1 Continuous Time Processes

1.1 Continuous Time Markov Chains

Let X_t be a family of random variables, parametrized by $t \in [0, \infty)$, with values in a discrete set S (e.g., $\mathbf{Z}$). To extend the notion of Markov chain to that of a *continuous time* Markov chain one naturally requires

$$P[X_{s+t} = j | X_s = i, X_{s_n} = i_n, \cdots, X_{s_1} = i_1] = P[X_{s+t} = j | X_s = i] \quad (1.1.1)$$

for all $t > 0$, $s > s_n > \cdots > s_1 \geq 0$ and $i, j, i_k \in S$. This is the obvious analogue of the *Markov property* when the discrete time variable l is replaced by a continuous parameter t. We refer to equation (1.1.1) as the Markov property and to the quantities $P[X_{s+t} = j | X_s = i]$ as *transition probabilities* or *matrices*.

We represent the transition probabilities $P[X_{s+t} = j | X_s = i]$ by a possibly infinite matrix P^s_{s+t}. Making the time homogeneity assumption as in the case of Markov chain, we deduce that the matrix P^s_{s+t} depends only on the difference $s + t - s = t$ and and therefore we simply write P_t instead of P^s_{s+t}. Thus for a continuous time Markov chain, the family of matrices P_t (generally an infinite matrix) replaces the single transition matrix P of a Markov chain.

In the case of Markov chains the matrix of transition probabilities after l units of time is given by P^l. The analogous statement for a continuous time Markov chain is

$$P_{s+t} = P_t P_s. \quad (1.1.2)$$

This equation is known as the *semi-group property*. As usual we write $P^{(t)}_{ij}$ for the $(i, j)^{\text{th}}$ entry of the matrix P_t. The proof of (1.1.2) is similar to that of the analogous statement for Markov chains, viz., that the matrix of transition probabilities after l units of time is given by P^l. Here the transition probability from state i to state j after $t + s$ units is given

$$\sum_k P^{(t)}_{ik} P^{(s)}_{kj} = P^{(t+s)}_{ij},$$

which means (1.1.2) is valid. Naturally $P_\circ = I$.

Just as in the case of Markov chains it is helpful to explicitly describe the structure of the underlying probability space Ω of a continuous time Markov chain. Here Ω is the space of step functions on $\mathbf{R}_+$ with values in the state space S. We also impose the additional requirement of right continuity on $\omega \in \Omega$ in the form

$$\lim_{t \to a^+} \omega(t) = \omega(a),$$

which means that $X_t(\omega)$, regarded as a function of t for each fixed ω, is right continuous function. This gives a definite time for transition from state i to state j in the sense that transition has occurred at time $t_\circ$ means $X_{t_\circ - \epsilon} = i$ for $\epsilon > 0$ and $X_{t_\circ + \delta} = j$ for $\delta \geq> 0$ near $t_\circ$. One often requires an initial condition $X_\circ = i_\circ$ which means that we only consider the subset of Ω consisting of functions ω with $\omega(0) = i_\circ$. Each $\omega \in \Omega$ is a *path* or a realization of the Markov chain. In this context, we interpret $P_{ij}^{(t)}$ as the probability of the set of paths ω with $\omega(t) = j$ given that $\omega(0) = i$.

The concept of accessibility and communication of states carries over essentially verbatim from the discrete case. For instance, state j is accessible from state i if for some t we have $P_{ij}^{(t)} > 0$, where $P_{ij}^{(t)}$ denotes the $(i, j)^{\text{th}}$ entry of the matrix P_t. The notion of communication is an equivalence relation and the set of states can be decomposed into equivalence classes accordingly.

The semi-group property has strong implications for the matrices P_t. For example, it immediately implies that the matrices P_s and P_t commute

$$P_s P_t = P_{s+t} = P_t P_s.$$

A continuous time Markov chain is determined by the matrices P_t. The fact that we now have a continuous parameter for time allows us to apply notions from calculus to continuous Markov chains in a way that was not possible in the discrete time chain. However, it also creates a number of technical issues which we treat only superficially since a thorough account would require invoking substantial machinery from functional analysis. We assume that the matrix of transition probabilities P_t is right continuous, and therefore

$$\lim_{t \to 0^+} P_t = I. \tag{1.1.3}$$

The limit here means entry wise for the matrix P_t. While no requirement of uniformity relative to the different entries of the matrix P_t is imposed, we use this limit also in the sense that for any vector v (in the appropriate function space) we have $\lim_{t \to 0} vP_t = v$. We define the *infinitesimal generator* of the continuous time Markov chain as the one-sided derivative

$$A = \lim_{h \to 0+} \frac{P_h - I}{h}.$$

A is a real matrix independent of t. For the time being, in a rather cavalier manner, we ignore the problem of the existence of this limit and proceed as if the matrix A exists and has finite entries. Thus we define the derivative of P_t at time t as

$$\frac{dP_t}{dt} = \lim_{h \to 0+} \frac{P_{t+h} - P_t}{h},$$

where the derivative is taken entry wise. The semi-group property implies that we can factor P_t out of the right hand side of the equation. We have two choices namely factoring P_t out on the left or on the right. Therefore we get the equations

$$\frac{dP_t}{dt} = AP_t, \qquad \frac{dP_t}{dt} = P_tA. \tag{1.1.4}$$

These differential equations are known as the *Kolmogorov backward* and *forward* equations respectively. They have remarkable consequences some of which we will gradually investigate.

The (possibly infinite) matrices P_t are Markov or stochastic in the sense that entries are non-negative and row sums are 1. Similarly the matrix A is not arbitrary. In fact,

Lemma 1.1.1 *The matrix $A = (A_{ij})$ has the following properties:*

$$\sum_j A_{ij} = 0, \quad A_{ii} \leq 0, \quad A_{ij} \geq 0 \text{ for } i \neq j.$$

Proof - Follows immediately from the stochastic property of P_h and the definition $A = \lim_{h\to 0} \frac{P_h - I}{h}$. ♣

So far we have not exhibited even a single continuous time Markov chain. Using (1.1.4) we show that it is a simple matter to construct many examples of stochastic matrices P_t, $t \geq 0$.

Example 1.1.1 Assume we are given a matrix A satisfying the properties of lemma 1.1.1. Can we construct a continuous time Markov chain from A? If A is an $n \times n$ matrix or it satisfies some boundedness assumption, we can in principle construct P_t easily. The idea is to explicitly solve the Kolomogorov (forward or backward) equation. In fact if we replace the matrices P_t and A by scalars, we get the differential equation $\frac{dp}{dt} = ap$ which is easily solved by $p(t) = Ce^{at}$. Therefore we surmise the solution $P_t = Ce^{tA}$ for the Kolmogorov equations where we have defined the exponential of a matrix B as the infinite series

$$e^B = \sum_j \frac{B^j}{j!} \tag{1.1.5}$$

where $B^\circ = I$. Substituting tA for B and differentiating formally we see that Ce^{tA} satisfies the Kolmogorov equation for any matrix C. The requirement $P_\circ = I$ (initial condition) implies that we should set $C = I$, so that

$$P_t = e^{tA} \tag{1.1.6}$$

is the desired solution to the Kolmogorov equation. Some boundedness assumption on A would ensure the existence of e^{tA}, but we shall not dwell on the issue of the existence and meaning of e^B which cannot be adequately treated in this context. An immediate implication of (1.1.6) is that

$$\det P_t = e^{t\mathrm{Tr}A} > 0,$$

assuming the determinant and trace exist. For a discrete time Markov chain $\det P$ can be negative. It is necessary to verify that the matrices P_t fulfill the requirements of a stochastic matrix. Proceeding formally (or by assuming the matrices in question are finite) we show that if the matrix A fulfills the requirements of lemma 1.1.1, then P_t is a stochastic matrix. To prove this let A and B be matrices with row sums equal to zero, then the sum of entries of the i^{th} row of AB is (formally)

$$\sum_k \sum_j A_{ij} B_{jk} = \sum_j A_{ij} \sum_k B_{jk} = 0.$$

From this and the definition of e^B it follows that that the row sums of entries of the matrix e^{tA} are 1. To prove non-negativity of the entries we make use of the formula (familiar from calculus for A a scalar)

$$e^{tA} = \lim_{n\to\infty} (I + \frac{tA}{n})^n. \tag{1.1.7}$$

It is clear that for n sufficiently large the entries of the $N \times N$ matrix $I + \frac{tA}{n}$ are non-negative (boundedness condition on entries of A) and consequently e^{tA} is a non-negative matrix. ♠

Nothing in the definition of a continuous time Markov chain ensures the existence of the infinitesimal generator A. In fact it is possible to construct continuous time Markov chains with diagonal entries of A being $-\infty$. Intuitively this means the transition out of a state may be instantaneous. For many Markov chains appearing in the analysis of problems of interest do not allow of instantaneous transitions. We eliminate this possibility by the requirement

$$P[X_{s+h} = i \text{ for all } h \in [0, \epsilon) \mid X_s = i] = 1 - \lambda_i h + o(h), \tag{1.1.8}$$

as $\epsilon \rightarrow 0$. Here λ_i is a non-negative real number and the notation $g(\epsilon) = o(\epsilon)$ means

$$\lim_{\epsilon \rightarrow 0} \frac{g(\epsilon)}{\epsilon} = 0.$$

Let T denote the time of first transition out of state i where we assume $X_\circ = 0$. Excluding the possibility of instantaneous transition out of i, the random variable T is necessarily memoryless for otherwise the Markov property, whereby assuming the knowledge of the current state the knowledge of the past is irrelevant to predictions about the future, will be violated. It is a standard result in elementary probability that the only memoryless continuous random variables are exponentials. Recall that the distribution function for the exponential random variable T with parameter λ is given by

$$P[T < t] = \begin{cases} 1 - e^{-\lambda t}, & \text{if } t > 0; \\ 0, & \text{if } t \leq 0. \end{cases}$$

The mean and variance of T are $\frac{1}{\lambda}$. It is useful to allow the possibility $\lambda_i = \infty$ which means that the state i is *absorbing*, i.e., no transition out of it is possible. The exponential nature of the transition time is compatible with the requirement (1.1.8). With this assumption one can rigorously establish existence of the infinitesimal generator A, but this is just a technical matter which we shall not dwell on.

While we have eliminated the case of instantaneous transitions, there is nothing in the definitions that precludes having an infinite number transitions in a finite interval of time. It in fact a simple matter to construct continuous time Markov chains where infinitely many transitions in finite interval occur with positive probability. In fact, since the expectation of an exponential random variable with parameter λ is $\frac{1}{\lambda}$, it is intuitively clear that if λ_i's increase sufficiently fast we should expect infinitely transitions in a finite interval. In order to analyze this issue more closely we consider a family $T_1, T_2, \ldots$ of independent exponential random variables with T_k having parameter λ_k. Then we consider the infinite sum $\sum_k T_k$. We consider the events

$$[\sum_k T_k < \infty] \quad \text{and} \quad [\sum_k T_k = \infty].$$

The first event means there are infinitely many transitions in a finite interval of time, and the second is the complement. It is intuitively clear that if the rates λ_k increases sufficiently rapidly we should expect infinitely many transitions in a finite interval, and conversely, if the rates do not increase too fast then only finitely many transitions are possible infinite time. More precisely, we have

Proposition 1.1.1 *Let $T_1, T_2, \ldots$ be independent exponential random variables with parameters $\lambda_1, \lambda_2, \ldots$. Then $\sum_k \frac{1}{\lambda_k} < \infty$ (resp. $\sum_k \frac{1}{\lambda_k} = \infty$) implies $P[\sum_k T_k < \infty] = 1$ (resp. $P[\sum_k T_k = \infty] = 1$).*

Proof - We have

$$\mathsf{E}[\sum_k T_k] = \sum_k \frac{1}{\lambda_k},$$

and therefore if $\sum_k \frac{1}{\lambda_k} < \infty$ then $P[\sum T_k = \infty] = 0$ which proves the first assertion. To prove the second assertion note that

$$\mathsf{E}[e^{-\sum T_k}] = \prod_k \mathsf{E}[e^{-T_k}].$$

Now

$$\mathsf{E}[e^{-T_k}] \quad = \quad \int_\circ^\infty P[-T_k > \log s]ds$$

$$
\begin{aligned}
&= \int_{\circ}^{\infty} P[T_k < -\log s]ds \\
&= \frac{\lambda_k}{1+\lambda_k}.
\end{aligned}
$$

Therefore, by a standard theorem on infinite products[1],

$$\mathsf{E}[e^{-\sum T_k}] = \prod \frac{1}{1+\frac{1}{\lambda_k}} = 0.$$

Since $e^{-\sum T_k}$ is a non-negative random variable, its expectation can be 0 only if $\sum T_k = \infty$ with probability 1. ♣

Remark 1.1.1 It may appear that the Kolmogorov forward and backward equations are one and the same equation. This is not the case. While A and P_t formally commute, the domains of definition of the operators AP_t and P_tA are not necessarily identical. The difference between the forward and backward equations becomes significant, for instance, when dealing with certain boundary conditions where there is instantaneous return from boundary points (or points at infinity) to another state. However if the infinitesimal generator A has the property that the absolute values of the diagonal entries satisfy a uniform bound $|A_{ii}| < c$, then the forward and backward equations have the same solution P_t with $P_\circ = I$. In general, the backward equation has more solutions than the forward equation and its minimal solution is also the solution of the forward equation. Roughly speaking, this is due to the fact that A can be an unbounded operator, while P_t has a smoothing effect. An analysis of such matters demands more technical sophistication than we are ready to invoke in this context. ♡

Remark 1.1.2 The fact that the series (1.1.5) converges is is easy to show for finite matrices or under some bounded assumption on the entries of the matrix A. If the entries A_{jk} grow rapidly with k, j, then there will be convergence problems. In manipulating exponentials of (even finite) matrices one should be cognizant of the fact that if $AB \neq BA$ then $e^{A+B} \neq e^A e^B$. On the other hand if $AB = BA$ then $e^{A+B} = e^A e^B$ as in the scalar case. ♡

[1]Let a_k be a sequence of positive numbers, then the infinite product $\prod(1+a_k)^{-1}$ diverges to 0 if and only if $\sum a_k = \infty$. The proof is by taking logarithms and expanding the log and can be found in many books treating infinite series and products, e.g. Titchmarsh - *Theory of Functions*, Chapter 1.

