DANIEL W. STROOCK
S. R. S. VARADHAN

Volume 233

Grundlehren
der mathematischen
Wissenschaften

A Series of
Comprehensive Studies
in Mathematics

MULTIDIMENSIONAL DIFFUSION PROCESSES

多维扩散过程

Springer

世界图书出版公司
www.wpcbj.com.cn

Daniel W. Stroock
S.R. Srinivasa Varadhan

Multidimensional Diffusion Processes

Springer

图书在版编目（CIP）数据

多维扩散过程=Multidimensional Diffusion Processes：英文/（美）斯出科（Stroock，d. w.）著.—北京：世界图书出版公司北京公司，2009．3
ISBN 978-7-5062-9262-7

Ⅰ.多… Ⅱ.斯… Ⅲ.扩散过程-英文 Ⅳ.0211.6

中国版本图书馆CIP数据核字（2009）第024056号

书　　名：	Multidimensional Diffusion Processes
作　　者：	Daniel W. Stroock & S. R. S. Varadhan
中 译 名：	多维扩散过程
责任编辑：	高蓉 刘慧
出 版 者：	世界图书出版公司北京公司
印 刷 者：	三河国英印务有限公司
发　　行：	世界图书出版公司北京公司（北京朝内大街 137 号　100010）
联系电话：	010-64015659
电子信箱：	kjb@wpcbj.com.cn
开　　本：	24开
印　　张：	15
版　　次：	2009 年 03月
版权登记：	图字:01-2009-0693
书　　号：	978-7-5062-9262-7 / O・643　　定　价: 55.00 元

Daniel W. Stroock
Massachusetts Institute of Technology
Department of Mathematics
77 Massachusetts Ave
Cambridge, MA 02139-4307
USA
e-mail: dws@math.mit.edu

S.R. Srinivasa Varadhan
New York University
Courant Institute of Mathematical Sciences
251 Mercer Street
New York, NY 10012
USA
e-mail: varadhan@cims.nyu.edu

Cataloging-in-Publication Data applied for

A catalog record for this book is available from the Library of Congress.

Bibliographic information published by Die Deutsche Bibliothek
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the Internet at http://dnb.ddb.de

Mathematics Subject Classification (2000): 60J60, 28A65

ISSN 0072-7830

ISBN 3-540-90353-4 Springer-Verlag Berlin Heidelberg New York

Springer-Verlag Berlin Heidelberg New York
a member of BertelsmannSpringer Science+Business Media GmbH

http://www.springer.de

To our parents:

Katherine W. Stroock
Alan M. Stroock
S.R. Janaki
S.V. Ranga Ayyangar

Frequently Used Notation

I. Topological Notation. Let (X, ρ) be a separable metric space.

1) B° is the interior of $B \subseteq X$.
2) $\bar{B}$ is the closure of $B \subseteq X$.
3) ∂B is the boundary of $B \subseteq X$.
4) $\mathscr{B}_X$ is the Borel field of subsets of X.
5) $C_b(X)$ is the set of bounded continuous functions $f: X \to R$.
6) $B(X)$ is the set of bounded $\mathscr{B}_X$-measurable $f: X \to R$.
7) $U_\rho(X)$ is the set of bounded ρ-uniformly continuous $f: X \to R$.
8) $M(X)$ is the set of probability measures on $(X, \mathscr{B}_X)$.
9) $\|f\| = \sup_{x \in X} |f(x)|$ for $f \in B(X)$.

II. Special Notation for Euclidean Spaces

1) R^d is d-dimensional Euclidean space.
2) $|x| = (\sum_1^d x_j^2)^{1/2}$ for $x \in R^d$.
3) $B(x, r) = \{y \in R^d: |x - y| < r\}$.
4) $\langle x, y \rangle = \sum_1^d x_j y_j$ for $x, y \in R^d$.
5) $S^{d-1} = \{x \in R^d: |x| = 1\}$.
6) $\hat{C}(R^d) = \{f \in C_b(R^d): \lim_{|x| \to \infty} f(x) = 0\}$.
7) $C_0(\mathscr{G})$ is the set of $f \in C_b(\mathscr{G})$ having compact support.
8) $C_b^m(\mathscr{G})$ is the set of $f: \mathscr{G} \to R$ possessing bounded continuous derivatives of order up to and including m.
9) $C_b^\infty(\mathscr{G}) = \bigcap_{m=0}^{\infty} C_b^m(\mathscr{G})$.
10) $C^\infty(\mathscr{G})$ is the set of $f: \mathscr{G} \to R$ possessing continuous derivatives of all orders.
11) $C_0^\infty(\mathscr{G}) = C^\infty(\mathscr{G}) \cap C_0(\mathscr{G})$.
12) $C_b^{m,n}(\mathscr{G})$ for $\mathscr{G} \subseteq [0, \infty) \times R^d$ is the set of $f: \mathscr{G} \to R$ such that f has m bounded continuous time derivatives and bounded continuous spacial derivatives of order less than or equal to n.
13) $L^p(\mathscr{G})$, $1 \le p \le \infty$, is the usual L^p-space defined in terms of Lebesgue measure on $\mathscr{G}$.

