Lectures on

Markov Processes and their Associated Semi Groups

by

Walter A. Rosenkrantz
Purdue University

Prepared with the Assistance of C. E. Ford

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series No. 488

April 1977

Lectures on

Markov Processes and their Associated Semi groups.

by

Walter A. Rosenkrantz

Prepared with the assistance of C. E. Ford

Preface

In his fundamental paper "Über die Analytischen methoden in der Wahrscheinlichkeitsrechnung" Math. Ann., V.104, 415-458, 1931, Kolmogorov derived his celebrated backward and forward differential equations for Markov processes x(t). If x(t) is a one dimensional Markov process which is homogeneous in time then the backward equation takes the form:

$$u_t(t,x) = (a(x)/2)u_{xx}(t,x) + b(x)u_x(t,x)$$

(1)

$$u(0,x) = f(x).$$

In practice, however, one is not given the Markov process, instead one is given only the diffusion and drift coefficients a(x) and b(x) respectively, and the problem is to construct the Markov process corresponding to these coefficients - this is called the existence problem. If the state space of the Markov process is R^1, a satisfactory theory can be constructed via the stochastic differential calculus of Ito. In genetics and analysis there occur Markov processes whose state space is a subinterval $I = [r_0,r_1]$ of R^1 and the question arises how to define the process when, if ever, it reaches the boundaries. In addition there are Markov processes whose associated Kolmogorov differential equation involve non-classical generalized second order operators. For such questions the Ito calculus is inadequate and a different approach due to Hille and Feller is, in the author's opinion, more successful. This approach is primarily analytic and relies heavily on semi-group theory, the essentials of which are given in Chapter III. The author's debt to Dynkin's masterful presentation of this material is obvious and need not be eleaborated on here. We do, however, include quite a bit of material not to be found in Dynkin or any other treatise on Markov processes e.g. the Trotter-Kato theorem, the Trotter product formula and some perturbation theory - ideas which play an important role in limit theorems for Markov processes, as well as existence theorems. Indeed one of the themes of these lectures is that a strong existence theorem leads to a strong limit theorem; this is why in Chapter IV we give a careful discussion, following Mandl, of the "stationary equation":

(2) $\lambda F_\lambda(x)-(a(x)/2)F''_\lambda(x)-b(x)F'_\lambda(x) = f(x).$

A noteworthy consequence of these methods is a counter example to the so-called "diffusion approximation". We construct a family of Markov processes $x_N(t)$, $1 \le N < \infty$ for which $\lim_{N\to\infty} a_N(x)=1$, $\lim_{N\to\infty} b_N(x) = 0$ and $\lim_{N\to\infty} x_N(t) = x(t)$ (in the sense of weak convergence of stochastic processes) and yet $x(t) \neq$ Brownian motion! The limit process $x(t)$ is what Ito-McKean call a skew Brownian motion.

We conclude these lecture notes with a theorem, 5.1.1, which may be regarded as a semi-group version of Ito's lemma. Several applications are given including some theorems of Burkholder-Gundy, Burkholder, Doob and Lai as special cases. The novelty here is that the drift coefficient has a singularity at one of the boundary points which precludes the use of Ito's lemma.

Because of time the author was unable to include other topics of great importance e.g. the L_2 theory of the Kolmogorov equation and the corresponding eigen-function expansions; applications of the perturbation theory and Trotter-Kato theorem to the equation of neutron transport; the diffusion approximation in genetics, etc. For these topics the reader should consult the items of the supplementary bibliography.

Table of contents

I. Preliminaries.

Let $\{\Omega, \mathcal{F}, P\}$ denote our basic probability space, i.e., $\mathcal{F}$ is a sigma field of subsets of Ω and P is a probability measure with domain $\mathcal{F}$. A stochastic process with index set J is merely a collection of random variables $\{x(t,\omega);\ t \in J\}$. In these lectures J will usually denote the half line $R_+ = [0,\infty)$ or a subinterval thereof e.g., $J = [0,1]$, $J = [\alpha, \beta]$, $0 \leq \alpha < \beta < \infty$.

We denote by $\mathcal{B}\{x(u,\omega);\ u \in J\}$ the smallest sigma field with respect to which all the random variables $x(u,\omega)$, $u \in J$ are measurable. If $J = [0,t]$, then we write $\mathcal{B}(t)$ for $\mathcal{B}\{x(u,\omega):\ u \in J\}$. To simplify the notation we shall often drop the ω and denote a stochastic process by $x(t)$ instead of $x(t,\omega)$. Notation: $X = Y$ a.s. means $P\{\omega:\ X(\omega) \neq Y(\omega)] = 0$ and "a.s." means "almost surely".

Definition 1.1. The stochastic process $x(t)$ is said to be stochastically continuous if

$$\lim_{h \to 0} P\{|x(t+h) - x(t)| \geq \varepsilon\} = 0 \text{ for every } \varepsilon > 0 \text{ and every } t \text{ in } J = [\alpha,\beta]. \tag{1.1}$$

At the end points we require

$$\lim_{h \to 0+} P\{|x(\alpha+h) - x(\alpha)| \geq \varepsilon\} = 0, \quad \lim_{h \to 0^+} P\{|x(\beta-h) - x(\beta)| \geq \varepsilon\} = 0.$$

Stochastic continuity is just a condition on the two dimensional joint distributions of the process $x(t)$ and in practice it is a very easy one to check.

So far we've assumed that $x(t,\omega)$ takes values in $R^1 = (-\infty,\infty)$. By a vector-valued stochastic process is meant a collection of random n-vectors which we also denote by $x(t,\omega) = (x_1(t,\omega),\ldots,x_n(t,\omega))$. More generally it is possible to define random variables with values in normed linear spaces or even in locally convex spaces, but such ideas will play only a small role in these lectures.

Definition 1.2. The stochastic process $\{x(s);\ 0 \leq s \leq t\}$ is said to be a Markov process if

(1.2) $$P\{x(t_n) \leq \lambda | x(t_1),\ldots,x(t_{n-1})\} = P\{x(t_n) \leq \lambda | x(t_{n-1})\} \text{ a.s.}$$

for $0 \leq t_1 < t_2 < \ldots < t_{n-1} < t_n \leq t$.

The following apparently more general definition of the Markov property (1.2) is actually equivalent to it (see Doob [9], p. 83 for the proof).

Definition 1.3. $\{x(s); 0 \leq s \leq t\}$ is a Markov process if for every bounded Borel measurable function f and $s_1 < s_2$ we have

(1.3) $$E\{f(x(s_2))|\mathscr{B}(s_1)\} = E\{f(x(s_2))|x(s_1)\} \text{ a.s.}$$

Notation: By $E_x\{\quad\}$ is meant $E\{\quad|x(0) = x\}$.

We've just defined a Markov process with state space R^1. To define an R^n valued Markov process we just replace the function f in (1.3) by a bounded Borel measurable function $g(x_1,\ldots,x_n)$ and replace x(s) by the random vector $\vec{x}(s) = (x_1(s),\ldots,x_n(s))$.

One of our concerns will be to determine the regularity properties of the sample functions $x(t,\omega)$ of the Markov process. For example, is $x(t,\omega)$ continuous in t with probability one? If not, is $x(t,\omega)$ right continuous with probability one with no discontinuities other than jumps? For the time being we shall content ourselves with the following weak regularity properties:

Definition 1.4. If $x(t,\omega)$: $J \times \Omega \to R$ is measurable with respect to the product sigma field $\mathscr{B}(J) \times \mathscr{F}$ then $x(t,\omega)$ is called a measurable stochastic process; $x(t,\omega)$ is progressively measurable if $x(u,\omega)$: $[0,t] \times \Omega \to R$ is measurable with respect to $\mathscr{B}[0,t] \times \mathscr{F}$.

Definition 1.5. The process $x(t,\omega)$ is called separable if there exists a countable sequence $\mathscr{T} = \{t_j\} \subset J$ and a subset $N \subset \Omega$ with $P(N) = 0$ such that $\omega \notin N$ implies $\{x(t,\omega) \in F \text{ for all } t \in I\} = \{x(t_j,\omega) \in F, \text{ all } t_j \in \mathscr{T} \cap I\}$ for any open subset I of J and any closed subset F of R^1 (or R^n).

Theorem 1.1. (Doob) Every stochastic process $x(t,\omega)$ has an equivalent version $\tilde{x}(t,\omega)$, i.e. $P\{\tilde{x}(t,\omega) \neq x(t,\omega)\} = 0$ for all $t \in J$, which is separable.

Theorem 1.2. (Doob [9], p. 60): If $x(t,\omega)$ is stochastically continuous and separable then $x(t,\omega)$ is a measurable stochastic process. In addition the sample paths $x(\cdot,\omega)$ are with probability one Lebesgue measurable functions of t. Moreover if $E\{x(t,\omega)\}$ exists for $t \in J$ it defines a Lebesgue measurable function of t (with probability one) and if $\int_A E\{|x(t,\omega)|\}dt < \infty$ then almost all sample functions $x(\cdot,\omega)$ are Lebesgue integrable over A.

The importance of the notion of separability is that without it we could not infer the measurability of certain functions of the stochastic process. For example, $\bar{x}(t,\omega) = \limsup_{t'\to t} x(t',\omega)$, $\underline{x}(t,\omega) = \liminf_{t'\to t} x(t',\omega)$, $\bar{x}(t\pm,\omega) = \limsup_{t'\to t\pm} x(t',\omega)$ and $\underline{x}(t\pm,\omega) = \liminf_{t'\to t\pm} x(t',\omega)$ are not in general measurable, since they are obtained as a limit of an uncountable number of random variables. If, however, $x(t,\omega)$ is a separable version then each of these random variables is measurable. A particularly useful application of these ideas is to continuous parameter martingales (or supermartingales or submartingales) which we may assume to be separable. Let us recall the definitions.

Definition 1.5. $\{x(t,\omega);\ t \in J\}$ is a martingale (supermartingale) if $E\{|x(t,\omega)|\} < \infty$, $t \in J$ and $E\{x(t_2)|\mathcal{B}(t_1)\} = x(t_1)$ a.s. $t_1 < t_2$ ($E\{X(t_2)|\mathcal{B}(t_1)\} \geq X(t_1)$ a.s. $t_1 < t_2$). If $-x(t,\omega)$ is a supermartingale then $x(t)$ is a submartingale.

Theorem 1.3. (Doob): Let $x(t,\omega)$ be a separable supermartingale, which is stochastically continuous on $[\alpha \leq t \leq \beta]$. Then there exists an equivalent process $y(t,\omega)$, i.e. $P\{x(t,\omega) \neq y(t,\omega)\} = 0$, such that $y(t,\omega)$ is right continuous, with probability one and $y(t-,\omega)$ exists with probability one.

II. Transition functions, semi-groups and the Kolmogorov differential equations.

Definition 2.1. The function $P(s,x;t,A)$ defined for all $0 \leq s < t$, $x \in [a,b]$, and A a Borel measurable subset of [a,b], is called a transition function if it satisfies the following three conditions:

(2.1)
- (i) $P(s,x;t,A)$ is a probability measure (as a function of the sets A),
- (ii) $x \to P(s,x;t,A)$ is a Borel measurable function of x for each fixed s, t, A and
- (iii) $\int P(t,x;s,dy)P(s,y;\tau,A) = P(t,x;\tau,A)$

where the integration is understood to be over the state space of the Markov process; in this case, over [a,b]. Condition (iii) is called the Chapman-Kolmogorov equation.

Let B[a,b] denote the Banach space of bounded Borel measurable functions with domain [a,b], and C[a,b] denote the subset of bounded continuous functions with domain [a,b]. The norm is denoted by $|f| = \sup_{a \leq x \leq b} |f(x)|$. For $f \in B[a,b]$ the operator $T(t,\tau)f(x) = \int f(y)P(t,x;\tau,dy)$ is well defined and satisfies the conditions listed below:

(2.2)
- (i) If $f(x) \geq 0$ all x, then $T(t,\tau)f(x) \geq 0$ all x.
- (ii) $|T(t,\tau)f| \leq |f|$.
- (iii) If $t < s < \tau$ then $T(t,s)T(s,\tau) = T(t,\tau)$,
- (iv) If $f \in B[a,b]$ then $T(t,\tau)\, f \in B[a,b]$.

Condition (iii) is called the generalized semi-group property. We shall call a semi-group of operators $T(t,\tau)$ satisfying (2.2)(i) - (iv) a positivity-preserving, contraction semi-group, or Markovian semi-group for short.

Theorem 2.1. Given a probability distribution $p(A)$ and transition function $P(t,x;\tau,A)$ there exists a Markov process $X(t)$ with initial distribution $P\{X(0) \in A\} = p(A)$ and transition function $P(t,x;\tau,A)$; more precisely we have $P\{X(\tau) \in A | X(t)\} = P(t,x(t);\tau,A)$ a.s. In particular if p is concentrated at the point x i.e., $p(A) = \delta_x(A)$, then the measure induced on function space by this process will be denoted by $P_x\{\ \}$ and expectations by $E_x\{\ \ \}$.

Proof: For each (n+1) tuple $0 = t_0 < t_1 < \dots < t_n$ define the family of cumulative distributon functions via the formula:

$$(2.3)\quad \begin{cases} F_{t_0,\dots,t_n}(x_0,\dots,x_n) = \int_{-\infty}^{x_n} \cdot\cdot \int_{-\infty}^{x_0} P(t_{n-1},y_{n-1};t_n,dy_n)\,..\,P(t_0,y_0;t_1,dy_1)\,p(dy_0) \\ F_{t_1,\dots,t_n}(x_1,\dots,x_n) = F_{t_0,t_1,\dots,t_n}(+\infty,x_1,\dots,x_n). \end{cases}$$

Remarks: We integrate with respect to dy_n first, then dy_{n-1}, etc. It suffices to show that the collection of distribution functions $F_{t_0,t_1,\dots,t_n}$ forms a consistentent family i.e.

$$(2.4)\qquad \lim_{x_k \uparrow +\infty} F_{t_0,..,t_n}(x_0,..,x_k,..,x_n) = F_{t_0,..,t_{k-1},..,t_{k+1},\ t_n}(x_0,..,x_{k-1},x_{k+1},..,x_n).$$

Set $f(y_{k+1}) = \int_{-\infty}^{x_n} \int_{-\infty}^{x_{k+2}} P(t_{n-1},y_{n-1},t_n,dy_n)\,..\,P(t_{k+1},y_{k+1};t_{k+2},dy_{k+2})$ and $g(y) = I_{(-\infty,x_{k+1}]}(y)\ f(y)$; I_A denotes the indicator function of the set A. The integral on the right hand side of (2.3) is evaluated by first computing

$$\int_{-\infty}^{\infty} P(t_{k-1},y_{k-1};t_k,dy_k) \int_{-\infty}^{\infty} g(y_{k+1})P(t_k,y_k;t_{k+1},dy_{k+1}) =$$

$$= T(t_{k-1},t_k)T(t_k,t_{k+1})g(y_{k-1})$$

$$= T(t_{k-1},t_{k+1})g(y_{k-1}).$$

$$= \int_{-\infty}^{x_{k+1}} f(y_{k+1})P(t_{k-1},y_{k-1};t_{k+1},dy_{k+1}).$$

Now continue the integration with respect to the remaining variables y_{k-1}, $y_{k-2},\ldots,y_0$. The result is clearly $F_{t_0,\ldots,t_{k-1},t_{k+1},\ldots,t_n}(x_0,\ldots,x_{k-1},x_{k+1},\ldots,x_n)$ as defined at (2.3). Hence by the Kolmogorov existence theorem (cf. Billingsly [2] pp. 228-230) there exists a stochastic process $x(t,\omega)$ such that

(2.5) $$P\{x(t_0) \in B_0,\ldots,x(t_n) \in B_n\} = \int_{B_n}\ldots\int_{B_0} P(t_{n-1},y_{n-1},t_n,dy_n)\ldots P(t_0,y_0;t_1,dy_1)P(dy_0)$$

All we have to do now is verify the Markov property i.e., it suffices to show

(2.6) $$P\,\{x(t_n) \le \lambda | x(t_0),\ldots,x(t_{n-1})\} = P(t_{n-1},\ x(t_{n-1});t_n,A) \text{ a.s.}$$

where $A = (-\infty,\lambda]$. Set $B = \{\omega:\ x(t_0) \in B_0,\ldots,x(t_{n-1}) \in B_{n-1}\}$. In the course of establishing (2.6) the following calculation is carried out:

(2.7) $$E\{I_B f(x(t_{n-1}))\} = \int_{B_{n-1}}\ldots\int_{B_0} f(y_{n-1})P(t_{n-1},y_{n-2};t_{n-1},dy_{n-1})P(t_0y_0;t_1dy_1)P(dy_0)$$

Now it is enough to prove (2.7) for $f(y) = I_F(y)$, in which case the left hand side is merely

$$P\{x(t_0) \in B_0,\ldots,x(t_{n-2}) \in B_{n-2},x(t_{n-1}) \in B_{n-1} \cap F\} =$$

$$\int_{\{B_{n-1}\cap F\}} \int_{B_{n-2}}\cdots\int_{B_0} P(t_{n-2},y_{n-2};t_{n-1},dy_{n-1})\cdots P(t_0,y_0,t_1,dy_1)P(dy_0) =$$

$$= \int_{B_{n-1}}\cdots\int_{B_0} I_F(y_{n-1})P(t_{n-2},y_{n-2};t_{n-1},dy_{n-1})\cdots P(t_0,y_0,t_1,dy_1)P(dy_0).$$

Let us return to (2.6), the left hand side of which is $E\{I_A(x(t_n))|x(t_0),\ldots,x(t_{n-1}\}$. We must show

(2.8) $$\int_B I_A(x(t_n))dP = \int_B P(t_{n-1},x(t_{n-1});t_n,A)dP.$$

Set $f(y) = P(t_{n-1},y;t_n,A)$ and apply (2.7) to the right hand side of (2.8) which becomes

(2.9) $$E\{I_B f(x(t_{n-1}))\} = \int_{B_{n-1}}\cdots\int_{B_0} P(t_{n-1},y_{n-1};t_n,A)\cdots P(t_0,y_0;t_1,dy_1)P(dy_0)$$

$$= P\{x(t_0)\in B_0,\ldots x(t_{n-1})\in B_{n-1},\ x(t_n)\in A\}.$$

$$= \int_B I_A(x(t_n))dP. \quad \text{q.e.d.}$$

EXAMPLE: For Brownian motion,

$$P(t,x;\tau,A) = \int_A [2\pi(\tau-t)]^{-1/2}\exp[-(y-x)^2/2(\tau-t)]dy$$

Remark: In practice an explicit formula for the transition function is not usually known; instead one first constructs by various methods, e.g. the Hille-Yosida theorem, a Markovian semi-group $T(t,\tau)$: $B[a,b] \to B[a,b]$. Given the semi-groups one can easily construct the transition function. More precisely we have

Theorem 2.2. To every Markovian semigroup $T(t,\tau)$ satisfying conditions (i)-(iv) of (2.2) there exists a transition function (and hence Markov process) such that $T(t,\tau)f(x) = \int f(y)P(t,x;\tau,dy)$.

Proof: Let $f \in C[a,b]$. If $[a,b]$ is not compact assume in addition that f has compact support. The mapping $f \to T(t,\tau)f(x)$ is a positive linear functional on $C[a,b]$ and hence by the Riesz Representation theorem (W. Rudin [27] p. 40) there exists a probability measure $P(t,x;\tau,A)$ such that $T(t,\tau)f(x) = \int f(y)P(t,x;\tau,dy)$. Since $P(t,x;\tau,A) = T(t,\tau)I_A(x) \in B[a,b]$, because $I_A \in B[a,b]$, we see at once that $x \to P(t,x;\tau,A)$ is Borel measurable in x. The Chapman-Kolmogorov equation is now a consequence of the semi-group property $T(t,s)T(s,\tau) = T(t,\tau)$. Thus $P(t,x;\tau,A)$ is a transition function in the sense of definition 2.1.

To proceed further and develop an interesting theory, additional conditions on the transition functions must be imposed. In particular, we assume the existence of functions $a(t,x)$ and $b(t,x)$ continuous on $R_+ \times [a,b]$ and satisfying the conditions:

For every $\delta > 0$

$$(2.10)\quad \begin{cases} \text{(i)} & \displaystyle\int_{|y-x|\geq\delta} P(t,x,t+h,dy) = o(h) \text{ as } h \to 0 \\ \text{(ii)} & \displaystyle\int_{|y-x|\leq\delta} (y-x)P(t,x,t+h,dy) = b(t,x)h + o(h) \text{ as } h \to 0 \\ \text{(iii)} & \displaystyle\int_{|y-x|\leq\delta} (y-x)^2P(t,x,t+h,dy) = a(t,x)h + o(h) \text{ as } h \to 0. \end{cases}$$

Remark: $a(t,x)$ is called the diffusion coefficient and $b(t,x)$ is the drift coefficient.

Theorem 2.3. Let $f \in C[a,b]$ and $u(t,x) = T(t,\tau)f(x) \in C^2[a,b]$ for all $0 \leq t \leq \tau$. If the transition function $P(t,x;\tau,A)$ satisfies conditions (i)-(iii) of (2.10) then

$u(t,x)$ satisfies the parabolic partial differential equation (P.D.E.).

(2.11)
$$\begin{cases} -u_t(t,x) = (a(t,x)/2)u_{xx}(t,x) + b(t,x)u_x(t,x) \\ \lim_{t\uparrow\tau} u(t,x) = u(\tau,x) = f(x) \end{cases}$$

Remark: The parabolic P.D.E. (2.11) is the Kolmogorov backward differential equation.

Lemma: Let $f \in C^2[a,b]$, Then

(2.2)
$$\lim_{h\to 0+} \frac{T(t,t+h)f(x) - f(x)}{h} = (a(t,x)/2)f''(x) + b(t,x)f'(x).$$

Proof: $T(t,t+h)f(x) - f(x) = \int(f(y) - f(x))P(t,x,t+h,dy)$

$$= \int_{|y-x|\leq\delta} (f(y) - f(x))P(t,x,t+h,dy) + 2|f|o(h).$$

Because $\left|\int_{|y-x|>\delta} (f(y) - f(x))P(t,x,t+h,dy)\right| \leq 2|f| \int_{|y-x|>\delta} P(t,x,t+h,dy)$

$$= 2|f|o(h) \text{ by } (2.10)(i).$$

Now expand f in a Taylor series:

$$f(y) - f(x) = f'(x)(y-x) + (1/2)f''(x)(y-x)^2 + R(x,y,\delta),$$

$$R(x,y,\delta) = (1/2)(f''(\xi) - f''(x))(y-x)^2$$

and ξ is a point between x and y. Given $\varepsilon > 0$ we can find a $\delta > 0$ small enough so that $|f''(\xi) - f''(x)| < 2\varepsilon$ for $|y-x| < \delta$ uniformly in x.

