Markov chains and mixing times (part 3)

Vaughn Climenhaga

March 6, 2013

The previous post introduced the idea of coupling for Markov chains as a method for estimating mixing times. Here we mention a particular example of a coupling that is often useful – this is the *classical* coupling, or *Doeblin coupling*, after Wolfgang Doeblin.

Let X_n be a Markov chain with state space S and transition probability matrix P. As usual, assume that P is irreducible and aperiodic, so that there is a unique stationary distribution π, and let λ be the initial distribution of X_n.

Let Y_n be another Markov chain over P, with initial distribution π. Let X_n and Y_n evolve independently of each other, and consider the stopping time

$$T = \min\{m \geq 0 \mid X_m = Y_m\}. \tag{1}$$

The time T is a random variable, and we can define a new stochastic process Z_n, depending on X_n, Y_n, and T, by

$$Z_n = \begin{cases} X_n & n < T, \\ Y_n & n \geq T. \end{cases} \tag{2}$$

Now the pair (Y_n, Z_n) is a coupling of the Markov chain. Each of Y_n, Z_n evolves according to the transition probabilities in P, but after time T the pair (Y_n, Z_n) lives on the diagonal of $S \times S$, and so in particular if we write λ_n for the distribution of X_n, the general coupling bound from the previous post gives

$$d_V(\lambda_n, \pi) \leq \mathbb{P}(T > n).$$

This bound can be computed directly as follows: for any $A \subset S$ we have

$$\begin{aligned}|\lambda_n(A) - \pi(A)| &= |\mathbb{P}(Z_n \in A) - \mathbb{P}(Y_n \in A)| \\ &= |\mathbb{P}(Z_n \in A \text{ and } T \leq n) + \mathbb{P}(Z_n \in A \text{ and } T > n) \\ &\quad - \mathbb{P}(Y_n \in A \text{ and } T \leq n) - \mathbb{P}(Y_n \in A \text{ and } T > n)| \\ &= |\mathbb{P}(Z_n \in A \text{ and } T > n) - \mathbb{P}(Y_n \in A \text{ and } T > n)| \\ &\leq \mathbb{P}(T > n),\end{aligned}$$

where the third equality uses the fact (from the definition of Z_n) that $Z_n = Y_n$ whenever $n \geq T$.

Now suppose that instead of beginning in the stationary distribution π, the Markov chain Y_n begins in another distribution λ'. We would like to estimate the distance between the distributions $\lambda_n = \lambda P^n$ and $\lambda'_n = \lambda' P^n$, which corresponds to *memory loss* in the Markov chain. If we define the hitting time T by (1) and the coupling Z_n by (2), then the same argument gives

$$d_V(\lambda_n, \lambda'_n) \leq \mathbb{P}(T > n),$$

and so we see that coupling techniques also estimate the memory loss in the Markov chain.

We remark that it can be shown that for each $x \in S$, not only does the quantity $\Delta_x(m) = d_V(p_x^m, \pi)$ converge to 0, but it does so monotonically.

Later on in this series, we will explore the connection between these techniques and smooth dynamics. For *uniformly hyperbolic* maps, Markov partitions can be used to connect the dynamics of a diffeomorphism to a finite-state Markov chain. For *non-uniformly hyperbolic* maps, the situation is more subtle and a countable-state Markov chain must be used – this will lead us to a discussion of Young towers and subexponential decay of correlations, for which coupling techniques can be used to get results even when spectral techniques fail because the transfer operator does not act with a spectral gap.