

中山大学课程讲义

# 随机过程

龚 杰

中山大学

广州

# 目 录

<b>第 1 章 随机过程的基本概念</b>	<b>3</b>
1.1 随机过程的定义与例子 . . . . .	3
1.2 相关函数 . . . . .	4
1.3 平稳性 . . . . .	5
1.4 遍历性 . . . . .	7
<b>第 2 章 概率论回顾</b>	<b>8</b>
2.1 随机变量与概率 . . . . .	8
2.2 随机向量 . . . . .	11
2.3 数字特征 . . . . .	13
2.3.1 数学期望 . . . . .	13
2.3.2 方差 . . . . .	16
2.3.3 协方差 . . . . .	16
2.4 特征函数 . . . . .	18
2.5 大数定律和中心极限定理 . . . . .	21
<b>第 3 章 离散时间 Markov 链</b>	<b>23</b>
3.1 定义与例子 . . . . .	23
3.1.1 离散时间 Markov 链的定义 . . . . .	23
3.1.2 Markov 链的表示方法 . . . . .	25
3.1.3 离散时间 Markov 链举例 . . . . .	26
3.2 状态的分类与性质 . . . . .	31
3.2.1 可达与相通 . . . . .	32
3.2.2 周期性 . . . . .	33
3.2.3 常返性 . . . . .	34
3.2.4 非常返态分析 . . . . .	45
3.3 极限行为与平稳分布 . . . . .	48
3.3.1 Markov 链的极限行为 . . . . .	48
3.3.2 平稳分布 . . . . .	49
3.3.3 应用举例——搜索引擎 PageRank 技术 . . . . .	53
<b>第 4 章 Poisson 过程</b>	<b>56</b>
4.1 Poisson 过程的定义 . . . . .	56
4.2 Poisson 过程的性质 . . . . .	59

4.2.1	联合分布及条件分布 . . . . .	59
4.2.2	到达时刻与事件间隔 . . . . .	60
4.3	Poisson 过程的拓广 . . . . .	64
4.3.1	非齐次 Poisson 过程 . . . . .	65
4.3.2	复合 Poisson 过程 . . . . .	66
4.3.3	条件 Poisson 过程 . . . . .	69
4.3.4	更新过程 . . . . .	71
<b>第 5 章</b>	<b>连续时间 Markov 链</b>	<b>75</b>
5.1	定义与例子 . . . . .	75
5.2	Q 矩阵 . . . . .	77
5.2.1	Q 矩阵的物理意义 . . . . .	77
5.2.2	Kolmogorov 向前-向后方程 . . . . .	80
5.3	转移概率的极限行为 . . . . .	82
5.4	排队论简介 . . . . .	85
5.4.1	排队模型基本概念 . . . . .	85
5.4.2	基本 Markov 排队模型: M/M/1 . . . . .	87
5.4.3	Little 定理 . . . . .	88
5.4.4	基本 Markov 排队模型的扩展 . . . . .	91
<b>第 6 章</b>	<b>Gauss 过程</b>	<b>93</b>
6.1	Gauss 过程的定义 . . . . .	93
6.1.1	多元 Gauss 分布的定义 . . . . .	93
6.1.2	多元 Gauss 分布的特征函数 . . . . .	94
6.2	多元 Gauss 分布的性质 . . . . .	95
6.2.1	线性性质 . . . . .	95
6.2.2	独立性 . . . . .	97
6.2.3	条件分布 . . . . .	99
6.2.4	应用举例——生成式人工智能 DDPM . . . . .	100
6.3	Brown 运动 . . . . .	102
6.3.1	定义与基本性质 . . . . .	102
6.3.2	最大值分布 . . . . .	105

# 第 1 章 随机过程的基本概念

## 1.1 随机过程的定义与例子

回顾概率论中的随机变量 (Random Variable), 它用来刻画单次随机试验的结果, 侧重单个随机事件的概率分布特性. 区别于此, 随机过程考察一系列随时间变化的随机变量, 侧重多个随机事件的关联性. 具体定义如下:

**定义1.1.** 随机过程 (Stochastic Process) 是指一组依赖于实参  $t$  的随机变量. 记为  $\{X(t), t \in \mathbb{R}\}$ , 或简记为  $X(t)$ , 其中  $t$  通常具有时间的含义.

为了描述的统一和方便, 把  $t$  称为**时间**, 把给定实参  $t$  对应的随机变量  $X(t)$  的取值称为**状态**.

由定义可以看出, 随机过程是一个二元函数的概念: (1) 固定时间  $t$ ,  $X(t)$  为一个依赖于  $t$  的随机变量. (2) 固定样本  $\omega$ ,  $X(t)$  为一个以  $t$  为自变量的函数, 是随机过程的一次“实现”, 称为**样本轨道** (sample path).

根据时间和状态的取值范围, 随机过程可以分为以下四类:

- (1) 离散时间离散状态随机过程. 时间参数取离散值时, 通常用  $n \in \mathbb{N}$  来表示. 相应的, 离散时间随机过程记为  $X_n$ .

**例1.1.** 连续抛掷一枚硬币, 若正面朝上则得一分, 若反面朝上则扣一分, 前  $n$  次抛掷后的总得分用  $X_n \in \mathbb{N}$  表示, 是一个离散时间离散状态随机过程, 也即后续章节要深入探讨的**随机徘徊** (Random Walk).

- (2) 离散时间连续状态随机过程, 仍记为  $X_n$ .

**例1.2.** 考察由以下递推方程给出的线性预测模型:

$$X_n = \alpha X_{n-1} + Z_n, \quad n \geq 1.$$

其中  $\alpha$  为给定常数,  $Z_n, n \geq 1$  为相互独立的零均值连续随机变量.  $X_n$  即为离散时间连续状态随机过程, 称为一阶自回归过程, 记为 AR(1).

- (3) 连续时间离散状态随机过程.

**例1.3.** 任意时刻  $t$ , 银行营业厅中排号等待办理业务的顾客数  $X(t) \in \mathbb{N}$  是一个连续时间离散状态随机过程. 由于顾客随机到达营业厅, 办理业务的时长也是随机的,  $X(t)$  由顾客达到过程和服务离去过程共同决定. 我们将在排队理论相关章节加以讨论.

(4) 连续时间连续状态随机过程.

**例1.4.** 分子在空间中随机运动的位置  $B(t)$  称为 Wiener 过程, 也称为 Brown 运动. 该过程是一个典型的连续时间连续状态随机过程, 在理论和应用中都非常重要. 后续也将深入讨论.

很多概率问题中, 多个随机变量之间的时间关联性起到至关重要的作用. 而关联性的一个特例是相互独立, 此时随机过程的分布特性由单个随机变量的分布唯一决定. 看下面的例子.

**例1.5.** 连续抛掷一枚硬币, 第  $n$  次抛掷的结果用  $Y_n$  表示,  $Y_n = 1$  表示正面朝上,  $Y_n = 0$  表示背面朝上. 则  $Y_n$  构成一个随机过程. 由于任意两次抛掷相互独立, 我们只需要考察任意给定  $n$  时随机变量  $Y_n$  的分布即可. 而在例 1.1 中,  $X_n$  则由历史结果和第  $n$  次实验的结果共同决定. 具体而言, 我们可以得到如下关系:

$$X_n = \sum_{k=1}^n Y_k$$

一般情况下, 由于存在相关性, 随机过程的分布特性可以用任意维联合分布来刻画, 即

$$F_{X(t_1), \dots, X(t_n)}(x_1, \dots, x_n), \quad \forall n, t_1, t_2, \dots, t_n.$$

任意维联合分布蕴含了丰富的关联信息. 但是用联合分布来刻画关联性并不方便, 为此引入相关函数的概念.

## 1.2 相关函数

**定义1.2.** 随机过程  $X(t)$  的自相关函数 (Auto-correlation, 简称为相关函数) 定义为

$$R_X(t, s) = E[X(t)X(s)], \quad \forall t, s.$$

自相关函数是一个二元确定性函数, 刻画了随机过程任意两个时刻对应的随机事件的相关程度. 与自相关函数相对应, 还可以定义两个随机过程  $X(t), Y(t)$  的互相关函数, 即  $R_{XY}(t, s) = E[X(t)Y(s)]$ .

**例1.6.** 计算随机徘徊 (例 1.1) 的相关函数.

**解:** 设  $Z_n \in \{-1, 1\}$  表示第  $n$  次抛掷的得分, 满足

$$P(Z_n = 1) = P(Z_n = -1) = \frac{1}{2},$$

则

$$X_n = \sum_{k=1}^n Z_k.$$

不妨设  $m < n$ , 有

$$\begin{aligned} R_X(n, m) &= E(X_n X_m) \\ &= E\left(\sum_{k=1}^n Z_k \sum_{l=1}^m Z_l\right) \\ &= E\left(\sum_{k=1}^m Z_k^2 + \sum_{k \neq l} Z_k Z_l\right) \\ &= \sum_{k=1}^m E(Z_k^2) + \sum_{k \neq l} E Z_k E Z_l \\ &= m. \end{aligned}$$

亦即  $R_X(n, m) = \min\{n, m\}$ ,  $\forall n, m$ . □

**定义1.3.** 随机过程  $X(t)$  的自协方差函数定义为

$$C_X(t, s) = E\{[X(t) - EX(t)][X(s) - EX(s)]\}, \quad \forall t, s.$$

通过简单计算可以发现, 如果随机过程的均值不变, 自相关函数与自协方差函数只相差一个常数, 即  $EX(t)EX(s)$ . 因此, 两者之间没有本质的区别.

### 1.3 平稳性

平稳性 (stationary) 用来刻画随机过程不随时间变化的某种统计特性. 针对不同的统计特性, 可以得到不同的平稳性的定义.

**定义1.4.** 若随机过程  $X(t)$  对  $\forall t, s, D$  都有

$$E(X(t)) = E(X(s))$$

$$R_X(t, s) = R_X(t + D, s + D),$$

则称随机过程  $X(t)$  具有**宽平稳性**.