Now

$$\left|\int_{|y-x|\le\delta} R(x,y,\delta)P(t,x,t+h,dy)\right| \le \varepsilon \int_{|y-x|\le\delta} (y-x)^2 P(t,x,t+h,dy) = \varepsilon[ha(t,x) + o(h)]$$

We thus arrive at the expansion

$$T(t,t+h)f(x) - f(x) = (a(t,x)/2)f''(x) + b(t,x)f'(x)$$

$$+ \varepsilon 0(h) + o(h)(|f'| + |f''|).$$

Put $G(t)f(x) = (a(t,x)/2)f''(x) + b(t,x)f'(x)$. Then

$$\left|\frac{T(t,t+h)f(x) - f(x)}{h} - G(t)f(x)\right| \le \varepsilon \frac{0(h)}{h} + \frac{o(h)}{h}(|f'| + |f''|).$$

Now let $h \to 0$ and the proof of the lemma is complete. The same reasoning shows that

(2.13) $$\lim_{h\to 0+} \frac{T(t-h,t)f(x) - f(x)}{h} = G(t)f(x).$$

We return to the proof of Theorem 2.3. From the semi-group property we see that

$$u(t-h,x) - u(t,x) = T(t-h,t)T(t,\tau)f(x) - T(t,\tau)f(x)$$

$$= T(t-h,t)u(t,x) - u(t,x).$$

Apply (2.13) to $u(t,x)$ (as a function of x) and get

$$\lim_{h\to 0+} \frac{u(t-h,x) - u(t,x)}{h} = \lim_{h\to 0+} \frac{T(t-h,t)u(t,x) - u(t,x)}{h} = G(t)u(t,x),$$

or

$$-u_t(t,x) = G(t)u(t,x).$$

Finally

$$u(\tau-h,x) - f(x) = \int_{|y-x|\leq\delta} (f(y)-f(x))P(\tau-h,x,\tau,dy) + o(h).$$

But $f \in C[a,b]$ implies $|f(y) - f(x)| < \varepsilon$ for $|y-x|\leq\delta$. Thus $|u(\tau-h,x) - f(x)| \leq \varepsilon + o(h)$. Clearly this implies $\lim_{h\to 0} u(\tau-h,x) = \lim_{t\uparrow\tau} u(t,x) = f(x)$. This completes the proof of Theorem 2.3.

Definition: The Markov process is said to be homogeneous in time if its transition function satisfies the condition

$$P(0,x;\tau-t,A) = P(t,x;\tau,A). \tag{2.14}$$

In particular this implies $P(t,x;t+h,A) = P(0,x;h,A)$. Put $P(t,x,A) = P(0,x;t,A)$ In this case the conditions (2.10)(i)-(iii) become

$$(2.15)\quad \begin{cases} \text{(i)} & \int_{|y-x|>\delta} P(h,x,dy) = o(h) \\ \text{(ii)} & \int_{|y-x|\leq\delta} (y-x)P(h,x,dy) = b(x)h + o(h) \\ \text{(iii)} & \int_{|y-x|\leq\delta} (y-x)^2 P(h,x,dy) = a(x)h + o(h) \\ & \qquad\qquad \text{as } h\to 0+. \end{cases}$$

In the time homogeneous case the Kolmogorov backward differential equation reduces to

$$(2.16)\quad \begin{cases} u_t(t,x) = (a(x)/2)u_{xx}(t,x) + b(x)u_x(t,x) \\ \lim_{t\downarrow 0} u(t,x) = u(0,x) = f(x) \text{ where} \\ u(t,x) = T(0,t)f(x) = \int f(y)P(t,x,dy) \end{cases}$$

Notation: We write $T(t)$ instead of $T(0,t)$ and put $Gf(x) = (a(x)/2)f''(x) + b(x)f'(x)$. $T(0) = I$, the identity operator.

The operators $T(t)$ form a semi-group acting on the Banach space $B[a,b]$; more precisely we have:

$$(2.17)\quad \begin{cases} T(t):\ B[a,b] \to B[a,b] \\ |T(t)f| \le |f| \\ \text{If } f \ge 0 \text{ then } T(t)f \ge 0 \\ T(t)T(s) = T(t+s), \quad t \ge 0,\ s \ge 0 \end{cases}$$

Let us look at the Kolmogorov differential equation (2.16) from the point of view of semi-group theory, which we shall study in detail in Chapter 3. In this approach a crucial role is played by the operator G and its domain $\mathcal{D}(G)$ defined by

$$(2.18)\qquad \mathcal{D}(G) = \{f:\ \lim_{h\to 0^+} \left|\frac{T(h)f - f}{h} - Gf\right| = 0\};$$ G is

called the infinitesimal generator of the semi-group $T(t)$. Suppose now $u(t,x) = T(t)f(x)$ and $f \in \mathcal{D}(G)$. Then

$$\lim_{h\to 0} \frac{u(t+h,x) - u(t,x)}{h} = \lim_{h\to 0} \frac{T(t+h)f(x) - T(t)f(x)}{h}$$

$$= \lim_{h\to 0} T(t)\left(\frac{T(h)f(x) - f(x)}{h}\right) = T(t)Gf(x)$$

$$= \lim_{h\to 0} \frac{(T(h) - I)T(t)f(x)}{h} = GT(t)f(x).$$

Formally then $u(t) = T(t)f$ satisfies the differential equation

(2.19) $$\begin{cases} u'(t) = Gu(t) \\ u(0) = f \end{cases}$$

Throwing caution to the winds we can "solve" (2.19) via the formula

(2.20) $$u(t) = \exp(tG)f = T(t)f.$$

Under suitable hypotheses we've shown that Markov processes, homogeneous in time, give rise to positivity-preserving contraction semi-groups $T(t)$ whose infinitesimal generators G are second order linear differential operators of the form

(2.21) $$Gf(x) = (a(x)/2)f''(x) + b(x)f'(x), \quad a(x) \geq 0.$$

In real life, one is usually given only the diffusion and drift terms and the question then becomes "Does there exist a Markovian semi-group $T(t): C[a,b] \to C[a,b]$, say, whose infinitesimal generator is G?" An affirmative answer to this question implies, by Theorem 2.2, the existence of a Markov process whose Kolmogorov differential equation is (2.16). In the next chapter we shall give necessary and sufficient conditions for G to generate a Markovian semi-group $T(t) = \exp(tG)$.

III. Semi-group theory

1. Let X denote a separable Banach space the elements of which are denoted by f, g,.. and norm $|f|$. Examples: $B[a,b]$, $C[a,b]$ with $|f| = \sup_{a \le x \le b} |f(x)|$; $C^k[a,b]$ with $|f| = \sum_{\ell=0}^{k} |f^{(\ell)}|$, here $f^{(\ell)}$ denotes the ℓ^{th} derivative; $L_p[a,b]$ with $|f| = \left(\int_a^b |f(x)|^p dx\right)^{1/p}$, $1 \le p \le \infty$.

We denote by X^* the class of continuous linear functionals $f^*: X \to R$. This means (i) $f^*(\alpha f + \beta g) = \alpha f^*(f) + \beta f^*(g)$ for every $f, g \in X$ and $\alpha, \beta \in R$ and (ii)

$$\lim_{n\to\infty} f^*(f_n) = f^*(f) \text{ if } \lim_{n\to\infty} |f_n - f| = 0.$$

Sometimes it is convenient, depending on the context, to set $\langle f^*, f\rangle = f^*(f)$.

Definition 3.1.1. A family of bounded linear operators $T(t): X \to X$ is called a semi-group if $T(t+s) = T(t)T(s)$, $s, t \ge 0$, $T(0) = I$. $T(t)$ is called strongly continuous if for every $f \in X$

$$\lim_{h\to 0} |T(h)f - f| = 0. \tag{3.1.1}$$

$T(t)f$ will be called weakly continuous, weakly right continuous or weakly measurable if for every $f^* \in X^*$ the corresponding real valued function $\langle f^*, T(t)f\rangle$ is continuous, right continuous or measurable in the ordinary sense. More generally if $u(t): [\alpha,\beta] \to X$ we shall say that $u(t)$ is strongly continuous at the point t if $\lim_{h\to 0} |u(t+h) - u(t)| = 0$ and we write $s - \lim_{h\to 0} u(t+h) = u(t)$. Similarly if $\langle f^*, u(t)\rangle$ is continuous, right continuous or measurable all $f^* \in X^*$ we shall say that $u(t)$ is weakly continuous, weakly right continuous or weakly measurable.

From the Hahn-Banach theorem we know that

$$(3.1.2)\quad \begin{cases} \text{(i)} & \langle f^*,f\rangle = \langle f^*,g\rangle \text{ for all } f^* \in X^* \text{ implies } f = g; \\ \text{(ii)} & \text{given any } f \in X \text{ there exists } f^* \in X^* \text{ such that } \langle f^*,f\rangle = \\ & |f| \text{ and } \langle f^*,g\rangle \le |g| \text{ , all } g \in X. \end{cases}$$

Let J denote a closed interval $[\alpha,\beta]$ and $u(t)$: $[\alpha,\beta] \to X$.

Definition 3.1.2. We say that $u(t)$ is strongly differentiable at t if

$$\operatorname*{s-lim}_{h\to 0}\ (u(t+h) - u(t)/h) \quad \text{exists;}$$

the limit is then denoted by $u'(t)$. If $u'(t)$ exists for all $t \in [\alpha,\beta]$ we say that $u(t)$ is strongly differentiable on $[\alpha,\beta]$.

Definition 3.1.3. We say that $u(t)$ is strongly integrable on $[\alpha,\beta]$ if

$$\operatorname*{s-lim}_{h\to 0} \sum_{k=0}^{n} u(t_k)(t_k - t_{k-1}) \qquad \text{exists where}$$

$$h = \operatorname*{Max}_{1\le k\le n} |t_k - t_{k-1}|,\ \alpha = t_0 < t_1 < ,\ldots, < t_n = \beta.$$

The limit is then denoted by $\int_\alpha^\beta u(t)dt$. If $\operatorname*{s-lim}_{\beta\to\infty} \int_\alpha^\beta u(t)dt$ exists then we write $\int_\alpha^\infty u(t)dt$ as the limit.

Properties of the integral $\int_\alpha^\beta u(t)dt$:

(3.1.3)

(i) If $u(t)$ is strongly continuous on $[\alpha,\beta]$ then $\int_\alpha^\beta u(t)dt$ exists and $|\int_\alpha^\beta u(t)dt| \leq \int_\alpha^\beta |u(t)|dt$.

(ii) Suppose A is a bounded linear mapping from X into the Banach space $\hat{X}$. If $u(t)$ is strongly differentiable then so is $Au(t)$ and $(Au(t))' = Au'(t)$.

(iii) If $u(t)$ is strongly integrable then so is $Au(t)$ and $A(\int_\alpha^\beta u(t)dt) = \int_\alpha^\beta Au(t)dt$.

(iv) If $u(t)$ is strongly integrable on $[\alpha,\alpha+h]$ and s-continuous from the right at α then

$$\underset{h\to 0}{\text{s-lim}} \ \frac{1}{h} \int_\alpha^{\alpha+h} u(t)dt = u(\alpha).$$

(v) If $u'(t)$ is strongly continuous then

$$\int_\alpha^\beta u'(t)dt = u(\beta) - u(\alpha)$$

(vi) If $u(t)$ is strongly integrable on $[\alpha,\beta]$ then $u(t-h)$ is strongly integrable on $[\alpha+h,\beta+h]$ and

$$\int_{\alpha+h}^{\beta+h} u(t-h)dt = \int_\alpha^\beta u(t)dt.$$

The proofs are routine. Consider for example (v)

$$\langle f^*, \int_\alpha^\beta u'(t)dt\rangle = \int_\alpha^\beta \langle f^*,u'(t)\rangle = \int_\alpha^\beta \langle f^*,u(t)\rangle' dt$$

$$= \langle f^*,u(\beta)\rangle - \langle f^*,u(\alpha)\rangle$$

$$= \langle f^*,u(\beta) - u(\alpha)\rangle.$$

An application of 3.1.2(i) completes the proof.

Set $a(t) = \log|T(t)|$ and observe that $a(t_1+t_2) \leq a(t_1) + a(t_2)$. Assume $T(t)$ is strongly continuous for $0 \leq t < \infty$. Since $|T(t)f|$ is continuous in t it follows that $|T(t)| = \sup_{|f|\leq 1} |T(t)f|$ is lower semicontinuous and hence measurable. So $a(t)$ is a measurable subadditive function, (cf Hille-Phillips [15]). Hence

$$\lim_{t\to\infty} t^{-1} \log|T(t)| = \Gamma \text{ exists}, \quad -\infty \leq \Gamma < \infty \tag{3.1.4}$$

$$\text{where } \Gamma = \inf_{t>0} t^{-1}\log|T(t)|.$$

Suppose $|\Gamma| < \infty$.

We have then $\limsup_{t\to\infty} t^{-1}\log |T(t)f| \leq \Gamma$. Hence for all $t \geq 1$, say, we can find a constant $M(f,\varepsilon)$ such that $\log|T(t)f|-(\Gamma+\varepsilon)t \leq M(f,\varepsilon)$. Thus

$$\exp(-t(\Gamma+\varepsilon))|T(t)f| \leq \exp(M(f,\varepsilon)).$$

By the uniform boundedness principle there exists a constant $M(\varepsilon)$, independent of f, such that

$$\exp(-t(\Gamma+\varepsilon))(T(t)f \leq M(\varepsilon) \text{ or}$$

$$|T(t)f| \leq M(\varepsilon)\exp(t(\Gamma+\varepsilon), \ t \geq 1.$$

On the other hand if $T(t)$ is strongly continuous on $0 \leq t \leq 1$ we must have $|T(t)| \leq M$, say. It suffices to show that $|T(t)f|$ is bounded on $[0,1]$ for all $f \in \mathbf{X}$. But $|T(t)f|$ is continuous in t, so it is obviously bounded. A similar argument works if $\Gamma = -\infty$. Summing up then we have

Theorem 3.1.1. If $T(t) : X \to X$ is a strongly continuous semigroup then there exist constants M and γ such that

$$\exp(-t\gamma)|T(t)| \leq M, \ 0 \leq t < \infty.$$

Remark: By considering the semi-group $S(t) = \exp(-t\gamma)T(t)$ we can therefore assume the semi group is bounded i.e. $|S(t)| \leq M$.

Definition 3.1.4: $X_0 = \{f: T(t)f \text{ is strongly continuous}\}$.

It follows at once from the semi-group property that $T(t)X_0 \subset X_0$ and moreover X_0 is a closed linear subspace of X. In general it is not an easy matter to find a nice subspace X_0 on which T(t) is strongly continuous. The following theorem is useful in questions of this sort (cf. Dynkin [11]).

Theorem 3.1.2: Suppose $T(t)$: $X \to X$ is weakly right continuous. Then T(t) is strongly continuous.

As an application consider the case where $X = C[a,b]$ and T(t) is a Markovian semi-group with transition function $P(t,x,A)$. Let $P(0,x,A) = \delta(x,A)$ = unit mass concentrated at x. Assume

$$(3.1.5)\quad \begin{cases} \text{(i)} & \lim\limits_{x\to x_0} P(t,x,A) = P(t,x_0,A) \text{ in the sense of weak convergence of measures and} \\ \text{(ii)} & \lim\limits_{t\downarrow 0} P(t,x,U) = 1 \text{ for every neighborhood } U \text{ of } x. \end{cases}$$

Lemma 3.1.1. If the transition function of a Markov process satisfies (3.1.5) (i) and (ii) then $T(t)$: $C[a,b] \to C[a,b]$ and T(t) is weakly right continuous and therefore T(t) is strongly continuous. (If [a,b] is noncompact then T(t) is strongly continuous on $C_0[a,b]$)

Proof: If $f \in C[a,b]$ and $\text{w-}\lim\limits_{x\to x_0} P(t,x,\cdot) = P(t,x_0,\cdot)$ then $\lim\limits_{x\to x_0} T(t)f(x) = T(t)f(x_0)$ i.e. $T(t)f \in C[a,b]$. Pick $f^* \in X^*$ which means f^* is a signed measure. Now $\lim\limits_{t\downarrow 0} \langle f^*,T(t)f\rangle = \langle f^*,f\rangle$ because (3.1.5) (ii) implies $\lim\limits_{t\downarrow 0} T(t)f(x) = f(x)$ pointwise and boundedly, since $|T(t)f| \leq |f|$. The proof is now completed by applying the Lebesgue Dominated Convergence Theorem. Actually we've only shown that T(t) is weakly right continuous at the origin, but the semi-group property immediately implies that T(t)f is weakly right continuous for all t.

Definition 3.1.5. The infinitesimal generator G of the semi-group T(t) is defined by the formula

$$Gf = \underset{h \to 0}{\text{s-lim}} \frac{T(h)f-f}{h}, \text{ provided the limit exists}$$

$$\mathscr{D}(G) = \{f: \ Gf = \underset{h \to 0}{\text{s-lim}} \frac{T(h)f-f}{h} \text{ exists}\}.$$ $\mathscr{D}(G)$ is called the domain of G.

Remarks: Clearly $\mathscr{D}(G) \subset X_0$ and $T(t)\mathscr{D}(G) \subset \mathscr{D}(G)$. In general G is an unbounded linear operator and formally at least we can recover the semi-group from its infinitesimal generator G via the formula

(3.1.6) $\quad T(t) = \exp(tG)$.

Since, as we've already remarked, G is an unbounded linear operator we cannot just put $\exp(tG) = \sum_{n=0}^{\infty} t^n G^n/n!$. What is true however is Hille's beautiful exponential formula

(3.1.7) $\quad T(t) = \exp(tG) = \underset{n \to \infty}{\text{s-lim}} \left(I - \frac{tG}{n}\right)^{-n}$ where $\left(I - \frac{tG}{n}\right)^{-1}$ denotes the inverse operator.

The standard method of recovering T(t) from its infinitesimal generator G is via the "Yosida approximation" which we present in III.2 . The representation (3.1.7) will be derived in III.4.

2. The Hille-Yosida Theorem

From now on we assume T(t) is a strongly continuous contraction semi-group on the Banach space X with infinitesimal generator G and domain $\mathscr{D}(G)$. Our plan is to characterize those linear operators G which generate contraction semi-groups. As a first step we derive necessary conditions that must be satisfied by G and $\mathscr{D}(G)$. We then show that these conditions are in fact sufficient - this is the Hille-Yosida theorem which, together with the Trotter-Kato Theorem, will play an important role in the existence theorems and limit theorems for Markov processes of Chapters IV and VI.

Lemma 3.2.1. $\mathcal{D}(G)$ is a dense subspace of X, i.e. the strong closure of $\mathcal{D}(G)$ is X.

Proof: By hypothesis T(t)f is strongly continuous in t for each $f \in X$. Thus $g(t) = \int_0^t T(s)f\,ds$ is well defined. We claim for each $t > 0$, $g(t) \in \mathcal{D}(G)$. Setting aside for a moment the proof of this fact we note that according to

(3.1.3) (iv) $\text{s-lim}_{h\to 0} \frac{g(h)}{h} = f$. But $\frac{g(h)}{h} \in \mathcal{D}(G)$ and since f was an arbitrary element of X this shows $\overline{\mathcal{D}(G)} = X$. Let us show that $g(t) \in \mathcal{D}(G)$.

$T(h)g(t) = \int_0^t T(s+h)f\,ds = \int_h^{t+h} T(s)f\,ds$. Hence

$$T(h)g(t)-g(t) = \int_h^{t+h} T(s)f\,ds - \int_0^h T(s)f\,ds = \int_0^{t+h} T(s)f\,ds - \int_0^h T(s)f\,ds$$

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

$T(h)g(t)-g(t) = \int_t^{t+h} T(s)f\,ds - \int_0^h T(s)f\,ds$. Thus

$\frac{T(h)g(t)-g(t)}{h} = \frac{1}{h}\int_t^{t+h} T(s)f\,ds - \frac{1}{h}\int_0^h T(s)f\,ds$. Let $h\to 0$ and involving (3.1.3)(iv) once again we deduce $\text{s-lim}_{h\to 0} \frac{T(h)g(t)-g(t)}{h} = T(t)f-f$. Hence $g(t) \in \mathcal{D}(G)$.

Lemma 3.2.2. If $f \in \mathcal{D}(G)$ then

(3.2.1) $T(t)f = f + \int_0^t T(s)Gf\,ds$. Conversely if $T(t)f = f + \int_0^t T(s)g\,ds$ then $f \in \mathcal{D}(G)$ and $Gf = g$.

Proof: If $U(t) = T(t)f$, $f \in \mathcal{D}(G)$ then U(t) is obviously s-differentiable with $U'(t) = T(t)Gf$. From 3.1.3(v) we obtain $U(t)-U(0) = \int_0^t U'(s)ds$ or equivalently $T(t)f-f = \int_0^t T(s)Gf\,ds$.

Conversely, suppose there exists $g \in X$ such that $T(t)f = f + \int_0^t T(s)g\,ds$. Applying 3.1.3(iv) yields $\text{s-lim}(T(t)f-f]/t = g$ so $f \in \mathcal{D}(G)$ and $Gf=g$.

Corollary: G is a closed linear operator.

Proof: Assume $\text{s-lim}_{n\to\infty} f_n = f$, $\text{s-lim}_{n\to\infty} G f_n = g$. Then $\text{s-lim}_{n\to\infty} T(t)f_n - f_n = \text{s-lim}_{n\to\infty} \int_0^t T(s)Gf_n\, ds$

As $n\to\infty$ the left hand side tends to $T(t)t-f$ and the right hand side tends to $\int_0^t T(s)g ds$.

Thus $T(t)f-f = \int_0^t T(s)g ds$. Thus $f \in \mathscr{D}(G)$ and $Gf=g$. For each $\lambda > 0$, $|\exp(-\lambda t)T(t)f| \le \exp(-\lambda t)$ and hence $\text{s-lim}_{\beta\to\infty} \int_0^\beta \exp(-\lambda t)T(t)f\, dt = \int_0^\infty \exp(-\lambda t)T(t)f dt$ exists.

Definition 3.2.1. The resolvent operator $R(\lambda)$ is defined via the formula $R(\lambda)f = \int_0^\infty \exp(-\lambda t)T(t)f dt$, $\lambda > 0$. The one parameter family of operators $R(\lambda)$, $\lambda > 0$ satisfy the following conditions:

(i) $|\lambda R(\lambda)f| \le |f|$ i.e. $\lambda R(\lambda)$ is a contraction

(ii) $g = R(\lambda)f$ is the unique solution to $(\lambda I-G)g=f$ satisfying the condition $g \in \mathscr{D}(G)$; we write

$$g = R(\lambda)f = (\lambda I-G)^{-1}f = (\lambda-G)^{-1}f.$$

(iii) $\text{s-lim}_{\lambda\to\infty} |\lambda R(\lambda)f-f| = 0$

(iv) $(\lambda-\mu)R(\lambda)R(\mu) = R(\mu)-R(\lambda)$, the resolvent equation.

Remark: In the literature the equation $\lambda g-Gg = f$ is sometimes referred to as the "stationary equation".

Proof: (i) $|R(\lambda)f| \le \int_0^\infty e^{-\lambda t}|T(t)f|dt \le |f| \int_0^\infty e^{-\lambda t}dt = \lambda^{-1}|f|$. Thus $|\lambda R(\lambda)| \le 1$.

