

[AI2613 随机过程][第三讲] 马尔可夫链基本定理，耦合合法

张驰豪

最后更新：2025 年 3 月 13 日

目录

1 耦合 (Coupling)	1
2 马尔可夫链的基本定理	4
3 马尔可夫链基本定理的证明	6
4 马尔可夫链的收敛速率与混合时间 (Mixing Time)	8
4.1 超立方体上的随机游走	8
4.2 随机至顶 (Random-to-Top) 洗牌问题	9

1 耦合 (Coupling)

今天我们将证明马尔可夫链的基本定理 (Fundamental theorem of Markov chains, FTMC)。如课堂所述，马尔可夫链可以从多种不同的角度来看待：作为一种随机过程，或是一个线性算子，亦或是一个矩阵等等。每一种视角都提供了不同的工具来证明 FTMC。在这门课中，我们将采用“耦合” (coupling) 的方法，这是一种在概率论中被广泛使用的技巧。

我们首先定义两个分布的全变差距离 (total variation distance)。

定义 1 (全变差距离). 在一个可数状态空间 Ω 上，两个分布 μ 和 ν 的全变差距离定义为

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

注意到，由于是离散（可数）空间，我们这里再次滥用了“分布”的定义。我们把分布看成在每个样本点上给一个概率值。在一般情况下这是不行的。

我们可以观察下图（概率质量函数的图像）中两个分布在同一样本空间上的关系。全变差距离等于两条曲线所围面积的一半。从图中我们可以直接看出如下命题成立。

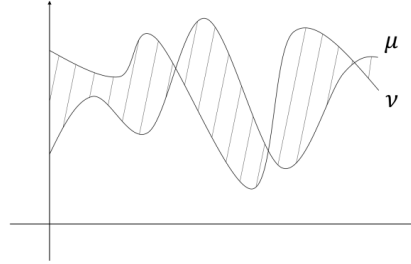


图 1: μ 和 ν 的全变差距离

命题 2. 定义 $\mu(A) = \sum_{x \in A} \mu(x)$, $\nu(A) = \sum_{x \in A} \nu(x)$, 则有

$$TV(\mu, \nu) = \max_{A \subseteq \Omega} |\mu(A) - \nu(A)|.$$

换句话说，两个分布 μ 和 ν 的全变差距离，可以看成那些所有 μ 比 ν 大的样本点上的概率差的和。

我们接着来介绍耦合。我们这儿说的耦合，是定义在两个分布上的。我们不妨先假设考虑的是离散分布，并且对应的事件集是样本点集的幂集。

定义 3 (耦合). 设 μ 和 ν 是分别定义在空间 Ω_1 和 Ω_2 上的两个分布。令 ω 为定义在空间 $\Omega_1 \times \Omega_2$ 上的分布。如果随机变量 $(X, Y) \sim \omega$ 满足 $X \sim \mu$ 且 $Y \sim \nu$, 则称 ω 为 μ 和 ν 的一个耦合。

我们有时候也会说一个新的（定义在另外的概率空间上的）二元随机变量 (\bar{X}, \bar{Y}) 随机变量是 X 和 Y 的耦合，它的意思是， (\bar{X}, \bar{Y}) 的分布是 X 的分布和 Y 的分布的耦合。在这门课里，在不会引起误解的情况下，我们语言上不太区分随机变量的耦合和分布的耦合。

耦合总是存在的，比如我们让 $\omega = \mu \otimes \nu$, 即 $(\bar{X}, \bar{Y}) \sim \omega$ 对应于 \bar{X} 和 \bar{Y} 各自独立的来自 μ 和 ν , 那显然 ω 是 μ 和 ν 的耦合。然而，在许多应用中，我们希望 \bar{X} 和 \bar{Y} 是相关的，同时保持它们各自的边缘概率正确。

示例 4. 我们现在给出耦合的一个简单例子。我们来定义分布 $\text{Ber}(\frac{1}{2})$ 和分布 $\text{Ber}(\frac{2}{3})$ 的耦合。我们把伯努利分布想象成掷硬币，其中 1 代表正面 (HEAD), 0 代表反面 (TAIL)。设 (X, Y)

