18.445 HOMEWORK 2 SOLUTIONS

Exercise 4.2. Let (a_n) be a bounded sequence. If, for a sequence of integers (n_k) satisfying

$$\lim_{k\to\infty} \frac{n_k}{n_{k+1}} = 1,$$

we have

$$\lim_{k\to\infty} \frac{a_1 + \cdots + a_{n_k}}{n_k} = a,$$

then

$$\lim_{n\to\infty} \frac{a_1 + \cdots + a_n}{n} = a.$$

Proof. For $n_k \leq n < n_{k+1}$, we can write

$$\begin{aligned} \frac{a_1 + \cdots + a_n}{n} &= \frac{a_1 + \cdots + a_{n_k}}{n} + \frac{a_{n_k+1} + \cdots + a_n}{n} \\ &= \frac{a_1 + \cdots + a_{n_k}}{n_k} \frac{n_k}{n} + \frac{a_{n_k+1} + \cdots + a_n}{n - n_k} \frac{n - n_k}{n}. \end{aligned} \tag{1}$$

As $n \to \infty$ and $k \to \infty$, by assumption

$$\frac{a_1 + \cdots + a_{n_k}}{n_k} \to a. \tag{2}$$

Since $\frac{n_k}{n_{k+1}} \leq \frac{n_k}{n} \leq 1$ and $\frac{n_k}{n_{k+1}} \to 1$, we have

$$\frac{n_k}{n} \to 1. \tag{3}$$

It follows that

$$\frac{n - n_k}{n} \to 0. \tag{4}$$

Also, (a_n) is bounded, so there exists constant $C > 0$ such that

$$\left|\frac{a_{n_k+1} + \cdots + a_n}{n - n_k}\right| \leq C. \tag{5}$$

Combining (2), (3), (4) and (5), we conclude that the formula in (1) converges to a as $n \to \infty$. □

Exercise 4.3. Let P be the transition matrix of a Markov chain with state space Ω and let μ and ν be any two distributions on Ω. Prove that

$$\|\mu P - \nu P\|_{\mathrm{TV}} \leq \|\mu - \nu\|_{\mathrm{TV}}.$$

(This in particular shows that $\|\mu P^{t+1} - \pi\|_{\mathrm{TV}} \leq \|\mu P^t - \pi\|_{\mathrm{TV}}$, that is, advancing the chain can only move it closer to stationary.)

Date: March 8, 2015. Prepared by Cheng Mao.

Proof. We have

$$\begin{aligned}
\|\mu P - \nu P\|_{\mathrm{TV}} &= \frac{1}{2}\sum_{x\in\Omega} |\mu P(x) - \nu P(x)| \\
&= \frac{1}{2}\sum_{x\in\Omega} \Big| \sum_{y\in\Omega} (\mu(y) - \nu(y))P(y,x) \Big| \\
&\le \frac{1}{2}\sum_{x,y\in\Omega} P(y,x)|\mu(y) - \nu(y)| \\
&= \frac{1}{2}\sum_{y\in\Omega} |\mu(y) - \nu(y)| \sum_{x\in\Omega} P(y,x) \\
&= \frac{1}{2}\sum_{y\in\Omega} |\mu(y) - \nu(y)| \\
&= \|\mu - \nu\|_{\mathrm{TV}}.
\end{aligned}$$

□

Exercise 4.4. Let P be he transition matrix of a Markov chain with stationary distribution π. Prove that for any $t \ge 0$,

$$d(t+1) \le d(t),$$

where $d(t)$ is defined by (4.22).

Proof. By Exercise 4.1 (see Page 329 of the book for its proof),

$$d(t) = \sup_{\mu\in\mathcal{P}} \|\mu P^t - \pi\|_{\mathrm{TV}}$$

where $\mathcal{P}$ is the set of probability distributions on Ω. By the remark in the statement of Exercise 4.3,

$$\|\mu P^{t+1} - \pi\|_{\mathrm{TV}} \le \|\mu P^t - \pi\|_{\mathrm{TV}}.$$

Therefore, we have

$$d(t+1) \le d(t).$$

□

Exercise 5.1. A mild generalization of Theorem 5.2 can be used to give an alternative proof of the Convergence Theorem.

