概率论笔记

田博宇

2018 年 5 月 12 日

目录

1 内容概要

本节课的主要内容有：

• 马尔科夫链中的 Coupling 方法的定义

• 引入概率分布之间的距离 Bounding Total Variation Distance

• 讲解了一些利用 Coupling 方法的例子，例如在 Hyper Cube 上的随机游走

2 问题引入

给定不可约非周期转移矩阵 P，是否存在 xP = x，即求 $\lim\limits_{n\to\infty}$ 是否存在

2.1 证明存在性

择一律

给定 x^{T} 和矩阵 A，以下两个条件必满足其一：

$$\begin{cases} x^{\mathrm{T}}A = b^{\mathrm{T}} \\ \exists y, s.t. Ay \leq 0, b^{\mathrm{T}}y > 0 \end{cases} \tag{1}$$

利用择一律证明存在性

令 $a = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}^{\mathrm{T}}$ 则有：

$$\begin{aligned} x^{\mathrm{T}}P = x^{\mathrm{T}} &\Longleftrightarrow x^{\mathrm{T}}(P-I) = 0 \\ &\Longleftrightarrow x^{\mathrm{T}}\left[\begin{array}{c|c} P-I & a \end{array}\right] = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \end{aligned}$$

其中 $\left[\begin{array}{c|c} P-I & a \end{array}\right]$ 就是择一律中的 A。若这个不成立，则必成立择一律中的另一条

$$\left[\begin{array}{c|c} P-I & a \end{array}\right] \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n+1} \end{bmatrix} \leq 0$$

由上式得：

$$p_{11}y_1 + p_{12}y_2 + \cdots + p_{1n}y_n < y_1$$

不妨设 $y_1 = \min\limits_{i=1\cdots n} y_i$，矛盾推出择一律中的第一条必定成立。

则必有 $x^{\mathrm{T}}P = x^{\mathrm{T}}$

3 Coupling 方法

3.1 定义

一对随机分布 μ 和 ν 的 coupling 指的是一对随机变量 (X,Y), 定义在同一个概率空间上, 使得 X 的 marginal distribution 是 μ, Y 的 marginal distribution 是 ν

marginal disribution 边缘分布

指在概率论和统计学的多维随机变量中，只包含其中部分变量的概率分布。

3.2 Example 12.1

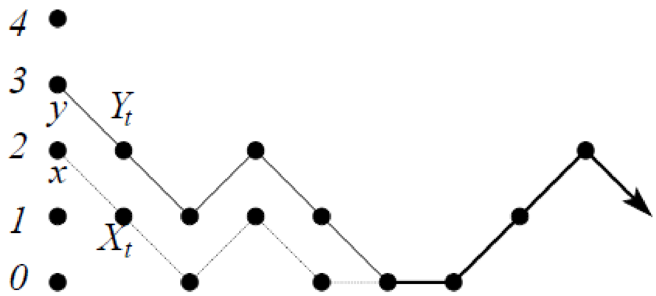
考虑如下马尔科夫链：

$$\begin{cases} P(i,i+1) = p(i,i-1) = \frac{1}{2} \\ P(0,1) = P(n,n-1) = \frac{1}{2} \\ P(0,0) = P(n,n) = \frac{1}{2} \end{cases} \tag{2}$$

对于不同的初始位置 x, y, 给予他们相同的转移 (这个过程即为耦合方法 coupling), 考虑 t 步转移概率：若 $x \leq y$, 则有 $P^t(x,n) \leq P^t(y,n)$

3.3 带转移矩阵的定义

Givena Markov chain on state space χ with transition matrix P, a *Markovian coupling* of two P-chains is a Markov chain with state space $\chi \times \chi$ which satisfies:



for all x, y, x', y',

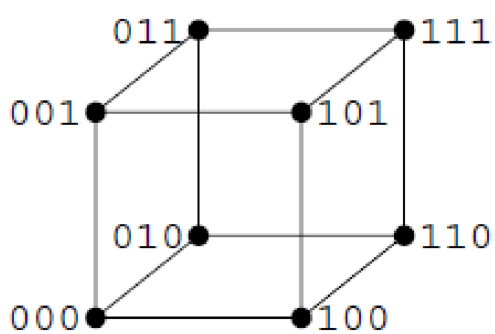
$$\begin{cases} P\{X_{t+1} = x' | X_t = x, Y_t = y\} = P(x, x') \\ P\{Y_{t+1} = x' | X_t = x, Y_t = y\} = P(y, y') \end{cases} \tag{3}$$

