Math 832, Assignment #2: Solutions

1. The claim is false. For example, let $S = \{1, 2\}$ and let $p(1, 2) = p(2, 1) = 1$, $p(1, 1) = p(2, 2) = 0$. Then, for any non-negative function f on S, the condition relating $p(x, y)$ to $p(x, y')$ is vacuously satisfied. However, the only stationary measures for $p(\cdot, \cdot)$ are counting measure and scalar multiples thereof. Thus, if we take $f(1) \neq f(2)$, we see that μ is not an stationary measure.

Here's what I should have given as the problem:

The heat bath Markov chain (corrected): Suppose we have a recurrent Markov chain on a finite state space S and a positive function f on S with the following properties: (a) $p(x, x) > 0$ for all x; (b) $p(x, y) > 0$ implies $p(y, x) > 0$; (c) $p(x, y) > 0$ and $p(y, z) > 0$ implies $p(x, z) > 0$; and (d) whenever $p(x, y)$ and $p(x, y')$ are both non-zero, $p(x, y)/p(x, y') = f(y)/f(y')$. Let $Z = \sum_{x \in S} f(x)$. Show that the probability distribution μ defined by $\mu(x) = f(x)/Z$ is an stationary measure for $p(.,.)$.

Solution: Fix x, and let $C = \{y \in S : p(x, y) > 0\}$. Note that $p(w, z) > 0$ for all $w, z \in C$. Since $p(x, y)$ is proportional to $f(y)$ for all $y \in C$, we have $p(x, y) = f(y)/Z_C$, where $Z_C = \sum_{z \in C} f(z)$. If $y \in C$, we have $\mu(x)p(x, y) = (f(x)/Z)(f(y)/Z_C) = (f(y)/Z)(f(x)/Z_C) = \mu(y)p(y, x)$. If $y \notin C$, we have $p(x, y) = p(y, x) = 0$ and $\mu(x)p(x, y) = 0 = \mu(y)p(y, x)$. Either way, we verify that μ satisfies detailed balance, whence μ is stationary for $p(\cdot, \cdot)$.

2. (a) $\sum_y \mu(x)q(x, y) = \sum_y \mu(y)p(y, x) = \mu(x)$ (by the stationarity of μ relative to $p(\cdot, \cdot)$). Hence $\mu q = \mu$, and μ is stationary relative to $q(\cdot, \cdot)$.

(b)(c) For both parts, the answer is No. For example, let $S = \{1, 2, 3\}$, $\mu(1) = 1$, $\mu(2) = \mu(3) = 0$, $p(1, 1) = p(2, 3) = p(3, 2) = 1$, and $q(1, 1) = q(2, 3) = q(3, 3) = 1$. If x and y are both 1, $\mu(x)q(x, y)$ and $\mu(y)p(y, x)$ are trivially equal. If one of x, y is 1 and the other is 2 or 3, then $q(x, y)$ and $p(y, x)$ both vanish, so the desired equality still holds. Lastly, if neither of x, y is 1, then $\mu(x) = \mu(y) = 0$, so equality again holds. On the other hand, note that all states are recurrent for $p(\cdot, \cdot)$ but that this is not the case for $q(\cdot, \cdot)$ (state 2 is not recurrent).

3. (a) $d(\mu p, \nu p) = \sum_x |\sum_y \mu(y)p(y, x) - \sum_y \nu(y)p(y, x)| \leq \sum_x \sum_y |\mu(y) - \nu(y)|\, p(y, x) = \sum_y |\mu(y) - \nu(y)| \sum_x p(y, x) = \sum_y |\mu(y) - \nu(y)|$.

(b) Suppose that S has n states, and that $p(x, y) \geq \epsilon > 0$ for all x, y in S. Then $p(x, y) = 1 - \sum_{y' \neq y} p(x, y') \leq 1 - \sum_{y' \neq y} \epsilon = 1 - (n - 1)\epsilon$, so if we define

$q(x,y) = (p(x,y) - \epsilon)/(1 - n\epsilon)$ we get $0 \le q(x,y) \le 1$. Moreover, $q(\cdot,\cdot)$ is a Markov kernel, since for all x, $\sum_y q(x,y) = 1$. Hence, by part (a), the ℓ^1 norm of $\mu q - \nu q$ is at most $d(\mu,\nu)$. Now $(\mu-\nu)(p) = (\mu-\nu)((1-n\epsilon)q + (\epsilon)r)$ where r is the all 1's matrix. Since $(\mu-\nu)r = 0$, we get $(\mu-\nu)(p) = (1-n\epsilon)(\mu-\nu)q$, whose ℓ^1-norm is $1-n\epsilon$ times the ℓ^1-norm of $(\mu-\nu)q$, and that norm is in turn at most $d(\mu,\nu)$. Putting this together, we get $d(\mu p, \nu p) \le (1-n\epsilon)d(\mu,\nu)$. Since this map $\mu \mapsto \mu p$ is a strict contraction on the simplex of probability distributions, there must exist a unique fixed point π satisfying $\pi = \mu p$.

