Homework 6

Benjamin Noland

Problems from the book

Exercise 4.1

By definition,

$$d(t) = \max_{x \in \mathcal{X}} \|P^t(x, \cdot) - \pi\|_{\text{TV}}.$$

Say the maximum is attained at $x' \in \mathcal{X}$, so that

$$d(t) = \|P^t(x', \cdot) - \pi\|_{\text{TV}}.$$

Then since $P^t(x', \cdot)$ is a probability distribution on $\mathcal{X}$, we see that

$$d(t) \leq \sup_{\mu} \|\mu P^t - \pi\|_{\text{TV}}.$$

Now for the opposite inequality. For any distribution μ on $\mathcal{X}$, we have the following:

$$\begin{aligned}
\|\mu P^t - \pi\|_{\text{TV}} &= \frac{1}{2} \sum_{z \in \mathcal{X}} \left|(\mu P)^t)(z) - \pi(z)\right| \\
&= \frac{1}{2} \sum_{z \in \mathcal{X}} \left| \sum_{w \in \mathcal{X}} \mu(w) P^t(w, z) - \sum_{w \in \mathcal{X}} \mu(w) \pi(z) \right| \\
&\leq \frac{1}{2} \sum_{z \in \mathcal{X}} \sum_{w \in \mathcal{X}} \left|\mu(w) P^t(w, z) - \mu(w)\pi(z)\right| \\
&= \frac{1}{2} \sum_{z \in \mathcal{X}} \sum_{w \in \mathcal{X}} \mu(w) \left|P^t(w, z) - \pi(z)\right| \\
&= \sum_{w \in \mathcal{X}} \mu(w) \left[\frac{1}{2} \sum_{z \in \mathcal{X}} \left|P^t(w, z) - \pi(z)\right| \right] \\
&= \sum_{w \in \mathcal{X}} \mu(w) \|P^t(w, \cdot) - \pi\|_{\text{TV}} \\
&\leq \sum_{w \in \mathcal{X}} \mu(w) \max_{x \in \mathcal{X}} \|P^t(x, \cdot) - \pi\|_{\text{TV}} = \max_{x \in \mathcal{X}} \|P^t(x, \cdot) - \pi\|_{\text{TV}} = d(t).
\end{aligned}$$

It then follows that

$$\sup_{\mu} \|\mu P^t - \pi\|_{\mathrm{TV}} \leq d(t),$$

and therefore

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

Now for the second part of the problem. By definition,

$$\bar{d}(t) = \max_{x,y\in\mathcal{X}} \|P^t(x,\cdot) - P^t(y,\cdot)\|_{\mathrm{TV}}.$$

Say the maximum is attained at $x', y' \in \mathcal{X}$, so that

$$\bar{d}(t) = \|P^t(x',\cdot) - P^t(y',\cdot)\|_{\mathrm{TV}}.$$

Then since $P^t(x',\cdot)$ and $P^t(y',\cdot)$ are probability distributions on $\mathcal{X}$, we see that

$$\bar{d}(t) \leq \sup_{\mu,\nu} \|\mu P^t - \nu P^t\|_{\mathrm{TV}}.$$

To establish the opposite inequality, first note that for any probability distributions μ and ν on $\mathcal{X}$, we have the following:

$$\begin{aligned}
\|\mu P^t - \nu P^t\|_{\mathrm{TV}} &= \frac{1}{2}\sum_{z\in\mathcal{X}} \left|(\mu P^t)(z) - (\nu P^t)(z)\right| \\
&= \frac{1}{2}\sum_{z\in\mathcal{X}} \left|\sum_{w\in\mathcal{X}} \nu(w)(\mu P^t)(z) - \sum_{w\in\mathcal{X}} \nu(w) P^t(w,z)\right| \\
&\leq \frac{1}{2}\sum_{z\in\mathcal{X}}\sum_{w\in\mathcal{X}} \left|\nu(w)(\mu P^t)(z) - \nu(w)P^t(w,z)\right| \\
&= \frac{1}{2}\sum_{z\in\mathcal{X}}\sum_{w\in\mathcal{X}} \nu(w)\left|(\mu P^t)(z) - P^t(w,z)\right| \\
&= \sum_{w\in\mathcal{X}} \nu(w)\left[\frac{1}{2}\sum_{z\in\mathcal{X}} \left|(\mu P^t)(z) - P^t(w,z)\right|\right] \\
&= \sum_{w\in\mathcal{X}} \nu(w)\|\mu P^t - P^t(w,\cdot)\|_{\mathrm{TV}} \\
&\leq \sum_{w\in\mathcal{X}} \nu(w)\max_{y\in\mathcal{X}} \|\mu P^t - P^t(y,\cdot)\|_{\mathrm{TV}} = \max_{y\in\mathcal{X}} \|\mu P^t - P^t(y,\cdot)\|_{\mathrm{TV}}.
\end{aligned}$$

