Statistics 251b/551b, spring 2009
Homework #4 solutions

For each real number x, the positive part x^+ is defined as $\max(0,x)$ and the negative part is defined as $\max(0,-x)$. Note that $x = x^+ - x^-$ and $|x| = x^+ + x^-$.

[1] *Suppose λ and μ are probability measures on a countable set $\mathcal{S}$. The total variation distance $\|\lambda-\mu\|_{\text{TV}}$ is defined as $\sup_{A\subseteq\mathcal{S}}|\lambda A - \mu A|$.*

(i) Show that the supremum in the definition must be achieved either by the set $A_0 = \{i\in\mathcal{S} : \lambda_i \ge \mu_i\}$ or by the set $A_1 = \{i\in\mathcal{S} : \lambda_i \le \mu_i\}$.

(ii) Deduce that $\|\lambda-\mu\|_{\text{TV}} = \max(\alpha_0,\alpha_1)$ where $\alpha_0 = \sum_{i\in\mathcal{S}}(\lambda_i-\mu_i)^+$ and $\alpha_1 = \sum_{i\in\mathcal{S}}(\lambda_i-\mu_i)^-$.

(iii) Show that $\alpha_0 = \alpha_1 = \frac{1}{2}\sum_{i\in\mathcal{S}}|\lambda_i-\mu_i|$.

For each $A\subseteq\mathcal{S}$,

$$\lambda A - \mu A = \sum\nolimits_{i\in A}(\lambda_i-\mu_i)$$

The right-hand side is maximized when A contains precisely those i for which $\lambda_i - \mu_i \ge 0$, that is, when $A = A_0$. It is minimized (that is, it is the most negative) when $A = A_1$. Note that

$$(\lambda_i-\mu_i)^+ = \begin{cases} \lambda_i-\mu_i & \text{if } i\in A_0 \\ 0 & \text{if } i\in A_0^c \end{cases}$$

and

$$(\lambda_i-\mu_i)^- = \begin{cases} -(\lambda_i-\mu_i) & \text{if } i\in A_1 \\ 0 & \text{if } i\in A_1^c \end{cases}$$

The equalities in part (iii) follow from

$$\sum\nolimits_{i\in\mathcal{S}}\left((\lambda_i-\mu_i)^+ + (\lambda_i-\mu_i)^+\right) = \sum\nolimits_{i\in\mathcal{S}}|\lambda_i-\mu_i|$$

and

$$\sum\nolimits_{i\in\mathcal{S}}\left((\lambda_i-\mu_i)^+ - (\lambda_i-\mu_i)^+\right) = \sum\nolimits_{i\in\mathcal{S}}\lambda_i - \sum\nolimits_{i\in\mathcal{S}}\mu_i = 1-1=0$$

[2] *Let $\mathcal{G}$ be a finite, connected graph with vertex set $\mathcal{S}$ and edge set $\mathcal{E}$. For each edge e suppose w_e is a strictly positive weight. Define $W_i = \sum_{\{i,j\}\in\mathcal{E}} w_{ij}$. The random walk on the weighted graph has transition probabilities*

$$Q(i,j) = w_{i,j}/W_i \qquad \textit{if } \{i,j\}\in\mathcal{E}.$$

Suppose λ is a probability distribution on $\mathcal{S}$ for which $\max_{i\in\mathcal{S}}\lambda_i/W_i > \min_{i\in\mathcal{S}}\lambda_i/W_i$.

(i) Explain why there must exist at least one edge $\{i,j\}$ for which $\lambda_i/W_i > \lambda_j/W_j$.

(ii) Explain why the the chain with transition probabilities

$$P(i,j) = Q(i,j)\min\left(1, \frac{\lambda_j Q(j,i)}{\lambda_i Q(i,j)}\right) \qquad \textit{for } \{i,j\} \in \mathcal{E}$$

is irreducible, aperiodic, and positive recurrent.

Many of you noticed that the assertions would fail, even if you could make sense of the definition of $P(i,j)$, if any of the λ_i's were zero. In fact, one runs Metropolis-Hastings only to get convergence to a stationary probability distribution. For an irreducible, finite state space, such a stationary probability distribution must give strictly positive weight to every state; one could not hope for λ to be a stationary distribution if λ_i were zero some i. So assume $\lambda_i > 0$ for every i in $\mathcal{S}$.

Connectedness of the graph ensures existence of a path $i_0 \mapsto i_1 \mapsto \ldots \mapsto i_k$ along the edges of the graph between each pair of vertices i_0 and i_k. For this path, each $w(i_{\alpha-1}, i_\alpha)$ is strictly positive, which makes each $Q(i_{\alpha-1}, i_\alpha)$, and hence each $P(i_{\alpha-1}, i_\alpha)$, strictly positive. The P-chain is irreducible.

Positive recurrence follows from HW 1.1.

