1 Some Mathematical Ideas and their Applications

1.1 An Idea from Analysis

Spectral theory is one of most fruitful ideas of analysis and has applications to probability theory. We do not intend to give a sophisticated treatment of the spectral theorem, rather we loosely explain the general idea and apply it to a random walk problems. Even understanding the broad outlines of the theory is a useful mathematical tool.

Let A be an $n \times n$ matrix with eigenvalues $\lambda_1, \cdots, \lambda_n$ counted with multiplicity, and assume A is diagonalizable. The diagonalization of the matrix A can be accurately interpreted as taking the underlying vector space to be the space $L(S)$ of functions (generally complex-valued) on a set $S = \{s_1, \cdots, s_n\}$ of cardinality n and the matrix A to be the operator of multiplication of an element $\psi \in L(S)$ by the function φ_A which has value λ_j at s_j. In this manner we have obtained the simplest form that a matrix can reasonably be expected to have. Of course not every matrix is diagonalizable but large classes of matrices including an open dense subset of them are. The idea of spectral theory is to try to do the same, to the extent possible, for infinite matrices or linear operators on infinite dimensional spaces. Even when diagonalization is possible for infinite matrices, several distinct possibilities present themselves which we now describe:

1. (*Pure Point Spectrum*) There is a countably infinite set $S = \{s_1, s_2, \cdots\}$ and a complex valued function ϕ_A defined on S such that the action of the matrix A is given by multiplication by φ_A. The underlying vector space is the vector space of square summable functions on S, i.e., if we set $\psi_k = \psi(s_k)$, then $\sum_k |\psi_k|^2 < \infty$. It often becomes necessary to take a weighted sum in the sense that there is a positive weight function $\frac{1}{c_k}$ and the undelying vector space is the space of sequences $\{\psi_k\}$ such that

$$\sum_k \frac{1}{c_k} |\psi_k|^2 < \infty.$$

The weight function $\frac{1}{c_k}$ is sometimes called *Plancherel measure.*

2. (*Absolutely Continuous Spectrum*) There is an interval (a, b) (closed or open, a and/or b possibly ∞), a function φ_A such that A is the operator of multiplication of functions on (a, b) by φ_A. Often the operator A acts on an infinite dimensional vector space $\mathcal{H}$ where there is a positive definite inner product $< .,. >$ is defined. In such a situation we can take a basis $e_1, e_2, \ldots$ for $\mathcal{H}$ such that $< e_i, e_j >= 0$ or 1 according as $i \neq j$ or $i = j$. The matrix A of a linear operator may be given relative to the basis $e_1, e_2, \ldots$. Then there are functions $\varphi_1, \varphi_2, \ldots$ corresponding to the basis $e_1, e_2, \ldots$ and a positive or non-negative function $\frac{1}{c(\lambda)}$ (called the *Plancherel measure*) such that the underlying vector space is the space of functions ψ on (a, b) with the property

$$\int_a^b |\psi(\lambda)|^2 \frac{d\lambda}{c(\lambda)} < \infty.$$

The functions φ_j satisfy

$$\int_a^b \varphi_j(\lambda)\overline{\varphi_k(\lambda)} \frac{d\lambda}{c(\lambda)} = \delta_{jk},$$

where δ_{jk} is 0 or 1 according as $j \neq k$ or $j = k$.

3. (*Singular Continuous Spectrum*) There is an uncountable set S of Lebesgue measure zero (such as the Cantor set) such that A can be realized as multiplication by a function φ_A on S. The underlying vector space is again the space of square integrable functions on S relative to some measure on S. One often hopes that the problem does not lead to this case.

4. None of the above cases covers the case of a matrix of the form $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. One can also consider an infinite dimensional analogue of it where $\mathcal{H}$ is an infinite dimensional space with basis $e_1, e_2, \ldots$ and $T : \mathcal{H} \rightarrow \mathcal{H}$ is the operator defined by

$$T(e_1) = 0, \quad T(e_{n+1}) = e_n \quad \text{for} \quad n \geq 1.$$

Of course this operator is not diagonalizable. In the broad outline of the theory that we describing we assume that for some reason we know that the operators in question are in fact diagonalizable. Operators given by symmetric or Hermitian matrices are always diagonalizable. We will dwell on this point.

5. It is possible for an operator to be a combination of the above cases. The cases of greatest interest are those when a problem can be reduced to cases 1 or 2. Singular spectrum occurs naturally in connection with fractals as stationary distributions associated to certain Markov processes.

The first two cases and their combination is of greatest interest to us. In order to demonstrate the idea we look at some familiar examples and demonstrate the diagonalization process.

Example 1.1.1 Let A be the differentiation operator $\frac{d}{dx}$ on the space of periodic functions with period 2π. Since

$$\frac{d}{dx}e^{inx} = ine^{inx},$$

in is an eigenvalue of $\frac{d}{dx}$. Writing a periodic function as a Fourier series

$$\psi(x) = \sum_{n\in\mathbf{Z}} a_n e^{inx},$$

we see that the appropriate set S is $S = \mathbf{Z}$ and

$$\varphi_{\frac{d}{dx}}(n) = in.$$

The Plancherel measure is $\frac{1}{c_k} = \frac{1}{2\pi}$ and from the basic theory of Fourier series (Parseval's theorem) we know that

$$\int_{-\pi}^{\pi} \psi \frac{d}{dx}\bar{\phi}dx = -\frac{i}{2\pi}\sum_{n\in\mathbf{Z}} na_n\bar{b}_n,$$

where $\psi(x) = \sum_{n\in\mathbf{Z}} a_n e^{inx}$ and $\phi(x) = \sum_{n\in\mathbf{Z}} b_n e^{inx}$. ♠

Example 1.1.2 Let f be a periodic function of period 2π and A_f be the operator of convolution with f, i.e.,

$$A_f : \psi \longrightarrow \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x-y)\psi(y)dy$$

Assume $f(x) = \sum_n f_n e^{inx}$ and $\psi(x) = \sum_n a_n e^{inx}$. Substituting the Fourier series for f and ψ in the defintiion of A_f we get

$$A_f(\psi) = \frac{1}{2\pi} \sum_{n,m} e^{imx} \int_{-\pi}^{\pi} f_n a_m e^{i(n-m)y} dy = \sum_m f_m a_m e^{imx}.$$

This means that in the diagonalization of the operator A_f, the set S is $\mathbf{Z}$, and the function φ_{A_f} is

$$\varphi_{A_f}(n) = f_n.$$

The underlying vector space and Plancherel measure is the same as in example 1.1.1. Thus Fourier series transforms convolutions into multiplication of Fourier coefficients. The fact that Fourier series simultaneously diagonalizes convolutions and differentiation reflects the fact that convolutions and differentiation commute and commuting diagonalizable matrices can be simultaneously diagonalized. Convolution operators occur frequently in many areas of mathematics and engineering. ♠

Example 1.1.3 Examples 1.1.1 and 1.1.2 for functions on $\mathbf{R}$ or $\mathbf{R}^n$ when the periodicity assumption is removed. For a function ψ of *compact support* (i.e., ψ vanishes outside a closed interval $[a, b]$) Fourier transform is defined by

$$\tilde{\psi}(\lambda) = \int_{\circ}^{\infty} e^{-i\lambda x} \psi(x) dx.$$

Integration by parts shows that under Fourier transform the operator of differentiation $\frac{d}{dx}$ becomes multiplication by $-i\lambda$:

$$\widetilde{\left(\frac{d\psi}{dx}\right)}(\lambda) = (-i\lambda)\tilde{\psi}(\lambda).$$

Similarly let A_f denotes the operator of convolution by an integrable function f:

$$A_f(\psi) = \int_{-\infty}^{\infty} f(x-y)\psi(y) dy.$$

Then a change of variable shows

$$\int_{-\infty}^{\infty} e^{-i\lambda x} \int_{-\infty}^{\infty} f(x-y)\psi(y) dy dx = \tilde{f}(\lambda)\tilde{\psi}(\lambda).$$

Therefore the diagonalization process of the differentiation and convolution operators on functions on $\mathbf{R}$ leads to case (2) with $S = \mathbf{R}$, $\frac{d}{dx} \leftrightarrow -i\lambda$, and $A_f \leftrightarrow \tilde{f}$. For the underlying vector space it is convenient to start with the space of compactly supported functions on $\mathbf{R}$ and then try to extend the operators to $L^2(S)$. There are technical points which need clarification, but for the time being we are going to ignore them. ♠

Example 1.1.4 Let us apply the above considerations to the simple random walk on $\mathbf{Z}$. The random walk is described by convolution with the function f on $\mathbf{Z}$ defined by

$$f(n) = \begin{cases} p, & \text{if } n = 1; \\ q, & \text{if } n = -1; \\ 0, & \text{otherwise.} \end{cases}$$

