Problem Set 2

Out: 13 Nov.; Due: 11 Dec.

Take time to write **clear** *and* **concise** *answers. None of the problem parts requires a long solution; if you find yourself writing a lot, you are either on the wrong track or giving an inappropriate level of detail. You are encouraged to form small groups (two or three people) to work through the problems, but you should write up your solutions on your own. Put your solutions in the box outside my office (677 Soda) by 5pm on Friday December 11.*

1. Recall the toy example in Lecture 9, namely the Markov chain with states $\{0, 1\}$ and transition probabilities $P(0, 1) = 1$, $P(1, 0) = P(1, 1) = \frac{1}{2}$. This is ergodic with stationary distribution $\pi = (\frac{1}{3}, \frac{2}{3})$.

(a) Show that there is a unique random map that is consistent with this Markov chain.

(b) Perform a direct, hands-on calculation which shows that Coupling from the Past on this example generates precisely the distribution π.

(c) Verify that the same protocol using forward simulation generates the (disastrous) distribution $(0, 1)$.

2. In this problem we will see that Coupling from the Past (CFTP) is in a certain sense an (almost) optimal technique when the Markov chain is described by a monotone coupling. Suppose we have an ergodic Markov chain with state space Ω, transition probabilities $P(\cdot\,, \cdot)$ and stationary distribution π. Suppose also that there is a complete coupling $f : \Omega \to \Omega$ that is consistent with P, and that f is monotone with respect to some partial order on Ω which has unique maximum and minimum elements $\top$ and $\perp$. Denote by $D(t)$ the quantity $\max_{x,y} \|P_x^t - P_y^t\|$ (the variation distance at time t maximized over initial states x, y), and by T the stopping time for CFTP as defined in Lecture 9 (i.e., $T = \min\{t :$ the composed map F_{-t}^0 is constant$\}$).

(a) Let T' be the stopping time for the *forward* simulation under f, i.e., $T' = \min\{t : F_0^t$ is constant$\}$. Argue that T and T' have the same distribution. (This conceptually simplifies arguments about the distribution of T.)

(b) Show that $D(t) \leq \Pr[T > t]$.

(c) Show that $D(t) \geq \frac{1}{h} \Pr[T > t]$, where h is the *height* of the partial order, i.e., the length of a longest chain of comparable elements in it. [HINT: Consider the random variable $H(X_t, Y_t)$, the height difference in the partial order between two coupled states at time t.]

(d) Deduce that, with probability at least $1 - \varepsilon$, CFTP will terminate within $O(\tau_{\text{mix}} \log(h\varepsilon^{-1}))$ steps.

3. The *random cluster model* is a "dual" of the Ising model. The input is as usual a graph $G = (V, E)$, but now configurations are subgraphs $H = (V, E')$, where $E' \subseteq E$. The Gibbs distribution takes the form

$$\pi(H) = \frac{1}{Z}\, p^{|H|}(1 - p)^{|E|-|H|} 2^{\mathcal{C}(H)},$$

where $|H|$ and $\mathcal{C}(H)$ denote the number of edges and the number of connected components in H respectively, $p \in [0, 1]$ is a parameter that depends on temperature, and Z is the partition function.

(a) Recall that the Gibbs distribution for the Ising model may be written as

$$\widehat{\pi}(\sigma) = \frac{1}{\widehat{Z}}\, \lambda^{\mathcal{U}(\sigma)},$$

where $\lambda \in [0, 1]$ depends on temperature and $\mathcal{U}(\sigma)$ is the number of *unaligned* neighbors in the spin configuration $\sigma \in \{\pm\}^V$. Show that the random cluster model is equivalent to the Ising model under the correspondence $p = 1 - \lambda$, in the following strong sense: the partition functions Z and $\widehat{Z}$ are equal, and moreover, given a random sample from π it is simple to generate a random sample from $\widehat{\pi}$, and vice versa.

(b) Write down the heat bath Markov chain for the random cluster model, based on flipping a single *edge* of G. Show that, with a suitable choice of partial order on the state space, this process can be described by a *monotone* coupling analogous to those we saw in Lecture 9.

4. Let P be an ergodic, reversible Markov chain, and τ_{mix} its mixing time. Show that there always exists a flow f with cost $\rho(f) \leq C\tau_{\mathrm{mix}}$, where C is a universal constant. [HINT: To construct f, use the actual random paths of length about τ_{mix} generated by the Markov chain itself. Also, for the analysis, you may find it useful to prove the following lemma: for all states x, y, and all $t \geq 2\tau_{\mathrm{mix}}$, we have $\frac{p_x^t(y)}{\pi(y)} \geq \frac{1}{8}$. (There is nothing special about $\frac{1}{8}$ here; any positive constant would do.)]

5. Here is a relatively simple combinatorial example that illustrates the use of the flow technique discussed in Lecture 12. Consider card shuffling by random transpositions (Section 2.2.3 of Lecture 2 and Example 5.2.4 in Lecture 5), where the state space Ω consists of all permutations of an n-card deck and transitions are made by picking two cards i, j u.a.r. (without replacement) and switching them. We also assume a self-loop probability of $\frac{1}{2}$. We will represent a permutation $x \in \Omega$ as $x = (x_1, x_2, \ldots, x_n)$, where x_i denotes the ith card in x.

Consider the following flow. For each pair of distinct permutations $x, y \in \Omega$, all the flow from x to y is routed along a single path γ_{xy} defined as follows. For each value $k = 1, 2, \ldots, n$ in turn, we move card x_k from its current position to position k (if it is not there already) by exchanging it with the current card at position k.

(a) Use the flow encoding technique (Lecture 12) to show that the number of paths γ_{xy} that pass through any particular transition of the Markov chain is at most $|\Omega|$. [NOTE: You should do this implicitly, using an injective mapping, not by explicit counting.]

(b) Hence deduce that the mixing time of this Markov chain, for any initial state x, satisfies $\tau_x(\epsilon) = O(n^3(n \log n + \log \epsilon^{-1}))$. [NOTE: Recall from class that the true mixing time is $O(n \log n)$, so the flow technique is rather wasteful here. However, it does provide a non-trivial polynomial bound on τ_{mix}.]