数学模型 Lecture Notes

Zhennan Zhou

2021 年 6 月 8 日

PRELIMINARY DRAFT. NOT FOR WIDE CIRCULATION.
参考：教材一第9，12章。

8 概率、随机模型

It is scientific only to say what's more likely or less likely, and not to be proving all the time what's possible or impossible.

—— Richard Feynman

8.1 果壳中的概率（Probability in a nutshell）（不直接考察）

概率空间(Ω, F, P)是一个总测度为1的测度空间（即$P(\Omega) = 1$）。

- Ω是一个非空集合，称为样本空间（Sample Space），他的元素称为样本输出（Outcome）。
- F是样本空间Ω的幂集的一个非空子集 (Ω的子集的集合)，它的元素称为事件（Event），事件是样本空间的子集。
- P称为概率(测度)。$P: F \to \mathbb{R}$。每个事件都被P赋予一个0和1之间的概率值。

随机变量$X : \Omega \to E$是从样本空间到可测空间E的可测函数。这门课里面，我们只考虑$E = \mathbb{R}$。随机变量取值$S \subset E$的概率我们记为

$$\Pr\big(X \in S\big) = P\big(\{\omega \in \Omega | X(\omega) \in S\}\big).$$

离散的随机变量可以被离散的概率密度刻画:

$$\mathbf{f} = \{f_i\}; \quad f_i \geq 0, \quad , i \in \mathbb{N}; \quad \sum_{i \in \mathbb{N}} f_i = 1.$$

如果X服从离散概率密度$\mathbf{f}$，我们记为$X \sim \mathbf{f}$。这个离散变量的累积分布函数$\mathbf{F} = \{F_i\}$定义为$F_n = \sum_{i \leq n} f_i$。我们易知，$0 \leq F_i \leq 1$, 而且 $F_n = \Pr(X \leq n)$。

如果（一维）随机变量 X 是连续的，那么我们有概率密度函数 $f(x)$，满足

$$f(x) \geq 0, \quad \int_{\mathbb{R}} f(x) dx = 1.$$

而相应的累计分布函数为 $F(x) = \int_{-\infty}^{x} f(s) ds$。

关于记号做一点说明：虽然事件是样本空间的子集，但是，我们也习惯的用随机变量相对应的表示，比如事件$\{\omega \in \Omega | u < X(w) \leq v\}$，这个事件也简写为$u < X \leq v$。

条件概率$P(A|B)$是指事件A在另外一个事件B已经发生条件下的发生概率，定义为

$$P(A|B)=\frac{P(A\cap B)}{P(B)}.$$

两个事件A和B是（统计）独立的，当且仅当 $P(A\cap B)=P(A)P(B)$。易知，如果A和B是独立事件，$P(A|B)=P(A), P(B|A)=P(B)$。

一般的，根据 $P(A\cap B)=P(A)P(B|A)=P(B)P(A|B)$，我们得到贝叶斯公式

$$P(A|B)=\frac{P(A)P(B|A)}{P(B)}.$$

（离散）随机变量的期望（或称均值），是随机变量在概率分布下的平均值。我们用 $\mathbf{E}$ 表示期望，

$$\mathbf{E}X=\sum_i X_i f_i.$$

有时候，我们也用 $\bar{X}$ 表示期望。我们也可以对随机变量的函数求期望。比如，方差定义为

$$\mathrm{Var}(X)=D(X)=\mathbf{E}(X-\mathbf{E}X)^2=\mathbf{E}(X^2)-(\mathbf{E}X)^2.$$

最后，随机过程是指如下的一族的随机变量

$$\{X(t):t\in T\},\quad \text{或写成}\quad \{X_t:t\in T\}.$$

这里，T 是一个指标集，可以是连续的，也可以离散的。如果 $t\in\mathbb{R}$，它常被理解为时间，而 $X(t)$ 是某个可观测量在时间 t 时对应的随机变量。有时候，人们也会把一个随机过程写成 $\{X(t,\omega):t\in T\}$，表明它其实是 $t\in T$ 和 $\omega\in\Omega$ 的二元函数。

需要注意的是，在使用随机过程来建立数学模型的时候，我们总有两个角度：

- $X(t)$ 的随 t 变化的变化规律；
- 考虑概率密度 $\{f_i(t)\}$ 或者 $f(x,t)$ 随时间的演化规律。

8.2 初等概率模型举例：随机人口模型

我们首先看一个随机变量取值离散，而时间连续变化的例子。时刻 t 的人口数量用随机变量 $X(t)$ 表示，$X(t)$ 只取非负整数值。记 $P_n(t)$ 是 $X(t)=n$ 的概率，$n=0,1,2,\cdots$。

下面我们对出生和死亡的概率做出适当的假设，寻求 $P_n(t)$ 的变化规律，并由此得到 $X(t)$的期望和方差。

若 $X(t)=n$，对于充分小的时间Δt，我们对人口在 t 到 $t+\Delta t$ 的出生和死亡做如下的假设：

- 出生一人的概率与 Δt 成正比，记为 $b_n\Delta t$，出生两人及以上的概率是$o(\Delta t)$。且 b_n 与 n 成正比，记为 $b_n=\lambda n$。
- 死亡一人的概率与 Δt 成正比，记为 $d_n\Delta t$，出生两人及以上的概率是$o(\Delta t)$。且 d_n 与 n 成正比，记为 $d_n=\mu n$。
- 出生死亡是相互独立的随机事件。

思考：如果给出 $\Pr(X(t+\Delta t)=\ell|X(t)=n)$？

对于概率密度，我们得到，

$$P_n(t+\Delta t)=P_{n-1}(t)b_{n-1}\Delta t+P_{n+1}(t)d_{n+1}\Delta t+P_n(t)(1-b_n\Delta t-d_n\Delta t)+o(\Delta t).$$

于是，我们得到如下的微分方程

$$\frac{dP_n}{dt} = \lambda(n-1)P_{n-1} + \mu(n+1)P_{n+1} - (\lambda+\mu)nP_n. \tag{8.1}$$