Convolution on $\mathbf{Z}$ is defined similar to the cases on $\mathbf{R}$ except that the integral is replaced by a sum:

$$A_f(\psi) = f \star \psi(n) = \sum_{k \in \mathbf{Z}} f(n-k)\psi(k).$$

Let e_j be the function on $\mathbf{Z}$ defined by $e_j(n) = \delta_{jn}$ where δ_{jn} is 1 if $j = n$ and 0 otherwise. It is straightforward to see that the matrix of the operator A_f relative to the basis e_j for the vector space of functions on $\mathbf{Z}$ is the matrix of transition probabilities for the simple random walk on $\mathbf{Z}$. Example 1.1.2 suggests that this situation is dual to one described in that example. For S we take the interval $[-\pi, \pi]$, to state j corresponds the periodic function e^{ijx} and the action of the transition matrix P is given by multiplication by the function

$$pe^{ix} + qe^{-ix} = (p+q)\cos x + i(p-q)\sin x.$$

The probability of being in state 0 at time $2l$ is the therefore the constant term in the Fourier expansion of $(pe^{ix} + qe^{-ix})^{2l}$, viz.,

$$\binom{2l}{l} p^l q^l,$$

which we had easily established before. ♠

Example 1.1.5 A more interesting example is the application of the idea of the spectral theorem to to the reflecting random walk on $\mathbf{Z}_+$ where the point 0 is a reflecting barrier. The matrix of transition probabilities is given by

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & & & & \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & \cdots & & & & \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & \cdots \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Assume the diagonalization can be implemented so that P becomes multiplication by the function $\varphi_P(x) = x$ on the space of functions on an interval which we take to be $[-1, 1]$. If the state n corresponds to the function ϕ_n then we must require

$$x\phi_\circ = \phi_1, \quad x\phi_n(x) = \frac{1}{2}\phi_{n-1}(x) + \frac{1}{2}\phi_{n+1}(x).$$

Using the elementary trigonometric identity

$$\cos\alpha\cos\beta = \frac{1}{2}\cos(\alpha - \beta) + \frac{1}{2}\cos(\alpha + \beta)$$

we obtain the following functions ϕ_n:

$$\phi_\circ(x) = 1, \quad \phi_n(x) = \cos n\theta, \quad \text{where} \quad \theta = \cos^{-1} x. \tag{1.1.1}$$

The polynomials $\phi_n(\theta) = \cos(n\cos^{-1}\theta)$ are generally called *Chebycheff polynomials*. The above discussion should serve as a good motivation for the introduction of these polynomials which found a number of applications. For the Plancherel measure we seek a function $\frac{1}{c(\theta}$ such that

$$\int_{-1}^{1} \phi_n(\theta)\phi_m(\theta)\frac{d\theta}{c(\theta)} = \delta_{mn}.$$

In terms of $\theta \in [0, \pi]$ and the orthogonality relations

$$\int_0^\pi \cos m\theta \cos n\theta d\theta = \frac{\pi}{2}\delta_{mn}.$$

To obtain the Plancherel measure we express $d\theta$ as a function of x using the relations

$$\theta = \tan^{-1}\frac{y}{x}, \qquad y = \sqrt{1 - x^2}.$$

We obtain

$$\frac{1}{c(\theta)} = \frac{1}{\pi\sqrt{1-\theta^2}}.$$

The coefficients of the matrix P^l can be computed in terms of the Chebycheff polynomials. In fact it is straightforward to see that

$$P_{jk}^{(l)} = \int_{-1}^{1} x^l \phi_j(\theta)\phi_k(\theta)\frac{d\theta}{c(\theta)}.$$

The idea of this simple example can be extended to considerably more complex Markov chains and the machinery of orthogonal polynomials can be used to this effect. ♠

Fourier transform or series, and in particular the fact that they convert convolutions to products, have many applications. Here we discuss an application to random walks. Although the following presentation, strictly speaking, can be made independent of Fourier transforms, yet in spirit we are making use of harmonic analysis.

Let δ_a be the delta function at $a \in \mathbf{R}$. δ_a is not a function but its Fourier transform is $e^{-i\lambda a}$. Therefore although δ_a is not a function, convolution of a function ϕ with δ_a is meaningfully defined as the inverse Fourier transform of $e^{-i\lambda a}\hat{\phi}(\lambda)$. One way of thinking about it is to take a sequence of Gaussians distributions ϕ_n with mean a and standard deviation $\frac{1}{n}$ and take the limit $n \rightarrow \infty$. Of particular interest is $\delta_\circ$ whose Fourier transform is the function which is identically 1. Therefore $\delta_\circ \star \phi = \phi$ or $\delta_\circ$ is identity relative to convolutions. The introduction of δ-function allows us to introduce an exponential function relative to the convolution operation. More precisely, we define the exponential of a function ϕ as the inverse Fourier transform of

$$1 + \hat{\phi} + \frac{\hat{\phi}^2}{2!} + \frac{\hat{\phi}^3}{3!} + \ldots$$

Although the inverse Fourier transform of the above expression is not a function, by subtracting $\delta_\circ$ from it we get a function. At any rate, we denote the exponential of ϕ, thus defined, by ε^ϕ or $\mathcal{E}(\phi)$. Similarly we define the log function as the inverse Fourier transform of

$$\hat{\phi} + \frac{\hat{\phi}^2}{2} + \frac{\hat{\phi}^3}{3} + \ldots$$

and denote it by $-\mathcal{L}(\delta_\circ - \phi)$. To make sure that no convergence problem occurs and the logarithm is meaningfully defined, we assume that

$$\int_{-\infty}^{\infty} |\phi(x)| dx < 1. \tag{1.1.2}$$

This assumption is made whenever we make use of $\mathcal{L}$ even if not explicitly stated. It is readily verified that $\mathcal{L}$ and $\mathcal{E}$ are inverses to each other in the sense that

$$\mathcal{E}\Big(\mathcal{L}(\delta_\circ - \phi)\Big) = \delta_\circ - \phi, \quad \mathcal{L}(\mathcal{E}(-\psi)) = -\psi. \tag{1.1.3}$$

Of course in the second identity we write $\mathcal{E}(-\psi)$ as $\delta_\circ - (\delta_\circ - \mathcal{E}(-\psi))$ to make the definition of $\mathcal{L}$ applicable. For a function ϕ on $\mathbf{R}$ we define

$$\phi_+(x) = \begin{cases} \phi(x), & \text{if } x \geq 0; \\ 0, & \text{otherwise.} \end{cases}$$

From the fact that if ϕ_+ and ψ_+ vanish for $x < 0$, then $\phi_+ \star \psi_+$ vanishes for $x < 0$

Following Dym and McKean-*Fourier Series and Integrals*, we introduce lemma 1.1.1 below which expresses the key mathematical fact for application to random walks:

Lemma 1.1.1 *(*(Spitzer's Identity*) With the above notation and assumption (1.1.2), the Fourier transform of* $\phi_+ + (\phi \star \phi_+)_+ + (\phi \star (\phi \star \phi_+)_+)_+ + \dots$ *is equal to*

$$-1 + \exp\Big[\sum_{n=1}^{\infty} \frac{1}{n} \int_\circ^{\infty} e^{-i\lambda x} (\phi \star \dots \star \phi)(x) dx\Big].$$

Equivalently it is the Fourier transform of

$$\mathcal{E}[-(\mathcal{L}(\delta_\circ - \phi))_+] - \delta_\circ.$$

One can more simply state the the Spitzer identity as

$$\delta_\circ + \phi_+ + (\phi \star \phi_+)_+ + (\phi \star (\phi \star \phi_+)_+)_+ + \dots = \mathcal{E}[-(\mathcal{L}(\delta_\circ - \phi))_+],$$

but the form given in the lemma is more useful for our purpose. Before giving the proof of lemma 1.1.1, we discuss an implication of it.

Let ϕ be a density function on $\mathbf{R}$ and assume that ϕ is symmetric, i.e., $f(x) = f(-x)$. Let $X_1, X_2, \ldots$ be a sequence of iid random variables with density ϕ, and set $S_\circ = 0, S_l = S_{l-1} + X_l$. Then $S_\circ, S_{\cdot}, S_2, \ldots$ is a random walk on $\mathbf{R}$ and the probability $P[X_1 \geq 0, S_2 \geq 0, \ldots, S_n \geq 0, a \leq S_l \leq b]$ is given by

$$\int_a^b (\phi_+ \star (\phi_+ \star \ldots (\phi \star \phi_+)_+ \ldots)_+ dx = \int \ldots \int \phi(x_1) \ldots \phi(x_l) dx_1 \ldots dx_l,$$

where the domain of the multiple integral is

$$\{x_1 \geq 0, x_1 + x_2 \geq 0, \ldots, x_1 + \ldots + x_l \geq 0, a \leq x_1 + \ldots + x_L \leq b\}.$$

For $\epsilon > 0$ sufficiently small, the Spitzer identity is applicable to the function $\epsilon\phi$. Noting that the integral of a function is equal to its Fourier transform at 0, we obtain

$$\begin{aligned}\sum_{l=1}^{\infty} \epsilon^l P[S_1 \geq 0, \ldots, S_l \geq 0] &= -1 + \exp\left[\sum_{l=1}^{\infty} \frac{\epsilon^l}{l} \int_\circ^\infty \phi^{\star l}\right] \\ &= -1 + \exp\left[\frac{\epsilon^l}{l} P[S_l \geq 0]\right].\end{aligned}$$

Notice that the right hand side is much easier to calculate and in effect we have removed the joint condition $[S_1 \geq 0, \ldots, S_l \geq 0]$ and reduced to a single condition $[S_l \geq 0]$. Let T be the first hitting time of the negative axis, then

$$p_l = P[T > l] = P[S_1 \geq 0, \ldots, S_l \geq 0].$$

Therefore

$$\sum_{l=1}^{\infty} \epsilon^l p_l = -1 + \exp\left[\sum \frac{\epsilon^l}{l} P[S_l \geq 0]\right]. \tag{1.1.4}$$

Since each $P[S_l \geq 0] = \frac{1}{2}$ by the symmetry assumption on the density ϕ, (1.1.4) is a significant reduction of the problem. We obtain

$$\begin{aligned}\sum_{l=1}^{\infty} p_l \epsilon^l &= -1 + \exp\left[\sum_{l=1}^{\infty} \frac{\epsilon^l}{2l}\right] \\ &= -1 + \exp[-\frac{1}{2}\log(1-\epsilon)] \\ &= -1 + \frac{1}{\sqrt{1-\epsilon}}.\end{aligned}$$

From the Taylor expansion of $\frac{1}{\sqrt{1-\epsilon}}$ we see that

$$P[T > l] = 4^{-l}\binom{2l}{l}. \tag{1.1.5}$$

One remarkable feature of this equation is its independence from the choice of the symmetric density ϕ. It is also similar to (**??**) derived in connection with the simple symmetric random walk on $\mathbf{Z}$.

