 T1.4a R1.3 Prob.2 T1.4a P2.1.5b P2.2.1a
 T2.2.13 R2.2.14 P2.2.16a P2.3.4 T3.8.6 T3.9.1 Prob.3.22c T1.2 Prob.7 T3.1 T3.3
 T1.4b P2.1.5b T2.2.13 R2.2.14 T3.8.6 T3.10.2a T1.1 Prob.7 PA.2.2 PA.2.3 T9.3.1
 R1.5 L2.1 P2.2 P2.6 P2.2 L2.1 R2.3 C2.4 C2.5 R2.3 P2.2 T3.1 C2.4 P2.2
 T3.1 C2.5 P2.2 P2.6 L2.1 C2.7 C2.8 T3.3 T12.4.1 C2.7 P2.6 T3.3 C2.8 P3.4.2
 P2.6 PA.4.5 T3.1 C3.3.3 T1.4a R2.3 C2.4 R3.2a R3.2b T3.3 C3.3.3 T1.4a
 P2.6 C2.7 R3.4 T4.1 P2.1.5b T2.2.13 R2.2.14 C3.1.9 T3.3.3 T3.8.6 T4.6.1 T5.2.9
 C4.2 C4.2 T4.1 Prob.13 T5.1 C5.2 C5.3 C5.2 T5.1 Prob.17 C5.3 L4.5.15 T5.1
 Prob.17 R5.4 C5.5 T11.3.1 C5.5 R5.4 T11.3.1 T5.6

Chapter 8

T1.1 E12.3.3 C1.2 Prob.1 R1.3 Prob.2 T1.4 T4.2.2 T1.5 R1.3.2 T4.2.2 T1.6
 P1.3.3 T1.7 P5.3.5 T2.1 P1.3.3 P1.3.7 L2.2 T9.1.3 L2.2 T2.1 T2.3 C5.3.4
 P5.3.5 T5.3.8 T5.3.10 P2.4 Prob.4 T2.5 T2.5 P1.3.7 P2.4 TA.5.1 T2.6 C5.3.4
 P5.3.5 T5.3.6 T5.3.10 R2.7 T2.8 P1.3.7 L2.9 PA.7.1 T10.1.1 L10.2.1 T10.3.5ab
 L2.9 T2.8 T3.1 P1.1.5c R1.3.2 P1.3.3 L1.3.6 T1.4.3 T1.6.1 P2.1.5b T2.2.13 R2.2.14
 T4.2.2 L4.2.3 T4.2.7 C4.2.8 TA.5.1 C3.2 C3.2 T3.1 T3.3 C4.3.7 C4.3.8 T3.4 T3.4
 P1.3.3 T3.3 T3.5 Prob.8 T3.6 Prob.8

Chapter 9

T1.1 PA.4.5 R1.2 T1.3 T1.6.5 C4.8.9 T8.2.1 TA.1.2 T1.4 T3.8.6 T3.10.2a
 Prob.3.26 T4.4.2c T6.1.5 L1.6 R1.5 C3.3.2 L1.6 T2.2.13 T1.4 T2.1a P2.2.1a
 P2.2.16b C4.8.6 Prob.3 T2.1b T2.1b C3.1.9 T2.1a T2.1c T2.1c T2.1b T3.1
 T2.2.16a C3.1.9 C3.7.4 P5.1.1 T5.1.2 T6.1.3 T7.1.4b E3.2 E3.2 T5.1.2b T5.2.9
 Prob.7.3 T3.1 L4.1 P2.2.16a L4.3.2 TA.5.1 T4.3 T4.2 P1.5.4 L4.3.4 T4.4.2a L4.1
 Prob.7 T4.3 P1.5.4 L4.3.4 T4.4.2a C4.8.16 L4.1 R4.4

Chapter 10

T1.1 T1.6.5 T4.2.7 C4.8.9 T8.2.8 Prob.1 T2.2 L2.1 P4.9.2 T8.2.8 T2.2 R2.3 T2.2
 T3.2.2 T1.1 L2.1 R2.3 L2.1 T2.4 Prob.3.5 P4.2.9 L2.10 Prob.3 PA.2.2 C2.6 C2.7
 R2.5 PA.2.3 C2.7 P2.8 C2.6 C3.1.9 T2.4 PA.2.3 C2.7 C3.1.9 T2.4 R2.5 P2.8
 T2.2.13 P4.1.7 R2.5 R2.9 R2.9 P2.8 L2.10 T2.2.13 T2.4 L3.1 P1.3.3 T3.5ab
 R3.2 L3.3 T3.5ab R3.4 T3.5a R1.7.4 R1.7.5 T1.7.6b C4.8.9 T8.2.8 L3.1 L3.3
 R3.6a E3.8 E3.9 T3.5b R1.7.4 R1.7.5 T1.7.6b C1.7.7b T8.2.8 L3.1 L3.3 R3.6a
 R3.6a T3.5ab R3.6b R3.7a E3.8 E3.9 R3.7b E3.8 T3.5a R3.7a E3.9 T3.5a
 R3.7a T4.1 P1.1.5b C1.1.6 T3.9.1 T3.9.4 L4.3.4 C4.3.7 T4.4.2a L4.5.3 C4.8.17
 TA.5.1 P4.2 E4.4 P4.2 C4.8.16 T4.1 E4.3 E4.4 T4.5 E4.3 P5.1.1 P4.2 E4.4 T4.1
 P4.2 T4.5 T4.6 T4.5 C2.2.17 Prob.2.29 T3.3.1 T5.2.9 P4.2 E4.4