(ii) $$T(h)R(\lambda)f = \int_0^\infty e^{-\lambda t}\, T(t+h)f dt = \int_h^\infty e^{-\lambda(t-h)}T(t)f dt$$

$$= e^{\lambda h}\int_h^\infty e^{-\lambda t}T(t)f dt$$

$$= e^{\lambda h}[R(\lambda)f - \int_0^h e^{-\lambda t}T(t)f dt]$$

We note in passing that $T(h)R(\lambda) = R(\lambda)T(h)$. Now $\frac{T(h)R(\lambda)f-R(\lambda)f}{h} = (\frac{e^{\lambda h}-1}{h}) R(\lambda)f - \frac{e^{\lambda h}}{h}\int_0^h e^{-\lambda t}\, T(t)f dt$.

Letting $h \to 0$ we get

$$GR(\lambda)f = \underset{h\to 0}{\text{s-lim}}\, h^{-1}(T(h)R(\lambda)f - R(\lambda)f) = \lambda R(\lambda)f - f$$

where we've used (3.1.3)(iv). Thus $(\lambda-G)R(\lambda)f=f$ as claimed. The same argument shows that if $f \in \mathscr{D}(G)$ then

$$GR(\lambda)f = R(\lambda)Gf.$$

(iii) Suppose $f \in \mathscr{D}(G)$. Then $\lambda R(\lambda)f-f = GR(\lambda)f = R(\lambda)Gf$. Hence $|\lambda R(\lambda)f-f| = |R(\lambda)Gf| \leq \lambda^{-1}|Gf|$. Thus $\lim_{\lambda\to\infty} |\lambda R(\lambda)f-f| \leq \lim_{\lambda\to\infty} \lambda^{-1}|Gf|=0$. Since $|\lambda R(\lambda)| \leq 1$ and $\mathscr{D}(G)$ is dense in X we conclude $\lim_{\lambda\to\infty} |\lambda R(\lambda)f-f|=0$ all $f \in X$.

(iv) Set $F_\lambda = R(\lambda)f$ and $F_\mu = R(\mu)f$.

Then
$$\begin{aligned}\lambda(F_\lambda - F_\mu)-G(F_\lambda - F_\mu) &= (\lambda F_\lambda - GF_\lambda) - \lambda F_\mu + GF_\mu \\ &= f - \lambda F_\mu + \mu f_\mu - f \\ &= (\mu-\lambda)F_\mu .\end{aligned}$$

But $F_\lambda - F_\mu \in \mathscr{D}(G)$ and so $F_\lambda - F_\mu$ is the unique solution to $\lambda g - Gg = (\mu-\lambda)F$ .i.e.

$$\begin{aligned}F_\lambda - F_\mu &= (\lambda-G)^{-1}(\mu-\lambda)F_\mu = (\mu-\lambda)R(\lambda)F \\ &= (\mu-\lambda)R(\lambda)R(\mu)f.\end{aligned}$$

But $F_\lambda - F_\mu = R(\lambda)f-R(\mu)f = (\mu-\lambda)R(\lambda)R(\mu)f$. This completes the proof of (iv).

Remarks: (i) It follows from the estimate $|R(\lambda)t| \leq \lambda^{-1}|f|$ that $|(\lambda-G)g| \geq \lambda|(\lambda-G)^{-1}(\lambda-G)g| \geq \lambda|g|$ if $g \in \mathscr{D}(G)$. Hence $|g| \neq 0$ and $g \in \mathscr{D}(G)$ implies $|(\lambda-G)g| \neq 0$. So $(\lambda-G)$ maps $\mathscr{D}(G)$ in a one-one fashion into X. In fact the map is onto because $f = (\lambda-G)R(\lambda)f$. Similarly $R(\lambda)$ is a one-one map of X onto $\mathscr{D}(G)$. Hence the range of $R(\lambda)$ is independent of λ and equals $\mathscr{D}(G)$.

(ii) It follows from (iv) that $R(\lambda)R(\mu) = R(\mu)R(\lambda)$.

We are now ready to state and prove the basic.

Theorem 3.2.1 (Hille-Yosida): The necessary and sufficient conditions for the linear operator G with domain $\mathscr{D}(G)$ to be the infinitesimal generator of a strongly continuous contraction semi-group $T(t)$ are:

(i) $\mathscr{D}(G)$ is dense in X

(ii) G is a closed operator

(iii) For every $f \in X$ there exists a unique solution $F_\lambda \in \mathscr{D}(G)$ of the equation $(\lambda - G)F_\lambda = f$ such that

(iv) $\lambda|F_\lambda| \leq |f|$, i.e. $(\lambda-G)^{-1}$ is a bounded linear operater of norm $\leq \lambda^{-1}$ and range $\mathscr{D}(G)$.

Remarks: (i) The necessity of conditions (iii) and (iv) has already been established when we derived the various properties of the resolvent family $R(\lambda)$ -see the discussion following Definition 3.2.1. The necessity of conditions (i) and (ii) follow from lemmas 3.2.1, 3.2.2.

(ii) The hypotheses of the theorem are not logically independent of one another. That G is a closed operator follows from (i), (iii) and (iv). To see this, let $F_n \in \mathscr{D}(G)$ denote a sequence such that $\text{s-}\lim_{n\to\infty} F_n = F$ and $\text{s-}\lim_{n\to\infty} GF_n = g$. Then, as we now show $F \in \mathscr{D}(G)$ and $GF=g$. Set $f_n = (\lambda-G)F_n$, then $\text{s-}\lim_{n\to\infty} f_n = \lambda F - g$ implies $\text{s-}\lim_{n\to\infty} (\lambda-G)^{-1} f_n = (\lambda-G)^{-1}(\lambda F - g) = \text{s-}\lim_{n\to\infty} F_n$. Thus $F \in \mathscr{D}(G)$ and $F = (\lambda-G)^{-1}(\lambda Fg)$ or $(\lambda-G)F = \lambda F - g$ which implies $GF=g$, q.e.d.

(iii) If we denote $(\lambda-G)^{-1}$ by $R(\lambda)$ then hypotheses (i)-(iv) of our theorem clearly imply that the family of operators $R(\lambda)$ satisfy (3.2.2) (i)-(iv)- the reader should check this assertion for himself. In the course of our proof use is made of

Lemma 3.2.3. Let A and B denote bounded linear operators with domain X and suppose that $AB=BA$. Then $\exp(A) = \sum_{n=0}^{\infty} A^n/n!$ is also a bounded linear operator and moreover

(i) $|\exp(A)| \leq \exp(|A|)$

(ii) $\exp(cI) = \exp(c)I$, c a constant

(iii) $\exp(A)\exp(B) = \exp(A+B)$

(iv) $\lim_{t\to 0} |t^{-1}(\exp(tA)-I)-A| = 0$

(v) If $|\exp(tA)| \leq 1$, $|\exp(tB)| \leq 1$ then $|\exp(tA)f - \exp(tB)f| \leq t|Af-Bf|$, all $f \in X$ and all $t \geq 0$.

Proof: The proofs of (i), (ii) and (iii) are obvious. Now $|\exp(tA)-I-tA| \leq \sum_{n=2}^{\infty} t^n|A|^n/n! = O(t^2)$, and this suffices to establish (iv). This proves that every bounded linear operator A generates a strongly continuous semi-group exp(tA). To prove (v) we write exp(tA)f-exp(tB)f as a "telescoping sum"

$$\sum_{k=1}^{n} \exp(\frac{(k-1)t}{n} A)\ \exp\ (\frac{(n-k)t}{n} B)(\exp(\frac{tA}{n})f-\exp(\frac{tB}{n})f)$$

and therefore

$$|\exp(tA)f-\exp(tB)f| \leq n|\exp(\frac{tA}{n})f-\exp(\frac{tB}{n})f|.$$

But $\exp(\frac{tA}{n})f-\exp(\frac{tB}{n})f = (\exp(\frac{tA}{n})f-f-(\exp(\frac{tB}{n})f-f))$

and so

$$n|\exp(\frac{tA}{n})f-\exp(\frac{tB}{n})f|=t|\ \frac{\exp(\frac{tA}{n})f-f}{t/n} - \frac{\exp(\frac{tB}{n})f-f}{t/n}\ |$$

Let $n\to\infty$ and use (iv) to deduce

$$|\exp(tA)f-\exp(tB)f| \leq t|Af-Bf|.$$

We turn now to the proof of the Hille-Yosida theorem. Our first step is to show that even though G is in general an unbounded linear operator it can be approximated by a family of bounded linear operators G_λ. More precisely we have the "Yosida approximation"

(3.2.3) $G_\lambda f = \lambda GR(\lambda)f = \lambda(\lambda R(\lambda)f-f)$.

Clearly $|G_\lambda f| \leq 2\lambda|f|$, since $|\lambda R(\lambda)| \leq 1$, and $G_\lambda G_\mu = G_\mu G_\lambda$ because $R(\lambda)R(\mu) = R(\mu)R(\lambda)$. Thus the operators G_λ satisfy the hypotheses of lemma 3.2.3. In addition if $f\in \mathcal{D}(G)$ then $\lim_{\lambda\to\infty} |G_\lambda f-Gf| = \lim_{\lambda\to\infty} |\lambda R(\lambda)Gf-Gf| = 0$.

This suggests that we define T(t) = exp(tG) via the limit: $\lim_{\lambda\to\infty} \exp(tG_\lambda) = \lim_{\lambda\to\infty} T_\lambda(t)$. According to lemma 3.2.3

$$|\exp(tG_\lambda)| = |\exp(-t\lambda)\exp(t\lambda^2R(\lambda))|$$
$$\leq \exp(-t\lambda)\exp(t\lambda|\lambda R(\lambda)|) = 1.$$ Therefore

$$|T_\lambda(t)f-T_\mu(t)f| \leq t|G_\lambda f-G_\mu f| \text{ and } f\in \mathcal{D}(G)$$

imply that $\underset{\lambda\to\infty}{\text{s-lim}}\, T_\lambda(t)f = T(t)f$ exists, and the convergence is uniform for $t \in$ compact subsets of R_+. Since $|T_\lambda(t)| \leq 1$ and $\mathcal{D}(G)$ is dense in X it follows that $|T(t)| \leq 1$ and $T(t)$ can be extended uniquely to all of X. All that remains to be done is (i) to establish the semi-group property $T(t+s) = T(t)T(s)$ (ii) the strong continuity of the semi-group and (iii) that G is the infinitesimal generator of $T(t)$.

But $T_\lambda(t)T_\lambda(s)f = T_\lambda(t+s)f$ imply that as $\lambda\to\infty$ the semi-group property is preserved in the limit. To prove strong continuity we begin with the inequality

$$|T(t)f-f| \leq |T(t)f-T_\lambda(t)f|+|T_\lambda(t)f-f|.$$

Given $\varepsilon > 0$ there exists λ_0 such that $\lambda \geq \lambda_0$ implies $|T(t)f-T_\lambda(t)f| < \varepsilon/2$ all $0 \leq t \leq 1$. Moreover $|T_\lambda(t)f-f| \leq \int_0^t |T_\lambda(t)G_\lambda f|ds \leq t|G_\lambda f| \leq 2\lambda t|f|$. By choosing $2\lambda t|f| < \frac{\varepsilon}{2}$ or equivalently $t < \varepsilon(4\lambda|f|)^{-1} = \delta$ we have $|T_\lambda(t)f-f| < \varepsilon/2$. Thus for $t < \delta$ we have $|T(t)f-f| < \varepsilon$.

Let $\tilde{G}$ denote the infinitesimal generator of $T(t)$; we want to prove that $\tilde{G} = G$. Now G_λ is certainly the infinitesimal generator of $\exp(tG_\lambda)$ so

$$T_\lambda(t)f-f = \int_0^t T_\lambda(s)G_\lambda f\, ds.$$ If $f \in \mathcal{D}(G)$ then

$$|T(s)Gf-T_\lambda(s)G_\lambda f| \leq |T(s)Gf-T_\lambda(s)Gf| + |T_\lambda(s)Gf-T_\lambda(s)G_\lambda f|$$
$$\leq |T(s)Gf-T_\lambda(s)Gf| + |Gf-G_\lambda f|$$

since $|T_\lambda(s)| \leq 1$. From this it follows at once that $\lim_{\lambda\to\infty} |T(s)Gf-T_\lambda(s)G_\lambda f| = 0$ uniformly for $s \in$ compact subintervals of R_+. Therefore $\lim_{\lambda\to\infty} \int_0^t T_\lambda(s)Gfds = \int_0^t T(s)Gfds$ $\lim_{\lambda\to\infty} T_\lambda(t)f-f = T(t)f-f = \int_0^t T(s)Gfds$. Thus if $f \in \mathcal{D}(G)$ then $f \in \mathcal{D}(\tilde{G})$ and $\tilde{G} f = G f$.

Finally we show $\mathcal{D}(\tilde{G}) \subset \mathcal{D}(G)$.

Suppose $\lambda \tilde{F}_\lambda - \tilde{G} \tilde{F}_\lambda = f$ and

$\lambda F_\lambda - G F_\lambda = f$. Since $F_\lambda \in \mathcal{D}(G) \subset \mathcal{D}(\tilde{G})$

we have $G F_\lambda = \tilde{G} F_\lambda$ and therefore F_λ is also a solution to the equation $\lambda F_\lambda - \tilde{G} F_\lambda = f$. But the solution to such an equation is unique, hence

$\tilde{F}_\lambda = F_\lambda$ and this implies $\mathcal{D}(\tilde{G}) \subset \mathcal{D}(G)$.

A refinement: For the applications to probability theory we need to know whether or not $T(t)$ is positivity-preserving. More precisely, if X is a function space, say $C[a,b]$, and $X^+ = \{f(x) : f(x) \geq 0\}$ is it true that $T(t)X^+ \subset X^+$?

Corollary: Suppose in addition to the hypotheses of the Hille-Yosida theorem we have a closed cone $X^+ \subset X$ (i.e. $f,g \in X^+$, $C_1, C_2 \geq 0$ then $C_1 f + C_2 g \in X^+$) such that $f \in X^+$ implies $R(\lambda) f \in X^+$. Then $T(t)\, X^+ \subset X^+$.

Proof: Since $T_\lambda(t) f = \exp(-t\lambda)\exp(t\lambda^2 R(\lambda)) f$, it follows that $R(\lambda) f \in X^+$ implies $T_\lambda(t) f \in X^+$ and hence $\text{s-}\lim_{\lambda\to\infty} T_\lambda(t) f = T(t) f \in X^+$.

Remarks: (i) The hypothesis $|T(t)| \leq 1$ is merely a convenience. We could just as well have assumed.

$$|T(t)| \leq M \text{ and } |(\lambda - G)^{-k}| \leq M\lambda^{-k}.$$

(ii) Suppose only that there exists a $\lambda_0 > 0$ and a constant M such that $\lambda > \lambda_0$ implies $(\lambda-G)^{-k} \leq M(\lambda-\lambda_0)^{-k}$. Now consider the operator $G = G - \lambda_0 I$. For every $\lambda > \lambda_0$

$$|(\lambda - \tilde{G})^{-k}| = |((\lambda+\lambda_0) - G)^{-k}| \leq M\lambda^{-k}.$$

hence $\tilde{G}$ generates a strongly continuous semi-group $\tilde{T}(t)$ of norm $\leq M$. Finally we observe that the existence of a strongly continuous semi-group $T(t)$ with infinitesimal generator G implies the existence and uniqueness of solutions to an abstract Cauchy Problem (ACP).

Theorem 3.2.2. Suppose $T(t)\colon X \to X$ is a strongly continuous contraction semi-group with infinitesimal generator G. .

If $f \in \mathcal{D}(G)$ then $U(t) = T(t)f$ is the unique solution of the equation of evolution

$U'(t) = GU(t)$, $U(0)=f$ which satisfies the conditions

(i) $U(t)$ is strongly differentiable and

$U'(t)$ is strongly continuous

(ii) $|U(t)| \leq C \exp(kt)$, C and k constants

(iii) $\operatorname{s-lim}_{t\to 0} U(t) = U(0) = f$.

Proof: Cf Dynkin [11] p.28, Theorem 1.3.

Examples: Brownian motion

An illuminating example of the general theory is provided by the standard Brownian motion process (also called the Wiener process) subjected to various boundary conditions e.g. reflection, absorption or adhesion, etc. Set $p(t,x,y) = (2\pi t)^{-\frac{1}{2}} \exp(-(y-x)^2/2t)$ and $P(t,x,A) = \int_A p(t,x,y)dy$. Then $P(t,x,A)$ is the transition function of the Brownian motion process $x(t)$. It is easily checked that $p(t,x,y)$ satisfies the heat equation

$$p_t(t,x,y) = (\tfrac{1}{2})p_{xx}(t,x,y)$$

(3.2.4)

$$\lim_{t\downarrow 0} p(t,x,y) = \delta(x-y)$$

Proceeding in a purely formal manner we see at once that the infinitesimal generator of the corresponding semi-group $T(t)$ is $Gf(x) = (\frac{1}{2})f''(x)$. More precisely we have the following result:

Theorem 3.2.3. Let $UC(-\infty,\infty)$ denote the set of bounded uniformly continuous functions. Then the Brownian motion semi-group $T(t)$: $UC(-\infty,\infty)\to UC(-\infty,\infty)$ is strongly continuous with $\mathcal{D}(G) = \{f: f,f'' \in UC(-\infty,\infty)\}$. It is easy to prove this directly using the explicit form of the $p(t,x,y)$ see Dynkin [11], v. 1, pp. 65-66.

Let us pretend, however, we are only given the operater $(\frac{1}{2})\frac{d^2}{dx^2}$ and our task is to construct the semi-group T(t) via the Hille-Yosida theorem. In particular this means we must study the "stationary equation"

(3.2.5) $\lambda g(x) - (\frac{1}{2})g''(x) = f(x), \quad f(x) \in C(-\infty,\infty)$

Lemma 3.2.4. Suppose $g \in C(-\infty,\infty)$ is a solution to the equation (3.2.5). Then $\lambda|g| \leq |f|$ and hence g is the unique bounded solution to (3.2.5).

Proof: Set $g_\varepsilon(x) = g(x)/(1+\varepsilon x^2), \varepsilon > 0$.

Then an easy calculation which we omit shows that g_ε satisfies the equation

(3.2.6) $$(\lambda - \frac{\varepsilon}{1+\varepsilon x^2})g_\varepsilon(x) - (\frac{2\varepsilon x}{1+\varepsilon x^2})g_\varepsilon'(x) - \frac{1}{2}g_\varepsilon''(x) = \frac{f(x)}{1+\varepsilon x^2}.$$

Choose $0 < \varepsilon < \lambda$ so $0 < \varepsilon/(1+\varepsilon x^2) < \lambda$ too. By hypothesis g(x) is bounded, thus $\lim_{|x|\to\infty} g_\varepsilon(x) = 0$. Let x_0 denote a point at which $\sup_{-\infty<x<\infty} g_\varepsilon(x)$ is attained.

Case 1: $x_0 = \pm\infty$, then $g_\varepsilon(x) \leq \sup_x g_\varepsilon(x) \leq 0$ since g_ε vanishes at $\pm\infty$. In particular then $g_\varepsilon(x) \leq \lambda^{-1}|f|$.

Case 2: $|x_0| < \infty$, then $g_\varepsilon'(x_0)=0$, $g_\varepsilon''(x_0) \leq 0$ imply $(\lambda - \frac{\varepsilon}{1+\varepsilon x_0^2})g_\varepsilon(x_0) \leq \frac{f(x_0)}{1+\varepsilon x_0^2} \leq \left|\frac{f(x)}{1+\varepsilon x_0^2}\right| \leq |f|$.

Since $(\lambda - \frac{\varepsilon}{1+\varepsilon x_0^2}) > 0$ we infer

$$g_\varepsilon(x) \leq g_\varepsilon(x_0) \leq (\lambda - \frac{\varepsilon}{1+\varepsilon x_0^2})^{-1}|f| \leq (\lambda-\varepsilon)^{-1}|f|.$$

Thus $g(x) \leq (1+\varepsilon x^2)(\lambda-\varepsilon)^{-1}|f|, \varepsilon < \lambda$. This is true for every $\varepsilon < \lambda$; let $\varepsilon \to 0$ and deduce $g(x) \leq \lambda^{-1}|f|$. A similar argument with -g and -f yields the estimate $-g(x) \leq \lambda^{-1}|-f| = \lambda|f|$. Thus $|g| \leq \lambda^{-1}|f|$. The solution g(x) is given explicitly by the formula:

$$(3.2.7) \qquad g(x) = (2\lambda)^{-\frac{1}{2}} \exp(-x\sqrt{2\lambda}) \int_{-\infty}^{x} \exp(y\sqrt{2\lambda}) f(y)dy + (2\lambda)^{-\frac{1}{2}} \exp(-x\sqrt{2\lambda}) \int_{x}^{\infty} \exp(-y\sqrt{2\lambda}) f(y)dy.$$

We leave it to the reader to check that g is bounded whenever f is. Since $g'' = 2\lambda g(x) - 2f(x)$ it follows that $g'' \in C(-\infty,\infty)$ and hence by Landau's inequality - $|g'| \leq 4|g||g''|$ - it follows that g' is bounded and therefore g is uniformly continuous. Thus the closure of this class of functions is again a uniformly continuous class of functions; to satisfy the conditions of the Hille-Yosida theorem we must choose $X = UC(-\infty,\infty)$ and then check separately that $\mathcal{D}(G) = \{f: f, f'' \in UC(-\infty,\infty)\}$ is dense in X. We leave this task to the reader. Note that (3.2.7) implies $g \geq 0$ if $f \geq 0$ and hence $T(t) = \exp(tG)$ is positivity-preserving.

Let us now consider reflecting Brownian motion. Analytically this corresponds to the equation

$$(3.2.8) \qquad \begin{cases} \lambda g(x) - (\frac{1}{2}) g''(x) = f(x), & 0 \leq x < \infty \\ g'(0) = 0 \end{cases}$$

Theorem 3.2.4. To every $f \in C[0,\infty)$ there exists a unique $g \in C[0,\infty)$ satisfying the equation (3.2.8) and the boundary condition $g'(0) = 0$; moreover $\lambda|g| \leq |f|$ holds.

