pfsl27.tex

Lecture 27. 11.12.2012

Let $N \to \infty$:

$$s := \sum_{j=1}^{\infty} \pi_j \leq 1.$$

Now

$$p_{ij}^{(n+1)} = \sum_k p_{ik}^{(n)} p_{kj} \geq \sum_{k=1}^{N} p_{ik}^{(n)} p_{kj},$$

for each N. Let $n \to \infty$:

$$\pi_j \geq \sum_{k=1}^{N} \pi_k p_{kj}.$$

Let $N \to \infty$:

$$\pi_j \geq \sum_k \pi_k p_{kj}. \qquad (*)$$

Sum over j:

$$s = \sum_j \pi_j \geq \sum_j \sum_k \pi_k p_{kj} = \sum_k \pi_k \sum_j p_{kj} = \sum_k \pi_k = s.$$

So the inequality we got by summing $(*)$ is an *equality* (the extreme left and right are both s). So $(*)$ must itself be an equality (as inequality would contradict this). This proves (iii).

That $\sum_j \pi_j = 1$ follows formally from $\sum_j p_{ij}^{n)} = 1$ and $p_{ij}^{(n)} \to \pi_j$ $(n \to \infty)$ on interchanging $n \to \infty$ and $\sum_j$. This follows by dominated convergence – or see e.g. [GS] 6.4, pp 207-217. //

The distribution $\pi = (\pi_j)$ is called the *limit distribution* of the chain. It is also an *invariant distribution*, or *stationary distribution*, in the sense that if π is the *initial* or *starting distribution*, the distribution after one step is πP, which is also π as $\pi = \pi P$, and similarly after n steps. So:

Cor. If an ergodic chain is started in its invariant or limit distribution π, it stays in distribution π for all time.

Examples.

1. *Gambler's ruin.* There is no limit distribution. The chain is not irreducible. The extreme states 0, a are absorbing; the others are transient. There are two different invariant distributions: 'start in 0 and 'start in a'.

2. *Ehrenfest urn.* Again, there is no limit distribution: the chain is periodic

with period 2. but apart from this, the chain comes as close to having a limit distribution as possible: it has an *invariant* distribution, the *binomial* distribution

$$\pi = (\pi_j), \qquad \pi_j = 2^{-d}\binom{d}{j}.$$

Recurrence time.

The mean recurrence time of state j is $\mu_j = 1/\pi_j$. So here

$$\mu_0 = 1/\pi_0 = 1/2^{-d} = 2^d.$$

Now d is of the order of Avogradro's number (6.02×10^{23}), so 2^d is astronomically vast. So π_0 is astronomically vast – effectively infinite. This means that in practice, we will not see the chain return to its starting position if started at 0 – even though it does so (infinitely often, almost surely).

Rate of convergence.

The distribution at time n is governed by P^n by the Chapman-Kolmogorov theorem. In the periodic case, the d e-values that are dth roots of unity do not have nth powers that $\to 0$, but in the aperiodic case every e-value other than the PF e-value 1 does. From the Perron-Frobenius theorem, the *rate of convergence* is determined by the *spectral gap* $1 - |\lambda_2|$, where as usual we order the e-values in decreasing modulus:

$$\lambda_1 = 1 > |\lambda_2| \geq \ldots \geq \ldots$$

(recall there is only one e-value of modulus 1, the PF e-value 1). See the handout for the algebraic details.

Reversibility.

The chain is *reversible* if its probabilistic structure is invariant under time-reversal (i.e., the chain looks the same if run backwards in time). We quote (Kolmogorov's theorem) that this is the same as *detailed balance*:

$$\pi_i p_{ij} = \pi_j p_{ji} \qquad \text{for all } i,\ j. \qquad (DB)$$

One can check (DB) here. So the Ehrenfest chain is reversible.