T4.6 T4.6 C3.1.9 T4.9.3 T4.9.12 T4.9.17 E4.4 T4.5 Prob.12 T4.7 T4.7 T4.6

Chapter 11

T2.1 T6.4.1c TA.5.1. T2.3 T4.1 R2.2 T2.3 C3.1.9 C3.3.3 T4.2.5 P5.1.1 T5.1.2a
 T2.1 T4.1 T3.1 T6.5.1 R7.5.4 C7.5.5 TA.5.1 T3.2 T5.3.11 R3.3 T4.1 T3.10.2a
 T2.1 T2.3 R4.2 Prob.5

Chapter 12

L2.1 P2.2 P1.3.3 C1.7.8 R1.7.9c R2.3 T2.4 R2.3 R1.7.9b P2.2 R2.5 T2.4 P1.4.1
 Prob.3.25 T4.2.7 P4.2.9 L4.3.4 C4.8.7 P2.2 E2.6 E2.7 R2.5 R2.3 E2.6 E2.6
 L1.7.3d T2.4 R2.5 E2.7 Prob.1.6a T2.4 T3.1 Prob.3.25 T4.2.7 P4.2.9 L4.3.4 C4.8.7
 E3.3 R3.2 E3.3 T8.1.1 T3.1 T4.1 L1.4.2 C2.2.10 C2.4.5 Prob.3.25 L4.2.1 T4.2.7
 P4.2.9 L4.3.4 C4.8.6 P7.2.6

Appendixes

P1.1 T1.2 T1.2 C3.1.9 T3.2.2 P3.2.4 P1.1 T9.1.3 P2.3 P2.1 T2.5.1 P2.2 P2.3
 P2.2 P2.1 T7.1.4b T10.2.4 P2.3 T1.2 P2.1 T7.1.4b R10.2.5 C10.2.6 P2.4 T4.2
 T2.5.1 R2.5 P3.1 P3.2 T4.1 T4.2 T2.4.2 C2.4.4 C2.4.5 T3.7.8b P4.1.2 P4.1.5
 P2.4 T4.3 T4.3 T4.2 P2.4.6 P3.4.2 C4.4 P4.5 C4.4 T4.3 P4.5 T4.3 P4.10.2
 C7.2.8 T9.1.1 T5.1 T5.3.7 T5.3.8 T8.2.5 T8.3.1 L9.4.1 T10.4.1 T11.2.1 T11.3.1
 T8.1 C6.3 R6.2 C6.3 T6.1 P7.1 T8.2.8 P7.2 P7.2 P7.1 T8.1 L3.2.1 T3.2.2
 T3.3.1 Prob.3.27 T4.9.17 R8.2 T9.1 T4.1.1 T10.1 L4.5.3 L5.3.2 C10.3 R10.2
 C10.3 T10.1 R10.4 T10.5 T4.4.6 P11.1 P11.2 T11.3 L6.2.1 T6.2.2a P6.2.4

WILEY SERIES IN PROBABILITY AND STATISTICS

ESTABLISHED BY WALTER A. SHEWHART AND SAMUEL S. WILKS

Editors: *David J. Balding, Noel A. C. Cressie, Nicholas I. Fisher,
Iain M. Johnstone, J. B. Kadane, Geert Molenberghs, Louise M. Ryan,
David W. Scott, Adrian F. M. Smith, Jozef L. Teugels*

Editors Emeriti: *Vic Barnett, J. Stuart Hunter, David G. Kendall*

The **Wiley Series in Probability and Statistics** is well established and authoritative. It covers many topics of current research interest in both pure and applied statistics and probability theory. Written by leading statisticians and institutions, the titles span both state-of-the-art developments in the field and classical methods.

Reflecting the wide range of current research in statistics, the series encompasses applied, methodological and theoretical statistics, ranging from applications and new techniques made possible by advances in computerized practice to rigorous treatment of theoretical approaches.

