

Stochastic Processes

Jacob Denson

September 24, 2015

Table Of Contents

1	Evolution	2
2	Discrete Markov Processes	6
2.1	Asymptotics of Markov chains	7

Chapter 1

Evolution

The theory of dynamic systems allows us to determine the motions of objects under deterministic actions. From Newton's development of Calculus, one could show that past and future are predicted exactly from the positions and velocities of particles at any one point. It is here that the theory of Differential equations takes over. Nonetheless, in reality one can never measure the data required to determine the state of a system in precision. Inexactness thus shrouds the determinism of a system, which invalidates Newton's model when applied to experiments. The theory of stochastic processes allows us to model the evolution of a system subject to uncertainty. They can be seen as the evolution of the concept of evolution in a deterministic system.

Definition. A Dynamical system is a tuple (S, M, φ) , where

1. S is an arbitrary set, the state space of the system.
2. M is a monoid, a fixed time system. (e.g. \mathbf{N} , \mathbf{R} , $[0, \infty)$, etc.)
3. $\varphi : S \times M \rightarrow S$ is a function, the evolution of the system. such that $\varphi(\varphi(t', s), t) = \varphi(s, t + t')$, and $\varphi(s, 0) = s$. If, in our model, we are at a point $s \in S$, then t seconds later we should arrive at $\varphi(s, t)$.
4. An initial state $s \in S$ determines a path $\varphi(s, \cdot)$ through S .

Definition. A **Random Dynamical System** is a tuple $(S, M, \Omega, \mathbf{P}, \varphi)$, where

1. S is a state space, M a time system, and (Ω, \mathbf{P}) a probability space.
2. φ is the ‘probabilistic flow’, a function $\varphi : S \times M \times \Omega \rightarrow S$, where for a fixed $s \in S$ and $t \in \Omega$, $\varphi(s, t, \cdot)$ is a random variable. We require that $\varphi(s, 0, \omega) = s$, for any $\omega \in \Omega$, and that for almost all ω , $\varphi(\varphi(s, t', \omega), t, \omega) = \varphi(s, t + t', \omega)$.
3. An initial state s determines a family of random variables $X_t = \varphi(s, t, \cdot)$. For any specific $\omega \in \Omega$, $\varphi(s, \cdot, \omega)$ is a path in state space.

Like in the theory of differential equations, we shall focus on studying what is formed out of a system of this type, namely, a family of random variables from a space.

Definition. Let T be a linearly ordered set. A **stochastic process** is a collection $\{X_t\}_{t \in T}$ of random variables defined over the same probability space Ω , with range in the same measurable space S .

Almost every problem in probability theory can be formulated as a statement about stochastic processes. This isn’t a good thing – we can’t say much about stochastic processes in their general form. The fun of stochastic processes results when we add additional relationships between the random variables, and study the resulting properties.

Example. Every random vector (X_1, X_2, \dots, X_n) is a stochastic process. In fact, any stochastic process over a finite T is effectively a random vector. Interesting examples of stochastic processes occur for infinite T , but random vectors can be useful for intuition.

Given a stochastic process $\{X_t\}_{t \in T}$, and a finite subset $T' \subset T$, the sub-collection $\{X_t\}_{t \in T'}$ is a random vector, known as a **finite dimensional dis-**

tribution of the process. It turns out that if we understand all finite distributions given by the finite subsets T' , we can build up a stochastic process $\{X_t\}_{t \in T}$, and any such choice has the same distribution. This lifesaver is known as the Kolmogorov extension theorem, and it makes it a deal more easy to define a random process. Some useful notation will simplify things. Let A is an index set, and $B \subset A$. Given a family of sets $\{C_i\}_{i \in A}$, we define $\pi_{A \rightarrow B}(\{x_i\}_{i \in A}) = \{x_i\}_{i \in B}$. If a fixed total index set T is given, we will let $\pi_A = \pi_{T \rightarrow A}$.