3.4 例子

3.4.1 hypercube 上的随机游走

hypercube

n 维的 hypercube 是一个顶点为二进制 n 元组 $\{0,1\}^n$ 的图，两个顶点之间有且仅有一维坐标不同，如图是一个三维 hypercube



lazy chain

每一步，停留在原本位置的概率为 $\frac{1}{2}$，另外有 $\frac{1}{2}$ 的概率移动到一个独立随机选取的相邻的点上。

构造方法 随机选取一维，并且用一个随机选取的 bit 更新这一位

游走过程

在 hypercube 上随机选取两个不同的起始点，每次选取他们两个的同一维，并且使用同样的 bit 更新。

τ 为第一次所有的维都被选择过的时间，

3.5 Mixing Time

首先定义 d(t):

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

定义 *mixing time*:

$$t_{mix}(\epsilon) := \min\{t : d(t) \leq \epsilon\}$$

3.6 Coupling Lemma

Let (X_t, Y_t) be a Markovian coupling based on a ground Markovian chain (Z_t) on Ω. Suppose $t : [0, 1] \to \mathbf{N}$ is a function satisfying:

$$P[X_{t(\epsilon)} \neq Y_{t(\epsilon)} | X_0 = x, Y_0 = y] \leq \epsilon, for\ all\ x,\ y\ \in \Omega\ and\ all\ \epsilon > 0$$

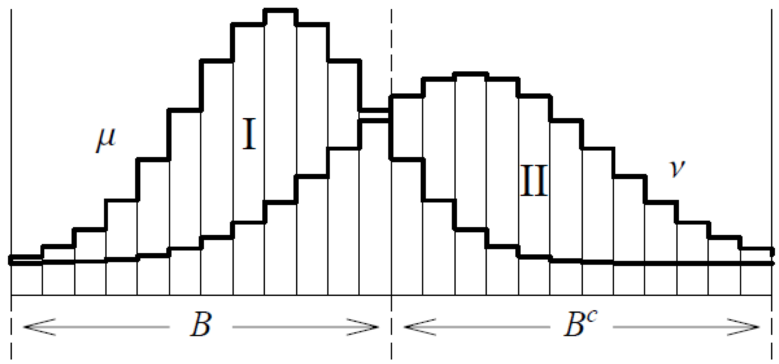
Then the mixing time $\tau(\epsilon)$ of Z_t is bounded above by $t(\epsilon)$

4 Total Variation Distance

4.1 定义

μ 和 ν 是状态空间 χ 上的两个概率分布，定义 μ 和 ν 的 Total Variation Distance 为：

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



total variation 的定义并不容易直接得出 distance，下面给出三个性质：

4.2 命题 1

μ 和 ν 是状态空间 χ 上的两个概率分布，则有：

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

证明

令 $B=\{x:\mu(x)\geq\nu(x)\}$, A 是 χ 中的任意事件,
任意 $x\in A\cap B^c$, 有 $\mu(x)-\nu(x)\leq 0$, 则有

$$\mu(A)-\nu(A)\leq\mu(A\cap B)-\nu(A\cap B)$$

而又有

$$\mu(A\cap B)-\nu(A\cap B)\leq\mu(B)-\nu(B)$$

故有

$$\mu(A)-\nu(A)\leq\mu(B)-\nu(B)$$

同理可得

$$\nu(A)-\mu(A)\leq\nu(B^c)-\mu(B^c)$$

由上图知，

$$\mu(B)-\nu(B)=\nu(B^c)-\mu(B^c)$$

同理把 A,B 互换，可得另外一边的结果，故得结论：

$$\begin{aligned}\|\mu-\nu\|_{TV} &= \frac{1}{2}[\mu(B)-\nu(B)+\nu(B^c)-\mu(B^c)] \\ &= \frac{1}{2}\sum_{x\in\chi}|\mu(x)-\nu(x)|\end{aligned}$$