Remark 1.1.3 In spite of remark 1.1.2 there is a formula for the exponential of the sum of two matrices which we will make use of later and has other applications especially to problems arising in mathematical physics. For symmetric or hermitian matrices A and B we have

$$e^{-(A+B)} = \lim_{n\to\infty} (e^{-\frac{A}{n}} e^{-\frac{B}{n}})^n. \tag{1.1.9}$$

The proof of this identity, known as *Trotter's Product Formula* is based on the following observation: Let $S_t = e^{-t(A+B)}$, and $U_t = e^{-tA}e^{-tB}$. Then for a vector v we have

$$(S_t - U^n_{\frac{t}{n}})v = \sum_{j=0}^{n-1} U^j_{\frac{t}{n}} (S_{\frac{t}{n}} - U_{\frac{t}{n}}) S^{n-j-1} v.$$

Setting $v_s = S_s v$ and taking norms we obtain

$$||(S_t - U^n_{\frac{t}{n}})v|| \leq Cn \sup_{0\leq s\leq t} ||(S_{\frac{t}{n}} - U_{\frac{t}{n}})v_s||, \tag{1.1.10}$$

for some constant C. Now write $S_r - U_r = (S_r - I) - (U_r - I)$ and note that

$$\lim_{r\to 0} \frac{S_r - I}{r} = -(A+B) = \lim_{r\to 0} \frac{U_r - I}{r},$$

to deduce

$$\lim_{n\to\infty} n \sup_{0\leq s\leq t} ||(S_{\frac{T}{n}} - U_{\frac{t}{n}})v_s|| = 0,$$

which proves the required identity at least for finite matrices. In actual applications it is necessary to apply this identity to infinite matrices where the entries may become infinitely large (unbounded operators). The proof in this case involves mathematical subtleties that we cannot discuss in this context, but we will make use of Trotter's Product Formula anyway. ♡

Recall that the stationary distribution played an important role in the theory of Markov chains. For a continuous time Markov chain we similarly define the stationary distribution as a row vector $\pi = (\pi_1, \pi_2, \cdots)$ satisfying

$$\pi P_t = \pi \quad \text{for all } t \geq 0, \quad \sum \pi_j = 1, \ \pi_j \geq 0. \tag{1.1.11}$$

The following lemma re-interprets $\pi P_t = \pi$ in terms of the infinitesimal generator A:

Lemma 1.1.2 *The condition $\pi P_t = \pi$ is equivalent to $\pi A = 0$.*

Proof - It is immediate that the condition $\pi P_t - \pi = 0$ implies $\pi A = 0$. Conversely, if $\pi A = 0$, then the Kolmogorov backward equation implies

$$\frac{d(\pi P_t)}{dt} = \pi \frac{dP_t}{dt} = \pi A P_t = 0.$$

Therefore πP_t is independent of t. Substituting $t = 0$ we obtain $\pi P_t = \pi P_\circ = \pi$ as required. ♣

Example 1.1.2 We apply the above considerations to an example from queuing theory. Assume we have a server which can service one customer at a time. The service times for customers are independent identically distributed exponential random variables with parameter μ. The arrival times are also assumed to be independent identically identically distributed exponential random variables with parameter λ. The customers waiting to be serviced stay in a queue and we let X_t denote the number of customers in the queue at time t. Our assumption regarding the exponential arrival times implies

$$P[X(t+h) = k+1 \mid X(t) = k] = \lambda h + o(h).$$

Similarly the assumption about service times implies

$$P[X(t+h) = k-1 \mid X(t) = k] = \mu h + o(h).$$

It follows that the infinitesimal generator of X_t is

$$A = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & 0 & \cdots \\ \mu & -(\lambda+\mu) & \lambda & 0 & 0 & \cdots \\ 0 & \mu & -(\lambda+\mu) & \lambda & 0 & \cdots \\ 0 & 0 & \mu & -(\lambda+\mu) & \lambda & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

The system of equations $\pi A = 0$ becomes

$$-\lambda\pi_\circ + \mu\pi_1 = 0, \cdots, \lambda\pi_{i-1} - (\lambda+\mu)\pi_i + \mu\pi_{i+1} = 0, \cdots$$

This system is easily solved to yield

$$\pi_i = (\frac{\lambda}{\mu})^i \pi_\circ.$$

For $\lambda < \mu$ we obtain

$$\pi_i = (1 - \frac{\lambda}{\mu})(\frac{\lambda}{\mu})^i.$$

as the stationary distribution. ♠

The semi-group property (1.1.2) implies

$$P_{ii}^{(t)} \geq \left(P_{ii}^{\frac{t}{n}}\right)^n.$$

The continuity assumption (1.1.3) implies that for sufficiently large n, $P_{ii}^{(\frac{t}{n})} > 0$ and consequently

$$P_{ii}^{(t)} > 0, \quad \text{for} \ \ t > 0.$$

More generally, we have

Lemma 1.1.3 *The diagonal entries $P_{ii}^{(t)} > 0$, and off-diagonal entries $P_{ij}^{(t)}$, $i \neq j$, are either positive for all $t > 0$ or vanish identically. The entries of the matrix P_t are right continuous as functions of t.*

Proof - We already know that $P_{jj}^{(t)} > 0$ for all t. Now assume $P_{ij}^{(t)} = 0$ where $i \neq j$. Then for $\alpha, \beta > 0$, $\alpha + \beta = 1$ we have

$$P_{ij}^{(t)} \geq P_{ii}^{(\alpha t)} P_{ij}^{(\beta t)}.$$

Consequently $P_{ij}^{(\beta t)} = 0$ for all $0 < \beta < 1$. This means that if $P_{ij}^{(t)} = 0$, the $P_{ij}^{(s)} = 0$ for all $s \leq t$. The conclusion that $P_{ij}^{(s)} = 0$ for all s is proven later (see corollary 1.4.1). The continuity property (1.1.3)

$$\lim_{h \to 0^+} P_{t+h} = P_t \Big(\lim_{h \to 0^+} P_h \Big) = P_t.$$

implies right continuity of $P_{ij}^{(t)}$. ♣

Note that in the case of a finite state Markov chain P_t has convergent series representation $P_t = e^{tA}$, and consequently the entries $P_{ij}^{(t)}$ are analytic functions of t. An immediate consequence is

Corollary 1.1.1 *If all the states of a continuous time Markov chain communicate, then the Markov chain has the property that $P^{(t)}_{ij} > 0$ for all $i, j \in S$ (and all $t > 0$). In particular, if S is finite then all states are aperiodic and recurrent.*

In view of the existence of periodic states in the discrete time case, this corollary stands in sharp contrast to the latter situation. The existence of the limiting value of $\lim_{l\to\infty} P^l$ for a finite state Markov chain and its implication regarding long term behavior of the Markov chain was discussed in §1.4. The same result is valid here as well, and the absence of periodic states for continuous Markov chains results in a stronger proposition. In fact, we have

Proposition 1.1.2 *Let X_t be a finite state continuous time Markov chain and assume all states communicate. Then P_t has a unique stationary distribution $\pi = (\pi_1, \pi_2, \cdots)$, and*

$$\lim_{t\to\infty} P^{(t)}_{ij} = \pi_i.$$

Proof - It follows from the hypotheses that for some $t > 0$ all entries of P_t are positive, and consequently for all $t > 0$ all entries of P_t are positive. Fix $t > 0$ and let $Q = P_t$ be the transition matrix of a finite state (discrete time) Markov chain. $\lim_l Q^l$ is the rank one matrix each row of which is the stationary distribution of the Markov chain. This limit is independent of choice of $t > 0$ since

$$\lim_{l\to\infty} P_t(P_s)^l = \lim_{l\to\infty} (P_s)^l$$

for every $s > 0$. ♣

EXERCISES

Exercise 1.1.1 *A hospital owns two identical and independent power generators. The time to breakdown for each is exponential with parameter λ and the time for repair of a malfunctioning one is exponential with parameter μ. Let $X(t)$ be the Markov process which is the number of operational generators at time $t \geq 0$. Assume $X(0) = 2$. Prove that the probability that both generators are functional at time $t > 0$ is*

$$\frac{\mu^2}{(\lambda+\mu)^2} + \frac{\lambda^2 e^{-2(\lambda+\mu)t}}{(\lambda+\mu)^2} + \frac{2\lambda\mu e^{-(\lambda+\mu)t}}{(\lambda+\mu)^2}.$$

Exercise 1.1.2 *Let $\alpha > 0$ and consider the random walk X_n on the non-negative integers with a reflecting barrier at 0 defined by*

$$p_{i\ i+1} = \frac{\alpha}{1+\alpha}, \quad p_{i\ i-1} = \frac{1}{1+\alpha}, \quad \text{for } i \geq 1.$$

1. *Find the stationary distribution of this Markov chain for $\alpha < 1$.*
2. *Does it have a stationary distribution for $\alpha \geq 1$?*

Exercise 1.1.3 *(Continuation of exercise 1.1.2) - Let $Y_\circ, Y_1, Y_2, \cdots$ be independent exponential random variables with parameters $\mu_\circ, \mu_1, \mu_2, \cdots$ respectively. Now modify the Markov chain X_n of exercise 1.1.2 into a Markov process by postulating that the holding time in state j before transition to $j-1$ or $j+1$ is random according to Y_j.*

1. *Explain why this gives a Markov process.*
2. *Find the infinitesimal generator of this Markov process.*
3. *Find its stationary distribution by making reasonable assumption on μ_j's and $\alpha < 1$.*

1.2 Inter-arrival Times and Poisson Processes

Poisson processes are perhaps the most basic examples of continuous time Markov chains. In this subsection we establish their basic properties. To construct a Poisson process we consider a sequence $W_1, W_2, \ldots$ of iid exponential random variables with parameter λ. W_j's are called *inter-arrival times*. Set $T_1 = W_1$, $T_2 = W_\circ + W_1$ and $T_n = T_{n-1} + W_n$. T_j's are called *arrival times*. Now define the Poisson process N_t with parameter λ as

$$N_t = \max\{n \mid W_1 + W_2 + \cdots + W_n \leq t\} \tag{1.2.1}$$

Intuitively we can think of certain events taking place and every time the event occurs the counter N_t is incremented by 1. We assume $N_\circ = 0$ and the times between consecutive events, i.e., W_j's, being iid exponentials with the same parameter λ. Thus N_t is the number of events that have taken place until time t. The validity of the Markov property follows from the construction of N_t and the exponential nature of the inter-arrival times, so that the Poisson process is a continuous time Markov chain.

The arrival and inter-arrival times can be recovered from N_t by

$$T_n = \sup\{t \mid N_t \leq n-1\}, \tag{1.2.2}$$

and $W_n = T_n - T_{n-1}$. One can similarly construct other counting processes[2] by considering sequences of independent random variables $W_1, W_2, \ldots$ and defining T_n and N_t just as above. The assumption that W_j's are exponential is necessary and sufficient for the resulting process to be Markov. What makes Poisson processes special among Markov counting processes is that the inter-arrival times have the same exponential law. The case where W_j's are not necessarily exponential (but iid) is very important and will be treated in connection with renewal theory later.

The underlying probability space Ω for a Poisson process is the space non-decreasing right continuous step function functions such that at each point of discontinuity $a \in \mathbf{R}_+$

$$\varphi(a) - \lim_{t \to a^-} \varphi(t) = 1,$$

reflecting the fact that from state n only transition to state $n+1$ is possible.

[2]By a *counting process* we mean a continuous time process N_t taking values in $\mathbf{Z}_+$ such that the only possible transition is from state n to state $n+1$.