The contraction property tells us that as n goes to infinity, $P_\pi(X_n = y)$ and $P_x(X_n = y)$ become arbitrarily close. Since the former is just $\pi(y)$, we have $p^n(x,y) = P_x(X_n = y) \to \pi(y)$.

4. $\{Y_n\}$ is a Markov chain. Proof: Let Z be the indicator random variable of the event $X_{\tau_0} = y$, and let $N = \tau_n$. Then by the strong Markov property, $P_\mu(Y_{n+1} = y|\mathcal{F}_n) = E_\mu(Z \circ \theta_N | \mathcal{F}_n) = E_{X_N} Z = E_{Y_n} Z$, which depends only on Y_n, so that $P_\mu(Y_{n+1} = y|\mathcal{F}_n)$ is $\sigma(Y_n)$-measurable. Since $\sigma(Y_0, Y_1, \ldots, Y_n)$ is intermediate between $\mathcal{F}_N$ and $\sigma(Y_n)$, we see that $P_\mu(Y_{n+1} = y|Y_0, Y_1, \ldots, Y_n)$ is $\sigma(Y_n)$-measurable. Hence $\{Y_n\}$ is Markovian, as claimed.

5. (a) Let s be the position in which the first $n-k$ positions are occupied by 0's and the rest by 1's. (Here we ignore the fact that the string is circular.) From any position t, one can get to s by a sequence of operations that moves the last 1 in t to the end of the string, followed by a sequence of operations that moves the second-to-last 1 in t to the second-to-last position of the string, and so on. Similarly, one can get from s to t by systematically advancing all the bits, starting from the rightmost, wrapping them all around from the end of the string to the beginning. Since all the constituent transitions have positive probability, we have $\rho_{s,t}, \rho_{t,s} > 0$. Since the state-space is finite, it follows that the Markov chain is irreducible and recurrent.

Now we construct the stationary measure. For any $s \in S$, let $N(s)$ be the number of occurrences of the substring 1,0 in the circular bit-string s. We have

$$p(s,t) = \begin{cases} 1/N(s) & \text{if } s \to t \\ 0 & \text{otherwise} \end{cases}$$

where $s \to t$ means that it is possible to convert s into t by replacing 1,0 by 0,1 in one location. Since for every t we have

$$\sum_s N(s)p(s,t) \quad = \quad \sum_s N(s)\frac{1}{N(s)}1_{s\to t}$$

$$
\begin{aligned}
&= |\{s : s \to t\}| \\
&= \#\text{ of 0,1's in } t \\
&= \#\text{ of 1,0's in } t \\
&= N(t),
\end{aligned}
$$

$N(\cdot)$ is a stationary measure for $p(\cdot,\cdot)$. To make it a probability measure, we will put $\mu(s) = N(s)/Z$ where $Z = \sum_t N(t)$. To evaluate Z, write $N(t) = N_1(t) + N_2(t) + \cdot + N_n(t)$, where $N_i(t) = 1$ if there is a 1,0 in the ith possible position and $N_i(t) = 0$ otherwise. The number of elements of S with 1,0 in any particular position is $\binom{n-2}{k-1}$, so $Z = \sum_t N(t) = n \sum_t N_1(t) = n\binom{n-2}{k-1}$.

(b) Irreducibility and recurrence for this case is much the same as in (a), since every move that is legal for the former Markov chain is also legal for the latter. However, the stationary measure is different. Here we have

$$
p(s,t) = \begin{cases} 2^{-N(s)} & \text{if } s \to t \\ 0 & \text{otherwise} \end{cases}
$$

where $N(s)$ is defined as before and $s \to t$ now means that it is possible to convert s into t by replacing some subset of 1,0's by 0,1's. Since for every t we have

$$
\begin{aligned}
\sum_s 2^{N(s)} p(s,t) &= \sum_s 2^{N(s)} 2^{-N(s)} 1_{s \to t} \\
&= |\{s : s \to t\}| \\
&= 2^{\#\text{ of 0,1's in } t} \\
&= 2^{\#\text{ of 1,0's in } t} \\
&= 2^{N(t)},
\end{aligned}
$$

$2^{N(\cdot)}$ is a stationary measure for $p(\cdot,\cdot)$. The normalization Z is trickier than I recalled; I have not been able to find a simple closed form for it. It can be expressed as the coefficient of $t^n x^k$ in the Taylor expansion of

$$
\frac{2 - t(x+1)}{1 - t(x+1) - t^2 x}.
$$