Name: Yuxuan Sun
Date: 2026-3-20

Lecture Notes — A Modern Introduction to Probability

1 Coupon collector's problem

The coupon collector problem deals with a game where, during each turn, the contestant is given one of n types of coupons, uniformly at random. When the contestant collects one of each type of coupon, they win the game. The question we are interested in is: "How many turns does it take us to collect all n types of coupon?"

More formally, let $\{U_i\}_{i\geq 1}$ be i.i.d. random variables, each of which is uniformly distributed on $\{1, 2, \ldots, n\}$. Define:

$$\tau := \min \{t : \{1, 2, \ldots, n\} \subset \{U_1, U_2, \ldots, U_t\}\}.$$

We are interested in the distribution of τ. To this end, notice that we can write τ as the sum of n independent random variables. Namely, let

$$\tau_k := \min \{t : |\{U_1, U_2, \ldots, U_t\}| = k\}.$$

Then, we can write:

$$\tau = \tau_1 + (\tau_2 - \tau_1) + \cdots + (\tau_n - \tau_{n-1})$$

Observe that the differences $\{\tau_k - \tau_{k-1}\}$ are independent geometric random variables with success probability $\frac{n-k+1}{n}$. Thus, by linearity of expectation:

$$\mathbb{E}[\tau] = 1 + \sum_{k=2}^{n} \mathbb{E}[\tau_k - \tau_{k-1}] = \sum_{k=1}^{n} \frac{n}{n-k+1} = nH_n$$

where $H_n = \sum\limits_{k=1}^{n} \frac{1}{k}$ is the n-th harmonic number. Comparing H_n to the integral $\int_1^n \frac{1}{x} dx$, we have the estimate:

$$\log(n+1) = \log n + O(n^{-1}) \leq H_n \leq \log n + 1.$$

The following proposition provides a right tail bound for τ:

Proposition 1.1.

$$\mathbb{P}[\tau > n \log n + cn] \leq e^{-c}.$$

Proof. Let A_i be the event that $U_j \neq i$ for all j such that $1 \leq j \leq n \log n + cn$. Using a union bound, we obtain:

$$\mathbb{P}[\tau > n \log n + cn] = \mathbb{P}\left[\bigcup_{i=1}^{n} A_i\right] \leq \sum_{i=1}^{n} \mathbb{P}[A_i] = n \left(1 - \frac{1}{n}\right)^{n \log n + cn} \leq n e^{-\log n - c} \leq e^{-c}.$$

Here we are using the fact that $\{(1 - \frac{1}{n})^n\}_{n\geq 1}$ is an increasing sequence of real numbers that converges to e^{-1} as $n \to \infty$. □

2 Lazy random walk

In this section, we use our analysis of the coupon collector problem from Section 1 to analyze the mixing time of lazy random walk on the hypercube $\{0,1\}^n$ and the n-cycle.

2.1 Lazy random walk on the n-dimensional hypercube: mixing time upper bound

A lazy random walk on the hypercube $\{0,1\}^n$ is a $\{0,1\}^n$-valued stochastic process, whose dynamics are as follows:

1. We start from some point $X_0 = (X_0(1), X_0(2), \dots, X_0(n)) \in \{0,1\}^n$.

2. At each time $t \geq 1$, if X_{t-1} was the position of our random walk at time $t-1$, we sample an index I_t uniformly at random from $\{1, 2, 3, ..., n\}$ as well as a Bernoulli random variable $J_t \sim \text{Bernoulli}\left(\frac{1}{2}\right)$, where all of the I_t's and J_t's are mutually independent.

3. To get X_t, take X_{t-1} and replace the I_t-th entry of X_{t-1} with J_t.

Equivalently, lazy simple random walk on $\{0,1\}^n$ is just the Markov chain on $\{0,1\}^n$ with transition probabilities:

$$P(x,y) = \begin{cases} \frac{1}{2}, & \text{if } x = y, \\ \frac{1}{2n}, & \text{if } x \text{ and } y \text{ differ at exactly one coordinate}, \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

Given an irreducible, aperiodic Markov chain with finite state space $\mathcal{X}$, and unique stationary distribution π, the mixing time of this Markov chain is given by:

$$t_{mix}(\varepsilon) = \inf\{t \geq 0 : d(t) \leq \varepsilon\}, \tag{2}$$

where:

$$d(t) = \max_{x \in \mathcal{X}} \|P^t(x, \cdot) - \pi\|_{TV}. \tag{3}$$

Additionally, we know that:

$$d(t) \leq \overline{d}(t) \leq 2d(t), \tag{4}$$

where:

$$\overline{d}(t) = \max_{x,y \in \mathcal{X}} \|P^t(x, \cdot) - P^t(y, \cdot)\|_{TV}. \tag{5}$$