In particular, if for any $x \in \mathcal{X}$ we take $\mu = \delta_x$ in the above, then $\mu P^t = \delta_x P^t = P^t(x,\cdot)$, and we get

$$\|P^t(x,\cdot) - \nu P^t\|_{\mathrm{TV}} \leq \max_{y\in\mathcal{X}} \|P^t(x,\cdot) - P^t(y,\cdot)\|_{\mathrm{TV}}.$$

Therefore,

$$\|\mu P^t - \nu P^t\|_{\text{TV}} \le \max_{x \in \mathcal{X}} \|P^t(x,\cdot) - \nu P^t\|_{\text{TV}} \le \max_{x,y \in \mathcal{X}} \|P^t(x,\cdot) - P^t(y,\cdot)\|_{\text{TV}} = \bar{d}(t).$$

Hence,

$$\sup_{\mu,\nu} \|\mu P^t - \nu P^t\|_{\text{TV}} \le \bar{d}(t),$$

so that

$$\bar{d}(t) = \sup_{\mu,\nu} \|\mu P^t - \nu P^t\|_{\text{TV}}.$$

Exercise 4.2

Let μ and ν be probability distributions on $\mathcal{X}$. Then we have the following:

$$\begin{aligned}
\|\mu P - \nu P\|_{\text{TV}} &= \frac{1}{2} \sum_{x \in \mathcal{X}} |(\mu P)(x) - (\nu P)(x)| \\
&= \frac{1}{2} \sum_{x \in \mathcal{X}} \left| \sum_{y \in \mathcal{X}} \mu(y) P(y,x) - \sum_{y \in \mathcal{X}} \nu(y) P(y,x) \right| \\
&= \frac{1}{2} \sum_{x \in \mathcal{X}} \left| \sum_{y \in \mathcal{X}} (\mu(y) - \nu(y)) P(y,x) \right| \\
&\le \frac{1}{2} \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{X}} |(\mu(y) - \nu(y)) P(y,x)| \\
&= \frac{1}{2} \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{X}} |\mu(y) - \nu(y)| \, P(y,x) \\
&= \frac{1}{2} \sum_{y \in \mathcal{X}} |\mu(y) - \nu(y)| \sum_{x \in \mathcal{X}} P(y,x) \\
&= \frac{1}{2} \sum_{y \in \mathcal{X}} |\mu(y) - \nu(y)| = \|\mu - \nu\|_{\text{TV}}.
\end{aligned}$$

In particular, if π is a stationary distribution for the chain, then

$$\|\mu P^{t+1} - \pi\|_{\text{TV}} = \|(\mu P^t) P - \pi P\|_{\text{TV}} \le \|\mu P^t - \pi\|_{\text{TV}}.$$

Therefore, for any $t \ge 0$,

$$d(t+1) = \sup_{\mu} \|\mu P^{t+1} - \pi\|_{\text{TV}} \le \sup_{\mu} \|\mu P^t - \pi\|_{\text{TV}} = d(t),$$

and

$$\begin{aligned}
\bar{d}(t+1) = \sup_{\mu,\nu} \|\mu P^{t+1} - \nu P^{t+1}\|_{\text{TV}} &= \sup_{\mu,\nu} \|(\mu P^t) P - (\nu P^t) P\|_{\text{TV}} \\
&\le \sup_{\mu,\nu} \|(\mu P^t) - (\nu P^t)\|_{\text{TV}} = \bar{d}(t).
\end{aligned}$$

Thus each of d and $\bar{d}$ are non-increasing in t.