To analyze Poisson processes we begin by calculating the density function for T_n. Recall that the distribution of a sum of independent exponential random variables is computed by convolving the corresponding density functions (or using Fourier transforms to convert convolution to multiplication.) Thus it is a straightforward calculation to show that $T_n = W_1 + \cdots + W_n$ has density function

$$f_{(n,\mu)}(x) = \begin{cases} \frac{\mu e^{-\mu x}(\mu x)^{n-1}}{(n-1)!} & \text{for } x \geq 0; \\ 0 & \text{for } x < 0. \end{cases} \tag{1.2.3}$$

One commonly refers to $f_{(n,\mu)}$ as Γ density with parameters (n, μ), so that T_n has Γ distribution with parameters (n, μ). From this we can calculate the density function for N_t, for given $t > 0$. Clearly $\{T_{n+1} \leq t\} \subset \{T_n \leq t\}$ and the event $\{N_t = n\}$ is the complement of $\{T_{n+1} \leq t\}$ in $\{T_n \leq t\}$. Therefore by (1.2.3) we have

$$P[N_t = n] = \int_\circ^t f_{(n,\mu)}(x)dx - \int_\circ^t f_{(n+1,\mu)}(x)dx = \frac{e^{-\mu t}(\mu t)^n}{n!}. \tag{1.2.4}$$

Thus N_t is a $\mathbf{Z}_+$-valued random variable whose distribution is Poisson with parameter μt, hence the terminology Poisson process. This suggests that we can interpret the Poisson process N_t as the number of arrivals at a server in the interval of time $[0, t]$ where the assumption is made that the number of arrivals is a random variable whose distribution is Poisson with parameter μt.

To compute the infinitesimal generator of the Poisson process we note that in view of (1.2.4) for $h > 0$ small we have

$$P[N_h = 0] - 1 = -\mu h + o(h), \quad P[N_h = 1] = \mu h + o(h), \quad P[N_h \geq 2] = o(h).$$

It follows that the infinitesimal generator of the Poisson process N_t is

$$A = \begin{pmatrix} -\mu & \mu & 0 & 0 & 0 & \cdots \\ 0 & -\mu & \mu & 0 & 0 & \cdots \\ 0 & 0 & -\mu & \mu & 0 & \cdots \\ 0 & 0 & 0 & -\mu & \mu & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \tag{1.2.5}$$

It is clear that N_t is *stationary* in the sense that $N_{s+t} - N_s$ has the same distribution as N_t. In addition to stationarity Poisson processes have another remarkable property. Let $0 \leq t_1 < t_2 \leq t_3 < t_4$, then the random variables $N_{t_2} - N_{t_1}$ and $N_{t_4} - N_{t_3}$ are independent. This property is called *independence of increments* of Poisson processes. The validity of this property can be understood intuitively without a formal argument. The essential point is that the inter-arrival times have the same exponential distribution and therefore the number of increments in the interval (t_3, t_4) is independent of how many transitions have occurred up to time t_3 an in particular independent of the number of transitions in the interval (t_1, t_2). This property will also follow more formally from the discussion below. To further analyze Poisson processes we recall the following elementary fact:

Lemma 1.2.1 *Let X_i be random variables with uniform density on $[0, a]$ with their indices re-arranged so that $X_1 < X_2 < \cdots < X_m$ (called order statistics. The joint distribution of $(X_1, X_2, \cdots, X_n)$ is*

$$f(x_1, \cdots, x_m) = \begin{cases} \frac{m!}{a^m}, & \text{if } x_1 \leq x_2 \leq \cdots \leq x_m; \\ 0, & \text{otherwise.} \end{cases}$$

Proof - The $m!$ permutations decompose $[0, a]^m$ into $m!$ subsets according to

$$x_{i_1} \leq x_{i_2} \leq \cdots \leq x_{im}$$

from which the required result follows.

Let $X_1, \cdots, X_m$ be (continuous) random variables with density function $f(x_1, \cdots, x_m)dx_1 \cdots dx_m$. Let $Y_1, \cdots, Y_m$ be random variables such that X_j's and Y_j's are related by invertible transformations. Thus the joint density of $Y_1, \cdots, Y_m$ is given by $h(y_1, \cdots, y_m)dy_1 \cdots dy_m$. The density functions f and h are then related by

$$h(y_1, \cdots, y_m) = f(x_1(y_1, \cdots, y_m), \cdots, x_m(y_1, \cdots, y_m))|\frac{\partial(x_1, \cdots, x_m)}{\partial(y_1, \cdots, y_m)}|,$$

where $\frac{\partial(x_1, \cdots, x_m)}{\partial(y_1, \cdots, y_m)}$ denotes the the familiar Jacobian determinant from calculus of several variables. In the particular case that X_j's and Y_j's are related by an invertible linear transformation

$$X_i = \sum \sum_j A_{ij} Y_j$$

we obtain

$$h(y_1, \cdots, y_m) = \det A f(\sum A_{1i} y_i, \cdots, \sum A_{mi} y_i).$$

We now apply these general considerations to calculate the conditional density of $T_1, \cdots, T_m$ given $N_t = m$.

Since $W_1, \cdots, W_m$ are independent exponentials with parameter μ, their joint density is

$$f(w_1, \cdots, w_m) = \mu^m e^{-\mu(w_1+\cdots+w_m)} \quad \text{for} \;\; w_i \geq 0.$$

Consider the linear transformation

$$t_1 = w_1,\; t_2 = w_1 + w_2, \cdots,\; t_m = w_1 + \cdots + w_m$$

Then the joint density of random variables $T_1, T_2, \cdots, T_{m+1}$ is

$$h(t_1, \cdots, t_{m+1}) = \mu^{m+1} e^{-\mu t_{m+1}}.$$

Therefore to calculate

$$P[A_m \mid N_t = m] = \frac{P[A_m, N_t = m]}{N_t = m},$$

where A_m denotes the event

$$A_m = \{0 < T_1 < t_1 < T_2 < t_2 < \cdots < t_{m-1} < T_m < t_m < t < T_{m+1}\}$$

we evaluate the numerator of the right hand side by noting that the condition $N_t = m$ is implied by the requirement $T_m < t_m < t < T_{m+1}$. Now

$$P[A_m] = \int_U \mu^{m+1} e^{-\mu t_{m+1}} ds_1 \cdots ds_{m+1}$$

where U is the region

$$U \;:\; (s_1, \cdots, s_{m+1}) \text{ such that } 0 < s_1 < t_1 < s_2 < t_2 < \cdots < s_m < t_m < t < s_{m+1}.$$

Carrying out the simple integration we obtain

$$P[A_m] = \mu^m t_1(t_2 - t_1) \cdots (t_m - t_{m-1}) e^{-\mu t}$$

Therefore

$$P[A_m \mid N_t = m] = t_1(t_2 - t_1) \cdots (t_m - t_{m-1}) \frac{m!}{t^m}. \tag{1.2.6}$$

To obtain the conditional joint density of $T_1, \cdots, T_m$ given $N_t = m$ we apply the differential operator $\frac{\partial^m}{\partial t_1 \cdots \partial t_m}$ to (1.2.6) to obtain

$$f_{T|N}(t_1, \cdots, t_m) = \frac{m!}{t^m}, \quad 0 \leq t_1 < t_2 < \cdots < t_m \leq t. \tag{1.2.7}$$

We deduce the following remarkable fact:

Proposition 1.2.1 *With the above notation and hypotheses, the conditional joint density of $T_1, \cdots, T_m$ given $N_t = m$ is identical with that of the order statistics of m uniform random variables from $[0, t]$.*

Proof - Follows from lemma 1.2.1 and (1.2.7). ♣.

A simple consequence of Proposition 1.2.1 is a proof of the independence of increments for a Poisson process. To do so let $t_1 < t_2 \leq t_3 < t_4$, $U = N_{t_2} - N_{t_1}$ and $V = N_{t_4} - N_{t_3}$. Then

$$\mathsf{E}[\xi^U \eta^V] = \mathsf{E}[\mathsf{E}[\xi^U \eta^V \mid N_{t_4}]]$$

From Proposition 1.2.1 we know that conditioned on $N_{t_4} = M$, the times t_1, t_2 and t_3 have the same distribution as the order statistics of three uniform random variables on $[0, t_4]$. In particular, the number of arrivals in an interval $(a, b) \subset (0, t_4)$ is proportional to the length $b - a$. Therefore the conditional expectation $\mathsf{E}[\xi^U \eta^V \mid N_{t_4}]$ is easily computed (see example **??**)

$$\mathsf{E}[\xi^U \eta^V \mid N_{t_4}] = \left[\Big(\frac{t_2 - t_1}{t_4}\Big)\xi + \Big(\frac{t_4 - t_3}{t_4}\Big)\eta + \Big(\frac{t_1 + t_3 - t_2}{t_4}\Big)\right]^M \tag{1.2.8}$$

Since for fixed t, N_t is a Poisson random variable with parameter μt, we can calculate the outer expectation to obtain

$$\begin{aligned} \mathsf{E}[\xi^U \eta^V] &= e^{-\mu[\xi(t_2 - t_1) + \eta(t_4 - t_3) + (t_1 + t_3 - t_2 - t_4)]} \\ &= e^{-\mu(t_2 - t_1)(\xi - 1)} e^{-\mu(t_4 - t_3)(\eta - 1)} \\ &= \mathsf{E}[\xi^U] \mathsf{E}[\eta^V], \end{aligned}$$

which proves the independence of increments of a Poisson process. Summarizing

Proposition 1.2.2 *The Poisson process N_t with parameter μ has the following properties:*

1. *For fixed $t > 0$, N_t is a Poisson random variable with parameter μt;*

2. *N_t is stationary (i.e., $N_{s+t} - N_s$ has the same distribution as N_t) and has independent increments;*

3. *The infinitesimal generator of N_t is given by (1.2.5).*

Property (3) of proposition 1.2.2 follows from the first two which in fact characterize Poisson processes. From the infinitesimal generator (1.2.5) one can construct the transition probabilities $P_t = e^{tA}$.

It is evident from the above arguments that the stationary nature of increments in a Poisson process was due to the fact that λ was a constant (independent of t). A generalization of the Poisson process that occurs in some applications is that of a *non-homogeneous Poisson process* where λ is no longer a constant but depends on t. More precisely one can define a non-homogeneous Poisson process with parameter $\lambda(t)$ as a counting process N_t such that

1. $N_\circ = 0$;

2. N_t has independent increments;

3. Set $m(t) = \int_\circ^t \lambda(s)ds$, then for $0 \leq s < t$ the random variable $N_t - N_s$ is Poisson with parameter $m(t) - m(s)$.

It is not difficult to show that for $h > 0$ small we have

$$P[N_{t+h} - N_t = 1] = \lambda(t)h + o(h), \quad P[N_{t+h} - N_t \geq 2] = O(h^2).$$

By a familiar mathematical device one can reduce the study non-homogeneous Poisson processes to that of the homogeneous case provided the function $\lambda(t)$ is positive everywhere and bounded. In fact define

$$m(t) = \int_\circ^t \lambda(s)ds.$$

The function $m(t)$ is a strictly increasing function of t. Now define

$$M_t = N_{m^{-1}(t)}.$$

It is not difficult to verify that M_t is a (homogeneous) Poisson process with parameter $\lambda = 1$ (see exercise 1.2.7 below.)

Example 1.2.1 In this example we show how to mathematically construct non-homogeneous Poisson processes. Let $X_1, X_2, \ldots$ be iid random variables with in $\mathbf{R}_+$, common density function f and distribution function F. Define the counting process N_t as follows: Set $N_\circ = 0$ and $N_{t+h} - N_t \geq 1$ if and only if the event

$$\{X_n \in (t, t+h],\ X_1 \leq t, X_2 \leq t, \ldots, X_{n-1} \leq t\}$$

has occurred. There is a more concrete way of describing the counting process N_t. Let $N_\circ = 0$ and set $N_t = 1$ when X_1 occurs. If X_2 occurs before X_1 then ignore it and look at X_3, and if it occurs after then increment N_t by 1 at the time that X_2 occurs. Similarly, if X_3 occurs before X_1 or X_2 does, then ignore it and look at X_4; otherwise increment N_t at the time that X_3 occurs. Continue in the obvious manner. We show that N_t is a non-homogeneous Poisson process with

$$\lambda(t) = \frac{f(t)}{1 - F(t)}.$$

Clearly, by the hypotheses on X_j's,

$$\begin{aligned}
P[N_{t+h} - N_t \geq 1] &= \sum_{n=1}^{\infty} P[X_1 \leq t, X_2 \leq t, \ldots, X_{n-1}]P[X_n \in (t, t+h]] \\
&= \sum_{n=1}^{\infty} F^{n-1}(t)[F(t+h) - F(t)] \\
&= \frac{F(t+h - F(t)}{1 - F(t)} \\
&= \frac{hf(t)}{1 - F(t)} + o(h),
\end{aligned}$$

from which the claim follows. Notice that the independence of increments implies that a non-homogeneous Poisson process is in fact Markov. In the special case where X_j's are exponential random variables with the same parameter λ, we have $\frac{f(t)}{1-F(t)} = \lambda$ and we obtain a (homogeneous) Poisson process. ♠

There is a general procedure for constructing continuous time Markov chains out of a Poisson process and a (discrete time) Markov chain. The resulting Markov chains are often considerably easier to analyze and behave somewhat like the finite state continuous time Markov chains. It is customary to refer to these processes as *Markov chains subordinated to Poisson processes*. Let Z_n be a (discrete time) Markov chain with transition matrix K, and N_t be a Poisson process with parameter μ. Let S be the state space of Z_n. We construct the continuous time Markov chain with state space S by postulating that the number of transitions in an interval $[s, s+t)$ is given by $N_{t+s} - N_s$ which has the same distribution as N_t. Given that there are n transitions in the interval $[s, s+t)$, we require the probability $P[X_{s+t} = j \mid X_s = i, N_{t+s} - N_s = n]$ to be

$$P[X_{s+t} = j \mid X_s = i, N_{t+s} - N_s = n] = K_{ij}^{(n)}.$$

Let $K^\circ = I$, then the transition probability $P_{ij}^{(t)}$ for X_t is given by

$$P_{ij}^{(t)} = \sum_{n=\circ}^{\infty} \frac{e^{-\mu t}(\mu t)^n}{n!} K_{ij}^{(n)}. \tag{1.2.9}$$

The infinitesimal generator is easily computed by differentiating (1.2.9) at $t = 0$:

$$A = \mu(-I + K). \tag{1.2.10}$$

From the Markov property of the matrix K it follows easily that the infinite series expansion of e^{tA} converges and therefore $P_t = e^{tA}$ is rigorously defined. The matrix Q of lemma 1.4.1 can also be expressed in terms of the Markov matrix K. Assuming no state is absorbing we get (see corollary 1.4.2)

$$Q_{ij} = \begin{cases} 0, & \text{if } i = j; \\ \frac{K_{ij}}{1-K_{ii}}, & \text{otherwise.} \end{cases} \tag{1.2.11}$$

Note that if (1.2.10) is satisfied then from A we obtain a continuous time Markov chain subordinate to a Poisson process.

