Homework 10 (due Friday, December 7)

Exercise 1. [Gibbs sampling]
Let $S = \{1, \ldots, N\}$ and $d \geq 1$. We would like to sample from a distribution π on S^d defined as

$$\pi(x) = \frac{g(x)}{Z}, \quad x \in S^d,$$

where g is some positive function on S^d and $Z = \sum_{x \in S^d} g(x)$ is the normalization constant, which we would like to avoid computing.

A possible way to handle this problem is the following.

1. Start from a state $x \in S^d$;

2. Choose an index $u \in \{1, \ldots, d\}$ uniformly at random;

3. Update the value of x_u to x'_u, which is sampled from the following conditional distribution:

$$\pi(x'_u | x_1, \ldots, x_{u-1}, x_{u+1}, \ldots, x_d) = \frac{\pi(x_1, \ldots, x_{u-1}, x'_u, x_{u+1}, \ldots, x_d)}{\sum_{y_u \in S} \pi(x_1, \ldots, x_{u-1}, y_u, x_{u+1}, \ldots, x_d)}$$

4. Repeat from 2.

What is the advantage of such a method? The above conditional probability can actually be rewritten as

$$\pi(x'_u | x_1, \ldots, x_{u-1}, x_{u+1}, \ldots, x_d) = \frac{g(x_1, \ldots, x_{u-1}, x'_u, x_{u+1}, \ldots, x_d)}{\sum_{y_u \in S} g(x_1, \ldots, x_{u-1}, y_u, x_{u+1}, \ldots, x_d)}$$

which only requires to compute one sum and not a multidimensional one, as required for computing the normalization constant Z.

Your task now is to formalize slightly the above algorithm by expressing it as a Markov chain $(X_n, \, n \geq 0)$ on S^d and

a) writing down its transition probabilities $p(x, y)$, $x, y \in S^d$;

b) showing that the detailed balance equation is satisfied, i.e. that $\pi(x) \, p(x, y) = \pi(y) \, p(y, x)$, for all $x, y \in S^d$.

Can therefore this algorithm be viewed as a Metropolis-Hastings algorithm?

% please turn the page

Exercise 2. On the state space $S = \{0, 1, 2\}$ and given $\beta > 0$, consider the following distribution:

$$\pi = \frac{1}{Z}\left(1, e^{-2\beta}, e^{-\beta}\right)$$

where the normalization constant $Z = 1 + e^{-2\beta} + e^{-\beta}$ is easy to compute in this case. For any given $\beta > 0$, we would like to sample from π, in order to obtain (by taking β large) an estimate of the global minimum of the function $f : S \to \mathbb{Z}$ defined as $f(0) = 0$, $f(1) = 2$ and $f(2) = 1$. Of course, in this situation, both finding the global minimum of f and sampling from the distribution π are trivial tasks, but the idea here is to get an idea of the performance (i.e. rate of convergence) of the Metropolis-Hastings algorithm in a simple case.

Consider the base chain on S with transition probabilities

$$\psi_{01} = \psi_{21} = 1 \quad \text{and} \quad \psi_{10} = \psi_{12} = \frac{1}{2}.$$

a) Compute the transition probabilities p_{ij} of the corresponding Metropolis chain.

b) Check that the detailed balance equation is satisfied.

c) Compute the eigenvalues $\lambda_0 \geq \lambda_1 \geq \lambda_2$ of P. (*Hint:* You already know that $\lambda_0 = 1$.)

d) Express the spectral gap γ as a function of β. How does it behave as β gets large?