It remains to prove Spitzer's identity which requires a preliminary result. It is straightforward that the quantity $Q = \delta_\circ + \phi_+ + (\phi \star \phi_+)_+ + \ldots$ satisfies the identity

$$Q = \delta_\circ + \Big(\phi \star Q\Big)_+. \tag{1.1.6}$$

On the other hand we have

Lemma 1.1.2 *For ϕ satisfying (1.1.2), the quantity $Q' = \mathcal{E}(-(\mathcal{L}(\delta_\circ - \phi))_+)$ satisfies the identity (1.1.6) with Q' replacing Q.*

Proof - We have

$$\varepsilon^{f_+} - \delta_\circ = \Big[\varepsilon^{f_+} \star (\delta_\circ - \varepsilon^{-f})\Big]_+$$

since both sides vanish for $x < 0$, $\varepsilon^{f_+ - f}$ vanishes for $x > 0$, and the validity of the identity is easily verified for $x \geq 0$. Now substitute

$$\phi = \delta_\circ - \varepsilon^{-f}$$

and use $f = -\mathcal{L}(\delta_\circ - \phi)$ to obtain (1.1.6) with Q' replacing Q. ♣
Proof of Spitzer's identity - Let $\psi = Q - Q'$, then ψ satisfies

$$\psi = (\phi \star \psi)_+.$$

Since $||\phi||_1 = \int |\phi(x)|dx < 1$, we necessarily have $\psi = 0$. Therefore

$$\delta_\circ + \phi_+ + (\phi \star \phi_+)_+ + (\phi \star (\phi \star \phi_+)_+)_+ + \ldots = \mathcal{E}(-[\mathcal{L}(1-\phi)]_+).$$

To complete the proof of Spitzer's identity we look at both sides of the identity in terms of of their definition via Fourier transforms. The required result follows immediately. ♣

EXERCISES

Exercise 1.1.1 *Use the fact that if* ϕ *and* ψ *vanish for* $x < 0$*, then* $\phi \star \psi$ *vanishes for* $x < 0$ *to deduce that* $\mathcal{E}(\phi_+)$ *(resp.* $\mathcal{E}(\phi_+ - \phi)$*) vanishes for* $x < 0$ *(resp. for* $x > 0$*).*

Exercise 1.1.2 *Assuming* ϕ *satisfies (1.1.2) show that*

$$-\mathcal{L}(1-\phi) = -\log(1-\hat{\phi}).$$

1.2 Laplace transforms

Laplace and Fourier transforms are quite useful mathematical tools in probability theory. In this subection we discuss some of the the basic properties of Laplace transforms and will gradually give several applications to probability theory. It is not our purpose to give a thorough treatment of the foundations of the theory of Laplace transforms, rather we will mention the basic properties and partially justify them to give some credibility and coherence to their applications. Let f be a continuous function on $[0,\infty)$ with *polynomial growth at infinity* which means there is $\rho < \infty$ such that for x sufficiently large

$$|f(x)| \leq x^\rho.$$

Then the Laplace transform of f is defined as

$$\hat{f}(\alpha) = \int_\circ^\infty e^{-\alpha x} f(x)dx,$$

where $\alpha \in [0,\infty)$. The growth condition on f makes the quantity $\hat{f}(\alpha)$ well-defined since the function $e^{-\alpha x}f(x)$ is rapidly decreasing at infinity. The requirement of continuity is not necessary as long as we make some assumption such as such as f is integrable on compact intervals in addition to the growth condition. If F is the probability distribution function of a random variable X taking values in $\mathbf{R}_+$, then we can also define $\hat{f}$ as

$$\hat{f}(\alpha) = \mathsf{E}[e^{-\alpha X}].$$

However we will also be interested in Laplace transforms of non-negative functions which are not necessarily distributions or density functions of random variables.

Example 1.2.1 Let X be a random variable with values in $\mathbf{Z}_+$ and f_X its density function. Although X takes values in $\mathbf{Z}_+$ we may regard it a random variable with values in $\mathbf{R}_+$ and accordingly define the distribution function

$$F_X(x) = P[X < x]$$

The distribution function F_X is no longer a continuous function. In fact it will have a jump discontinuity at a positive integer n equal to $P[X = n]$ and is constant on each intrval $[n, n+1)$. The Laplace transform of of F_X is given by

$$\int_{\circ}^{\infty} e^{-\alpha x} dF(x) = \sum_{j=0}^{\infty} e^{-\alpha n} P[X = n]. \qquad (1.2.1)$$

The validity of the second equality is a simple and instructive exercise using either of the definitions of $\int_{\circ}^{\infty} e^{-\alpha x} dF(x)$ given in §3.4. It is clear that for $\xi = e^{-\alpha}$, the Laplace transform $\hat{f}(\alpha)$ is simply the probability generating function for X. Therefore it is not surprising that Laplace transforms will be quite useful in investigating random processes. ♠

It is clear that Laplace transforms of non-negative functions are positive for all α (unless $f \equiv 0$). Furthermore, differentiation under the integral sign shows that $\hat{f}$ is infinitely differentiable and satisfies the *strong monotonicity condition*:

$$(-1)^n \frac{d^n \hat{f}}{d\alpha^n} \geq 0. \qquad (1.2.2)$$

For applications to stochastic processes it is sometimes necessary to give meaning to $\int_{\circ}^{\infty} e^{-\alpha x} dF(x)$, where F is a non-negative function. Integration by parts allows us to define this as

$$\int_{\circ}^{\infty} e^{-\alpha x} dF(x) = F(0) + \alpha \int_{\circ}^{\infty} e^{-\alpha x} F(x) dx.$$

We denote this quantity as $\hat{f}(\alpha)$ in spite of the fact that the natural candidate for f, namely $\frac{dF}{dx}$ may not even exist as a function. It is clear that $\hat{f}$ still satisfies the strong monotonicity condition (1.2.2).

By a *normalized function* F we mean a function F such that its its left and right limits exist at every point and

$$F(0) = 0, \quad F(x) = \frac{F(x^-) + F(x^+)}{2}.$$

The second condition means that at points of discontinuity, the value of the function is the average of its limiting values. Let F be a normalized function of bounded variation (or for simplicity assume F is discontinuous only at a discrete set of points). Assume that the Laplace transform

$$\hat{f}(\alpha) = \int_{\circ}^{\infty} e^{-\alpha x} dF(x) \qquad (1.2.3)$$

exists. Then its Laplace transform determines F uniquely. A formula for the inversion of the Laplace transform is given in theorem 1.2.1 (4) below.

In our applications of Laplace transform it is necessary to relate the asymptotic behavior of the function $F(t)$ as $t \to \infty$ to that $\hat{f}(\alpha)$ as $\alpha \to 0$. Theorems deducing the asymptotic behavior of $\hat{f}$ from that of F are often called P*Abelian* theorems. The converse implication where the asymptotic behavior of F is deduced from that $\hat{f}$ is called generally called Tauberian theorem(s). The latter results are generally more difficult to establish. Parts (5) and (6) of theorem 1.2.1 below are examples of Abelian and Tauberian theorems for which we have immediate application.

A good reference for the theory of Laplace transforms is the classic monograph by D. V. Widder entitled *The Laplace Transform.* The basic properties of Laplace transforms which we will make use of are summarized in the following theorem:

Theorem 1.2.1 *Let F be a non-negative function on $[0, \infty)$ with polynomial growth at infinity and integrable on bounded intervals. Then*

1. *The Laplace transform of $\hat{f}(\alpha)$ exists and is a completely monotone function.*

2. *A completely monotone (and therefore infinitely differentiable) function φ is the Laplace transform of some non-negative function*[1] *in the sense $\varphi(\alpha) = \int_\circ^\infty e^{-\alpha x} dF(x)$.*

3. *F is a probability distribution function if and only if $\hat{f}(0) = 1$.*

4. *The function F is uniquely determined by its Laplace transform $\hat{f}$ and at points of continuity of F the inversion is given by*

$$F(x) = \lim_{a \to \infty} \sum_{n \leq ax} \frac{(-a)^n}{n!} \frac{d^n \hat{f}}{d\alpha^n}(a).$$

5. *Assume $F(t)$ grows like $\frac{At^\gamma}{\Gamma(\gamma+1)}$ as $t \to \infty$ where $\gamma \geq 0$. Then*

$$\hat{f}(\alpha) = \int_\circ^\infty e^{-\alpha t} dF(t)$$

[1]The precise statement is that such φ is the Laplace transform of a measure. We are only trying to avoid the use of the dreaded word *measure*.

grows like $\frac{A}{\alpha^\gamma}$ *as* $\alpha \to 0^+$.