如果我们假设，在初始时刻（t=0）人口为确定的数量 n_0，则 $P_n(t)$ 的初始条件为

$$P_{n_0}(0) = 1, \quad P_n(0) = 0, \quad n \neq n_0. \tag{8.2}$$

从随机过程的角度，我们得到了一个马氏链模型（见下一节），但是从概率密度函数的角度看，我们其实得到了一个无穷维的变化率模型，这里的状态变量是 $\{P_n(t)\}_{n=0}^{\infty}$。

求解这些方程非常复杂，但是如果我们只关心 $X(t)$ 的期望（以下简记 $E(t)$ ）和方差(以下简记 $D(t)$)，则我们可以由(8.1)和(8.2)直接得到。根据期望的定义，$E(t) = \sum_n nP_n(t)$。我们可以得到 $E(t)$ 满足的方程

$$\frac{dE}{dt} = \lambda\sum_{n=1}^{\infty} n(n-1)P_{n-1} + \mu\sum_{n=1}^{\infty} n(n+1)P_{n+1} - (\lambda+\mu)\sum_{n=1}^{\infty} n^2 P_n.$$

经过化简（化简中，我们不妨假设 $P_0 = P_1 = 0$ ），我们得到

$$\frac{dE}{dt} = (\lambda-\mu)E.$$

而它的初始条件为$E(0) = n_0$。所以，我们得到方程的解

$$E(t) = n_0 e^{(\lambda-\mu)t}.$$

注意，这个形式就和非随机的模型完全一致了。

对于方差 $D(t)$, 按照定义$D(t) = \sum_{n=1}^{\infty} n^2 P_n(t) - E^2(t)$. 可以推出（练习）

$$D(t) = n_0 \frac{\lambda+\mu}{\lambda-\mu} e^{(\lambda-\mu)t}[e^{(\lambda-\mu)t} - 1].$$

最后，我们做一个观察，如果我们假设没有出生，只有死亡，那么从期望上看，我们得到期望人口所满足的方程是

$$E(t) = n_0 e^{-\mu t}, \quad 即 \quad \frac{dE}{dt} = -\mu E, \quad E(0) = n_0.$$

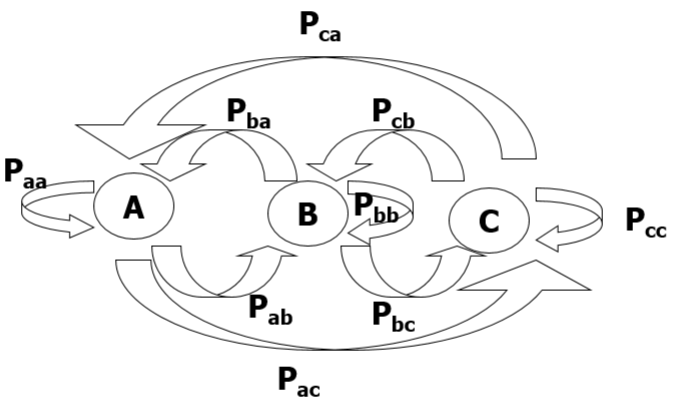
这样，我们就理解了第二章的 Decay 过程其实是一个随机过程的平均行为。

8.3 马氏链模型和数学理论

在考虑随机动力系统时，系统在每个时刻所处的状态是随机的，从这个时期到下一个时期的状态按照一定的概率进行转移，并且下个时期的状态只与这个时期的状态和转移概率有关，与以前各个时期的状态无关，这种性质称为无后效性，或马尔科夫（Markov）性。这种随机转移过程通常用马氏链模型来描述。注意，在本小节中，我们考虑的时间都是离散的。

8.3.1 引例

某大学有三个食堂A、B、C。调查显示：在食堂A就餐的人中 p_{aa} 部分仍然回到食堂A，有 p_{ab} 部分选择食堂B，p_{ac} 部分选择食堂C；在食堂B就餐的人中 p_{bb} 部分仍然回到食堂B，有 p_{ba} 部分选择食堂A，p_{bc} 部分选择食堂C；在食堂C就餐的人中 p_{cc}部分仍然回到食堂C，有 p_{ca} 部分选择食堂A，p_{cb}部分选择食堂B；如图所示。



- 令A_n为第n天在食堂A就餐的人数比例;
- 令B_n为第n天在食堂B就餐的人数比例;
- 令C_n为第n天在食堂C就餐的人数比例。

于是我们有

$$\pi_{n+1}^T := \begin{pmatrix} A_{n+1} \\ B_{n+1} \\ C_{n+1} \end{pmatrix} = \begin{pmatrix} p_{aa}A_n + p_{ba}B_n + p_{ca}C_n \\ p_{ab}A_n + p_{bb}B_n + p_{cb}C_n \\ p_{ac}A_n + p_{bc}B_n + p_{cc}C_n \end{pmatrix} = \begin{pmatrix} p_{aa} & p_{ba} & p_{ca} \\ p_{ab} & p_{bb} & p_{cb} \\ p_{ac} & p_{bc} & p_{cc} \end{pmatrix} \begin{pmatrix} A_n \\ B_n \\ C_n \end{pmatrix} =: P^T \pi_n^T = (\pi_n P)^T.$$

8.3.2 马氏链选讲

我们回忆到，对于离散的时间 $t = 0, 1, \cdots$ 的每一个 t 对应一个随机变量 $\xi_t(\omega)$，我们把 $\{\xi_0, \xi_1, \cdots, \xi_n, \cdots\}$ 这样一个随机变量的序列叫做离散时间的随机过程。

所有 $\xi_n(\omega)$ 具有公共的取值集合, 我们把此集合叫做状态空间，记为 S，其元素称为状态。

离散时间离散状态的随机过程$\{\xi_n, n \geq 0\}$，状态空间 S 为有限或者可数集合，如果满足

$$\Pr(\xi_{n+1} = j | \xi_n = i, \xi_{n-1} = i_{n-1}, \cdots, \xi_0 = i_0) = \Pr(\xi_{n+1} = j | \xi_n = i),$$

称其为一个离散时间马氏链，其中的条件概率 $\Pr(\xi_{n+1} = j | \xi_n = i)$ 称为其在时刻n处的转移概率 $p_{ij}(n)$，而 $P(n) = (p_{ij}(n))$ 称为时刻n处的转移概率矩阵。

如果马氏链的转移矩阵与出发时刻无关，即$P(n)$与n无关，则称此马氏链是时齐的。这时可将转移概率矩阵简记为 $P = (p_{ij})$。通常不特别说明, 马氏链就指时齐马氏链。

