Some exercises

1 Coupon collector's problem. A company issues n types of coupons. A collector desires a complete set. We suppose each coupon she acquires is chosen uniformly and independently from the set of n possible types, and denote by τ the random number of coupons collected when the collection first contains one of each type. Show that

$$\mathbb{E}[\tau] = n\sum_{k=1}^{n}\frac{1}{k}.$$

2 The total variation distance between two probability distributions μ and ν on a discrete space Ω is

$$\|\mu - \nu\|_{\mathrm{TV}} = \max_{A\subset\Omega} |\mu(A) - \nu(A)|.$$

Show that

$$\|\mu - \nu\|_{\mathrm{TV}} = \frac{1}{2}\sum_{x\in\Omega} |\mu(x) - \nu(x)|.$$

3 Consider an irreducible and aperiodic Markov chain with transition matrix P and invariant distribution π, and denote by $\mathbb{P}^t(x,A)$ the probability that the chain started at x is in the set A at time t. We want to control

$$d(t) := \max_{x\in\Omega} \|\mathbb{P}^t(x,\cdot) - \pi\|_{\mathrm{TV}}.$$

It is often easier to bound $\|\mathbb{P}^t(x,\cdot) - \mathbb{P}^t(y,\cdot)\|_{\mathrm{TV}}$, uniformly over all pairs of states (x,y). Define

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

Show that

$$d(t) \leq \bar{d}(t) \leq 2d(t).$$

4 A coupling of two probability measures μ and ν is a pair of random variables (X,Y) defined on a single probability space such that the marginal distribution of X is μ and the marginal distribution of Y is ν. It turns out (elementary but not easy) that

$$\|\mu - \nu\|_{\mathrm{TV}} = \inf\left\{\mathbb{P}(X \neq Y) : (X,Y) \text{ is a coupling of } \mu \text{ and } \nu\right\}.$$

A coupling of Markov chains with transition matrix P is a process $(X_t, Y_t)_{t=0}^{\infty}$ with the property that both (X_t) and (Y_t) are Markov chains with transition matrix P. A coupling can be modified so that the chains stay together after the first time they meet: if $\tau = \inf\{t \geq 0 : X_t = Y_t\}$ then $X_s = Y_s$ for $s \geq \tau$. Show that

$$\|\mathbb{P}^t(x,\cdot) - \mathbb{P}^t(y,\cdot)\|_{\mathrm{TV}} \leq \mathbb{P}_{(x,y)}(\tau > t),$$

where $\mathbb{P}_{(x,y)}$ denotes the probability on the space where both (X_t) and (Y_t) are defined.