根据定义, 宽平稳随机过程的均值不依赖于时间, 相关函数只依赖于时间差  $t - s$ , 而与绝对时刻无关, 即在任意时刻观察, 只要时间间隔不变, 相关函数就保持不变. 因此, 常常把宽平稳随机过程的相关函数写成一元函数的形式, 即

$$R_X(t, s) = R_X(t - s) = R_X(\tau).$$

**例1.7.** 考虑随机过程

$$X(t) = A \cos(\omega t + \Theta), t \in \mathbb{R}$$

的宽平稳性, 其中  $A$  和  $\Theta$  是相互独立的随机变量, 且  $\Theta \sim U(0, 2\pi)$ ,  $\omega$  为常数.

**解:**  $X(t)$  的期望为

$$\begin{aligned} E(X(t)) &= E(A \cos(\omega t + \Theta)) \\ &= E A E(\cos(\omega t + \Theta)) \\ &= E A \int_0^{2\pi} \cos(\omega t + \theta) \frac{1}{2\pi} d\theta \\ &= 0. \end{aligned}$$

再计算  $X(t)$  的相关函数

$$\begin{aligned} R_X(t, s) &= E(X(t)X(s)) = E(A \cos(\omega t + \Theta) A \cos(\omega s + \Theta)) \\ &= E(A^2 (\cos(\omega t + \Theta) \cos(\omega s + \Theta))) \\ &= E(A^2) E(\cos(\omega t + \Theta) \cos(\omega s + \Theta)) \\ &= \frac{E(A^2)}{2} E(\cos(\omega(t-s)) + \cos(\omega(t+s) + 2\Theta)) \\ &= \frac{E(A^2)}{2} \cos(\omega(t-s)). \end{aligned}$$

可以看到,  $X(t)$  的均值为常数, 相关函数只依赖于时间差  $t-s$ , 所以是宽平稳的.  $\square$

**定义1.5.** 若随机过程  $X(t)$  对  $\forall n, t_1, \dots, t_n, D$  都有

$$F_{t_1, \dots, t_n}(x_1, \dots, x_n) = F_{t_1+D, \dots, t_n+D}(x_1, \dots, x_n)$$

其中  $F_{t_1, \dots, t_n}$  表示  $X(t_1), \dots, X(t_n)$  的联合概率分布函数, 则称随机过程  $X(t)$  具有**严平稳性**.

顾名思义, 严平稳的要求比宽平稳要严苛得多, 后者只要求相关函数的时间不变性, 而前者对有限维联合分布提出了要求. 因此, 严平稳蕴含了宽平稳, 反之则不然.

**定义1.6.** 若两个宽平稳随机过程  $X(t), Y(t)$  对  $\forall t, s, D$  都有

$$R_{XY}(t, s) = R_{XY}(t+D, s+D),$$

则称随机过程  $X(t)$  和  $Y(t)$  具有**联合宽平稳性**.

## 1.4 遍历性

由于随机过程  $X(t)$  是二元函数  $X(t, \omega)$ , 即自变量包含时间  $t$  和样本空间的元素  $\omega$ , 因此可以从两个角度计算平均. 固定一条样本轨道  $\omega$ , 可以计算时间平均值, 即

$$\langle X(t) \rangle = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} X(t) dt.$$

固定时间  $t$ , 得到随机变量的期望, 称为集平均, 即

$$E(X(t)) = \int_{-\infty}^{\infty} x dF(x).$$

两者之间的联系是遍历理论所关注的问题. 我们给出遍历的如下定义.

**定义1.7.** 如果宽平稳随机过程  $X(t)$  满足

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} X(t) dt = E(X(t))$$

则称该过程是**均值遍历**的.

宽平稳性是均值遍历的必要条件, 因为时间平均不依赖于时间变量  $t$ , 按照定义, 集平均也应不依赖于时间. 从下面的例子可以看出, 时间平均和集平均一般是不等的. 也就是说, 遍历性需要满足一定的条件.

**例1.8.** 考察随机过程  $X(t) = Z$ , 其中  $Z$  是任意给定分布的随机变量. 该随机过程在任意时刻的取值都相等, 即样本轨道是一条平行于时间轴的直线. 因此,  $X(t)$  的集平均等于随机变量  $Z$  的期望  $EZ$ , 而时间平均则是  $Z$  本身, 仍然是一个随机变量.



## 第 2 章 概率论回顾

### 2.1 随机变量与概率

首先回顾一下用于描述随机性试验的相关概念.

- **样本  $\omega$** : 随机试验的每一个可能结果;
- **样本空间  $\Omega$** : 一个随机试验的所有可能结果的全体;
- **事件  $A$** : 随机试验某类结果的集合, 是样本空间的子集  $A \subseteq \Omega$ ;
- **事件类  $\mathcal{F}$** : 一个随机试验的所有事件的全体;
- **随机变量  $X$** : 按某种确定的方式将样本  $\omega$  对应于一个数量  $X(\omega)$ , 即从样本空间到实数域的映射:

$$X : \Omega \rightarrow \mathbb{R}$$

**例2.1.** 投掷一枚硬币, 投掷的结果要么为正面, 要么为反面, 即样本  $\omega = \text{正面/反面}$ . 于是, 该随机试验的样本空间为  $\Omega = \{\text{正面}, \text{反面}\}$ , 结果为正面组成的集合是一个事件  $A = \{\text{正面}\}$ . 定义映射

$$X(\omega) = \begin{cases} 1, & \omega = \text{正面}, \\ -1, & \omega = \text{反面}. \end{cases}$$

则  $X$  为一个离散型随机变量.

一次随机试验中, 某个事件  $A$  发生的可能性的的大小称为**概率**, 记为  $P(A)$ , 是从事件类到实数的一个映射:

$$P : \mathcal{F} \rightarrow \mathbb{R}$$

概率满足以下三条基本公理:

- **非负性**: 对任何事件  $A$ ,  $P(A) \geq 0$ ;
- **规范性**: 对必然事件  $\Omega$ ,  $P(\Omega) = 1$ ;
- **可数可加性**: 对不相容事件  $A_1, A_2, \dots$  有

$$P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \dots.$$

样本空间、事件类、以及概率组成的三元组  $(\Omega, \mathcal{F}, P)$  是一个**概率空间**, 包含了一个随机试验的全部信息. 概率论中重点讨论随机事件的概率, 也对随机事件之间的关系作了探讨, 很重要的一个概念就是条件概率. 设  $(\Omega, \mathcal{F}, P)$  是概率空间,  $B \in \mathcal{F}$  且  $P(B) > 0$ , 则对任意  $A \in \mathcal{F}$ ,

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

为在事件  $B$  发生的条件下, 事件  $A$  发生的**条件概率**. 条件概率满足以下几条重要性质:

- **乘法公式:** 对任意两个事件  $A_1, A_2 \in \mathcal{F}$ , 有

$$P(A_1 A_2) = P(A_1)P(A_2|A_1).$$

- **全概率公式:** 设  $B_1, B_2, \dots, B_n$  是样本空间  $\Omega$  的一个划分, 满足  $B_1 \cup B_2 \cup \dots \cup B_n = \Omega$ , 并且  $B_i \cap B_j = \emptyset, \forall i \neq j$ , 则对任意事件  $A \in \mathcal{F}$ , 有

$$P(A) = \sum_{k=1}^n P(B_k)P(A|B_k).$$

- **Bayes公式:** 对任意两个事件  $A, B \in \mathcal{F}$ , 满足  $P(B) > 0$ , 有

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

由乘法公式可以引出独立的概念. 对任意两个事件  $A, B \in \mathcal{F}$ , 如果  $P(AB) = P(A)P(B)$ , 则称事件  $A, B$  **相互独立**. 两个事件相互独立等价于  $P(B) = P(B|A)$ , 即事件  $A$  发生与否不会影响事件  $B$  发生的概率.

引入随机变量的概念之后, 关于概率的讨论便可以脱离具体的随机试验, 而聚集于其背后的数学模型. 随机变量的概率特性可以用概率分布函数统一刻画, 如果对任何  $x \in \mathbb{R}$ ,  $\{\omega \in \Omega : X(\omega) \leq x\}$  是  $\mathcal{F}$  中的事件, 则称函数

$$F_X(x) = P(X \leq x)$$

为随机变量  $X$  的**概率分布函数**.