6. *Conversely, if* $\hat{f}$ *grow like* $\frac{A}{\alpha^\gamma}$ *as* $\alpha \to 0^+$, *then* $F(t)$ *grows like* $\frac{At^\gamma}{\Gamma(\gamma+1)}$ *as* $t \to \infty$.

Laplace transform has the remarkable property of transforming convolutions into products. More precisely, let f and h be continuous functions on $\mathbf{R}_+$ and assume for convenience that they vanish outside a bounded interval. Then

$$\widehat{f \star h}(\alpha) = \int_\circ^\infty \int_{-\infty}^\infty e^{-\alpha x} f(x-y)h(y)dydx.$$

Making the change of variable $x = y + z$ and noting that $f(z) = 0$ for $z \leq 0$ we obtain

$$\widehat{f \star h}(\alpha) = \hat{f}(\alpha)\hat{h}(\alpha). \tag{1.2.4}$$

For convenience we made this calculation by assuming f and h vanish outside a bounded interval. The result is valid considerably more generally. In our applications the validity of this transformation will not be an issue.

Let us apply (1.2.4) to a renewal-type equation

$$z(t) = h(t) + \int_\circ^t z(t-s)dF(s), \tag{1.2.5}$$

where z and h are functions on $[0, \infty)$ and F is a probability distribution vanishing on the negative axis. Taking Laplace transform of this equation and solving the resulting linear equation we obtain

$$\hat{z} = \frac{\hat{h}}{1 - \hat{f}}.$$

Expanding the fraction $\frac{1}{1-\hat{f}}$ formally as the geometric series $\sum \hat{f}^n$ we obtain the expression

$$\hat{z} = \hat{h} + \hat{h}\hat{f} + \hat{h}\hat{f}^2 + \hat{h}\hat{f}^3 + \cdots \tag{1.2.6}$$

Thus we have an algebraic way of solving the renewal type equation (1.2.5). Our interest in really in z not $\hat{z}$ and therefore it is necessary to invert the Laplace transform to obtain an expression for z. By making use of theorem 1.2.1 (5) and (6), sometimes we can obtain useful information without actually inverting the Laplace transform.

INSERT EXAMPLES OF APPLICATIONS OF LAPLACE TRANSFORMS, ABELIAN AND TAUBERIAN THEOREMS, AND EASY PART OF RENEWAL THEOREM

Rather than discussing the validity of these formal manipulations, we use this expression to guess and verify that the solution to the integral equation (1.2.5) is given by

Proposition 1.2.1 *For a bounded function h the eqaution (1.2.5) has a unique bounded solution which is given by*

$$z(t) = h(t) + \int_\circ^t h(t-s)d\mathsf{m}(s).$$

Proof - Defining $z(t)$ as in the proposition and using the fact that

$$\mathsf{m}(t) = F(t) = \int_\circ^t \mathsf{m}(t-s)dF(s)$$

we see that $z(t)$ satisfies (1.2.5). Now suppose z_i, $i = 1, 2$ are two solutions to (1.2.5), then $y(t) = z_1(t) - z_2(t)$ satisfies the equation

$$y(t) = \int_\circ^t y(t-s)dF(s),$$

which we write in the form $y = y \star F$. Therefore by iteration $y = y \star F_n$ for all $n \geq 1$ where $F_1 = F$ and $F_n = F_{n-1} \star F$. Now $F_n(t)$ is the probability the n^{th} event (or arrival) has taken place within time t and therefore for t fixed it tends to 0 as $n \rightarrow \infty$. Now

$$|y(t)| \leq F_n(t) \sup_{0 \leq s \leq t} |y(s)|.$$

From bounded assumption on solutions and $\lim_{n\rightarrow\infty} F_n(t) = 0$ it follows that $y(t) = 0$. ♣

THIS SUBSECTION IS INCOMPLETE

EXERCISES

Exercise 1.2.1 *Let X be a Poisson random variable with parameter λt. Show that*

$$P[X \le \lambda x] = e^{-\lambda t} \sum_{k \le \lambda x} \frac{(\lambda t)^k}{k!} \overset{\lambda \to \infty}{\longrightarrow} \begin{cases} 0, & \text{if } t > x; \\ 1, & \text{if } t < x. \end{cases}$$

Let F be a probability distribution on $[0, \infty)$, and

$$\hat{f}(\lambda) = \int_\circ^\infty e^{-\lambda t} dF(t)$$

be its Laplace transform. Show that

$$\sum_{k \le \lambda x} \frac{(-1)^k \lambda^k}{k!} \frac{d^k \hat{f}}{d\lambda^k}(\lambda) \longrightarrow F(x)$$

as $\lambda \to \infty$.

Exercise 1.2.2 *Assume the distribution F on $[0, \infty)$ has moments $\mu_1, \cdots, \mu_{2n}$. Show that the Laplace transform $\hat{f}$ satisfies*

$$\sum_{k=\circ}^{2n-1} \frac{(-1)^k \mu_k \lambda^k}{k!} \le \hat{f} \le \sum_{k=\circ}^{2n} \frac{(-1)^k \mu_k \lambda^k}{k!}.$$

(Hint - Use the infinite series expansion for e^{-t} and compare the sums of $2n-1$ and $2n$ terms. *Deduce that if all momemnts of F exist, then*

$$\hat{f}(\lambda) = \sum_{k=\circ}^{\infty} \frac{(-1)^k \mu_k \lambda^k}{k!}$$

on any interval $[0, a)$ where the series converges. (Note that this exercise gives a sufficient condition for moments to determine a distribution uniquely. In general moments may not uniquely determine the distribution.)

1.3 Discrete Laplace Transforms

In this subsection we see how the ideas developed in connection with the Laplace transform can be adopted to the case of discrete time Markov chains. As an application we prove theorem **??**. The terminology of discrete Laplace transform is not standard and it will be used in a manner somewhat different from the continuous case, nevertheless it seems appropriate. For a function ϕ on $\mathbf{Z}_+$ we define its *discrete Laplace transform* as the function on $(0,\infty)$ defined as

$$\tilde{\phi}(\alpha) = \sum_{l=\circ}^{\infty} e^{-\alpha l}\phi(l).$$

For ϕ a bounded function, the sum converges for $\alpha \in (0,\infty)$. The value at 0 may or may not be finite and should be dealt with separately.

Let $X_\circ, X_1, X_2, \cdots$ be a Markov chain with P the matrix of transition probabilities. The Laplace transform of P is naturally defined as

$$\tilde{P}^\alpha = \sum_{l=\circ}^{\infty} e^{-\alpha l}P^l,$$

so that $\tilde{P}^\alpha$ is a matrix. Let f a function on the state space S. In analogy with the continuous case we define the Laplace transform of f as

$$\tilde{f}^\alpha(i) = \mathsf{E}_i[\sum_{l=\circ}^{\infty} e^{-\alpha l}f(X_l)],$$

where E_i means conditional expectation relative to $X_\circ = i$. We often identify S with $\mathbf{Z}_+$ and f with a column vector whose i^{th} entry is $f(i)$. With this provision it is clear that

$$\tilde{f}^\alpha(i) = \sum_{j\in S}\tilde{P}^\alpha_{ij}f(j) = \sum_{j\in S}\sum_{l=\circ}^{\infty} e^{-\alpha l}P^{(l)}_{ij}f(j). \tag{1.3.1}$$

In this representation $\tilde{f}^\alpha$ is also a column vector.