事实上，对于一般（不一定离散）的分布 μ 和 ν , 我们用这个式子定义它们的全变差距离，即

$$TV(\mu, \nu) := \sup_A |\mu(A) - \nu(A)|.$$

其中 A 来自于所有的可测集。

严格的说，我们有两个概率空间 $(\Omega_1, \mathcal{F}_1, \mu)$ 和 $(\Omega_2, \mathcal{F}_2, \nu)$, 那么我们说概率空间 $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \omega)$ 上的概率测度 ω 是 μ 和 ν 的耦合，当且仅当如果随机变量 (X, Y) 的分布是 ω , 那么它们的边缘分布 X 和 Y 分别是 μ 和 ν 。

(X, Y)	TAIL	HEAD
TAIL	1/6	1/3
HEAD	1/6	1/3

表 1: 独立扔硬币产生的耦合

(X, Y)	TAIL	HEAD
TAIL	1/3	1/6
HEAD	0	1/2

表 2: 使对角线概率最大的耦合

是投掷两枚硬币的结果，我们可以耦合如下（行代表 X ，列代表 Y ）：容易看到，至少在离散场合，所谓 μ 和 ν 的耦合，就是对一个表格进行填数，使得行和与列和对应的边缘分布分别对应了 μ 与 ν 。

在上面的例子里，第一张图对应的是两个独立的硬币 $X \sim \text{Ber}(\frac{1}{2})$, $Y \sim \text{Ber}(\frac{2}{3})$ 。而在第二张图里，我们“尽量”的把联合分布的概率放在了对角线上。容易验证，这是使得 $\mathbb{P}_{(X,Y) \sim \omega} [X = Y]$ 最大的耦合 ω 。

示例 5 (随机图的比较). Erdős-Rényi 随机图模型 $\mathcal{G}(n, p)$ 是指这样一个包含 n 个顶点的随机图的分布，我们通过如下操作来得到一个随机图 $G \sim \mathcal{G}(n, p)$ ：对于每一对点 $\{i, j\}$ ，我们独立的扔一个 $\text{Ber}(p)$ 的硬币，如果是正面，就加入边 $\{i, j\}$ ，如果是反面，则不加入。我们现在想证明， $\mathbb{P}_{G \sim \mathcal{G}(n, 1/2)} [G \text{ is connected}] \leq \mathbb{P}_{H \sim \mathcal{G}(n, 2/3)} [H \text{ is connected}]$ 。这看似是一个很显然的结论，但似乎不太好严格的证明。我们现在用耦合的想法，来巧妙的说明。

我们首先看一下，如果要从 $\mathcal{G}(n, p)$ 中真的采样一个图，应该如何操作。一个很自然的做法就是依次检查每一个 $\{i, j\}$ ，对每一个点对，在 $[0, 1]$ 中均匀随机的选取一个实数 r ，然后看如果 $r \leq p$ ，则加上边 $\{i, j\}$ 否则不加。我们现在想象有两个人同时从 $\mathcal{G}(n, 1/2)$ 和 $\mathcal{G}(n, 2/3)$ 中采样图 G 和 H 。她们依次检查同一个 $\{i, j\}$ ，然后对每一个点对，在 $[0, 1]$ 中均匀随机的选取一个实数 r ，然后两个人使用同一个 r ，来决定自己的图里面是否有这一条边。显然，这样获得的 G 和 H 的分布是 $\mathcal{G}(n, 1/2)$ 和 $\mathcal{G}(n, 2/3)$ 的耦合。我们把这个耦合称作 ω 。此外，我们的构造保证了，如果 $\{i, j\}$ 是 G 的边，那么它一定也是 H 的边（因为 $r \leq 1/2 \implies r \leq 2/3$ ）。所以我们有下面的结论：

$$\mathbb{P}_{G \sim \mathcal{G}(n, 1/2)} [G \text{ is connected}] = \mathbb{P}_{(G,H) \sim \omega} [G \text{ is connected}] \leq \mathbb{P}_{(G,H) \sim \omega} [H \text{ is connected}] = \mathbb{P}_{H \sim \mathcal{G}(n, 2/3)} [H \text{ is connected}].$$

实际上，我们这里在每条边上构造的耦合就是示例 4 里第二个表格里的耦合。

实际上，我们会特别关心这种使 $X = Y$ 概率最大的耦合，因为它与两个分布的全变差距离有密切联系。下面这个结论，称为耦合引理 (coupling lemma)，会在我们未来的学习中扮演关键的角色。同样，我们不妨假设我们考虑的是有限的概率空间。这个引理对于一般的概率空间也是正确的，但其证明留作练习。