对于状态空间有限的情况（不妨假设只有 k 个状态），令 $a_j(n) = \Pr(\xi_n = j)$，则马氏链的基本方程为

$$a_i(n+1) = \sum_{j=1}^{k} a_j(n) p_{ji}, \quad i = 1, \cdots, k.$$

这里，我们要求

$$\sum_{j=1}^{k} a_j(n) = 1, \quad n = 0, 1, \cdots$$

$$p_{ij} \geq 0, \quad i, j = 1, \cdots, k. \quad \sum_{j=1}^{k} p_{ij} = 1, \quad i = 1, \cdots, k.$$

如果我们记状态概率向量

$$\mathbf{a}(n) = (a_1(n), \cdots, a_k(n)), \quad n = 0, 1, \cdots$$

（注意，我们一般认为状态概率向量是横向量）则我们有

$$\mathbf{a}(n+1)=\mathbf{a}(n)P,\quad \mathbf{a}(n)=\mathbf{a}(0)P^n.$$

状态空间 S 上的一个概率分布称为转移概率矩阵P的不变概率分布(简称不变分布)，如果

$$\pi P=\pi,\quad \sum_{j=1}^{k}\pi_j=1,\quad \pi_j\ge 0,\quad j=1,\cdots,k.$$

注意，这时候，π^T可以看作是 P^T 特征值为 1 的特征向量，即

$$P^T\pi^T=1\pi^T.$$

下面通过几个例子来理解马氏链模型的数学结构。在第一个例子中，我们考虑一个非常简单的情况，状态的个数为 2。我们考察当 $n\to\infty$ 时，状态概率向量 $\mathbf{a}(n)$ 的渐进行为。

例子 1. 考虑如下转移矩阵

$$P=\begin{pmatrix}1-a & a\\ b & 1-b\end{pmatrix},\quad 0\le a,b\le 1.$$

特征值为1和$1-(a+b)$, 相应的特征向量为

$$\begin{pmatrix}1\\1\end{pmatrix},\quad \begin{pmatrix}a\\-b\end{pmatrix}.$$

于是有相似变换

$$\begin{aligned}P&=\begin{pmatrix}1-a & a\\ b & 1-b\end{pmatrix}\\ &=\begin{pmatrix}1 & a\\ 1 & -b\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & 1-a-b\end{pmatrix}\begin{pmatrix}1 & a\\ 1 & -b\end{pmatrix}^{-1}.\end{aligned}$$

通过计算，我们容易得到

$$\begin{pmatrix}1 & a\\ 1 & -b\end{pmatrix}^{-1}=\frac{1}{a+b}\begin{pmatrix}b & a\\ 1 & -1\end{pmatrix}.$$

于是，

$$\begin{aligned}P^n&=\begin{pmatrix}1 & a\\ 1 & -b\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & (1-a-b)^n\end{pmatrix}\begin{pmatrix}1 & a\\ 1 & -b\end{pmatrix}^{-1}\\ &=\begin{pmatrix}1 & a\\ 1 & -b\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & (1-a-b)^n\end{pmatrix}\frac{1}{a+b}\begin{pmatrix}b & a\\ 1 & -1\end{pmatrix}.\end{aligned}$$

下面我们分三种情况讨论，

- Case A. $0<a+b<2, a\ne 1, b\ne 1$, 则所有元素非0,

$$\begin{aligned}P^n&\to\begin{pmatrix}1 & a\\ 1 & -b\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}\frac{1}{a+b}\begin{pmatrix}b & a\\ 1 & -1\end{pmatrix}\\ &=\frac{1}{a+b}\begin{pmatrix}b & a\\ b & a\end{pmatrix}.\end{aligned}$$

容易验证，此时，不变分布为 $\pi=[\frac{b}{a+b},\frac{a}{a+b}]$.

- Case B. $a=b=0$, P 为单位矩阵，$P^n=P$.
- Case C. $a=b=1$, 则

$$P=\begin{pmatrix}0&1\\1&0\end{pmatrix},\quad P^{2k}=\begin{pmatrix}1&0\\0&1\end{pmatrix}=I,\quad P^{2k+1}=\begin{pmatrix}0&1\\1&0\end{pmatrix}=P.$$

此时 P^n 极限不存在。但平均极限存在

$$\lim_{n\to+\infty}\frac{1}{n}\sum_{k=1}^{n}P^k=\begin{pmatrix}\frac{1}{2}&\frac{1}{2}\\\frac{1}{2}&\frac{1}{2}\end{pmatrix}.$$

当状态空间的元素比较多的时候，考查状态概率向量的渐进行为比较困难了，很多的分析方法大家会在《随机过程》等相关课程中学习。不过，对于一些复杂的情况，我们可以直接求解马氏链模型的不变分布。下面，我们考虑一个有实际背景的例子。

例子 2. Ehrenfest模型。 容器内有 $2a$ 个粒子，一张薄膜将容器分成对称的A,B 两部分。将粒子穿过薄膜时占用的时间忽略不计。用 X_0 表示初始时 A 中的粒子数，X_n 表示有 n 个粒子穿过薄膜后 A 中的粒子数。假设 $\{X_n\}$ 是马氏链，有状态空间 $I=\{0,1,\cdots,2a\}$。设转移概率

$$p_{ij}=\begin{cases}\dfrac{2a-i}{2a}, & ,0\le i\le 2a-1,\ j=i+1,\\ \dfrac{i}{2a}, & 1\le i\le 2a,\ j=i-1,\\ 0, & \text{otherwise.}\end{cases}$$

已知该马氏链的不变分布唯一存在，计算不变分布 π。

补充定义 $\pi_{-1}=\pi_{2a+1}=0$, 则方程组 $\pi P=\pi$ 可以写成

$$\pi_i=\pi_{i-1}p_{i-1,i}+\pi_{i+1}p_{i+1,i},\quad 0\le i\le 2a.$$

注意到，这其实是一族的代数方程（参考第6章差分方程的一般形式），我们可以把它写成

$$\pi_{i+1}=\frac{\pi_i-\pi_{i-1}p_{i-1,i}}{p_{i+1,i}}.$$

于是，我们可以顺次计算

$$\begin{aligned}\pi_1&=\frac{\pi_0}{p_{10}}=2a\pi_0=C_{2a}^1\pi_0.\\ \pi_2&=\frac{\pi_1-\pi_0p_{01}}{p_{21}}=\frac{(\pi_1-\pi_0)2a}{2}=(2a-1)a\pi_0=C_{2a}^2\pi_0.\\ \pi_3&=\cdots=C_{2a}^3\pi_0.\\ &\cdots\\ \pi_{2a}&=C_{2a}^{2a}\pi_0.\end{aligned}$$