EXERCISES

Exercise 1.2.1 *For a Poisson process N_t with parameter λ compute* $\mathsf{Cov}(N_t, N_s)$. (Use i8ndependence of increments.)

Exercise 1.2.2 *Let N_t and M_t be independent Poisson processes with parameters λ and μ. Determine which of the following are Poisson processes:*

1. $N_t + M_t$;
2. $\max(N_t, M_t)$;
3. $\min(N_t, M_t)$.

Exercise 1.2.3 *Let N_t and M_t be independent Poisson processes with parameters λ and μ. Let T_k^N (resp. T_k^M) denote the time of k^{th} arrival for the process N_t (resp. M_t). Show that*

$$P[T_1^N < T_1^M] = \frac{\lambda}{\lambda + \mu}.$$

Exercise 1.2.4 *An oscilloscope is receiving electrical impulses according to a Poisson process with parameter λ. Let the arrival times be $T_1, T_2, \ldots$. The initial amplitudes $A_1, A_2, \ldots$ of the impulses are iid random variables and each impulse decays exponentially with time so that the total impulse at time t is*

$$A(t) = \sum_{k=1}^{N_t} A_k e^{-\alpha(t-T_k)},$$

where $\alpha > 0$ is a fixed number. Show that

1. $\mathsf{E}[A(t) \mid N_t = n] = \frac{n(1-e^{-\alpha t})}{\alpha} \mathsf{E}[A_j]$;
2. $\mathsf{E}[A(t)] = \frac{\lambda(1-e^{-\alpha t})}{\alpha} \mathsf{E}[A_j]$.

Exercise 1.2.5 *Arrivals for a demonstration follow a Poisson process N_t with parameter λ, and let T_k denote the time of the k^{th} arrival. Then the total number of demonstrator hours is*

$$X(t) = \sum_{k=1}^{N_t} (t - T_k).$$

Show that

1. $\mathsf{E}[\sum_{k=1}^{N_t} T_k \mid N_t = n] = \frac{nt}{2}$

2. $\mathsf{E}[\sum_{k=1}^{N_t} T_k] = \frac{\lambda t^2}{2} = \mathsf{E}[X(t)]$.

Exercise 1.2.6 *Let $X, X_1, X_2, \ldots$ be id random variables and N_t a Poisson process independent of X_k's and with parameter λ. Set*

$$S(t) = \sum_{k=1}^{N_t} X_k.$$

Let $\Phi_Y = \mathsf{E}[e^{i\xi Y}]$ denote the characteristic function of a random variable Y. Show that the characteristic function of $S(t)$ is

$$\Phi_{S(t)}(\xi) = \Phi_{N_t}(\Phi_X(\xi)).$$

(Immitate the proof of the corresponding statement in the subsection "A Branching Process." The process $S(t)$ is called a *Compound Poisson Process.*) *In particular, if X is 0 or 1 with probabilities p and q respectively, show that the generating function for $S(t)$ is $e^{\lambda qt(\xi-1)}$.*

Exercise 1.2.7 *Assume the function $\lambda(t)$ of a non-homogeneous Poisson processes is positive everywhere and bounded. Set*

$$m(t) = \int_\circ^t \lambda(s)ds.$$

The function $m(t)$ is a strictly increasing function of t, and define

$$M_t = N_{m^{-1}(t)}.$$

Show that M_t is a (homogeneous) Poisson process with parameter $\lambda = 1$

Exercise 1.2.8 *For the two state Markov chain with transition matrix*

$$K = \begin{pmatrix} p & q \\ p & q \end{pmatrix},$$

show that the continuous time Markov chain subordinate to the Poisson process of rate μ has trasnsition matrix

$$P_t = \begin{pmatrix} p + qe^{-\mu t} & q - qe^{-\mu t} \\ p - pe^{-\mu t} & q + pe^{-\mu t} \end{pmatrix}$$

1.3 The Kolomogorov Equations

To understand the significance of the Kolmogorov equations we analyze a simple continuous time Markov chain. The method is applicable to many other situations and it is most easily demonstrated by an example. Consider a simple *pure birth process* by which we mean we have a (finite) number of organisms which independently and randomly split into two. We let X_t denote the number of organisms at time t and it is convenient to assume that $X_\circ = 1$. The law for splitting of a single organism is given by

$$P[\text{splitting in } h \text{ units of time}] = \lambda h + o(h)$$

for $h > 0$ a small real number. This implies that the probability that a single organism splits more than once in h units of time is $o(h)$. Now suppose that we have n organisms and A_j denote the event that organism number j splits (at least once) and A be the event that in h units of time there is exactly one split. Then

$$P[A] = \sum_j P[A_j] - \sum_{i<j} P[A_i \cap A_j] + \sum_{i<j<k} P[A_i \cap A_j \cap A_k] - \cdots$$

The probability of each of the events $A_i \cap A_j$, $A_i \cap A_j \cap A_k$ etc. is clearly $o(h)$ for $h > 0$ small. Therefore

$$P[A] = n\lambda h + o(h). \tag{1.3.1}$$

Note that the exact value of the terms incorporated into $o(h)$ is quite complicated. We shall see that in spite of ignoring these complicated terms, we can recover exact information about our continuous time Markov chain by using Kolmogorov forward equation. Let B be the event that in h units of time there are at least two splits, and C the event that there are no split of the n organisms. Then

$$P[B] = o(h), \quad P[C] = 1 - \lambda nh + o(h). \tag{1.3.2}$$

Equations (1.3.1) and (1.3.2) imply that the infinitesimal generator A of the continuous time Markov chain X_t is

$$A = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & 0 & \cdots \\ 0 & -2\lambda & 2\lambda & 0 & 0 & \cdots \\ 0 & 0 & -3\lambda & 3\lambda & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

For any initial distribution $\pi^\circ = (\pi_1^\circ, \pi_2^\circ, \cdots)$ let $q(t) = (q_1(t), q_2(t), \cdots)$ be the row vector $q(t) = \pi^\circ P_t$ which describes the distribution of states at time t. In fact,

$$q_k(t) = P[X_t = k \mid X_\circ = \pi^\circ].$$

Thus a basic problem about the Markov chain X_t is the calculation of $\mathsf{E}[X_t]$ or more generally of the generating function

$$\mathsf{F}_X(t,\xi) = \mathsf{E}[\xi^{X_t}] = \sum_{k=1} P[X_t = k]\xi^k.$$

We now use Kolomogorov's forward equation to solve this problem. From (1.1.4) we obtain

$$\frac{dq(t)}{dt} = \pi^\circ \frac{dP_t}{dt} = \pi^\circ P_t A = q(t)A.$$

Expanding the $q_t A$ we obtain

$$\frac{dq_k(t)}{dt} = -k\lambda q_k(t) + (k-1)\lambda q_{k-1}(t). \tag{1.3.3}$$

With a little straightforward book-keeping, one shows that (1.3.3) implies

$$\frac{\partial \mathsf{F}_X}{\partial t} = \lambda(\xi^2 - \xi)\frac{\partial \mathsf{F}_X}{\partial \xi}. \tag{1.3.4}$$

The fact that F_X a linear partial differential equation makes the calculation of $\mathsf{E}[X_t]$ very simple. In fact, since $\mathsf{E}[X_t] = \frac{\partial \mathsf{F}_X}{\partial \xi}$ evaluated at $\xi = 1^-$, we differentiate both sides of (1.3.4) with respect to ξ, change the order of differentiation relative to ξ and t on the left side, and set $\xi = 1$ to obtain

$$\frac{d\mathsf{E}[X_t]}{dt} = \lambda \mathsf{E}[X_t]. \tag{1.3.5}$$

The solution to this ordinary differential equation is $Ce^{\lambda t}$ and the constant C is determined by the initial condition $X_\circ = 1$ to yield

$$\mathsf{E}[X_t] = e^{\lambda t}.$$

The partial differential equation (1.3.4) tells us considerably more than just the expectation of X_t. The basic theory of a single linear first order partial differential equation is well understood. Recall that the solution to a first order ordinary differential equation is uniquely determined by specifying one initial condition. Roughly speaking, the solution to a linear first order partial differential equation in two variables is uniquely determined by specifying a function of one variable. Let us see how this works for our equation (1.3.4). For a function $g(s)$ of a real variable s we want to substitute for s a function of t and ξ such that (1.3.4) is necessarily valid regardless of the choice of g. If for s we substitute $\lambda t + \phi(\xi)$, then by the chain rule

$$\frac{\partial g(\lambda t + \phi(\xi))}{\partial t} = \lambda g'(\lambda t + \phi(\xi)), \quad \frac{\partial g(\lambda t + \phi(\xi))}{\partial \xi} = \phi'(\xi) g'(\lambda t + \phi(\xi)).$$

where g' denote the derivative of the function g. Therefore if ϕ is such that $\phi'(\xi) = \frac{1}{\xi^2 - \xi}$, then, regardless of what function we take for g, equation (1.3.4) is satisfied by $g(\lambda t + \phi(\xi))$. There is an obvious choice for ϕ (by solving the differential equation $\phi'(\xi) = \frac{1}{\xi^2 - \xi}$) namely the function

$$\phi(\xi) = \log \frac{1 - \xi}{\xi},$$

for $0 < \xi < 1$. Now we incorporate the initial condition $X_\circ = 1$ which in terms of the generating function F_X means $\mathsf{F}_X(0, \xi) = \xi$. In terms of g this translates into

$$g(\log \frac{1 - \xi}{\xi}) = \xi.$$

That is, g should be the inverse to the mapping $\xi \rightarrow \log \frac{1-\xi}{\xi}$. It is easy to see that

$$g(s) = \frac{1}{1 + e^s}$$

is the required function. Thus we obtain the expression

$$\mathsf{F}_X(t, \xi) = \frac{\xi}{\xi + (1 - \xi) e^{\lambda t}} \tag{1.3.6}$$

for the probability generating function of X_t. If we change the initial condition to $X_\circ = N$, then the generating function becomes

$$\mathsf{F}_X(\xi, t) = \frac{\xi^N}{[\xi + (1-\xi)e^{\lambda t}]^N}.$$

From this we deduce that for $j \geq N$

$$P_{Nj}^{(t)} = e^{-N\lambda t}\binom{j-1}{N-1}(1 - e^{-\lambda t})^{j-N}$$

for the transition probabilities. This formula would have been quite difficult to derive without using the Kolmogorov differential equation. The method of derivation of (1.3.6) is remarkable and instructive. We made essential use of the Kolmogorov forward equation in obtaining a linear first order partial differential equation for the probability generating function. This was possible because we have an infinite number of states and the coefficients of $q_k(t)$ and $q_{k-1}(t)$ in (1.3.3) were linear in k. Different kind of dependence of q_k on k would have resulted in other equation where the analysis more difficult or simpler. At any rate the fact that we have an explicit differential equation for the generating function gave us a fundamental new tool for understanding it. In the exercises this method will be further demonstrated.