Theorem 1.1 (Kolmogorov's extension theorem). *Suppose that T is a linearly ordered set, Σ is a fixed σ -algebra on a hausdorff, locally compact topological space S , and for each finite subcollection $A \subset T$, there is a probability function \mathbf{P}_A defined on Σ^A , such that if $A \subset B$, $\mathbf{P}_B = \mathbf{P} \circ \pi_{B \rightarrow A}$. Then there exists a probability space Ω , and a process $\{X_t : \Omega \rightarrow S\}_{t \in T}$, where $\{X_{t_i}\}_{i=1}^n \sim \mathbf{P}_{\{t_1, \dots, t_n\}}$.*

Proof. A stochastic process is pretty much just a probability distribution over Σ^T , so we are really just defining a distribution \mathbf{P} on S^T . We must define

$$\mathbf{P}(\pi_B^{-1}(U)) = \mathbf{P}_B(U)$$

Suppose

$$W = \pi_B^{-1}(U) = \pi_A^{-1}(V)$$

Then it follows that

$$\pi_{A \cup B}^{-1}(\pi_{A \cup B}(W)) = W$$

so by the consistency condition above,

$$\mathbf{P}_B(\pi_B(W)) = \mathbf{P}_{A \cup B}(\pi_{A \cup B}(W)) = \mathbf{P}_A(\pi_A(W))$$

Thus \mathbf{P} is well defined.

It is easy to see that \mathbf{P} is defined on an algebra, and is finitely additive. It is also countably additive. Let E_1, E_2, \dots, E_n be disjoint measurable by \mathbf{P} . If we define

$$F_n := \bigcup_{i=1}^{\infty} E_i - \bigcup_{i=1}^n E_i$$

we must show $\mathbf{P}(F_n) \rightarrow 0$. Let each $F_n = \pi_{A_n}^{-1}(G_n)$, for some indices A_n and event G_n . Each \mathbf{P}_{A_n} is inner regular, and so we may find a compact $K_n \subset G_n$ in each Σ^{A_n} for which $0 < \mathbf{P}_{A_n}(G_n - K_n) < \varepsilon/2^n$. Since $\pi_{A_n \rightarrow A_m}(G_n) \subset$

G_m . Defining $H_n = \bigcup_{i=1}^n \pi_{A_i \rightarrow A_n}^{-1}(K_n)$, we see that each H_n is compact, that $\pi_{A \rightarrow B}^{-1}(H_B) \subset H_A$, and that

$$\mathbf{P}(G_n - H_n) \leq \mathbf{P}(G_n - K_n) < \varepsilon$$

But also $\pi_n^{-1}(H_n) \rightarrow \emptyset$. Assigning the product measure to S^T , each $\pi_n^{-1}(H_n)$ is compact. But this implies that some H_N is empty. Thus $\mathbf{P}(F_N - H_N) = \mathbf{P}(F_N) < \varepsilon$. So $\mathbf{P}(F_n) \rightarrow 0$.

$0 < \mathbf{P}(F_n - K'_n) < \varepsilon$. We have $\bigcap K'_n = \emptyset$, since each $K'_n \subset F_n$. But then some finite intersection is empty, so there is some empty K'_M . This implies that $\mathbf{P}(K'_M) = 0$, so $\mathbf{P}(F_M - K'_M) = \mathbf{P}(F_M) < \varepsilon$. This implies that $\mathbf{P}(F_n) \rightarrow 0$. Thus \mathbf{P} defines a pre-measure. By the Hahn-Kolmogorov theorem, \mathbf{P} can be extended to a unique measure on all of Σ^T . If we let $X_t = \pi_{\{t\}}$, then $\{X_t\}_{t \in T}$ is the distribution needed. \square

The Kolmogorov theorem is used to construct measures, most importantly when T is uncountable. To begin if, we will study processes with countable T , for which most paradoxes are avoidable.

Chapter 2

Discrete Markov Processes

In this chapter, we assume all probability and state spaces are countable or finite, and probabilities defined on every subset of event space.

Definition. Let $\{X_t\}_{t \in T}$ be a stochastic process. We say this process satisfies the **Markov property** if, for any $x_{t_1}, x_{t_2}, \dots, x_{t_n}, x_{t_{n+1}} \in S$, and $t_1 < t_2 < \dots < t_n < t_{n+1}$, we have

$$\mathbf{P}(X_{t_{n+1}} = x_{t_{n+1}} | X_{t_n} = x_{t_n}, \dots, X_{t_1} = x_{t_1}) = \mathbf{P}(X_{t_{n+1}} = x_{t_{n+1}} | X_{t_n} = x_{t_n})$$

For any choice for which this equation is not inadmissible. A **Markov chain** is a stochastic process $\{X_n\}_{n \in \mathbb{N}}$ satisfying the Markov property.