利用

$$\pi_0+\pi_1+\cdots+\pi_{2a}=2^{2a}\pi_0=1,$$

得到 $\pi_0=2^{-2a}$。于是我们得到不变分布

$$\pi_i=C_{2a}^i\left(\frac{1}{2}\right)^{2a}=C_{2a}^i\left(\frac{1}{2}\right)^i\left(\frac{1}{2}\right)^{2a-i},\quad i=0,\cdots,2a.$$

这是一个二项分布，说明在不变分布下，或者时间充分长以后，各个粒子的位置是相互独立的，每个粒子在 A 的概率是1/2。

最后我们再看一个比较复杂的马氏链模型。特别是，我们在模型中引入了迁入和迁出的机制。

例子 3. 等级结构。 设一个社会从低到高分成了 k 个等级。时间以年为单位离散化，即每年只进行一次调动。等级记作 $i = 1, \cdots, k$。时间记作 $t = 0, 1, \cdots$ 我们引入如下的记号：

- $\mathbf{n}(t) = (n_1(t), \cdots, n_k(t))$，$n_i(t)$ 为第 t 年属于等级 i 的人数；$N(t) = \sum_{i=1}^k n_i(t)$ 为第 t 年的总人数。
- $\mathbf{a}(t) = (a_1(t), \cdots, a_k(t))$ 称等级结构，$a_i(t) = \frac{n_i(t)}{N(t)}$ 为第 t 年属于等级 i 的人数比例，$\sum_{i=1}^k a_i(t) = 1$。
- $Q = \{p_{ij}\}_{k\times k}$，$p_{ij}$ 为每年从等级 i 转移至等级 j（的人口在等级 i 人口中占）的比例。
- $\mathbf{w} = (w_1, \cdots, w_k)$ 为每年从等级 i 退出（在等级 i 中占）的比例。第 t 年退出的总人数为
$$W(t) = \sum_{i=1}^k w_i n_i(t) = \mathbf{n}(t)\mathbf{w}^T.$$
而且，我们可以看出
$$\sum_{i=1}^k p_{ji} + w_j = 1, \quad j = 1, \cdots, k.$$
- $\mathbf{r} = (r_1, \cdots, r_k)$ 为每年调入等级 i 的在总调入人数的比例。第 t 年调入的总人数为 $R(t)$，则第 t 年调入等级 i 的人数为 $r_i R(t)$。显然，$\sum_{i=1}^k r_i = 1$。

为了导出 $\mathbf{n}(t)$ 的变化规律，我们先注意到
$$N(t+1) = N(t) + R(t) - W(t).$$
而且
$$n_j(t+1) = \sum_{i=1}^k p_{ij} n_i(t) + r_j R(t), \quad j = 1, \cdots, k.$$

用矩阵来表示，则有
$$\mathbf{n}(t+1) = \mathbf{n}(t)Q + R(t)\mathbf{r}.$$
但是需要注意的是，Q 并非概率转移矩阵（想想为什么）。我们用 $M(t)$ 表示 t 到 t+1 年的净增长量，则
$$R(t) = W(t) + M(t) = \mathbf{n}(t)\mathbf{w}^T + M(t)$$
我们代入可得，
$$\mathbf{n}(t+1) = \mathbf{n}(t)(Q + \mathbf{w}^T\mathbf{r}) + M(t)\mathbf{r} =: \mathbf{n}(t)P + M(t)\mathbf{r}.$$
注意，P 的行和是 1，所以 P 是一个概率转移矩阵。

下面我们考虑一个特殊的形式，$M(t) = \alpha N(t)$，即每年的总人数按照百分比 α 增长，可以推出等级结构满足的方程为
$$\mathbf{a}(t+1) = (1+\alpha)^{-1}(\mathbf{a}(t)P + \alpha\mathbf{r}).$$

特别的，如果我们进一步假设每年总人数不变，那么 $\alpha = 0$，于是我们有
$$\mathbf{a}(t+1) = \mathbf{a}(t)P = \mathbf{a}(t)(Q + \mathbf{w}^T\mathbf{r}). \tag{8.3}$$

接下来，我们将调入比例 $\mathbf{r}$ 看作是一个可以控制的自由度，从优化和控制的角度进一步研究这个模型。首先，我们注意到，不同的调入比例 $\mathbf{r}$ 会对应上述马氏链模型不同的不变分布 $\mathbf{a}$。于是我们引入如下的定义。

根据这个方程，对于某个等级结构 $\mathbf{a}$，如果存在调入比例 $\mathbf{r}$，使得

$$\mathbf{a} = \mathbf{a}P = \mathbf{a}(Q + \mathbf{w}^T\mathbf{r}). \tag{8.4}$$

则称 $\mathbf{a}$ 为稳定结构。注意，这里的 $\mathbf{r}$ 需要满足

$$\sum_{i=1}^{k} r_i = 1, \quad r_i \geq 0, \quad i = 1, \cdots, k.$$

由(8.4)，我们得到

$$\mathbf{r} = \frac{\mathbf{a} - \mathbf{a}Q}{\mathbf{a}\mathbf{w}^T}.$$

得以验证 $\sum_{i=1}^{k} r_i = 1$ 是满足的，但是为了 $r_i \geq 0, \quad i = 1, \cdots, k$，我们可以稳定结构 $\mathbf{a}$ 的范围由

$$\mathbf{a} \geq \mathbf{a}Q$$

决定，这称为等级结构的稳定域。

最后我们讨论通过控制调入比例 r 对等级结构进行动态调节。设理想的等级结构是 $\mathbf{a}^*$，并假设 $\mathbf{a}^*$ 在稳定域里面。已知 $\mathbf{a}(0)$，求调入比例 $\mathbf{r}$ 使得 $\mathbf{a}(1)$ 尽可能的接近 $\mathbf{a}^*$。再求一个新的调入比例 $\mathbf{r}$ 使得 $\mathbf{a}(2)$ 尽可能的接近 $\mathbf{a}^*$，以此类推。

