AI2613 Stochastic Processes Homework 2

Xiangyuan Xue (521030910387)

1 Optimal Coupling

First, we give an explicit construction of ω. For simplicity, let $p(x, y)$ denote the union probability of X and Y, namely

$$p(x,y) \stackrel{\Delta}{=\!=} \mathbf{Pr}_{(X,Y)\sim\omega}\left[X = x \wedge Y = y\right]$$

For any $a \in \Omega$, let

$$p(a,a) = \min\left\{\mu(a), \nu(a)\right\}$$

For any $a, b \in \Omega$ such that $a \neq b$, let

$$p(a,b) = \varphi(a,b) \stackrel{\Delta}{=\!=} \frac{\left[\mu(a) - p(a,a)\right]\left[\nu(b) - p(b,b)\right]}{1 - \sum\limits_{c\in\Omega} p(c,c)}$$

Such $p(x, y)$ forms a distribution ω.

Second, we prove that such ω is a legal coupling. Note that for any $a \in \Omega$, either $p(a, a) = \mu(a)$ or $p(a, a) = \nu(a)$ holds, thus

$$\varphi(a,a) = 0$$

Then for any $a \in \Omega$, it holds that

$$\begin{aligned}
\sum_{b\in\Omega} p(a,b) &= p(a,a) + \sum_{b\neq a} p(a,b) \\
&= p(a,a) + \frac{\mu(a) - p(a,a)}{1 - \sum\limits_{c\in\Omega} p(c,c)} \sum_{b\neq a}\left[\nu(b) - p(b,b)\right] \\
&= p(a,a) + \frac{\mu(a) - p(a,a)}{1 - \sum\limits_{c\in\Omega} p(c,c)} \left\{1 - \sum_{c\in\Omega} p(c,c) - \left[\nu(a) - p(a,a)\right]\right\} \\
&= p(a,a) + \mu(a) - p(a,a) - \phi(a,a) \\
&= \mu(a)
\end{aligned}$$

Similarly for any $b \in \Omega$, it holds that

$$\sum_{a\in\Omega} p(a,b) = \nu(b)$$

Hence, such ω is a coupling of μ and ν.

Finally, we prove that ω is an optimal coupling. By equivalent definition of total variation distance, we have

$$\begin{aligned}
D_{\mathrm{TV}}(\mu,\nu) &= \max_{A\subseteq\Omega} |\mu(A)-\nu(A)| \\
&= \sum_{a\in\Omega} \left[\mu(a) - \min\{\mu(a),\nu(a)\}\right] \\
&= \sum_{a\in\Omega} \mu(a) - \sum_{a\in\Omega} p(a,a) \\
&= 1 - \mathbf{Pr}_{(X,Y)\sim\omega}\left[X = Y\right] \\
&= \mathbf{Pr}_{(X,Y)\sim\omega}\left[X \neq Y\right]
\end{aligned}$$

Therefore, we claim that there exists an optimal coupling ω of μ and ν.

2 Total Variation Distance is Non-Increasing

Let $X_t \sim \mu_t$ and $Y_t \sim \pi$ be two random variables. By coupling lemma, there exists a coupling ω_t of μ and π such that

$$\mathbf{Pr}_{(X_t,Y_t)\sim\omega_t}\left[X_t \neq Y_t\right] = D_{\mathrm{TV}}(\mu_t,\pi) = \Delta(t)$$

Then we construct ω_{t+1} as follow

- If $X_t = Y_t$ for some $t \geq 0$, then let $X_{t'} = Y_{t'}$ move synchronously.
- If $X_t \neq Y_t$ for any $t \geq 0$, then let $X_{t'}$ and $Y_{t'}$ move independently.

By coupling lemma, it holds that

$$\mathbf{Pr}_{(X_{t+1},Y_{t+1})\sim\omega_{t+1}}\left[X_{t+1} \neq Y_{t+1}\right] \geq D_{\mathrm{TV}}(\mu_{t+1},\pi) = \Delta(t+1)$$

According to the construction of ω_{t+1}, we have

$$\begin{aligned}
&\mathbf{Pr}_{(X_{t+1},Y_{t+1})\sim\omega_{t+1}}\left[X_{t+1} \neq Y_{t+1}\right] \\
=&1 - \mathbf{Pr}_{(X_{t+1},Y_{t+1})\sim\omega_{t+1}}\left[X_{t+1} = Y_{t+1}\right] \\
=&1 - \mathbf{Pr}_{(X_t,Y_t)\sim\omega_t}\left[X_t = Y_t\right] - \sum_{x\neq y} \mathbf{Pr}_{(X_t,Y_t)\sim\omega_t}\left[X_t = x \wedge Y_t = y\right] \sum_{z\in\Omega} P(x,z)P(y,z) \\
\leq&1 - \mathbf{Pr}_{(X_t,Y_t)\sim\omega_t}\left[X_t = Y_t\right] \\
=&\mathbf{Pr}_{(X_t,Y_t)\sim\omega_t}\left[X_t \neq Y_t\right]
\end{aligned}$$