Intutively, the markov property tells us we cannot gain information about the future of our process except from the most recent information.

Example. All independent random variables $\{X_t\}_{t \in T}$ satisfy the Markov property,

$$\mathbf{P}(X_{t_{n+1}} = x_{t_{n+1}} | X_{t_n} = x_{t_n}, \dots, X_{t_1} = x_{t_1}) = \mathbf{P}(X_{t_{n+1}} = x_{t_{n+1}}) = \mathbf{P}(X_{t_{n+1}} = x_{t_{n+1}} | X_{t_n} = x_{t_n})$$

Of course, independant processes are some of the least interesting examples.

We will begin by focusing on ‘time homogenous’ Markov chains.

Definition. A Markov chain can be defined in the following manner. First, we specify two functions $P(\cdot, \cdot) : S^2 \rightarrow [0, 1]$ and $I : S \rightarrow [0, 1]$ such that, for a fixed $s \in S$, $P(s, \cdot)$ is a probability distribution, and I specifies an ‘initial probability’ distribution. Then we may define

$$\mathbf{P}_{\{1,2,\dots,n\}}(x_1, \dots, x_n) = I(x_1) \prod_{k=1}^{n-1} P(x_k, x_{k+1})$$

and extend this to a probability density on S^n . These distributions are sufficient for Kolmogorov’s theorem to apply, and the stochastic processes which form are called **time-homogenous markov chains**.

Now suppose S is finite. Without loss of generality, we may assume $S = \{1, 2, \dots, n\}$, for some natural number n . We may then form the **stochastic matrix** P , where $P_{ij} = P(i, j)$. The rows of P sum to one, and any such matrix with these rows can specify a time-homogenous markov chain. An initial distribution can be seen as a row vector v whose entries sum to one, and then $\mathbf{P}(X_n = x_i) = (vP^n)_i = \sum_k v_k P_{ik}$. Thus studying finite time-homogenous markov chains reduces to matrix algebra. Determining the long term behaviour of this markov chain is found just by taking matrix limits.

2.1 Asymptotics of Markov chains

We want to determine the behaviour of a Markov chain $\{X_1, X_2, \dots\}$ as the X_i tend to infinity. It is most easy to do this analysis in the case of a finite-state space time-homogenous Markov chain. In some cases, we can determine a limiting probability distribution. We have

$$\mathbf{P}(X_n = x_k) = (vP^n)_k$$

If $m = \lim_{n \rightarrow \infty} (vP^n) = v \lim_{n \rightarrow \infty} P^n$ exists, then

$$\lim_{n \rightarrow \infty} \mathbf{P}(X_n = x_k) = \lim_{n \rightarrow \infty} (vP^n)_k \rightarrow m_k$$

Thus the Markov chain $\{X_n\}$ converges in distribution. In most cases, this limit will exist. In extraneous circumstances, this need not hold.

Example (Inverter). Consider the stochastic matrix

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

Then $P^k = P$, for odd k , and $P^k = I$, for even k . Thus P^k cannot converge.

Example (Bernoulli Pair). Fix two numbers $0 < p, q < 1$. Consider the stochastic matrix

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$$

Defining the matrices

$$Q := \begin{pmatrix} 1 & -p \\ 1 & q \end{pmatrix} \quad Q^{-1} := \frac{1}{p+q} \begin{pmatrix} q & p \\ -1 & 1 \end{pmatrix}$$

we see that

$$QPQ^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1-p-q \end{pmatrix}$$

and so

$$\lim_{n \rightarrow \infty} P^n = Q^{-1} \left[\lim_{n \rightarrow \infty} (QPQ^{-1})^n \right] Q = Q^{-1} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} Q = \frac{1}{p+q} \begin{pmatrix} q & p \\ q & p \end{pmatrix}$$

Notice that all rows of the limiting matrix are the same, so that regardless of the initial probability distribution v we choose, the probability values will converge to the same result. Our specific computation works since the eigenvalue with largest modulus is 1, and that the matrix is fully diagonalizable. Because of this, we can multiply the diagonal matrix with itself, and all eigenvalues will vanish except for eigenvalues of value 1. It shall turn out that our general strategy is only a simple extension.