Although there is no analogue of the infinitesimal generator A in the discrete case, yet the following result which emulates proposition **??** is valid in this case:

Lemma 1.3.1 *Let f be a bounded function on the state space S, $\alpha \in (0, \infty)$, then $u = \tilde{f}^\alpha(i)$ is the unique solution to the linear system*

$$(I - e^{-\alpha} P)u = f.$$

Proof - Since the entries of P are bounded we can expand $(I - e^{-\alpha}P)^{-1}$ in a geometric series

$$(I - e^{-\alpha}P)^{-1} = I + e^{-\alpha}P + e^{-2\alpha}P^2 + e^{-3\alpha}P^3 + \cdots$$

The fact that $u = \tilde{f}^\alpha(i)$ is a solution follows from (1.3.1). Uniqueness is an immediate consequence of the fact all eigenvalues of P are bounded above by 1 and therefore $I - e^{-\alpha}P$ is invertible. Equivalently, if u_1 and u_2 are solutions, then $v = u_1 - u_2$ is a solution of $(I - e^{-\alpha}P)v = 0$ and consequently

$$v = e^{-\alpha}Pv = e^{-2\alpha}P^2 v = \cdots = e^{-l\alpha}P^l v = \cdots$$

Now let $l \to \infty$ to obtain $v = 0$. ♣

In the application of the discrete Laplace transforms it is essential to obtain the analogue of lemma 1.4.1 which in this case becomes

Lemma 1.3.2 *Let T be a stopping time for the Markov chain $X_\circ, X_1, \cdots$ Then*

$$\tilde{f}^\alpha(i) = \mathsf{E}_i[\sum_{l=0}^{T-1} e^{-\alpha l} f(X_l)] + \mathsf{E}_i[e^{-\alpha T}\tilde{f}^\alpha(X_T)].$$

Proof - The statement of the lemma is equivalent to

$$\mathsf{E}_i[e^{-\alpha T}\tilde{f}^\alpha(X_T)] = \mathsf{E}_i[\sum_{l=T}^{\infty} e^{-\alpha l} f(X_l)]. \tag{1.3.2}$$

The quantity inside $\mathsf{E}_i[.]$ on the right hand side of (1.3.2) can be written as

$$\sum_{l=T}^{\infty} e^{-\alpha l} f(X_l) = e^{-\alpha T} \sum_{l=\circ}^{\infty} e^{-\alpha l} f(X_{T+l}).$$

Therefore

$$
\begin{aligned}
\mathsf{E}_i[\sum_{l=T}^{\infty} e^{-\alpha l} f(X_l)] &= \mathsf{E}_i[\mathsf{E}[e^{-\alpha T}\sum_{\circ}^{\infty} e^{-\alpha l} f(X_{T+l}) \mid X_j,\ j\leq T,\ T]] \\
&= \mathsf{E}_i[e^{-\alpha T}\mathsf{E}[\sum_{\circ}^{\infty} e^{-\alpha l} f(X_{T+l}) \mid X_u,\ u\leq T,\ T]] \\
(T \text{ is a stopping time}) \quad &= \mathsf{E}_i[e^{-\alpha T}\mathsf{E}[\sum_{\circ}^{\infty} e^{-\alpha l} f(X_{T+l}) \mid X_j,\ j\leq T]] \\
(\text{Strong Markov property}) \quad &= \mathsf{E}_i[e^{-\alpha T}\mathsf{E}[\sum_{l=\circ}^{\infty} e^{-\alpha l} f(X_{T+l}) \mid X_T]] \\
&= \mathsf{E}_i[e^{-\alpha T}\tilde{f}^{\alpha}(X_T)],
\end{aligned}
$$

proving the lemma. ♣

It is convenient to introduce some definitions. A function f on the state space of the Markov chain $X_\circ, X_1, \cdots$ is called *α-excessive* where $0 < \alpha \leq \infty$ if

1. $f(i) \geq 0$ for all $i \in S$;
2. $f - e^{-\alpha}Pf \geq 0$.

The case $\alpha = 0$ is called *excessive*.

Lemma 1.3.3 *For $\alpha > 0$ a bounded α-excessive function f can be written as $f = \tilde{h}^{\alpha}$ for a non-negative bounded function h.*

Proof - For an α-excessive function f with $\alpha > 0$ we have $f - e^{-\alpha}Pf = h \geq 0$ and consequently such functions can be written as

$$
f = \sum_{l=\circ}^{\infty} e^{-\alpha l} P^l h = (1 - e^{-\alpha}P)^{-1} h = \tilde{h}^{\alpha}.
$$

proving the claim. ♣

α-excessive functions are well-behaved relative to stopping times in the sense that

Lemma 1.3.4 *Let T be a stopping time for the Markov chain $X_\circ, X_1, \cdots$, and f an α-excessive function. Then*

$$
f(i) \geq \mathsf{E}_i[e^{-\alpha T} f(X_T)].
$$

Proof - Let $\gamma > 0$ be any real number $\geq \alpha$. Then it follows from (1.3.3) that

$$f(i) = \sum_{l=\circ}^{T-1} \mathsf{E}_i[e^{-\gamma l}h(X_l)] + \mathsf{E}_i[e^{-\gamma T}f(X_T)] \geq \mathsf{E}_i[e^{-\gamma T}f(X_T)],$$

which proves the lemma (for $\alpha = 0$ we let $\gamma \to 0$). ♣

Lemma 1.3.5 *Let f be an α-excessive function and $T \leq T'$ stopping times. Then*

$$\mathsf{E}_i[e^{-\alpha T}f(X_T)] \geq \mathsf{E}_i[e^{-\alpha T'}f(X_{T'})]$$

Proof - Let $\gamma > 0$ be any real number $\geq \alpha$. Then, for some $h \geq 0$,

$$\begin{aligned} \mathsf{E}_i[e^{-\gamma T}f(X_T)] &= f(i) - \mathsf{E}_i[\sum_{l=\circ}^{T-1} e^{-\gamma l}h(X_l)] \\ &\geq f(i) - \mathsf{E}_i[\sum_{l=\circ}^{T'-1} e^{-\gamma l}h(X_l)] \\ \text{By (1.3.3)} \quad &= \mathsf{E}_i[e^{-\gamma T'}f(X_{T'})]. \end{aligned}$$

Taking $\lim \gamma \to \alpha$ if $\alpha = 0$ we obtain the desired result. ♣

An important application of the concept of excessive function is to the determination of the value $V(i)$ of the game described in §2.4. Recall that the value of the game is defined as

$$V(i) = \sup_T \mathsf{E}_i[e^{-\alpha T}f(X_T) - \sum_{l=0}^{T-1} e^{-\alpha l}g(X_l)]. \tag{1.3.3}$$

The first step is to reduce the calculation to the case where the cost function $g \equiv 0$. For this purpose we set

$$f' = f + \tilde{g}^\alpha, \qquad V' = V + \tilde{g}^\alpha.$$

Then equation (1.3.3) is equivalent to

$$V'(i) = \sup_T \mathsf{E}_i[e^{-\alpha T}f'(X_T)], \tag{1.3.4}$$

and therefore our problem is mathematically equivalent to the special case where $g \equiv 0$. To understand this point clearly, consider

$$\begin{aligned} \mathsf{E}_i[e^{-\alpha T} f'(X_T)] &= \mathsf{E}_i[e^{-\alpha T} f(X_T) + e^{-\alpha T}\tilde{g}^\alpha(X_T)] \\ \text{(By lemma 1.3.2)} \qquad &= \mathsf{E}_i[e^{-\alpha T} f(X_T) - \sum_{j=\circ}^{T-1} e^{-\alpha j} g(X_j)] + \tilde{g}^\alpha(i). \end{aligned}$$

Taking $\sup_T$, the first term gives $V(i)$ which together with $\tilde{g}^\alpha(i)$ gives $V'(i)$. For this reason from now on we asume $g \equiv 0$.

Lemma 1.3.6 *With the above notation (and under the assumption $g \equiv 0$), the value of the game is an excessive function.*

Proof - Let $\epsilon > 0$ and for each state k let T_k be a stopping time such that

$$\mathsf{E}_k[e^{-\alpha T_k} f(X_{T_k})] \geq V(k) - \epsilon.$$

Let T be the stopping time which on a path $\omega \in \Omega$ is $1 + T_k(\omega)$ if $X_1(\omega) = k$. Then

$$\begin{aligned} \mathsf{E}_i[e^{-\alpha T} f(X_T)] &= \sum_k e^{-\alpha} P_{ik} \mathsf{E}_k[e^{-\alpha T_k} f(X_{T_k}] \\ &\geq \sum_k e^{-\alpha} P_{ik}(V(k) - \epsilon) \\ &= e^{-\alpha} PV(i) - e^{-\alpha}\epsilon. \end{aligned}$$

Since V is obtained by taking $\sup_T$, we have

$$V(i) \geq e^{-\alpha} PV(i) - e^{-\alpha}\epsilon$$

and $V(i)$ is an α-excessive function. ♣

Lemma 1.3.7 *With the above notation and hypotheses, V is the minimal excessive function $\geq f$.*

Proof - Since $T = 0$ is a stopping time we have

$$V(i) \geq \mathsf{E}_i[e^{-\alpha T} f(X_T)] = f(i)$$

proving $V \geq f$. Now if g is any excessive function then by lemma 1.3.4 (taking $T = 0$)

$$g(i) \geq \mathsf{E}_i[e^{-\alpha T} g(X_T)]$$

for any stopping time T. If $g \geq f$ then $\mathsf{E}_i[e^{-\alpha T} g(X_T)] \geq \mathsf{E}_i[e^{-\alpha T} f(X_T)]$. Consequently

$$g(i) \geq \sup_T \mathsf{E}_i[e^{-\alpha T} f(X_T)] = V(i)$$

proving the lemma. ♣

Notice that *a priori* it is not even clear that a minimal excessive function $\geq f$ exists. Lemma 1.3.7 establishes its existence and as noted in theorem **??** it gives an algorithmic way of calculating it. To complete the proof of theorem **??** it remains to show that the exhibited stopping time realizes the desired optimum. We begin with a lemma of a general nature.