(a). Show that when (X_t, Y_t) is a coupling satisfying (5.2) for which $X_0 \sim \mu$ and $Y_0 \sim \nu$, then

$$\|\mu P^t - \nu P^t\|_{\mathrm{TV}} \le \mathbb{P}[\tau_{\text{couple}} > t]. \tag{6}$$

Proof. Note that (X_t, Y_t) is a coupling of μP^t and νP^t. By Proposition 4.7 and (5.2),

$$\|\mu P^t - \nu P^t\|_{\mathrm{TV}} \le \mathbb{P}_{x,y}[X_t \ne Y_t] = \mathbb{P}_{x,y}[\tau_{\text{couple}} > t].$$

□

(b). If in (a) we take $\nu = \pi$, where π is the stationary distribution, then (by definition) $\pi P^t = \pi$, and (6) bounds the difference between μP^t and π. The only thing left to check is that there exists a coupling guaranteed to coalesce, that is, for which $\mathbb{P}[\tau_{\text{couple}} < \infty] = 1$. Show that if the chains (X_t) and (Y_t) are taken to be independent of one another, then they are assured to eventually meet.

Proof. Since P is aperiodic and irreducible, by Proposition 1.7, there is an integer r such that $P^r(x,y) > 0$ for all $x, y \in \Omega$. We can find $\varepsilon > 0$ such that $\varepsilon < P^r(x,y)$ for all $x, y \in \Omega$. Hence for a fixed $z \in \Omega$, wherever (X_t) and (Y_t) start from, they meet at z after r steps with probability at least ε^2 as they are independent. If they are not at z after r steps (which has probability at most $1 - \varepsilon^2$), then they meet at z after another r steps with probability at least ε^2. Hence they have not met at z after $2r$ steps with probability at most $(1-\varepsilon^2)^2$. Inductively, we see that (X_t) and (Y_t) have not met at z after nr steps with probability at most $(1-\varepsilon^2)^n$. It follows that $\mathbb{P}[\tau_{\text{couple}} > nr] \le (1-\varepsilon^2)^n$ which goes to 0 as $n \to \infty$. Thus $\mathbb{P}[\tau_{\text{couple}} < \infty] = 1$. □

Exercise 5.3. Show that if $X_1, X_2, \dots$ are independent and each have mean μ and if τ is a $\mathbb{Z}^+$-valued random variable independent of all the X_i's, then

$$\mathbb{E}[\sum_{i=1}^{\tau} X_i] = \mu\mathbb{E}[\tau].$$

Proof. Since τ is independent of (X_i),

$$\begin{aligned}
\mathbb{E}[\sum_{i=1}^{\tau} X_i] &= \sum_{n=1}^{\infty} \mathbb{P}[\tau = n]\mathbb{E}[\sum_{i=1}^{n} X_i | \tau = n] \\
&= \sum_{n=1}^{\infty} \mathbb{P}[\tau = n] \sum_{i=1}^{n} \mathbb{E}[X_i] \\
&= \sum_{n=1}^{\infty} \mathbb{P}[\tau = n] n\mu \\
&= \mu\mathbb{E}[\tau].
\end{aligned}$$

□

Exercise 6.2. Consider the top-to-random shuffle. Show that the time until the card initially one card from the bottom rises to the top, plus one more move, is a strong stationary time, and find its expectation.

Proof. Let this time be denoted by τ. We consider the top-to-random shuffle chain (X_t) as a random walk on $\mathcal{S}_n$. Let (Z_t) be an i.i.d. sequence each having the uniform distribution on the locations to insert the top card. Let $f(X_{t-1}, Z_t)$ be the function defined by inserting the top card of X_{t-1} at the the position determined by Z_t. Hence X_0 and $X_t = f(X_{t-1}, Z_t)$ define the chain inductively.