We shall rely on a theorem which can be proved by the methods of analytical linear algebra.

Theorem 2.1 (The Perron-Frobenius Theorem). Let M be a positive square $n \times n$ matrix (that is, $M_{ij} > 0$ for all $1 \leq i, j \leq n$). There is a unique positive eigenvalue λ , called the **Perron root** of M , such that

1. If γ is an eigenvalue for M , then $|\gamma| \leq |\lambda|$
2. The eigenspace of λ is one-dimensional, and contains some vector v , all of whose coordinates are positive, and the only other eigenvectors with positive coordinates is a scalar multiple of v .

Many more properties hold for λ , but these properties are the two which we will utilize.

It shall be helpful to fix a stochastic matrix P , and vary the initial distributions v . Some of these distributions work very nicely when taking limits of the stochastic matrix: Suppose v is a left eigenvector of P ($vP = v$), which also defines an initial distribution ($\sum v_i = 1$, and each $v_i \geq 0$). Then $vP^n = v$, and thus $\lim_{n \rightarrow \infty} vP^n = \lim_{n \rightarrow \infty} v = v$, and so v is the limiting distribution of the Markov chain it generates. In the matrix in the last example, we see that v must be equal to the row repeated in the limiting matrix. Identifying these vectors therefore seems important in order to identify the limiting distribution of the matrix.

Definition. Fix a stochastic matrix P . A **invariant probability distribution** for P is a positive row vector v whose coordinates sum to one, such that $vP = v$. Note that this is simply a left eigenvector of eigenvalue 1.

Now suppose P is a stochastic, positive matrix. Then we may apply Perron-Frobenius to P^t , obtaining a Perron root λ . Now every eigenvalue of P^t is an eigenvalue for P , so that λ must also be the Perron root for P as well. Then $(1, 1, \dots, 1)^t$ is a right eigenvector for P of eigenvalue 1, which implies that $\lambda = 1$. Thus P can be modified, under some change of basis matrix Q , such that

$$QPQ^{-1} = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{M} \end{pmatrix}$$

Where \mathbf{M} is a square matrix such that $\lim_{n \rightarrow \infty} \mathbf{M}^n = 0$ (Use the Jordan Canonical Form, and the fact that 1 is the maximal eigenvalue). But then

$$\lim_{n \rightarrow \infty} P^n = Q^{-1} \left(\lim_{n \rightarrow \infty} QPQ^{-1} \right)^n Q = Q^{-1} \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q = \begin{pmatrix} \pi \\ \vdots \\ \pi \end{pmatrix}$$

where π is a row vector which sums to one – an invariant probability vector for P . Note that since the left eigenvector is an invariant probability, this left eigenvector is exactly π , so we need only find a positive normalized left eigenvector to eigenvalue 1.

This argument can be considerably strengthened from the fact that the argument above works if P is any stochastic matrix with P^n positive, for some n , because the eigenvalue of P^n are simply the eigenvalues of P to the power of n , as are the eigenvectors. Since we know 1 is the Perron root of P^n (since P^n is stochastic), 1 is a maximal eigenvalue for P as well, and that if we pick an eigenvector v for P of eigenvalue 1, then v is also an eigenvector for P^n , and thus we may multiply v by a scalar such that all entries are positive.

Our problem thus reduces to finding those stochastic matrices P for which P^n is positive, for some n . In the terms of a time homogenous markov-chain, for some n , and for any states s, m , we have

$$\mathbf{P}(X_n = s | X_0 = m) > 0$$

This will reduce to two concepts, that of periodicity and irreducibility.

Definition. Let $s, m \in \mathcal{S}$ be two states in an arbitrary Markov chain. We say s communicates with m if there is some n with $\mathbf{P}(X_n = s | X_0 = m) > 0$. If we divide \mathcal{S} into classes of states, all of which communicate between one another, then we obtain a set of **communication classes**. A Markov chain with one communication class is **irreducible**.