对于等级结构，我们首先引入任意两个等级结构 $\mathbf{a}^{(1)} = (a_1^{(1)}, \cdots, a_k^{(1)})$ 和 $\mathbf{a}^{(2)} = (a_1^{(2)}, \cdots, a_k^{(2)})$ 之间的距离

$$D(\mathbf{a}^{(1)}, \mathbf{a}^{(2)}) = \sum_{i=1}^{k} \lambda_i (a_i^{(1)} - a_i^{(2)})^2,$$

其中 $\lambda_i \geq 0$ 为加权因子。

我们可以把问题总结为 (E1)

$$\begin{aligned} &\min_{r_i} \quad D(\mathbf{a}(1), \mathbf{a}^*) \\ &\text{s.t.} \quad \mathbf{a}(1) = \mathbf{a}(0)(Q + \mathbf{w}^T\mathbf{r}), \quad \sum_{i=1}^{k} r_i = 1, \quad r_i \geq 0, \quad i = 1, \cdots, k. \end{aligned}$$

注意到，

$$\mathbf{a}^* - \mathbf{a}(1) = \mathbf{a}(0)\mathbf{w}^T \left[\frac{\mathbf{a}^* - \mathbf{a}(0)Q}{\mathbf{a}(0)\mathbf{w}^T} - \mathbf{r} \right].$$

如果记

$$\mathbf{y} = \frac{\mathbf{a}^* - \mathbf{a}(0)Q}{\mathbf{a}(0)\mathbf{w}^T}, \quad \mathbf{y} = (y_1, \cdots, y_k),$$

则 $\mathbf{a}^* - \mathbf{a}(1)$ 与 $\mathbf{y} - \mathbf{r}$ 成正比，而且容易验证 $\sum_{i=1}^{k} y_i = 1$.

于是，我们把问题转化为 (E2)

$$\begin{aligned} &\min_{r_i} \quad \sum_{i=1}^{k} \lambda_i (y_i - r_i)^2 \\ &\text{s.t.} \quad \sum_{i=1}^{k} r_i = 1, \quad r_i \geq 0, \quad i = 1, \cdots, k. \end{aligned}$$

不妨令 $\mathbf{r}^*$ 表示最优值。

我们注意到，如果 $y_i \geq 0, \quad i = 1, \cdots, k$，则（E2）的解显然为

$$r_i = y_i, \quad i = 1, \cdots, k.$$

否则，我们将$\{1,\cdots,k\}$按按照 y_i 的符号分成两部分 I_1 和 I_2，即

$$y_j > 0, \quad \forall j \in I_1; \quad y_l \le 0, \quad \forall l \in I_2.$$

于是，我们有

$$\sum_{i=1}^{k} \lambda_i (y_i - r_i)^2 = \sum_{j\in I_1} \lambda_j (y_j - r_j)^2 + \sum_{l\in I_2} \lambda_l (|y_l| + r_l)^2.$$

注意到，$\sum_{j\in I_1} y_j > 1$，易知若使和式最小，应有

$$r_j \le y_j, \quad \forall j \in I_1; \quad r_l = r_l^* = 0, \quad \forall l \in I_2.$$

最后，我们得到问题（E3）

$$\begin{aligned} \min_{r_j, j\in I_1} \quad & \sum_{j\in I_1} \lambda_j (y_j - r_j)^2 \\ \text{s.t.} \quad & \sum_{j\in I_1} r_j = 1, \quad y_j \ge r_j \ge 0, \quad j \in I_1. \end{aligned}$$

利用拉格朗日因子法，我们可以求解这个条件极值问题。我们引入因子 α，（E3）的目标函数为

$$G = \sum_{j\in I_1} \lambda_j (y_j - r_j)^2 + 2\alpha \sum_{j\in I_1} r_j.$$

利用 $\partial G/\partial r_j = 0$，可知

$$r_j = y_j - \frac{\alpha}{\lambda_j}.$$

再利用 $\sum_{j\in I_1} r_j = 1$ 可以得到，如果令

$$\alpha = \frac{\sum_{j\in I_1} y_j - 1}{\sum_{j\in I_1} \lambda_j^{-1}},$$

于是我们得到

$$r_j = y_j - \frac{\alpha}{\lambda_j}, \quad j \in I_1.$$

注意，我们在求解中并没有要求 $y_j \ge r_j \ge 0$。显然，$y_j \ge r_j$ 对所有的 j 是满足的，但是可能存在 $j' \in I_1$，使得 $r_{j'} < 0$。这时候，我们令 $r_{j'}^* = 0$，然后利用 $\sum_{j\in I_1, j\ne j'} r_j = 1$ 重新定义 α, 然后再次求解剩下的 r_j。重复以上操作，直到得到非空的 $I_1^* \subset I_1$，使得 $\forall j \in I_1^*, r_j = y_j - \frac{\alpha}{\lambda_j} \in [0, y_j]$。此时，

$$\alpha = \alpha^* = \frac{\sum_{j\in I_1^*} y_j - 1}{\sum_{j\in I_1^*} \lambda_j^{-1}},$$

最终，（E3）的解为

$$r_j^* = \begin{cases} y_j - \dfrac{\alpha^*}{\lambda_j}, & j \in I_1^*; \\ 0, & \text{otherwise.} \end{cases}$$

8.3.3 一些马氏链的相关概念*

对于马氏链，我们进一步引入几个相关概念，更多的概念大家会在随机过程相关的课程中学习到

- 可达：状态i称为可达状态j, 如果存在一个指标序列 $i_0 = i, i_1, \cdots, i_n = j$，使得

$$p_{i_k, i_{k+1}} > 0 \quad k = 0, 1, \cdots, n-1.$$

  用转移概率矩阵来刻画 i可达 j:

$$\exists\, n > 0, \quad (p^n)_{ij} > 0.$$

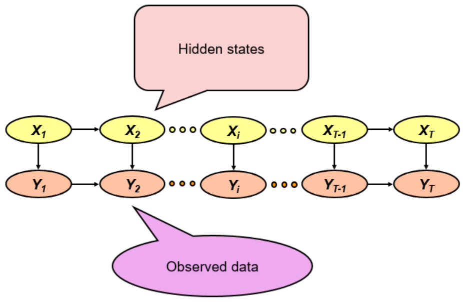
  这里，$(p^n)_{ij} = \Pr(\xi_n = j | \xi_0 = i)$ 是n步转移概率。