This series provides essential and invaluable reading for all statisticians, whether in academia, industry, government, or research.

- † ABRAHAM and LEDOLTER · Statistical Methods for Forecasting
AGRESTI · Analysis of Ordinal Categorical Data
AGRESTI · An Introduction to Categorical Data Analysis
AGRESTI · Categorical Data Analysis, *Second Edition*
ALTMAN, GILL, and McDONALD · Numerical Issues in Statistical Computing for the Social Scientist
AMARATUNGA and CABRERA · Exploration and Analysis of DNA Microarray and Protein Array Data
ANDĚL · Mathematics of Chance
ANDERSON · An Introduction to Multivariate Statistical Analysis, *Third Edition*
* ANDERSON · The Statistical Analysis of Time Series
ANDERSON, AUQUIER, HAUCK, OAKES, VANDAELE, and WEISBERG · Statistical Methods for Comparative Studies
ANDERSON and LOYNES · The Teaching of Practical Statistics
ARMITAGE and DAVID (editors) · Advances in Biometry
ARNOLD, BALAKRISHNAN, and NAGARAJA · Records
* ARTHANARI and DODGE · Mathematical Programming in Statistics
* BAILEY · The Elements of Stochastic Processes with Applications to the Natural Sciences
BALAKRISHNAN and KOUTRAS · Runs and Scans with Applications
BARNETT · Comparative Statistical Inference, *Third Edition*
BARNETT and LEWIS · Outliers in Statistical Data, *Third Edition*
BARTOSZYNSKI and NIEWIADOMSKA-BUGAJ · Probability and Statistical Inference
BASILEVSKY · Statistical Factor Analysis and Related Methods: Theory and Applications
BASU and RIGDON · Statistical Methods for the Reliability of Repairable Systems
BATES and WATTS · Nonlinear Regression Analysis and Its Applications
BECHHOFER, SANTNER, and GOLDSMAN · Design and Analysis of Experiments for Statistical Selection, Screening, and Multiple Comparisons
BELSLEY · Conditioning Diagnostics: Collinearity and Weak Data in Regression

*Now available in a lower priced paperback edition in the Wiley Classics Library.

†Now available in a lower priced paperback edition in the Wiley-Interscience Paperback Series.

- † BELSLEY, KUH, and WELSCH · Regression Diagnostics: Identifying Influential Data and Sources of Collinearity
- BENDAT and PIERSOL · Random Data: Analysis and Measurement Procedures, *Third Edition*
- BERRY, CHALONER, and GEWEKE · Bayesian Analysis in Statistics and Econometrics: Essays in Honor of Arnold Zellner
- BERNARDO and SMITH · Bayesian Theory
- BHAT and MILLER · Elements of Applied Stochastic Processes, *Third Edition*
- BHATTACHARYA and WAYMIRE · Stochastic Processes with Applications
- † BIEMER, GROVES, LYBERG, MATHIOWETZ, and SUDMAN · Measurement Errors in Surveys
- BILLINGSLEY · Convergence of Probability Measures, *Second Edition*
- BILLINGSLEY · Probability and Measure, *Third Edition*
- BIRKES and DODGE · Alternative Methods of Regression
- BLISCHKE AND MURTHY (editors) · Case Studies in Reliability and Maintenance
- BLISCHKE AND MURTHY · Reliability: Modeling, Prediction, and Optimization
- BLOOMFIELD · Fourier Analysis of Time Series: An Introduction, *Second Edition*
- BOLLEN · Structural Equations with Latent Variables
- BOROVKOV · Ergodicity and Stability of Stochastic Processes
- BOULEAU · Numerical Methods for Stochastic Processes
- BOX · Bayesian Inference in Statistical Analysis
- BOX · R. A. Fisher, the Life of a Scientist
- BOX and DRAPER · Empirical Model-Building and Response Surfaces
- * BOX and DRAPER · Evolutionary Operation: A Statistical Method for Process Improvement
- BOX, HUNTER, and HUNTER · Statistics for Experimenters: Design, Innovation, and Discovery, *Second Edition*
- BOX and LUCEÑO · Statistical Control by Monitoring and Feedback Adjustment
- BRANDIMARTE · Numerical Methods in Finance: A MATLAB-Based Introduction
- BROWN and HOLLANDER · Statistics: A Biomedical Introduction
- BRUNNER, DOMHOF, and LANGER · Nonparametric Analysis of Longitudinal Data in Factorial Experiments
- BUCKLEW · Large Deviation Techniques in Decision, Simulation, and Estimation
- CAIROLI and DALANG · Sequential Stochastic Optimization
- CASTILLO, HADI, BALAKRISHNAN, and SARABIA · Extreme Value and Related Models with Applications in Engineering and Science
- CHAN · Time Series: Applications to Finance
- CHARALAMBIDES · Combinatorial Methods in Discrete Distributions
- CHATTERJEE and HADI · Sensitivity Analysis in Linear Regression
- CHATTERJEE and PRICE · Regression Analysis by Example, *Third Edition*
- CHERNICK · Bootstrap Methods: A Practitioner's Guide
- CHERNICK and FRIIS · Introductory Biostatistics for the Health Sciences
- CHILÈS and DELFINER · Geostatistics: Modeling Spatial Uncertainty
- CHOW and LIU · Design and Analysis of Clinical Trials: Concepts and Methodologies, *Second Edition*
- CLARKE and DISNEY · Probability and Random Processes: A First Course with Applications, *Second Edition*
- * COCHRAN and COX · Experimental Designs, *Second Edition*
- CONGDON · Applied Bayesian Modelling
- CONGDON · Bayesian Statistical Modelling
- CONOVER · Practical Nonparametric Statistics, *Third Edition*
- COOK · Regression Graphics
- COOK and WEISBERG · Applied Regression Including Computing and Graphics