Lemma 1.3.8 *Let* $A\imath S$ *be a subset of the state space,* T *the first hitting time of* A *and* h *be an* α*-excessive function. Then* $h'(i) = \mathsf{E}_i[e^{-\alpha T} h(X_T)]$ *is also an* α*-excessive function.*

Proof - Noting that $T(\omega) = 0$ if $\omega(0) \in A$ we define the stopping time T' by

$$T'(\omega) = \min\{l \geq 1 \mid X_l(\omega) \in A\}.$$

Since $T' \geq T$ we have

$$h'(i) \geq \mathsf{E}_i[e^{-\alpha T'} h(X_{T'})]$$

where $=$ holds if $i \notin A$ and for $i \in A$ we $\geq$ in view of lemma 1.3.5. Conditioning on X_1 we obtain

$$\begin{aligned} h'(i) &\geq \mathsf{E}_i[\mathsf{E}[e^{-\alpha T'} h(X_{T'}) | X_1]] \\ &= \sum_j e^{-\alpha} P_{ij} h'(j) \end{aligned}$$

which proves that h' is α-excessive. ♣

To complete the proof of theorem **??** of chapter 2 it is necessary to show that under the assumption of finiteness of the state space S and for $A = \{j \in S \mid f(j) = V(j)\}$, the first hitting time T of A is the optimal strategy. To this end we have to prove that for $h(i) = \mathsf{E}_i[e^{-\alpha T} f(X_T)]$ we have $h(i) = V(i)$. It follows from the definition of V that $h \leq V$. The reverse inequality is established in two steps

1. h is an α-excessive function;

2. $h \geq f$.

Since $f(X_T) = V(X_T)$ and V is an α-excessive function, the first statement follows[2] from lemma 1.3.8. To prove the second statement first note that if $i \in A$ then $T = 0$ and so $h(i) = \mathsf{E}_i[e^{-\alpha T} f(X_T)] = f(i)$. For $i \notin A$ assume $f(i) > h(i)$ and let

$$a = \max_{j \notin A}(f(j) - h(j)),$$

which exists by finiteness of S. Let the maximum be attained at $k \notin A$ Then $a + h \geq f$ and and $a + h$ is an excessive function. since V is the minimal excessive function $\geq f$ we have $a + h \geq V$. Hence

$$V(k) \leq a + h(k) = f(k) - h(k) + h(k) = f(k)$$

which implies $V(k) = f(k)$ contradicting the assumption $k \notin A$. This completes the proof. ♣

[2]This requires the assumption that $P[T < \infty] = 1$. However, by a general argument involving the introduction of a state ϕ with $X_\infty = \phi$ and extending f and v to ϕ by setting $f(\phi) = 0 = V(\phi)$ we circumvent this problem.

EXERCISES

Exercise 1.3.1 *Let f and h be α-excessive functions, show that* $\min(f, h)$ *is α-excessive.*

Exercise 1.3.2 *Let $X_\circ, X_1, X_2, \ldots$ be a Markov chain with state space S, $A \subset S$, f a function on A vanishing on A, and T be the first hitting time of A. Show that the function*

$$u(i) = \mathsf{E}_i[\sum_{l=0}^{T-1} e^{-\alpha l} f(X_l)]$$

satisfies the equation $(I - e^{-\alpha}P)u(i) = f(i)$, for $i \in A'$ where $\alpha \in [0, \infty)$ and A' denotes the complement of A in S. This problem is the like the discrete analogue of solving a boundary value problem.

Exercise 1.3.3 *Show that for an irreducible and recurrent Markov chain an excessive function is a constant.*(Hint - Use the inequality $f(i) \geq \mathsf{E}_i[f(X_T)]$ and let T be the first hitting time of state j to deduce $f(i) \geq f(j)$.)

Exercise 1.3.4 *Show that an excessive function f can be written in the form*

$$f = h + \sum_{l=\circ}^{\infty} P^l g,$$

where $h \geq 0$, $g \geq 0$ and $h = Ph$. Assume the underlying Markov chain is finite and all recurrent states communicate. What can you say about h?

Exercise 1.3.5 (Continuation of exercise 1.3.4) *Let $\mathcal{R}\psi = \sum_{l=\circ}^{\infty} P^l \psi$ where ψ is a non-negative function on the state space. Let f be an excessive function and $g = \mathcal{R}\psi$ for some non-negative function ψ. Show that* $\min(f, g)$ *is of the form $\mathcal{R}\varphi$ for some non-negative function φ.* (Hint - Use the fact that $\min(f, g)$ is excessive and apply exercise 1.3.4.

1.4 The Linear Equation $(\alpha I - A)u = f$

In accordance with our general definition, the the Laplace transforms of the transition probabilities $P_{ij}^{(t)}$ are

$$\tilde{P}_{ij}^{\alpha} = \int_{\circ}^{\infty} e^{-\alpha t} P_{ij}^{(t)} dt.$$

Let f be a bounded function on the state space S of the continuous Markov chain X_t. Then $f(X_t)$ is a function of the Markov chain and its Laplace transform is a function on the state space S and is defined by

$$\tilde{f}^{\alpha}(i) = \mathsf{E}_i[\int_{\circ}^{\infty} e^{-\alpha t} f(X_t) dt],$$

where as noted earlier E_i means conditional expectation conditioned on $X_\circ = i$. Representing f as a column vector we have

$$\mathsf{E}_i[f(X_t)] = \sum_{j \in S} P_{ij}^{(t)} f(j) = (P_t f)_i, \quad \text{where } f = \begin{pmatrix} \vdots \\ f(j) \\ \vdots \end{pmatrix}.$$

It follows that the Laplace transform of $f(X_t)$ can be written as

$$\tilde{f}^{\alpha}(i) = \Big(\sum_{j \in S} \Big[\int_{\circ}^{\infty} e^{-\alpha t} P_t dt\Big] f\Big)_i = \sum_{j \in S} \tilde{P}_{ij}^{\alpha} f(j). \tag{1.4.1}$$

More generally we define the Laplace transform of f at a Markov time T as

$$\tilde{f}^{\alpha}(X_T) = \mathsf{E}_{X_T}[\int_{\circ}^{\infty} e^{-\alpha t} f(X_t) dt], \tag{1.4.2}$$

where we recall that E_{X_T} means conditional expectation conditioned on X_T. The following lemma is an important technical tool in the application of Laplace transforms and relates Laplace transforms to Markov times:

Lemma 1.4.1 *Let T be a Markov time and f be a bounded function on the state space S of a continuous time Markov chain X_t. Then*

$$\tilde{f}^{\alpha}(i) = \mathsf{E}_i\Big[\int_{\circ}^{T} e^{-\alpha t} f(X_t) dt\Big] + \mathsf{E}_i[e^{-\alpha T} \tilde{f}^{\alpha}(X_T)\Big],$$

for all $i \in S$ and $\alpha \geq 0$.

Proof - The statement of the lemma is equivalent to

$$\mathsf{E}_i[\int_T^\infty e^{-\alpha t} f(X_t)dt] = \mathsf{E}_i[e^{-\alpha T}\tilde{f}^\alpha(X_T)]. \qquad (1.4.3)$$

After making the change of variable $t = s + T$, the quantity inside $\mathsf{E}_i[.]$ on left hand side of (1.4.3) can be written as

$$\int_T^\infty e^{-\alpha t} f(X_t)dt = e^{-\alpha T}\int_\circ^\infty e^{-\alpha s} f(X_{T+s})ds.$$

Therefore,

$$\begin{aligned}
\mathsf{E}_i[\int_T^\infty e^{-\alpha t} f(X_t)dt] &= \mathsf{E}_i[\mathsf{E}[e^{-\alpha T}\int_\circ^\infty e^{-\alpha s} f(X_{T+s})ds \mid X_u,\ u \leq T,\ T]] \\
&= \mathsf{E}_i[e^{-\alpha T}\mathsf{E}[\int_\circ^\infty e^{-\alpha s} f(X_{T+s})ds \mid X_u,\ u \leq T,\ T]] \\
(T \text{ is a stopping time}) \quad &= \mathsf{E}_i[e^{-\alpha T}\mathsf{E}[\int_\circ^\infty e^{-\alpha s} f(X_{T+s})ds \mid X_u,\ u \leq T]] \\
(\text{StrongMarkov property}) \quad &= \mathsf{E}_i[e^{-\alpha T}\mathsf{E}[\int_\circ^\infty e^{-\alpha s} f(X_{T+s})ds \mid X_T]] \\
&= \mathsf{E}_i[e^{-\alpha T}\tilde{f}^\alpha(X_T)],
\end{aligned}$$

proving the lemma. ♣

The following application of lemma 1.4.1 relates the Laplace transform to the infinitesimal generator of the continuous time Markov chain:

Proposition 1.4.1 *For a bounded function f on the state space S, $\tilde{f}^\alpha$ is the unique solution to the linear equation*

$$(\alpha I - A)\tilde{f} = f.$$

Proof - Set $T = T_1$ the first transition time in lemma 1.4.1. Then $X_t = X_\circ$ for $t < T$ and the first term on right hand side of the formula for $\tilde{f}(i)$ becomes

$$\begin{aligned}
\mathsf{E}_i[\textstyle\int_\circ^T e^{-\alpha t} f(X_t)dt] &= f(i)\mathsf{E}_i[\textstyle\int_\circ^T e^{-\alpha t}dt] \\
&= \mathsf{E}_i[\tfrac{1-e^{-\alpha T}}{\alpha}] \\
&= \tfrac{f(i)}{\alpha}\left[1 - \textstyle\int_\circ^\infty \lambda_i e^{-\lambda_i t} e^{-\alpha t}dt\right] \\
&= \tfrac{f(i)}{\alpha+\lambda_i}.
\end{aligned}$$