Define $M = \max_{i\in\mathcal{S}} \lambda_i/W_i$ and $m = \min_{i\in\mathcal{S}} \lambda_i/W_i$. By irredicibility there must exist some path leading from the set $\{i \in \mathcal{S} : \lambda_i/W_i = M\}$ to the set $\{i \in \mathcal{S} : \lambda_i/W_i = m\}$. Somewhere along the path there must be an edge $\{i,j\}$ with $\lambda_i/W_i > \lambda_j/W_j$. For this edge,

$$\frac{\lambda_j Q(j,i)}{\lambda_i Q(i,j)} = \frac{\lambda_j w_{i,j}/W_j}{\lambda_i w_{i,j}/W_i} < 1$$

so that $P(i,j) < Q(i,j)$. Let $\mathcal{N}_i$ denote the set of all vertices k for which $\{i,k\} \in \mathcal{E}$. For Metropolis-Hastings,

$$P(i,i) = 1 - \sum\nolimits_{k\in\mathcal{N}_i} P(i,k),$$

which is strictly greater than $1 - \sum_{k\in\mathcal{N}_i} Q(i,k) \geq 0$ because the vertex j belongs to $\mathcal{N}_i$ and $P(i,k) \leq Q(i,k)$ for all k. It follows that state i is aperiodic, which makes the whole (irreducible) chain aperiodic.

[3] *Chang Problem 2.20. [Facts about the top-in-at-random shuffle: irreducible; aperiodic; stationary probability distribution is the uniform distribution on $\mathcal{S}$.]*

To avoid notational confusion, I will write $i, j, \ldots$ to denote only integers between 1 and d, reserving $\sigma, \tau, \ldots$ for elements of $\mathcal{S}$, the set of all $d!$ permutations of the numbers $1, 2, \ldots, d$.

Aperiodicity is easy because $P(\sigma,\sigma) = 1/d$ for every σ.

For ***irreducibility*** , consider two different permutations, $\sigma = (\sigma_1, \ldots, \sigma_d)$ and τ. We need to find a path with positive probability that leads from σ

to τ. The following steps, which each have probability $1/d$, provide such a path.

First move the card labelled σ_1 below the card labelled σ_d. Then move the new top card, which is labelled σ_2, below the σ_d card into the one of the two slots that puts cards σ_1 and σ_2 into the same relative order they have in τ. For example, if $\sigma_1 = 5, \sigma_2 = 11, \sigma_d = 3$ and if card 5 comes somewhere after card 11 in the τ permutation, then card number 11 is placed below card number 3 and above card number 5 at the end of the deck.

And so on. At the kth step, the card number σ_k is placed below the card numbered σ_d into the slot that gives the cards numbered $\sigma_1, \sigma_2, \dots, \sigma_k$ the same relative order as the corresponding cards in the τ permutation.

When the card numbered σ_d is finally placed in its appropriate slot then the ordering of the deck is given by the permutation τ.

Finally, I need to show that the uniform distribution, π, on $\mathcal{S}$ is ***stationary*** . It might seem intuitively obvious that uniform followed by a mindless shift of the top card should equal uniform, but the question asked for a more rigorous argument.

I need to show $\mathbb{P}_\pi\{X_1 = \sigma\} = 1/d!$ for each σ in $\mathcal{S}$. Condition on the value of X_0 and on the slot number, N, into which the top card is moved. (If $N = 1$ then the card stays on top; if $N = d$ it is moved to the bottom of the deck.) By assumption, X_0 and N are independent,

$$\mathbb{P}_\pi\{X_0 = \tau, N = k\} = \frac{1}{d!} \times \frac{1}{d} \qquad \text{for } \tau \in \mathcal{S} \text{ and } 1 \leq k \leq d.$$

For a fixed σ, let $\tau^{(k)}$ denote the permutation obtained by moving the kth element of σ back to position 1. For example, if $\sigma = (2, 5, 3, 1, 4)$ then $\tau^{(4)} = (1, 2, 5, 3, 4)$. Note: if $X_0 = \tau$ and $N = k$ then $X_1 = \sigma$ if and only if $\tau = \tau^{(k)}$. Put another way,

$$\mathbb{P}_\pi\{X_1 = \sigma \mid X_0 = \tau, N = k\} = \begin{cases} 1 & \text{if } \tau = \tau^{(k)} \\ 0 & \text{otherwise.} \end{cases}$$

Now I can condition.

$$\begin{aligned} &\mathbb{P}_\pi\{X_1 = \sigma\} \\ &\quad = \sum\nolimits_{k=1}^{d} \sum\nolimits_{\tau \in \mathcal{S}} \mathbb{P}_\pi\{X_1 = \sigma \mid X_0 = \tau, N = k\} \mathbb{P}_\pi\{X_0 = \tau, N = k\}. \end{aligned}$$

For each k, the only nonzero τ term is $\tau^{(k)}$. Thus

$$\mathbb{P}_\pi\{X_1 = \sigma\} = \sum\nolimits_{k=1}^{d} 1 \times \mathbb{P}_\pi\{X_0 = \tau^{(k)}, N = k\} = \frac{1}{d!}.$$

The stationary probability distribution is π.