MARKOV PROCESSES

Thomas Wieting
Reed College, 2006

1° Let a be any positive integer $(2 \leq a)$ and let A be the finite set:

$$A = \{1, 2, 3, \ \dots \ , a\}$$

Let P be a probability vector:

$$P = (P_1, P_2, P_3, \ \dots \ , P_a)$$

where:

$$0 \leq P_j \qquad (1 \leq j \leq a)$$

and:

$$\sum_{j=1}^{a} P_j = 1$$

Let Π be a stochastic matrix:

$$\Pi = \begin{pmatrix} \Pi_{11} & \cdots & \Pi_{1a} \\ \vdots & & \vdots \\ \Pi_{a1} & \cdots & \Pi_{aa} \end{pmatrix}$$

where:

$$0 \leq \Pi_{jk} \qquad (1 \leq j \leq a, \ 1 \leq k \leq a)$$

and:

$$\sum_{k=1}^{a} \Pi_{jk} = 1 \qquad (1 \leq j \leq a)$$

We assume that:

$$P\Pi = P \tag{1}$$

that is, that:

$$\sum_{j=1}^{a} P_j \Pi_{jk} = P_k \qquad (1 \leq k \leq a)$$

2° Now let X be the set of all sequences:

$$x = (x_0, x_1, x_2, \ \dots \ , x_n, \ \dots)$$

with entries in A:

$$1 \leq x_n \leq a \qquad (0 \leq n)$$

For any nonnegative integer r and for any finite sequence:

$$w = (w_0, w_1, w_2, \ldots, w_r)$$

with entries in A, let C_w be the *cylinder* in X comprised of all sequences x for which:

$$x_0 = w_0, x_1 = w_1, \ldots, x_r = w_r$$

We specify a probability measure μ on X by defining the values of μ on the cylinders in X, as follows:

$$\mu(C_w) := P_{w_0} \Pi_{w_0 w_1} \Pi_{w_1 w_2} \cdots \Pi_{w_{r-1} w_r}$$

One can readily extend μ to the various borel subsets of X. Finally, let T be the mapping carrying X to itself, defined as follows:

$$T((x_0, x_1, x_2, \ldots, x_n, \ldots)) := (x_1, x_2, x_3, \ldots, x_{n+1}, \ldots) \qquad (x \in X)$$

By relation (1), we find that μ is *invariant* under T:

$$T_*(\mu) = \mu \tag{2}$$

that is, that:

$$\mu(T^{-1}(C_w)) = \mu(C_w)$$

where C_w is any cylinder in X. At this point, we have assembled the initial ingredients A, P, and Π to produce a dynamical system:

$$(X, \mu, T)$$

One refers to this system as a *markov* system.

3° Let j be a member of A for which $P_j = 0$. One can easily show that:

$$\mu(\bigcup_{\ell=0}^{\infty} T^{-\ell}(C_j)) = 0$$

Hence, one may excise j from A without loss of significance. Hereafter, we will assume that:

$$0 < P_j \qquad (1 \leq j \leq a) \tag{3}$$

4° Let us say that the stochastic matrix Π is *irreducible* iff, for any members j and k of A, there is some positive integer ℓ such that:

$$0 < \Pi^{\ell}_{jk} \tag{4}$$

We plan to prove that the markov system (X, μ, T) is ergodic iff the stochastic matrix Π is irreducible.

5° Let C_w be a cylinder in X, where:

$$w = (w_0, w_1, w_2, \ \dots \ , w_r)$$

Let 1_w be the characteristic function for C_w. Applying the Ergodic Theorem, we introduce the limit function:

$$\hat{1}_w$$

as follows:

$$\hat{1}_w(x) := \lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} 1_w(T^m(x)) \qquad (x \in X)$$

One knows that:

$$\int_X \hat{1}_w(x)\mu(dx) = \int_X 1_w(x)\mu(dx) = \mu(C_w) \tag{5}$$

If (X, μ, T) is ergodic then in fact:

$$\hat{1}_w(x) = \mu(C_w) \qquad (x \in X) \tag{6}$$

In turn, let C_u and C_v be cylinders in X, where:

$$u = (u_0, u_1, u_2, \ \dots \ , u_p)$$

and:

$$v = (v_0, v_1, v_2, \ \dots \ , v_q)$$

Clearly:

$$1_u(x)\hat{1}_v(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} 1_u(x)1_v(T^m(x)) \qquad (x \in X)$$