14) $L^p_{loc}(\mathscr{G})$ is the set of $f: \mathscr{G} \to R$ (or $\mathbb{C}$) such that $f \in L^p(K)$ for all compact K in $\mathscr{G}$.

III. Path Spaces Notation

1) $C(I, R^d)$ for $I \subseteq [0, \infty)$ is the set of R^d-valued functions on I into R^d.
2) $\Omega_d(\Omega)$ (see p. 30).
3) $\mathscr{M}_t(\mathscr{M})$ (see p. 30).
4) $x(t, \omega)$ (see p. 30).

IV. Miscellaneous Notation

1) $a \wedge b$ is the smaller of the numbers $a, b \in R$.
2) $a \vee b$ is the larger of the numbers $a, b \in R$.
3) S_d is the set of symmetric non-negative definite $d \times d$ real matrices.
4) S_d^+ is the set of nondegenerate elements of S_d.
5) $\|A\|$, where A is a square matrix, and is the operator norm of A.
6) $\sigma(\mathscr{C})$, where $\mathscr{C}$ is a collection of subsets of X, and is the smallest σ-algebra over X containing $\mathscr{C}$.
7) $\sigma(\mathscr{F})$, where $\mathscr{F}$ is a set of functions on X into a measurable space, and is the smallest σ-algebra over X with respect to which every element of $\mathscr{F}$ is measurable.
8) $[\lambda]$, $\lambda \in R$, is the integral part of λ.
9) $\xi \sim \Gamma_d^s(a, b)$ (see p. 92).

Contents

Chapter 0

Introduction

The main purpose of this book is to elucidate the martingale approach to the theory of Markov processes. Needless to say, we believe that the approach has many advantages over previous ones, and it is our hope that the present book will convince some other people that this is indeed so. When we began this project we were uncertain whether to proselytize by intimidating the reader with myriad examples demonstrating the full scope of the techniques or by persuading him with a careful treatment of just one problem to which they apply. We have decided on the latter plan in the belief that it is preferable to bore than to batter. The result is that we have devoted what may seem like an inordinate number of pages to a rather special topic. On the other hand, we have endeavoured to present our proofs in such a way that the techniques involved should lend themselves to easy adaptation in other contexts. Only time will tell if we have succeeded.

The topic which we have chosen is that of diffusion theory in R^d. In order to understand how this subject fits into the general theory of Markov processes, it is best to return to Lévy's ideas about "stochastic differentials." Let $x(\cdot)$ be a Markov process with values in R^d and suppose that for $t \geq 0$ and test functions $\varphi \in C_0^\infty(R^d)$

$$(0.1) \quad E[\varphi(x(t+h)) - \varphi(x(t)) \mid x(s), s \leq t] = hL_t\varphi(x(t)) + o(h), \qquad h > 0,$$

where, for each $t \geq 0$, L_t is a linear operator on $C_0^\infty(R^d)$ into $C_b(R^d)$. It is obvious that L_t must satisfy the weak maximum principle, since if φ achieves its maximum at x^0 then $E[\varphi(x(t+h)) - \varphi(x(t)) \mid x(t) = x^0] \leq 0$. Moreover, if $x^0 \in R^d$ and $\eta \in C_0^\infty(R^d)$ is such that $\eta(x^0) = 1$, $0 \leq \eta \leq 1$, and η is identically zero outside an ε-neighborhood of x^0, then $P(|x(t+h) - x^0| \geq \varepsilon \mid x(t) = x^0) \leq E[(1 - \eta(x(t+h))) \mid x(t) = x^0] \leq h|L_t\eta(x^0)| + o(h)$. Thus if $\varphi \in C_0^\infty(R^d)$ vanishes in a neighborhood of x^0, then $|L_t\varphi(x^0)| \in C_\varepsilon\|\varphi\|$. That is to say, L_t is "quasilocal." Therefore, if we now define the operator L_{t,x^0} by the relation