Exercise 4.3

Let $s, t \geq 0$. Let $x \in \mathcal{X}$, and let π be a stationary distribution for the chain. Then by the proof of Proposition 4.7, there exists an optimal coupling (X, Y) for the distributions $P^t(x, \cdot)$ and π, that is, $\|P^t(x, \cdot) - \pi\|_{\text{TV}} = \mathrm{P}\{X \neq Y\}$, and $X \sim P^t(x, \cdot)$ and $Y \sim \pi$. Thus, for any $w \in \mathcal{X}$,

$$P^{s+t}(x, w) = \sum_{z \in \mathcal{X}} P^t(x, z) P^s(z, w) = \sum_{z \in \mathcal{X}} \mathrm{P}\{X = z\} P^s(z, w) = \mathrm{E}[P^s(X, w)],$$

and, since π is stationary,

$$\pi(w) = (\pi P^s)(w) = \sum_{z \in \mathcal{X}} \pi(z) P^s(z, w) = \sum_{z \in \mathcal{X}} \mathrm{P}\{Y = z\} P^s(z, w) = \mathrm{E}[P^s(Y, w)].$$

Thus, for any $A \subseteq \mathcal{X}$,

$$\begin{aligned}
P^{s+t}(x, A) - \pi(A) &= \mathrm{E}[P^s(X, A) - P^s(Y, A)] \\
&= \mathrm{E}[I\{X \neq Y\}(P^s(X, A) - P^s(Y, A))] \\
&\leq \mathrm{E}[I\{X \neq Y\} \bar{d}(s)] \\
&= \bar{d}(s)\, \mathrm{P}\{X \neq Y\} \\
&= \bar{d}(s) \|P^t(x, \cdot) - \pi\|_{\text{TV}} \\
&\leq \bar{d}(s) d(t).
\end{aligned}$$

Therefore, in particular we have

$$\|P^{s+t}(x, \cdot) - \pi\|_{\text{TV}} = \sum_{\substack{w \in \mathcal{X} \\ P^{s+t}(x, w) \geq \pi(w)}} [P^{s+t}(x, w) - \pi(w)] \leq \bar{d}(s) d(t),$$

from which it follows that

$$d(s + t) = \max_{x \in \mathcal{X}} \|P^{s+t}(x, \cdot) - \pi\|_{\text{TV}} \leq \bar{d}(s) d(t).$$

For the final part of the problem, I will proceed by induction on k. When $k = 2$, the statement reduces to

$$t_{\text{mix}}(2^{-2}) = t_{\text{mix}}(1/4) = t_{\text{mix}} = (2 - 1) t_{\text{mix}},$$

and so $t_{\text{mix}}(2^{-2}) \leq (2 - 1) t_{\text{mix}}$ holds trivially. Now let $k > 2$, and assume that $t_{\text{mix}}(2^{-k}) \leq (k - 1) t_{\text{mix}}$. Then since d is non-increasing, $d((k - 1) t_{\text{mix}}) \leq d(t_{\text{mix}}(2^{-k}))$. Thus, using the first result of this problem, we get

$$\begin{aligned}
d(kt_{\mathrm{mix}}) &= d((k-1)t_{\mathrm{mix}} + t_{\mathrm{mix}}) \\
&\leq \bar{d}(t_{\mathrm{mix}}) d((k-1)t_{\mathrm{mix}}) \\
&\leq \bar{d}(t_{\mathrm{mix}}) d(t_{\mathrm{mix}}(2^{-k})) \\
&\leq 2d(t_{\mathrm{mix}}) d(t_{\mathrm{mix}}(2^{-k})) \\
&\leq \frac{1}{2} d(t_{\mathrm{mix}}(2^{-k})),
\end{aligned}$$

where the inequality $\bar{d}(t_{\mathrm{mix}}) \leq 2d(t_{\mathrm{mix}})$ is due to Lemma 4.10, and in the last step I used the fact that $t_{\mathrm{mix}} = 1/4$ by definition. Finally, note that since $d(t_{\mathrm{mix}}(2^{-k})) \leq 2^{-k}$ by definition,

$$\frac{1}{2} d(t_{\mathrm{mix}}(2^{-k}) \leq 2^{-(k+1)},$$

so that

$$\frac{1}{2} d(t_{\mathrm{mix}}(2^{-k}) \leq d(t_{\mathrm{mix}}(2^{-(k+1)}).$$

Therefore $d(kt_{\mathrm{mix}}) \leq d(t_{\mathrm{mix}}(2^{-(k+1)}))$, and hence since d is non-increasing, $t_{\mathrm{mix}}(2^{-(k+1)}) \leq kt_{\mathrm{mix}}$. Thus the result holds by induction.

