pfsl24.tex

Lecture 24. 4.12.2012

Proof. For $n = 2$:

$$\begin{aligned}
p_{ij}^{(2)} &= P(i \to j \text{ in 2 steps}) \\
&= \textstyle\sum_k P(i \to k \to j) \\
&= \textstyle\sum_k P(i \to k \text{ on first step}).P(k \to j \text{ on second step}|i \to k \text{ on first step}) \\
&= \textstyle\sum_k P(i \to k).P(k \to j),
\end{aligned}$$

using the Markov property in the second term. This says that

$$p_{ij}^{(2)} = \textstyle\sum_k p_{ik} p_{kj},$$

the (i, j) element of the second matrix power P^2.

For the general case we can use induction on the power n. Alternatively, we can argue as follows. The probability of going from i to j in n steps is, summing over all possible paths from i to j in n steps,

$$p_{ij}^{(n)} = \textstyle\sum_{k_1,\ldots,k_{n-1}} P(i \to k_1).P(k_1 \to k_2|i \to k_1).P(k_2 \to k_3|i \to k_1 \to k_2)$$

$$\ldots P(k_{n-1} \to j|i \to k_1 \to \ldots \to k_{n-1}),$$

by iterated conditional expectation. Using the Markov property,, the RHS simplifies to

$$p_{ij}^{(n)} = \textstyle\sum_{k_1,\ldots,k_{n-1}} P(i \to k_1).P(k_1 \to k_2).P(k_2 \to k_3) \ldots P(k_{n-1} \to j).$$

The LHS is the (i, j) element of $P^{(n)}$, while the RHS is the (i, j) element of the nth matrix power P^n of P. Since this holds for all i and j, the two matrices are equal, as required. //

This result is vital. It shows one of the great advantages of Markov chain theory – that it is perfectly adapted to the theory of matrices and Linear Algebra, which is very well developed.

Note. The result is named after Sydney CHAPMAN (1888-1970), an English applied mathematician (paper of 1928) and Kolmogorov (paper of 1931).

Initial distribution. Suppose that the position at time $t = 0$ is random, with

$$p_i := P(X_0 = i).$$

Form the *row-vector*

$$p := (p_0, p_1, \ldots).$$

Then

$$\begin{aligned} P(X_n = j) &= \textstyle\sum_i P(X_n = j \& X_0 = i) \\ &= \textstyle\sum_i P(X_0 = i) P(X_n = j | X_0 = i) \\ &= \textstyle\sum_i p_i p_{ij}^{(n)} \\ &= (pP^{(n)})_j. \end{aligned}$$

That is, the *row-vector* $pP^{(n)} = pP^n$ gives the distribution of the chain at time n.

Note. 1. Because it is natural to specify where we are at one time (at i with probability p_i), and then where we go to next (go from i to j with probability p_{ij}), it is *row*-vectors, rather than column-vectors, that are more useful in Markov chain theory.

This is worth bearing in mind, as in Linear Algebra the convention is often adopted that vectors are *column*-vectors (by default – i.e., unless otherwise specified), in which case one needs to use a transpose sign (A^T denotes the transpose of a matrix A) to obtain a row-vector. This is actually unnecessary here: vectors, row or column, are special cases of matrices, and it is better not to clutter things up with unnecessary transpose signs.

2. Precisely for this reason, one sometimes sees p_{ji} used for what we call p_{ij}, as in e.g. [M], Ch. 3: Markov processes.

Beware of this if using this otherwise excellent book!

Stationary distribution.

Suppose that the initial distribution π satisfies the linear equations

$$\pi P = \pi. \qquad (SD)$$

Then by above, its distribution after one step is $\pi P = \pi$. Similarly, its distribution after n steps is

$$\pi P^{(n)} = \pi P^n = \pi P.P^{n-1} = \pi P^{n-1} = \pi P^{n-2} = \ldots = \pi P = \pi :$$

the distribution stays the same for all time. Such a distribution is called *stationary*, or *invariant*, or an *equilibrium distribution.* We shall return to such distributions later, when we shall see that they are (under broad conditions) *limiting distributions*, to which the chain settles down as time passes.