除了概率分布函数, 根据随机变量的取值范围, 还可以使用不同的概率描述. 离散型随机变量  $X$  的取值有限或可数, 用分布律描述较为方便, 即  $X$  取某一具体值的概率大小. 对于连续型随机变量, 其取值有连续实数, 则存在非负可积函数  $f$  (称为**概率密度函数**) 使得

$$F_X(x) = \int_{-\infty}^x f(u)du, \quad \forall x \in \mathbb{R}.$$

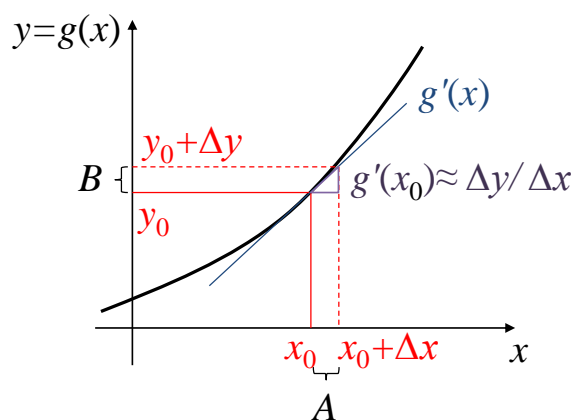


图 2.1 随机点坐标在二维平面中的关系

随机变量很重要的一个问题是计算随机变量的函数的分布. 设  $X$  是任意给定的连续型随机变量,  $Y = g(X)$  是关于  $X$  的函数, 也是一个随机变量, 其概率密度  $f_Y(y)$  可以借助概率分布函数  $F_Y(y)$  得到, 具体来说, 由

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = \int_{x: g(x) \leq y} f_X(x) dx,$$

于是可以得到

$$f_Y(y) = F'_Y(y).$$

事实上, 我们可不借助概率分布函数而直接计算概率密度函数. 考虑二维平面上的随机点  $(X, Y)$ , 其中  $Y = g(X)$ , 函数  $y = g(x)$  单调且可导, 则随机点只会出现在函数  $y = g(x)$  的图像上, 如图 2.1 所示. 考察随机点  $(X, Y)$  落在曲线上一点  $(x_0, y_0)$  附近的小弧段上的事件, 我们可以通过  $x$  坐标和  $y$  坐标观察这个现象. 从  $x$  坐标看,  $X$  落在  $x_0$  附近的一个小线段  $A$  上. 从  $y$  坐标看,  $Y$  落在  $y_0$  附近的一个小线段  $B$  上, 满足  $B = g(A)$ . 因此,

$$P(X \in A) = P(Y \in B).$$

取  $A = (x_0, x_0 + \Delta x)$ ,  $B = (y_0, y_0 + \Delta y)$ , 则当  $\Delta x$  和  $\Delta y$  足够小时, 上式近似可表为

$$f_X(x_0)|\Delta x| \approx f_Y(y_0)|\Delta y|.$$

于是

$$\begin{aligned} f_Y(y_0) &= \lim_{\Delta x \rightarrow 0} f_X(x_0) \left| \frac{\Delta x}{\Delta y} \right| \\ &= f_X(g^{-1}(y_0)) |(g^{-1})'(y_0)|. \end{aligned}$$

上述方法称为**概率微元法**, 可以方便地拓展到函数  $y = g(x)$  非单调的情况. 下一节即将看到, 这一方法也可以拓展到高维空间中随机向量间的映射.

## 2.2 随机向量

设  $X_1, \dots, X_n$  是同一概率空间上的随机变量, 称  $\mathbf{X} = (X_1, \dots, X_n)^T$  为一个随机向量, 称  $F_{\mathbf{X}} = F_{X_1, \dots, X_n} : \mathbb{R}^n \rightarrow \mathbb{R}$ ,

$$F_{\mathbf{X}}(\mathbf{x}) = F_{X_1, \dots, X_n}(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n)$$

为  $X_1, \dots, X_n$  的联合概率分布函数. 称  $X_1, \dots, X_n$  中部分随机变量的联合概率分布函数为随机向量  $\mathbf{X}$  的边缘概率分布函数.

对于连续型随机向量, 可以类似地定义概率密度. 设  $\mathbf{X} = (X_1, \dots, X_n)^T$  是  $n$  维连续型随机变量, 如果存在  $n$  元函数  $f(x_1, \dots, x_n)$  满足

- (1)  $f(x_1, \dots, x_n) \geq 0$ ;
- (2)  $\int_{\mathbb{R}^n} f(x_1, \dots, x_n) dx_1 \cdots dx_n = 1$ ;
- (3)  $\mathbf{X}$  落在  $\mathbb{R}^n$  中任一区域  $G$  的概率为

$$P(\mathbf{X} \in G) = \int_G f(x_1, \dots, x_n) dx_1 \cdots dx_n,$$

则称  $f(x_1, \dots, x_n)$  为  $n$  维连续型随机变量  $\mathbf{X}$  的联合概率密度函数.

利用联合概率密度函数, 进一步可以定义边缘概率密度函数. 对于  $n$  维随机变量  $\mathbf{X}$ , 任意低于  $n$  维的随机向量均为其边缘. 比如, 其 1 维边缘概率密度为

$$f_{X_1}(x_1) = \int_{\mathbb{R}^{n-1}} f(x_1, \dots, x_n) dx_2 \cdots dx_n,$$

$k$  维边缘概率密度可表为

$$f_{X_1, \dots, X_k}(x_1, \dots, x_k) = \int_{\mathbb{R}^{n-k}} f(x_1, \dots, x_n) dx_{k+1} \cdots dx_n.$$

高维情形下的条件概率密度函数可按如下方式定义. 设  $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)^T$ , 则给定  $\mathbf{X}_2 = \mathbf{x}_2$  时  $\mathbf{X}_1$  的条件概率密度为

$$f_{\mathbf{X}_1|\mathbf{X}_2}(\mathbf{x}_1|\mathbf{x}_2) = \frac{f_{\mathbf{X}_1, \mathbf{X}_2}(\mathbf{x}_1, \mathbf{x}_2)}{f_{\mathbf{X}_2}(\mathbf{x}_2)}.$$

我们考察随机向量的函数的分布. 当两个随机向量之间存在可微的一一映射时, 他们的联合概率密度存在如下换算关系.

**定理2.1.** 设  $\mathbf{X} = (X_1, \dots, X_n)^T$  为连续型随机向量,  $f_{\mathbf{X}}(x_1, \dots, x_n)$  是联合概率密度函数.  $T: \mathbb{R}^n \rightarrow \mathbb{R}^n$  是一一映射, 且  $T, T^{-1}$  均可微. 则随机向量  $\mathbf{Y} = T(\mathbf{X})$  联合概率密度函数为

$$f_{\mathbf{Y}}(y_1, \dots, y_n) = f_{\mathbf{X}}(x_1, \dots, x_n) \left| \frac{\partial(x_1, \dots, x_n)}{\partial(y_1, \dots, y_n)} \right|,$$

其中,  $(x_1, \dots, x_n) = T^{-1}(y_1, \dots, y_n)$ ,

$$\frac{\partial(x_1, \dots, x_n)}{\partial(y_1, \dots, y_n)} = \det \begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \dots & \frac{\partial x_1}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial y_1} & \dots & \frac{\partial x_n}{\partial y_n} \end{pmatrix}$$

是逆变换  $\mathbf{x} = T^{-1}(\mathbf{y})$  的 Jacobi 矩阵的行列式.

**证明:** 设  $\mathbf{x} \in \mathbb{R}^n$  是  $n$  维空间中的任意一点, 且  $\mathbf{y} = T(\mathbf{x})$ . 则映射  $T$  将  $\mathbf{x}$  的邻域  $O_{\mathbf{x}}$  变换为  $\mathbf{y}$  的邻域  $O_{\mathbf{y}} = T(O_{\mathbf{x}})$ , 且满足  $O_{\mathbf{x}} = T^{-1}(O_{\mathbf{y}})$ . 于是有

$$P(\mathbf{Y} \in O_{\mathbf{y}}) = P(\mathbf{X} \in T^{-1}(O_{\mathbf{y}})).$$

可以得到

$$f_{\mathbf{Y}}(\mathbf{y})|O_{\mathbf{y}}| + o(|O_{\mathbf{y}}|) = f_{\mathbf{X}}(\mathbf{x})|T^{-1}(O_{\mathbf{y}})| + o(|T^{-1}(O_{\mathbf{y}})|),$$

其中,  $|O_{\mathbf{y}}|$  和  $|T^{-1}(O_{\mathbf{y}})|$  分别表示邻域  $O_{\mathbf{y}}$  和  $T^{-1}(O_{\mathbf{y}})$  的  $n$  维体积. 于是

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y}) &= \lim_{|O_{\mathbf{y}}| \rightarrow 0} f_{\mathbf{X}}(\mathbf{x}) \frac{|T^{-1}(O_{\mathbf{y}})|}{|O_{\mathbf{y}}|} \\ &= \lim_{|O_{\mathbf{y}}| \rightarrow 0} f_{\mathbf{X}}(\mathbf{x}) \frac{|(T^{-1})' O_{\mathbf{y}}| + o(|O_{\mathbf{y}}|)}{|O_{\mathbf{y}}|} \\ &= f_{\mathbf{X}}(\mathbf{x}) \left| \frac{\partial(x_1, \dots, x_n)}{\partial(y_1, \dots, y_n)} \right|. \quad \square \end{aligned}$$

**例2.2.** 设随机变量  $X$  和  $Y$  的联合概率密度为  $f(x, y)$ , 求  $U = \alpha_1 X + \beta_1 Y$  和  $V = \alpha_2 X + \beta_2 Y$  的联合概率密度函数. 其中  $\alpha_1 \beta_2 - \alpha_2 \beta_1 \neq 0$ .

**解:** 将映射关系写成矩阵形式有

$$\begin{pmatrix} U \\ V \end{pmatrix} = \begin{pmatrix} \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}.$$

逆变换为

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \frac{1}{\alpha_1 \beta_2 - \alpha_2 \beta_1} \begin{pmatrix} \beta_2 & -\beta_1 \\ -\alpha_2 & \alpha_1 \end{pmatrix} \begin{pmatrix} U \\ V \end{pmatrix}.$$

根据定理 2.1,

$$\begin{aligned} f_{U,V}(u, v) &= f(x, y) \left| \frac{\partial(x, y)}{\partial(u, v)} \right| \\ &= \frac{1}{|\alpha_1\beta_2 - \alpha_2\beta_1|} f\left(\frac{\beta_2u - \beta_1v}{\alpha_1\beta_2 - \alpha_2\beta_1}, \frac{\alpha_1v - \alpha_2u}{\alpha_1\beta_2 - \alpha_2\beta_1}\right). \end{aligned}$$

□

利用上述结果, 可以方便地计算两个随机变量的和的概率密度. 具体来说, 令  $\alpha_1 = \beta_1 = \beta_2 = 1, \alpha_2 = 0$ , 则

$$f_{X+Y,Y}(u, v) = f(u - v, v).$$

于是得到,

$$f_{X+Y}(u) = \int_{-\infty}^{\infty} f(u - v, v) dv.$$

此即随机变量的和的概率密度公式. 事实上, 通过不同的映射, 还可以得到不同的随机变量的和的概率密度公式. 如果取  $\alpha_1 = \beta_1 = \alpha_2 = 1, \beta_2 = -1$ , 则

$$f_{X+Y,X-Y}(u, v) = \frac{1}{2} f\left(\frac{u+v}{2}, \frac{u-v}{2}\right).$$

从而,

$$f_{X+Y}(u) = \frac{1}{2} \int_{-\infty}^{\infty} f\left(\frac{u+v}{2}, \frac{u-v}{2}\right) dv.$$