*Now available in a lower priced paperback edition in the Wiley Classics Library.

†Now available in a lower priced paperback edition in the Wiley-Interscience Paperback Series.

- COOK and WEISBERG · An Introduction to Regression Graphics
- CORNELL · Experiments with Mixtures, Designs, Models, and the Analysis of Mixture Data, *Third Edition*
- COVER and THOMAS · Elements of Information Theory
- COX · A Handbook of Introductory Statistical Methods
- * COX · Planning of Experiments
- CRESSIE · Statistics for Spatial Data, *Revised Edition*
- CSÖRGÖ and HORVÁTH · Limit Theorems in Change Point Analysis
- DANIEL · Applications of Statistics to Industrial Experimentation
- DANIEL · Biostatistics: A Foundation for Analysis in the Health Sciences, *Eighth Edition*
- * DANIEL · Fitting Equations to Data: Computer Analysis of Multifactor Data, *Second Edition*
- DASU and JOHNSON · Exploratory Data Mining and Data Cleaning
- DAVID and NAGARAJA · Order Statistics, *Third Edition*
- * DEGROOT, FIENBERG, and KADANE · Statistics and the Law
- DEL CASTILLO · Statistical Process Adjustment for Quality Control
- DEMARIS · Regression with Social Data: Modeling Continuous and Limited Response Variables
- DEMIDENKO · Mixed Models: Theory and Applications
- DENISON, HOLMES, MALLICK and SMITH · Bayesian Methods for Nonlinear Classification and Regression
- DETTE and STUDDEN · The Theory of Canonical Moments with Applications in Statistics, Probability, and Analysis
- DEY and MUKERJEE · Fractional Factorial Plans
- DILLON and GOLDSTEIN · Multivariate Analysis: Methods and Applications
- DODGE · Alternative Methods of Regression
- * DODGE and ROMIG · Sampling Inspection Tables, *Second Edition*
- * DOOB · Stochastic Processes
- DOWDY, WEARDEN, and CHILKO · Statistics for Research, *Third Edition*
- DRAPER and SMITH · Applied Regression Analysis, *Third Edition*
- DRYDEN and MARDIA · Statistical Shape Analysis
- DUDEWICZ and MISHRA · Modern Mathematical Statistics
- DUNN and CLARK · Basic Statistics: A Primer for the Biomedical Sciences, *Third Edition*
- DUPUIS and ELLIS · A Weak Convergence Approach to the Theory of Large Deviations
- * ELANDT-JOHNSON and JOHNSON · Survival Models and Data Analysis
- ENDERS · Applied Econometric Time Series
- † ETHIER and KURTZ · Markov Processes: Characterization and Convergence
- EVANS, HASTINGS, and PEACOCK · Statistical Distributions, *Third Edition*
- FELLER · An Introduction to Probability Theory and Its Applications, Volume I, *Third Edition*, Revised; Volume II, *Second Edition*
- FISHER and VAN BELLE · Biostatistics: A Methodology for the Health Sciences
- FITZMAURICE, LAIRD, and WARE · Applied Longitudinal Analysis
- * FLEISS · The Design and Analysis of Clinical Experiments
- FLEISS · Statistical Methods for Rates and Proportions, *Third Edition*
- † FLEMING and HARRINGTON · Counting Processes and Survival Analysis
- FULLER · Introduction to Statistical Time Series, *Second Edition*
- FULLER · Measurement Error Models
- GALLANT · Nonlinear Statistical Models
- GEISSER · Modes of Parametric Statistical Inference
- GEWEKE · Contemporary Bayesian Econometrics and Statistics
- GHOSH, MUKHOPADHYAY, and SEN · Sequential Estimation

*Now available in a lower priced paperback edition in the Wiley Classics Library.