Example 1.3.1 We continue with example 1.1.2. We set $q(t) = (q_\circ(t), q_1(t), \ldots)$ where

$$q(t) = \pi^\circ P_t, \quad q_k(t) = P[X_t = k \mid X_\circ \pi^\circ],$$

and form the generating function $\mathsf{F}_X(\xi, t) = \sum_{k=0}^\infty q_k(t)\xi^k$. It is convenient to set $q_{-1} \equiv 0$. Then Kolmogorov's forward equation implies that F_X satisfies the differential equation

$$\xi \frac{d\mathsf{F}_X}{dt} = (\lambda\xi^2 - (\lambda + \mu)\xi + \mu)\mathsf{F}_X + \mu q_\circ, \tag{1.3.7}$$

which is a linear ordinary differential equation of first order. However we have the additional term $\mu q_\circ$ since the first column of the infinitesimal generator is different from other columns. Now $q_\circ$ is the probability that there are no customers in queue at time t. Assuming that time $t = 0$ there were no customers in queue, $q_\circ(t)$ is the probability that the number of customers arrived is equal to the number of customers serviced. Therefore

$$q_\circ(t) = e^{-(\lambda+\mu)t}\sum_{l=0}^{\infty}\frac{(\lambda\mu)^k t^{2k}}{(k!)^2}.$$

In principle an equation of the form (1.3.7), with a known function $q_\circ$, can be solved by using Laplace transforms, a subject which we will discuss in connection with renewal theory later. We will not pursue the solution of this equation any further here. ♠

Example 1.3.2 The birth process described above can be easily generalized to a birth-death process by introducing a positive parameter $\mu > 0$ and replacing equations (1.3.1) and (1.3.2) with the requirement

$$P[X_{t+h} = n + a \mid X_t = n] = \begin{cases} n\mu h + o(h), & \text{if } a = -1; \\ n\lambda h + o(h), & \text{if } a = 1; \\ o(h), & \text{if } |a| > 1. \end{cases} \tag{1.3.8}$$

The probability generating function for X_t can be calculated by an argument similar to that of pure birth process given above and is delegated to exercise 1.3.5. It is shown there that for $\lambda \neq \mu$

$$\mathsf{F}_X(\xi, t) = \left(\frac{\mu(1-\xi) - (\mu - \lambda\xi)e^{-t(\lambda-\mu)}}{\lambda(1-\xi) - (\mu - \lambda\xi)e^{-t(\lambda-\mu)}}\right)^N. \tag{1.3.9}$$

where N is given by the initial condition $X_\circ = N$. For $\lambda = \mu$ this expression simplifies to

$$\mathsf{F}_X(\xi, t) = \left(\frac{\mu t(1-\xi) + \xi}{\mu t(1-\xi) + 1}\right)^N. \tag{1.3.10}$$

From this it follows easily that $\mathsf{E}[X_t] = Ne^{(\lambda-\mu)t}$. Let $\zeta(t)$ denote the probability that the population is extinct at time t, i.e., $\zeta(t) = P[X_t = 0 \mid X_\circ = N]$. Therefore $\zeta(t)$ is the constant term, as a function of ξ for fixed t, of the generating function $\mathsf{F}_X(\xi, t)$. In other words, $\zeta(t) = \mathsf{F}_X(0, t)$ and we obtain

$$\lim_{t\to\infty} = \begin{cases} 1, & \text{if } \mu \geq \lambda; \\ \frac{\mu^N}{\lambda^N}, & \text{if } \mu < \lambda; \end{cases} \tag{1.3.11}$$

for the probability of eventual extinction. ♠

EXERCISES

Exercise 1.3.1 ($M/M/1$ queue) *A server can service only one customer at a time and the arriving customers form a queue according to order of arrival. Consider the continuous time Markov chain where the length of the queue is the state space, the time between consecutive arrivals is exponential with parameter μ and the time of service is exponential with parameter λ. Show that the matrix $Q = (Q_{ij})$ of lemma 1.4.1 is*

$$Q_{ij} = \begin{cases} \frac{\mu}{\mu+\lambda}, & \text{if } j = i+1; \\ \frac{\lambda}{\mu+\lambda}, & \text{if } j = i-1; \\ 0, & \text{otherwise.} \end{cases}$$

Exercise 1.3.2 *The probability that a central networking system receives a call in the time period $(t, t+h)$ is $\lambda h + o(h)$ and calls are received independently. The service time for the calls are independent and identically distributed each according to the exponential random variable with parameter μ. The service times are also assumed independent of the incoming calls. Consider the Markov process $X(t)$ with state space the number of calls being processed by the server. Show that the infinitesimal generator of the Markov process is the infinite matrix*

$$\begin{pmatrix} -\lambda & \lambda & 0 & 0 & 0 & \cdots \\ \mu & -(\lambda+\mu) & \lambda & 0 & 0 & \cdots \\ 0 & 2\mu & -(\lambda+2\mu) & \lambda & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Let $F_X(\xi, t) = \mathsf{E}(\xi^{X(t)})$ denote the generating function of the Markov process. Show that F_X satisfies the differential equation

$$\frac{\partial F_X}{\partial t} = (1-\xi)[-\lambda F_X + \mu \frac{\partial F_X}{\partial \xi}].$$

Assuming that $X(0) = m$, use the differential equation to show that

$$\mathsf{E}[X(t)] = \frac{\lambda}{\mu}(1 - e^{-\mu t}) + m e^{-\mu t}.$$

Exercise 1.3.3 *(Continuation of exercise 1.3.2) - With the same notation as exercise 1.3.2, show that the substitution*

$$F_X(\xi,t) = e^{-\frac{\lambda(1-\xi)}{\mu}} G(\xi,t)$$

gets rid of the term involving F_X on the right hand side of the differential equation for F_X. More precisely, it transforms the differential equation for F_X into

$$\frac{\partial G}{\partial t} = \mu(1-\xi)\frac{\partial G}{\partial \xi}.$$

Can you give a general approach for solving this differential equation? Verify that

$$F_X(\xi,t) = e^{-\lambda(1-\xi)(1-e^{-\mu t})/\mu}[1-(1-\xi)e^{-\mu t}]^m$$

is the desired solution to equation for F_X.

Exercise 1.3.4 *The velocities V_t of particles in a quantized field are assumed to take only discrete values $n+\frac{1}{2}$, $n \in \mathbf{Z}_+$. Under the influence of the field and mutual interactions their velocities can change by at most one unit and the probabilities for transitions are given by*

$$P[V_{t+h} = m + \frac{1}{2} \mid V_t = n + \frac{1}{2}] = \begin{cases} (n+\frac{1}{2})h + o(h), & \textit{if } m = n+1; \\ 1-(2n+1)h + o(h), & \textit{if } m = n; \\ (n-\frac{1}{2})h + o(h), & \textit{if } m = n-1. \end{cases}$$

Let $\mathsf{F}_V(\xi,t)$ be the probability generating function

$$\mathsf{F}_V(\xi,t) = \sum_{n=0}^{\infty} P[V_t = n + \frac{1}{2}]\xi^n.$$

Show that

$$\frac{\partial \mathsf{F}_V}{\partial t} = (1-\xi)^2 \frac{\partial \mathsf{F}_V}{\partial s} - (1-\xi)\mathsf{F}_X.$$

Assuming that $V_\circ = 0$ deduce that

$$\mathsf{F}_V(\xi,t) = \frac{1}{1+t-\xi t}.$$

Exercise 1.3.5 *Consider the birth-death process of example 1.3.2*

1. *Show that the generating function* $\mathsf{F}_X(\xi, t) = \mathsf{E}[\xi^{X_t}]$ *satisfies the partial differential equation*

$$\frac{\partial \mathsf{F}_X}{\partial t} = (\lambda\xi - \mu)(\xi - 1)\frac{\partial \mathsf{F}_X}{\partial \xi},$$

   with the initial condition $\mathsf{F}_X(\xi, 0) = \xi^N$.

2. *Deduce the validity of (1.3.8) and (1.3.9).*

3. *Show that*

$$\mathsf{Var}[X_t] = N\frac{\lambda + \mu}{\lambda - \mu}e^{(\lambda-\mu)t}(e^{(\lambda-\mu)t} - 1).$$

4. *Let* $N = 1$ *and* Z *denote the time of the extinction of the process. Show that for* $\lambda = \mu$, $\mathsf{E}[Z] = \infty$.

5. *Show that* $\mu > \lambda$ *we have*

$$\mathsf{E}[Z \mid Z < \infty] = \frac{1}{\lambda}\log\frac{\mu}{\mu - \lambda}.$$

6. *Show that* $\mu < \lambda$ *we have*

$$\mathsf{E}[Z \mid Z < \infty] = \frac{1}{\mu}\log\frac{\mu}{\mu - \lambda}.$$

(Use the fact that

$$\mathsf{E}[Z] = \int_\circ^\infty P[Z > t]dt = \int_\circ^\infty [1 - \mathsf{F}_X(0, t)]dt,$$

to calculate $\mathsf{E}[Z \mid Z < \infty]$.)

1.4 Discrete vs. Continuous Time Markov Chains

In this subsection we show how to assign a discrete time Markov chain to one with continuous time, and how to construct continuous time Markov chains from a discrete time one. We have already introduced the notions of Markov and stopping times for for Markov chains, and we can easily extend them to continuous time Markov chains. Intuitively a *Markov time* for the (possibly continuous time) Markov chain is a random variable T such that the event $[T \leq t]$ does not depend on X_s for $s > t$. Thus a Markov time T has the property that if $T(\omega) = t$ then $T(\omega') = t$ for all paths which are identical with ω for $s \leq t$. For instance, for a Markov chain X_l with state space $\mathbf{Z}$ and $X_\circ = 0$ let T be the first hitting time of state $1 \in \mathbf{Z}$. Then T is a Markov time. If T is Markov time for the continuous time Markov chain X_t, the fundamental property of Markov time, generally called *Strong Markov Property*, is

$$P[X_{T+s} = j \mid X_t, t \leq T] = P^{(s)}_{X_T j} \tag{1.4.1}$$

This reduces to the Markov property if we take T to be a constant. To understand the meaning of equation (1.4.1), consider $\Omega_u = \{\omega \mid T(\omega) = u\}$ where $u \in \mathbf{R}_+$ is any fixed positive real number. Then the left hand side of (1.4.1) is the conditional probability of the set of paths ω that after s units of time are in state j given Ω_u and X_t for $t \leq u = T(\omega)$. The right hand side states that the information X_t for $t \leq u$ is not relevant as long we know the states for which $T(\omega) = u$, and this probability is the probability of the paths which after s units of time are in state j assuming at time 0 they were in a state determined by $T = u$. One can also loosely think of the strong Markov property as allowing one to reparametrize paths so that all the paths will satisfy $T(\omega) = u$ at the same constant time T and then the standard Markov property will be applicable. Examples that we encounter will clarify the meaning and significance of this concept. The validity of (1.4.1) is quite intuitive, and one can be convinced of its validity by looking at the set of paths with the required properties and using the Markov property. It is sometimes useful to make use of a slightly more general version of the strong Markov property where a function of the Markov time is introduced. Rather than stating a general theorem, its validity in the context where it is used will be clear.

The notation $\mathsf{E}_i[Z]$ where the random variable Z is a function of of the continuous time Markov chain X_t means that we are calculating conditional expectation conditioned on $X_\circ = i$. Naturally, one may replace the subscript $_i$ by a random variable to accommodate a different conditional expectation. Of course, instead of a subscript one may write the conditioning in the usual manner $\mathsf{E}[\star \mid \star]$. The strong Markov property in the context of conditional expectations implies

$$\mathsf{E}[g(X_{T+s}) \mid X_u, u \leq T] = \mathsf{E}_{X_T}[g] = \mathsf{E}[\sum_{j\in S} P^{(s)}_{X_T j} g(j)]. \tag{1.4.2}$$

The Markov property implies that transitions between states follows a memoryless random variable. It is worthwhile to try to understand this statement more clearly. Let $X_\circ = i$ and define the random variable Yas

$$Y(\omega) = \inf\{t \mid \omega(t) \neq i\}$$

Then Y is a Markov time. The assumption (1.1.8) implies that except for a set of paths of probability 0, $Y(\omega) > 0$, and by the right continuity assumption, the infimum is actually achieved. The strong Markov property implies that the random variable Y is memoryless in the sense that

$$P[Y \geq t + s \mid Y > s] = P[Y \geq t \mid X_\circ = i].$$

It is a standard result in elementary probability that the only memoryless continuous random variables are exponentials. Recall that the distribution function for the exponential random variable T with parameter λ is given by

$$P[T < t] = \begin{cases} 1 - e^{-\lambda t}, & \text{if } t\ 0; \\ 0, & \text{if } t \leq 0. \end{cases}$$

The mean and variance of T are $\frac{1}{\lambda}$. Therefore

$$P[Y \geq t \mid X_\circ = i] = P[X_s = i \text{ for } s \in [0, t) \mid X_\circ = i] = e^{-\lambda_i t}. \tag{1.4.3}$$

This equation is compatible with (1.1.8). Note that for an absorbing state i we have $\lambda_i = 0$.