The interpretation of this in Statistical Mechanics is that μ_0 is the mean recurrence time of state 0, when all the $2d$ gas molecules are in one half of the container. Although this state is certain to recur, its mean recurrence time is so vast as to be effectively infinite – which explains why we do not see such states recurring in practice! This reconciles the theoretical reversibility

of the model with the irreversible behaviour we observe when gases diffuse, etc. This was the Ehrenfests' motivation for their model, in 1912.
Note. Relevant here is the concept of *entropy* – a measure of disorder. For details, see Problems and Solutions 9 and 10.
Markov Chain Monte Carlo (MCMC).

The area of *Markov chain Monte Carlo (MCMC)*, for which see e.g. Häggström [Hag] Ch. 7, 8, originated in physics, but has since become extremely important in statistics, particularly Bayesian statistics (for which see e.g. SMF, IV). The idea is to sample, or simulate, from a distribution π. If this is straightforward, fine (see e.g. IS II for simulation) – but it may not be. In this case, the method of MCMC is to find a Markov chain $X = (X_n)$ with π as its limit distribution. Then we can run the chain, knowing that its distribution for large n will approximate π. How long we have to wait for the approximation to be good enough for our purposes depends on the transition matrix P of the chain – and in particular, on its spectral gap.
Note. The two most important developments in Statistics in recent decades have been MCMC and wavelets.

4. Finite and infinite chains

Finite chains have special and useful properties.

Theorem. For a finite Markov chain, it is impossible for all states to be transient: a finite chain must contain at least one persistent (= recurrent) state.

Proof. If the state-space is $\{1, \ldots, N\}$, for each i and each n

$$1 = \sum_{j=1}^{N} p_{ij}(n). \qquad (a)$$

Let $n \to \infty$: if j is transient, the total expected time in it is finite: $\sum_n p_{ij}(n) < \infty$. So

$$p_{ij}(n) \to 0 \qquad (n \to \infty). \qquad (b)$$

If *all* states were transient, then letting $n \to \infty$ in (a) and using (b) would give the contradiction $1 = 0$. So not all states in a finite chain can be transient. //

Note. 1. An infinite chain can easily consist of only transient states. A trivial example is walk to the right on the integers: $p_{i,i+1} = 1$, with the other p_{ij} 0.

A non-trivial example is given by Pólya's theorem: simple symmetric random walk on the integer lattice $\mathbf{Z}^d$ is transient for $d \geq 3$ (but recurrent for $d = 1, 2$). See e.g. [F], XIV.7, [GS], 13.11 p.560.
2. The sum $\sum_n p_{ij}(n)$ is the expected total time spent in state j, starting from i. With only finitely many states, and infinite total time altogether, at least one of these sums must thus be infinite.

Theorem. A persistent state j in a finite chain is positive (= non-null).

Proof. If the finite chain has state-space $\{1, \ldots, N\}$, assume there is a null state. Let C be the equivalence class containing it. Since C is closed, we can consider the subchain induced on C. Then

$$1 = \sum_{k \in C} p_{ik}(n) \qquad \text{(finite sum)}.$$

Let $n \to \infty$: each $p_{ik}(n) \to 0$, so the sum on the RHS $\to 0$, giving $1 = 0$. This contradiction gives the non-existence of null states in a finite chain. //

The restriction to *finite* chains is essential here: e.g., simple symmetric random walk on the integers has all states persistent null.

The limit theorem above is due to Kolmogorov in 1936. The algebraic treatment we have given is in terms of matrices – and in the case of an infinite chain, these will be infinite matrices. Dealing with infinite rather than finite matrices is possible (with care, and under suitable conditions) – but belongs to Functional Analysis rather than to Linear Algebra. Infinite-dimensional versions of the Perron-Frobenius theorem exist, such as the *Krein-Rutman theorem* for positive operators. But this leads beyond the scope of this course.
Continuous state-space

It turns out that, although the language of matrices is so useful in the above, one can extend much of the treatment above to situations where the state space is *continuous* rather than discrete. It turns out also that it is this case that is most useful in applications, particularly MCMC. For a full treatment, see e.g. Meyn & Tweedie [MT]. Much of the theory above extends to the continuous-state case. Again the transience-recurrence dichotomy is crucial, but there are now various possible types of recurrence. One of the most important is *Harris recurrence* (T. E. HARRIS (1919-2005) in 1956).