Hence

$$\Delta(t+1) \leq \mathbf{Pr}_{(X_{t+1},Y_{t+1})\sim\omega_{t+1}}\left[X_{t+1} \neq Y_{t+1}\right] \leq \mathbf{Pr}_{(X_t,Y_t)\sim\omega_t}\left[X_t \neq Y_t\right] = \Delta(t)$$

Therefore, $\Delta(t+1) \leq \Delta(t)$ holds for every $t \geq 0$, namely $\Delta(t)$ is non-increasing.

3 Tossing Coin

For the pattern HH, we construct a finite Markov chain with 3 states (S, H, HH). The transition matrix is specified as

$$\boldsymbol{P} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix}$$

Let $\mathbf{E}\left(t_Q \middle| P\right)$ denote the expected time to reach state Q from state P, we have

$$\begin{aligned} \mathbf{E}\left(t_{\mathrm{HH}} \middle| \mathrm{S}\right) &= P(\mathrm{S}, \mathrm{S}) \cdot \left[1 + \mathbf{E}\left(t_{\mathrm{HH}} \middle| \mathrm{S}\right)\right] + P(\mathrm{S}, \mathrm{H}) \cdot \left[1 + \mathbf{E}\left(t_{\mathrm{HH}} \middle| \mathrm{H}\right)\right] \\ &= \frac{1}{2}\left[1 + \mathbf{E}\left(t_{\mathrm{HH}} \middle| \mathrm{S}\right)\right] + \frac{1}{2}\left[1 + \mathbf{E}\left(t_{\mathrm{HH}} \middle| \mathrm{H}\right)\right] \\ &= 1 + \frac{1}{2}\mathbf{E}\left(t_{\mathrm{HH}} \middle| \mathrm{S}\right) + \frac{1}{2}\mathbf{E}\left(t_{\mathrm{HH}} \middle| \mathrm{H}\right) \end{aligned}$$

together with

$$\begin{aligned} \mathbf{E}\left(t_{\mathrm{HH}} \middle| \mathrm{H}\right) &= P(\mathrm{H}, \mathrm{HH}) \cdot \left[1 + \mathbf{E}\left(t_{\mathrm{HH}} \middle| \mathrm{HH}\right)\right] + P(\mathrm{H}, \mathrm{S}) \cdot \left[1 + \mathbf{E}\left(t_{\mathrm{HH}} \middle| \mathrm{S}\right)\right] \\ &= \frac{1}{2} + \frac{1}{2}\left[1 + \mathbf{E}\left(t_{\mathrm{HH}} \middle| \mathrm{S}\right)\right] \\ &= 1 + \frac{1}{2}\mathbf{E}\left(t_{\mathrm{HH}} \middle| \mathrm{S}\right) \end{aligned}$$

which yields $\mathbf{E}\left(t_{\mathrm{HH}} \middle| \mathrm{S}\right) = 6$ and $\mathbf{E}\left(t_{\mathrm{HH}} \middle| \mathrm{H}\right) = 4$. Hence $\mathbf{E}(T_1) = \mathbf{E}\left(t_{\mathrm{HH}} \middle| \mathrm{S}\right) = 6$.

For the pattern HT, we construct a finite Markov chain with 3 states (S, H, HT). The transition matrix is specified as

$$\boldsymbol{P} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix}$$

Similarly, we have

$$\begin{aligned} \mathbf{E}\left(t_{\mathrm{HT}} \middle| \mathrm{S}\right) &= P(\mathrm{S}, \mathrm{S}) \cdot \left[1 + \mathbf{E}\left(t_{\mathrm{HT}} \middle| \mathrm{S}\right)\right] + P(\mathrm{S}, \mathrm{H}) \cdot \left[1 + \mathbf{E}\left(t_{\mathrm{HT}} \middle| \mathrm{H}\right)\right] \\ &= \frac{1}{2}\left[1 + \mathbf{E}\left(t_{\mathrm{HT}} \middle| \mathrm{S}\right)\right] + \frac{1}{2}\left[1 + \mathbf{E}\left(t_{\mathrm{HT}} \middle| \mathrm{H}\right)\right] \\ &= 1 + \frac{1}{2}\mathbf{E}\left(t_{\mathrm{HT}} \middle| \mathrm{S}\right) + \frac{1}{2}\mathbf{E}\left(t_{\mathrm{HT}} \middle| \mathrm{H}\right) \end{aligned}$$