Applying the Dominated Convergence Theorem, we obtain:

$$\int_X 1_u(x)\hat{1}_v(x)\mu(dx) = \lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} \mu(C_u \cap T^{-m}(C_v)) \tag{7}$$

Now one can readily verify that (X, μ, T) is ergodic iff, for any cylinders C_u and C_v in X:

$$\mu(C_u)\mu(C_v) = \lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} \mu(C_u \cap T^{-m}(C_v)) \tag{8}$$

6° Let j and k be any members of A. Taking C_u and C_v to be C_j and C_k, we may apply relation (7) to obtain:

$$\int_X 1_j(x)\hat{1}_k(x)\mu(dx) = \lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} P_j \Pi_{jk}^m$$

so that:

$$P_j^{-1} \int_X 1_j(x)\hat{1}_k(x)\mu(dx) = \lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} \Pi_{jk}^m \tag{9}$$

Now we may define the stochastic matrix Q as follows:

$$Q_{jk} := \lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} \Pi_{jk}^m \qquad (1 \le j \le a,\ 1 \le k \le a) \tag{10}$$

that is:

$$Q := \lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} \Pi^m$$

Clearly, $\Pi Q = Q = Q\Pi$, $QQ = Q$, and $PQ = P$.

7° If (X, μ, T) is ergodic then $\hat{1}_k$ is constant with constant value $\mu(C_k) = P_k$. Hence, by relation (9):

$$P_k = Q_{jk} \qquad (1 \le j \le a,\ 1 \le k \le a)$$

so all the rows of Q coincide with P. Conversely, if all the rows of Q coincide with P then, for any cylinders C_u and C_v in X:

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} \mu(C_u \cap T^{-m}(C_v)) \\
&\qquad = \lim_{n\to\infty} \frac{1}{n} \sum_{m=p+1}^{n-1} \mu(C_u \cap T^{-m}(C_v)) \\
&\qquad = \lim_{n\to\infty} \frac{1}{n} \sum_{m=p+1}^{n-1} P_{u_0}\Pi_{u_0u_1}\cdots\Pi_{u_{p-1}u_p}\Pi_{u_pv_0}^{m-p}\Pi_{v_0v_1}\cdots\Pi_{v_{q-1}v_q} \\
&\qquad = P_{u_0}\Pi_{u_0u_1}\cdots\Pi_{u_{p-1}u_p}P_{v_0}\Pi_{v_0v_1}\cdots\Pi_{v_{q-1}v_q} \\
&\qquad = \mu(C_u)\mu(C_v)
\end{aligned}$$

Hence, by relation (8), (X, μ, T) is ergodic. We conclude that (X, μ, T) is ergodic iff all the rows of Q coincide with P.

8° Let us assume that all the entries in Q are positive. Since $QQ = Q$, it is plain that all the columns of Q must be constant. Indeed, for each column K of Q, if the smallest entry g in K is strictly less than the largest entry h then, for each row L in Q, $g < LK < h$, which contradicts the fact that $QK = K$. Since $PQ = P$, it follows in turn that all the rows of Q coincide with P. We conclude that all the rows of Q coincide with P iff all the entries in Q are positive.

9° Let us assume that all the entries in Q are positive. By relation (10) (that is, by the definition of Q), it is plain that Π is irreducible. Let us assume that Π is irreducible. Let j and k be any members of A. There must be some member j' of A such that $0 < Q_{jj'}$. There must then be some positive integer ℓ such that $0 < \Pi^{\ell}_{j'k}$. Hence, $0 < (Q\Pi^{\ell})_{jk}$. However, $Q\Pi^{\ell} = Q$. We conclude that all the entries in Q are positive iff Π is irreducible.

10° Finally, we conclude that (X, μ, T) is ergodic iff Π is irreducible. Moreover, in such a case, all the rows of Q coincide with P.