- 吸收：如果 $p_{ii} = 1$, 则称 i 是吸引（吸收）状态。
- 互通：状态i可达状态j，而且状态j可达状态i.
- 不可约：如果所有状态之间是互通的.
- 首达概率：令 $f_{ij}(n)$ 为由 i 出发在 n 步后首次达到 j 的概率，简称首达概率。即

$$f_{ij}^{(n)} = \Pr\big(\xi_n = j, \xi_m \neq j, m = 1, \cdots, n-1 \,|\, \xi_0 = i\big).$$

- 令 f_{ij}^* 是从 i 出发到达 j 的概率，即

$$f_{ij}^* = \sum_{n=1}^{\infty} f_{ij}^{(n)} \leq 1.$$



在很多实际问题中，我们并不能直接知道状态变量的取值，但是我们也许可以对状态变量做一些观察。为此，我们需要引入隐马氏链的数学模型：

- 隐过程为 $X = \{X_1, \cdots, X_T\}$， $X_i \in \{1, \cdots, N\}$.
- 观察过程为 $Y = \{Y_1, \cdots, Y_T\}$， $Y_i \in \{1, \cdots, M\}$.
- 模型参数为 $\lambda = \{\pi, A, B\}$，其中
  - 初始分布 $\pi = (\pi_i), \pi_i = \Pr(X_1 = i)$.
  - 转移矩阵 $A = (a_{ij}), a_{ij} = \Pr(X_{n+1} = j | X_n = i)$.
  - 给定某时间的隐状态的条件下，观测的分布矩阵 $B = (b_{i\ell}), b_{i\ell} = b_i(l) = \Pr(Y_n = \ell | X_n = i)$.

该模型在统计模型中有很重要的地位，我们在本课程中就不展开介绍了。

8.4 随机微分方程模型简介

我们回顾，在第2章中，我们学习了基于一阶微分方程的变化率模型。事实上，在模型的构成和使用中，随机性是普遍存在。一方面，模型在建立的过程中，不可避免地会存在数学上的简化假设和非第一性原理的经验结论的使用；另一方面，模型中的各类参数通常不能通过严格的数学推导和实验数据建立精准的关系，而且误差和噪声在实验测量中也是无所不在。

有“噪声”的变化率模型（常微分方程）可以写成如下的形式：

$$\frac{\mathrm{d}X_t}{\mathrm{d}t} = b(t, X_t) + \sigma(t, X_t)\eta(t).$$

数学上，我们常对噪声 $\eta(t)$ 做如下的假设

1. $\mathbb{E}\eta(t) = 0$，即对于任意固定的时刻，噪声的期望是 0;
2. （*）$\eta(t)$ 是平稳的。平稳的严格定义超出了本课的范围，我们直观的理解为这个随机过程统计特性不随时间的推移而变化;
3. $\mathbb{E}\eta(t)\eta(s) = 0$ if $t \neq s$.

这样的噪声被称为**白噪声**。事实上，可以证明，满足这样的 $\eta(t)$ 是不会有连续的路径的。为了理解这类方程，我们来考虑一个离散的形式：在时间点$t_0 < t_1 \cdots$，我们令 $X_k = X(t_k)$:

$$X_{k+1} - X_k = b(t_k, X_k)\Delta t_k + \sigma(t_k, X_k)\int_{t_k}^{t_{k+1}} \eta(s)\,\mathrm{d}s. \tag{8.5}$$

我们不妨引入$V_t = V(t)$，使得

$$\int_{t_k}^{t_{k+1}} \eta(s)\,\mathrm{d}s = V(t_{k+1}) - V(t_k) =: \Delta V_k. \tag{8.6}$$

根据$\eta(t)$的性质，我们知道$\mathbb{E}V_t = 0$。而且，因为我们希望 X_t 至少是连续的，所以我们要求 V_t 有连续的路径。事实上，在一定意义下，唯一可能的满足上述条件的过程是布朗运动（Brownian motion, or Wiener process）。我们用 W_t 表示布朗运动，它除了$\mathbb{E}W_t = 0$ 还满足

- $\mathbb{E}W_tW_s = \min(s,t)$;
- for $t > s$, $W_t - W_s \sim N(0, t-s)$.

于是，我们可设，

$$X_k = X_0 + \sum_{j=0}^{k-1} b(t_j, X_j)\Delta t_j + \sum_{j=0}^{k-1} \sigma(t_j, X_j)\Delta W_j, \tag{8.7}$$

然后考虑 $\Delta t_j \to 0$ 的极限。假设极限存在，则我们可写成

$$X_t = X_0 + \int_0^t b(s, X_s)\,\mathrm{d}s + \int_0^t \sigma(s, X_s)\,\mathrm{d}W_s, \tag{8.8}$$

或者他的微分形式

$$\mathrm{d}X_t = b(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}W_t. \tag{8.9}$$

注意，布朗运动 W_t 其实是关于时间不可微的，因此，我们常避免写 $\frac{\mathrm{d}W_t}{\mathrm{d}t}$。虽然事实上，一些人还是会这么写。

我们需要理解在什么意义下，对于哪些随机过程 f，$\int_0^t f\,\mathrm{d}W_s$ 是存在的。因此，我们先考虑如下的积分

$$I = \int_0^t W_s\,\mathrm{d}W_s. \tag{8.10}$$

让我们尝试在 $[k\Delta t, (k+1)\Delta t)$ 等区间上用黎曼积分计算这个积分。我们考虑两种特殊的选择

A. 被积函数在区间的左端点取值

$$I^{(n)} = \sum_{k=0}^{n-1} W_{k\Delta t}\big(W_{(k+1)\Delta t} - W_{k\Delta t}\big), \tag{8.11}$$

于是，我们计算得到

$$\mathbb{E} I^{(n)} = \sum k\Delta t - k\Delta t = 0. \tag{8.12}$$

B. 被积函数在区间的中点取值

$$I^{(n)} = \sum_{k=0}^{n-1} W_{(k+\frac{1}{2})\Delta t}\big(W_{(k+1)\Delta t} - W_{k\Delta t}\big), \tag{8.13}$$