Proof: Let $\tilde{f}(x) = f(x)$ if $x \geq 0$ and set $\tilde{f}(x) = f(-x)$ if $x \leq 0$. Then $\tilde{f} \in C(-\infty,\infty)$ and if $\tilde{g}$ is the solution to (3.2.5) then $g_1(x) = \tilde{g}(x)$, $x \geq 0$ is a particular integral of (3.2.8). To satisfy the boundary condition bring in $\exp(-x\sqrt{2\lambda})$ which is a bounded solution to (3.2.8) on R_+. All we have to do is choose the constant C so that $g'(0) = 0$ where $g(x) = g_1(x) + C\exp(-x\sqrt{2\lambda})$; the choice $C = (2\lambda)^{-\frac{1}{2}} g_1'(0+)$ serves. We're still not done because we have yet to establish the estimate $\lambda|g| \leq |f|$. Familiar reasoning as in lemma 3.2.4 with the maximum principle yields the estimate $g(x) \leq (1 + \varepsilon x^2)(\lambda-\varepsilon)^{-1}|f|$. The only tricky point is what happens if $g_\varepsilon(x)$ has a maximum at $x_0 = 0$. Clearly in this case $g_\varepsilon'(0) = 0$ and $g_\varepsilon''(0) \leq 0$, for otherwise $g_\varepsilon(x)$ would be increasing in a neighborhood of 0, contrary to the hypotheses that 0 is a point at which $\sup_{x\geq 0} g_\varepsilon(x) = g_\varepsilon(0)$ is attained. Thus $\mathscr{D}(G) = \{f: f, f'' \in UC[0,\infty),\ f'(0) = 0\}$ and it is easy to check that $\overline{\mathscr{D}(G)} = UC[0,\infty)$. The Markov process corresponding to this semi-group is called the reflecting Brownian motion and is equivalent in law to $|x(t)|$. In Chapter IV we shall pursue this topic in a more systematic manner for operators of the form

$$Gf(x) = (a(x)/2)f''(x)+b(x)f'(x).$$

3. The Trotter-Kato Theorem

In our proof of the Hille-Yosida theorem in III.2 we derived, interalia, the following proposition:

Let $T(t)\colon X\to X$ be a strongly continuous contraction semi-group with infinitesimal generator G. Put $G_\lambda = \lambda(\lambda R(\lambda)-I)$. Then for every $f \in \mathscr{D}(G)$ $\operatorname{s-lim}_{\lambda\to\infty} G_\lambda f = Gf$ and $\operatorname{s-lim}_{\lambda\to\infty} T_\lambda(t)f = T(t)f$ i.e. $\operatorname{s-lim}_{\lambda\to\infty} \exp(tG_\lambda)f = \exp(tG)f$.

This suggests the following problem of independent interest.

Problem: Given a sequence of strongly continuous contraction semi-groups $T_n(t)\colon X\to X$, under what conditions is it true that $\operatorname{s-lim}_{n\to\infty} T_n(t)f = T(t)f$ where $T(t)$ is also a strongly continuous contraction semi-group? Let us proceed to derive a necessary condition. Clearly, if

(3.3.1) $$\operatorname{s-lim}_{n\to\infty} T_n(t)f = T(t)f \text{ all } f \in X,$$

uniformly for $t \in$ compact subsets of R_+, then

(3.3.2) $$\operatorname{s-lim}_{n\to\infty} \int_0^\infty e^{-\lambda t}\, T_n(t)f\,dt = \int_0^\infty e^{-\lambda t}\, T(t)f\,dt,\ \lambda > 0.$$

To avoid confusion with the Yosida approximation G_λ we denote the infinitesimal generators of $T_n(t)$ and $T(t)$ by A_n and A respectively. Thus a necessary condition for (3.3.1) to hold is that the resolvents converge strongly i.e.

(3.3.3) $$\operatorname{s-lim}_{n\to\infty} (\lambda-A_n)^{-1} f = (\lambda-A)^{-1} f, \text{ every } \lambda > 0.$$

Theorem 3.3.1. (Trotter-Kato): A necessary and sufficient condition for $\underset{n\to\infty}{\text{s-lim}}\ T_n(t)f = T(t)f$ all $f \in X$ is $\underset{n\to\infty}{\text{s-lim}}\ (\lambda-A_n)^{-1} f = (\lambda-A)^{-1} f$ for all f in a dense subset of X and some $\lambda > 0$.

Proof: The necessity has already been established. We first show that if $f = (\lambda-A)^{-1}g$ and $\underset{n\to\infty}{\text{s-lim}}\ (\lambda-A_n)^{-1}f = (\lambda-A)f$ then $\underset{n\to\infty}{\text{s-lim}}\ T_n(t)f = T(t)f$ uniformly for $t \in$ compact subsets of R_+. This is enough because such f are dense in X. Now $T_n(t)f-T(t)f = T_n(t)(\lambda-A)^{-1}g - T(t)(\lambda-A)^{-1}g$

$$= I_n + II_n + III_n \text{ where}$$

$$I_n = T_n(t)(\lambda-A)^{-1}g - (\lambda-A_n)^{-1}T_n(t)g$$

$$II_n = (\lambda-A_n)^{-1}T_n(t)g - (\lambda-A_n)^{-1}T(t)g$$

$$III_n = (\lambda-A_n)^{-1}T(t)g - T(t)(\lambda-A)^{-1}g$$

We show that $I_n\to 0$, $III_n\to 0$ and $II_n\to 0$ in the indicated order. $I_n = T_n(t)$ $\{(\lambda-A)^{-1}g - (\lambda-A_n)^{-1}g\}$ where we've used the fact that $T_n(t)$ commutes with $(\lambda-A_n)^{-1}$. Thus $|I_n| \le |(\lambda-A^{-1})g - (\lambda-A_n)^{-1}g| \to 0$ as $n\to\infty$ by hypothesis. Now the resolvents are uniformly bounded by λ^{-1}, so if we have strong convergence to 0 on a dense subset of X we have strong convergence to 0 on all of X.

$$|III_n| = |(\lambda-A_n)^{-1}\ T(t)g - (\lambda-A)^{-1}\ T(t)g|$$

where we've used the fact that $T(t)(\lambda-A)^{-1} = (\lambda-A)^{-1}\ T(t)$. Clearly $\lim_{n\to\infty}$ $|(\lambda-A_n)^{-1}\ T(t)g - (\lambda-A)^{-1}\ T(t)g| = 0$ for each t. Actually we have uniform convergence for $t \in$ compact subintervals of R_+ because $T(t)g$ is strongly continuous.

Finally we come to the term II_n:

$$II_n = (\lambda-A_n)^{-1}\ T_n(t)g - (\lambda-A_n)^{-1}\ T(t)g.$$

We may as well assume that g is of the form $(\lambda-A)^{-1}\tilde{g}$ since such g are dense in X.

Lemma 3.3.1.

(3.3.4) $$(\lambda-A_n)^{-1}(T(t)-T_n(t))(\lambda-A)^{-1}\tilde{g} = II_n$$
$$= \int_0^t T_n(t-s)\{(\lambda-A_n)^{-1} - (\lambda-A)^{-1}\}T(s)\tilde{g}\, ds.$$

$|II_n| \le \int_0^t |(\lambda-A_n)^{-1}T(s)\tilde{g} - (\lambda-A)^{-1}T(s)\tilde{g}|ds$, and $\lim_{n\to\infty} |(\lambda-A_n)^{-1}T(s)\tilde{g} - (\lambda-A)^{-1}T(s)\tilde{g}|$ $= 0$ in the sense of bounded pointwise convergence. So by the Lebesgue dominated convergence theorem $\lim_{n\to\infty} |II_n| = 0$. We turn now to the proof of the lemma. Set

$u(s) = T_n(t-s)(\lambda-A_n)^{-1}$, $v(s) = T(s)(\lambda-A)^{-1}$. Strictly speaking we should write $u(s)=T_n(t-s)(\lambda-A_n)^{-1} f_1$ and $v(s) = T(s)(\lambda-A)^{-1}f_2$ where $f_i \in X$. Clearly u and v are strongly differentiable and so is their product with derivative computed in the standard fashion:

$$(u(s)v(s))' = u'(s)v(s) + u(s)v'(s).$$

Now

$$u'(s) = (-1)T_n'(t-s)(\lambda-A_n)^{-1} = (-1)T_n(t-s)A_n(\lambda-A_n)^{-1}$$
$$= (-1)T_n(t-s)[\lambda(\lambda-A_n)^{-1}-I], \text{ and}$$

Similarly

$$v'(s) = T(s)[\lambda(\lambda-A)^{-1}-I].$$

So $u'(s)v(s) = (-1)T_n(t-s)[\lambda(\lambda-A_n)^{-1}-I]T(s)(\lambda-A)^{-1}$

$$u(s)v'(s) = T_n(t-s)(\lambda-A_n)^{-1}T(s)[\lambda(\lambda-A)^{-1}-I].$$

Adding and making the obvious cancellation we obtain

(3.3.5) $$(u(s)v(s))' = T_n(t-s)T(s)(\lambda-A)^{-1} - T_n(t-s)(\lambda-A_n)^{-1}T(s)$$
$$= T_n(t-s)[(\lambda-A)^{-1} - (\lambda-A_n)^{-1}]T(s) \text{ where}$$

we've used the fact that $T(s)(\lambda-A)^{-1} = (\lambda-A)^{-1}T(s)$.

Integrating both sides of (3.3.5) from 0 to t yields

$$u(t)v(t)-u(0)v(0) = \int_0^t T_n(t-s)[(\lambda-A)^{-1}-(\lambda-A_n)^{-1}]T(s)ds.$$

But $u(t)v(t) = (\lambda-A_n)^{-1} T(t)(\lambda-A)^{-1}$ and

$u(0)v(0) = T_n(t)(\lambda-A_n)^{-1}(\lambda-A)^{-1}$ so

$u(t)v(t)-u(0)v(0) = (\lambda-A_n)^{-1}(T(t)-T_n(t))(\lambda-A)^{-1}$

where we've used again the fact that a semi-group commutes with its resolvent. This completes the proof of the lemma and hence the Trotter-Kato Theorem.

Definition 3.3.1. Let $\mathscr{C}$ be a dense subset of $\mathscr{D}(A)$. We shall say that $\mathscr{C}$ is a core if for every $f \in \mathscr{D}(A)$ there exists a sequence $f_n \in \mathscr{C}$ such that

$$\operatorname*{s-lim}_{n\to\infty} f_n = f \text{ and } \operatorname*{s-lim}_{n\to\infty} Af_n = Af.$$

Although the necessary and sufficient condition of the Trotter-Kato theorem is in general indispensable the condition itself is not always easy to check in practice. For some purposes the following condition suffices:

Lemma 3.3.1. Let $\mathscr{C}$ denote a core for the operator A and suppose $A, A_1, A_2, \ldots A_n$ are generators of strongly continuous contraction semi-groups $T(t), T_1(t), \ldots T_n(t)$, respectively. Suppose for all $f \in \mathscr{C}$, $A_n f$ is defined and $\operatorname*{s-lim}_{n\to\infty} A_n f = Af$.

Then for every $\lambda > 0$ we have

$$\operatorname*{s-lim}_{n\to\infty} (\lambda-A_n)^{-1} f = (\lambda-A)^{-1} f, \text{ all } f \in X.$$

Proof: First we show that $(\lambda-A)\mathscr{C}$ is a dense subset of X. Clearly $(\lambda-A)\mathscr{D}(A) = X$. Now pick $f_n \in \mathscr{C}$ with the property $\operatorname*{s-lim}_{n\to\infty} f_n = f$, $\operatorname*{s-lim}_{n\to\infty} Af_n = Af$, which exists for every $f \in \mathscr{D}(A)$, to deduce that $\operatorname*{s-lim}_{n\to\infty} (\lambda-A)f_n = (\lambda-A)f$. Put $g = (\lambda-A)f$ where $f \in \mathscr{C}$; such g are dense in X and for g of this form we shall prove $\operatorname*{s-lim}_{n\to\infty} (\lambda-A_n)^{-1}g = (\lambda-A)^{-1}g$.

Observe that $(A_n-A)f = [(\lambda-A)-(\lambda-A_n)]f = g-(\lambda-A_n)(\lambda-A)^{-1}g$. So $(\lambda-A_n)^{-1}(A_n-A)f = (\lambda-A_n)^{-1}g-(\lambda-A)^{-1}g$. Thus $|(\lambda-A_n)^{-1}g-(\lambda-A)^{-1}g| \leq \lambda^{-1}|(A_n-A)f|$. By hypothesis $\lim_{n\to\infty} |(A_n-A)f| = 0$ for $f \in \mathscr{C}$ and hence $\lim_{n\to\infty} |(\lambda-A_n)^{-1}g-(\lambda-A)^{-1}g| = 0$ for all g in a dense subset of X and therefore for all $g \in X$.

Extensions: typical applications of the Trotter-Kato theorem include (i) numerical solutions of partial differential equations and (ii) the so-called "diffusion approximation in genetics. The preceeding theory is not, however, directly applicable here, since the semi-groups $T_n(t)$ act on Banach spaces X_n which differ from X. This leads quite naturally to the notion of an "Approximating sequence of Banach spaces". Suppose we are given a sequence of Banach spaces X_n and a sequence of bounded linear operators π_n: $X \to X_n$ with the property $\lim_{n\to\infty} |\pi_n f| = |f|$ all $f \in X$.

The uniform boundedness principle then implies $|\pi_n| \leq M$, $n=1,2,\ldots$

Definition 3.3.2. (i) A sequence of Banach spaces X_n together with a sequence of bounded linear maps $\pi_n: X \to X_n$ is said to approximate the Banach space X if $\lim_{n\to\infty} |\pi_n f| = |f|$ all $f \in X$. (ii) We say that $\lim_{n\to\infty} f_n = f$, $f_n \in X_n$ if $\lim_{n\to\infty} |\pi_n f - f_n| = 0$.

Theorem 3.3.2. Suppose $\{X_n, \pi_n\}$ approximates the Banach space X and $T_n(t)$: $X_n \to X_n$, $T(t)$: $X \to X$ are strongly continuous contraction semi-groups with infinitesimal generaters A_n and A respectively. If for some $\lambda > 0$

(3.3.5) $$\lim_{n\to\infty} |\pi_n(\lambda - A)^{-1} g - (\lambda - A_n)^{-1} \pi_n g| = 0 \text{ all } g \in X,$$

or a dense subset of X, then $\lim_{n\to\infty} |T_n(t)\pi_n f - \pi_n T(t) f| = 0$ uniformly for $t \in$ compact subsets of R_+.

We omit the proof because it is similar to the proof of the original Trotter-Kato theorem, a careful discussion of which we've already given. For additional variants the reader is advised to consult the treatise of Kato [17] or the papers of Kurtz [18], [19].

4. Perturbation theory - an introduction

Suppose it is known that B and C generate strongly continuous semigroups exp(tB), exp(t C) respectively: under what conditions can it be inferred that B + C generates a strongly continuous semigroups exp(t(B+C))? Since many of the operators that occur in mathematical physics, e.g., transport theory see, and analysis are of the form A = B + C, the results on this problem to be presented below will turn out to have immediate and interesting applications - see [1], [16]. But, first, we need a modest amount of new terminology.

Definition 3.4.1. The graph of a linear operator A with domain $\mathcal{D}(A)$ is the set $G(A) = \{(f,Af): f \in \mathcal{D}(A)\}$; G(A) is a linear subspace of X x X. We endow X x X with the product topology derived from the norm $|(f,g)| = (|f|^2 + |g|^2)^{1/2}$. With this norm, as is easily checked, X x X is again a Banach Space.

Remark: The closure $\overline{G(A)}$ of the graph of an operator A is not in general the graph of an operator. If, however $\operatorname{s-lim}_{n\to\infty} f_n = 0$ and $\operatorname{s-lim}_{n\to\infty} Af_n = g$ together imply g = 0, then it is clear that $\overline{G(A)}$ is the graph of an operator $\tilde{A}$, which is closed. $\tilde{A}$ is sometimes denoted by $\overline{A|\mathcal{D}(A)}$. We shall say that a linear subspace D is a core of A (A is assumed to be a closed operator) if $A = \overline{A|D}$.

Notation: B(X,X) = set of all bounded linear operators from X to X. We begin our introduction to perturbation theory with a result of independent interest due to P. Chernoff [7].

Theorem 3.4.1: Let $F(t)$: $[0,\infty) \to B(X,X)$ be strongly continuous, i.e. $F(t)f$: $[0,\infty) \to X$ is for each $f \in X$ strongly continuous, with $F(0) = I$, $|F(t)| \leq 1$. Let $A = F'(0)$ denote the s-derivative at 0 and suppose that the closure of A generates a strongly continuous contraction semigroup $\exp(tA)$. Then

$$\text{s-}\lim_{n\to\infty} F(\tfrac{t}{n})^n = \exp(tA).$$

Before proceeding to the proof we derive the Hille-exponential formula as a consequence. We assume therefore that A generates a strongly continuous contraction semigroup $\exp(tA)$. Set $F(t) = (I-tA)^{-1} = t^{-1}R(t^{-1})$ where $R(\lambda) = (\lambda I-A)^{-1}$. We claim that $F(t)$ satisfies the conditions of the Chernoff theorem. Clearly $|F(t)| \leq 1$ and is strongly continuous for $t > 0$. Moreover $\text{s-}\lim_{t\to\infty} F(t) = \text{s-}\lim_{\lambda\to\infty} \lambda R(\lambda) = I$, so $F(t)$ is strongly continuous on $[0,\infty)$. Finally, $t^{-1}(F(t) - I) = \lambda[\lambda R(\lambda)-I) = A_\lambda$, $t^{-1} = \lambda$, where A_λ denotes the Yosida approximation of A. Hence $\text{s-}\lim_{t\to 0} t^{-1}(F(t)-I)t = Af$ for $f \in \mathscr{D}(A)$, so $F'(0) = A$. Therefore,

$$\text{s-}\lim_{n\to\infty} F(t/n)^n = \text{s-}\lim_{n\to\infty} (I-tA/n)^{-n} = \exp(tA),$$

and this completes the derivation of the Hille-exponential formula. Another application is to the Trotter product formula:

Theorem 3.4.2: Suppose $\exp(tB)$, $\exp(tC)$ are strongly continuous contraction semigroups on X. Suppose the closure of $A = B + C$, with $\mathscr{D}(A) = \mathscr{D}(B) \cap \mathscr{D}(C)$, generates a strongly continuous contraction semigroup exp(tA). Then

$$\underset{n\to\infty}{\text{s-lim}}\ (\exp(\frac{tB}{n})\ \exp(\frac{tC}{n}))^n = \exp(t(B+C)).$$

Proof: Set $F(t) = \exp(tB)\exp(tC)$. If $f \in \mathcal{D}(B+C)$ then

$$t^{-1}(F(t)-I) = t^{-1}\{\exp(tB)(\exp(tC)-I) + \exp(tB)-I\}$$

implies

$$\underset{t\to 0^+}{\text{s-lim}}\ t^{-1}(F(t)-t)f = Cf + Bf = Af. \quad \text{Thus}$$

$$\underset{n\to\infty}{\text{s-lim}}\ F(\frac{t}{n})^n = \underset{n\to\infty}{\text{s-lim}}\ (\exp(\frac{tB}{n})\exp(\frac{tC}{n}))^n$$

$$= \exp(tA).$$

We return now to the proof of Chernoff's result.

Lemma 3.4.1. Let H denote a bounded linear operator of norm $|H| \leq 1$. Then $\exp(t(H-I))$ is a contraction semigroup. For all $f \in X$ we have

$$|(\exp n(H-I) - H^n)f| \leq n^{1/2}|(H-I)f|.$$

Proof: $\exp(t(H-I))$ is a semigroup because $|H-I| \leq 2$. In addition, we have

$$|\exp(t(H-I))| = \exp(-t)\left|\sum_{k=0}^{\infty} (t\ H)^k/k!\right|$$

$$\leq \exp(-t)\exp(t) = +1.$$

To complete the proof we need some elementary facts concerning the Poisson distribution. Suppose ξ is a random variable with Poisson distribution $P(\xi = k) = \exp(-n)n^k/k!$ for $k = 0,1,2,\ldots$ Then $\sum_{k=0}^{\infty} kP(\xi = k) = n$ and $\sum_{k=0}^{\infty} (k-n)^2 P(\xi = k) = n$.

By Schwarz' inequality

$$E(|\xi-n|) \leq E(|\xi-n|^2)^{1/2} = n^{1/2} \qquad \text{which means}$$

$$\exp(-n) \sum_{k=0}^{\infty} |k-n| n^k/k! \leq n^{1/2}.$$

Now

$$|\exp[n(H-I)]f-H^n f| = \exp(-n)\left|\sum_{k=0}^{\infty} n^k(H^k-H^n)f/k!\right|$$

$$\leq \exp(-n)\left|\sum_{k=0}^{\infty} (n^k/k!)[(H^k-H^n)f]\right|$$

$$\leq \exp(-n)\left|\sum_{k=0}^{\infty} (n^k/k!)|(H^{|k-n|}-I)f\right|$$

But $|H^{|k-n|}-I| = |H^{|k-n|}-I^{|k-n|}| \leq |k-n||H-I|$.

Thus

$$\exp(-n)\{\sum_{k=0}^{\infty} (n^k/k!)|(H^{|k-n|}-I)f|\} \leq$$

$$\exp(-n)\{\sum_{k=0}^{\infty} (n^k/k!)|k-n|\}|(H-I)f| \leq n^{1/2}|(H-I)f|.$$

This completes the proof of the lemma. We turn now to the proof of the theorem itself.

$\mathcal{D}(A) = \{f: \operatorname{s-lim}_{t\to 0} t^{-1}(F(t)f-f) = AF \text{ exists}\}$

Set $A_n = (n/t)(F(t/n) - I)$ for fixed $t > 0$ and note that $\exp(tA_n/n) = \exp(F(t/n) - I)$ is a contraction by Lemma 3.4.1. Thus $\exp(tA_n/n)$ is a semigroup of contractions and therefore so is $\exp(tA_n)$. In addition,

$$\text{s-}\lim_{n\to\infty} \exp(tA_n)f = \text{s-}\lim_{n\to\infty} \exp(n(F(t/n) - I))f$$

$$= \exp(tA)f, \text{ because}$$

$\text{s-}\lim_{n\to\infty} A_n f = Af$ all $f \in \mathscr{D}_{(A)}$; now apply lemma 3.3.2. On the other hand

$$|\exp(tA_n)f - F(t/n)^n f| = |\exp n(F(t/n) - I)f - F(t/n)^n f|$$

$$\leq n^{1/2}|(F(t/n) - I)|f|$$

$$= tn^{-1/2}|(t/n)^{-1}(F(t/n) - I)f|.$$

For $f \in \mathscr{D}(A)$ then

$\text{s-}\lim_{n\to\infty} |\exp(tA_n)f - F(t/n)^n f| = 0$. Thus for all $f \in \mathscr{D}(A)$

$$\text{s-}\lim_{n\to\infty} |F(t/n)^n f - \exp(tA)f| = 0.$$

Since $\mathscr{D}(A)$ is dense in X, the result holds for all $f \in X$. q.e.d.

We complete our brief introduction to the perturbation theory of semigroups of operators by presenting a theorem of Hille-Phillips [15]; the proof, however, is due to Kato [17] pp. 495-496.

Theorem 3.4.2. Suppose B generates a strongly continuous semigroup $U(t) = \exp(tB)$ with $|U(t)| \leq M \exp(\beta t)$. Let C denote a bounded linear operator with norm $|C|$. Then $V(t) = \exp(t(B+C))$ is a strongly continuous semigroup such that $|V(t)| \leq M \exp t(\beta + M|C|)$, and $\mathscr{D}(B+C) = \mathscr{D}(B)$.

The proof relies on the following

Lemma 3.4.2. Suppose u(t) is a solution to the (inhomogeneous) equation of evolution

$$\begin{cases} u'(t) = Bu(t) + f(t), \ f(t) \text{ is s-continuous} \\ u(0) = u_0, \end{cases}$$

Then $u(t) = U(t)u_0 + \int_0^t U(t-s)f(s)ds.$

Proof:
$$(U(t-s)u(s))' = -U'(t-s)u(s) + U(t-s)u'(s)$$
$$= U(t-s)Bu(s) + U(t-s)[Bu(s) + f(s)]$$
$$= U(t-s)f(s).$$

Integrating both sides from 0 to t yields

$$u(t) - U(t)u_0 = \int_0^t U(t-s)f(s)ds. \quad \text{q.e.d.}$$

To motivate the proof we observe that if B + C generates a strongly continuous semigroup $V(t) = \exp(t(B+C))$ then the equation of evolution

$$\begin{cases} v'(t) = (B+C)v(t) \\ v(0) = v_0 \in \mathcal{D}(B+C) = \mathcal{D}(B) \end{cases}$$

has the unique solution $v(t) = V(t)v_0$.

