Math 578B – Fall 2015 – Homework #2

Peter Ralph

due 8 September

1. Let X_n be the simplified "polymerase complex assembly" Markov chain defined in class, with transition matrix (where "†" means transcription):

$$P = \begin{array}{c} \\ \varnothing \\ \alpha \\ \beta \\ \alpha+\beta \\ \text{pol} \\ \dagger \end{array} \begin{array}{c} \begin{array}{cccccc} \varnothing & \alpha & \beta & \alpha+\beta & \text{pol} & \dagger \end{array} \\ \begin{bmatrix} * & k_\alpha & k_\beta & 0 & 0 & 0 \\ k_\alpha & * & 0 & k_\beta & 0 & 0 \\ k_\beta & 0 & * & k_\alpha & 0 & 0 \\ 0 & k_\beta & k_\alpha & * & k_{\text{pol}} & 0 \\ 0 & 0 & 0 & 0 & * & 1 \\ 0 & 0 & 0 & 1 & 0 & * \end{bmatrix} \end{array} \tag{1}$$

Here the "$*$"s on the diagonal are set so that rows sum to 1. For each state a, define the first *hitting* times

$$\tau_a = \min\{n \geq 0 : X_n = a\}$$

a. Since the factors α and β can stay on the DNA, there may be more than one transcript per "burst". Let N denote the number of transcripts before the next complete disassembly, i.e.

$$N = \sum_{k=0}^{\tau_\varnothing} \mathbf{1}_{X_k=\dagger}.$$

Find the set of linear equations solved by

$$u(a) = \mathbb{E}[N \mid X_0 = a].$$

b. Set $k_\alpha = k_\beta = 0.2$ and $k_{\text{pol}} = 0.5$, and compute numerically $u(a)$ for each a.

c. Verify your answer by simulation.

2. Suppose that X is a Markov chain on a state space $\mathcal{A}$, and that $\varnothing$ and † are two distinct sets of states in $\mathcal{A}$. Define

$$h(a) = \mathbb{P}^a\{\tau_\varnothing < \tau_\dagger\},$$

Show that

$$\mathbb{P}\{X_{n+1} = b \mid X_n = a \text{ and } \tau_\varnothing < \tau_\dagger\} = \frac{h(b)}{h(a)} P_{ab}.$$

3. Let X be a Markov chain on k-subsets of $[n] := \{1, 2, \ldots, n\}$, for some fixed $k \leq n/2$, defined as follows: Suppose the chain is in state $X_t = A$. With probability $1/2$, do nothing, so $X_{t+1} = A$. Otherwise, pick an element $a \in A$ uniformly at random, and another $b \in [n] \backslash A$ uniformly at random, and let $X_{t+1} = (A \cup \{b\}) \backslash \{a\}$. Show that the unique stationary distribution of X is uniform. (And, say why the "do nothing" step is necessary!)

4. *(soft TSP)* Suppose we have a distance matrix D with pairwise distances between each of n points, so that $D_{ij} = D_{ji}$ is the distance between point i and point j. The *length* of a given ordering of points $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$ is

$$L(\sigma) = \sum_{i=1}^{n-1} D_{\sigma(i),\sigma(i+1)}.$$

Note that this is not a *tour*, i.e., it doesn't loop back to the start. For a given value of the *temperature* $T > 0$, define

$$\pi_T(\sigma) = \exp\left(-L(\sigma)/T\right).$$

Write (and, explain) computer code to do the following.

a. Sample $n = 100$ points uniformly from the unit square $[0, 1]^2$, and compute D for this configuration. Use this D in subsequent problems.

b. Implement the Metropolis algorithm to sample from π_T using the proposal distribution from class: pick $j < k$ uniformly at random, and reverse the order of $\sigma(j), \ldots, \sigma(k)$. For $T = 10$ and $T = 0.05$ show a histogram of the distribution of $L(\sigma)$, where $\sigma \sim \pi_T$, and pictures of a few representative paths. Make sure to say what aspects of the output convince you that your algorithm is working.

c. For each of the Markov chains at $T = 10$ and $T = 0.05$, estimate $\operatorname{cor}\left[L(X_t), L(X_{t+n})\right]$ for a range of relevant n (large enough n that they become uncorrelated). Plot, and explain the differences between the two chains you see.