于是我们计算得到

$$\mathbb{E} I^{(n)} = \sum (k+\frac{1}{2})\Delta t - k\Delta t = \frac{t}{2}. \tag{8.14}$$

可见，在不同的逼近选择下，我们得到了不同的极限。所以，黎曼积分是并不存在的。为了解决这个问题，我们需要先决定被积函数在每个区间的逼近方式。这里，有两个常见的选择：

定义 1 *Itô integral* 是黎曼和中被积函数在左端点取值的极限，记为

$$\int_0^t f(s,\omega)\,\mathrm{d}W_s.$$

定义 2 (*) *Stratonovich integral* 是黎曼和中被积函数利用梯形公式取值的极限，即

$$\sum_j \frac{1}{2}\big(f(t_j,\omega) + f(t_{j+1},\omega)\big)\Delta W_j.$$

记为

$$\int_0^t f(s,\omega)\circ \mathrm{d}W_s.$$

在下文中，我们只考虑 Itô integral。注意 $\int_0^t f(s,\omega)\,\mathrm{d}W_s$ 是一个随机变量。这里，我们列举 Itô Calculus 的几个常用的性质，其中前两个是对 Itô integral 的刻画，而最后一个著名的 Itô formula，可以看作是 Itô Calculus 中的链锁法则。

- 对于 Itô integral，我们有如下的一般性结果

$$\mathbb{E}\int_0^t f(s,\omega)\,\mathrm{d}W_s = 0. \tag{8.15}$$

- 对于 Itô integral，我们有如下的 Itô isometry

$$\mathbb{E}\Big(\int_0^t f(s,\omega)\,\mathrm{d}W_s\Big)^2 = \mathbb{E}\int_0^t f(s,\omega)^2\,\mathrm{d}s. \tag{8.16}$$

- Itô formula (Itô lemma):

  Let X_t be the solution to

$$\mathrm{d}X_t = b(t,\omega)\,\mathrm{d}t + \sigma(t,\omega)\,\mathrm{d}W_t$$

  where b, σ are functions of (t,ω). Given $g(t,x) \in C^2([0,\infty)\times\mathbb{R})$. Then $Y_t = g(t,X_t)$ satisfies the equation

$$\mathrm{d}Y_t = \frac{\partial g}{\partial t}(t,X_t)\,\mathrm{d}t + \frac{\partial g}{\partial x}(t,X_t)\,\mathrm{d}X_t + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(t,X_t)(\mathrm{d}X_t)^2, \tag{8.17}$$

  where $(\mathrm{d}X_t)^2$ is computed according to the rules

$$\mathrm{d}t\cdot\mathrm{d}t = \mathrm{d}t\cdot\mathrm{d}W_t = \mathrm{d}W_t\cdot\mathrm{d}t = 0, \qquad \mathrm{d}W_t\cdot\mathrm{d}W_t = \mathrm{d}t.$$

例子1. 计算 $\mathrm{d}(\frac{1}{2}W_t^2)$。令 $g(t,x)=\frac{1}{2}x^2$，则

$$\mathrm{d}Y_t = W_t\,\mathrm{d}W_t + \frac{1}{2}(\mathrm{d}W_t)^2 = W_t\,\mathrm{d}W_t + \frac{1}{2}\,\mathrm{d}t.$$

例子2. Ornstein-Uhlenbeck process. 考虑方程

$$\mathrm{d}X_t = -aX_t\,\mathrm{d}t + \sigma\,\mathrm{d}W_t, \quad X_0 = \xi. \tag{8.18}$$

注意到，右端的第一项代表指数型衰减，而第二项表示“波动”变化。

我们来利用积分因子法求解此问题。方程两边乘以 e^{at}，我们得到

$$\mathrm{d}\big(e^{at}X_t\big) = e^{at}\,\mathrm{d}X_t + ae^{at}X_t\,\mathrm{d}t = \sigma e^{at}\,\mathrm{d}W_t, \tag{8.19}$$

于是，积分得到

$$e^{at}X_t - X_0 = \int_0^t \sigma e^{as}\,\mathrm{d}W_s. \tag{8.20}$$

所以，方程的解是

$$X_t = e^{-at}X_0 + \sigma\int_0^t e^{-a(t-s)}\,\mathrm{d}W_s. \tag{8.21}$$

我们注意到，方程对初值的“记忆”会指数型衰减。并且，我们做如下的计算

- Mean. $\mathbb{E}X_t = e^{-at}\mathbb{E}X_0$ 指数型衰减到 0。
- Variance. (假设初值和 W_t 无关)

$$\begin{aligned}\operatorname{Var}X_t = \mathbb{E}(X_t)^2 - (\mathbb{E}X_t)^2 &= e^{-2at}\big(\mathbb{E}X_0^2 - (\mathbb{E}X_0)^2\big) + \sigma^2\mathbb{E}\left[\left(\int_0^t e^{-a(t-s)}\,\mathrm{d}W_s\right)^2\right]\\ &= e^{-2at}\operatorname{Var}X_0 + \sigma^2\int_0^t e^{-2a(t-s)}\,\mathrm{d}s\\ &= e^{-2at}\operatorname{Var}X_0 + \frac{\sigma^2}{2a}\big(1-e^{-2at}\big).\end{aligned}$$

特别的，我们注意到 $\lim_{t\to\infty}\operatorname{Var}X_t = \frac{\sigma^2}{2a}$.