†Now available in a lower priced paperback edition in the Wiley-Interscience Paperback Series.

- GIESBRECHT and GUMPERTZ · Planning, Construction, and Statistical Analysis of Comparative Experiments
- GIFI · Nonlinear Multivariate Analysis
- GIVENS and HOETING · Computational Statistics
- GLASSERMAN and YAO · Monotone Structure in Discrete-Event Systems
- GNANADESIKAN · Methods for Statistical Data Analysis of Multivariate Observations, *Second Edition*
- GOLDSTEIN and LEWIS · Assessment: Problems, Development, and Statistical Issues
- GREENWOOD and NIKULIN · A Guide to Chi-Squared Testing
- GROSS and HARRIS · Fundamentals of Queueing Theory, *Third Edition*
- † GROVES · Survey Errors and Survey Costs
- * HAHN and SHAPIRO · Statistical Models in Engineering
- HAHN and MEEKER · Statistical Intervals: A Guide for Practitioners
- HALD · A History of Probability and Statistics and their Applications Before 1750
- HALD · A History of Mathematical Statistics from 1750 to 1930
- † HAMPEL · Robust Statistics: The Approach Based on Influence Functions
- HANNAN and DEISTLER · The Statistical Theory of Linear Systems
- HEIBERGER · Computation for the Analysis of Designed Experiments
- HEDAYAT and SINHA · Design and Inference in Finite Population Sampling
- HELLER · MACSYMA for Statisticians
- HINKELMANN and KEMPTHORNE · Design and Analysis of Experiments, Volume 1: Introduction to Experimental Design
- HINKELMANN and KEMPTHORNE · Design and Analysis of Experiments, Volume 2: Advanced Experimental Design
- HOAGLIN, MOSTELLER, and TUKEY · Exploratory Approach to Analysis of Variance
- HOAGLIN, MOSTELLER, and TUKEY · Exploring Data Tables, Trends and Shapes
- * HOAGLIN, MOSTELLER, and TUKEY · Understanding Robust and Exploratory Data Analysis
- HOCHBERG and TAMHANE · Multiple Comparison Procedures
- HOCKING · Methods and Applications of Linear Models: Regression and the Analysis of Variance, *Second Edition*
- HOEL · Introduction to Mathematical Statistics, *Fifth Edition*
- HOGG and KLUGMAN · Loss Distributions
- HOLLANDER and WOLFE · Nonparametric Statistical Methods, *Second Edition*
- HOSMER and LEMESHOW · Applied Logistic Regression, *Second Edition*
- HOSMER and LEMESHOW · Applied Survival Analysis: Regression Modeling of Time to Event Data
- † HUBER · Robust Statistics
- HUBERTY · Applied Discriminant Analysis
- HUNT and KENNEDY · Financial Derivatives in Theory and Practice
- HUSKOVA, BERAN, and DUPAC · Collected Works of Jaroslav Hajek—with Commentary
- HUZURBAZAR · Flowgraph Models for Multistate Time-to-Event Data
- IMAN and CONOVER · A Modern Approach to Statistics
- † JACKSON · A User's Guide to Principle Components
- JOHN · Statistical Methods in Engineering and Quality Assurance
- JOHNSON · Multivariate Statistical Simulation
- JOHNSON and BALAKRISHNAN · Advances in the Theory and Practice of Statistics: A Volume in Honor of Samuel Kotz
- JOHNSON and BHATTACHARYYA · Statistics: Principles and Methods, *Fifth Edition*
- JOHNSON and KOTZ · Distributions in Statistics

*Now available in a lower priced paperback edition in the Wiley Classics Library.

†Now available in a lower priced paperback edition in the Wiley-Interscience Paperback Series.