The second term on the right hand side of the formula for $\tilde{f}(i)$ in lemma **??** gives

$$\begin{aligned} \mathsf{E}_i[e^{-\alpha T}\tilde{f}^\alpha(X_T)] &= \sum_{j\in S}\lambda_i Q_{ij}\tilde{f}^\alpha(j)\int_\circ^\infty e^{-\lambda_i t}e^{-\alpha t}dt \\ &= \tfrac{\lambda_i}{\alpha+\lambda_i}Q\tilde{f}^\alpha. \end{aligned}$$

Putting these together we get after a little algebra

$$(\alpha+\lambda_i)\tilde{f}^\alpha(i) - \lambda_i(Q\tilde{f}^\alpha)(i) = f(i).$$

Comparing with the expression for the infinitesimal generator of a continuous time Markov chain given in corollary **??** we see that $\tilde{f}^\alpha$ is a solution of the equation in question. To prove uniqueness assume $(\alpha I - A)g = 0$ for a bounded function g on the state space. Then in view of Kolmogorov's forward equation we have

$$\frac{d}{dt}(e^{-\alpha t}P_t g) = e^{-\alpha t}P_t(\alpha I - A)g = 0.$$

Integrating we obtain

$$0 = \int_\circ^u \frac{d}{dt}(e^{-\alpha t}P_t g)dt = e^{-\alpha u}P_u g - g.$$

Since g is a bounded function, $P_u g$ is uniformly (in u) bounded and taking $\lim_{u\to\infty}$ we obtain $g = 0$ proving uniqueness. ♣

Propositions of the type 1.4.1 are the continuous time analogues of lemma 1.3.1 and have far-reaching implications. For example, if we replace the discrete state space S with $\mathbf{R}^n$, then the continuous time Markov chain should be replaced with a diffusion process and the infinitesimal generator A generally becomes a second order partial differential operator. Then the analogue of proposition **??** gives an explicit expression for the solution of the linear partial differential equation in terms of integration on path spaces (the appearance of expectation E_i). By realizing the Schrödinger's operator as the infinitesimal generator of a diffusion process one obtains the celebrated Feynman-Kac formula.

1.5 Don't be Afraid of Measure Theory

We have already encountered examples of measures. While measure theory does not solve problems arising in random processes, it does provide a several useful theorems and points of view which play important roles in the development of the theory. For example, theorems on convergence of functions or random variables are essential for probability theory. We present here the general outline of measure theory in relation to random processes which are of interest to us. This will not be a systematic treatment, however, it may serve as a bridge between sophisticated treatments of random processes based on measure theory and informal approaches where the word measure is entirely avoided.

We have already encountered examples of measures. The probability spaces Ω of interest to us have been generally spaces of paths. To certain subsets of Ω which have been specified by the values of paths at specific points in time we assigned probabilities. For instance, we considered the infinite coin tossing experiment where 0's appear with probability p and 1's with with probability $q = 1 - p$ in which case Ω is the set of all sequences 0's and 1's. Let $\Omega_{i_1,\cdots,i_n;j_1,\cdots,j_m}\imath\Omega$ be the subset consisting of all sequences $\{\omega(l)\}$ where

$$\omega(i_1) = \cdots = \omega(i_n) = 0, \quad \text{and } \omega_{j_1} = \cdots = \omega_{j_m} = 1.$$

Then to $\Omega_{i_1,\cdots,i_n;j_1,\cdots,j_m}$ we assigned probability $P[\Omega_{i_1,\cdots,i_n;j_1,\cdots,j_m}] = p^n q^m$. Sets of this form often called *cylinder sets*. In practice it was essential to extend assignement of probability to subsets which had a more complex description. For instance, we considered the set of paths which visited a transient state in a Markov chain infinitely often. By exhibiting this set through countable unions and intersections of simple cylinder sets we were able to conclude that the probability of visiting a transient state infinitely often is 0. Since measure is a generalization of the notion of probability it thus reasonable to begin by defining a *measure* (or *measure space*) as a triple $(\Omega, \mathcal{A}, \mu)$ where Ω is a set (e.g. a set of paths), $\mathcal{A}$ a family of subsets of Ω (e.g. cylinder sets, their complements, their countable unions and intersections) and μ a function which to every set $A \in \mathcal{A}$ assigns a non-negative real number $\mu(A)$. Since the probability of union of mutually exclusive events is the sum of their probabilities, we also require

$$\mu(\bigcup A_n) = \sum \mu(A_n),$$

where A_n's are pairwise disjoint sets in $\mathcal{A}$. We always assume Ω and the emptyset $\emptyset$ are elements of $\mathcal{A}$. It is also clear that we should require complements, countable unions and intersections of sets in $\mathcal{A}$ are also in $\mathcal{A}$ in order to ensure that $\mu(.)$ is such sets is defined. A probability space is thuerefore a measure space $(\Omega, \mathcal{A}, \mu)$ such that $\mu(\Omega) = 1$. In practice it is important to consider measure spaces where $\mu(\Omega) = \infty$. In such situations we require that $\Omega = \cup\Omega_i$ with $\mu(\Omega_i) < \infty$ for otherwise measure spaces will become rather unwieldy. Since such pathological situations do not occur in our context and because we will not provide complete proofs of basic results in measure theory, there is no need to dwell on this condition. It is implicitly assumed to hold in this text. A function $f\Omega \to \mathbf{R}$ is called *measurable* if for every interval I (which may be open, closed half closed) the set

$$f^{-1}(I) \stackrel{\text{def}}{\equiv} \{\omega \in \Omega \mid f(\omega) \in I\} \in \mathcal{A}$$

A random variable in this context is simply a measurable function on a probability space. In the special case where the random variables in question are discrete valued, the sets $f^{-1}(I)$, as I runs over sufficiently small intervals define a partition of the space Ω. Therefore the family $\mathcal{A}$, in the context of discrete random variables, can be replaced by a partition of the space Ω and the condition of measurability becomes $\mathcal{A}$-admissibility.

Given a measure space $(\Omega, \mathcal{A}, \mu)$ and an $\mathcal{A}$-measurable function f on Ω, then it is possible to define $\int_\Omega f d\mu$. In the special case where $\mathcal{A}$ is a partition of Ω, then $\mathcal{A}$-measurability means f is constant on pieces of the partition. Then we define

$$\int_\Omega f d\mu = \sum_j f_j \mu(\Omega_j)$$

where $\Omega = \cup\Omega_j$ (disjoint union) and the constant f_j is the value of f on Ω_j. This definition can be extended by approximation to general measure spaces, but overcoming the technical points is rather lengthy and by a leap of faith we simply assume that via approximation and imitating the ideas of Riemann integral the definition can be extended to the general case. A function f is called *p-integrable* or *p-summable* if

$$\int_\Omega |f| d\mu < \infty$$

The space of p-summable functions is denoted by L^p. We are only interested in the case $1 \leq p < \infty$, and we do not distinguish between functions which are equal outside a set of measure 0. The cases $p = 1$ and $p = 2$ are called integrable and square integrable functions.

It is useful to clarify the meaning of some of the notions of convergence which are commonly used in probability and understand their relationship. Let $X_1, X_2, \cdots$ be a sequence of random variables (defined on a probability space Ω).

1. (*L^p Convergence*) - X_n converges to a random variable X (defined on Ω) in L^p, $1 \leq p < \infty$, if

$$\lim_{n\to\infty} \mathsf{E}[|X_n - X|^p] = 0.$$

   We will be mainly interested in $p = 1, 2$. The Cauchy-Schwartz inequality in this context can be stated as

$$|\mathsf{E}[WZ]| \leq \sqrt{\mathsf{E}[W^2]\mathsf{E}[Z^2]},$$

   where W and Z are real-valued random variables. Substituting $W = |X_n - X|$ and $Z = 1$ we deduce that convergence in L^2 implies convergence in L^1. In general if $X_n \to X$ in L^p and $1 \leq q \leq p < \infty$ then $X_n \to X$ is L^q. Note that if $X_n \to X$ in L^1 then $\mathsf{E}[X_n]$ converges to $\mathsf{E}[X]$.

2. (*Pointwise Convergence*) - X_n converges to X pointwise if for every $\omega \in \Omega$ the sequence of numbers $X_n(\omega)$ converges to $X(\omega)$. Pointwise convergence does not imply convergence in L^1 and $\mathsf{E}[X_n]$ may not converge to $\mathsf{E}[X]$ as shown in example **??**. It is often more convenient to relax the notion of pointwise convergence to that of *almost pointwise* convergence or *almost sure* convergence. This means there is a subset $\Omega_\circ \subset \Omega$ such that $P[\Omega_\circ] = 1$ and $X_n \to X$ pointwise on $\Omega_\circ$. We have already seen how deleting a set of probability zero makes it possible to make precise statements about the behavior of a sequence of random variables. For example, with probability 1 (i,e,, in the complement of a set of paths of probability zero) a transient state is visited only finitely many times, or with probability any given pattern appears infinitely often. Or the (strong) law of large numbers states that for an iid sequence of random variables $X_1, X_2, \cdots$ with mean $\mu = \mathsf{E}[X_j]$, we have almost pointwise convergence of the sequence $\frac{X_1+\cdots+X_n}{n}$ to μ.