## 2.3 数字特征

### 2.3.1 数学期望

数学期望 (expectation) 是随机变量最重要的数字特征之一, 表示随机变量取值的概率加权平均. 对于离散型随机变量, 期望为

$$EX = \sum_n x_n P(X = x_n).$$

对于连续型随机变量, 期望为

$$EX = \int_{-\infty}^{\infty} xf(x)dx,$$

其中  $f(x)$  是随机变量  $X$  的概率密度函数. 期望又称为一阶矩, 具有以下性质:

- 随机变量的函数的期望计算公式为:  $Eg(X) = \sum_x g(x)P(X=x)$  或  $Eg(X) = \int_{-\infty}^{\infty} g(x)f(x)dx$
- 线性性质:  $E(\alpha X + \beta Y) = \alpha EX + \beta EY$
- 若  $X, Y$  相互独立, 则  $E(XY)$  存在, 且  $E(XY) = EXEY$

在讨论多个随机变量时, 可进一步引入条件期望的概念. 已知  $Y = y$  条件下  $X$  的条件期望记为  $E(X|Y=y)$ , 对于离散型随机变量, 有

$$E(X|Y=y) = \sum_x xP(X=x|Y=y).$$

对于连续型随机变量, 有

$$E(X|Y=y) = \int_{-\infty}^{+\infty} xf_{X|Y}(x|y)dx.$$

每给定一个  $Y$  的取值  $y$ , 可以计算如上定义的条件期望. 不同的  $y$  值映射到不同的期望值, 于是可以定义函数  $g(y) = E(X|Y=y)$ . 那么  $E(X|Y) = g(Y)$  是关于  $Y$  的函数, 称为随机变量  $X$  关于  $Y$  的条件期望, 记为  $E(X|Y)$ . 条件期望有如下重要结论:

**定理2.2. (全期望公式)** 设  $(X, Y)$  是二维随机变量,  $EX$  存在, 则

$$E(E(X|Y)) = EX.$$

**证明:** 对于离散型随机变量, 有

$$\begin{aligned} E(E(X|Y)) &= \sum_y E(X|Y=y)P(Y=y) \\ &= \sum_y \sum_x xP(X=x|Y=y)P(Y=y) \\ &= \sum_x x \sum_y P(X=x, Y=y) \\ &= \sum_x xP(X=x) = EX. \end{aligned}$$

对于连续型随机变量, 有

$$\begin{aligned}
 E(E(X|Y)) &= \int_{-\infty}^{\infty} E(X|Y=y)f_Y(y)dy \\
 &= \int_{-\infty}^{\infty} \left( \int_{-\infty}^{\infty} xf_{X|Y}(x|y)dx \right) f_Y(y)dy \\
 &= \int_{-\infty}^{\infty} \left( \int_{-\infty}^{\infty} xf_{X|Y}(x|y)f_Y(y)dy \right) dx \\
 &= \int_{-\infty}^{\infty} x \left( \int_{-\infty}^{\infty} f_{XY}(x,y)dy \right) dx \\
 &= \int_{-\infty}^{\infty} xf_X(x)dx = EX.
 \end{aligned}$$

对于混合型随机变量 ( $X$  为离散型、 $Y$  为连续型, 或者反过来), 可以类似的证明.  $\square$

全期望公式类似于全概率公式, 将期望的运算分解为若干子问题, 每个子问题是在给定条件下缩小的样本空间上的条件期望运算, 从而分解简化问题. 下面给出一个应用的例子.

**例2.3. (马格达囚徒问题)** 从 A 地出发有 3 条路, 沿其中一条路走  $t_1$  小时可到 B 地, 沿另一条路走  $t_2$  小时可回到 A 地, 沿第三条路走  $t_3$  小时可回到 A 地. 从 A 地出发等可能地选择这三条道路中的一条, 若走向 A 地, 仍等可能地任选一条. 问到达 B 地平均用多少时间?

**解:** 令  $X$  为从 A 地出发走到 B 地时间,  $Y$  为第一次所选择的路, 则  $Y \in \{1, 2, 3\}$ . 由全期望公式, 有

$$EX = E(E(X|Y)). \quad (2.3. 1)$$

易得

$$\begin{aligned}
 E(X|Y=1) &= t_1, \\
 E(X|Y=2) &= t_2 + EX, \\
 E(X|Y=3) &= t_3 + EX.
 \end{aligned}$$

代入 (2.3. 1) 得

$$EX = \frac{1}{3}t_1 + \frac{1}{3}(t_2 + EX) + \frac{1}{3}(t_3 + EX).$$

解得

$$EX = t_1 + t_2 + t_3.$$

$\square$



### 2.3.2 方差

如果二阶矩  $EX^2$  存在, 则  $X$  的方差定义为

$$DX = E[(X - EX)^2].$$

方差具有以下常用性质:

- $DX = EX^2 - (EX)^2$ . 利用此式往往可以简化方差的计算.
- $D(\alpha X) = \alpha^2 DX$ , 即常数与方差运算交换时, 需要进行平方.
- 若  $X_1, \dots, X_n$  相互独立, 则

$$D(X_1 + \dots + X_n) = DX_1 + \dots + DX_n.$$

### 2.3.3 协方差

协方差用于表征两个随机变量之间的相关性. 两个随机变量  $X, Y$  的协方差定义为

$$\text{Cov}(X, Y) = E[(X - EX)(Y - EY)].$$

协方差具有以下常用性质:

- $\text{Cov}(X, Y) = E(XY) - EXEY$ .
- 对称双线性性, 即

$$\text{Cov}(X, Y) = \text{Cov}(Y, X),$$

$$\text{Cov}(\alpha X + \beta Y, Z) = \alpha \text{Cov}(X, Z) + \beta \text{Cov}(Y, Z).$$

由于具有线性性质, 用协方差表征相关性时, 会附带有随机变量自身的绝对大小的信息. 可以将其归一化, 得到**相关系数**:

$$r_{X,Y} = \frac{\text{Cov}(X, Y)}{\sqrt{DXDY}}.$$

可以证明,  $-1 \leq r_{X,Y} \leq 1$ .

对于多维随机变量, 可以定义**协方差矩阵**. 设  $\mathbf{X}$  为  $n$  维随机向量,  $\mathbf{Y}$  为  $m$  维随机向量,  $\mathbf{X}$  和  $\mathbf{Y}$  的协方差矩阵为

$$\begin{aligned} \Sigma_{\mathbf{X}\mathbf{Y}} &= \text{Cov}(\mathbf{X}, \mathbf{Y}) = E((\mathbf{X} - E\mathbf{X})(\mathbf{Y} - E\mathbf{Y})^T) \\ &= (\text{Cov}(X_i, Y_j))_{1 \leq i \leq n, 1 \leq j \leq m} = \begin{pmatrix} \text{Cov}(X_1, Y_1) & \cdots & \text{Cov}(X_1, Y_m) \\ \vdots & \ddots & \vdots \\ \text{Cov}(X_n, Y_1) & \cdots & \text{Cov}(X_n, Y_m) \end{pmatrix}. \end{aligned}$$

下面的定理列举了协方差矩阵的主要性质.

**定理2.3.** 设  $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$  为任意多维随机向量, 则

- (1)  $\mathbf{X}$  的自协方差矩阵  $\Sigma_{\mathbf{X}} = \text{Cov}(\mathbf{X}, \mathbf{X})$  是对称非负定的;
- (2)  $\Sigma_{\mathbf{XY}} = \Sigma_{\mathbf{YX}}^T$ ;
- (3) 对给定矩阵  $\mathbf{A}, \mathbf{B}$ , 有

$$\text{Cov}(\mathbf{AX} + \mathbf{BY}, \mathbf{Z}) = \mathbf{A}\Sigma_{\mathbf{XZ}} + \mathbf{B}\Sigma_{\mathbf{YZ}},$$

$$\text{Cov}(\mathbf{Z}, \mathbf{AX} + \mathbf{BY}) = \Sigma_{\mathbf{ZX}}\mathbf{A}^T + \Sigma_{\mathbf{ZY}}\mathbf{B}^T.$$

**证明:** (1) 协方差矩阵的对称性可由协方差的对称性直接得到. 对于非负定性, 即对  $\forall \mathbf{v} \in \mathbb{R}^n$ , 都有  $\mathbf{v}^T \Sigma_{\mathbf{X}} \mathbf{v} \geq 0$ . 事实上, 令  $E\mathbf{X} = (EX_1, \dots, EX_n)^T$ , 有

$$\begin{aligned} \mathbf{v}^T \Sigma_{\mathbf{X}} \mathbf{v} &= \mathbf{v}^T \begin{pmatrix} \text{Cov}(X_1, X_1) & \cdots & \text{Cov}(X_1, X_n) \\ \vdots & \ddots & \vdots \\ \text{Cov}(X_n, X_1) & \cdots & \text{Cov}(X_n, X_n) \end{pmatrix} \mathbf{v} \\ &= \mathbf{v}^T E \begin{pmatrix} (X_1 - EX_1)^2 & \cdots & (X_1 - EX_1)(X_n - EX_n) \\ \vdots & \ddots & \vdots \\ (X_n - EX_n)(X_1 - EX_1) & \cdots & (X_n - EX_n)^2 \end{pmatrix} \mathbf{v} \\ &= \mathbf{v}^T E \left( \begin{pmatrix} X_1 - EX_1 \\ \vdots \\ X_n - EX_n \end{pmatrix} (X_1 - EX_1, \dots, X_n - EX_n) \right) \mathbf{v} \\ &= E(\mathbf{v}^T (\mathbf{X} - E\mathbf{X})(\mathbf{X} - E\mathbf{X})^T \mathbf{v}) \\ &= E(\mathbf{v}^T (\mathbf{X} - E\mathbf{X}))^2 \geq 0, \end{aligned}$$

其中, 第二个等号将期望运算提出来放在矩阵前面, 是协方差矩阵的等价写法, 最后一个等号是因为  $\mathbf{v}^T (\mathbf{X} - E\mathbf{X}) = (\mathbf{X} - E\mathbf{X})^T \mathbf{v}$  为标量.