Note that $\tau = t$ if and only if there exists a subsequence $Z_{t_1}, \dots, Z_{t_{n-2}}$ where $t_1 < \cdots < t_{n-2} = t - 1$ such that Z_{t_i} chooses one of the bottom $i+1$ locations to insert the top card. Hence $\mathbb{1}_{\{\tau = t\}}$ is a function of $(Z_1, \dots, Z_t)$, so τ is a stopping time for (Z_t) . That is, τ is a randomized stopping time for (X_t).

Next, denote by $\mathcal{C}$ the card initially one card from the bottom. We show inductively that at a time t the $k!$ possible orderings of the k cards below $\mathcal{C}$ are equally likely. At the beginning, there is only the bottom card below $\mathcal{C}$. When we have k cards below $\mathcal{C}$ and insert a top card below $\mathcal{C}$, since the insertion is uniformly random, the possible orderings of the $k+1$ cards below $\mathcal{C}$ after insertion are equally likely. Therefore, when $\mathcal{C}$ is at the top, the possible orderings of the remaining $n-1$ cards are uniformly distributed. After we make one more move, the order of all n cards is uniform over all possible arrangements. That is, X_τ has the stationary distribution π. In particular, the above process shows that the distribution of X_τ is independent of τ. Hence we conclude that τ is a strong stationary time.

Finally, we compute the expectation of τ. For $1 \leq i \leq n-2$, when $\mathbb{C}$ is i cards from the bottom, then the probability that the top card is inserted below $\mathbb{C}$ is $\frac{i+1}{n}$. Hence if τ_i denotes the time it takes for $\mathbb{C}$ to move from i cards from the bottom to $i+1$ cards from the bottom, then $\mathbb{E}[\tau_i] = \frac{n}{i+1}$. It is easily seen that $\tau = \tau_1 + \cdots + \tau_{n-2} + 1$, so

$$\mathbb{E}[\tau] = \mathbb{E}[1 + \sum_{i=1}^{n-2} \tau_i] = 1 + \sum_{i=1}^{n-2} \frac{n}{i+1} = n\sum_{i=1}^{n-1} \frac{1}{i+1}.$$

□

Exercise 6.6. (Wald's Identity). Let (Y_t) be a sequence of independent and identically distributed random variables such that $\mathbb{E}[|Y_t|] < \infty$.

(a). Show that if τ is a random time so that the event $\{\tau \geq t\}$ is independent of Y_t and $\mathbb{E}[\tau] < \infty$, then

$$\mathbb{E}[\sum_{t=1}^{\tau} Y_t] = \mathbb{E}[\tau]\mathbb{E}[Y_1]. \tag{7}$$

Hint: Write $\sum_{t=1}^{\tau} Y_t = \sum_{t=1}^{\infty} Y_t \mathbb{1}_{\{\tau \geq t\}}$. First consider the case where $Y_t \geq 0$.

Proof. Using the monotone convergence theorem and that $\{\tau \geq t\}$ is independent of Y_t, we see that

$$\mathbb{E}[\sum_{t=1}^{\tau}|Y_t|] = \sum_{t=1}^{\infty}\mathbb{E}[|Y_t|\mathbb{1}_{\{\tau\geq t\}}] = \mathbb{E}[|Y_1|]\sum_{t=1}^{\infty}\mathbb{P}[\tau \geq t] = \mathbb{E}[|Y_1|]\mathbb{E}[\tau] < \infty.$$

Therefore, we can then apply the dominated convergence theorem to get that

$$\mathbb{E}[\sum_{t=1}^{\tau}Y_t] = \sum_{t=1}^{\infty}\mathbb{E}[Y_t\mathbb{1}_{\{\tau\geq t\}}] = \mathbb{E}[Y_1]\sum_{t=1}^{\infty}\mathbb{P}[\tau \geq t] = \mathbb{E}[Y_1]\mathbb{E}[\tau].$$

□

(b). Let τ be a stopping time for the sequence (Y_t). Show that $\{\tau \geq t\}$ is independent of Y_t, so (7) holds provided that $\mathbb{E}[\tau] < \infty$.

