
Shuffling Cards and Genes

Let us consider the following modified version of the card shuffle. As before, the card shuffle is a Markov chain on the permutation group S_n , but this time the permutation σ_t obeys the following Markov chain: at each time step, pick an index i_{t+1} uniformly from $\{1, \ldots, n-1\}$ and set

$$\sigma_{t+1}(j) = \begin{cases} \sigma_t(j+1) & j = i_{t+1} \\ \sigma_t(j-1) & j = i_{t+1} + 1 \\ \sigma_t(j) & \text{otherwise} \end{cases}$$

If we were allowed to transpose any two elements at random, we found that this had a mixing time $\tau \sim n \log n$. In this problem, we will show that the above Markov chain, with only neighboring transpositions allowed, has a mixing time $\tau \leq n^3 \log n$.

To solve this problem, we will use coupling. The chains we will couple will be the forward and backward chains $\sigma_t^{\rm F}$ and $\sigma_t^{\rm B}$, which start from the initial conditions $\sigma_0^{\rm F}(a) = \sigma_0^{\rm B} = (n-a)$. The chains will each evolve as the chain above, *except* in the case where $\sigma_t^{\rm F}(i_{t+1}) = \sigma_t^{\rm B}(i_{t+1}+1)$ or $\sigma_t^{\rm B}(i_{t+1}) = \sigma_t^{\rm F}(i_{t+1}+1)$. In this case, one of the chains is picked uniformly at random, and only that chain is switched.

(a) Let us denote $D_t(a)$ as the (positive-valued) distance between a in the F and B chains. Show that $|D_{t+1}(a) - D_t(a)| \le 1$, and that

$$P\left(D_{t+1}(a) = D_t(a) + 1 \left| D_t(a), \sigma_t^{\mathrm{F}}, \sigma_t^{\mathrm{B}} \right) = \frac{1}{n-1} \quad (0 < D_t(a) \le n-1)$$
$$P\left(D_{t+1}(a) = D_t(a) + 1 \left| D_t(a), \sigma_t^{\mathrm{F}}, \sigma_t^{\mathrm{B}} \right) \le \frac{1}{n-1} \quad (0 < D_t(a) < n-1)$$

and show that $D_t(a) = 0$ is an absorbing state.

(b) Let X_t be a random walk on $\{0, \ldots, n-1\}$ with the same transition rules as above (but with the latter equation replaced by an equality): i.e., with probability $(n-1)^{-1}$, the walker will move either left or right, except at a boundary or at 0. Let

$$\tau_a = \inf(t : D_t(a) = 0),$$

$$\tau_a^X = \inf(t : X_t = 0 | X_0 = n + 1 - 2a).$$

Show that $\tau_a \leq \tau_a^X$, if we appropriately match the steps of the random walk to the shuffle.

- (c) Using techniques that we used for random walks on the line, find the explicit expression for τ_a^X .
- (d) Use Markov's inequality to show that $P(\tau_a > n^3) < 1/2$ for any a, and use this result to determine a good upper bound for $P(\tau_a > 2n^3 \log_2 n)$.
- (e) Conclude that the mixing time $\tau \leq 2n^3 \log n$ in the limit of n large.

With some clever techniques for estimating lower bounds on mixing times, one can show that $\tau \sim n^3 \log n$ precisely, although this is above the scope of this problem.

This is also a toy model for the biological problem of mutations in DNA. As DNA. which is a long sequence of base pairs, is replicated in cells, sometimes the cell makes a mistake, and may transpose two base pairs. We thus have a rough time scale for how long it should take for us to "lose track" of what our ancestors' DNA was.