together with

$$\begin{aligned} \mathbf{E}\left(t_{\mathrm{HT}} \middle| \mathrm{H}\right) &= P(\mathrm{H}, \mathrm{HT}) \cdot \left[1 + \mathbf{E}\left(t_{\mathrm{HT}} \middle| \mathrm{HT}\right)\right] + P(\mathrm{H}, \mathrm{H}) \cdot \left[1 + \mathbf{E}\left(t_{\mathrm{HT}} \middle| \mathrm{H}\right)\right] \\ &= \frac{1}{2} + \frac{1}{2}\left[1 + \mathbf{E}\left(t_{\mathrm{HT}} \middle| \mathrm{H}\right)\right] \\ &= 1 + \frac{1}{2}\mathbf{E}\left(t_{\mathrm{HT}} \middle| \mathrm{H}\right) \end{aligned}$$

which yields $\mathbf{E}\left(t_{\mathrm{HT}} \middle| \mathrm{S}\right) = 4$ and $\mathbf{E}\left(t_{\mathrm{HT}} \middle| \mathrm{H}\right) = 2$. Hence $\mathbf{E}(T_2) = \mathbf{E}\left(t_{\mathrm{HT}} \middle| \mathrm{S}\right) = 4$. Since $\mathbf{E}(T_1) > \mathbf{E}(T_2)$, we can conclude that the pattern HT is expected to occur faster than the pattern HH.

To generalize the result, we can draw longer Markov chains.

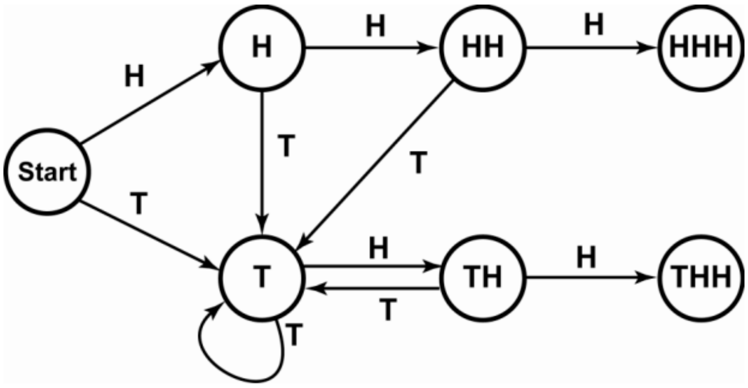

Figure 1: Markov chain

We can see that although the patterns HHH and THH have the same length, the previous states of THH contribute more degrees, which makes THH more likely to occur. In other words, states like T and TH are relatively stable and not easy to change. Similarly, we can specify the transition matrix $\boldsymbol{P}$ and list the equations $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$. The solution will give the expectation of waiting time. However, the closed-form solution is beyond my discussion.[1]

4 Path Coupling

1. According to the random process, Z_0' is random sampled from the Markov chain conditioned on Z_0, so $X_{t+1}=z_0$ is simply obtained by running the Markov chain from X_t for one step.

   For an arbitrary $i \in [k-1]$, since $\left(Z_i', Z_{i+1}'\right)$ is sampled from the promised coupling of (Z_i, Z_{i+1}), Z_{i+1}' can be considered as obtained by running the Markov chain from Z_{i+1} for one step, which is independent from Z_i'. Note that Z_0' is actually sampled from the Markov chain. By induction on i, $Z_1', Z_2', \dots, Z_k'$ are considered as sampled from the Markov chain. Hence, $Y_{t+1}=Z_k'$ is no different from being obtained by running the Markov chain from $Y_t=Z_k$ for one step, which is independent from X_{t+1}.

   Therefore, $(X_{t+1}, Y_{t+1})=(z_0, z_k)$ is a legal coupling of (X_t, Y_t).