这也可能是耦合之所以被称为耦合的原因。在文献中，我们往往把 $X = Y$ 这个事件称为“ X 和 Y 耦合住了”。

引理 6 (耦合引理). 设 μ 和 ν 是定义在同一个 (离散) 样本空间 Ω 上的两个分布。则对于 μ 和 ν 的任何耦合 ω , 有

$$\mathbb{P}_{(X,Y)\sim\omega} [X \neq Y] \geq \text{TV}(\mu, \nu).$$

并且, 存在一个“最优”耦合 ω^* , 使得

$$\mathbb{P}_{(X,Y)\sim\omega^*} [X \neq Y] = \text{TV}(\mu, \nu).$$

证明. 对于有限的 Ω , 设计一个耦合等价于填充一个 $\Omega \times \Omega$ 的矩阵, 使得其边缘分布正确。显然有

$$\mathbb{P} [X = Y] = \sum_{t \in \Omega} \mathbb{P} [X = Y = t] \leq \sum_{t \in \Omega} \mu(t) \wedge \nu(t).$$

回忆我们定义过记号 $a \wedge b := \min\{a, b\}$,
 $a \vee b := \max\{a, b\}$ 。

因此,

$$\mathbb{P} [X \neq Y] \geq 1 - \sum_{t \in \Omega} \mu(t) \wedge \nu(t) = \sum_{t \in \Omega} (\mu(t) - \mu(t) \wedge \nu(t)).$$

由此,

$$\mathbb{P} [X \neq Y] = \max_{A \subseteq \Omega} \{\mu(A) - \nu(A)\} = \text{TV}(\mu, \nu).$$

为了构造“最优”的耦合 ω^* 使得上述不等式取到等号, 对于每个 $t \in \Omega$, 我们可以“贪心”的令 $\mathbb{P}_{(X,Y)\sim\omega^*} [X = Y = t] = \min\{\mu(t), \nu(t)\}$ 。当然, 我们还要说明剩下的格子能够合法的填下去。至少存在这样一种填法: 把剩下的格子里的概率定义成对于 μ 和 ν 去掉对角线概率然后进行合理归一化后各自的概率乘积。□

请务必想明白剩下的各自概率如何构造。

耦合引理为两个分布之间的距离提供了一种上界: 对于任意两个分布 μ 和 ν 及其任意耦合 ω , $\mathbb{P}_{(X,Y)\sim\omega} [X \neq Y]$ 的上界也是 $\text{TV}(\mu, \nu)$ 的上界。这是一种非常有用的方法来约束全变差距离。

2 马尔可夫链的基本定理

我们现在回忆一下在上一讲中介绍的马尔可夫链的基本定理。

定理 7 (马尔可夫链的基本定理). 若有限马尔可夫链 $P \in \mathbb{R}^{n \times n}$ 是不可约且非周期的, 则它有唯一的平稳分布 $\pi \in \mathbb{R}^n$ 。并且, 对于任意分布 $\mu \in \mathbb{R}^n$, 有

$$\lim_{t \rightarrow \infty} \mu^\top P^t = \pi^\top.$$

今天我们将给出该定理的证明。为此, 我们首先研究不可约且非周期链的转移矩阵 P 的性质。我们将使用以下引理。

引理 8. 设 c_1, c_2, \dots, c_s 是一组正整数, 满足 $\gcd(c_1, \dots, c_s) = 1$ 。则对于任何足够大的整数 b , 存在 $y_1, y_2, \dots, y_s \in \mathbb{N}$ 使得

$$c_1 y_1 + c_2 y_2 + \dots + c_s y_s = b.$$

证明. 根据 **Bézout 恒等式**, 存在 $x_1, x_2, \dots, x_s \in \mathbb{Z}$, 使得

$$c_1 x_1 + c_2 x_2 + \dots + c_s x_s = 1.$$

我们对 s 进行归纳。 $s = 1$ 的情形显然成立。假设 $s \geq 2$ 且引理对更小的 s 成立。令 $g = \gcd(c_1, \dots, c_{s-1})$ 。根据归纳假设, 我们知道

$$\frac{c_1}{g} \cdot x_1 + \frac{c_2}{g} \cdot x_2 + \dots + \frac{c_{s-1}}{g} \cdot x_{s-1} = b' \iff c_1 \cdot x_1 + c_2 \cdot x_2 + \dots + c_{s-1} x_{s-1} = g \cdot b'$$