$$L_{t,x^0}\varphi(x) = [L_t(\tau_{x-x^0}\varphi)](x^0), \qquad x \in R^d,$$

where $\tau_y\varphi(\cdot) = \varphi(\cdot + y)$, then L_{t,x^0} is a quasi-local, translation invariant operator satisfying the maximum principle. This class of operators is well-known and can be shown to coincide with the class of generators of time homogeneous independent increment processes (cf. Hille and Philips [1957]). In particular, we can conclude that

$$L_t\varphi(x^0) = \frac{1}{2}\sum_{i,j=1}^{d} a^{ij}(t, x^0)\frac{\partial^2\varphi}{\partial x_i\,\partial x_j}(x^0) + \sum_{i=1}^{d} b^i(t, x^0)\frac{\partial\varphi}{\partial x_i}(x^0)$$
$$+ \int\left(\varphi(x^0 + y) - \varphi(x^0) + \frac{\langle y, \nabla\varphi(x^0)\rangle}{1 + |y|^2}\right)M(t, x^0; dy)$$

where $((a^{ij}(t, x^0)))$ is non-negative definite and $M(t, x^0; \cdot)$ is a Lévy measure. More important, we develop from these considerations the intuitive picture of the process $x(\cdot)$ leaving $x(t)$ like the independent increment process with characteristics $a(t, x(t))$, $b(t, x(t))$, and $M(t, x(t); \cdot)$. Throughout this book we will be restricting ourselves to continuous Markov processes. For a continuous process, the Lévy measure M must be absent. That is, if $x(\cdot)$ is a continuous Markov process and (0.1) obtains, then

(0.2) $$L_t = \frac{1}{2}\sum_{i,j=1}^{d} a^{ij}(t, x)\frac{\partial^2}{\partial x_i\,\partial x_j} + \sum_{1}^{d} b^i(t, x)\frac{\partial}{\partial x_i}$$

and for small $h > 0$, $x(t + h) - x(t)$ is like the Gaussian independent increment process having mean $b(t, x(t))$ and covariance $a(t, x(t))$. (A slightly different presentation of these ideas is given in the introduction to Itô [1951]. We recommend Itô's discussion to the interested reader.)

The structure of this book can now be explained in terms of the ideas introduced in the preceding paragraph. Starting from (0.1), various tacks toward an understanding of the process $x(\cdot)$ suggest themselves. The most analytic of these is the following. Let $P(s, x; t, \cdot)$ denote the transition probability function determined by $x(\cdot)$ (i.e., $P(s, x; t, \Gamma) = P(x(t) \in \Gamma \mid x(s) = x)$). From (0.1), we see that

$$\frac{\partial}{\partial t}\int P(s, x; t, dy)\varphi(y) = \lim_{h\downarrow 0}\int P(s, x; t, dz) \times \int P(t, z; t + h, dy)(\varphi(y) - \varphi(z))$$
$$= \int P(s, x; t, dy)L_t\varphi(y).$$

Of course, we have used the Chapman-Kolmogorov equation. From this we derive the formal relation:

(0.3) $$\frac{\partial}{\partial t}P(s, x; t, \cdot) = L_t^* P(s, x; t, \cdot), \qquad t > s,$$

where L_t^* is the formal adjoint of L_t. Equation (0.3) is called the *forward equation* (in physics and engineering literature it is often referred to as the *Fokker-Planck* equation). Since it is clear that

(0.3′) $$\lim_{t\downarrow s} P(s, x; t, \cdot) = \delta_x(\cdot),$$

it is reasonable to suppose that one can recapture $P(s, x; t, \cdot)$ from (0.3) and (0.3′). Indeed, this was done with great success by Kolmogorov [1931] and Feller [1936] in their pioneering work on this subject. However, there are severe technical problems with (0.3). In particular, one must tacitly assume that $P(s, x; t, \cdot)$ admits a density $p(s, x; t, y)$ and think of (0.3) as being an equation for $p(s, x; t, y)$ as a function of t and y; and even when such an assumption is justified, there remain inherent difficulties in the interpretation of L_t^* unless the coefficients are smooth. For this reason, people turned their attention to the *backward equation*. Namely, starting once again from (0.1) we have