From a continuous time Markov chain one can construct a (discrete time) Markov chain. Let us assume $X_\circ = i \in S$. A simple and not so useful way is to define the transition matrix P of the Markov chain as $P^{(1)}_{ij}$. A more useful approach is to let T_n be the time of the n^{th} transition. Thus $T_1(\omega) = s > 0$ means that there is $j \in S$, $j \neq i$, such that

$$\omega(t) = \begin{cases} i & \text{for } t < s; \\ j & \text{for } s = t. \end{cases}$$

T_1 is a stopping time if we assume that i is not an absorbing state. We define Q_{ij} to be the probability of the set of paths that at time 0 are in state i and at the time of the first transition they move to state j. Therefore

$$Q_{kk} = 0, \quad \text{and} \quad \sum_{j \neq i} Q_{ij} = 1.$$

Let $W_n = T_{n+1} - T_n$ denote the time elapsed between the n^{th} and $(n+1)^{st}$ transitions. We define a Markov chain $Z_\circ = X_\circ, Z_2, \cdots$ by setting $Z_n = X_{T_n}$. Note that the strong Markov property for X_t is used in ensuring that $Z_\circ, Z_1, Z_2, \cdots$ is a Markov chain since transitions occur at different times on different paths. The following lemma clarifies the transition matrix of the Markov chain Z_n and sheds light on the transition matrices P_t.

Lemma 1.4.1 *For a non-absorbing state k we have*

$$P[Z_{n+1} = j, W_n > u \mid Z_\circ = i_\circ, Z_1, \cdots, Z_n = k, T_1, \cdots, T_n] = Q_{kj} e^{-\lambda_k u}.$$

Furthermore $Q_{kk} = 0$, $Q_{kj} \geq 0$ and $\sum_j Q_{kj} = 1$, so that $Q = (Q_{kj})$ is the transition matrix for the Markov chain Z_n. For an absorbing state k, $\lambda_k = 0$ and $Q_{kj} = \delta_{kj}$.

Proof - Clearly the left hand side of the equation can be written in the form

$$P[X_{T_n+W_n} = j, W_n > u \mid X_{T_n} = k, X_t, t \leq T].$$

By the strong Markov property[3] we can rewrite this as

$$P[X_{T_n+W_n} = j, W_n > u \mid X_{T_n} = k, X_t, t \leq T] = P[X_{W_\circ} = j, W_\circ > u \mid X_\circ = k].$$

Right hand side of the above equation can be written as

$$P[X_{W_\circ} = j \mid W_\circ > u, X_\circ = k] P[W_\circ > u \mid X_\circ = k].$$

[3]We are using a slightly more general version than the statement (1.4.1), but its validity is equally clear.

We have

$$\begin{aligned} P[X_{W_\circ} = j \mid W_\circ > u, X_\circ = k] &= P[X_{W_\circ} = j \mid X_s = k \text{ for } s \leq u] \\ &= P[X_{u+W_\circ} = j \mid X_u = k] \\ &= P[X_{W_\circ} = j \mid X_\circ = k]. \end{aligned}$$

The quantity $P[X_{W_\circ} = j \mid X_\circ = k]$ is independent of u and we denote it by Q_{kj}. Combining this with (1.4.3) (exponential character of elapsed time W_n between consecutive transitions) we obtain the desired formula. The validity of stated properties of Q_{kj} is immediate. ♣

A immediate corollary of lemma 1.4.1 is that it allows to fill in the gap in the proof of lemma 1.1.3.

Corollary 1.4.1 *Let $i \neq j \in S$, then either $P_{ij}^{(t)} > 0$ for all $t > 0$ or it vanishes identically.*

Proof - If $P_{ij}^{(t)} > 0$ for some t, then $Q_{ij} > 0$, and it follows that for all $t > 0$ $P_{ij}^{(t)} > 0$. ♣

This process of assigning a Markov chain to a continuous time Markov chain can be reversed to obtain (infinitely many) continuous time Markov chains from a discrete time one. In fact, for every $j \in S$ let $\lambda_j > 0$ be a positive real number. Now given a Markov chain Z_n with state space S and transition matrix Q, let W_j be an exponential random variable with parameter λ_j. If j is not an absorbing state, then the first transition out of j happens at time $s > t$ with probability $e^{-\lambda_j t}$ and once the transition occurs the probability of hitting state k is

$$\frac{Q_{jk}}{\sum_{i\neq j} Q_{ji}}.$$

Lemma 1.4.1 does not give direct and adequate information about the behavior of transition probabilities. However, combining it with the strong Markov property yields an important integral equation satisfied by the transition probabilities.

Lemma 1.4.2 *Assume X_t has no absorbing states. For $i, j \in S$, the transition probabilities $P_{ij}^{(t)}$ satisfy*

$$P_{ij}^{(t)} = e^{-\lambda_i t}\delta_{ij} + \lambda_i \int_\circ^t e^{-\lambda_i s} \sum_k Q_{ik} P_{kj}^{(t-s)} ds$$

Proof - We may assume i is not an absorbing state. Let T_1 be the time of first transition. Then trivially

$$P_{ij}^{(t)} = P[X_t = j, T_1 > t \mid X_\circ = i] + P[X_t = j, T_1 \leq t \mid X_\circ = i].$$

The term containing $T_1 > t$ can be written as

$$P[X_t = j, T_1 > t \mid X_\circ = i] = \delta_{ij} P[T_1 > t \mid X_\circ = i] = e^{-\lambda_i t}\delta_{ij}.$$

By the strong Markov property the second term becomes

$$P[X_t = j, T_1 \leq t \mid X_\circ = i] = \int_\circ^t P[T_1 = s < t \mid X_\circ = i] P_{X_s j}^{(t-s)} ds.$$

To make a substitution for $P[T_1 = s < t \mid X_\circ = i]$ from lemma 1.4.1 we have to differentiate[4] the the expression $1 - Q_{ik}e^{-\lambda_i s}$ with respect to s. We obtain

$$P[X_t = j, T_1 \leq t \mid X_\circ = i] = \lambda_i \int_\circ^t \sum_k e^{-\lambda_i s} Q_{ik} P_{kj}^{(t-s)} ds,$$

from which the required result follows. ♣

An application of this lemma will be given in the next subsection. Integral equations of the general form given in lemma 1.4.2 occur frequently in probability. Such equations generally result from conditioning on a Markov time together with the strong Markov property.

A consequence of lemma 1.4.2 is an explicit expression for the infinitesimal generator of a continuous time Markov chain.

Corollary 1.4.2 *With the notation of lemma 1.4.2, the infinitesimal generator of the continuous time Markov chain X_t is*

$$A_{ij} = \begin{cases} -\lambda_i, & \text{if } i = j; \\ \lambda_i Q_{ij}, & \text{if } i \neq j. \end{cases}$$

Proof - Making the change of variable $s = t - u$ in the integral in lemma 1.4.2 we obtain

$$P_{ij}^{(t)} = e^{-\lambda_i t}\delta_{ij} + \lambda_i \int_\circ^t e^{-\lambda_i u} \sum_k Q_{ik} P_{kj}^{(u)} du.$$

[4]This is like the connection between the density function and distribution function of a random variable.

Differentiating with respect to t yields

$$\frac{dP_{ij}^{(t)}}{dt} = -\lambda_i e^{-\lambda_i t}\delta_{ij} + \lambda_i e^{-\lambda_i t}\sum_k Q_{ik}P_{kj}^{(t)}.$$

Taking $\lim_{t\rightarrow 0^+}$ we obtain the desired result. ♣

1.5 Brownian Motion

Brownian motion is the most basic example of a Markov process with continuous state space and continuous time. In order to motivate some of the develpments it may be useful to give an intuitive description of Brownian motion based on random walks on $\mathbf{Z}$. In this subsection we give an intuitive approach to Brownian motion and show how certain quantities of interest can be practically calculated. In particular, we give heuristic but not frivolous arguments for extending some of the properties of random walks and Markov chains to Brownian motion.

Recall that if $X_1, X_2, \cdots$ is a sequence of iid random variables with values in $\mathbf{Z}$, then $S_\circ = 0, S_1 = X_1, \cdots$, and $S_n = S_{n-1} + X_n$, $\cdots$, is the general random walk on $\mathbf{Z}$. Let us assume that $\mathsf{E}[X_i] = 0$ and $\mathsf{Var}[X_i] = \sigma^2 < \infty$, or even that X_i takes only finitely many values. The final result which is obtained by a limiting process will be practically independent of which random walk, subject to the stated conditions on mean and variance, we start with, and in particular we may even start with the simple symmetric random walk. For $t > 0$ a real number define

$$Z_t^{(n)} = \frac{1}{\sqrt{n}} S_{[nt]},$$

where $[x]$ denotes the largest integer not exceeding x. It is clear that

$$\mathsf{E}[Z_t^{(n)}] = 0, \quad \mathsf{Var}[Z_t^{(n)}] = \frac{[nt]\sigma^2}{n} \simeq t\sigma^2, \tag{1.5.1}$$

where the approximation $\simeq$ is valid for large n. The interesting point is that with re-scaling by $\frac{1}{\sqrt{n}}$, the variance becomes approximately independent of n for large n. To make reasonable paths out of those for the random walk on $\mathbf{Z}$ in the limit of large n, we further rescale the paths of the random walk S_{nt} in the time direction by $\frac{1}{n}$. This means that if we fix a positive integer n

and take for instance $t = 1$, then a path ω between 0 and n will be squeezed in the horizontal (time) direction to the interval $[0, 1]$ and the values will be multiplied by $\frac{1}{\sqrt{n}}$. The resulting path will still consist of broken line segments where the points of nonlinearity (or non-differentiability) occur at $\frac{k}{n}$, $k = 1, 2, 3, \cdots$. At any rate since all the paths are continuous, we may surmise that the path space for $\lim_{n\to\infty}$ is the space $\Omega = \mathcal{C}_{x_\circ}[0, \infty)$ of continuous function on $[0, \infty)$ and we may require $\omega(0) = x_\circ$, some fixed number. Since in the simple symmetric random walk, a path is just as likely to up as down we expect, the same is true of the paths in the Brownian motion. A differentiable path on the other hand has a definite preference at each point, namely, the direction of the tangent. Therefore it is reasonable to expect that with probability 1 the paths in Brownian are nowhere differentiable in spite of the fact that we have not yet said anything about how probabilities should be assigned to the appropriate subsets of Ω. The assignment of probabilities is the key issue in defining Brownian motion.

Let $0 < t_1 < t_2 < \cdots < t_m$ and we want to see what we can say about the joint distribution of $(Z^{(n)}_{t_1}, Z^{(n)}_{t_2} - Z^{(n)}_{t_1}, \cdots, Z^{(n)}_{t_m} - Z^{(n)}_{t_{m-1}})$. Note that these random variables are independent while $Z^{(n)}_{t_1}, Z^{(n)}_{t_2}, \cdots$ are not. By the central limit theorem, for n sufficiently large,

$$Z^{(n)}_{t_k} - Z^{(n)}_{t_{k-1}} = \frac{1}{\sqrt{n}}\Big(S_{[nt_k]} - S_{[tn_{k-1}]}\Big)$$

is approximately normal with mean 0 and variance $(t_k - t_{k-1})\sigma^2$. We assume that taking the limit of $n \to \infty$ the process $Z^{(n)}_t$ tends to a limit. Of course this requires specifying the sense in which convergence takes place and proof, but because of the applicability of the central limit theorem we assign probabilities to sets of paths accordingly without going through a convergence argument. More precisely, to the set of paths, starting at 0 at time 0, which are in the open subset $B \subset \mathbf{R}$ at time t, it is natural to assign the probability

$$P[Z_t \in B] = \frac{1}{\sqrt{2\pi t}\sigma}\int_B e^{-\frac{u^2}{2t\sigma^2}}du. \tag{1.5.2}$$

In view of independence of Z_{t_1} and $Z_{t_2} - Z_{t_1}$, the probability that $Z_{t_1} \in (a_1, b_1)$ and $Z_{t_2} - Z_{t_1} \in (a_2, b_2)$, is

$$\frac{1}{2\pi\sigma^2\sqrt{t_1(t_2 - t_1)}}\int_{a_1}^{b_1}\int_{a_2}^{b_2} e^{-\frac{u_1^2}{2\sigma^2 t_1}} e^{-\frac{u_2^2}{2\sigma^2(t_2-t_1)}} du_1 du_2.$$

Note that we are evaluating the probability of the event $[Z_{t_1} \in (a_1, b_1),\ Z_{t_2} - Z_{t_1} \in (a_2, b_2)]$ and not $[Z_{t_1} \in (a_1, b_1),\ Z_{t_2} \in (a_2, b_2)]$ since the random variables Z_{t_1} and Z_{t_2} are not independent. This formula extends to probability of any finite number of increments. In fact, for $0 < t_1 < t_2 < \cdots < t_k$ the joint density function for $(Z_{t_1},\ Z_{t_2} - Z_{t_1}, \cdots,\ Z_{t_k} - Z_{t_{k-1}})$ is the product

$$\frac{e^{-\frac{u_1^2}{2\sigma^2 t_1}} e^{-\frac{u_2^2}{2\sigma^2(t_2-t_1)}} \cdots e^{-\frac{u_k^2}{2\sigma^2(t_k-t_{k-1})}}}{\sigma^k \sqrt{(2\pi)^k t_1(t_2-t_1)\cdots(t_k-t_{k-1})}}$$

One refers to the property of independence of $(Z_{t_1},\ Z_{t_2} - Z_{t_1}, \cdots,\ Z_{t_k} - Z_{t_{k-1}})$ as *independence of increments*. For future reference and economy of notation we introduce

$$p_t(x;\sigma) = \frac{1}{\sqrt{2\pi t}\sigma} e^{-\frac{x^2}{2t\sigma^2}}. \tag{1.5.3}$$

For $\sigma = 1$ we simply write $p_t(x)$ for $p_t(x;\sigma)$.