By the lemma just proved v(t) satisfies the integral equation

(3.4.1) $$v(t) = U(t)v_0 + \int_0^t U(t-s)Cv(s)ds.$$

Under our hypotheses we can actually solve the integral eqation (3.4.1) by the method of successive approximations. To see this, set

$$V(t) = \sum_{n=0}^{\infty} U_n(t) \quad \text{where} \quad U_0(t) = U(t),$$

$$U_n(t) = \int_0^t U(t-s)CU_{n-1}(s)ds, \; n = 1,2,\ldots$$

Proceeding formally we have

$$V(t) = U_0(t) + \int_0^t U(t-s)\; C\{\sum_{n=0}^{\infty} U_n(s)\}\; ds$$

(3.4.2) $$V(t) = U_0(t) + \int_0^t U(t-s)C\; V(s)ds.$$

To make this rigorous all we have to do is to establish the convergence of the infinite series $\sum_{n=0}^{\infty} U_n(t)$. We claim: $|U_n(t)| \le M^{n+1}|C|^n \exp(\beta t)\; t^n/n!$, $n = 0, 1,\ldots$

The case n = 0 is just the hypotheses of theorem 3.4.2. We proceed by induction on n.

$$\left|\int_0^t U(t-s)CU_n(s)ds\right| \le \frac{1}{n!} \int_0^t M \exp(t(\beta-s))|C|M^{n+1}|C|^n \exp(\beta s)s^n ds$$

$$= M^{n+2}|C|^{n+1} \exp(\beta t) \int_0^t (s^n/n!)ds$$

$$= M^{n+2}|C|^{n+1} \exp(\beta t)/(n+1)!.$$

Therefore $\sum_{n=0}^{\infty} |U_n(t)|$ is absolutely convergent and less than or equal to $M \exp(\beta t) \sum_{n=0}^{\infty} (M|C|)^n t^n/n! = M \exp(\beta + M|C|)t$. The next step is to show that $R_1(\lambda) = \int_0^{\infty} \exp(-\lambda t)V(t)dt$ for $\lambda > \lambda_0 = \beta + M|C|$ is in fact a a resolvent family. In fact. we shall show that if $f \in X$, $\lambda' > \lambda_0$ then

$F_\lambda = R_1(\lambda)f$ satisfies the equation $(\lambda-B-C)F_\lambda = f$ and the inequalities $|F_\lambda|^k \leq M(\lambda-\lambda_0)^{-k}$, $k = 1,2,\ldots$ To see this multiply both sides of (3.4.2) by $\exp(-\lambda t)$ and integrate to obtain

$$R_1(\lambda) = \int_0^\infty \exp(-\lambda t)U(t)dt + \int_0^\infty \exp(-\lambda t)\{\int_0^t U(t-s)CV(s)ds\}dt$$

$$= R(\lambda) + \int_0^\infty \exp(-\lambda t) \{\int_0^\infty U(t-s)CV(s)ds\}dt,$$

$$= R(\lambda) + \int_0^\infty \{\int_s^\infty \exp(-\lambda t)U(t-s)CV(s)dt\}ds.$$

But

$$\int_0^\infty \{\int_s^\infty \exp(-\lambda t)U(t-s)CV(s)dt\}ds =$$

$$\int_0^\infty \{\int_s^\infty \exp(-\lambda(t-s)U(t-s)\exp(-\lambda s)CV(s)dt\}ds =$$

$$\int_0^\infty \exp(-\lambda t)U(t)dt \cdot \int_0^\infty \exp(-\lambda s)CV(s)ds.$$

Thus

(3.4.3) $$R_1(\lambda) = R(\lambda) + R(\lambda)CR_1(\lambda) = R(\lambda)(I+CR_1(\lambda)).$$

In particular $R_1(\lambda): X \to \mathscr{D}(B)$. Apply $(\lambda-B)$ to both sides of 3.4.3 and obtain $(\lambda-B)R_1(\lambda) = I + CR_1(\lambda)$ or equivalently $(\lambda-B-C)R_1(\lambda) = I$. Thus $R_1(\lambda) = [\lambda-(B+C)]^{-1}$. We now proceed to estimate

$$|R_1(\lambda)^{k+1}| = (1/k!)|R_1(\lambda)^{k+1}| = (1/k!)|D^kR_1(\lambda)|$$ (where D^k denotes the k^{th} derivative)

$$\leq (1/k!) \int_0^\infty t^k\exp(-\lambda t)|V(t)|dt$$

$$\leq (M/k!) \int_0^\infty t^k\exp(-t(\lambda-\lambda_0))dt$$

$$= (M/k!)(\lambda-\lambda_0)^{-(k+1)}\Gamma(k+1) = M(\lambda-\lambda_0)^{-(k+1)};$$

Γ denotes the Gamma function. According to remark(ii) after the proof of Theorem 3.2.1, it follows that $R_1(\lambda)$ is the resolvent family of a semigroup $U_1(t)$. The uniqueness of the Laplace transform now implies $U_1(t) = V(t)$, so $V(t)$ itself is a semigroup.

IV. Applications to one dimensional Markov processes

1. The equation $\lambda F_\lambda(x)-(\frac{1}{2})a(x)F_\lambda''(x)-b(x)F_\lambda'(x) = f(x)$- existence and boundary behavior of F_λ.

Intuitively it is clear what a probabilist means by reflecting or absorbing Brownian motion on $[0,\infty)$: in the interior of $(0,\infty)$ the process is ordinary Brownian motion until it reaches the boundary 0, at which time it is either immediately reflected back into the interior of $(0,\infty)$ or it is absorbed at 0. It is possible, in fact, to impose more complicated boundary conditions than reflection or absorption. For the general class of diffusion processes studied in Chapter II the situation is less obvious. In genetics, for example, there occur diffusion processes with state space $[0,1]$ for which it is not possible to impose a reflecting boundary condition at 0, say. It is our purpose in this chapter to study the possible boundary conditions that can be imposed upon a one-dimensional Markov process $x(t,\omega)$ with state space $[r_0,r_1]$ and diffusion and drift coefficients $a(x)$ and $b(x)$. It was shown in Chapter II that the existence of such a process is equivalent to the existence of a Markovian semi-group $T(t)$ with infinitesimal generater $Gf(x) = (a(x)/2)f''(x)+b(x)f'(x)$. Analytically, then, our problem is to describe all those domains $\mathscr{D}(G)$ for which $\overline{G|\mathscr{D}(G)}$ generates a Markovian semi-group; and thus will be done by means of a careful study of the "stationary equation"

$$\lambda F_\lambda(x)-GF_\lambda(x) = f(x),\ \lambda > 0. \tag{4.1.1}$$

Now this is just a second order linear differential equation of the sort discussed in any good undergraduate text on ordinary differential equations. The existence of a Markov process $x(t)$ satisfying various "boundary conditions" is thereby reduced to the study of solutions to equation (4.1.1) satisfying boundary conditions in the classical sense. For example, the boundary condition corresponding to reflection is $F_\lambda'(r_0) = 0$ and the boundary condition for absorption is $F_\lambda(r_0) = 0$, etc.

Our plan then is to construct the most general solution to equation (4.1.1) - a routine but tedious process - and then construct resolvent families $(\lambda-G)^{-1}$ satisfying the Hille-Yosida theorem. Before proceeding to the study of solutions of (4.1.1), however we make the following notational conventions and assumptions:

(i) $[r_0,r_1]$ and (r_0,r_1) denote the closed and open intervals with end points r_0,r_1 with $-\infty \leq r_0 < r_1 \leq \infty$; "a" will denote a fixed but arbitrary point in the interior of (r_0,r_1).

(ii) $C[r_0,r_1]$ is the Banach space of bounded continuous functions on $[r_0,r_1]$ with $\lim_{x\to r_i} f(x) = f(r_i)$ and normed with the sup norm

(4.1.2) $|f| = \sup_{r_0 \leq x \leq r_1} |f(x)|$; $C(r_0,r_1)$ is the set of functions continuous on the open interval (r_0,r_1), but not necessarily bounded;

$C^k[r_0,r_1] = \{f: \; f^{(\ell)} \in C[r_0,r_1],\; 0 \leq \ell \leq k\}$

(iii) $a(x) > 0$ on (r_0,r_1), $a(x) \in C(r_0,r_1)$;

however we do allow $a(r_i) = 0$.

(iv) $b(x) \in C(r_0,r_1)$

Examples:

(i) Brownian motion: $r_0 = -\infty$, $r_1 = +\infty$, $a(x) = 1, b(x) = 0$

(ii) Ornstein-Uhlenbeck process: $r_0 = -\infty$, $r_1 = +\infty$, $a(x) = 1$, $b(x) = -kx$.

(iii) Radial component of n-dimensional Brownian motion: $r_0 = 0$, $r_1 = +\infty$, $a(x) = 1$, $b(x) = (n-1)/2x$.

(iv) Markov processes in genetics: $r_0 = 0, r_1 = 1$, $a(x) = Ax(1-x)$, $b(x) = Bx(1-x)$.

Following Feller we define the monotone increasing functions $p(x)$ and $m(x)$ via the formulae:

$$(4.1.3)\quad \begin{cases} p(x) = \int_a^x \exp(-B(y))dy \\ m(x) = 2\int_a^x a(y)^{-1} \exp(B(y))dy \\ B(x) = 2\int_a^x a(y)^{-1} \cdot b(y)dy \end{cases}$$

Remark: p is called the scale and m the speed measure. Since p and m are monotone increasing functions they induce measures on (r_0, r_1) denoted by $dp(x) = \exp(-B(x))dx$ and $dm(x) = 2a(x)^{-1} \exp(B(x))dx$; it is obvious that these measures are independent of the choice of a. It was observed by Feller that the allowable domains $\mathcal{D}(G)$ depend on the boundary behavior of the functions

$$(4.1.4)\quad \begin{cases} \text{(i)} \quad u_1(x) = \int_a^x m(s)dp(s) & \text{and} \\ \text{(ii)} \quad v_1(x) = \int_a^x p(x)dm(s) \end{cases}$$

Definition 4.1.1.

(i) We say that r_i is an accessible boundary if $u_1(r_i) < \infty$ and inaccessible if $u_1(r_i) = +\infty$.

(ii) An accessible boundary is regular if $v_1(r_i) < \infty$ and is exit if $v_1(r_i) = +\infty$.

(iii) an inaccessible boundary is entrance if $v_1(r_i) < \infty$ and is natural if $v_1(r_i) = +\infty$.

Examples:

(1) Brownian motion, $r_0 = -\infty$, $r_1 = +\infty$ are both natural.

(2) Brownian motion on R_+, $r_0 = 0$, $r_1 = \infty$; here r_0 is regular, r_1 is natural.

(3) Radial component of n-dimensional Brownian motion with $a(x) = 1$, $b(x) = \frac{n-1}{2x}$; $r_0 = 0$ is entrance, $r_1 = \infty$ is natural.

Remark: It is to be observed that r_i is regular if and only if both $m(r_i)$ and $p(r_i)$ are finite. If r_i is an exit boundary then $p(r_i)$ is finite but $|m(r_i)| = +\infty$. If r_i is an entrance boundary then $|m(r_i)| < \infty$ but $|p(r_i)| = +\infty$.

A trivial calculation shows that

$$Gf(x) = (a(x)/2)\exp(-B(x))(\exp(B(x))f'(x))'$$

$$= \frac{(f'(x)/p'(x))'}{m'(x)};$$ thus suggests we define operators D_p and D_m via the formulae:

$$D_p f(x) = f'(x)/p'(x) = \lim_{h\to 0} \frac{f(x+h)-f(x)}{p(x+h)-p(x)}$$

(4.1.5)

$$D_m g(x) = g'(x)/m'(x) = \lim_{h\to 0} \frac{g(x+h)-g(x)}{m(x+h)-m(x)}$$

from which we arrive at the "Feller form"

(4.1.6) $\quad Gf(x) = D_m D_p f(x).$

The following properties are easily established and left to the reader -cf Mandl [21] p. 22.

(i) If $D_m D_p f(x) = g(x)$, $f,g \in C(r_0,r_1)$ then

$$f(x) = \int_a^x \int_a^y g(s)dm(s)dp(y) + f(a) + D_p f(a)\cdot p(x).$$

(4.1.7)

(ii) $\int_x^y D_p f(x)dp(s) = f(y)-f(x)$ and

$$\int_x^y D_m g(s)dm(s) = g(y)-g(x)$$

We now return to the study of the stationary equation (4.1.1) which can be completely solved, provided, we can construct two linearly independent solutions $w_1(x,\lambda)$ and $w_2(x,\lambda)$ of the homogeneous equation

(4.1.8) $\quad (\lambda - D_m D_p)w_i(x,\lambda) = 0.$ Since $\lambda > 0$ will be held fixed throughout the remainder of our discussion we shall often suppress the explicit dependence of w_i on λ. Our first step is to solve the homogeneous equation $(\lambda - D_m D_p)u(x) = 0$

subject to the initial conditions $u(a) = 1$, $D_p u(a) = 0$. This is equivalent to the integral equation

(4.1.9) $u(x) = 1 + \lambda Lu(x)$ where

$$Lf(x) = \int_a^x \int_a^y f(s)dm(s)dp(y).$$

This can be solved in the usual way by setting $u_0(x) = 1$, $u_{n+1}(x) = Lu_n(x)$ and defining $u(x) = \sum_{n=0}^{\infty} \lambda^n u_n(x)$. The convergence of this series follows at once from the

Lemma 4.1.1: $u_k(x) \leq u_j(x)u_1(x)^{k-j}/(k-j)!$ for $j \leq k$.

Proof: The lemma is obviously true for k=j and to prove it in the general case we proceed by mathematical induction on **k**. We use the fact, easily established, that $u_n(x)$ is monotone increasing on $[a,r_1]$ and monotone decreasing on $[r_0,a]$ with $u_n(a) = 0$, $n = 1,2,.,.$. In addition, $u_1(x) = Lu_0(x) = \int_a^x \int_a^y dm(s)dp(y) = \int_a^x m(y)dp(y)$, which is the way $u_1(x)$ is defined in definition 4.1.1; in particular $du_1(y) = m(y)dp(y)$. From the induction hypothesis and the monotonicity properties of $u_1(x)$ we infer

$$u_{k+1}(x) \leq u_j(x) \int_a^x \frac{u_1(y)^{k-j}}{(k-j)!} \int_a^y dm(s)dp(y)$$

$$= u_j(x) \int_a^x \frac{u_1(y)^{k-j}}{(k-j)!} du_1(y) = u_j(x)\ u_1(x)^{k-j}/(k-j)!$$

The proof of the lemma is now complete.

Corollary: $u(x) = u(x,\lambda) = \sum_{n=0}^{\infty} \lambda^n u_n(x)$ converges uniformly on compact subsets of (r_0,r_1) and the following inequality holds:

(4.1.10) $1+\lambda u_1(x) \leq u(x) \leq \exp(\lambda u_1(x))$.

In addition

$$D_m D_p\ u(x) = \lambda u(x),\ u(a) = 0,\ D_p\ u(a) = +1.$$

The only nontrivial statement to be proven is that u is a solution to the homogeneous equation (4.1.8). But $D_m D_p u_n(x) = u_{n-1}(x)$ implies $D_m D_p(\sum_{n=0}^{N} \lambda^n u_n(x)) = \lambda \sum_{n=0}^{N-1} \lambda^n u_n(x)$; also $\sum_{n=0}^{N} \lambda^n u_n(a) = 1$, $D_p(\sum_{n=0}^{N} \lambda^n u_n(a)) = 0$ and this together with (4.1.7)(i) implies $\sum_{n=0}^{N} \lambda^n u_n(x) = 1+\lambda\ L(\sum_{n=0}^{N-} \lambda^n u_n(x))$.

Passing to the limit we conclude u solves the equivalent integral equation $u(x) = 1+\lambda\ Lu(x)$. qed For future reference we note that

$$(4.1.11)\quad \begin{cases} \lim_{x\to r_i} |D_p u(x)| < \infty \text{ if } r_i \text{ is a regular boundary} \\ \lim_{x\to r_i} |D_p u(x)| = +\infty \text{ if } r_i \text{ is exit.} \end{cases}$$

This follows at once from the representation $D_p u(x) = \lambda \int_a^x u(y)dm(y)$ and the fact that at a regular boundary $u(r_i)$ is finite and $|m(r_i)| < \infty$, whereas at an exit boundary $|m(r_i)| = +\infty$.

Definition 4.1.2: Define the non-negative functions $w_i(x)$, i= ,2, by

$$(4.1.12)\quad \begin{cases} w_1(x) = w_1(x,\lambda) = u(x,\lambda) \int_{r_0}^{x} u\ (y,\lambda)^{-2} dp(y) \\ w_2(x) = w_2(x,\lambda) = u(x,\lambda) \int_{x}^{r_1} u\ (y,\lambda)^{-2} dp(y). \end{cases}$$

Lemma 4.1.2. w_1 is monotone increasing and continuous on $[r_0,r_1)$, w_2 is monotone decreasing and continuous on $(r_0,r_1]$; $(\lambda-D_m D_p)w_i(x) = 0$ and the Wronskian $W(x) = D_p w_1(x)\ w_2(x) - w_1(x) D_p w_2(x) = W > 0$, where W is a constant.

Proof: We first show $w_i(x)$ are well defined. Consider, for example, w_2: from (4.1.10) we obtain the inequality (for x > a)

$$w_2(x) \le u(x) \int_x^{r_1} (1+\lambda u_1(y))^{-2} dp(y) \le u(x)m(x)^{-1} \int_x^{r_1} (1+\lambda u_1(y))^{-2} m(y)dp(y)$$

$$= u(x)m(x)^{-1} \int_x^{r_1} (1+\lambda u_1(y))^{-2}\ du_1(y); \text{ thus}$$

$$(4.1.13)\quad w_2(x) \le u(x)m(x)^{-1}\lambda^{-1}\{(1+\lambda u_1(x_1))^{-1} - (1+\lambda u_1(r_1))^{-1}\}$$ - so w_2 is well defined

as claimed. The proof for $w_1(x)$ is similar and omitted. We now show that $D_p w_2(x) - D_p w_2(a) = \lambda \int_a^x w_2(y)dm(y)$, from which it follows at once that $D_m D_p w_2 = \lambda w_2$. By direct calculation we obtain

$$D_p w_2(x) - D_p w_2(a) = D_p u(x) \int_x^{r_1} u(y)^{-2} dp(y) + u(a)^{-1} - u(x)^{-1}.$$

on the other hand

$$\lambda \int_a^x w_2(y)dm(y) = \int_a^x \lambda u(y)\{\int_y^{r_1} u(s)^{-2} dp(s)\}\, dm(y)$$

$$= \int_a^x D_m D_p u(y)\{\int_y^{r_1} u(s)^{-2} dp(y)\}\, dm(y).$$

Since $d(D_p u(y)) = D_m D_p\, u(y)dm(y)$ the last integral can be transformed into $\int_a^x \int_y^{r_1} u(s)^{-2} dp(s) dD_p u(y)$ which, after an integration by parts, becomes

$$D_p u(x) \int_x^{r_1} u(s)^{-2} dp(y) + \int_a^x D_p u(y)\; u(y)^{-2} dp(y) =$$

$$D_p u(x) \int_x^{r_1} u(s)^{-2} dp(s) + \int_a^x u(y)^{-2} du(y) =$$

$$D_p u(x) \int_x^{r_1} u(s)^{-2} dp(s) + u(a)^{-1} - u(x)^{-1};\ \text{thus}$$

$D_m D_p w_2(x) = \lambda w_2(x)$. Our next task is to show that $w_2(x)$ is monotone decreasing. Now $w_2(x) > 0$ and $D_m D_p w_2(x) = \lambda w_2(x)$ together imply $D_p w_2(x)$ is monotone increasing. Thus

$$D_p w_2(x) = D_p u(x) \int_x^{r_1} u(y)^{-2} dp(y) - u(x)^{-1}$$

$$\leq \int_x^{r_1} D_p u(y)\; u(y)^{-2} dp(y) - u(x)^{-1};\ \text{hence}$$

(4.1.14) $$D_p w_2(x) \leq \int_x^{r_1} u(y)^{-2} du(y) - u(x)^{-1} = -u(r_1)^{-1} \leq 0.$$

In a similar fashion we can show that $w_1(x)$ and $D_p w_1(x)$ are well defined, monotone increasing on $[r_0, r_1)$ and $D_m D_p w_1 = \lambda w_1$. The proof of the lemma is now completed by showing that the Wronskian $W(x)$ is constant. From $W(x)-W(y) = \int_y^x dW(s)$ we infer

$$\begin{aligned} W(x)-W(y) &= \int_y^x w_2(s)d(D_p w_1(s)) + \int_y^x D_p w_1(s)dw_2(s) \\ &- \int_y^x w_1(s)d(D_p w_2(s)) - \int_y^x D_p w_2(s)dw_1(s) = \\ &= \lambda\int_y^x w_2(s)w_1(s)dm(s) + \int_y^x D_p w_1(s)D_p w_2(s)dp(s) \\ &- \lambda\int_y^x w_1(s)w_2(s)dm(s) - \int_y^x D_p w_2(s)D_p w_1(s)dp(s) = 0; \end{aligned}$$

hence $W(x) = W$ as claimed. The verification that $W > 0$ is left to the reader. We have the expected consequences that $w_1(x), w_2(x)$ form a fundamental set of solutions to the homogeneous equation (4.1.8).

Lemma 4.1.3: Suppose $D_m D_p w(x) = \lambda w(x)$; then there exist constants c_1, c_2 such that $w(x) = c_1 w_1(x) + c_2 w_2(x)$.

Proof: The proof is standard - see Mandl [21]. Note that $W(u,w_1) \neq 0$, $W(u,w_2) \neq 0$; thus (u,w_1) and (u,w_2) also form a fundamental set of solutions for the homogeneous equation.

Corollary: If r_i is an accessible boundary then every solution w of the homogeneous equation has a finite limit at r_i.

Proof: Say r_1 is accessible. It follows from inequality (4.1.10) that $u(x)$ possesses a finite limit at r_1; and obviously so does $w_2(x)$. But

$w(x) = c_1 u(x) + c_2 w_2(x)$. q.e.d.

To construct the general solution to $(\lambda - D_m D_p)F_\lambda = f$ we need only find a particular integral to the inhomogeneous equation which we shall denote by $g_\lambda(x)$; thus

$F_\lambda(x) = g_\lambda(x) + c_1 w_1(x) + c_2 w_2(x)$ where

$(\lambda - D_m D_p) g_\lambda(x) = f(x)$. We define a kernel

$R(\lambda, x, y)$ by

$$(4.1.15) \quad R(\lambda,x,y) = \begin{cases} W^{-1} w_1(x) w_2(y), & r_0 < x \le y < r_1 \\ W^{-1} w_1(y) w_2(x), & r_0 < y \le x < r_1. \end{cases}$$

Theorem 4.1.1: $g_\lambda(x) = \int_{r_0}^{r} R(\lambda,x,y) f(y) dm(y)$ is a particular solution of the inhomogeneous equation $(\lambda - D_m D_p) g_\lambda(x) = f(x)$. Assume $f \in C[r_0, r_1]$.