3. *Convergence in Probability* - Related to almost pointwise convergence is the weaker notion of convergence probability. A sequence $X_n \to X$ *in probability* if for every $\epsilon > 0$ we have

$$\lim_{n\to\infty} P[\{\omega \mid |X_n(\omega) - X(\omega)| \geq \epsilon\}] = 0.$$

This notion is strictly weaker than almost sure convergence in the sense that there are sequences which converge in probability but do not converge almost surely. In fact, example **??** gives an example where the sequence does not converge anywhere pointwise! The Weak Law of Large Numbers is the statement that $\frac{S_n}{n}$ converges to μ in probability.

4. (*Convergence in Distribution*) - Let F_n be the distribution function of X_n and F that of X. Assume F is a continuous function. X_n converges to X in distribution means for every $x \in \mathbf{R}$ we have

$$P[X_n \leq x] = F_n(x) \longrightarrow F(x) = P[X \leq x].$$

The standard statement of the Central Limit Theorem is about convergence in distribution. Convergence in distribution does not imply almost pointwise convergence, but almost pointwise convergence implies convergence in distribution.

The relationship between these four modes of convergence is summarized as follows ($1 \leq q \leq p < \infty$):

$$\begin{array}{ccccc} \text{Conv. in } L^p & \Longrightarrow & \text{Conv. in } L^q & & \\ & & \Downarrow & & \\ & & \text{Conv. in Prob.} & \Longrightarrow & \text{Conv. in Dist.} \\ & & \Uparrow & & \\ & & \text{Almost Sure Conv.} & & \end{array}$$

Example 1.5.1 Consider the sequence of functions $\{f_n\}$ defined on $[0,1]$ defined as (draw a picture to see a sequence of *spike* functions)

$$f_n(x) = \begin{cases} 2^{2n}x & \text{if } 0 \leq x \leq 2^{-n} \\ -2^{2n}x + 2^{n+1} & \text{if } 2^{-n} < x \leq 2^{-n+1} \\ 0 & \text{otherwise.} \end{cases}$$

It is clear that

$$\int_{\circ}^{1} f_n(x)dx = 1,$$

and the sequence f_n converges to the zero function everywhere on $[0,1]$. Therefore

$$\lim_{n\to\infty}\int_{\circ}^{1} f_n(x)dx \neq \int_{\circ}^{1}\lim_{n\to\infty} f_n(x)dx.$$

This is obviously an undesirable situation which should be avoided since when dealing with limiting values we want the integrals (e.g. expectations) to converge to the *right* values. The assumption of uniform integrability (described below) eliminates cases like this when we cannot interchange limit and integral. It will be immediate that the uniform integrability condition is not satisfied for the sequence $\{f_n\}$. ♠

Let X_l on Ω be a sequence of random variables on Ω. For every c let $\Omega_{l;c}$ be the set points where $|X_l| > c$ and $\chi_{l;c}$ be the indicator function of the set $\Omega_{l;c}$. The sequence Ω_l is called *uniformly integrable* if

$$\lim_{c\to\infty}\sup_l \mathsf{E}[X_l\chi_{l;c}] = 0.$$

It is clear that the sequence in example 1.5.1 is not uniformly integrable.

Example 1.5.2 We construct an example of a sequence of functions $\{f_n\}$ proves that convergene in probability is strictly weaker than convergence almost surely, and convergence in L^p does not imply convergence almost surely. We let $\Omega = [0,1]$. We describe the definition of f_n algorithmically rather by a formula since it is easier to understand them in this manner. Let $f_\circ \equiv 1$. Define f_1 and f_2 by subdividing $[0,1]$ into $[0,\frac{1}{2})$ and $[\frac{1}{2},1]$ and defining

$$f_1(x) = \begin{cases} 1, & \text{if } x \in [0,\frac{1}{2}); \\ 0, & \text{otherwise.} \end{cases}, \quad f_2(x) = \begin{cases} 1, & \text{if } x \in [\frac{1}{2},1]; \\ 0, & \text{otherwise.} \end{cases}$$

Next we consider the subdivision

$$[0,1] = [0,\frac{1}{3}) \cup [\frac{1}{3},\frac{2}{3}) \cup [\frac{2}{3},1],$$

and define f_3, f_4 and f_5 to be 1 on $[0, \frac{1}{3})$, $[\frac{1}{3}, \frac{2}{3})$ and $[\frac{2}{3}, 1]$ and 0 elsewhere. Thus having defined f_j for $j < \frac{n(n+1}{2}$ we look at the subdivision

$$[0,1] = [0, \frac{1}{n+1}) \cup [\frac{1}{n+1}, \frac{2}{n+1}) \cup \cdots \cup [\frac{n}{n+1}, 1]$$

and define $f_{k+\frac{n(n+1}{2}}$, for $k = 0, 1, \cdots, n$, to be 1 on $[k, \frac{k+1}{n+1})$ and 0 elsewhere. It is immediate that the sequence f_n converges to the 0 function in probability and in L^1, however, it does not converge to 0 anywhere! Therefore convergence in probability is strictly weaker than almost sure convergence. While f_n does not converge to 0 anywhere, a subsequence of it will converge almost surely to the 0 function. This is a general phenomenon, i.e., convergence in L^1 implies that a subsequence converges almost surely. ♠

The most important facts about convergence of sequences of functions can be summarized as follows:

Theorem 1.5.1 *Let f_n be a sequence of integrable functions*

1. *If f_n are non-negative, then*

$$\int_\Omega \liminf f_n \leq \liminf \int_\Omega f_n.$$

2. *If the sequence f_n is monotone in the sense that $f_n(\omega) \leq f_{n+1}(\omega)$ (almost surely) for all $\omega \in \Omega$ and all n, then*

$$\lim_n \int_\Omega f_n d\mu = \int_\Omega (\lim f_n) d\mu,$$

where both sides of the equation are allowed to be ∞.

3. *Assume there is an integrable function g such that $|f_n(\omega)| \leq g(\omega)$ (almost surely) for all $\omega \in \Omega$. Then*

$$\lim_n \int_\Omega f_n d\mu = \int_\Omega (\lim f_n) d\mu.$$

The same conclusion hold if the sequence is uniformly integrable.

We had noted earlier that even when the distribution function F of a random variable is not differentiable, it is possible to make sense out of dF by for example assigning the value $F(b) - F(a)$ to an interval $I = [a, b]$ or more precisely by defining

$$\int \chi_I(x) dF(x) = F(b) - F(a).$$

This means that while dF may not make sense as a function, it is meaningful as a measure μ assigning the value $\mu(I) = F(b) - F(a)$ to the interval $I = [a, b]$. For this measure space $(\mathbf{R}, \mathcal{A}, \mu)$, $\mathcal{A}$ is the family of subsets of $\mathbf{R}$ which can be written as countable unions and/or intersections of intervals (open,closed or half closed). It is customary to refer to this $\mathcal{A}$ as the family of *Borel sets* in $\mathbf{R}$.

It is useful to cast our previous considerations on infinite visits to a transient state and infinite appearance of a pattern in a more measure theoretic framework. This requires no new ideas and we will repeat the same arguments. The general question can be formulated as follows: Suppose we have an infinite sequence of events A_l, then what is the probability that infinitely many of A_l's occur? The event that infinitely many of A_l's occur is represented as

$$A = \bigcap_l \bigcup_{m=l}^{\infty} A_m. \tag{1.5.1}$$

Lemma 1.5.1 *With the above notation and hypotheses*

1. $P[A] = 0$ *if* $\sum P[A_l] < \infty$.
2. $P[A] = 1$ *if* $\sum P[A_l] 1 \infty$ *and* $A_1, A_2, \ldots$ *are independent.*

Proof - Since $A \subset \cup_{m=l}^{\infty} A_m$

$$P[A] \leq \sum_{l=m}^{\infty} P[A_m] \longrightarrow 0$$

which proves the first assertion. To prove the second assertion let $A' = \cup_l \cap_{m=l}^{\infty} A'_m$ denote the complement of A. Then

$$\begin{aligned} P[\cap_{m=l}^{\infty} A'_m] &= \lim_{N \to \infty} P[\cap_{m=l}^{\infty} A'_m] \\ \text{(by independence)} \quad &= \prod (1 - P[A_m]) \\ &\leq \prod e^{-P[A_m]} \\ &= \exp(-\sum P[A_m]) = 0, \end{aligned}$$

proving the lemma. ♣

EXERCISES

Exercise 1.5.1 *A computer is generating)'s and 1's with probabilities* p *and* $1-p$*. Let* A_l *be the event that there is sequence of length* l *of* 0*'s in the time period* $2^l, 2^l+1, \ldots, 2^{l+1}$*. Show that*

$$P[A_l \text{ infinitely often}] = \begin{cases} 1, & \textit{if } p \geq \frac{1}{2}; \\ 0, & \textit{otherwise.} \end{cases}$$

(Hint - Let $B_i^{(l)}$ be the event that there is a sequence of l consecutive 0's beginning at time $2^l + (i-1)l$.)