Proof. Since τ is a stopping time, $\mathbb{1}_{\{\tau\geq t\}} = \mathbb{1}_{\{\tau\leq t-1\}^c}$ is a function of $Y_0, \dots, Y_{t-1}$. Since Y_t is independent of $Y_0, \dots, Y_{t-1}$, we conclude that $\{\tau \geq t\}$ is independent of Y_t. □

Exercise 7.1. Let $\mathbf{X}_t = (X_t^1, \dots, X_t^n)$ be the position of the lazy random walker on the hypercube $\{0,1\}^n$, started at $\mathbf{X}_0 = \mathbf{1} = (1, \dots, 1)$. Show that the covariance between X_t^i and X_t^j is negative. Conclude that if $W(\mathbf{X}_t) = \sum_{i=1}^n X_t^i$, then $\mathrm{Var}(W(\mathbf{X}_t)) \leq n/4$.

Hint: It may be easier to consider the variables $Y_t^i = 2X_t^i - 1$.

Proof. Let $Y_t^i = 2X_t^i - 1$. Then $\mathrm{Cov}(Y_t^i, Y_t^j) = 4\,\mathrm{Cov}(X_t^i, X_t^j)$, so it suffices to show that $\mathrm{Cov}(Y_t^i, Y_t^j) < 0$ for $i \neq j$ and $t > 0$. If the ith coordinate is chosen in the first t steps, then the conditional expectation of Y_t^i is 0. Hence

$$\mathbb{E}[Y_t^i] = (1 - \frac{1}{n})^t \quad \text{and} \quad \mathbb{E}[Y_t^iY_t^j] = (1 - \frac{2}{n})^t.$$

It follows that for $t > 0$,

$$\mathrm{Cov}(Y_t^i, Y_t^j) = \mathbb{E}[Y_t^iY_t^j] - \mathbb{E}[Y_t^i]\mathbb{E}[Y_t^j] = (1 - \frac{2}{n})^t - (1 - \frac{1}{n})^{2t} < 0.$$

On the other hand,

$$4\,\mathrm{Var}(X_t^i) = \mathrm{Var}(Y_t^i) = \mathbb{E}[(Y_t^i)^2] - \mathbb{E}[Y_t^i]^2 = 1 - (1 - \frac{1}{n})^{2t} \leq 1.$$

Therefore,

$$\mathrm{Var}(W(\mathbf{X}_t)) = \mathrm{Var}(\sum_{i=1}^n X_t^i) = \sum_{i=1}^n \mathrm{Var}(X_t^i) + \sum_{i\neq j}\mathrm{Cov}(X_t^i, X_t^j) \leq \frac{n}{4}.$$

□

Exercise 7.2. Show that $Q(S, S^c) = Q(S^c, S)$ for any $S \subset \Omega$. (This is easy in the reversible case, but holds generally.)

Proof. We have

$$\begin{aligned}
Q(S,S^c) &= \sum_{x\in S}\sum_{y\in S^c}\pi(x)P(x,y) \\
&= \sum_{y\in S^c}\Big(\sum_{x\in\Omega}\pi(x)P(x,y) - \sum_{x\in S^c}\pi(x)P(x,y)\Big) \\
&= \sum_{y\in S^c}\sum_{x\in\Omega}\pi(x)P(x,y) - \sum_{x\in S^c}\pi(x)\sum_{y\in S^c}P(x,y) \\
&= \sum_{y\in S^c}\pi(y) - \sum_{x\in S^c}\pi(x)\Big(1-\sum_{y\in S}P(x,y)\Big) \\
&= \sum_{y\in S^c}\pi(y) - \sum_{x\in S^c}\pi(x) + \sum_{x\in S^c}\sum_{y\in S}\pi(x)P(x,y) \\
&= \sum_{x\in S^c}\sum_{y\in S}\pi(x)P(x,y) \\
&= Q(S^c,S).
\end{aligned}$$

□