对足够大的 b' 有非负解。因此, 我们只需证明以下方程

$$g \cdot b' + c_s \cdot x_s = b \tag{1}$$

对于任何足够大的 b , 存在非负解 (b', x_s) , 且 $b' > 0$ 。换句话说, 我们需要证明, 对于任意 $b_0 > 0$, 当 b 足够大时, 方程 (1) 有 $b' > b_0$ 的非负解。

注意到 $\gcd(g, c_s) = 1$, 我们可以找到整数 (y, x) 使得

$$g \cdot y + c_s \cdot x = 1 \iff g \cdot (by) + c_s \cdot (bx) = b.$$

注意, 对于任意 $k \in \mathbb{Z}_{\geq 0}$, 有 $g \cdot (by + kc_s) + c_s \cdot (bx - kg) = b$ 。我们需要 $by + kc_s > b_0$ 且 $bx - kg \geq 0$, 这等价于

$$\frac{bx}{g} \geq k > \frac{b_0 - by}{c_s}.$$

当 $b \geq g(b_0 + c_s)$ 时, 总能找到这样的整数 k 。 \square

接下来我们利用这一引理证明不可约且非周期链的以下性质。

命题 9. 设 $P \in \mathbb{R}^{n \times n}$ 是一个不可约且非周期的马尔可夫链。则有

$$\exists t^* : \forall i, j \in [n] : P^{t^*}(i, j) > 0.$$

证明. 不可约的性质意味着

$$\forall i, j : \exists t : P^t(i, j) > 0.$$

假设从状态 i 出发并返回 i 的环路有 s 条, 其长度为 c_1, c_2, \dots, c_s 。根据非周期性, 我们有

$$\gcd(c_1, c_2, \dots, c_s) = 1.$$

对于任意足够大的 m 和任意状态对 (i, j) , 根据上述引理和不可约性, 存在从 i 到 j 的恰好 m 步的路径。因此, 存在 $t^* > 0$, 使得对于任意状态对 (i, j) , $P^{t^*}(i, j) > 0$ 。此外, 对于任何 $t > t^*$, $P^t(i, j) > 0$ 对任意 $i, j \in \Omega$ 都成立。 \square

3 马尔可夫链基本定理的证明

我们已知 P 有一个平稳分布 π 。我们需要证明，对于任意初始分布 μ_0 ，成立

$$\lim_{t \rightarrow \infty} \text{TV}(\mu_t, \pi) = 0,$$

其中 $\mu_t^\top = \mu_0^\top P^t$ 。我们使用耦合引理，通过构造 μ_t 和 π 的一个耦合，来给出它们的全变差距离的一个上界。为了方便阐述，我们引入马尔可夫链的耦合的概念。

设 $(X_t)_{t \geq 0}, \{Y_t\}_{t \geq 0}$ 是两个马尔可夫链，它们的转移矩阵分别是 P 和 Q 。我们说一个（二元取值）的随机过程 $(\bar{X}_t, \bar{Y}_t)_{t \geq 0}$ 是 $(X_t)_{t \geq 0}$ 和 $(Y_t)_{t \geq 0}$ 的耦合，如果对于任意 $t \geq 1, \forall i, j$:

$$\begin{aligned} \mathbb{P}_{(\bar{X}_t, \bar{Y}_t)} [\bar{X}_t = j \mid \bar{X}_{t-1} = i] &= P(i \rightarrow j) \\ \mathbb{P}_{(\bar{X}_t, \bar{Y}_t)} [\bar{Y}_t = j \mid \bar{Y}_{t-1} = i] &= Q(i \rightarrow j). \end{aligned}$$

换句话说，如果我们单独盯着 \bar{X}_t 或者 \bar{Y}_t 看，它的演化行为分别和 X_t 和 Y_t 是一样的。如果我们用 μ_t 表示 X_t 的分布，用 ν_t 表示 Y_t 的分布，那么显然 (\bar{X}_t, \bar{Y}_t) 的分布是 μ_t 和 ν_t 的一个耦合。

请验证这件事!