The total variation distance between two probability distributions is a measure of how well they can be coupled. Thus, using the estimate in Equation 5, we can get an upper bound for the mixing time of lazy simple random walks on the hypercube by constructing a coupling $(\widetilde{X}_t, \widetilde{Y}_t)$ of two lazy random walks with different starting points. Suppose $X_0 = (X_0(1), X_0(2), ..., X_0(n))$ and $Y_0 = (Y_0(1), Y_0(2), ..., Y_0(n)$. Let $(X_t)_{n \geq 0}$, $(Y_t)_{n \geq 0}$ be two lazy random walks on $\{0,1\}^n$ started from X_0 and Y_0, and then evolved using the same I_t's and J_t's.

Observe that once the i-th coordinates $X_t(i)$ and $Y_t(i)$ match, they will remain equal forever. Consequently, $X_t = Y_t$ holds if the sequence $\{I_s\}_{s=1}^t$ has visited all elements in $\{1, 2, \ldots, n\}$. Therefore, we have:

$$d(t) \le \overline{d}(t) \le \mathbb{P}[X_t \ne Y_t] \le \mathbb{P}[\tau > t],$$

where τ is the stopping time with the same name that we saw in Section 1 in the context of the coupon collector's problem. By Proposition 1.1, we have:

$$d(n \log n + cn) \le \mathbb{P}[\tau > n \log n + cn] \le e^{-c}, \tag{6}$$

which implies that:

$$t_{\text{mix}}(\epsilon) \le n\big(\log n + \log(1/\epsilon)\big). \tag{7}$$

2.2 Lazy random walk on the n-dimensional hypercube: mixing time lower bound

In Section 2.1, we saw that because the total variation distance between $P^t(x, \cdot)$ and $P^t(y, \cdot)$ can be written as the minimum of $\mathbb{P}(X_t \ne Y_t)$ over all couplings (X_t, Y_t) of $P^t(x, \cdot)$ and $P^t(y, \cdot)$, any such coupling gives us an upper bound for the mixing time. However, we also know that:

$$d(t) = \max_{x \in \mathcal{X}} \max_{A \subseteq \mathcal{X}} |P^t(x, A) - \pi(A)|. \tag{8}$$

Because this formula expresses the mixing time as a maximum, evaluating $|P^t(x, A) - \pi(A)|$ for any $x \in \mathcal{X}$ and any $A \subseteq \mathcal{X}$ gives us a lower bound for the mixing time of our chain. This is exactly what we will do now, to get a lower bound for the mixing time of lazy simple random walk on the hypercube.

Choose the starting point $X_0 = (0, 0, \ldots, 0)$, and define:

$$Z_t := \sum_{j=1}^{n} (-1)^{X_t(j)}. \tag{9}$$

It is easy to see that the stationary distribution π of our Markov chain is just the uniform distribution on $\{0, 1\}^n$. Define:

$$Z_\pi := \sum_{i=1}^{n} (-1)^{w_i}, \qquad \text{where } (w_1, w_2, \ldots, w_n) \sim \pi. \tag{10}$$

It is easy to check that $\mathbb{E}[Z_\pi] = 0$ and $\mathrm{Var}(Z_\pi) = n$. In fact, $Z_\pi \overset{(d)}{=} S_n$, where S_n is a simple random walk on $\mathbb{Z}$. On the other hand, observe that because the J_i's are independent of the I_i's, if $U_i = j$, the probability that $(-1)^{X_i(j)}$ and $(-1)^{X_{i-1}(j)}$ agree, is the same as the probability that $(-1)^{X_i(j)}$ and $(-1)^{X_{i-1}(j)}$ are different. Hence:

$$\mathbb{E}[(-1)^{X_t(j)}] = \mathbb{P}[U_i \ne j \text{ for all } i \le t] = \left(1 - \frac{1}{n}\right)^t, \tag{11}$$

and:

$$\mathbb{E}[(-1)^{X_t(j)+X_t(l)}] = \mathbb{P}[U_i \neq j \text{ or } l \text{ for all } i \leq t] = \left(1 - \frac{2}{n}\right)^t \tag{12}$$

provided that $j \neq l$. Therefore, for any $j \neq l$,

$$\mathrm{Cov}((-1)^{X_t(j)}, (-1)^{X_t(l)}) = \left(1 - \frac{2}{n}\right)^t - \left(1 - \frac{1}{n}\right)^{2t} < 0. \tag{13}$$