For both discrete and continuous time Markov chains the transition probabilities were give by matrices P_t. Here the transition probabilities are encoded in the Gaussian density function $p_t(x;\sigma)$. It is easier to introduce the analogue of P_t for Brownian motion if we look at the dual picture where the action of the semi-group P_t on functions on the state space is described. Just as in the case of continuous time Markov chains we set

$$(P_t\psi)(x) = \mathsf{E}[\psi(Z_t) \mid Z_\circ = x] = \int_{-\infty}^{\infty} \psi(y) p_t(x-y;\sigma)dy, \tag{1.5.4}$$

which is completely analogous to (**??**). The operators P_t (acting on whatever the appropriate function space is) still have the semi-group property $P_{s+t} = P_s P_t$. In view of (1.5.4) the semi-group is equivalent to the statement

$$\int_{-\infty}^{\infty} p_s(y-z;\sigma) p_t(x-y;\sigma)dy = p_t(x-z;\sigma). \tag{1.5.5}$$

Perhaps the simplest way to see the validity of (1.5.5) is by making use of Fourier analysis which transforms convolutions into products as explained earlier in subsection (XXXX). It is a straightforward calculation that

$$\int_{-\infty}^{\infty} e^{-i\lambda x} \frac{e^{-\frac{x^2}{2t\sigma^2}}}{\sqrt{2\pi t}\sigma} dx = \frac{1}{\pi} e^{-\frac{\lambda^2\sigma^2 t}{2}}. \tag{1.5.6}$$

From (**??**) and (1.5.6), the desired relation (1.5.5) and the semi-group property follow. An important feature of continuous time Markov chains is that P_t satisfied the the Kolmogorov forward and backward equations. In view of the semi-group property the same is true for Brownian motion and we will explain in example 1.5.2 below what the infinitesimal generator of Brownian motion is. With some of the fundamental definitions of Brownian motion in place we now calculate some quantities of interest.

Example 1.5.1 Since the random variables Z_s and Z_t are dependent, it is reasonable to calculate their covaraince. Assume $s < t$, then we have

$$\begin{array}{rcl} \mathsf{Cov}(Z_s, Z_t) & = & \mathsf{Cov}(Z_s, Z_t - Z_s + Z_s) \\ \text{(By independence of increments)} & = & \mathsf{Cov}(Z_s, Z_s) \\ & = & s\sigma^2. \end{array}$$

This may appear counter-intuitive at first sight since one expects Z_s and Z_t to become more independent as $t - s$ increases while the covariance depends only on $\min(s,t) = s$. However, if we divide $\mathsf{Cov}(Z_s, Z_t)$ by $\sqrt{\mathsf{Var}[Z_s]\mathsf{Var}[Z_t]}$ we see that the correlation tends to 0 as t increases for fixed s. ♠

Example 1.5.2 One of the essential features of continuous time Markov chains was the existence of the infinitesimal generator. In this example we derive a formula for the infinitesimal generator of Brownian motion. For a function ψ on the state space $\mathbf{R}$, the action of the semi-group P_t is given by (1.5.4). We set $u_\psi(t,x) = u(t,x) = P_t\psi$. The Gaussian $p_t(x;\sigma)$ has variance $t\sigma^2$ and therefore it tends to the δ-function supported at the origin as $t \to 0$, i.e.,

$$\lim_{t\to 0}(P_t\psi)(x) = \lim_{t\to 0} u(t,x) = \psi(x).$$

It is straightforward to verify that

$$\frac{\partial u}{\partial t} = \frac{\sigma^2}{2}\frac{\partial^2 u}{\partial x^2}. \tag{1.5.7}$$

Therefore from the validity of Kolomogorov's backward equation ($\frac{dP_t}{dt} = AP_t$) we conclude that the infinitesimal generator of Brownian motion is given by

$$A = \frac{\sigma^2}{2}\frac{d^2}{dx^2}. \tag{1.5.8}$$

Thus the matrix A is now replaced by a differential operator. ♠

Example 1.5.3 The notion of hitting time of a state played an important role in our discussion of Markov chains. In this example we calculate the density function for hitting time of a state $a \in \mathbf{R}$ in Brownian motion. The trick is to look at the identity

$$P[Z_t > a] = P[Z_t > a \mid T_a \leq t]P[T_a \leq t] + P[Z_t > a \mid T_a > t]P[T_a > t].$$

Clearly the second term on the right hand side vanishes, and by symmetry

$$P[Z_t > a \mid T_a < t] = \frac{1}{2}.$$

Therefore

$$P[T_a < t] = 2P[Z_t > a]. \tag{1.5.9}$$

The right hand side is easily computable and we obtain

$$P[T_a < t] = \frac{2}{\sqrt{2\pi t}\sigma} \int_a^\infty e^{-\frac{x^2}{2t\sigma^2}} dx = \frac{2}{\sqrt{2\pi}} \int_{\frac{a}{\sqrt{t}\sigma}}^\infty e^{-\frac{u^2}{2}} du.$$

The density function for T_a is obtained by differentiating this expression with respect to t:

$$f_{T_a}(t) = \frac{a}{\sqrt{2\pi}\sigma} \frac{1}{t\sqrt{t}} e^{-\frac{a^2}{2t\sigma^2}}. \tag{1.5.10}$$

Since $tf_{T_a}(t) \sim c\frac{1}{\sqrt{t}}$ as $t \to \infty$ for a non-zero constant c, we obtain

$$\mathsf{E}[T_a] = \infty, \tag{1.5.11}$$

which is similar to the case of simple symmetric random walk on $\mathbf{Z}$. ♠

The reflection principle which we introduced in connection with the simple symmetric random walk on $\mathbf{Z}$ and used to prove the arc-sine law is also valid for Brownian motion. However we have already accumulated enough information to prove the arc-sine law for Brownian motion without reference to the reflection principle. This is the substance of the following example:

Example 1.5.4 For Brownian motion with $Z_\circ = 0$ the event that it crosses the line $-a$, where $a > 0$, between times 0 and t is identical with the event $[T_{-a} < t]$ and by symmetry it has the same probability as $P[T_a < t]$. We calculated this latter quantity in example 1.5.3. Therefore the probablity P

that the Brownian motion has at least one 0 in the interval (t_1, t_2) can be written as

$$P = \frac{1}{\sqrt{2\pi t\sigma}} \int_{-\infty}^{\infty} P[T_a < t_1 - t_\circ] e^{-\frac{a^2}{2t_\circ\sigma^2}} da. \tag{1.5.12}$$

Let us explain the validity of this assertion. At time $t_\circ$, $Z_{t_\circ}$ can be at any point $a \in \mathbf{R}$. The Gaussian exponential factor $\frac{1}{\sqrt{2\pi t\sigma}} e^{-\frac{a^2}{2t_\circ\sigma^2}}$ is the density function for $Z_{t_\circ}$. The factor $P[T_a < t_1 - t_\circ]$ is equal to the probability that starting at a, the Brownian motion will assume value 0 in the ensuing time interval $(t_1 - t_\circ)$. The validity of the assertion follows from these facts put together. In view of the symmetry between a and $-a$, and the density function for T_a which we obtained in example 1.5.3, (1.5.12) becomes

$$\begin{aligned} P &= \frac{2}{\sqrt{2\pi t_\circ\sigma}} \int_\circ^\infty e^{-\frac{a^2}{2t_\circ\sigma}} \frac{a}{\sqrt{2\pi\sigma}} \left(\int_\circ^{t_1-t_\circ} e^{-\frac{a^2}{2u}} \frac{1}{u\sqrt{u}} du\right) da \\ &= \frac{1}{\pi\sigma^2\sqrt{t_\circ}} \int_\circ^{t_1-t_\circ} u^{-\frac{3}{2}} \left(\int_\circ^\infty a e^{-\frac{a^2}{2\sigma^2}\left(\frac{1}{u}+\frac{1}{t_\circ}\right)} da\right) du \\ &= \frac{\sqrt{t_\circ}}{\pi} \int_\circ^{t_1-t_\circ} \frac{du}{(u+t_\circ)\sqrt{u}} \end{aligned}$$

The substitution $\sqrt{u} = x$ yields

$$P = \frac{2}{\pi} \tan^{-1} \sqrt{\frac{t_1 - t_\circ}{t_\circ}} = \frac{2}{\pi} \cos^{-1} \sqrt{\frac{t_1}{t_\circ}}, \tag{1.5.13}$$

for the probability of at least one crossing of 0 betwen times $t_\circ$ and t_1. It follows that the probability of no crossing in the time interval $(t_\circ, t_1)$ is

$$\frac{2}{\pi} \sin^{-1} \sqrt{\frac{t_1}{t_\circ}}$$

which is the arc-sine law for Brownian motion. ♠

So far we have only considered Brownian motion in dimension one. By looking at m copies of independent Brownian motions $Z_t = (Z_{t;1}, \cdots, Z_{t;m})$ we obtain Brownian motion in $\mathbf{R}^m$. While m-dimensional Brownian motion is defined in terms of coordinates, there is no preference for any direction in space. To see this more clearly, let $A = (A_{jk})$ be an $m \times m$ orthogonal matrix. This means that $AA' = I$ where superscript $'$ denotes the transposition of the matrix, and geometrically it means that the linear transformation A preserves lengths and therefore necessarily angles too. Let

$$Y_{t;j} = \sum_{k=1}^m A_{jk} Z_{t;k}.$$

Since a sum independent Gaussian random variables is Gaussian, $Y_{t;j}$'s are also Gaussian. Furthermore,

$$\mathsf{Cov}(Y_{t;j}, Y_{t;k}) = \sum_{l=1}^{m} A_{jl}A_{kl} = \delta_{jk}$$

by orthogonality of the matrix A. It follows that the components $Y_{t;j}$ of Y_t, which are Gaussian, are independent Gaussian random variables just as in the case of Brownian motion. This invariance property of Brownian motion (or independent Gaussians) under orthogonal transformations in particular implies that the distribution of the first hitting points on the sphere $S_\rho^{m-1} \subset \mathbf{R}^m$ of radius ρ centered at the origin, is uniform on S_ρ^{m-1}. This innocuous and almost obvious observation together with the standard fact from analysis that the solutions of Laplace's equation (*harmonic functions*)

$$\Delta_m u \stackrel{\text{def}}{=} \frac{\partial^2 u}{\partial x_1^2} + \cdots + \frac{\partial^2 u}{\partial x_m^2} = 0$$

are characterized by their mean value property has an interesting consequence, viz.,

Proposition 1.5.1 *Let M and N be disjoint compact hypersurfaces in $\mathbf{R}^m$ such that N is contained in the interior of M or vice versa. Let D denote the region bounded by M and N. Denote by $p(x)$ the probability that m-dimensional Brownian motion with $Z_\circ = x \in D$ hits N before it hits M. Then p is the unique solution to Laplace's equation in D with*

$$p \equiv 1 \quad \text{on} \quad N; \qquad p \equiv 0 \quad \text{on} \quad M.$$

(See remark 1.5.1 below.)