$$\begin{aligned} -\frac{\partial u}{\partial s}(s, x) &= -\frac{\partial}{\partial s}\int P(s, x; t, dy)\varphi(y) \\ &= \lim_{h\downarrow 0}\frac{1}{h}\int P(s-h, x; s, dy)(u(s, y) - u(s, x)) \\ &= L_s u(s, x), \end{aligned}$$

where $u(s, x) = \int P(s, x; t, dy)\varphi(y)$, $0 \le s < t$. (Notice that the preceding computation is not fully justified since we do not know that $u(s, \cdot)$ is in our test-function space. Nonetheless, the argument is correct in spirit.) Hence we arrive at

$$\frac{\partial}{\partial s}P(s, x; t, \cdot) + L_s P(s, x; t, \cdot) = 0, \qquad 0 \le s < t, \tag{0.4}$$

$$\lim_{s\uparrow t} P(s, x; t, \cdot) = \delta_x(\cdot). \tag{0.4′}$$

(It should be clear why (0.3) is the forward equation and (0.4) is the backward equation: (0.3) involves the forward (i.e., future) variables whereas (0.4) involves the backward (i.e., past) variables.) Again one might suspect that (0.4) and (0.4′) determine $P(s, x; t, \cdot)$ and now there are no problems about interpretation. The study of diffusion theory via the backward equation has been one of the more powerful and successful approaches to the subject and we have included a sketch of this procedure in Chapters 2 and 3.

The major objection to the study of diffusion theory by the method just described is that the hard machinery used comes from the theory of partial differential equations and the probabilistic input is relatively small. A more probabilistically satisfactory approach was suggested by Lévy and carried out by Itô [1951]. The idea here is to return to the intuitive picture of $x(t + h) - x(t)$, for small $h > 0$, looking like the Gaussian independent increment process with drift $b(t, x(t))$ and covariance $a(t, x(t))$. In differential form, this intuitive picture means that

$$dx(t) = \sigma(t, x(t))\, d\beta(t) + b(t, x(t))\, dt \tag{0.5}$$

where $\beta(\cdot)$ is a d-dimensional Brownian motion and σ is a square root of a. Indeed, $\sigma(t, x(t))(\beta(t + h) - \beta(t)) + b(t, x(t))$ will be just such a Gaussian process;

and if $\{x(s),\ 0 \le s < t\}$ is $\{\beta(s): 0 \le s \le t\}$-measurable, then $\sigma(t, x(t)) \times (\beta(t + h) - \beta(t)) + b(t, x(t))$ will be conditionally independent of $\{x(s): 0 \le s < t\}$ given $x(t)$. There are two problems of considerable technical magnitude raised by (0.5). First and foremost is the question of interpretation. Since a Brownian path is nowhere differentiable it is by no means obvious that sense can be made out of a differential equation like (0.5). Secondly, even if one knows what (0.5) means, one still has to learn how to solve such an equation before it can be considered to be useful. Both these problems were masterfully handled by Itô, a measure of the success of his solution is the extent to which it is still used. We develop Itô's theory of stochastic integration in Chapter 4 and apply it to equations like (0.5) in Chapter 5.

With Chapter 6 we begin the study of diffusion theory along the lines initiated by us in Stroock and Varadhan [1969]. In order to understand this approach, we return once again to (0.1). From (0.1), it is easy to deduce that:

$$\begin{aligned}\frac{d}{dt_2} & E[\varphi(x(t_2)) \mid x(s),\ s \le t_1] \\ &= \lim_{h \downarrow 0} \frac{1}{h} E[E[\varphi(x(t_2 + h)) - \varphi(x(t_2)) \mid x(s),\ s \le t_2] \mid x(s),\ s \le t_1] \\ &= E[L_{t_2}\varphi(x(t_2)) \mid x(s),\ s \le t_1].\end{aligned}$$

Thus

$$E[\varphi(x(t_2)) - \varphi(x(t_1)) - \int_{t_1}^{t_2} L_t\varphi(x(t))\, dt \mid x(s),\ s \le t_1] = 0;$$

or in other words

$$X_\varphi(t) \equiv \varphi(x(t)) - \int_0^t L_s\varphi(x(s))\, ds \tag{0.6}$$

is a martingale for all test functions φ. (Notice that the line of reasoning leading from (0.1) to (0.6) is essentially the same as that from (0.1) to the forward equation.) One can now ask if the property that $X_\varphi(\cdot)$ is a martingale for all test functions φ uniquely characterizes the process $x(\cdot)$ apart from specifying $x(0)$. To be more precise, given L_t, consider the following problems:

(*i*) Is there for each $x \in R^d$ a probability measure P on $C([0, \infty), R^d)$ such that $P(x(0) = x) = 1$ and $X_\varphi(\cdot)$ is a martingale for all test functions φ?

and

(*ii*) Is there at most one such P for each x?