Summing over j in Equation (11), we get:

$$\mathbb{E}[Z_t] = n\left(1 - \frac{1}{n}\right)^t, \tag{14}$$

and so:

$$\mathbb{E}[Z_t^2] = n + 2\sum_{j \neq l} \mathrm{Cov}((-1)^{X_t(j)}, (-1)^{X_t(l)}) < n + 0 = n. \tag{15}$$

Choose $t = \left\lfloor \frac{1}{2} n \log n - cn \right\rfloor$. Then:

$$\mathbb{E}[Z_t] \geq n\left(1 - \frac{1}{n}\right)^{n\left(\frac{\log n}{2} - c\right)} = ne^{n \log\left(1 - \frac{1}{n}\right)\left(\frac{\log n}{2} - c\right)} = ne^{(-1+O(n^{-1}))\left(\frac{\log n}{2} - c\right)} \geq Be^c\sqrt{n}, \tag{16}$$

where $B > 0$ is an absolute constant. Define:

$$A := \left\{ v \in \{0,1\}^n : \sum_{j=1}^n (-1)^{v(j)} > \frac{1}{2} Be^c \sqrt{n} \right\} \tag{17}$$

This is the set we will plug into Equation (8) to get a lower bound for the mixing time of lazy simple random walk on the hypercube. By Chebyshev's inequality:

$$\mathbb{P}\left(|Z_t - \mathbb{E}[Z_t]| > \frac{1}{2} Be^c \sqrt{n}\right) \leq \frac{4\,\mathrm{Var}(Z_t)}{B^2 e^{2c} n} \leq \frac{4\mathbb{E}[Z_t^2]}{B^2 e^{2c} n} = \frac{4}{B^2 e^{2c}} \tag{18}$$

By Equation (16), $\mathbb{E}[Z_t] \geq Be^c\sqrt{n}$. Hence:

$$A^c \subseteq \left\{ |Z_t - \mathbb{E}Z_t| > \frac{1}{2} Be^c \sqrt{n} \right\}, \tag{19}$$

and so:

$$\mathbb{P}(X_t \notin A) \leq \frac{4}{B^2 e^{2c}}. \tag{20}$$

Under the stationary distribution, Z_π is distributed like a simple random walk on $\mathbb{Z}$ run for n steps. Thus, by Hoeffding's inequality, we have that:

$$\pi(A) = \mathbb{P}\left(Z_\pi > \frac{1}{2} Be^c \sqrt{n}\right) \leq \exp\left(-\frac{B^2 e^{2c}}{8}\right). \tag{21}$$

Alternatively, by Chebyshev's inequality:

$$\pi(A) = \mathbb{P}\left(Z_\pi > \frac{1}{2}Be^c\sqrt{n}\right) \leq \frac{4}{B^2e^{2c}} \tag{22}$$

While the estimate for $\pi(A)$ in (21) is superior to the one in (22), we will use the latter estimate going forward, since it will simplify our algebra.

Putting all this together, we have that:

$$d(t) \geq [\mathbb{P}(X_t \in A) - \pi(A)] \geq \left(1 - \frac{4}{B^2e^{2c}}\right) - \frac{4}{B^2e^{2c}} \geq 1 - \frac{8}{B^2e^{2c}}. \tag{23}$$

It follows that for any $\epsilon \in (0,1)$ sufficiently small, there is a constant $c_\epsilon > 0$ such that:

$$t_{\text{mix}}(\epsilon) \geq \frac{n}{2}\log n - c_\epsilon n.$$

2.3 Lazy random walk on the n-cycle: mixing time upper bound

Lazy random walk on the n-cycle is the Markov chain with state space $\mathbb{Z}/n\mathbb{Z}$ and transition rule:

$$X_{t+1} = \begin{cases} X_t, & \text{with probability } \frac{1}{2}, \\ (X_t+1) \bmod n, & \text{with probability } \frac{1}{4}, \\ (X_t-1) \bmod n, & \text{with probability } \frac{1}{4}. \end{cases} \tag{24}$$

In other words, at each time t, we either move one step to the right (with probability $\frac{1}{4}$), one step to the left (with probability $\frac{1}{4}$), or else we stay in place (with probability $\frac{1}{2}$). Just like in Section 2.1, getting an upper bound for the mixing time of this chain corresponds to finding a good coupling of our Markov chain started from x and our Markov chain started from y for any pair of distinct points x and y.