- JOHNSON and KOTZ (editors) · Leading Personalities in Statistical Sciences: From the Seventeenth Century to the Present
- JOHNSON, KOTZ, and BALAKRISHNAN · Continuous Univariate Distributions, Volume 1, *Second Edition*
- JOHNSON, KOTZ, and BALAKRISHNAN · Continuous Univariate Distributions, Volume 2, *Second Edition*
- JOHNSON, KOTZ, and BALAKRISHNAN · Discrete Multivariate Distributions
- JOHNSON, KOTZ, and KEMP · Univariate Discrete Distributions, *Second Edition*
- JUDGE, GRIFFITHS, HILL, LÜTKEPOHL, and LEE · The Theory and Practice of Econometrics, *Second Edition*
- JUREČKOVÁ and SEN · Robust Statistical Procedures: Asymptotics and Interrelations
- JUREK and MASON · Operator-Limit Distributions in Probability Theory
- KADANE · Bayesian Methods and Ethics in a Clinical Trial Design
- KADANE AND SCHUM · A Probabilistic Analysis of the Sacco and Vanzetti Evidence
- KALBFLEISCH and PRENTICE · The Statistical Analysis of Failure Time Data, *Second Edition*
- KASS and VOS · Geometrical Foundations of Asymptotic Inference
- † KAUFMAN and ROUSSEEUW · Finding Groups in Data: An Introduction to Cluster Analysis
- KEDEM and FOKIANOS · Regression Models for Time Series Analysis
- KENDALL, BARDEN, CARNE, and LE · Shape and Shape Theory
- KHURI · Advanced Calculus with Applications in Statistics, *Second Edition*
- KHURI, MATHEW, and SINHA · Statistical Tests for Mixed Linear Models
- * KISH · Statistical Design for Research
- KLEIBER and KOTZ · Statistical Size Distributions in Economics and Actuarial Sciences
- KLUGMAN, PANJER, and WILLMOT · Loss Models: From Data to Decisions, *Second Edition*
- KLUGMAN, PANJER, and WILLMOT · Solutions Manual to Accompany Loss Models: From Data to Decisions, *Second Edition*
- KOTZ, BALAKRISHNAN, and JOHNSON · Continuous Multivariate Distributions, Volume 1, *Second Edition*
- KOTZ and JOHNSON (editors) · Encyclopedia of Statistical Sciences: Volumes 1 to 9 with Index
- KOTZ and JOHNSON (editors) · Encyclopedia of Statistical Sciences: Supplement Volume
- KOTZ, READ, and BANKS (editors) · Encyclopedia of Statistical Sciences: Update Volume 1
- KOTZ, READ, and BANKS (editors) · Encyclopedia of Statistical Sciences: Update Volume 2
- KOVALENKO, KUZNETZOV, and PEGG · Mathematical Theory of Reliability of Time-Dependent Systems with Practical Applications
- LACHIN · Biostatistical Methods: The Assessment of Relative Risks
- LAD · Operational Subjective Statistical Methods: A Mathematical, Philosophical, and Historical Introduction
- LAMPERTI · Probability: A Survey of the Mathematical Theory, *Second Edition*
- LANGE, RYAN, BILLARD, BRILLINGER, CONQUEST, and GREENHOUSE · Case Studies in Biometry
- LARSON · Introduction to Probability Theory and Statistical Inference, *Third Edition*
- LAWLESS · Statistical Models and Methods for Lifetime Data, *Second Edition*
- LAWSON · Statistical Methods in Spatial Epidemiology
- LE · Applied Categorical Data Analysis
- LE · Applied Survival Analysis
- LEE and WANG · Statistical Methods for Survival Data Analysis, *Third Edition*

*Now available in a lower priced paperback edition in the Wiley Classics Library.

†Now available in a lower priced paperback edition in the Wiley-Interscience Paperback Series.