4.3 命题 2

μ 和 ν 是状态空间 χ 上的两个概率分布，则有：

$$\|\mu-\nu\|_{TV} = \frac{1}{2}sup\left\{\sum_{x\in\chi}f(x)\mu(x)-\sum_{x\in\chi}f(x)\nu(x) : \max_{x\in\chi}|f(x)\leq 1\right\}$$

证明

先证右边大于等于左边:

若有 $\max_{x\in\chi}|f(x)\leq 1$，

$$\frac{1}{2}|\sum_{x\in\chi}f(x)\mu(x)-\sum_{x\in\chi}f(x)\nu(x)| \leq \frac{1}{2}\sum_{x\in\chi}|\mu(x)-\nu(x)| = \|\mu-\nu\|_{TV}$$

再证左边大于等于右边:

构造

$$f^*(x) = \begin{cases} 1 & \mu(x)\geq\nu(x) \\ -1 & \mu(x)\leq\nu(x)\end{cases}$$

则有

$$\begin{aligned}\frac{1}{2}[\sum_{x\in\chi}f(x)\mu(x)-\sum_{x\in\chi}f(x)\nu(x)] &= \frac{1}{2}\sum_{x\in\chi}f^*(x)[\mu(x)-\nu(x)] \\ &= \frac{1}{2}[\sum_{\substack{x\in\chi\\ \mu(x)\geq\nu(x)}}[\mu(x)-\nu(x)]+\sum_{\substack{x\in\chi\\ \nu(x)\geq\mu(x)}}[\nu(x)-\mu(x)]]\end{aligned}$$

原命题得证.

4.4 命题 3

μ 和 ν 是状态空间 χ 上的两个概率分布，则有：

$$\|\mu-\nu\|_{TV} = inf\{\mathbf{P}\{X\neq Y\}) : (X,Y)\ is\ a\ coupling\ of\ \mu\ and\ \nu\}$$

证明

先证右边大于等于左边：

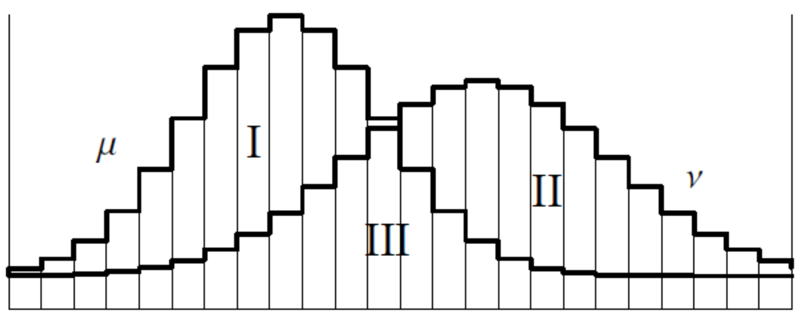
对任意 coupling(X, Y) 和任意事件 $A \in \chi$，有：

$$\begin{aligned}\mu(A)-\nu(A) &= P\{X\in A\}-P\{Y\in A\}\\ &\le P\{X\in A, Y\notin A\}\\ &\le P\{X\neq Y\}\end{aligned}$$

则有

$$\|\mu-\nu\|_{TV} \le inf\{\mathbf{P}\{X\neq Y\}) : (X,Y)\ is\ a\ coupling\ of\ \mu\ and\ \nu\}$$