Proof - Let $\rho > 0$ be sufficiently small so that the sphere $S_\rho^{m-1}(x)$ of radius $\rho > 0$ centered at $x \in D$ in contained entirely in D. Let T be the first hitting of the sphere $S_\rho^{m-1}(x)$ given $Z_\circ = x$. Then the distribution of the points y defined by $Z_T = y$ is uniform on the sphere. Let $\mathcal{B}_x$ be the event that starting x the Brownian motion hits N before M. Consequently in view of the Markov property (see remark 1.5.2 below) we have for $B \subset \mathbf{R}^m$

$$P[\mathcal{B}_x] = \int_{S_\rho^{m-1}(x)} P[\mathcal{B}_y \mid Z_T = y] \frac{1}{\mathrm{vol}(S_\rho^{m-1}(x))} dv_S(y),$$

where $dv_S(y)$ denotes the standard volume element on the sphere $S_\rho^{m-1}(x)$. Therefore we have

$$p(x) = \int_{S_\rho^{m-1}(x)} \frac{p(y)}{\mathrm{vol}(S_\rho^{m-1}(x))} dv_S(y), \tag{1.5.14}$$

which is precisely the mean value property of harmonic functions. The boundary conditions are clearly satisfied by p and the required result follows. ♣

Remark 1.5.1 The condition that N is contained in the interior of M is intuitively clear in low dimensions and can be defined precisely in higher dimensions but it is not appropriate to dwell on this point in this context. This is not an essential assumption since the same conclusion remains valid even if D is an unbounded region by requiring the solution to Laplace's equation to vanish at infinity. ♡

Remark 1.5.2 We are in fact using the strong Markov property of Brownian motion. This application of this property is sufficiently intuitive that we did give any further justification. ♡

Example 1.5.5 We specialize proposition 1.5.1 to dimensions 2 and 3 with N a sphere of radius $\epsilon > 0$ and M a sphere of radius R, both centered at the origin. In both cases we can obtain the desired solutions by writing the Laplacian Δ_m in polar coordinates:

$$\Delta_2 = \frac{1}{r}\frac{\partial}{\partial r}(r\frac{\partial}{\partial r}) + \frac{1}{r^2}\frac{\partial^2}{\partial\theta^2},$$

and

$$\Delta_3 = \frac{1}{r^2}\Big[\frac{\partial}{\partial r}r^2\frac{\partial}{\partial r} + \frac{1}{\sin\theta}\frac{\partial}{\partial\theta}(\sin\theta\frac{\partial}{\partial\theta}) + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\Big].$$

Looking for spherically symmetric solutions p_m (i.e., depending only on the variable r) the partial differential equations reduce to ordinary differential equations which we easily solve to obtain the solutions

$$p_2(x) = \frac{\log r - \log R}{\log\epsilon - \log R}, \quad \text{for} \quad x = (r, \theta), \tag{1.5.15}$$

and

$$p_3(x) = \frac{\frac{1}{r} - \frac{1}{R}}{\frac{1}{\epsilon} - \frac{1}{R}}, \quad \text{for} \ \ x = (r, \theta, \phi), \tag{1.5.16}$$

for the given boundary conditions. Now notice that

$$\lim_{R\to\infty} p_2(x) = 1, \qquad \lim_{R\to\infty} p_3(x) = \frac{\epsilon}{r}. \tag{1.5.17}$$

The difference between the two cases is naturally interpreted as Brownian motion being recurrent in dimension two but transient in dimensions ≥ 3. ♠

Remark 1.5.3 The functions $u = \frac{1}{r^{m-2}}$ satisfies Laplace's equation in $\mathbf{R}^m$ for $m \geq 3$, and can be used to rstablish the analogue of example 1.5.5 in dimensions ≥ 4. ♡

Brownian motion has the special property that transition from a starting point x to a set A is determined by the integration of a function $p_t(x-y;\sigma)$ with respect to y on A. The fact that the integrand is a function of $y-x$ reflects a space homogeneity property which Brownian shares with random walks. On the other hand, Markov chains do not in general enjoy such space homogeneity property. Naturally there are many pocesses that are Markovian in nature but do not have space homogeneity property. We present some such examples constructed from Brownian motion.

Example 1.5.6 Consider one dimensional Brownian motion with $Z_\circ = x > 0$ and impose the condition that if for a path ω, $\omega(t_\circ) = 0$, then $\omega(t) = 0$ for all $t > t_\circ$. This is *absorbed Brownian motion* which we denote by $\tilde{Z}_t$. Let us compute the transition probabilities for $\tilde{Z}_t$. Unless stated to the contrary, in this example all probabilities and events involving Z_t are conditioned on $Z_\circ = x$. For $y > 0$ let

$$\mathcal{B}_t(y) = \{\omega \mid \omega(t) > y, \ \min_{0\leq s\leq t} \omega(s) > 0\},$$
$$\mathcal{C}_t(y) = \{\omega \mid \omega(t) > y, \ \min_{0\leq s\leq t} \omega(s) < 0\}.$$

We have

$$P[Z_t > y] = P[\mathcal{B}_t(y)] + P[\mathcal{C}_t(y)]. \tag{1.5.18}$$

By the reflection principle

$$P[\mathcal{C}_t(y)] = P[Z_t < -y].$$

Therefore

$$\begin{array}{rcl} P[\mathcal{B}_t(y)] & = & P[Z_t > y] - P[Z_t < -y] \\ & = & P[Z_t > y - x \mid Z_\circ = 0] - P[Z_t > x + y \mid Z_\circ = 0] \\ & = & \frac{1}{\sqrt{2\pi t}\sigma} \int_{y-x}^{y+x} e^{-\frac{u^2}{2t\sigma^2}} du \\ & = & \int_y^{y+2x} p_t(u - x; \sigma) du. \end{array}$$

Therefore for $\tilde{Z}_t$ we have

$$\begin{array}{rcl} P[\tilde{Z}_t = 0 \mid \tilde{Z}_\circ = x] & = & 1 - P[\mathcal{B}_t(0)] \\ & = & 1 - \int_{-x}^{x} p_t(u; \sigma) du \\ & = & 2 \int_\circ^\infty p_t(x + u; \sigma) du. \end{array}$$

Similarly, for $0 < a < b$,

$$\begin{array}{rcl} P[a < \tilde{Z}_t < b] & = & P[\mathcal{B}_t(a)] - P[\mathcal{B}_t(b)] \\ & = & \int_a^b [p_t(u - x; \sigma) - p_t(u + x; \sigma)] du. \end{array}$$

Thus we see that the transition probability has a discrete part $P[\tilde{Z}_t = 0 \mid \tilde{Z}_\circ = 0]$ and a continuous part $P[a < \tilde{Z}_t < b]$ and is not a function of $y - x$. ♠

Example 1.5.7 Let $Z_t = (Z_{t;1}, Z_{t;2})$ denote two dimensional Brownian motion with $Z_{t;i}(0) = 0$, and define

$$R_t = \sqrt{Z_{t;1}^2 + Z_{t;2}^2}.$$

This process means that for every Brownian path ω we consider its distance from the origin. This is a one dimensional process on $\mathbf{R}_+$, called *radial Brownian motion* or a *Bessel process* and its Markovian property is intuitively reasonable and will analytically follow from the calculation of transition probabilities presented below. Let us compute the transition probabilities. We have

$$\begin{array}{rcl} P[R_t \leq b] \mid Z_\circ = (x_1, x_2)] & = & \int \int_{y_1^+ y_2^2 \leq b^2} \frac{1}{2\pi t\sigma^2} e^{-\frac{(y_1 - x_1)^2 + (y_2 - x_2)^2}{2t\sigma^2}} dy_1 dy_2 \\ & = & \frac{1}{2\pi t\sigma^2} \int_\circ^b \int_\circ^{2\pi} e^{-\frac{(r\cos\theta - x_1)^2 + (r\sin\theta - x_2)^2}{2t\sigma^2}} d\theta r dr \\ & = & \frac{r}{2\pi t\sigma^2} \int_\circ^b e^{-\frac{r^2 + \rho^2}{2t\sigma^2}} I(r, x) dr. \end{array}$$

where (r, θ) are polar coordinates in $y_1 y_2$-plane, $\rho = ||x||$ and

$$I(r, x) = \int_\circ^{2\pi} e^{\frac{r}{t\sigma^2}[x_1 \cos\theta + x_2 \sin\theta]} d\theta.$$

Setting $\cos\phi = \frac{x_1}{\rho}$ and $\sin\phi = \frac{x_2}{\rho}$, we obtain

$$I(r, x) = \int_\circ^{2\pi} e^{\frac{r\rho}{t\sigma^2}\cos\theta} d\theta.$$

The *Bessel function* $I_\circ$ is defined as

$$I_\circ(\alpha) = \frac{1}{2\pi}\int_\circ^{2\pi} e^{\alpha\cos\theta}d\theta.$$

Therefore the desired transition probability

$$P[R_t \leq b \mid Z_\circ = (x_1, x_2)] = \int_\circ^b \tilde{p}_t(\rho, r; \sigma)dr, \qquad (1.5.19)$$

where

$$\tilde{p}_t(\rho, r; \sigma) = \frac{r}{t\sigma^2}e^{-\frac{r^2+\rho^2}{2t\sigma^2}}I_\circ(\frac{r\rho}{t\sigma^2}).$$

The Markovian property of radial Brownian motion is a consequence of the expression for transition probabilities since they depends only on (ρ, r). From the fact that $I_\circ$ is a solution of the differential equation

$$\frac{d^2u}{dz^2} + \frac{1}{z}\frac{du}{dz} - u = 0,$$

we obtain the partial differential differential equation satisfied $\tilde{p}$:

$$\frac{\partial\tilde{p}}{\partial t} = \frac{\sigma^2}{2}\frac{\partial^2\tilde{p}}{\partial r^2} + \frac{\sigma}{2r}\frac{\partial\tilde{p}}{\partial r},$$

which is the radial heat equation. ♠

The analogue of non-symmetric random walk ($\mathsf{E}[X] \neq 0$) is Brownian motion with *drift* μ which one may define as

$$Z_t^\mu = Z_t + \mu t$$

in the one dimensional case. It is a simple exercise to show

Lemma 1.5.1 *Z_t^μ is normally distributed with mean μt and variance $t\sigma^2$, and has stationary independent increments.*

In particular the lemma implies that, assuming $Z_\circ^\mu = 0$, the probability of the set of paths that at time t are in the interval (a, b) is

$$\frac{1}{\sqrt{2\pi t}\sigma}\int_a^b e^{-\frac{(u-\mu t)^2}{2t\sigma^2}}du.$$

Example 1.5.8 Let $-b < 0 < a$, $x \in (-b, a)$ and $p(x)$ be the probability that Z_t^μ hits a before it hits $-b$. This is similar to proposition 1.5.1. Instead of using the mean value property of harmonic functions (which is no longer valid here) we directly use our knowledge of calculus to derive a differential equation for $p(x)$ which allows us to calculate it. The method of proof has other applications (see exercise 1.5.5). Let h be a small number and $\mathcal{B}$ denote the event that Z_t^μ hits a before it hits $-b$. By conditioning on Z_h^μ, and setting $Z_h^\mu = x + y$ we obtain

$$p(x) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi h}\sigma} e^{-\frac{(y-\mu h)^2}{2h\sigma^2}} p(x+y)dy \tag{1.5.20}$$

The Taylor expansion of $p(x+y)$ gives

$$p(x+y) = p(x) + yp'(x) + \frac{1}{2}y^2 p''(x) + \cdots$$

Now $y = Z_h - Z_\circ$ and therefore

$$\int_{-\infty}^{\infty} y \frac{e^{-\frac{(y-\mu h)^2}{2h\sigma^2}}}{\sqrt{2\pi h}\sigma} dy = \mu h, \quad \int_{-\infty}^{\infty} y^2 \frac{e^{-\frac{(y-\mu h)^2}{2h\sigma^2}}}{\sqrt{2\pi h}\sigma} dy = \sigma^2 h + h^2\mu^2.$$

It is straightforward to check that contribution of terms of the Taylor expansion containing y^k, for $k \geq 3$, is $O(h^2)$. Substituting in (1.5.20), dividing by h and taking $\lim_{h\to 0}$ we obtain

$$\frac{\sigma^2}{2}\frac{d^2 p}{dx^2} + \mu \frac{dp}{dx} = 0.$$

The solution with the required boundary conditions is

$$p(x) = \frac{e^{\frac{2\mu b}{\sigma^2}} - e^{-\frac{2\mu x}{\sigma^2}}}{e^{\frac{2\mu b}{\sigma^2}} - e^{-\frac{2\mu a}{\sigma^2}}}.$$

The method of solution is applicable to other problems. ♠

EXERCISES

Exercise 1.5.1 *Formulate the analogue of the reflection principle for Brownian motion and use it to give an alternative proof of (1.5.9).*

Exercise 1.5.2 *Discuss the analogue of example 1.5.5 in dimension 1.*

Exercise 1.5.3 *Generate ten paths for the simple symmetric random walk on* $\mathbf{Z}$ *for* $n \leq 1000$. *Rescale the paths in time direction by* $\frac{1}{1000}$ *and in the space direction by* $\frac{1}{\sqrt{1000}}$, *and display them as graphs.*

Exercise 1.5.4 *Display ten paths for two dimensional Brownian motion by repeating the computer simulation of exercise 1.5.3 for each component. The paths so generated are one dimensional curves in three dimensional space (time + space). Display only their projections on the space variables.*

Exercise 1.5.5 *Consider Brownian motion with drift* Z_t^μ *and assume* $\mu > 0$. *Let* $-b < 0 < a$ *and let* T *be the first hitting of the boundary of the interval* $[-b, a]$ *and assume* $Z_\circ^\mu = x \in (-b, a)$. *Show that* $\mathsf{E}[T] < \infty$. *Derive a differential equation for* $\mathsf{E}[T]$ *and deduce that for* $\sigma = 1$

$$\mathsf{E}[T] = \frac{a-x}{\mu} + (a+b)\frac{e^{-2\mu b} - e^{-2\mu x}}{\mu(e^{2\mu b} - e^{-2\mu a})}.$$

Exercise 1.5.6 *Consider Brownian motion with drift* Z_t^μ *and assume* $\mu > 0$ *and* $a > 0$. *Let* T_a^μ *be the first hitting of the point* a *and*

$$F_t(x) = P[T_x^\mu < t \mid Z_\circ^\mu = 0].$$

Using the method of example 1.5.8, derive a differential equation for $F_t(x)$.