11° One says that (X, μ, T) is (strongly) *mixing* iff, for any cylinders C_u and C_v in X:

$$\mu(C_u)\mu(C_v) = \lim_{n\to\infty} \mu(C_u \cap T^{-n}(C_v))$$

Clearly, if (X, μ, T) is (strongly) mixing then it is ergodic. One says that Π is *primitive* iff there is a positive integer ℓ such that all the entries in Π^{ℓ} are positive. Clearly, if Π is primitive then it is irreducible. Show that (X, μ, T) is (strongly) mixing iff Π is primitive. Moreover, show that, in such a case:

$$Q = \lim_{n\to\infty} \Pi^n$$

From the relation just stated, it follows that, for any probability vector L:

$$\lim_{n\to\infty} L\Pi^n = P$$

12° Let us compute the *entropy* of the markov system (X, μ, T). We make no assumptions about A, P, and Π other than those expressed in 1°. To connect with the theory of entropy, let us introduce the following *markov process*, based on (X, μ):

$$F_0,\, F_1,\, F_2,\, \dots\, ,\, F_{\ell},\, \dots$$

where:

$$F_0(x) := x_0 \qquad (x \in X)$$

and:

$$F_n(x) := F_0(T^n(x)) \qquad (x \in X,\ 0 \le n)$$

By the conventional definitions of entropy and of conditional entropy, we have:

$$\begin{aligned}&\eta(F_0 \times F_1 \times F_2 \times \cdots \times F_{n-1})\\ &\quad = \eta(F_0) + \eta(F_1|F_0) + \eta(F_2|F_0 \times F_1) + \cdots + \eta(F_{n-1}|F_0 \times F_1 \times \cdots \times F_{n-2})\end{aligned}$$

where n is any positive integer. However:

$$\begin{aligned}&\eta(F_3|F_0 \times F_1 \times F_2)\\ &= -\sum_{j=1}^{a}\sum_{k=1}^{a}\sum_{\ell=1}^{a}\mu(F_0 = j, F_1 = k, F_2 = \ell)\\ &\cdot \sum_{m=1}^{a}\frac{\mu(F_0 = j, F_1 = k, F_2 = \ell, F_3 = m)}{\mu(F_0 = j, F_1 = k, F_2 = \ell)} log \frac{\mu(F_0 = j, F_1 = k, F_2 = \ell, F_3 = m)}{\mu(F_0 = j, F_1 = k, F_2 = \ell)}\\ &= -\sum_{j=1}^{a}\sum_{k=1}^{a}\sum_{\ell=1}^{a} P_j \Pi_{jk}\Pi_{k\ell} \cdot \sum_{m=1}^{a}\frac{P_j \Pi_{jk}\Pi_{k\ell}\Pi_{\ell m}}{P_j \Pi_{jk}\Pi_{k\ell}} log \frac{P_j \Pi_{jk}\Pi_{k\ell}\Pi_{\ell m})}{P_j \Pi_{jk}\Pi_{k\ell}}\\ &= -\sum_{\ell=1}^{a} P_\ell \sum_{m=1}^{a} \Pi_{\ell m} log\, \Pi_{\ell m}\end{aligned}$$

In general:

$$\eta(F_{n-1}|F_0 \times F_1 \times \cdots \times F_{n-2}) = -\sum_{\ell=1}^{a} P_\ell \sum_{m=1}^{a} \Pi_{\ell m} log\, \Pi_{\ell m}$$

Hence:

$$\eta(F_0 \times F_1 \times F_2 \times \cdots \times F_{n-1}) = \eta(F_0) - (n-1)\sum_{\ell=1}^{a} P_\ell \sum_{m=1}^{a} \Pi_{\ell m} log\, \Pi_{\ell m}$$

Now it is plain that:

$$\lim_{n\to\infty}\frac{1}{n}\eta(F_0 \times F_1 \times F_2 \times \cdots \times F_{n-1}) = -\sum_{\ell=1}^{a} P_\ell \sum_{m=1}^{a} \Pi_{\ell m} log\, \Pi_{\ell m}$$

which (by definition) is the entropy of the markov process.