- LePAGE and BILLARD · Exploring the Limits of Bootstrap
 LEYLAND and GOLDSTEIN (editors) · Multilevel Modelling of Health Statistics
 LIAO · Statistical Group Comparison
 LINDVALL · Lectures on the Coupling Method
 LINHART and ZUCCHINI · Model Selection
 LITTLE and RUBIN · Statistical Analysis with Missing Data, *Second Edition*
 LLOYD · The Statistical Analysis of Categorical Data
 LOWEN and TEICH · Fractal-Based Point Processes
 MAGNUS and NEUDECKER · Matrix Differential Calculus with Applications in Statistics and Econometrics, *Revised Edition*
 MALLER and ZHOU · Survival Analysis with Long Term Survivors
 MALLOWS · Design, Data, and Analysis by Some Friends of Cuthbert Daniel
 MANN, SCHAFER, and SINGPURWALLA · Methods for Statistical Analysis of Reliability and Life Data
 MANTON, WOODBURY, and TOLLEY · Statistical Applications Using Fuzzy Sets
 MARCHETTE · Random Graphs for Statistical Pattern Recognition
 MARDIA and JUPP · Directional Statistics
 MASON, GUNST, and HESS · Statistical Design and Analysis of Experiments with Applications to Engineering and Science, *Second Edition*
 McCULLOCH and SEARLE · Generalized, Linear, and Mixed Models
 McFADDEN · Management of Data in Clinical Trials
 * McLACHLAN · Discriminant Analysis and Statistical Pattern Recognition
 McLACHLAN, DO, and AMBROISE · Analyzing Microarray Gene Expression Data
 McLACHLAN and KRISHNAN · The EM Algorithm and Extensions
 McLACHLAN and PEEL · Finite Mixture Models
 McNEIL · Epidemiological Research Methods
 MEEKER and ESCOBAR · Statistical Methods for Reliability Data
 MEERSCHAERT and SCHEFFLER · Limit Distributions for Sums of Independent Random Vectors: Heavy Tails in Theory and Practice
 MICKEY, DUNN, and CLARK · Applied Statistics: Analysis of Variance and Regression, *Third Edition*
 * MILLER · Survival Analysis, *Second Edition*
 MONTGOMERY, PECK, and VINING · Introduction to Linear Regression Analysis, *Third Edition*
 MORGENTHALER and TUKEY · Configural Polysampling: A Route to Practical Robustness
 MUIRHEAD · Aspects of Multivariate Statistical Theory
 MULLER and STOYAN · Comparison Methods for Stochastic Models and Risks
 MURRAY · X-STAT 2.0 Statistical Experimentation, Design Data Analysis, and Nonlinear Optimization
 MURTHY, XIE, and JIANG · Weibull Models
 MYERS and MONTGOMERY · Response Surface Methodology: Process and Product Optimization Using Designed Experiments, *Second Edition*
 MYERS, MONTGOMERY, and VINING · Generalized Linear Models. With Applications in Engineering and the Sciences
 † NELSON · Accelerated Testing, Statistical Models, Test Plans, and Data Analyses
 † NELSON · Applied Life Data Analysis
 NEWMAN · Biostatistical Methods in Epidemiology
 OCHI · Applied Probability and Stochastic Processes in Engineering and Physical Sciences
 OKABE, BOOTS, SUGIHARA, and CHIU · Spatial Tesselations: Concepts and Applications of Voronoi Diagrams, *Second Edition*
 OLIVER and SMITH · Influence Diagrams, Belief Nets and Decision Analysis
 PALTA · Quantitative Methods in Population Health: Extensions of Ordinary Regressions

*Now available in a lower priced paperback edition in the Wiley Classics Library.

†Now available in a lower priced paperback edition in the Wiley-Interscience Paperback Series.

- PANKRATZ · Forecasting with Dynamic Regression Models
- PANKRATZ · Forecasting with Univariate Box-Jenkins Models: Concepts and Cases
- * PARZEN · Modern Probability Theory and Its Applications
- PEÑA, TIAO, and TSAY · A Course in Time Series Analysis
- PIANTADOSI · Clinical Trials: A Methodologic Perspective
- PORT · Theoretical Probability for Applications
- POURAHMADI · Foundations of Time Series Analysis and Prediction Theory
- PRESS · Bayesian Statistics: Principles, Models, and Applications
- PRESS · Subjective and Objective Bayesian Statistics, *Second Edition*
- PRESS and TANUR · The Subjectivity of Scientists and the Bayesian Approach
- PUKELSHEIM · Optimal Experimental Design
- PURI, VILAPLANA, and WERTZ · New Perspectives in Theoretical and Applied Statistics
- † PUTERMAN · Markov Decision Processes: Discrete Stochastic Dynamic Programming
- QIU · Image Processing and Jump Regression Analysis
- * RAO · Linear Statistical Inference and Its Applications, *Second Edition*
- RAUSAND and HØYLAND · System Reliability Theory: Models, Statistical Methods, and Applications, *Second Edition*
- RENCHER · Linear Models in Statistics
- RENCHER · Methods of Multivariate Analysis, *Second Edition*
- RENCHER · Multivariate Statistical Inference with Applications
- * RIPLEY · Spatial Statistics
- RIPLEY · Stochastic Simulation
- ROBINSON · Practical Strategies for Experimenting
- ROHATGI and SALEH · An Introduction to Probability and Statistics, *Second Edition*
- ROLSKI, SCHMIDLI, SCHMIDT, and TEUGELS · Stochastic Processes for Insurance and Finance
- ROSENBERGER and LACHIN · Randomization in Clinical Trials: Theory and Practice
- ROSS · Introduction to Probability and Statistics for Engineers and Scientists
- † ROUSSEEUW and LEROY · Robust Regression and Outlier Detection
- * RUBIN · Multiple Imputation for Nonresponse in Surveys
- RUBINSTEIN · Simulation and the Monte Carlo Method
- RUBINSTEIN and MELAMED · Modern Simulation and Modeling
- RYAN · Modern Regression Methods
- RYAN · Statistical Methods for Quality Improvement, *Second Edition*
- SALTELLI, CHAN, and SCOTT (editors) · Sensitivity Analysis
- * SCHEFFE · The Analysis of Variance
- SCHIMEK · Smoothing and Regression: Approaches, Computation, and Application
- SCHOTT · Matrix Analysis for Statistics, *Second Edition*
- SCHOUTENS · Levy Processes in Finance: Pricing Financial Derivatives
- SCHUSS · Theory and Applications of Stochastic Differential Equations
- SCOTT · Multivariate Density Estimation: Theory, Practice, and Visualization
- * SEARLE · Linear Models
- SEARLE · Linear Models for Unbalanced Data
- SEARLE · Matrix Algebra Useful for Statistics
- SEARLE, CASELLA, and McCULLOCH · Variance Components
- SEARLE and WILLETT · Matrix Algebra for Applied Economics
- SEBER and LEE · Linear Regression Analysis, *Second Edition*
- † SEBER · Multivariate Observations
- † SEBER and WILD · Nonlinear Regression
- SENNOTT · Stochastic Dynamic Programming and the Control of Queueing Systems
- * SERFLING · Approximation Theorems of Mathematical Statistics
- SHAFER and VOVOK · Probability and Finance: It's Only a Game!