Definition. For any $s \in \mathcal{S}$, let $\mathcal{J}(s) = \{n \in \mathbf{N} : \mathbf{P}(X_n = s | X_0 = s) > 0\}$. Then $\mathcal{J}(s)$ is closed under addition. The greatest common divisor of $\mathcal{J}(s)$ is known as the **period** of s . A Markov chain for which every state has period one is known as **aperiodic**.

Theorem 2.2. *If s and m are in the same communication class, then the period of s is the same as the period of m .*

Theorem 2.3. *If the period of a state s is m , then for some M , if $n > m$, and $n|m$, then $n \in \mathcal{J}(s)$.*

Theorem 2.4. *Let P be a stochastic matrix, which determines an aperiodic, irreducible Markov chain. Then there is a unique vector π for which $\pi P = \pi$, and for any other probability distribution v , $\lim_{n \rightarrow \infty} v P^n = \pi$.*

Proof. We just need to verify that P^n is positive for some n . Since P is aperiodic, there is m for each i such that $P_{ii}^n > 0$ for $n > m$. Taking $M = \max_i(m)$, we get $P_{ii}^M > 0$ for all choices of i . If $P_{ij}^m > 0$, for some m , and $m > M$, then FINISH PROOF LATER. \square

This effectively completes our understanding of the asymptotic behaviour of these types of Markov chains. We can obtain more general results by further classifying the communication states of a state space. One way this can be done by understanding how often a state returns to itself, or to other states.

Definition. Let $\{X_i\}$ be a Markov chain. Fix some state $s \in \mathcal{S}$. Define a new random variable H_s , on the same probability space as the X_i by the formula

$$H_s(\omega) := \inf_i (X_i(\omega) = s)$$

This is known as the hitting time of the chain. Fix a state $s \in \mathcal{S}$.

1. We say s is **recurrent** if $\mathbf{P}(H_s = \infty | X_0 = s) = 0$.
2. s is **transient** if it is not recurrent: $\mathbf{P}(H_s = \infty | X_0 = s) > 0$.

Lemma 2.5. *Let s, m be two states in the same communication class. Then s is recurrent if and only if m is.*

This means we may divide the communication classes of a Markov chain up into recurrent classes, and transient classes. $\mathcal{S} = \bigcup \mathcal{R} \cup \bigcup \mathcal{T}$, where each $R \in \mathcal{R}$ is recurrent, and each $T \in \mathcal{T}$ is transient. For any $R \in \mathcal{R}$, and n , $\mathbf{P}(X_n \notin R | X_0 \in R) = 0$, since otherwise

$$\mathbf{P}(\{\omega : \forall k > m : X_k(\omega) \notin R\} | X_0 \in R) = \mathbf{P}(\{\})$$

then R_i would be transient (there is no way for our process to reenter R_i afterward). Thus if we consider our stochastic matrix P , and consider only the entries involving elements of R_i , we obtain a submatrix P_i which itself is stochastic. By reordering the entries of P , we may assume

$$P = \begin{pmatrix} P_1 & 0 & \dots & 0 \\ 0 & P_2 & \ddots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & \dots & P_n & 0 \\ \dots & S_1 & \dots & Q \end{pmatrix}$$

And for any m ,

$$P^m = \begin{pmatrix} P_1^m & 0 & \dots & 0 \\ 0 & P_2^m & \ddots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & \dots & P_n^m & 0 \\ \dots & S_m & \dots & Q^m \end{pmatrix}$$

We can work with each P_i as if it were its own stochastic process, and then put them together to form a stochastic process in full. For instance, assuming each P_i is aperiodic, we can find an invariant probability π_i for P_i . Swapping P_i out for $\lim_{n \rightarrow \infty} P^n$, we obtain the first part of convergence. Secondly, note that if s is a transient state, then any chain that starts at s must end up in a recurrent state (for otherwise it would keep up ending up in different transient states, and then one of these must return to itself with probability 1). Thus if j is any other transient state, then $\lim_{n \rightarrow \infty} \mathbf{P}(X_n = j | X_0 = i) = 0$. Let $\alpha_k(i)$ be the probability that a chain starting in state i eventually ends up in recurrence class R_k . Once a chain ends up in R_k , it will settle down to the equilibrium distribution on R_k . If $j \in R_k$, then

$$\lim_{n \rightarrow \infty} \mathbf{P}(X_n = j | X_0 = i) = \alpha_k(i) \pi_k(j)$$