(i) If r_i is accessible then $\lim_{X \to r_i} g_\lambda(x) = 0$.

(ii) If r_i is natural then $\lim_{X \to r_i} g_\lambda(x) = \lambda^{-1} f(r_i)$.

In general $f(x) \ge 0$ implies $g_\lambda(x) \ge 0$ and in all cases $|g_\lambda| \le \lambda^{-1}|f|$.

Proof: We use repeatedly the representation

$$(4.1.16) \quad W g_\lambda(x) = w_2(x) \int_{r_0}^{x} w_1(y) f(y) dm(g) + w_1(x) \int_{x}^{r_1} w_2(y) f(y) dm(y).$$

Recall that $D_p w_1(x)$ is of bounded variation on $[r_0, x]$ while $D_p w_2(x)$ is of bounded variation on $[x, r_1]$. In particular $w_1(y) dm(y) = \lambda^{-1} d(D_p w_1(y))$ and $w_2(y) dm(y) = \lambda^{-1} d(D_p w_2(y))$.

From this and the representation

$$W g_\lambda(x) = \lambda^{-1} w_2(x) \int_{r_0}^{x} f(y) d(D_p w_1(y) +$$

$$\lambda^{-1} w_1(x) \int_{x}^{r_1} f(y) d(D_p w_2(y))$$ we see at once that g_λ is well defined.

A routine calculation shows that

$$Wdg_\lambda(x) = \left(\int_{r_0}^{x} w_1(y)f(y)dm(y)\right)dw_2(x) + \left(\int_{x}^{r_1} w_2(y)f(y)dm(y)\right)dw_1(x)$$

and therefore

$$W(g_\lambda(x'')-g\ (x')) = W\int_{x'}^{x''} dg\ (x) =$$

$$\int_{x'}^{x''}\left\{\int_{r_0}^{x} w_1(y)f(y)dm(y)\right\}dw_2(x) + \int_{x'}^{x''}\left\{\int_{x}^{r_1} w_2(y)f(y)dm(y)\right\}dw_1(x).$$

Since $dw_i(x) = D_p w_i(x)dp(x)$ we can transform these integrals into

$$\int_{x'}^{x''}\left\{\int_{r_0}^{x} w_1(y)f(y)dm(y)\right\} D_p w_2(y)dp(x) + \int_{x'}^{x''}\left\{\int_{x}^{r_1} w_2(y)f(y)dm(y)\right\}D_p w_1(x)dp(x).$$

Hence

$$WD_p g_\lambda(x) = D_p w_2(x)\int_{r_0}^{x} w_1(y)f(y)dm(y) + D_p w_1(x)\int_{x}^{r_1} w_2(y)f(y)dm(y),$$

from which we infer, without too much difficulty, that

$$WD_m D_p g_\lambda(x) = \lambda(-Wf + Wg_\lambda) \text{ or } (\lambda - D_m D_p)g_\lambda(x) = f(x).$$

We now turn our attention to the boundary behavior of $g_\lambda(x)$ as $x \to r_i$. We shall carry out the analysis for the boundary point r_1 as the analysis at r_0 is similar.

Lemma 4.1.4. $w_2(x) \geq \lambda \int_{x}^{r_1}\int_{y}^{r_1} w_2(s)dm(s)dp(y)$.

Proof: $D_p w_2(r_1) - D_p w_2(y) = \lambda\int_{g}^{r_1} w_2(s)dm(s)$. But $D_p w_2(r_1) \leq 0$ implies $-D_p w_2(y) \geq \lambda\int_{y}^{r_1} w_2(s)dm(s)$ and integrating both sides of this inequality with respect to $dp(y)$ yields $w_2(x) \geq w_2(x) - w_2(r_1) \geq \lambda\int_{x}^{r_1}\int_{y}^{r_1} w_2(s)dm(s)dp(y)$.

Corollary: If r_1 is natural, exit or regular then $\lim_{x\to r_i} w_2(x) = 0$. Similarly, if r_0 is natural, exit or regular then $\lim_{x\to r_0} w_1(x) = 0$.

Proof: If r_1 is exit or natural then $\lim_{x\to r_1} v_1(x) = +\infty$; thus,

$$\lim_{x\to r_1} \int_a^x p(y)\,dm(y) = \lim_{x\to r_1} \{m(x)p(x) - \int_a^x m(y)\,dp(y)\}$$
$$= \lim_{x\to r_1} \int_a^x [m(x)-m(y)]\,dp(y)$$
$$= \int_a^{r_1} [m(r_1)-m(y)]\,dp(y) = +\infty.$$

On the other hand if $\lim_{x\to r_1} w_2(x) = w_2(r_1) = \delta > 0$ then lemma 4.1.4 implies $w_2(x) \geq \lambda\delta \int_x^{r_1} [m(r_1)-m(y)]\,dp(y)$, i.e. $w_2(x) = +\infty$, which is absurd. So $w_2(r_1) = 0$. At a regular boundary $\lim_{x\to r_1} \int_x^{r_1} u(y)^{-2}dp(y) = 0$, because $p(r_1)$ is finite. Therefore $\lim_{x\to r_1} w_2(x) = \lim_{x\to r_1} u(x) \int_x^{r_1} u(y)^{-2}dp(y) = 0$, since $\lim_{x\to r_1} u(x) = u(r_1)$ is finite.

Lemma 4.1.5. If r_1 is inaccessible then $\lim_{x\to r_1} D_p w_2(x) = 0$; similarly if r_0 is inaccessible then $\lim_{x\to r_0} D_p w_1(x) = 0$.

Proof: $D_p w_2(x) = D_p u(x) \int_x^{r_1} u(y)^{-2}dp(y) - u(x)^{-1} \geq -u(x)^{-1}$, $x > a$. Therefore $0 \leq -D_p w_2(x) \leq u(x)^{-1}$. Since $\lim_{x\to r_1} u(x) = +\infty$ when r_1 is inaccessible the assertion of the lemma follows at once.

Lemma 4.1.6. If r_1 is accessible then $\lim_{x\to r_1} w_2(x) D_p w_1(x) = 0$. Similarly if r_0 is accessible then $\lim_{x\to r_0} w_1(x) D_p w_2(x) = 0$.

Proof:

$$w_2(x)D_pw_1(x) = \{u(x)\int_x^{r_1} u(y)^{-2}dp(y)\}\{u(x)^{-1} + D_pu(x)\int_{r_0}^{x} u(y)^{-2}dp(y)\}$$

$$= \int_x^{r_1} u(y)^{-2}dp(y) + u(x)D_pu(x)\int_x^{r_1} u(y)^{-2}dp(y)\int_{r_0}^{x} u(y)^{-2}dp(y).$$

At an accessible boundary $u(r_1) < \infty$ and it is always the case that

$$\lim_{x\to r_1}\int_x^{r_1}u(y)^{-2}dp(y) = 0 \text{ , } \lim_{x\to r_1}\int_{r_0}^{x} u(y)^{-2}dp(y) \text{ is finite.}$$

So all we have to do is show that $\lim_{x\to r} D_pu(x) \int_x u(y)^{-2}dp(y) = 0$. But

$$D_pu(x) = \lambda\int_a^x u(y)dm(y) \le m(x)u(x) \text{ and}$$

$$\int_x^{r_1} u(y)^{-2}dp(y) \le \lambda^{-1}m(x)^{-1}\{(1+\lambda u_1(x_1))^{-1}-(1+ u_1(r_1)^{-1}\}$$ (see the proof of

lemma 4.1.2.). Thus

$$0 \le D_pu(x)\int_x^{r_1} u(y)^{-2}dp(y) \le u(x)\{(1+ u_1(x))^{-1}-(1+ u_1(r_1))^{-1}\}.$$

Now $\lim_{x\to r_1}\{(1+\lambda u_1(x))^{-1}-(1+\lambda u_1(r_1))^{-1}\} = 0$ and this completes the proof. We are

now ready to prove theorem 4.1.1.

Step 1: Assume $f(x) \equiv 1$. Then

$$wg_\lambda(x) = w_2(x)\int_{r_0}^{x} w_1(y)dm(y) + w_1(x)\int_x^{r_1} w_2(y)dm(y)$$

$$= w_2(x)\lambda^{-1}\int_{r_0}^{x} D_mD_pw_1(y)dm(y) +$$

$$w_1(x)\lambda^{-1}\int_x^{r_1} D_mD_pw_2(y)dm(y)$$

$$= \lambda^{-1}\{w_2(x)(D_pw_1(x)-D_p(w_1(r_0)) + w_1(x)(D_pw_2(r_1)-D_pw_2(x))\} = \lambda^{-1}\{W-H(x)\}$$

where $H(x) = w_2(x)D_pw_1(r_0)-w_1(x)D_pw_2(r_1)$.

Let us assume r_1 is inaccessible, in which case $D_p w_2(r_1) = 0$, so $H(x) = w_2(x)D_p w_1(r_0)$. If r_1 is natural then $\lim_{x\to r_1} w_2(x) = w_2(r_1) = 0$ and so $\lim_{x\to r_1} g_\lambda(x) = \lambda^{-1}$. If, on the other hand, r_1 is entrance then $\lim_{x\to r_1} w_2(x) \geq 0$ and from this we infer $\lim_{x\to r_1} g_\lambda(x) \leq \lambda^{-1}$. Suppose now r_1 is an accessible boundary. Then $W=\lim_{x\to r_1} (w_2(x)D_p w_1(x)-w_1(x)D_p w_2(x)) = \lim_{x\to r_1} (w_1(x)D_p w_2(x)$ (by lemma 4.1.6) $= -w_1(r_1)D_p w_2(r_1)$. In addition, if r_1 is accessible, we have $\lim_{x\to r_1} w_2(x) = 0$ (by the corollary to Lemma 4.1.4); thus $\lim_{x\to r_1} H(x) = -w_1(r_1)D_p w_2(r_1)$. Therefore when r_1 is accessible $\lim_{x\to r_1} (W-H(x)) = 0$ and so $\lim_{x\to r_1} g_\lambda(x) = 0$. The analogous assertions for the boundary r_0 are obvious and left to the reader. Putting all these facts together we obtain the following result, which we state as a lemma.

Lemma 4.1.7. If $f(x) \equiv 1$ then $g_\lambda(x) = \int_{r_0}^{r_1} R(\lambda,x,y)dm(y)$ satisfies the equation $(\lambda-D_m D_p)g_\lambda = 1$ and the a-priori estimate $0 \leq g_\lambda(x) \leq \lambda^{-1}$, i.e. $\lambda|g_\lambda| \leq 1$. More generally, if $f \in C[r_0,r_1]$ then $g_\lambda(x) = \int_{r_0}^{r_1} R(\lambda,x,y)dm(y)$ satisfies the equation $(\lambda-D_m D_p)g_\lambda = f$ and the a-priori estimate $\lambda|g_\lambda| \leq |f|$.

Proof. It suffices to prove the lemma for the case $f \equiv 1$, because $|g_\lambda| \leq |f| \int_{r_0}^{r_1} R(\lambda,x,y)dm(y) \leq |f|\ \lambda^{-1}$. Clearly $g_\lambda \in C[r_0,r_1]$ and $g_\lambda \geq 0$ on $[r_0,r_1]$.

Let x' denote a point at which $g_\lambda(x') = |g_\lambda|$.

Case 1: $r_0 < x' < r_1$. Then $D_m D_p g_\lambda(x') \leq 0$ implies $\lambda|g_\lambda| \leq \lambda g_\lambda(x')-D_m D_p g_\lambda(x') = 1$, so $\lambda|g_\lambda| \leq 1$.

Case 2: $x' = r_1$ or r_0. Assume $x' = r_1$. If r_1 is accessible $\lim_{x\to r_1} g_\lambda(x) = 0$ implies $|g_\lambda| = 0$ which is impossible. In this case then the maximum of g_λ cannot occur at an accessible boundary. Suppose r_1 is natural or entrance. In the first case $\lim_{x\to r_1} g_\lambda(x) = \lambda^{-1}$ and in the second case $\lim_{x\to r_1} g_\lambda(x) \le \lambda^{-1}$. Thus in all cases $|g_\lambda| = \lim_{x\to r_1} g_\lambda(x) \le \lambda^{-1}$. This completes the proof of the lemma. Now return to the proof of Theorem 4.1.1. Every function $f \in C[r_0,r_1]$ can be written as a sum $f = f_1(x) + f_2(x)$ where $f_1(x) \equiv f(r_1)$ and $f_2(x)$ vanishes at r_1; in particular we can even assume f_2 vanishes in a neighborhood of r_1 and pass to the general case in the usual way. Note that

$g_\lambda(x) = \hat{g}_1(x) + \hat{g}_2(x)$ where

$\hat{g}_i(x) = \int_{r_0}^{r_1} R(\lambda,x,y)f_i(y)dm(y)$. Clearly,

$\hat{g}_1(x) \in C[r_0,r_1]$, so all we have to do is show that $\hat{g}_2(x) \in C[r_0,r_1]$. But $W\hat{g}_2(x) = w_2(x) \int_{r_0}^{r'} w_1(y)f_2(y)dm(y)$ where $f_2(x) = 0$ for $r' \le x \le r_1$. But $\lim_{x\to r_1} w_2(x)$ always exists, hence $\lim_{x\to r_1} g_2(x)$ exists also. The proof of Theorem 4.1.1 is now complete.

Definition 4.1.3: $D(D_mD_p) = \{f: f \in C[r_0,r_1], D_mD_pf \in C[r_0,r_1]\}$.

Lemma 4.1.8: The set $D(D_mD_p)$ is dense in $C[r_0,r_1]$.

Notation: If $p(x) = x$ we set $D_p = D_x$.

Proof:

Step 1 - A change of variable: Since $p(x)$ is continuous and strictly increasing it has an inverse p^{-1} which is also continuous and strictly increasing. Set $g(y) = f(p^{-1}(y))$ and $n(y) = m(p^{-1}(y))$. Then it is easily checked that $f \in D(D_m D_p)$ if and only if $g \in D(D_n D_y)$ and $D_mD_pf(x) =$

$D_n\, D_y g(p^{-1}(y))$. It therefore suffices to establish the lemma for operators of the form $D_m\, D_x$.

Step 2: We recall that bounded $C^\infty[r_0,r_1]$ functions are dense in $C[r_0,r_1]$. So given any $f \in C[r_0,r_1]$ and any $\varepsilon > 0$ we can find a function $j \in C^\infty[r_0,r_1]$ such that $|f-j| < \varepsilon$. From the representation $j(x) = j(a) + \int_a^x j'(y)dy$ we infer that $j' \in C^\infty[r_0,r_1] \cap L_1[r_0,r_1]$.

Let $\varphi(x) = 1$ if $\alpha \leq x \leq \beta$ and $\varphi(x) = 0$ otherwise, i.e., $\varphi(x)$ is a step function with support the interval $[\alpha,\beta]\subset(r_0,r_1)$. Let $\varphi_i(x) \in C[r_0,r_1]$, i=1,2 with the properties $\int_{r_0}^{r_1} \varphi_i(x)\,dm(x) = 1$, $\varphi_i(x) \geq 0$ and $\varphi_0(x) = 0$ for $x \notin [\alpha-\delta,\alpha]$, $\varphi_1(x) = 0$ for $x \notin [\beta,\beta+\delta]$. Set $\psi(x) = \int_{r_0}^{x} [\varphi_2(y)-\varphi_1(y)]dm(y)$; clearly $0 \leq \psi(x) \leq 1$, $\psi(x) = 1$ for $\alpha \leq x \leq \beta$ and $\psi(x) = 0$ for $r_0 \leq x \leq \alpha-\delta$ and $\beta+\delta \leq x \leq r_1$. Moreover $\int_{r_0}^{r_1} |\psi(y)-\varphi(y)|\,dy < 2\delta$; thus, we've shown that linear combinations of step functions may be approximated in the L_1 sense by functions of the form $\psi(x) = \int_{r_0}^{x} \ell(y)dm(y)$ where $\ell \in C[r_0,r_1]$. Hence, arbitrary functions in $C[r_0,r_1] \cap L_1[r_0,r_1]$ can be approximated in L_1 by functions of the form $\psi(x) = \int_{r_0}^{x} \ell(y)\,dm(y)$. Given $\varepsilon > 0$ and $j' \in C[r_0,r_1] \cap L_1[r_0,r_1]$ pick ψ such that

$$\int_{r_0}^{r_1} |j'(y)-\psi(y)|dy < \varepsilon.$$

Set $h(x) = j(a) + \int_a^x \int_{r_0}^{y} \ell(s)dm(s)dy$ and note that $|h(x)-j(x)| \leq \int_a^x |\psi(y)-j'(y)|dy < \varepsilon$.

Thus $|f-h| \leq |f-j| + |j-h| < 2\varepsilon$. It is obvious that $h \in C[r_0,r_1]$ and $D_m\, D_x\, h \in C[r_0,r_1]$-the proof of lemma is now complete.

2. Markov processes satisfying boundary conditions

The general solution to the stationary equation (4.1.1) is of the form $F_\lambda(x) = g_\lambda(x) + c_1 w_1(x) + c_2 w_2(x)$. If r_i is inaccessible then $w_i(x)$ is unbounded and hence if both r_0 and r_1 are inaccessible the only bounded solution of (4.1.1) is (by Theorem 4.1.1) $F_\lambda(x) = g_\lambda(x) = \int_{r_0}^{r_1} R(\lambda,x,y)f(y)\,dm(y)$.

From this and lemma 4.1.7 we arrive at the following result:

Theorem 4.2.1. Suppose r_0, r_1 are both inaccessible boundaries. Then there exists a unique Markovian semi-group $T(t)$: $C[r_0,r_1] \to C[r_0,r_1]$ with infinitesimal generator

$$Gf(x) = (a(x)/2)f''(x) + b(x)f'(x) \text{ and domain}$$

$$\mathscr{D}(G) = \{f:\ f \in C[r_0,r_1],\ Gf \in C[r_0,r_1]\}$$

In addition $T(t)1 = 1$-so the corresponding Markov process $x(t)$ is conservative, i.e., $P(t,x,[r_0,r_1]) = 1,\ t \geq 0$.

Proof: All we need to do is check the hypotheses of the Hille-Yosida theorem (Theorem 3.2.1). That $\mathscr{D}(G)$ is dense in a consequence of lemma 4.1.8. In addition, to every $f \in C[r_0,r_1]$ there exists a unique $F_\lambda \in \mathscr{D}(G)$ given by $F_\lambda(x) = \int_{r_0}^{r_1} R(\lambda,x,y)f(y)\,dm(y)$ satisfying the equation $(\lambda - G)F_\lambda = f$ and the estimate $\lambda|F_\lambda| \leq |f|$ - this is a consequence of lemma 4.1.7. Obviously $f \geq 0$ implies $F_\lambda \geq 0$. So $\overline{G|\mathscr{D}(G)}$ generates a Markovian semi-group $T(t)$. We note that $(\lambda-G)^{-1}\,1 = \lambda^{-1} = \int_0^\infty e^{-\lambda t}\,P(t,x,dy)$; from the uniqueness of the Laplace transform we infer $P(t,x,dy) = 1,\ t \geq 0$.

Examples of processes to which the above theorem applies:

(i) Brownian motion on $(-\infty,\infty)$

(ii) Ornstein-Uhlenbeck process on $(-\infty,\infty)$

(iii) Radial component of n-dimensional Brownian motion on $[0,\infty)$; here, 0 is an entrance boundary; ∞, a natural boundary.

More interesting possibilites occur when at least one of the boundaries is accessible, say r_0. To simplify matters let us assume r_0 is a regular boundary and r_1 is inaccessible. Then the most general bounded solution to equation (4.1.1) is given by $F_\lambda(x) = g_\lambda(x) + c\, w_2(x)$, because $w_1(x)$ is unbounded. To specify c we must impose a boundary condition at r_0.

Definition 4.2.1: For r_0 a regular boundary, r_1 inaccessible, set

$$\mathcal{D}_1(G) = \{f: \ f \in C[r_0,r_1],\ Gf \in C[r_0,r_1],\ D_p f(r_0) = 0\}$$

$$\mathcal{D}_2(G) = \{f: \ f \in C[r_0,r_1],\ Gf \in C[r_0,r_1],\ Gf(r_0) = 0\}$$

$$\mathcal{D}_3(G) = \{f: \ f \in C[r_0,r_1],\ Gf \in C[r_0,r_1],\ f(r_0) = 0,\ Gf(r_0) = 0\}.$$

Theorem 4.2.2.

(a) $\overline{G|\mathcal{D}_1(G)}$ generates a Markovian semi-group $T_1(t)$: $C[r_0,r_1] \to C[r_0,r_1]$. The corresponding Markov process is conservative; this is the process with reflection at r_0.

(b) $\overline{G|\mathcal{D}_2(G)}$ generates a Markovian semi-group $T_2(t)$: $C[r_0,r_1] \to C[r_0,r_1]$. The corresponding process is conservative; this is the process with an adhesive boundary at r_0.

(c) Let $\hat{C}[r_0,r_1] = \{f: \ f \in C[r_0,r_1],\ f(r_0) = 0\}$. $\overline{G|\mathcal{D}_3(G)}$ generates a Markovian semi-group $T_3(t)$: $\hat{C}[r_0,r_1] \to \hat{C}[r_0,r_1]$. The corresponding Markov process is non-conservative; this is the process with an absorbing boundary at r_0.

Proof: (a) Set $F_\lambda(x) = g_\lambda(x) + c\, w_2(x)$ where c is so chosen that $D_p F_\lambda(r_0) = D_p g_\lambda(r_0) + c\, D_p w_2(r_0) = 0$. Of course, one must show that $D_p w_2(r_0)$ and $D_p g_\lambda(r_0)$ are well defined; a routine calculation, which we omit, (but see (4.1.11) where a similar calculation is carried out) yields the explicit

$$\text{value } c = \frac{D_p w_1(r_0) \int_{r_0}^{r_1} w_2(y) f(y)\, dm(y)}{-W D_p w_2(r_0)} .$$

Thus $f \geq 0$ implies $c \geq 0$, since $-D_p w_2(r_0) > 0$, and in particular $F_\lambda \geq 0$. Set $F_\lambda = R_1(\lambda)f$. All we need do now is establish the inequality $\lambda|F_\lambda| \leq |f|$. Now $h = R_1(\lambda)1-\lambda^{-1}$ is a bounded solution to the homogeneous equation satisfying the boundary condition $D_p(r_0) = 0$. But $D_p w_2(r_0) \neq 0$ implies $c = 0$ and hence $R_1(\lambda)1 = \lambda^{-1}$. This together with the fact that $R_1(\lambda)$ is a positive operator implies $|R_1(\lambda)f| \leq |f|R_1(\lambda)1 = \lambda^{-1}|f|$. It is obvious that $\mathscr{D}_1(G)$ is dense in $C[r_0,r_1]$. Putting these pieces together in the usual way we get that $R_1(\lambda)$ is the resolvent of a Markovian semi-group $T_1(t)$. Clearly $1 \in \mathscr{D}_1(G)$, $G1 = 0$ and $T_1(t)1=1$, so the corresponding Markov process is conservative.