Problems (*i*) and (*ii*) constitute what we call the *martingale problem for* L_t. Of course problems (*i*) and (*ii*) are interesting only if one can also answer

(*iii*) If (*i*) and (*ii*) have affirmative answers, what conclusions can be drawn?

To convince oneself that these are reasonable questions, one should recall that in the case when $d = 1$ and $L_t = \frac{1}{2}d^2/dx^2$, Lévy (cf. Doob [1953] or Exercise 4.6.6) characterized Wiener measure as the unique probability measure P on

$C([0,\infty), R^1)$ such that $P(x(0)=0)=1$ and $x(t)$ and $x^2(t)-t$ are martingales. That is, he showed that in this case one only needs the functions $\varphi(x)=x$ and $\psi(x)=x^2$. (Actually [cf. Exercise 4.6.6], this is a general phenomenon, since under general conditions one can show that $X_\varphi(\cdot)$ is a martingale for all test functions φ if $X_{\varphi_j}(\cdot)$ and $X_{\psi_{ij}}(\cdot)$ are martingales for $\varphi_j(x)=x_j$ and $\psi_{ij}(x)=x_i x_j$, $1\le i,j\le d$. When $d=1$, this general phenomenon was already pointed out in Chapter 9 of Doob [1953].) Furthermore, one should remember that much of Doob's beautiful work in potential theory relies heavily on the observation that $X_\varphi(\cdot)$ is a martingale when P is the process associated with L_t. Thus what if anything is truly new about our approach is that we have made this observation the cornerstone of our theory and asked if in fact it does not underlie the whole structure of diffusion theory.

In Chapter 6 we lay the foundation for everything that follows. In particular, we prove there a basic existence theorem for solutions to the martingale problem. Once this is done, we start laying the groundwork for out attack on the question of uniqueness by deriving general conclusions that can be drawn about a P on $C([0,\infty), R^d)$ under the assumption that $X_\varphi(\cdot)$ is a martingale for all test functions φ. These include the relationship between the martingale problem and the (strong) Markov property, as well as the formula of Cameron, Martin and Girsanov.

Chapter 7 contains a proof of our best general theorem about uniqueness for the martingale problem. What we show is that if the coefficients a and b in L_t are bounded and measurable and for each $T>0$ and $R>0$, a satisfies

$$\inf_{\substack{0\le s\le T\\ |x|\le R}} \inf_{\theta\in R^d\setminus\{0\}} \frac{\langle\theta, a(s,x)\theta\rangle}{|\theta|^2} > 0 \tag{0.7}$$

and

$$\lim_{\delta\downarrow 0} \sup_{\substack{0\le s\le T\\ |x^1|,|x^2|\le R\\ |x^1-x^2|<\delta}} \|a(s,x^1)-a(s,x^2)\| = 0, \tag{0.8}$$

then the martingale problem for L_t is well-posed (i.e., existence and uniqueness hold). As a dividend of our proof, we show that L_t determines a strong Markov, strongly Feller continuous process.

The contents of Chapter 8 are somewhat tangential to the main thrust of our development. What we do there is expand on the theme initiated in Watanabe and Yamada [1971] in their investigation of the relationship between Itô's approach and the martingale problem.

In Chapter 9 we return to L_t's having coefficients of the sort studied in Chapter 7. Here we take advantage of certain analytic relations and estimates upon which our proof of uniqueness in Chapter 7 turns. In brief, the results of these considerations are various L^p-estimates for the transition probability function of the process determined by L_t.

Chapter 10 extends the martingale problem approach to unbounded coefficients. The point made here is that this extension is elementary, provided one can show that the diffusion process does not "explode." We give some standard conditions that can be used to test for explosion.

Again in Chapter 11 we deal with L_t's of the sort studied in Chapters 7 and 10. This time we are interested in stability results for the associated processes. These results can be naturally divided into two categories: convergence of Markov chains to diffusions (i.e., invariance principles of the sort initiated by Erdös and Kac and perfected by Donsker) and convergence of diffusions to other diffusions. Both categories are surprisingly easy to handle given the results of Chapters 7 and 9.