*Now available in a lower priced paperback edition in the Wiley Classics Library.

†Now available in a lower priced paperback edition in the Wiley-Interscience Paperback Series.

- SILVAPULLE and SEN · Constrained Statistical Inference: Inequality, Order, and Shape Restrictions
- SMALL and McLEISH · Hilbert Space Methods in Probability and Statistical Inference
- SRIVASTAVA · Methods of Multivariate Statistics
- STAPLETON · Linear Statistical Models
- STAUDTE and SHEATHER · Robust Estimation and Testing
- STOYAN, KENDALL, and MECKE · Stochastic Geometry and Its Applications, *Second Edition*
- STOYAN and STOYAN · Fractals, Random Shapes and Point Fields: Methods of Geometrical Statistics
- STYAN · The Collected Papers of T. W. Anderson: 1943–1985
- SUTTON, ABRAMS, JONES, SHELDON, and SONG · Methods for Meta-Analysis in Medical Research
- TANAKA · Time Series Analysis: Nonstationary and Noninvertible Distribution Theory
- THOMPSON · Empirical Model Building
- THOMPSON · Sampling, *Second Edition*
- THOMPSON · Simulation: A Modeler's Approach
- THOMPSON and SEBER · Adaptive Sampling
- THOMPSON, WILLIAMS, and FINDLAY · Models for Investors in Real World Markets
- TIAO, BISGAARD, HILL, PEÑA, and STIGLER (editors) · Box on Quality and Discovery: with Design, Control, and Robustness
- TIERNEY · LISP-STAT: An Object-Oriented Environment for Statistical Computing and Dynamic Graphics
- TSAY · Analysis of Financial Time Series
- UPTON and FINGLETON · Spatial Data Analysis by Example, Volume II: Categorical and Directional Data
- VAN BELLE · Statistical Rules of Thumb
- VAN BELLE, FISHER, HEAGERTY, and LUMLEY · Biostatistics: A Methodology for the Health Sciences, *Second Edition*
- VESTRUP · The Theory of Measures and Integration
- VIDAKOVIC · Statistical Modeling by Wavelets
- VINOD and REAGLE · Preparing for the Worst: Incorporating Downside Risk in Stock Market Investments
- WALLER and GOTWAY · Applied Spatial Statistics for Public Health Data
- WEERAHANDI · Generalized Inference in Repeated Measures: Exact Methods in MANOVA and Mixed Models
- WEISBERG · Applied Linear Regression, *Third Edition*
- WELSH · Aspects of Statistical Inference
- WESTFALL and YOUNG · Resampling-Based Multiple Testing: Examples and Methods for *p*-Value Adjustment
- WHITTAKER · Graphical Models in Applied Multivariate Statistics
- WINKER · Optimization Heuristics in Economics: Applications of Threshold Accepting
- WONNACOTT and WONNACOTT · Econometrics, *Second Edition*
- WOODING · Planning Pharmaceutical Clinical Trials: Basic Statistical Principles
- WOODWORTH · Biostatistics: A Bayesian Introduction
- WOOLSON and CLARKE · Statistical Methods for the Analysis of Biomedical Data, *Second Edition*
- WU and HAMADA · Experiments: Planning, Analysis, and Parameter Design Optimization
- YANG · The Construction Theory of Denumerable Markov Processes
- * ZELLNER · An Introduction to Bayesian Inference in Econometrics
- ZHOU, OBUCHOWSKI, and McCLISH · Statistical Methods in Diagnostic Medicine

*Now available in a lower priced paperback edition in the Wiley Classics Library.

†Now available in a lower priced paperback edition in the Wiley-Interscience Paperback Series.