我们接下来，通过构造马尔可夫链的耦合的方式，来证明马尔可夫链的基本定理。假设 $\{X_t\}$ 和 $\{Y_t\}$ 是从不同分布出发的两个相同的马尔可夫链，其中 $Y_0 \sim \pi$ 而 X_0 由任意分布 μ_0 生成。

在课堂上，我们直接考虑 t^* 步的转移矩阵 $Q := P^{t^*}$ ，并对 Q 构造马尔可夫链的耦合来进行证明。这样做会更加简洁一点（请大家自己尝试）。但我在讲义上保留直接对 P 做耦合的证明，其原因在于，我们最后会需要一个“耦合的耦合”的论述，我觉得很有意思。

现在我们有两列随机变量：

$$\begin{array}{ccccccccccc} \mu_0 & & \mu_1 & & & & \mu_t & & & & \\ \wr & & \wr & & & & \wr & & & & \\ X_0 & \rightarrow & X_1 & \rightarrow & X_2 & \rightarrow & \cdots & \rightarrow & X_t & \rightarrow & X_{t+1} & \rightarrow & \cdots \\ & & & & & & & & & & & & \\ Y_0 & \rightarrow & Y_1 & \rightarrow & Y_2 & \rightarrow & \cdots & \rightarrow & Y_t & \rightarrow & Y_{t+1} & \rightarrow & \cdots \\ \wr & & \wr & & & & \wr & & & & & & \\ \pi & & \pi & & & & \pi & & & & & & \end{array}$$

我们现在构造 $(X_t)_{t \geq 0}$ 和 $(Y_t)_{t \geq 0}$ 的这两个马尔可夫链的耦合 $(\bar{X}_t, \bar{Y}_t)_{t \geq 0}$:

- 初始情况，分别独立的从 μ_0 和 π 中采样: $\bar{X}_0 \sim \mu_0, \bar{Y}_0 \sim \pi$;
- 对于 $t \geq 1$ ，如果 $\bar{X}_{t-1} = \bar{Y}_{t-1}$ ，则令 $\bar{X}_t = \bar{Y}_t$ ，否则分别独立的按照 P 从 \bar{X}_{t-1} 和 \bar{Y}_{t-1} 演化。

根据耦合引理，我们有

$$\text{TV}(\mu_t, \pi) \leq \mathbb{P}_{(\bar{X}_t, \bar{Y}_t)} [\bar{X}_t \neq \bar{Y}_t].$$

令 t^* 为上面命题中提到的 t^* 。令 δ 为一个正常数, 使得对于任意状态对 (i, j) , $P^{t^*}(i, j) \geq \delta > 0$ 。我们首先断言, 对于任意初始状态 \bar{X}_0, \bar{Y}_0 , 成立

$$\mathbb{P}[\bar{X}_{t^*} \neq \bar{Y}_{t^*}] \leq 1 - \delta^2$$

如果上述断言成立, 我们可以对 \bar{X}_{2t^*} 和 \bar{Y}_{2t^*} 未相遇的概率进行如下估计:

$$\begin{aligned} \mathbb{P}[\bar{X}_{2t^*} \neq \bar{Y}_{2t^*}] &= \mathbb{P}[\bar{X}_{2t^*} \neq \bar{Y}_{2t^*} | \bar{X}_{t^*} = \bar{Y}_{t^*}] \mathbb{P}[\bar{X}_{t^*} = \bar{Y}_{t^*}] \\ &\quad + \mathbb{P}[\bar{X}_{2t^*} \neq \bar{Y}_{2t^*} | \bar{X}_{t^*} \neq \bar{Y}_{t^*}] \mathbb{P}[\bar{X}_{t^*} \neq \bar{Y}_{t^*}] \\ &\leq (1 - \delta^2) \cdot \mathbb{P}[\bar{X}_{t^*} \neq \bar{Y}_{t^*}] \\ &\leq (1 - \delta^2)^2 \end{aligned}$$

通过归纳法可以得到 $\mathbb{P}[\bar{X}_{kt^*} \neq \bar{Y}_{kt^*}] \leq (1 - \delta^2)^k$ 。根据耦合引理,

$$\text{TV}(\mu_{kt^*}, \pi) \leq \mathbb{P}[\bar{X}_{kt^*} \neq \bar{Y}_{kt^*}] \leq (1 - \delta^2)^k.$$