The final chapter, Chapter 12, takes up the question of what can be done in those circumstances when existence of solutions to a martingale problem can be proved but uniqueness cannot. The idea here, is to make a careful "selection" of solutions so that they fit together into a Markov family. The procedure that we use goes back to Krylov [1973]. We also show in Chapter 12 that every solution to a given martingale problem can, in some sense, be built out of those solutions which are part of a Markov family.

The only parts of the book which we have not yet discussed are the beginning and the end. Chapter 1 provides an introduction to those parts of measure and probability theories which we consider most important for an understanding of this book. Although the material here is not new, much of it has been reworked. In particular, our criteria for compactness in Section 1.4 strikes us as a useful variation on the ideas of Prohorov.

Finally, in spite of our attempt to make it look as if it were, the appendix is not probability theory. Instead, it is that part of the theory of singular integrals on which we rely in Chapters 7 and 9. At the present time, one has to depend on these results from outside probability theory and we have provided a proof in the Appendix in order to make the book self-contained.

It is now time for us to thank the many people and organizations to whom we are deeply indebted. The original work out of which this book grew was performed while both of us were at the Courant Institute of Mathematical Sciences. During that period we were encouraged and stimulated by many people, particularly: M. Kac, H. P. McKean Jr., S. Sawyer, M. D. Donsker and L. Nirenberg; and we were supported by grants from the Air Force, the Sloan and Ford foundations as well as general C.I.M.S. funds. Whether this book would ever see the light of day was cast into considerable doubt by the departure from C.I.M.S. of one of us to the Rocky Mountains in 1972. At that time not a sentence of it had been written. However, in 1976 we had the good fortune to visit Paris together under the auspices of Professors Neveu and Revuz; and it was at that time (much to the dismay of an accompanying wife) that we actually began to write this book. Progress from that point on has been slow but steady. During the interim we have incurred a considerable debt of gratitude to several people: wives Lucy and Vasu; secretaries Janice Morgenstern, Gloria Lee, Susan Parris and Helen Samoraj; students Marty Day and Pedro Echeverria; colleagues Richard Holley, G. Papanicolaou, M. D. Donsker, and E. Fabes; gadfly J. Doob, and publisher Springer Verlag. To all these we extend our heart felt thanks along with the promise that they do not necessarily have to read what we have written.

Chapter 1

Preliminary Material: Extension Theorems, Martingales, and Compactness

1.0. Introduction

As mentioned in the Introduction, the point of view that we take will involve us in a detailed study of measures on function spaces. There are a few basic tools which are necessary for the construction of such measures. The purpose of this chapter is to develop these tools. In the process, we will introduce some notions (e.g., conditioning and martingales) which will play an important role in what follows. Section 1.1 contains the basic theorem of Prohorov and Varadarajan characterizing weakly compact families of measures on a Polish space. Using their results, we prove the existence of conditional probability distributions. The final topics in Section 1.1 are the extension theorems of Tulcea and Kolmogorov. Section 1.2 introduces the notions of progressively measurable functions and martingales. In connection with martingales we prove Doob's inequality, his stopping time theorem and a useful integration by parts formula. Finally we prove a result connecting martingale theory and conditioning.

In Section 1.3 we specialize the results of Section 1.1 to the case when our Polish space is $C([0, \infty); R^d)$ (i.e., the space of R^d valued continuous functions on $[0, \infty)$ with the natural topology induced by uniform convergence on bounded intervals). Section 1.4 contains a useful sufficient condition for compactness of a family of measures on $C([0, \infty); R^d)$ in terms of certain martingales associated with them.

1.1. Weak Convergence, Conditional Probability Distributions, and Extension Theorems

Throughout this section (X, D) will stand for a Polish space (i.e., a complete separable metric space) and $\mathscr{B} = \mathscr{B}_X$ its Borel σ-field. We denote by $M(X)$ the set of all probability measures on $(X, \mathscr{B})$ and by $C_b(X)$ the set of all bounded continuous functions on X. We will view $M(X)$ as a subset of the dual space of $C_b(X)$ and give it the inherited weak* topology. It will turn out that this topology makes $M(X)$ into a metric space.

1.1.1 Theorem. *Let $\mu_n \in M(X)$ for each $n \geq 1$. Given $\mu \in M(X)$, the following are equivalent:*