(2) 由定义可得

$$\Sigma_{\mathbf{XY}} = E((\mathbf{X} - E\mathbf{X})(\mathbf{Y} - E\mathbf{Y})^T) = E((\mathbf{Y} - E\mathbf{Y})(\mathbf{X} - E\mathbf{X})^T)^T = \Sigma_{\mathbf{YX}}.$$

(3) 这两条性质可由协方差的线性性质以及第 (2) 条性质直接得到.  $\square$

## 2.4 特征函数

特征函数是研究随机变量的重要工具，首先给出特征函数的定义.

**定义2.1.** 对任意随机变量  $X$ ，它的特征函数为

$$\phi_X(\omega) = E(e^{i\omega X}),$$

其中,  $\omega \in \mathbb{R}, i^2 = -1$ .

特征函数可以看作是随机变量  $X$  的函数的期望, 由于  $|e^{i\omega X}| = 1$ , 特征函数一定存在.

**定理2.4.** 任意随机变量  $X$  的特征函数具有以下性质:

- (1)  $\phi_{a+bX}(\omega) = e^{ja\omega} \phi_X(b\omega)$ ;
- (2) 若  $X_1, \dots, X_n$  相互独立, 则  $\phi_{X_1+\dots+X_n}(\omega) = \phi_{X_1}(\omega) \cdots \phi_{X_n}(\omega)$ ;
- (3)  $E[X^n] = \frac{\phi_X^{(n)}(0)}{i^n}$ ;
- (4) 概率分布与特征函数一一对应.

**证明:** (1)  $\phi_{a+bX}(\omega) = E(e^{i\omega(a+bX)}) = e^{i\omega a} E(e^{i\omega bX}) = e^{ja\omega} \phi_X(b\omega)$ .

(2)

$$\begin{aligned}\phi_{X_1+\dots+X_n}(\omega) &= E(e^{i\omega(X_1+\dots+X_n)}) \\ &= E(e^{i\omega X_1} \cdots e^{i\omega X_n}) \\ &= E(e^{i\omega X_1}) \cdots E(e^{i\omega X_n}) = \phi_{X_1}(\omega) \cdots \phi_{X_n}(\omega)\end{aligned}$$

(3)  $\phi_X^{(n)}(\omega)|_{\omega=0} = E((iX)^n e^{i\omega X})|_{\omega=0} = i^n E X^n$ , 于是得证.

(4) 利用特征函数与傅里叶变换的联系可以证明. 以连续型随机变量为例, 根据定义, 有

$$\phi_X(\omega) = \int_{-\infty}^{\infty} f_X(x) e^{i\omega x} dx.$$

所以, 特征函数  $\phi_X(\omega)$  与概率密度函数  $f_X(x)$  是傅里叶变换对的关系. 根据傅里叶变换对的唯一性可知, 特征函数与概率密度函数一一对应, 即也概率分布一一对应.  $\square$

**例2.4.** (离散型) 计算二项分布、几何分布、以及泊松分布的特征函数.

解：二项分布的分布律为

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n.$$

则特征函数为

$$\begin{aligned}\phi_X(\omega) &= \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} e^{i\omega k} \\ &= \sum_{k=0}^n \binom{n}{k} (pe^{i\omega})^k (1-p)^{n-k} \\ &= (1-p + pe^{i\omega})^n = (pe^{i\omega} + q)^n.\end{aligned}$$

其中,  $q = 1 - p$ .

几何分布的分布律为

$$P(X = k) = (1-p)^{k-1} p, \quad k = 1, 2, \dots$$

特征函数为

$$\begin{aligned}\phi_X(\omega) &= \sum_{k=1}^{\infty} (1-p)^{k-1} p e^{i\omega k} \\ &= \sum_{k=1}^{\infty} ((1-p)e^{i\omega})^{k-1} p e^{i\omega} \\ &= \frac{pe^{i\omega}}{1 - qe^{i\omega}},\end{aligned}$$

其中,  $q = 1 - p$ .

泊松分布的分布律为

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

特征函数为

$$\begin{aligned}\phi_X(\omega) &= \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} e^{i\omega k} \\ &= \sum_{k=0}^{\infty} \frac{(\lambda e^{i\omega})^k}{k!} e^{-\lambda} \\ &= e^{\lambda(e^{i\omega} - 1)}.\end{aligned}$$

□

**例2.5.** (连续型) 计算均匀分布、指数分布、以及正态分布的特征函数.

**解:** 均匀分布的概率密度为

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b, \\ 0, & \text{其他.} \end{cases}$$

则特征函数为

$$\phi_X(\omega) = \int_a^b \frac{1}{b-a} e^{i\omega x} dx = \frac{e^{i\omega b} - e^{i\omega a}}{e^{i\omega}(b-a)}.$$

指数分布的概率密度为

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & \text{其他.} \end{cases}$$

则特征函数为

$$\phi_X(\omega) = \int_0^\infty \lambda e^{-\lambda x} e^{i\omega x} dx = \frac{\lambda}{\lambda - i\omega}.$$

正态分布的概率密度为

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

则特征函数为

$$\begin{aligned} \phi_X(\omega) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} e^{i\omega x} dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu-i\omega\sigma^2)^2 + \mu^2 - (\mu+i\omega\sigma^2)^2}{2\sigma^2}} dx \\ &= e^{i\omega\mu - \frac{\sigma^2\omega^2}{2}}. \end{aligned}$$

□

对随机向量  $\mathbf{X} = (X_1, \dots, X_n)^T$  和任意  $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^T \in \mathbb{R}^n$ ,

$$\phi_{\mathbf{X}}(\boldsymbol{\omega}) = E(e^{i\boldsymbol{\omega}^T \mathbf{X}}) = E(e^{i(\omega_1 X_1 + \dots + \omega_n X_n)})$$

为  $\mathbf{X}$  的特征函数 ( $X_1, \dots, X_n$  的联合特征函数). 可以验证, 随机变量  $X_1, \dots, X_n$  相互独立当且仅当

$$\phi_{\mathbf{X}}(\boldsymbol{\omega}) = \prod_{k=1}^n \phi_{X_k}(\omega_k).$$

当随机变量取非负整数值时, 可定义**概率母函数** (也叫**生成函数**):

$$G_X(z) = E(z^X) = \sum_{n=0}^{+\infty} P(X = n)z^n.$$

可以看到, 离散型随机变量的分布律与其母函数是一对  $z$  变换, 特征函数是概率母函数自变量取  $z = e^{j\omega}$  的特例, 也就是傅里叶变换.

## 2.5 大数定律和中心极限定理

大数定律表明, 用一个随机变量的一组独立观测值的算术平均作为其数学期望 (概率意义上的平均值) 是合理的, 即使不知道这个随机变量的具体概率分布. 这个结论可以得到严格证明, 是一个严格的数学定理. 而事实上, 大数定理结论是符合我们的日常经验的, 在严格证明之前已被广泛地使用着, 因此仍称为“定律”. 为了表述大数定律, 首先引入依概率收敛的概念.

**定义2.2.** 设  $\{X_n\}$  为一随机变量序列, 若对任意  $\epsilon > 0$ , 有

$$\lim_{n \rightarrow \infty} P(|X_n - X| < \epsilon) = 1,$$

则称  $\{X_n\}$  **依概率收敛** 于  $X$ , 记为

$$X_n \xrightarrow{P} X$$

**定理2.5.** 设  $X_1, \dots, X_n, \dots$  是一列独立同分布的随机变量, 存在数学期望 (记为  $\mu$ ). 则

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow{P} \mu, \quad n \rightarrow \infty$$

即对任意  $\epsilon > 0$ ,

$$\lim_{n \rightarrow +\infty} P\left(\left|\frac{X_1 + \dots + X_n}{n} - \mu\right| < \epsilon\right) = 1.$$

上述用依概率收敛表述的大数定律, 称为**弱大数定律**. 事实上大数定律还有更强的结论, 称为**强大数定律**, 其表述需要引入以概率 1 收敛的概念.

**定义2.3.** 设  $\{X_n\}$  为一随机变量序列, 若有

$$P(\omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)) = 1,$$

则称  $\{X_n\}$  **几乎处处收敛** (亦称为**以概率1收敛**) 于  $X$ , 记为

$$X_n \xrightarrow{a.s.} X.$$

**定理2.6.** 设  $X_1, \dots, X_n, \dots$  是一列独立同分布的随机变量, 存在数学期望 (记为  $\mu$ )。则

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow{a.s.} \mu, \quad n \rightarrow \infty.$$

进一步的, 下面的中心极限定理刻画了大量随机因素综合影响下的分布特性.

**定理2.7.** 设  $X_1, \dots, X_n, \dots$  是一列独立同分布的随机变量, 存在数学期望 (记为  $\mu$ ) 和有限的方差 (记为  $\sigma^2$ )。则

$$\lim_{n \rightarrow +\infty} \sup_{x \in \mathbb{R}} \left| P \left( \frac{X_1 + \dots + X_n - n\mu}{\sqrt{n}\sigma} \leq x \right) - \Phi(x) \right| = 0,$$

其中  $\Phi(x)$  标准正态 (Gauss) 分布的概率分布函数.