(b) This time we choose c so that

$$G\,F_\lambda(r_0) = G\,g_\lambda(r_0) + c\,G\,w_2(r_0) = 0.$$

But $Gg_\lambda(r_0) = \lambda g_\lambda(r_0)-f(r_0)$. However, if r_0 is a regular boundary $g_\lambda(r_0) = 0$ (Theorem 4.1.1) and so $GF_\lambda(r_0) = -f(r_0) + c\,Gw_2(r_0) = -f(r_0) + c\lambda w_2(r_0)=0$. So we must choose $c = \lambda^{-1} f(r_0)/w_2(r_0)$; clearly $f \geq 0$ implies $c \geq 0$ and thus $F_\lambda = R_2(\lambda)f$ is again a positive operator. Moreover using the same reasoning as in part (a) we infer $R_2(\lambda)1 = \lambda^{-1}$ and hence $\lambda|F_\lambda| \leq |f|$. Finally it is a triviality to check that $\mathscr{D}_2(G)$ is dense in $C[r_0,r_1]$.

(c) Although $\mathscr{D}_3(G)$ is not dense in $C[r_0,r_1]$ it is dense in $\hat{C}[r_0,r_1]$; also note that $1 \notin \hat{C}[r_0,r_1]$. From theorem (4.1.1) we see that $F_\lambda(x) = \int_{r_0}^{r_1} R(\lambda,x,y)f(y)dm(y)$ satisfies the boundary condition $F_\lambda(r_0) = 0$ and the estimate $\lambda|F_\lambda| \leq |f|$. In addition it is easy to see that $R_3(\lambda)1 < \lambda^{-1}$ (where $F_\lambda = R_3(\lambda)f$) and hence

$$\int_0^\infty \exp(-\lambda t)P(t,x,[r_0,r_1])dt = R_3(\lambda)1 < \lambda^{-1},$$

from which it is easily inferred that

$$P(t,x,[r_0,r_1]) < 1, \quad t > 0.$$

This completes our brief introduction to the possible boundary conditions that can be imposed on a one dimensional Markov process. For a more complete discussion we refer the reader to Ito-McKean [5] of supplementary bibliography, Mandl [21] and the references therein to the papers of W. Feller.

3. Feller's generalized second order differential operators

It was observed in IV.1 that every second order linear differential operator $Gf(x) = (a(x)/2)f''(x)+b(x)f'(x)$, $r_0 < x < r_1$, can be put into the so-called "Feller form" $G = D_m D_p$. It is an interesting and extremely useful fact that $D_m D_p$ remains well defined even if we drop the differentiability assumptions on p and the continuity assumptions on m. More precisely we have

Definition 4.3.1. Let p(x) be strictly increasing and continuous on (r_0,r_1) and m(x) be strictly increasing and right continuous on (r_0,r_1). We define, provided the limits exist,

$$(4.3.1)\quad \begin{cases} D_p+ f(x) = \lim_{y\to x+} \dfrac{f(y)-f(x)}{p(y)-p(x)} \\ \\ D_m g(x) = \lim_{y\to x} \dfrac{g(y)-g(x)}{m(y)-m(x)} \end{cases}$$

A generalized second order differential operator is defined by the formula $D_m D_p + f(x)$. We say that $f \in D(D_m D_p +)$ if $f \in C[r_0,r_1]$ and $D_m D_p + f \in C[r_0,r_1]$.

Notation: To simplify the typography we drop the symbol + and continue to write these operators as $D_m D_p$.

Remark: Every definition and theorem of IV, 1 and 2 remains valid for these generalized second order differential operators. In addition we note that if $f \in D(D_m D_p)$ and has an interior maximum at the point x then $D_m D_p f(x) \leq 0$. Thus these generalized operators of Feller, subject to appropriate boundary conditions as given in IV.2 for example, generate Markovian semi-groups. It is a noteworthy fact that under suitable hypotheses on the Markov process x(t) the converse is also true - c.f Dynkin [11], V. II p.143 or Varadhan [32] p. 174. Specifically, suppose x(t) is a one dimensional Markov process with continuous paths and regular, meaning that the probability of going from x to y (and from y to x) in finite time is positive, and suppose the strong

Markov property holds (cf Dynkin V.I [11] p.99); then the infinitesimal generator of the corresponding semi-group is a restriction of an operator of the form $D_m D_p$. In addition to their obvious theoretical importance the generalized operators of Feller play an important role in limit theorems for Markov processes. In the next section we shall exhibit a weakly convergent family of one dimensional Markov processes $x_N(t)$, $1 \leq N < \infty$, whose infinitesimal generators are classical operators of the form $G_N f(x) = (a_N(x)/2)f''(x)+b_N(x)f'(x)$ and yet the limiting process $x(t)$ has as its infinitesimal generator a generalized operater $D_m D_p$ - see also the author's papers [23], [24].

4. Limit theorems for one dimensional Markov processes

Many of the stochastic models that occur in genetics, transport theory, learning theory etc., lead to a family of Markov processes $x_N(t)$ depending on a real parameter N, $1 \leq N < \infty$, with corresponding diffusion and drift coefficients $a_N(x)$ and $b_N(x)$. Under various hypotheses one shows that $\lim_{N\to\infty} a_N(x) = a(x)$, $\lim_{N\to\infty} b_N(x) = b(x)$ and one would like to infer that the Markov processes $x_N(t)$ converge in distribution to the Markov process $x(t)$ whose diffusion and drift coefficients are $a(x)$ and $b(x)$ respectively; this passage to the limit is the so-called "diffusion approximation". Such a theorem is made even more plausible if one looks at it from the point of view of semi-group theory.

Set $G_N f(x) = (a_N(x)/2)f''(x)+b_N(x)f'(x)$ and

$Gf(x) = (a(x)/2)f''(x)+b'(x)f'(x)$. Then

$\lim_{N\to\infty} a_N(x) = a(x)$, $\lim_{N\to\infty} b_N(x) = b(x)$ imply (at least formally) that $\lim_{N\to\infty} G_N f(x) = Gf(x)$ and (hopefully) $\lim_{N\to\infty} \exp(tG_N)f = \exp(tG)f$; from this the weak convergence of the finite dimensional distributions is easily deduced.

Of course our problem now is to convert this heuristic reasoning into rigorous mathematics. Thanks to the Trotter-Kato theorem the justification of the passage to the limit is not particularly difficult, although the proofs are lengthy and admittedly somewhat tedious. There is, however, an unexpected and noteworthy consequence of our methods. We shall exhibit a family of Markov processes $x_N(t)$ which converge weakly to the Markov process $x(t)$ (see Billingsley [2] for the precise meaning of weak convergence) such that $\lim_{N\to\infty} a_N(x) = a(x)$, $\lim_{N\to\infty} b_N(x) = b(x)$, but the infinitesimal generator of the limit process $x(t)$ is a generalized second order differential operator $D_m D_p^+ f(x) \neq (a(x)/2)f''(x) + b(x)f'(x)$. For a more precise statement of this counterexample as well as others we refer the reader to Theorem 4.4.3 and its consequences. This counterexample suggests that it might be more profitable to set $G_N = D_{m_N} D_{p_N}$ and study the asymptotic behavior as $N\to\infty$ of the resolvents $(\lambda - D_{m_N} D_{p_N})^{-1}$ as a function of p_N and m_N instead of studying the asymptotic behavior of G_N directly. One justification that comes to mind is the fact that the resolvents are uniformly bounded (in norm) by λ^{-1} whereas G_N is an unbounded operator. The main result of this section is that if $\lim_{N\to\infty} p_N(x) = p(x)$ and $\lim_{N\to\infty} m_N(x) = m(x)$ at all continuity points of the latter than $\lim_{N\to\infty} (\lambda - D_{m_N} D_{p_N})^{-1} = (\lambda - D_m D_p)^{-1}$. Of course when one or both boundaries are accessible these hypotheses must be supplemented by imposing additional boundary conditions; to simplify matters we will assume that r_0, r_1 are both natural boundaries.

Theorem 4.4.1. Assume r_0, r_1 are both natural boundaries for the operators $D_{m_N} D_{p_N}$, $1 \leq N < \infty$, and $D_m D_p$. Suppose $\lim_{N\to\infty} p_N(x) = p(x)$ and $\lim_{N\to\infty} m_N(x) = m(x)$ at all continuity points of m.

Then

(4.4.1) $\lim_{N\to\infty} |(\lambda - D_{m_N} D_{p_N})^{-1} f - (\lambda - D_m D_p)^{-1} f| = 0$ for all $f \in C[r_0, r_1]$.

Before proceeding to the proof we give an application to a central limit theorem first obtained by Gihman-Skorohod [13] p.152, but see also A. Friedman [12].

Theorem 4.4.2 (Gihman-Skorohod): Let $x(t)$ be a one dimensional Markov process with diffusion coefficient $a(x) \equiv 1$, and drift coefficient satisfying the condition $\int_{-\infty}^{\infty} b(x)dx = 0$ (it is assumed that $b \in C[-\infty,\infty] \cap L_1(-\infty,\infty)$).

Then

(4.4.2) $\lim_{t\to\infty} P_x(\frac{x(t)}{\sqrt{t}} < z) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{z} \exp(-y^2/2)dy.$

Proof (via theorem 4.4.1): Set $x_N(t) = x(N^2t)/N$ and observe that $x_N(1) = x(N^2)/N = x(t)/\sqrt{t}$, if we set $N^2=t$. We will show that the finite dimensional distributions of the $x_N(t)$ process converge to those of the Brownian motion. We recall that the infinitesimal generator of Brownian motion is $Gf = (\frac{1}{2})f''$ and its scale and speed measures are $p(x) = cx$, $m(x) = 2c^{-1}x$ with $c > 0$, but otherwise arbitrary. An easy calculation shows that the infinitesimal generator G_N of $x_N(t)$ is

(4.4.3) $G_N f(x) = (\frac{1}{2})f''(x) + Nb(Nx)f'(x)$ and its corresponding scale and speed measures are given by

$$(4.4.4) \quad \begin{cases} p_N(x) = \int_0^x \exp(-B_N(y))dy \\ m_N(x) = 2\int_0^x \exp(B_N(y))dy \\ B_N(y) = 2\int_0^y Nb(Nz)dz = 2\int_0^{Ny} b(z)dz. \end{cases}$$

Put $c = \exp(2\int_0^\infty b(z)dz)$. We shall establish by means of a very elementary computation that $\lim_{N\to\infty} p_N(x) = cx$, $\lim_{N\to\infty} m_N(x) = 2c^{-1}x$. By theorem 4.4.1 and the Trotter-Kato Theorem we conclude $\lim_{N\to\infty} T_N(t)f(x) = T(t)f(x)$ where $T(t)$ is the Brownian motion semi-group; clearly this proves the Theorem of Gihman-Skorohod. In fact we can prove an even more general result.

Assume

$$(4.4.5)\quad \begin{cases} \int_{-\infty}^{\infty} b(x)dx=\alpha, \text{ where } \alpha \text{ need not equal zero,} \\ c_1 = \exp(-2\int_0^\infty b(z)dz) \text{ and } c_2 = c_1 \exp(2\alpha), \\ \text{so } \alpha \neq 0 \text{ implies } c_1 \neq c_2. \end{cases}$$

Then

$$(4.4.6)\quad \begin{cases} \lim_{N\to\infty} p_N(x) = c_1\, x,\ x \geq 0;\ \lim_{N\to\infty} m_N(x) = 2c_1^{-1}\, x,\ x \geq 0 \\ \lim_{N\to\infty} p_N(x) = c_2\, x,\ x \leq 0;\ \lim_{N\to\infty} m_N(x) = 2c_2^{-1},\ x \leq 0. \end{cases}$$

We now prove (4.4.6)- the special case $\alpha=0$ is the hypothesis of Theorem 4.4.2. From the hypotheses on $b(z)$ we get

$$\int_{-\infty}^{0} b(z)dz + \int_0^\infty b(z)dz = \alpha. \quad \text{Thus}$$

$$(4.4.7)\quad \begin{cases} \lim_{N\to\infty} B_N(y) = 2\int_0^\infty b(z)dz \text{ if } y > 0 \text{ and} \\ \lim_{N\to\infty} B_N(y) = 2\int_0^{-\infty} b(z)dz = 2\int_0^\infty b(z)dz-2\alpha \text{ if } y < 0. \end{cases}$$

Therefore $\lim_{N\to\infty} \exp(-B_N(y)) = c_1$, if $y > 0$ and

$$\lim_{N\to\infty} \exp(-B_N(y)) = c_1 \exp(2\alpha), \text{ if } y < 0;$$

this completes the proof of (4.4.6).

Definition 4.4.1. Define $p_\alpha(x)$, $m_\alpha(x)$ via the formulae

$$(4.4.8)\quad \begin{cases} p_\alpha(x) = cx,\ x \geq 0;\ m_\alpha(x) = 2c^{-1}x,\ x \geq 0 \\ p_\alpha(x) = c'x,\ x \leq 0,\ m_\alpha(x) = 2c'x,\ x \leq 0 \text{ where } c' = c\exp(2\alpha). \end{cases}$$

We note that (i) the boundary points $\pm\infty$ are natural boundaries for the operator $D_{m_\alpha} D_{p_\alpha}$ and

(ii) $p'_\alpha(0)$, $m'_\alpha(0)$ do not exist for $\alpha \neq 0$; in particular $D_{m_\alpha} D_{p_\alpha}$ cannot be put in the form $D_{m_\alpha} D_{p_\alpha} f(x) = (a_\alpha(x)/2)f''(x)+b_\alpha(x)f'(x)$.

Now let $y_\alpha(t)$ denote the Markov process whose corresponding semi-group will be denoted by $\exp(tD_{m_\alpha} D_{p_\alpha}) = S_\alpha(t)$. The calculation (4.4.6) together with Theorem 4.4.1 lead to the following result which is a considerable generalization of Theorem 4.4.2.

Theorem 4.4.3. Let $x(t)$ be a one dimensional Markov process with diffusion coefficient $a(x) = 1$ and drift coefficient $b(x)$ satisfying the condition $\int_{-\infty}^{\infty} b(x)dx = \alpha$. Then the family of stochastic processes $x_N(t) = x(N^2t)/N$, $1 \leq N < \infty$, converges weakly to the Markov process $y_\alpha(t)$.

Remark: We've only established the convergence of the finite dimensional distributions; for the proof of the more general result we refer the reader to the author's paper [23].

Of, perhaps, greater interest than the theorem itself are the following counterexamples which we now proceed to derive.

Examples: (i) Choose $\alpha \neq 0$, let b have compact support and assume $b(0) = 0$. Then $\lim_{N\to\infty} N\, b(Nx) = 0$ all x. We have thus constructed a family of Markov processes $x_N(t)$ converging weakly to a limit $y_\alpha(t) \neq w(t)$ (Brownian motion) and yet $\lim_{N\to\infty} a_N(x) = 1$, $\lim_{N\to\infty} b_N(x) = 0$!

(ii) Choose $\alpha=0$ and assume $|b(0)|\neq 0$. In this case $x_N(t)$ converges weakly to Brownian motion but $\lim_{N\to\infty}|b_N(x)| = +\infty$!

These examples show that the hypotheses of Theorem 4.4.1 cannot be dispensed with. We now proceed to sketch the proof, referring the reader to the author's paper [23] for a more detailed account.

Set $F_\lambda(x) = (\lambda - D_m D_p)^{-1} f(x)$ and

$F_{\lambda,N}(x) = (\lambda - D_{m_N} D_{p_N})^{-1} f(x).$

The boundaries r_0, r_1 are natural so

(4.4.9) $\quad F_\lambda(x) = \int_{r_0}^{r_1} R(\lambda,x,y) f(y) dm(y)$ and

(4.4.10) $\quad F_{\lambda,N}(x) = \int_{r_0}^{r_1} R_N(\lambda,x,y) f(y) dm_N(y)$

where $R(\lambda,x,y)$ and $R_N(\lambda,x,y)$ are defined as in (4.1.15). In addition we set

$u_N(x) = \sum_{n=0}^{\infty} \lambda^n u_{n,N}(x)$ where

$$u_{0,N}(x) \equiv 1, \; u_{n+1,N}(x) = L\, u_{n,N}(x) \text{ and}$$

(4.4.11)
$$w_{1,N}(x) = u_N(x) \int_{r_0}^{x} u_N(y)^{-2}\, dp_N(y)$$
$$w_{2,N}(x) = u_N(x) \int_{x}^{r_1} u_N(y)^{-2}\, dp_N(y).$$

$W_N = W(w_{1,N}, w_{2,N})$, the Wronskian of $w_{1,N}$ and $w_{2,N}$.

Lemma 4.4.1. If $\lim_{N\to\infty} p_N(x) = p(x)$ and $\lim_{N\to\infty} m_N(x) = m(x)$ (at all continuity points of m)

then

(4.4.12) $\lim_{N\to\infty} u_N(x) = u(x)$

(4.4.13) $\lim_{N\to\infty} w_{1,N}(x) = w_1(x)$; $\lim_{N\to\infty} w_{2,N}(x) = w_2(x)$

(4.4.14) $\lim_{N\to\infty} W_N = W$;

the convergence being uniform for x in compact subsets of (r_0,r_1).

Proof: Our first step is to show that for a sufficiently large class of functions

(4.4.15) $\lim_{N\to\infty} \int_a^x f_N(y)dp_N(y) = \int_a^x f(y)dp(y)$ uniformly for x in compact subsets

of (r_0,r_1). Now $\int_a^x f_N(y)dp_N(y) = \int_{p_N(a)}^{p_N(x)} f_N(p_N^{-1}(z))dz$ implies that when

$\lim_{N\to\infty} f_N(y) = f(y)$ in the sense of bounded point wise convergence a.e. for x in compact subsets of (r_0,r_1), then (4.4.15) holds. In particular this implies

$$\lim_{N\to\infty} u_{1,N}(x) = \lim_{N\to\infty} \int_a^x m_N(y)dp_N(y) = \int_a^x m(y)dp(y) = u_1(x)$$

uniformly for x in compact subsets of (r_0,r_1). Similarly

$$\lim_{N\to\infty} \int_a^y u_{n,N}(s)dm_N(s) = \int_a^y u_n(s)dm(s)$$

for all continuity points y of m. Applying (4.4.15) once again we infer

$$\lim_{N\to\infty} \int_a^x \int_a^y u_{n,N}(s)dm_N(s)dp_N(y) = \int_a^x \int_a^y u_n(s)dm(s)dp(y).$$

Thus $\lim_{N\to\infty} u_{n,N}(x) = u_n(x)$ uniformly on compact subsets. By lemma 4.1.1 we have

$u_{n,N}(x) \le \frac{u_{1,N}(x)^n}{n!}$ (with a similar estimate for $u_n(x)$) and from this (4.4.12) is an easy consequence. (4.1.13) is proved in a similar fashion except we now use estimate (4.1.13); the proof of (4.4.14) is also routine - see [23] pp.619-620 for the complete details. We now have at our disposal all the tools we need to prove $\lim_{N\to\infty} |F_{\lambda,N}(x) - F_\lambda(x)| = 0$. Assume f has compact support, say

$[\alpha,\beta]$.

Then $F_{\lambda,N}(x) = W_N^{-1}\, w_{2,N}(x) \int_{\alpha}^{x} w_{1,N}(y) f(y)\, dm_N(y)$
$+ W_N^{-1}\, w_{1,N}(x) \int_{x}^{\beta} w_{2,N} f(y)\, dm_N(y)$. Since r_0, r_1 are natural boundaries we know that $\lim_{x\to r_1} w_{2,N}(x) = 0$, $\lim_{x\to r_0} w_{1,N}(x) = 0$

and therefore in a neighborhood of r_i we can make $|F_{\lambda,N}(x)| < \varepsilon$ uniformly in N; we may take this neighborhood to be the complement of $[\alpha,\beta]$ itself. Now apply lemma 4.4.1 to conclude $\lim_{N\to\infty} F_{\lambda,N}(x) = F_{\lambda}(x)$ uniformly for $\alpha < x < \beta$. Summing up, then, we've proved that if f has compact support then $\lim_{N\to\infty} |F_{\lambda,N} - F_{\lambda}| = 0$. Next assume $f(x) \equiv 1$ and use the representation of step 1 on p.55 to conclude $\lim_{N\to\infty} |F_{\lambda,N} - F| = 0$ in this case too. Finally, every function in $C[r_0, r_1]$ is a uniform limit of a linear combination of functions of the above sort.

Remarks: (i) Theorem 4.4.1 as given here is an improvement of Theorem 2 of [23] in that we need only assume m is right continuous.

(ii) That the hypotheses of theorem 4.4.1 are the most natural was first pointed out by C. Stone [29]. The applicability of the Trotter-Kato theorem to this problem was first noted by author [23]. The counterexamples following theorem 4.4.3 are implicit in [23] and further elaborated in [24].

V. Markov processes and their associated martingales

1. Martingale functions.

Call a function $U(t,x)$ a <u>martingale function</u> for the Markov process $x(t)$ if $U(t,x(t))$ is a martingale; supermartingale and submartingale functions are defined similarly. Such functions have been studied by several authors including Doob [10], Stroock-Varadhan [30], Lai [20].

A classic example is furnished by $U(t,x) = x^2-t$ with $x(t)$ Brownian motion. More generally, assume the Markovian semi-group $T(t)f(x) = E_x f(x(t))$ has G for its infinitesimal generator. Then, provided $U(t,x)$, $a(x)$ and $b(x)$ satisfy certain regularity conditions - the precise nature of which need not concern us here - it is known that

$$(5.1.1) \qquad Z(t) = U(t,x(t)) - \int_0^t \{U_s(s,x(s)) + GU(s,x(s))\}ds$$

is a martingale. This is a consequence of Ito's lemma - see [12].

These regularity conditions exclude many interesting Markov processes e.g. Bessel processes - to be defined below; processes whose infinitesimal generators are nonclassical operaters of the form $D_m D_p$; and processes satisfying boundary conditions. To illustrate the sort of difficulties that can arise suppose $x(t) = |w(t)|$ - the reflecting Brownian motion - and set $U(t,x) = x$. In this case $Gf(x) = (\frac{1}{2})f''(x)$ so $U_t(t,x)+ GU(t,x) = 0$ implies, formally, that $U(t,x(t)) = |w(t)|$ is a martingale; this is obviously false and the question is why? One explanation is that $U(t,x) \notin \mathcal{D}(G)$. Recall that $f \in \mathcal{D}(G)$ implies $f'(0) = 0$, but $U_x(t,x) = 1$ and in particular $U_x(t,0) = 1 \neq 0$. This example illustrates that a function $U(t,x)$ satisfying the differential equation

$$(5.1.2) \qquad U_t(t,x)+ GU(t,x) = 0$$

is not necessarily a martingale function; additional conditions depending on $\mathcal{D}(G)$ must be imposed. One such set of sufficient conditions is presented in

Theorem 5.1.1 below - a theorem which has been shown to be useful in cases where the hypotheses of Ito's lemma are not satisfied - see the author's papers [25], [26].