8.4.1 更多的例子*

例子3. Geometric Brownian motion. 考虑方程

$$\mathrm{d}N_t = rN_t\,\mathrm{d}t + \alpha N_t\,\mathrm{d}W_t. \tag{8.22}$$

这是另外一个人口增长模型，这里 r 是（相对）增长系数，α 是波动系数。这个例子也跟金融数学里面的Black-Scholes model 有关。这个意义下，N_t 可以理解称资产的定价，α 可以理解为波动性。

为了求解此方程，我们在两端除以N_t，便得到

$$\frac{\mathrm{d}N_t}{N_t} = r\,\mathrm{d}t + \alpha\,\mathrm{d}W_t. \tag{8.23}$$

注意，利用 Itô formula，我们有

$$\mathrm{d}(\log N_t) = \frac{1}{N_t}\,\mathrm{d}N_t - \frac{1}{2N_t^2}(\mathrm{d}N_t)^2 = \frac{1}{N_t}\,\mathrm{d}N_t - \frac{\alpha^2}{2}\,\mathrm{d}t. \tag{8.24}$$

于是我们得到，

$$\mathrm{d}(\log N_t) = \big(r - \frac{\alpha^2}{2}\big)\,\mathrm{d}t + \alpha\,\mathrm{d}W_t, \tag{8.25}$$

容易得出方程的解

$$N_t = N_0 e^{(r-\frac{\alpha^2}{2})t+\alpha W_t}. \tag{8.26}$$

接下来，我们来计算一下解的均值（期望）。注意到，问题的关键在于求 $Y_t := e^{\alpha W_t}$ 的均值。利用 Itô formula，我们有

$$\mathrm{d}Y_t = \alpha e^{\alpha W_t}\,\mathrm{d}W_t + \frac{\alpha^2}{2}e^{\alpha W_t}\,\mathrm{d}t = \alpha Y_t\,\mathrm{d}W_t + \frac{\alpha^2}{2}Y_t\,\mathrm{d}t, \tag{8.27}$$

两边积分，并求期望，我们得到

$$\mathbb{E}Y_t = \mathbb{E}Y_0 + \frac{\alpha^2}{2}\int_0^t \mathbb{E}Y_s\,\mathrm{d}s, \tag{8.28}$$

于是有

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}Y_t = \frac{\alpha^2}{2}\mathbb{E}Y_t. \tag{8.29}$$

显然，$\mathbb{E}Y_0 = 1$，于是我们有$\mathbb{E}Y_t = e^{\frac{\alpha^2}{2}t}$。

最终，我们得到，

$$\mathbb{E}N_t = (\mathbb{E}N_0)e^{rt},$$

也就是说，人口数量期望的增长和变化率模型的人口增长是一样的。（注意，按照第二章的模型建立方法，我们有简单的没有随机因素的人口增长模型 $\dot{N} = rN$.）

8.4.2 Fokker-Planck equation

最后，我们换一个角度看随机微分方程。我们考虑 X_t 满足

$$\mathrm{d}X_t = b(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}W_t. \tag{8.30}$$

注意，对于固定的 t，X_t 是一个随机变量，我们令它的概率密度函数为 $\rho(x,t)$，那么已知了 X_t 的时间演化，我们如何得到 $\rho(x,t)$ 的时间演化呢？

我们先回忆一下，在第三章的 Reynolds Transport Theorem（或者第一章的守恒律模型），如果 $\sigma \equiv 0$，则 X_t 的方程退化成一个确定的特征线方程

$$\mathrm{d}X_t = b(t, X_t)\,\mathrm{d}t. \tag{8.31}$$

我们知道，这时候密度函数 $\rho(x,t)$ 满足连续性方程

$$\frac{\partial}{\partial t}\rho(t,x) + \frac{\partial}{\partial x}\big(b(x)\rho(t,x)\big) = 0. \tag{8.32}$$

那么噪声对密度函数的演化产生了什么影响呢？下面只给出一个形式上的推导。考虑任何一个光滑函数 $g(x)$, 由 Ito's formula，我们有

$$\begin{aligned} \mathrm{d}g(X_t) &= \frac{\partial g(X_t)}{\partial x}\,\mathrm{d}X_t + \frac{1}{2}\frac{\partial^2 g(X_t)}{\partial x^2}\sigma(X_t)^2\,\mathrm{d}t \\ &= \left(\frac{\partial g(X_t)}{\partial x}b(X_t) + \frac{1}{2}\frac{\partial^2 g(X_t)}{\partial x^2}\sigma(X_t)^2\right)\mathrm{d}t + \frac{\partial g(X_t)}{\partial x}\sigma(X_t)\,\mathrm{d}W_t. \end{aligned} \tag{8.33}$$

通过取期望，我们得到

$$\mathrm{d}\mathbb{E}g(X_t) = \mathbb{E}\left(\frac{\partial g(X_t)}{\partial x}b(X_t) + \frac{1}{2}\frac{\partial^2 g(X_t)}{\partial x^2}\sigma(X_t)^2\right)\mathrm{d}t. \tag{8.34}$$

利用，概率密度函数 $\rho(x,t)$，可以把上式表达为

$$\begin{aligned}\frac{\partial}{\partial t}\int \rho(t,x)g(x)\,\mathrm{d}x &= \int \Big(g'(x)b(x)+\frac{1}{2}g''(x)\sigma(x)^2\Big)\rho(t,x)\,\mathrm{d}x \\ &= \int g(x)\Big(-\frac{\partial}{\partial x}\big(b(x)\rho(t,x)\big)+\frac{1}{2}\frac{\partial^2}{\partial x^2}\big(\sigma^2(x)\rho(t,x)\big)\Big)\,\mathrm{d}x.\end{aligned} \tag{8.35}$$

如果我们把 $g(x)$ 看作一个试验函数，由 g 的任意性，我们得到

$$\frac{\partial}{\partial t}\rho(t,x) = -\frac{\partial}{\partial x}\big(b(x)\rho(t,x)\big)+\frac{1}{2}\frac{\partial^2}{\partial x^2}\big(\sigma^2(x)\rho(t,x)\big). \tag{8.36}$$

这就是所谓的 forward Kolmogorov equation，又叫做 Fokker-Planck equation。

我们注意到，Fokker-Planck equation 其实是一种对流扩散方程。下面观察一些特例：

- 如果没有随机微分方程没有噪声项的话（即 $\sigma = 0$，随机微分方程变成了常微分方程），那么 Fokker-Planck equation 其实退化成了第一章（或者第三章）的守恒律方程。
- 而微观运动中的噪声项引入了密度函数的扩散项。特别的，如果 $b\equiv 0$ 而 $\sigma=\sqrt{2\beta}$，则 Fokker-Planck equation 变成了我们熟悉的扩散方程

$$\frac{\partial}{\partial t}\rho(t,x) = \beta\frac{\partial^2}{\partial x^2}\rho(t,x).$$

本学期的课程内容全部结束了！但是数学模型的学习道路是没有终点的……