**证明:** 作为一个典型应用, 中心极限定理可以利用特征函数加以证明. 不妨设  $\mu = 0, \sigma^2 = 1$ . 实际上, 对于一般的情形, 考察规范化随机变量序列  $Y_k = \frac{X_k - \mu}{\sigma}$  即可. 令

$$Z_n = \frac{X_1 + \dots + X_n}{\sqrt{n}},$$

考察  $Z_n$  的特征函数为

$$\begin{aligned} \phi_{Z_n}(\omega) &= E(\exp(j\omega Z_n)) \\ &= E \left( \exp \left( j\omega \frac{X_1 + \dots + X_n}{\sqrt{n}} \right) \right) \\ &= E \left( \prod_{k=1}^n \exp \left( j\omega \frac{X_k}{\sqrt{n}} \right) \right) \\ &= \prod_{k=1}^n E \left( \exp \left( j \frac{\omega}{\sqrt{n}} X_k \right) \right) \\ &= \left( \phi_{X_1} \left( \frac{\omega}{\sqrt{n}} \right) \right)^n. \end{aligned} \tag{2.5. 1}$$

对于给定的  $\omega$ , 当  $n \rightarrow \infty$  时,  $\frac{\omega}{\sqrt{n}} \rightarrow 0$ , 于是有

$$\begin{aligned} \phi_{X_1} \left( \frac{\omega}{\sqrt{n}} \right) &= 1 + \phi'_{X_1}(0) \frac{\omega}{\sqrt{n}} + \frac{1}{2} \phi''_{X_1}(0) \left( \frac{\omega}{\sqrt{n}} \right)^2 + o \left( \frac{1}{n} \right) \\ &= 1 - \frac{\omega^2}{2n} + o \left( \frac{1}{n} \right), \end{aligned}$$

代入式 (2.5. 1) 并令  $n \rightarrow \infty$ , 得到

$$\lim_{n \rightarrow \infty} \phi_{Z_n}(\omega) = \lim_{n \rightarrow \infty} \left( 1 - \frac{\omega^2}{2n} + o \left( \frac{1}{n} \right) \right)^n = \exp \left( -\frac{\omega^2}{2} \right).$$

这正是标准正态分布  $N(0, 1)$  的特征函数. □

## 第 3 章 离散时间 Markov 链

从随机过程的定义可知, 随机变量之间的关联是随机过程的重要研究对象, 而这种关联与时间关系紧密. 本章考虑离散时间随机过程, 引入  $X_k$  与最接近的过去  $X_{k-1}$  之间的依赖, 即离散时间 Markov 链. 一方面, 这种依赖关系易于分析, 一般具有良好的性质 (Markov 性); 另一方面, 对最接近的过去的依赖, 是对客观世界时间相关性的良好近似.

### 3.1 定义与例子

#### 3.1.1 离散时间 Markov 链的定义

**定义 3.1.** 设具有可数样本空间  $\mathcal{E}$  的随机过程  $X_n$  满足  $\forall n, k_1 < \cdots < k_n < k_{n+1}$ ,

$$\begin{aligned} &P(X_{k_{n+1}} = i_{n+1} | X_{k_n} = i_n, X_{k_{n-1}} = i_{n-1}, \cdots, X_{k_1} = i_1) \\ &= P(X_{k_{n+1}} = i_{n+1} | X_{k_n} = i_n), \end{aligned} \quad (3.1.1)$$

则称其为**离散时间 Markov 链**.

离散时间 Markov 链未来任意时刻的分布仅仅依赖于当前时刻的状态. 换句话说, 在给定当前时刻的状态取值的条件下, 未来的事件与过去的事件相互 (条件) 独立. 可以用下面的方式直观描述: 设  $A, B, C$  依次表示“过去”、“现在”、“将来”, 则

$$P(C|BA) = P(C|B) \Leftrightarrow P(CA|B) = P(C|B)P(A|B).$$

**定义 3.2.** 离散时间 Markov 链  $X_n$  的  $n$  步转移概率定义为

$$P_{ij}(k, k+n) = P(X_{k+n} = j | X_k = i), \quad \forall i, j \in \mathcal{E}, k = 0, 1, \cdots.$$

在离散时间 Markov 链中, 有一类非常重要的随机过程, 其转移概率不随时间变化, 具有某种平稳性. 我们给出以下定义.

**定义 3.3.** 若 Markov 链的  $n$  步转移概率不依赖于时间  $k$ , 即

$$P_{ij}(k, k+n) = P(X_{k+n} = j | X_k = i) = p_{ij}^{(n)}, \quad \forall k,$$

则称该链为**齐次 Markov 链**.

在本章中, 如无特别说明, 默认所讨论的 Markov 链均是齐次的. 特别的, 当  $n = 1$  时, 得到**一步转移概率**, 简记为

$$P(X_{k+1} = j | X_k = i) = p_{ij}.$$



在 Markov 链中, 一步转移概率起到至关重要的作用. 事实上, 对于  $\forall m > 0$ ,  $i_0, \dots, i_m \in \mathcal{E}$ , 有

$$\begin{aligned} & P(X_0 = i_0, X_1 = i_1, \dots, X_m = i_m) \\ &= P(X_0 = i_0)P(X_1 = i_1|X_0 = i_0) \cdots P(X_m = i_m|X_{m-1} = i_{m-1}, \dots, X_0 = i_0) \\ &= P(X_0 = i_0)P(X_1 = i_1|X_0 = i_0) \cdots P(X_m = i_m|X_{m-1} = i_{m-1}) \\ &= p_0(i_0)p_{i_0 i_1} \cdots p_{i_{m-1} i_m}, \end{aligned}$$

即任意有限维联合分布由初始分布  $p_0(i_0) = P(X_0 = i_0)$  和一步转移概率完全决定. 事实上, 由下面的 C-K 方程可知, 任意  $n$  步转移概率也由一步转移概率完全决定.

**定理3.1.** 齐次 Markov 链的状态转移概率满足 Chapman-Kolmogorov 方程

$$p_{ij}^{(n+m)} = \sum_{k \in \mathcal{E}} p_{ik}^{(n)} p_{kj}^{(m)}.$$

矩阵形式为

$$\mathbf{P}^{(n+m)} = \mathbf{P}^{(n)} \mathbf{P}^{(m)}.$$

特别的,  $n$  步转移概率矩阵可由一步转移概率矩阵求得

$$\mathbf{P}^{(n)} = \mathbf{P}^n, \quad (3.1. 2)$$

其中,  $\mathbf{P}^{(n)} = (p_{ij}^{(n)})_{i,j \in \mathcal{E}}$ ,  $\mathbf{P} = (p_{ij})_{i,j \in \mathcal{E}}$  分别表示  $n$  步转移概率矩阵和一步转移概率矩阵.

**证明:**

$$\begin{aligned} p_{ij}^{(n+m)} &= P(X_{l+n+m} = j | X_l = i) \\ &= \sum_{k \in \mathcal{E}} P(X_{l+n} = k | X_l = i) P(X_{l+n+m} = j | X_{l+n} = k, X_l = i) \\ &= \sum_{k \in \mathcal{E}} P(X_{l+n} = k | X_l = i) P(X_{l+n+m} = j | X_{l+n} = k) \\ &= \sum_{k \in \mathcal{E}} p_{ik}^{(n)} p_{kj}^{(m)}, \end{aligned}$$

其中第二个等号是以第  $n$  步所处的所有可能的状态对  $n + m$  步转移概率做全概率展开, 第三个等号成立是由于该链具有齐次性.

矩阵形式可由上式直接得到. 令  $m = 1$  得到

$$\mathbf{P}^{(n+1)} = \mathbf{P}^{(n)} \mathbf{P}.$$

利用上式迭代计算可得 (3.1. 2). □

可以说, 一步转移概率矩阵包含了 Markov 链的全部信息, 在 Markov 链的研究中起到关键的作用. 因此, 通常将其简称为**转移概率矩阵**.

### 3.1.2 Markov 链的表示方法

Markov 链通常有三种表示方法. 首先, Markov 链可以用转移概率矩阵表示. 具体来说, 将转移概率排列成矩阵形式, 其中行表示当前时刻的状态, 列表示下一时刻的状态. 若状态空间为  $\mathcal{E} = \{0, 1, \dots, n\}$ , 则转移概率矩阵可表示为

$$\mathbf{P} = (p_{ij})_{i,j \in \mathcal{E}} = \begin{pmatrix} p_{00} & \cdots & p_{0n} \\ \vdots & \ddots & \vdots \\ p_{n0} & \cdots & p_{nn} \end{pmatrix}.$$

第二种表示方法是**状态转移图**, 即以状态为节点, 以状态转移为有向线段连成的有向图. 我们将在下面的例子中加以讨论.

第三种表示方法是用递推函数, 即

$$X_{n+1} = f(X_n, Z_{n+1}), \quad (3.1. 3)$$

其中  $f$  是二元确定性函数, 随机变量序列  $Z_1, Z_2, \dots$  满足下列条件之一

- $Z_1, Z_2, \dots$  独立同分布, 且与初始状态  $X_0$  独立;
- $Z_1, Z_2, \dots$  同分布, 且在给定  $X_n$  的条件下,  $Z_{n+1}$  与  $X_0, Z_1, \dots, Z_n$  独立.