2. Note that $\left(Z_i', Z_{i+1}'\right)$ is sampled from the coupling of (Z_i, Z_{i+1}), it holds that

$$\mathbf{E}\left[d\left(Z_i', Z_{i+1}'\right) \,\middle|\, (Z_i, Z_{i+1})\right] \leq (1-\alpha)\cdot \mathbf{E}\left[d\left(Z_i, Z_{i+1}\right)\right]$$

   Another inequality holds that

$$d\left(Z_0, Z_k\right) \leq \sum_{i=0}^{k-1} d\left(Z_i, Z_{i+1}\right)$$

Hence

$$\begin{aligned}
\mathbf{E}\left[d\left(z_0, z_k\right)\right] &= \mathbf{E}\left[d\left(Z_0', Z_k'\right)\right] \\
&\leq \mathbf{E}\left[\sum_{i=0}^{k-1} d\left(Z_i', Z_{i+1}'\right)\right] \\
&= \sum_{i=0}^{k-1} \mathbf{E}\left[d\left(Z_i', Z_{i+1}'\right) \,\middle|\, \left(Z_i, Z_{i+1}\right)\right] \\
&\leq \sum_{i=0}^{k-1} (1-\alpha) \cdot d\left(Z_i, Z_{i+1}\right) \\
&= (1-\alpha) \cdot k
\end{aligned}$$

Therefore, we can conclude that $\mathbf{E}\left[d\left(z_0, z_k\right)\right] \leq (1-\alpha) \cdot k$.

3. Let $X \sim \mu_0$ and $Y \sim \pi$ be two colorings of G where X and Y only differ at vertex v. Let $c_X(v)$ and $c_Y(v)$ denote the color of vertex v in X and Y respectively. Let $N(v)$ denote the set of neighbours of vertex v in G. We construct a coupling ω of (X, Y) as follow

   - Select a uniformly random vertex u and a uniformly random color c_X'.
   - Modify $c_X(u)$ into c_X'.
   - Consider the selection of c_Y' in 2 cases.
     (a) $u \notin N(v)$: simply let $c_Y' = c_X'$.
     (b) $u \in N(v)$: if $c_X' = c_X(v)$, let $c_Y' = c_Y(v)$; if $c_X' = c_Y(v)$, let $c_Y' = c_X(v)$; otherwise, let $c_Y' = c_X'$.
   - Modify $C_Y(u)$ into c_Y'.

   Note that X' and Y' are not necessarily proper colorings. Despite of that, the uniform distribution π only contains proper colorings, so improper colorings will finally become proper a few steps later. In addition, X' and Y' can be considered as obtained by running the Markov chain for one step independently, so such ω is a legal coupling.

   Consider case 3a of the coupling ω. Since $c_X(u)$ and $c_Y(u)$ change simultaneously, $d(X, Y)$ is non-increasing. Only when $u = v$ and the modification is valid, $d(X, Y)$ decreases by 1, which indicates that

$$\mathbf{Pr}_{(X,Y)\sim\omega}\left[d\left(X', Y'\right) = d\left(X, Y\right) - 1\right] \geq \frac{1}{n} \cdot \frac{q-\Delta}{q}$$

   Consider case 3b of the coupling ω. It is likely that $d(X, Y)$ will remain the same. Only when $c_X' = c_Y(v)$ and $c_Y' = c_X(v)$, $d(X, Y)$ increases by 1, which indicates that

$$\mathbf{Pr}_{(X,Y)\sim\omega}\left[d\left(X', Y'\right) = d\left(X, Y\right) + 1\right] \leq \frac{\Delta}{n} \cdot \frac{1}{q}$$

Since $q > 2\Delta$, the inequality holds that

$$\begin{aligned}\mathbf{E}\left[d\left(X', Y'\right) \,\middle|\, (X, Y)\right] &\le \mathbf{E}\left[d\left(X, Y\right)\right]\left(1 - \frac{1}{n} \cdot \frac{q-\Delta}{q} + \frac{\Delta}{n} \cdot \frac{1}{q}\right) \\ &= \mathbf{E}\left[d\left(X, Y\right)\right]\left(1 - \frac{q-2\Delta}{qn}\right) \\ &\le \mathbf{E}\left[d\left(X, Y\right)\right]\left(1 - \frac{1}{qn}\right)\end{aligned}$$

Suppose $X_0 \sim \mu_0$ and $Y_0 \sim \pi$, we have $d(X_0, Y_0) \le n$, which can easily be reduced to $d(X_k, Y_k) = 1$ by path coupling technique. Hence

$$\mathbf{E}\left[d\left(X_t, Y_t\right)\right] \le \left(1 - \frac{1}{qn}\right) \cdot n \cdot \left(1 - \frac{1}{qn}\right)^{t-n} < n \cdot \left(1 - \frac{1}{qn}\right)^{t-n}$$

The mixing time is bounded by

$$\tau(\varepsilon) \le n + qn \log \frac{n}{\varepsilon}$$

Therefore, the Markov chain is guaranteed to mix within $O(qn \log n)$ steps.[2]

References

[1] Nickerson R S. Penney Ante: Counterintuitive probabilities in coin tossing[J]. The UMAP Journal, 2007, 28(4): 503-532.

[2] Shayan Oveis Gharan. Graph coloring using path coupling. Washington University.