另一方面, 我们还有以下引理。

引理 10. 对于任意 $t \geq 0$, 有 $\text{TV}(\mu_{t+1}, \pi) \leq \text{TV}(\mu_t, \pi)$ 成立。

证明. 这里我们简要说明证明思路。我们再次使用耦合引理。根据耦合引理, 我们只需构造一个 μ_{t+1} 和 π 的耦合, 使得当从该耦合中抽取 (X_{t+1}, Y_{t+1}) 时, $\mathbb{P}[X_{t+1} \neq Y_{t+1}]$ 被 $\text{TV}(\mu_t, \pi)$ 所控制。我们可以通过以下方式构造耦合:

- 从 μ_t 和 π 的最优耦合中抽取 (X_t, Y_t) ;
- 如果 $X_t = Y_t$, 则令 X_{t+1} 从 X_t 演化, 且 $Y_{t+1} = X_{t+1}$;
- 如果 $X_t \neq Y_t$, 则令 X_{t+1} 和 Y_{t+1} 分别从 X_t 和 Y_t 独立演化。

□

结合上述所有结果, 我们可以得出结论:

$$\lim_{t \rightarrow \infty} \text{TV}(\mu_t, \pi) = 0$$

最后, 我们需要证明上述断言。直观上, 我们有 $\mathbb{P}[X_{t^*} = 1] \geq \delta$ 以及 $\mathbb{P}[Y_{t^*} = 1] \geq \delta$, 并且在两链尚未相遇时, 它们是独立的, 如果相遇了, 那自然永远相等。但如果要严格的说明, 我们使用一个“耦合的耦合”的论证。我们把具体细节留作作业。

4 马尔可夫链的收敛速率与混合时间 (Mixing Time)

我们现在开始研究马尔可夫链的收敛速率。我们从混合时间的概念开始。对于任意 $\varepsilon > 0$, 马尔可夫链 P 的 (误差不超过 ε) 的混合时间定义为最小的步数 t , 使得从任意初始分布运行马尔可夫链后, 其与平稳分布的全变差距离至多为 ε 。形式化地定义为:

$$\tau_{\text{mix}}(\varepsilon) := \min_t \sup_{\mu_0} \text{TV}(\mu_t, \pi) \leq \varepsilon.$$

在有的文献里, 人们会定义 $\tau_{\text{mix}} := \tau_{\text{mix}}(1/4)$, 并直接把 τ_{mix} 称为混合时间。

回忆刚才我们用耦合法证明马尔可夫链基本定理的时候, 我们用到了不等式

$$\text{TV}(\mu_t, \pi) \leq \mathbb{P}_{(\bar{X}_t, \bar{Y}_t)} [\bar{X}_t \neq \bar{Y}_t].$$

其中 (\bar{X}_t, \bar{Y}_t) 是一个耦合。显然, 如果我们能够定量的控制上式右边的概率, 比如说证明对于任意初始的 (\bar{X}_0, \bar{Y}_0) , $\mathbb{P}_{(\bar{X}_t, \bar{Y}_t)} [\bar{X}_t \neq \bar{Y}_t] \leq \varepsilon$, 那么根据定义, 马上就能得到 $\tau_{\text{mix}}(\varepsilon) \leq t$ 。换句话说

为什么这个不等式可以蕴含混合时间上界? 请自己严格证明。

一个好的耦合可以给出马尔可夫链的混合时间的好的上界

因此, 耦合法是我们将要介绍的证明马尔可夫链收敛速度上界的第一个方法。下面是几个例子。

4.1 超立方体上的随机游走

考虑在 n -立方体上的随机游走。状态空间为 $\Omega = \{0, 1\}^n$, 当且仅当 $\|x - y\|_1 = 1$ 时, 状态 x 和 y 之间存在一条边。容易验证, 这个马尔可夫链的唯一平稳分布是 Ω 上的均匀分布。我们从一个点 $X_0 \in \Omega$ 出发, 在每一步中, 如果当前的位置是 $X_t \in \{0, 1\}^n$:

这是个可逆马尔可夫链, 因此使用细致平衡条件即可简单验证其平稳分布为均匀分布。

这里以 $\frac{1}{2}$ 概率不动的主要目的是为了使得马尔可夫链非周期化。在每个点上都加上自环的马尔可夫链也被称为 *lazy chain*。

- 以概率 $\frac{1}{2}$ 保持不动;
- 否则, 随机选择一个 $i \in [n]$ 并翻转 $X_t(i)$ 。

现在我们用耦合分析该过程的混合时间。我们可以把它每一步的移动写成下面这个等价形式。

- 随机选择一个 $i \in [n]$ 和一个 $b \in \{0, 1\}$;
- 将 $X_t(i)$ 改为 b 。

我们构造一组马尔可夫链的耦合 (X_t, Y_t) , 其中 X_0 和 Y_0 是任意状态。构造方法是

两个随机游走 X_t 和 Y_t 在每一步选择相同的 i 和 b 。

于是，一旦某个位置 $i \in [n]$ 被选择， $X_t(i)$ 和 $Y_t(i)$ 将永远相同。因此，这个问题等价于我们研究过的奖券收集问题。我们可以计算出对于 $t \geq n \log n + cn$ ，未选择第 i 维的概率为：

$$\left(1 - \frac{1}{n}\right)^{n \log n + cn} \leq \frac{e^{-c}}{n}.$$

于是，至少存在一个维度未被选择的概率不超过 e^{-c} 。我们希望这个概率小于 ε ，因此选择 $c > \log \frac{1}{\varepsilon}$ 。于是，

$$\tau_{\text{mix}}(\varepsilon) \leq n \log \frac{n}{\varepsilon}.$$

接下来我们稍作修改，将随机游走里每一步不动的概率由 $\frac{1}{2}$ 改为 $\frac{1}{n+1}$ ，即以概率 $\frac{1}{n+1}$ 保持不动，使得这个 lazy walk 略为活跃一点。在这种情况下，我们描述另一种 X_t 和 Y_t 的耦合方法。我们总可以不失一般性的排列 X_t 和 Y_t 的坐标，使得所有取值不同的坐标排在前面。也就是说，存在一个索引 k ，使得当 $1 \leq i \leq k$ 时， $X_t(i) \neq Y_t(i)$ ，而当 $i > k$ 时， $X_t(i) = Y_t(i)$ 。我们的耦合规则如下：

- 如果 $k = 0$ ， Y_t 的行为与 X_t 相同。
- 如果 $k = 1$ ，当 X_t 翻转第一位时， Y_t 保持不动；当 X_t 保持不动时， Y_t 翻转第一位；其余时间两者行为一样。
- 如果 $k \geq 2$ ，我们区分 X 是否翻转了 $[k]$ 中的位置：
 - 如果 X 保持不动或翻转了一个 $i > k$ 的位置： Y 的行为相同；
 - 如果 X 翻转了 $1 \leq i \leq k$ ： Y 翻转 $(i \bmod k) + 1$ ，即将 $1 \mapsto 2, 2 \mapsto 3, \dots, k-1 \mapsto k, k \mapsto 1$ 。

显然，上述方法确实是一个耦合。事实上，这种耦合的行为类似于“倍速奖券收集问题”，因为在 $k \geq 2$ 的情况下，如果 $i \in [k]$ ，我们每次可以“收集”两个奖券。因此，可以相信，

$$\tau_{\text{mix}} \leq \frac{1}{2} n \log n + O(n).$$

4.2 随机至顶 (Random-to-Top) 洗牌问题

给定一副 n 张牌的牌堆，考虑以下洗牌规则：

- 随机抽取一张牌；
- 将这张牌放到牌堆的顶端。

这一洗牌规则可以视为在 $n!$ 种排列上的随机游走，且容易验证，均匀分布是其平稳分布。我们设计一个此马尔可夫链的耦合方法。即，设 X_t 和 Y_t 分别表示两副牌堆，我们通过以下方式构造 X_{t+1} 和 Y_{t+1} ：

抽取相同的牌。

注意，这里是抽取“相同的牌”，而不是位于相同位置的牌。换句话说，一旦我们在 X_t 中抽取了 $\heartsuit K$ ，我们也在 Y_t 中抽取 $\heartsuit K$ 。

显然，这是一个耦合，而且一旦某张牌（比如 $\heartsuit K$ ）被抽取过，那么 $\heartsuit K$ 在两副牌堆中的位置将始终相同。因此，如果我们问需要多少轮 T 使得 $X_T = Y_T$ ，这个问题就再次等价于奖券收集问题。我们有

$$\tau_{\text{mix}}(\epsilon) \leq n \log \frac{n}{\epsilon}.$$