接下来验证满足第一个条件的随机过程  $X_n$  是齐次 Markov 链, 第二个条件请读者自行验证. 由于任意  $n$  步转移概率由一步转移概率决定, 可以证明, 要验证定义式 (3.1. 1), 只需验证

$$P(X_{n+1} = i_{n+1} | X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_1 = i_1) = P(X_{n+1} = i_{n+1} | X_n = i_n)$$

即可. 一方面,

$$\begin{aligned} & P(X_{n+1} = i_{n+1} | X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) \\ &= P(f(X_n, Z_{n+1}) = i_{n+1} | X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) \\ &= P(f(i_n, Z_{n+1}) = i_{n+1}), \end{aligned}$$

其中, 第二个等号成立是因为根据递推关系式(3.1. 3),  $X_k$  可以表为  $X_0$  和  $Z_1, \dots, Z_k$  的函数

$$X_k = g_k(X_0, Z_1, \dots, Z_k).$$

于是条件  $\{X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0\}$  可以由  $\{X_0, Z_1, \dots, Z_k\}$  表示, 从而与  $Z_{n+1}$  独立. 类似的,

$$P(X_{n+1} = i_{n+1} | X_n = i_n) = P(f(i_n, Z_{n+1}) = i_{n+1}).$$

所以  $X_n$  是一个 Markov 链. 另外, 由于序列  $Z_n$  独立同分布, 所以

$$\begin{aligned} P(X_{n+1} = j | X_n = i) &= P(f(i, Z_{n+1}) = j) \\ &= P(f(i, Z_{k+1}) = j) \\ &= P(X_{k+1} = j | X_k = i), \quad \forall n, k. \end{aligned}$$

即一步转移概率不依赖于时间  $n$ , 因此是齐次的. 如果  $Z_n$  不同分布, 则是一个非齐次 Markov 链.

递推表示法也是验证一个随机过程是 Markov 链的有效方法. 只要写出形如 (3.1. 3) 的迭代方程, 且满足一定的条件, 就可以说明该过程是 Markov 链.

### 3.1.3 离散时间 Markov 链举例

**例3.1.** (简单随机徘徊) 质点在直线上左右运动, 如果某一时刻处于位置  $i$ , 下一时刻以概率  $p$  向右运动到  $i + 1$ , 以概率  $q = 1 - p$  向左运动到  $i - 1$ , 则质点的位置构成随机过程  $X_n$ .

- (1) 验证该过程为 Markov 链, 并给出 Markov 链描述.
- (2) 考察带一个反射壁的简单随机徘徊. 在  $-\frac{1}{2}$  位置处放置一个反射壁, 质点初始处于壁右边, 其从 0 位置向左运动时, 会被反射回 0 位置.
- (3) 考察带两个反射壁的简单随机徘徊. 在  $-\frac{1}{2}$  和  $N + \frac{1}{2}$  位置处各放置一个反射壁, 质点初始处于两壁之间, 碰壁后反射回原位置.
- (4) 考察带一个吸收壁的简单随机徘徊. 在 0 位置处放置一个吸收壁, 质点初始处于壁右边, 当它运动到 0 位置后, 将一直停留在该位置.
- (5) 考察带两个吸收壁的简单随机徘徊. 在 0 和  $N$  位置处各放置一个吸收壁, 质点初始处于两壁之间, 到达任一吸收壁后一直停留在该位置.

**解:** (1) 令  $Z_1, Z_2, \dots$  为独立同分布的随机变量序列, 满足

$$\begin{aligned} P(Z_k = 1) &= p, \\ P(Z_k = -1) &= q. \end{aligned}$$

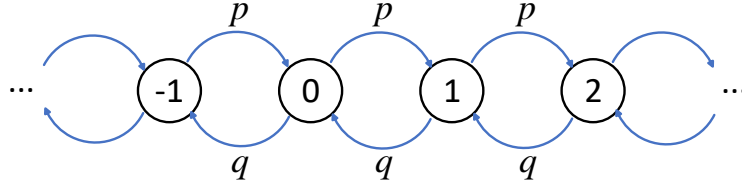


图 3.1 简单随机徘徊的状态转移图

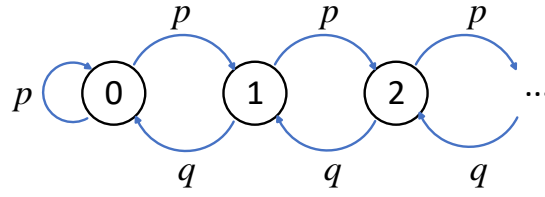


图 3.2 带一个反射壁的简单随机徘徊

则

$$X_{n+1} = X_n + Z_{n+1}.$$

根据递推表示法知,  $X_n$  是一个齐次 Markov 链, 其转移概率矩阵可形式化表为:

$$\mathbf{P} = (p_{ij})_{i,j \in \mathcal{E}} = \begin{pmatrix} \ddots & \ddots & \ddots & & \\ & q & 0 & p & \\ & & q & 0 & p \\ & & & \ddots & \ddots & \ddots \end{pmatrix}.$$

其状态转移图如图 3.1 所示.

(2) 放置一个反射壁后, 状态空间从全体整数变为  $\mathcal{E} = \{0, 1, 2, \dots\}$ , 可以用下面的递推式表示

$$X_{n+1} = \max\{0, X_n + Z_{n+1}\}.$$

因此,  $X_n$  仍是一个齐次 Markov 链, 其转移概率矩阵可形式化表为:

$$\mathbf{P} = \begin{pmatrix} q & p & & \\ q & 0 & p & \\ & \ddots & \ddots & \ddots \end{pmatrix}.$$

其状态转移图如图 3.2 所示.

(3) 具有两个反射壁的随机徘徊的状态空间为  $\mathcal{E} = \{0, 1, \dots, N\}$ , 迭代方程为

$$X_{n+1} = \min\{\max\{0, X_n + Z_{n+1}\}, N\}.$$

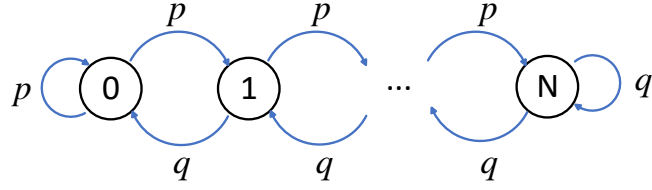


图 3.3 带两个反射壁的简单随机徘徊

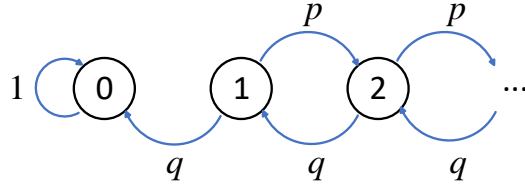


图 3.4 带一个吸收壁的简单随机徘徊

其转移概率矩阵可表为:

$$\mathbf{P} = \begin{pmatrix} q & p & & \\ q & 0 & p & \\ & \ddots & \ddots & \ddots \\ & & q & 0 & p \\ & & & q & p \end{pmatrix}.$$

其状态转移图如图 3.3 所示.

(4) 放置一个吸收壁后, 状态空间从全体整数变为  $\mathcal{E} = \{0, 1, 2, \dots\}$ , 递推关系为

$$X_{n+1} = \begin{cases} X_n + Z_{n+1}, & X_n > 0, \\ 0, & X_n = 0. \end{cases}$$

其转移概率矩阵可形式化表为:

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & & \\ q & 0 & p & \\ & \ddots & \ddots & \ddots \end{pmatrix}.$$

其状态转移图如图 3.4 所示.

(5) 具有两个吸收壁的随机徘徊的状态空间为  $\mathcal{E} = \{0, 1, \dots, N\}$ , 迭代方程为

$$X_{n+1} = \begin{cases} X_n + Z_{n+1}, & 0 < X_n < N, \\ 0, & X_n = 0, \\ N, & X_n = N. \end{cases}$$

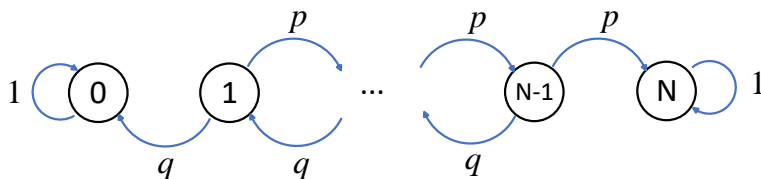


图 3.5 带两个反射壁简单随机徘徊

其转移概率矩阵可表为:

$$P = \begin{pmatrix} 1 & 0 & & & \\ q & 0 & p & & \\ & \ddots & \ddots & \ddots & \\ & & q & 0 & p \\ & & & 0 & 1 \end{pmatrix}.$$

其状态转移图如图 3.5 所示. □

**例3.2. (Ehrenfest 模型)** 设一个坛子里装有  $M$  个球, 或者是红色球, 或者是黑色球. 从坛中随机地取出一个球并换入一个颜色与之不同的球, 称为一次操作. 经过  $n$  次操作后, 以坛中所装的黑球数目作为状态, 构成一个随机过程. 验证该过程为 Markov 链, 并给出 Markov 链描述.

**解:** 设  $X_n \in \{0, 1, \dots, M\}$  为  $n$  次操作后坛中黑球的数目, 引入随机变量  $Z_n \in \{1, -1\}$  表示坛中黑球数目的增量, 其概率分布依赖于  $X_n$ , 即

$$P(Z_{n+1} = 1 | X_n = i) = \frac{M-i}{M},$$

$$P(Z_{n+1} = -1 | X_n = i) = \frac{i}{M}.$$

于是有迭代方程

$$X_{n+1} = X_n + Z_{n+1}.$$

由于给定  $X_n$  时  $Z_{n+1}$  与历史状态无关, 则该过程为齐次 Markov 链. 转移概率矩阵为

$$P = \begin{pmatrix} 0 & 1 & & & \\ \frac{1}{M} & 0 & \frac{M-1}{M} & & \\ & \ddots & \ddots & \ddots & \\ & & \frac{M-1}{M} & 0 & \frac{1}{M} \\ & & & 1 & 0 \end{pmatrix}.$$

其状态转移图如图 3.6 所示. □

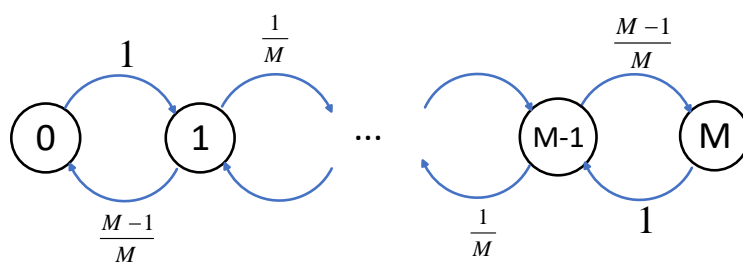


图 3.6 Ehrenfest模型的状态转移图

**例3.3. (设备维修问题)** 设第  $n$  天有  $Z_n$  台损坏的设备送到维修站接收维修, 且  $Z_1, Z_2, \dots$  是独立同分布的随机变量, 满足  $P(Z_n = k) = a_k, k = 0, 1, \dots$ . 维修站每天可以修复一台设备, 令  $X_n$  表示第  $n$  天维修站中待维修的设备数. 验证该过程为 Markov 链, 并给出  $X_n$  的 Markov 链描述.