Coupling

Exercise 1

Let $\mu \sim \mathrm{Poisson}(b)$ and $\nu \sim \mathrm{Poisson}(a)$ denote the two Poisson distributions in question. To construct a coupling of μ and ν, first let $X \sim \mathrm{Poisson}(a)$ and $Y \sim \mathrm{Poisson}(b-a)$ be independent. Then since X and Y are independent, $X+Y \sim \mathrm{Poisson}(b)$, and thus $(X+Y, X)$ is a coupling of μ and ν, where $X+Y \sim \mu$ and $Y \sim \nu$. We have

$$\mathrm{P}\{X \neq X+Y\} = 1 - \mathrm{P}\{X = X+Y\} = 1 - \mathrm{P}\{Y = 0\} = 1 - \exp[-(b-a)].$$

Thus by Proposition 4.7, it follows that

$$\|\mu - \nu\|_{\mathrm{TV}} \leq \mathrm{P}\{X \neq X+Y\} = 1 - \exp[-(b-a)] \leq b - a,$$

where the final inequality follows from the fact that $\exp(x) \geq 1 + x$ for any $x \in \mathbb{R}$.

Exercise 2

Let $\mu \sim \mathrm{binomial}(n, 1 - e^{-\lambda/n})$ and $\nu \sim \mathrm{Poisson}(\lambda)$ denote the two distributions in question. To construct a coupling of μ and ν, define random variables $P_1, \ldots, P_n \overset{\mathrm{IID}}{\sim} \mathrm{Poisson}(\lambda/n)$, and $B_i = \min\{1, P_i\}$ for every $1 \leq i \leq n$. Then since the P_i's are independent, we have

$$Y = \sum_{i=1}^{n} P_i \sim \mathrm{Poisson}(\lambda).$$

In addition, for any $1 \leq i \leq n$,

$$\mathrm{P}\{B_i = 0\} = \mathrm{P}\{P_i = 0\} = e^{-\lambda/n}.$$

Thus, since the P_i's are independent, it follows that the $B_i's$ are independent, so that $B_1, \ldots, B_n \overset{\text{IID}}{\sim} \text{Bernoulli}(1 - e^{-\lambda/n})$. Hence,

$$X = \sum_{i=1}^{n} B_i \sim \text{binomial}(n, 1 - e^{-\lambda/n}).$$

Thus (X, Y) is a coupling of μ and ν, where $X \sim \mu$ and $Y \sim \nu$. Thus by Proposition 4.7 it follows that

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

We have the following:

$$\begin{aligned}
\mathrm{P}\{X = Y\} &= \sum_{k=0}^{\infty} \mathrm{P}\{X = Y | X = k\}\, \mathrm{P}\{X = k\} \\
&= \sum_{k=0}^{\infty} \mathrm{P}\{Y = k\}\, \mathrm{P}\{X = k\} \\
&= \sum_{k=0}^{n} \frac{\lambda^k e^{-\lambda}}{k!} \binom{n}{k} (1 - e^{-\lambda/n})^k (e^{-\lambda/n})^{n-k}.
\end{aligned}$$

Note that if we define

$$g(k) = \frac{\lambda^k e^{-\lambda}}{k!} \quad \text{for every } 0 \leq k \leq n,$$

then we can write

$$\mathrm{P}\{X = Y\} = \sum_{k=0}^{n} g(k) \binom{n}{k} (1 - e^{-\lambda/n})^k (e^{-\lambda/n})^{n-k} = \mathrm{E}[g(X)].$$

Moreover, since $g(X)$ is a non-negative random variable,

$$\mathrm{E}[g(X)] = \int_0^{\infty} \mathrm{P}\{g(X) \geq x\}\, dx = \sum_{k=0}^{n} \mathrm{P}\{g(X) \geq g(k)\}.$$

(I derived a bound from this, but then realized at the last minute that it was erroneous).

Coupling of two binomials

See the file `binomial_coupling.R` for the code, and the file `binomial_coupling_plots.pdf` for the output (both have been uploaded to Sakai).