Our coupling is as follows. Fix initial states $X_0 = x$ and $Y_0 = y$, and let $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ be independent lazy simple random walks on the n-cycle, started from x and y respectively. Define:

$$\tau := \inf\{t \geq 0 : X_t = Y_t\}, \tag{25}$$

and define the new process $(\widetilde{Y}_t)_{t\geq 0}$ as follows:

$$\widetilde{Y}_t = \begin{cases} Y_t, & \text{if } t < \tau, \\ X_t, & \text{if } t \geq \tau. \end{cases} \tag{26}$$

Clearly $(\widetilde{Y}_t)_{t\geq 0}$ is also a lazy simple random walk on the n-cycle, started at y. Informally, X_t and $\widetilde{Y}_t$ evolve like independent lazy simple random walks on the n-cycle, up until the point when these random walks collide. After the two random walks collide, they move in unison. With this coupling, it is clear that:

$$d(t) \leq \max_{x,y\in(\mathbb{Z}/n\mathbb{Z})} \mathbb{P}_{x,y}(X_t \neq \widetilde{Y}_t) = \max_{x,y\in(\mathbb{Z}/n\mathbb{Z})} \mathbb{P}_{x,y}(\tau > t). \tag{27}$$

Furthermore, observe that the difference of the two random walks:

$$\Delta_t := (X_t - \widetilde{Y}_t) \bmod n, \tag{28}$$

is itself a simple random walk on $\mathbb{Z}/n\mathbb{Z}$, with transition rule:

$$\Delta_{t+1} = \begin{cases} (\Delta_t + 2) \bmod n, & \text{with probability } \frac{1}{16}, \\ (\Delta_t + 1) \bmod n, & \text{with probability } \frac{1}{4}, \\ \Delta_t \bmod n, & \text{with probability } \frac{3}{8} \\ (\Delta_t - 1) \bmod n, & \text{with probability } \frac{1}{4}, \\ (\Delta_t - 2) \bmod n, & \text{with probability } \frac{1}{16}, \end{cases} \tag{29}$$

and τ is exactly the hitting time of 0 mod n by this random walk. Consider the expectation of this hitting time. That is, for any $0 \leq z \leq n-1$, let $\mathbb{E}_z\tau$ be the hitting time of 0 by the random walk $(\Delta_t)_{t\geq 0}$, started from z mod n. Then $\mathbb{E}_0\tau = 0$, and by the strong Markov property, we have that:

$$\mathbb{E}_z\tau = \frac{1}{16}\Big(\mathbb{E}_{(z+2)}\tau + 1\Big) + \frac{1}{16}\Big(\mathbb{E}_{(z-2)}\tau + 1\Big) + \frac{1}{4}\Big(\mathbb{E}_{(z+1)}\tau + 1\Big) + \frac{1}{4}\Big(\mathbb{E}_{(z-1)}\tau + 1\Big) + \frac{3}{8}\Big(\mathbb{E}_z\tau + 1\Big) \tag{30}$$

for $z \not\equiv 0 \bmod n$. Equivalently, if $(\xi_t = \Delta_t - \Delta_{t-1})_{t\geq 0}$ are the i.i.d. increments of our random walk $(\Delta_t)_{t\geq 0}$, then:

$$\sum_{j=-2}^{2} \mathbb{P}(\xi_1 \equiv j \bmod n)\big(\mathbb{E}_{(z+j)}\tau - \mathbb{E}_z\tau\big) = -1 \qquad \text{for all } z \not\equiv 0 \bmod n \tag{31}$$

Equation (31), along with the fact that $\mathbb{E}_0\tau = 0$, uniquely characterizes the function $z \mapsto \mathbb{E}_z\tau$. Consider the function $k \mapsto k(n-k)$ for $k = 0, 1, 2, ..., (n-1)$. Because it evaluates to 0 at 0 and satisfies Equation (31), it follows that for any k such that $0 \leq k \leq n-1$,

$$\mathbb{E}_k\tau = k(n-k) \leq \frac{n^2}{4}. \tag{32}$$

Therefore:

$$d(t) \leq \max_{x,y\in\mathbb{Z}_n} \mathbb{P}_{(x-y)}[\tau > t] \leq \frac{1}{t} \max_{x,y\in\mathbb{Z}_n} \mathbb{E}_{(x-y)}[\tau] \leq \frac{n^2}{4t}, \tag{33}$$

which gives:

$$t_{\text{mix}}\left(\frac{1}{4}\right) \leq n^2. \tag{34}$$