Theorem 5.1.1. Suppose $U(t,x)$ and $U_t(t,x)$ are both jointly continuous for $(t,x) \in R_+ \times [r_0,r_1]$ and in addition $U(t,x) \in \mathcal{D}(G)$, $U_t(t,x) \in \mathcal{D}(G)$ all $t \geq 0$. Then $Z(t)$, defined at 5.1.1, is a martingale.

Corollary: If, in addition to the hypotheses of theorem (5.1.1),

(a) U satisfies the differential inequality

(5.1.3) $\quad U_t(t,x) + GU(t,x) \leq 0$

then $U(t,x(t)$ is a supermartingale;

(b) If U satisfies the differential equation

(5.1.2) then $U(t,x(t))$ is a martingale.

Let us illustrate the theorem with examples some of which, strictly speaking, do not exactly satisfy the hypotheses of Theorem 5.1.1. The conclusions of the theorem remain valid, however, and we shall indicate later on how to adapt the Theorem to these cases.

Examples:

(1): $\quad U(t,x) = \exp(-\lambda t)R(\lambda)f(x),\ f \geq 0,\ \lambda > 0.$

We have $U_t(t,x) = -\lambda U(t,x)$ and $GU(t,x) =$

$$\exp(-\lambda t)GR(\lambda)f = \exp(-\lambda t)[\lambda R(\lambda)f - f] =$$

$$\lambda U(t,x) - \exp(-\lambda t)f.$$

Hence $U_t + GU = -\exp(-\lambda t)f \leq 0$, so U satisfies the differential inequality (5.1.3). In addition $R(\lambda)f \in \mathcal{D}(G)$ and the hypotheses of theorem 5.1.1 and its corollary are satisfied. Thus $\exp(-\lambda t)R(\lambda)f(x(t))$ is a positive supermartingale.

(2) $x(t)$ is Brownian motion, $Gf(x) = (\frac{1}{2})f''(x)$ and $U(t,x) = \exp(\theta x - \theta^2 t/2)$, θ a real parameter, $U(t,x)$ is a martingale function.

(3) $x(t)$ is the radial component of n-dimensional Brownian motion and $U(t,x) = x^2 - nt$, U is a martingale function.

Remark: The martingale of example (3) leads to a simple proof of $E_0\{\tau_R\} = R^2/n$ where τ_R is the first passage time to the surface of sphere of radius R. Assuming Doob's optional stopping theorem holds in this case (which it does) one sees immediately $E_0(x^2(\tau_R) - n\tau_R) = 0$; but $x^2(\tau_R) = R^2$, q.e.d. We turn now to the proof of theorem 5.1.1.

Lemma 5.1.1. If $f \in \mathscr{D}(G)$ then $y(t) = f(x(t)) - \int_0^t Gf(x(u))du$ is a martingale.

(This is obviously a special case of the theorem).

Proof: From the Markov property and lemma 3.2.2 we obtain

$$E\{f(x(t)) | \mathscr{B}(s)\} = T(t-s)f(x(s)) =$$

$$f(x(s)) + \int_0^{t-s} T(u)Gf(x(s))du$$

But $T(u)Gf(x(s)) = E\{Gf(x(s+u)) | \mathscr{B}(s)\}$ and therefore,

$$E\{f(x(t)) - f(x(s)) | \mathscr{B}(s)\} = E\{\int_0^{t-s} Gf(x(s+u))du | \mathscr{B}(s)\} = \tag{5.1.4}$$

$$= E\{\int_0^t Gf(x(u))du | \mathscr{B}(s)\}.$$

On the other hand

$$y(t) = y(s) + f(x(t)) - f(x(s)) - \int_s^t Gf(x(u))du;$$

Clearly (5.1.4) now implies $E\{y(t) | \mathscr{B}(s)\} = y(s)$; the proof is complete.

Proof of Theorem 5.1.1.

Step 1: The following decomposition holds for $t_1 < t_2$:

$$E\{U(t_2,x(t_2))-U(t_1,x(t_1))|\mathscr{B}(t_1)\} = I + II + III,$$

where

$$I = E\{\int_{t_1}^{t_2} [U_s(s,x(s)) + GU(x,x(s))]ds|\mathscr{B}(t_1)\}$$

$$II = E\{\int_{t_1}^{t_2} [U_s(s,x(t_2))-U_s(s,x(s))]ds|\mathscr{B}(t_1)\}$$

$$III = E\{\int_{t_1}^{t_2} [GU(t_1,x(s))-GU(s,x(s))]ds|\mathscr{B}(t_1)\}.$$

This decomposition will be derived below. Assuming, for the moment, its validity we come to

Step 2: $II + III = 0$. Deferring the proof of this as well we conclude

$E\{U(t_2,x(t_2))-U(t_1,x(t_1))|\mathscr{B}(t_1) = I$ where I can be rewritten as

$$I = E\{\int_{0}^{t_2} [U_s(s,x(s)) + GU(s,x(s))]ds$$

$$- \int_{0}^{t_1} [U_s(s,x(s)) + GU(s,x(s))]ds|\mathscr{B}(t_1)\}.$$

But this is clearly equivalent to the statement $E\{Z(t_2)|\mathscr{B}(t_1)\} = Z(t_1)$. We now proceed to prove the assertions of steps 1 and 2. From lemma 5.1.1. and the hypothesis $U(t,x) \in \mathscr{D}(G)$ we get

$$(5.1.5) \quad E\{U(t_1,x(t_2))-\int_{0}^{t_2} GU(t_1,x(s))ds|\mathscr{B}(t_1)\} =$$

$$U(t_1,x(t_1))-\int_{0}^{t_1} GU(t_1,x(s))ds.$$

Next observe - taking into account the obvious cancellations - that

$$I + II + III = E\{\int_{t_1}^{t_2} U_s(s,x(t_2))ds|\mathscr{B}(t_1)\}$$

$$+ E\{\int_{t_1}^{t_2} GU(t_1,x(s))ds|\mathscr{B}(t_1)\}.$$

But

$$(5.1.6)\qquad E\{\int_{t_1}^{t_2} U_s(s,x(t_2))ds|\mathscr{B}(t_1)\} =$$

$$E\{U(t_2,x(t_2))-U(t_1,x(t_2))|\mathscr{B}(t_1)\},$$ and from (5.1.5)

$$(5.1.7)\qquad E\{\int_{t_1}^{t_2} GU(t_1,x(s))ds|\mathscr{B}(t_1)\} =$$

$$E\{U(t_1,x(t_2))-U(t_1,x(t_1))|\mathscr{B}(t_1)\}.$$

Adding (5.1.6) and (5.1.7) yields the decomposition of step 1.

Proof of step 2. We first show that

$$(5.1.8)\qquad II = E\{\int_{t_1}^{t_2}\int_{s}^{t_2} GU_s(s,x(u))duds|\mathscr{B}(t_1)\} =$$

$$= \int_{t_1}^{t_2} E\{\int_{s}^{t_2} GU_s(s,x(u))duds|\mathscr{B}(t_1)\}$$

$$= \int_{t_1}^{t_2} E\{U_s(s,x(t_2))-U_s(s,x(s))|\mathscr{B}(t_1)\}ds$$

$$= E\{\int_{t_1}^{t_2} [U_s(s,x(t_2))-U_s(s,x(s))]ds|\mathscr{B}(t_1)\}.$$

On the other hand interchanging the order of integration in (5.1.8) yields

$$II = E\{\int_{t_1}^{t_2}\int_{t_1}^{u} GU_s(s,x(u))dsdu|\mathscr{B}(t_1)\} =$$

$$= E\{\int_{t_1}^{t_2}\int_{t_1}^{u} \frac{\partial}{\partial s} GU(s,x(u))dsdu|\mathscr{B}(t_1)\} =$$

$$= E\{\int_{t_1}^{t_2} [GU(u,x(u))-GU(t_1,x(u))]du|\mathscr{B}(t_1)\} = -III,$$ q.e.d.

2. A theorem of Doob and its generalization.

Let $x(t)$ be a one dimensional Markov process with state space $[r_0,r_1]$, associated semi-group $T(t)f(x) = E_x\{f(x(t))\}$ and infinitesimal generator $Gf(x) = (a(x)/2)f''(x) + b(x)f'(x)$. Using the stochastic differential calculus of Ito and the notion of semi-group of type Γ, Doob [10] proved the following theorem (under slightly different hypotheses).

Theorem 5.2.1 (Doob): Assume r_0,r_1 are natural boundaries; $a(x),b(x)$ are both bounded and Lipschitz continuous on compact subintervals of (r_0,r_1).

(i) If $Gf(x) = cf(x)$ then $f(x(t))-ct$ is a martingale

(ii) If $Gg(x) = \lambda g(x)$ then $\exp(-\lambda t)g(x(t))$ is a martingale

Example: $x(t)$ is Brownian motion, $f(x) = x^2$ then $Gf \equiv 1$ so $x^2(t)-t$ is a martingale. If $g(x) = \exp(\theta x)$ then $Gg(x) = (\theta^2/2)g(x)$ so (ii) holds with $\lambda = \theta^2/2$, c.f. example 2 after statement of Theorem 5.1.1. Note once again that neither f nor g is bounded and so theorem 5.1.1. as it stands cannot be directly applied. Nevertheless, by means of a suitable argument to be sketched below, we can derive the following generalization of Doob's theorem.

Theorem 5.2.2. Assume $G = D_m D_p$ with r_0,r_1 natural boundaries. Then

(i) $Gf(x) = cf(x)$ implies $f(x(t))-ct$ is a martingale

(ii) $Gf(x) = \lambda g(x)$ implies $\exp(-\lambda t)g(x(t))$ is a martingale.

Remark: The methods by which these results are obtained can easily be extended to the case where r_i are entrance or even regular provided the functions f and g satisfy the appropriate boundary conditions - see the author's papers [25], [26] where such calculations are carried out.

Proof: If $f \in C[r_0,r_1]$, $g \in C[r_0,r_1]$ then there is no problem, because both f and g belong to $\mathcal{D}(G)$ and we may apply Theorem 5.1.1 directly. Unfortunately, we have only the result that $f \in C(r_0,r_1)$, $g \in C(r_0,r_1)$.

This is a standard technical difficulty and is circumvented by means of the following device: truncate the function g (or f) in such a way that its truncation $\hat{g}(x) \in \mathscr{D}(G)$ and $\hat{g}(x) = g(x)$ on a compact subinterval of (r_0,r_1). If G is a classical second order operator this can be accomplished by setting $g(x) = g(x)\varphi(x)$ where $\varphi(x)$ is a $C^\infty[r_0,r_1]$ function such that $\varphi(x) = 1$, $x \in I$ and φ vanishes at the boundaries. If G is a generalized operator then a similar truncation can be carried out - see lemma 1, p.273 of [25]. Theorem 5.1.1 can now be applied to the function $U(t,x) = \exp(-\lambda t)\hat{g}(x)$; a routine application of Doob's optional stopping theorem ([4] p.373 and "note added in proof" p.379) now yields the useful result that $U(t\wedge\tau, x(t\wedge\tau))$ is a martingale where τ denotes the first exit time of the x(t) process from the interval I[(1)]. In particular then we have

(5.2.1) $$U(0,x) = E_x\{\exp(-\lambda(t\wedge\tau))g(x(\tau\wedge\tau)\}.$$

We would of course like to pass to the limit by choosing an increasing sequence of compact subintervals I_n, $\cup I_n = (r_0,r_1)$ and a corresponding sequence of stopping times $\tau_n \uparrow \infty$ and thereby conclude

(5.2.2) $$U(0,x) = E_x\{\exp(-\lambda t)g(x(t))\}.$$

More generally we would like to prove

(5.2.3) $$E_x\{U(t,x(t-s))\} = U(s,x), \quad 0 \leq s < t$$

for then

(5.2.4) $$E\{U(t,x(t))|x(s)\} = E_{x(s)}\{U(t,x(t-s))\} = U(s,x(s)).$$

In many cases the reasoning used to establish (5.2.2) may also be applied to derive (5.2.3). In the present instance set $V(t',x) = U(s+t',x)$, $0 \leq t' < \infty$ and check that V satisfies the differential equation (5.1.2). Thus

---

(1) $t\wedge\tau = \min(t,\tau)$

$$(5.2.5)\quad U(s,x) = V(0,x) = E_x\{V(t\wedge\tau_n,\ x(t\wedge\tau_n)\}$$
$$= E_x\{U(s+\ t\wedge\tau_n,\ x(t\wedge\tau_n))\}$$

Now let $\tau_n \uparrow \infty$ as before to infer $U(s,x) = E_x\{U(s+t,x(t))\}$ which is equivalent to (5.2.3). We shall not carry out the justification of the passage to the limit, referring the reader instead to an article by T. L. Lai [20] to whom Theorem 5.2.2 is due. The (simpler) proof of (5.2.1) is due to the author - see [25].

3. L_p estimates for the stopping times of Brownian motion and Bessel processes; theorems of Burkholder-Gundy and Burkholder.

Let τ denote a bounded stopping time for the Markov process $x(t)$. A number of authors including Skorokhood [28], Rosenkrantz [22], Burkholder-Gundy [6], Burkholder [4] Rosenkrantz-Sawyer [26] have derived inequalities of the following kind:

$$(5.3.1)\quad a(p)E_x(\tau^p) \leq E_x(|x(\tau)|^{2p}) \leq A(p)E_x(\tau^p),$$

where $a(p)$ and $A(p)$ are independent of τ and $1 \leq p < \infty$, and $x = 0$. If $x(t)$ is Brownian motion then (5.3.1) plays an important role in the so-called Skorokhod embedding [28], Chapter 7 . In a related development B. Davis [8] has actually characterized the best possible constants in the case of Brownian motion. It is not our purpose to enter into such delicate calculations but merely to indicate in a purely formal manner how estimates of the kind displayed at (5.3.1) are easily derived used the ideas of V.1 and V.2.

Example 1: $x(t)$ is Brownian motion with $r_0 = -\infty$, $r_1 = +\infty$. Set $U(t,x) = t^p - C(p)x^2t^{p-1}$ where $1 < p < \infty$ and $C(p)$ remains to be chosen. A routine calculation shows that the choice $C(p) = p$ implies $U(t,x)$ satisfies the differential inequality (5.1.3 a) and hence $U(t,x(t))$ is a supermartingale. Thus for any bounded stopping time τ we have

$$\text{(5.3.2)} \qquad 0 \geq E_0\{U(\tau,x(\tau))\} = E_0\{\tau^p - px^2(\tau)\tau^{p-1}\}$$

Thus
$$E_0\{\tau^p\} \leq p\, E_0\{x^2(\tau)\ \tau^{p-1}\}$$
$$\leq p\, E_0\{x^{2p}(\tau)\}^{\frac{1}{p}}\, E_0\{\tau^p\}^{(p-1)/p}$$

where we use Holder's inequality at the last step; dividing both sides by $E_0\{\tau^p\}^{(p-1)/p}$

(5.3.3) yields $E_0\{\tau^p\} \leq (p)^p\, E_0\{x^{2p}(\tau)\}$, so $a(p) = p^{-p}$ in this case.

To get the inequality in the other direction pick $U(t,x) = x^{2p} - c(p)tx^{2p-2}$ with $2 \leq p$. The choice $c(p) = p(2p-1)$ makes $U(t,x)$ a supermartingale function and hence

$$E_0\{x^{2p}(\tau) - p(2p-1)x^{2p-2}(\tau)\cdot\tau\} \leq 0 \text{ or}$$

$$\text{(5.3.4)} \qquad E_0\{x^{2p}(\tau)\} \leq p(2p-1)E_0\{x^{2p-2}(\tau)\cdot\tau\}.$$

The proof is now completed by using Holder's inequality again with exponents $\alpha = p-1$ and $\beta = {}^p/p-1$. Here $A(p) = (p(2p-1))^p$.

Remark: These arguments are due to L. Gordon [14] who made use of Ito's lemma instead of theorem 5.1.1. The proofs can be extended into the range $0 < p < 1$, but at the cost of some additional technicalities which we choose to avoid here.

Recently, Burkholder [5], has extended the inequalities (5.3.1) to the radial component of n-dimensional Brownian motion. By a different and considerably simpler method Rosenkrantz and Sawyer [26] have extended Burkholder's results to the so-called Bessel processes of order $\gamma+1$. More precisely, if we set $Gf(x) = (\frac{1}{2})f''(x) + (\gamma/x)f'(x)$ where $r_0 = 0 \leq x < \infty = r_1$ and $-(\frac{1}{2}) < \gamma$ and define

$$\mathscr{D}(G) = \{f:\ f \in C_0(R_+) \cap C_0^2(R_+),\ Gf \in C_0(R_+), f'(0) = 0\}$$

then $\overline{G|\mathcal{D}(G)}$ generates a Markovian semi-group whose corresponding Markov process we call the Bessel process of order $\gamma+1$ - see [3] for the details. It is noteworthy that the functions $t^p - C(p)x^2t^{p-1}$, $x^{2p} - c(p)tx^{p-2}$ are also supermartingale functions for the Bessel process of order $\gamma+1$. However one must choose $C(p) = p/(2\gamma+1)$, $1 < p$ and $c(p) = 2p\gamma + p(2p-1)$, $2 \leq p$. In the latter case $1 < p < 2$ requires a separate argument - see [26] for the details. Thanks to Theorem (5.5.1) the arguments leading up to (5.3.3) and (5.3.4) can be repeated word for word to yield inequalities (5.3.1). The case $\gamma=(n-1)/2$ corresponds to the radial component of n-dimensional Brownian motion and so includes Burkholder's results as a special case.

BIBLIOGRAPHY

[1] C. Bardos, Problems aux limites pour les équations aux dérivées partielles du premiere ordre a coefficients reel; Théorèmes d'approximation Application a l'equation de transport, Ann. Scient. Ecole Normale, Sup 4e serie, t.3, 185-233 (1970).

[2] P. Billingsley, Convergence of Probability measures, Wiley, New York, 1968.

[3] H. Brezis, W. Rosenkrantz and B. Singer, with an appendix by Peter D. Lax, On a degenerate elliptic-parabolic equation occurring in the theory of probability, Comm. Pure and Applied Math. Vol.XXIV, 395-416, 1971.

[4] D. Burkholder, Distribution functions inequalities for martingales, Ann. of Prob. Vol.1, 19-42, 1973.

[5] D. Burkholder, Exit times of Brownian motion, harmonic majorization and Hardy Spaces, Advances in Mathematics (to appear).

[6] D. Burkholder and R. F. Gundy, Extrapolation and interpolation of quasi-linear operators on martingales, Acta Math. Vol.124, 249-304, 1970.

[7] P. Chernoff, Note on product formulas for operator semi-groups, Journ. of Functional Analysis, V.2, 238-242, 1968.

[8] B. Davis, On the L_p norms of stochastic integrals and other martingales, Duke Math. Journ. Vol.43, No.4, 697-704, December, 1976.

[9] J. L. Doob, Stochastic Processes, Wiley, New York, 1953.

[10] J. L. Doob, Martingales and one dimensional diffusion, Trans. Amer. Math. Soc. Vol.78, No.1, 168-208, January, 1955.

[11] E. B. Dynkin, Markov processes, Vol.1 and 2, Academic Press, New York, 1965.

[12] A. Friedman, Stochastic differential equations and applications, V.1, Academic Press, New York, 1975.

[13] I. I. Gihman and A. V. Skorokhod, Stochastic differential equations, Springer Verlag, 1972.

[14] L. Gordon, An equivalent to the martingale square function inequality, Ann. Math. Stat. 43, 1927-1934, 1972.

[15] E. Hille and R. S. Phillips, Functional analysis and semi-groups, Amer. Math. Soc. Coll. Publications Vol.XXXI, Published by Amer. Math. Soc. Providence, R.I., 1957.

[16] K. Jorgens, An asymptotic expansion in the theory of neutron transport, Comm. Pure and Applied Math. Vol.11, 219-242, 1958.

[17] T. Kato, Perturbation theory for linear operators, Springer-Verlag, 1966.

[18] T. Kurtz, Extensions of Trotter's operator semi-group approximation theorems, Journ. of Functional Analysis, Vol.3, No.3, 354-375, June, 1969.

[19] T. Kurtz, A limit theorem for perturbed operator semi-groups with application to random evolutions, Journ. of Functional analysis 12, 55-67, 1973.

[20] T. L. Lai, Space-time processes, parabolic functions and one dimensional diffusions, Trans. Amer. Math. Soc. V.175, 409-438, 1973.

[21] P. Mandl,, Analytical treatment of one dimensional Markov processes, Springer-Verlag, 1968.

[22] W. Rosenkrantz, On rates of convergence for the invariance principle, Trans. Amer. Math. Soc., Vol.129, 542-552, 1967.

[23] W. Rosenkrantz, Limit theorems for solutions to a class of stochastic differential equations, Indiana University Mathematics Journal, Vol. 24, No.7, 613-625, 1975.

[24] W. Rosenkrantz, A convergent family of diffusion processes whose diffusion coefficients diverge Bull. of Amer. Math. Soc. Vol.80, No.5, 973-976, September, 1974.

[25] W. Rosenkrantz, Martingales and diffusion processes satisfying boundary conditions, Proceedings of the summer research institute on statistical inference for stochastic processes, Vol.1, 267-282, Academic Press, 1975.

[26] W. Rosenkrantz and S. Sawyer, L_p estimates for stopping times of Bessel processes, Zeit. fur Wahr. (to appear).

[27] W. Rudin, Real and complex analysis, McGraw-Hill, 1966.

[28] A. V. Skorokhod, Studies in the theory of random processes (translated from the Russian), Addison-Wesley, Reading, Mass., 1965.

[29] C. Stone, Limit theorems for birth and death processes and diffusion processes, Ph.D. thesis, Dept. of Mathematics, Stanford University, 1961.

[30] D. W. Stroock and S. R. S. Varadhan, Diffusion processes with continuous coefficients, I and II, Comm. Pure and Applied Math. Vol.XXll, 345-400, 479-530, 1969.

[31] H. Trotter, Approximation of semi-groups of operators, Pac. J. Math., V.8, 887-919, 1958.

[32] S. R. S. Varadhan, Stochastic Processes, Lecture notes, New York University, Courant Institute of Mathematical Sciences, 1968.

Supplementary Bibliography. Readers interested in additional applications of semi-group methods to the so-called "diffusion approximation" as well as other complements of a theoretical character may consult the papers and treatises listed below.

[1] C. C. Y. Dorea, Differentiability preserving properties of a class of semi-groups, Zeit. fur Wahr and verw. Gebiete, v.36, 13-26, 1976.

[2] S. Ethier, A class of degenerate diffusion processes occurring in population genetics, Comm. Pure and Appl. Math. Vol. XXIX, 483-493, 1976.

[3] R. Ellis and W. Rosenkrantz, Diffusion Approximation for transport processes satisfying boundary conditions, Indiana University Mathematics Journal (to appear).

[4] H. A. Guess, On the weak convergence of Wright-Fisher models, Stochastic Processes Appl. Vol.1, 287-306, 1973.

[5] K. Ito and H. P. McKean Jr., Diffusion processes and their sample paths, 2nd printing corrected, Springer-Verlag, 1974.

[6] M. F. Norman, Markov processes and learning models, Academic Press, New York, 1972.

[7] G. Papanicolau, Asymptotic Analysis of transport processes, Bull. of Amer. Math. Soc. Vol.81, No.2, 330-392, March 1975.