Observe that the linear equations (SD) are homogeneous: if π is a solution, then so is $c\pi$ for any scalar c. We are only interested in solutions $\pi = (\pi_j)$ which are probability distributions, i.e. $\pi_j \geq 0$, $\sum_j \pi_j = 1$. There may well be solutions but *not* solutions of this type; we shall meet examples of this below.

Examples.

1. *Two states.* This is the simplest possible case:

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}.$$

There are two common interpretations:

(i) Motion on the line with constant speed,

$$\alpha = P(\text{change direction to left}|\text{going right}), \quad \beta = P(\text{change direction to right}|\text{going left}).$$

(ii) Rainfall. This chain has been used to model rainfall data, with days in Tel Aviv being classified as dry (if no rain falls) and wet otherwise. It gives a reasonable fit to the Tel Aviv rainfall data. For details, see [CM], 3.2.

2. *Gambler's ruin: Random walk with absorbing barriers on a finite set.* Here

$$P = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ q & 0 & p & \dots & 0 & 0 & 0 \\ 0 & q & 0 & p & \dots & 0 & 0 \\ \ddots & \ddots & \ddots & \ddots & & & \\ \dots & \dots & \dots & \dots & q & 0 & p \\ \dots & \dots & \dots & \dots & 0 & 0 & 1 \end{pmatrix}.$$

Random walk is given by an infinite matrix on the integers, with the tri-diagonal structure above (0 diagonal, p in the super-diagonal, q in the sub-diagonal throughout).

3. *Gambling for fun: Random walk with reflecting barriers on a finite set.* If our gamblers are playing for fun rather than for money, they may decide that to avoid the game stopping when a player is ruined, his last stake is

returned to him so that he can continue playing. The matrix is replaced by

$$P = \begin{pmatrix} q & p & 0 & \dots & 0 & 0 & 0 \\ q & 0 & p & \dots & 0 & 0 & 0 \\ 0 & q & 0 & p & \dots & 0 & 0 \\ \ddots & \ddots & \ddots & \ddots & & & \\ \dots & \dots & \dots & \dots & q & 0 & p \\ \dots & \dots & \dots & \dots & 0 & q & p \end{pmatrix}.$$

4. *Cyclic random walk.* Suppose the states represent positions on a circle:

$$P = \begin{pmatrix} q_0 & q_1 & \dots & \dots & q_{a-1} \\ q_{a-1} & q_0 & \dots & \dots & q_{a-2} \\ \ddots & \ddots & \ddots & \ddots & \\ q_1 & q_2 & \dots & q_{a-1} & q_0 \end{pmatrix}.$$

5. *Ehrenfest model of diffusion: Ehrenfest urn.* Suppose that N balls are distributed between two urns. At each stage, a ball is chosen at random (each with probability $1/N$) and changed to the *other* urn. The state is the number of balls in Urn 1. Then

$$p_{i,i-1} = i/N, \qquad p_{i,i+1} = 1 - i/N, \qquad p_{i,j} = 0 \quad \text{otherwise}$$

(the first represents the chance that a ball in Urn 1 is chosen, and changed to Urn 2, the second that a ball in Urn 2 is chosen, with the complementary probability, and changed to Urn 1). The matrix is again tri-diagonal:

$$P = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 & 0 & 0 \\ 1/N & 0 & 1-1/N & \dots & 0 & 0 & 0 \\ 0 & 2/N & 0 & 1-2/N & \dots & 0 & 0 \\ \ddots & \ddots & \ddots & \ddots & & & \\ \dots & \dots & \dots & \dots & 1-1/N & 0 & 1/N \\ \dots & \dots & \dots & \dots & 0 & 1 & 0 \end{pmatrix}.$$

The motivation for this model is Statistical Mechanics (Paul EHRENFEST (1880-1933) and Tatyana Ehrenfest, in 1907, published in 1911). The balls represent molecules of a gas (so for a physically observable system, will be present in enormous numbers – recall *Avogradro's number*, c. 6.02×10^{23}, is the number of gas molecules per standard volume under standard conditions).