再证存在 (X,Y) 使得右边等于左边：



由上图知，区域 I 和 II 的面积是 $\|\mu-\nu\|_{TV}$，则区域 III 的面积是 $1-\|\mu-\nu\|_{TV}$

令 $p=1-\|\mu-\nu\|_{TV}$, 考虑掷一枚正面向上概率为 p 的硬币

(i) 若向上，则按照如下概率分布选取 Z:

$$\gamma_{III}(x)=\frac{\mu(x)\wedge\nu(x)}{p}$$

并且令 $X=Y=Z$

(ii) 若向下，则按照如下概率分布选取 X:

$$\gamma_I(x)=\begin{cases}\frac{\mu(x)-\nu(x)}{1-p} & \mu(x)>\nu(x)\\ 0 & otherwise\end{cases}$$

按照如下概率分布选取 Y：

$$\gamma_I(x) = \begin{cases} \frac{\mu(x)-\nu(x)}{1-p} & \nu(x) > \mu(x) \\ 0 & otherwise \end{cases}$$

则有：

$$p\gamma_{III} + (1-p)\gamma_I = \mu$$

$$p\gamma_{III} + (1-p)\gamma_{II} = \mu$$

则当且仅当硬币正面朝上时，$X = Y$，则有

$$P\{X \neq Y\} = \|\mu - \nu\|_{TV}$$

原命题得证.

5 Ergodicity 定理的表述和证明

（不是课堂上的内容，是一道习题，可直接解答问题引入中的问题，同时这份证明也上传在问题解决部分）If P is irreducible and aperiodic, then there is a unique stationary distribution π such that

$$\forall x, \lim_{t\to\infty} P^t(x, \cdot) = \pi$$

PROOF:

构造 coupling distribution 如下：
• 若 $X_t \neq Y_t$, 根据 P, 独立地选取 X_{t+1} 和 Y_{t+1}
• 若 $X_t = Y_t$，根据 P, 选取 X_{t+1}，并且令 $Y_{t+1} = X_{t+1}$
由定理（可参照 Class12 的 notes）:

$$\|X^t - Y^t\|_{TV} = inf(P(X^t \neq Y^t))$$

存在 t^*, 使得 $\forall x, y$, $P^{t^*} > 0$. 故存在 $\epsilon > 0$ 满足对任意初始状态 X_0, Y_0

$$P(X^{t^*} \neq Y^{t^*} | X_0, Y_0) \leq 1 - \epsilon$$

同理得：

$$P(X^{2t^*} \neq Y^{2t^*} | X^{t^*} \neq Y^{t^*}) \leq 1 - \epsilon$$

则：

$$\begin{aligned} P(X^{2t^*} \neq Y^{2t^*}|X_0, Y_0) &= P(X^{t^*} \neq Y^{t^*} \wedge X^{2t^*} \neq Y^{2t^*}|X_0, Y_0) \\ &= P(X^{2t^*} \neq Y^{2t^*}|X^{t^*} \neq Y^{t^*})P(X^{t^*} \neq Y^{t^*}|X_0, Y_0) \\ &\leq (1-\epsilon)^2 \end{aligned}$$

则对任意正整数 k >0，有

$$P(X^{kt^*} \neq Y^{kt^*}|X_0, Y_0) \leq (1-\epsilon)^k$$

当 $k \to \inf$, $P(X^{kt^*} \neq Y^{kt^*}|X_0, Y_0) \to 0$ 则有：

$$\lim_{t\to\infty} P(X^t \neq Y^t|X_0, Y_0) \to 0$$

由 coupling lemma, 有：

$$\|P^t(x, \cdot) - P^t(y, \cdot)\|_{TV} \leq P(X^t \neq Y^t) \to 0, \ when\ t\ \to 0$$

令 $\sigma = \lim_{t\to\infty} P^t(x, \cdot)$，则有

$$\begin{aligned} \sum_x \sigma(x)P(x,y) &= \sum_x \lim_{t\to\infty} P^t(z,x)P(x,y) \ \ \forall z \\ &= \lim_{t\to\infty} P^{t+1}(z,y) = \sigma(y) \end{aligned}$$

即为 $\sigma P = \sigma$

同时，由于 $\|\lim_{t\to\infty} P^t(x, \cdot) - \lim_{t\to\infty} P^t(y, \cdot)\|_{TV} \to 0$, σ 唯一

原命题得证