**解:** 根据题意, 可以得到以下迭代方程

$$X_{n+1} = \max\{0, X_n - 1\} + Z_{n+1}.$$

其转移概率矩阵为

$$P = \begin{pmatrix} a_0 & a_1 & a_2 & \cdots \\ a_0 & a_1 & a_2 & \cdots \\ & a_0 & a_1 & \cdots \\ & & \ddots & \ddots \end{pmatrix}.$$

由于状态转移边数较多, 图略. □

根据 C-K 方程, 已知 Markov 链的 (一步) 转移概率矩阵, 便可通过代数运算, 求得任意  $n$  步转移概率矩阵. 看下面的例子.

**例3.4.** 天气预报中, 考虑晴天和阴雨天两状态的简单情形, 设

$$P(\text{明天阴雨}|\text{今天天晴}) = \alpha,$$

$$P(\text{明天天晴}|\text{今天阴雨}) = \beta,$$

计算  $n$  步转移概率.

**解:** 天气状态的转移概率矩阵为

$$P = \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix}.$$

根据 (3.1. 2), 要计算  $n$  步转移概率, 只需计算  $\mathbf{P}^n$ . 可利用相似变换将  $\mathbf{P}$  对角化, 再计算  $\mathbf{P}^n$ . 具体来说, 首先计算  $\mathbf{P}$  的特征值, 解方程

$$|\mathbf{P} - \lambda \mathbf{I}| = 0,$$

得到特征值为

$$\lambda_1 = 1, \quad \lambda_2 = 1 - \alpha - \beta.$$

再由

$$\mathbf{P}\mathbf{v} = \lambda\mathbf{v}$$

解得对应于两个特征值的特征向量分别为

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix} \alpha \\ -\beta \end{pmatrix}.$$

于是

$$\mathbf{P} = \mathbf{V} \begin{pmatrix} 1 & 0 \\ 0 & 1 - \alpha - \beta \end{pmatrix} \mathbf{V}^{-1},$$

其中,

$$\mathbf{V} = (\mathbf{v}_1 \mathbf{v}_2) = \begin{pmatrix} 1 & \alpha \\ 1 & -\beta \end{pmatrix}, \quad \mathbf{V}^{-1} = \frac{1}{\alpha + \beta} \begin{pmatrix} \beta & \alpha \\ 1 & -1 \end{pmatrix}.$$

所以有,

$$\begin{aligned} \mathbf{P}^{(n)} = \mathbf{P}^n &= \mathbf{V} \begin{pmatrix} 1 & 0 \\ 0 & (1 - \alpha - \beta)^n \end{pmatrix} \mathbf{V}^{-1} \\ &= \frac{1}{\alpha + \beta} \begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix} + \frac{(1 - \alpha - \beta)^n}{\alpha + \beta} \begin{pmatrix} \alpha & -\alpha \\ -\beta & \beta \end{pmatrix}. \end{aligned}$$

□

## 3.2 状态的分类与性质

Markov 链的性质很大程度上依赖于状态的性质, 而各个状态的性质往往有所区别. 由于状态数目可能较多, 甚至是可数的, 因此有必要将状态按性质进行分类研究.



### 3.2.1 可达与相通

**定义3.4.** 设  $i, j \in \mathcal{E}$  是 Markov 链中的两个状态, 如果  $\exists n > 0$ , 使得

$$p_{ij}^{(n)} > 0,$$

则称  $i$  可达  $j$  (accessible), 记作  $i \rightarrow j$ .

如果  $i$  不可达  $j$ , 即  $\forall n > 0, p_{ij}^{(n)} = 0$ , 记为  $i \nrightarrow j$ .

**定理3.2. (传递性):** 设  $i, j \in \mathcal{E}$  是 Markov 链中的两个状态, 若  $i \rightarrow j, j \rightarrow k$ , 则  $i \rightarrow k$ .

**证明:** 由定义知,  $\exists n > 0, m > 0$ , 使得

$$p_{ij}^{(n)} > 0, \quad p_{jk}^{(m)} > 0.$$

根据 C-K 方程及概率的非负性, 有

$$p_{ik}^{(n+m)} = \sum_{s \in \mathcal{E}} p_{is}^{(n)} p_{sk}^{(m)} \geq p_{ij}^{(n)} p_{jk}^{(m)} > 0.$$

如此, 我们找到了一个正数  $n + m$ , 使得  $p_{ik}^{(n+m)} > 0$ , 即  $i \rightarrow k$ . □

**定义3.5.** 设  $i, j \in \mathcal{E}$  是 Markov 链中的两个状态, 如果  $\exists n > 0, m > 0$ , 使得

$$p_{ij}^{(n)} > 0, \quad p_{ji}^{(m)} > 0,$$

则称  $i$  和  $j$  相通 (或互通, communicate), 记作  $i \leftrightarrow j$ .

不难验证, 相通具有以下性质:

- 自反性:  $i \leftrightarrow i$ ;
- 对称性: 如果  $i \leftrightarrow j$ , 那么  $j \leftrightarrow i$ ;
- 传递性: 如果  $i \leftrightarrow j, j \leftrightarrow k$ , 那么  $i \leftrightarrow k$ .

因此, 相通是一种等价关系, 可以将状态按是否相通分为多个等价类, 每个等价类具有相同的“类性质”.

**注3.1.** 我们约定相通具有自反性, 使其成为等价关系. 事实上, 由于  $p_{ii}^{(0)} = P(X_n = i | X_n = i) = 1$ , 约定状态  $i$  与自身相通是平凡的.

**定义3.6.** 对 Markov 链的状态集合  $\mathcal{S} \subseteq \mathcal{E}$ , 如果任意  $i \in \mathcal{S}, j \notin \mathcal{S}$ , 有  $i \nrightarrow j$ , 则称  $\mathcal{S}$  为闭集.

直观来说, 闭集是一种“进得去、出不来”的集合. 特别的, 对于单状态闭集  $\{i\}$ ,  $i$  称为吸收态.

显然,  $\emptyset$  和  $\mathcal{E}$  是闭集, 而且闭集本身也是一个 Markov 链. 对于一个 Markov 链, 可以找出很多闭集, 或者说存在多个“子链”. 其中有一类特殊的闭集, 它不再包含更小的子链.

**定义3.7.** 设 Markov 链的状态空间为  $\mathcal{E}$ , 如果闭集  $\mathcal{C} \subseteq \mathcal{E}$  没有闭的真子集, 即

$$\{S \subseteq \mathcal{C}, S \text{ 是闭集}\} \Leftrightarrow \{S = \emptyset \text{ 或者 } S = \mathcal{C}\},$$

则称  $\mathcal{C}$  不可约 (irreducible). 如果  $\mathcal{E}$  本身不可约, 则称该 Markov 链不可约.

**定理3.3.** 设  $\mathcal{C} \subseteq \mathcal{E}$  是 Markov 链的不可约闭子集, 则其中任意两个状态都相通.

**证明:**  $\forall i \in \mathcal{C}$ , 令  $\mathcal{S}_i = \{j \in \mathcal{C} : i \rightarrow j\}$  为从状态  $i$  可达的所有状态的集合, 利用反证法可以证明  $\mathcal{S}_i$  是闭集. 事实上,  $\forall j \in \mathcal{S}_i, k \notin \mathcal{S}_i$ , 如果  $j \rightarrow k$ , 由于  $i \rightarrow j$ , 则有  $i \rightarrow k$ . 从而有  $k \in \mathcal{S}_i$ , 与条件  $k \notin \mathcal{S}_i$  矛盾, 所以  $j \nrightarrow k$ , 因此  $\mathcal{S}_i$  是闭集. 又由于  $\mathcal{C}$  不可约, 于是  $\mathcal{S}_i = \mathcal{C}$ . 由  $i$  的任意性知,  $\mathcal{C}$  中任意两个状态都相通.  $\square$

**例3.5.** 考察以下 Markov 链的状态类型及性质

$$(1) \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{3} & \frac{2}{3} \end{pmatrix}, \quad (2) \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad (3) \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{4} & \frac{3}{4} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

**解:** 画出各链的状态转移图, 容易看出: 第一个链不可约, 所有状态都相通; 第二个链可约, 其中状态  $\{1, 2\}$  组成不可约的子链, 状态 4 是吸收态; 第三个链也可约, 其中状态  $\{1, 2\}$  与  $\{3, 4, 5\}$  分别组成不可约的子链.  $\square$

### 3.2.2 周期性

**定义3.8.** 状态  $i$  的周期  $d_i$  是集合  $\mathcal{T}_i = \{n : p_{ii}^{(n)} > 0\}$  的最大公约数, 即

$$d_i = \gcd\{n : p_{ii}^{(n)} > 0\}.$$

如果  $d_i = 1$ , 则称状态  $i$  为非周期态. 如果  $d_i > 1$ , 则称状态  $i$  为周期态 (periodic).

下面的定理表明, 周期性也是一种类性质.

**定理3.4.** 对于 Markov 链的两个状态  $i, j \in \mathcal{E}$ , 如果  $i \leftrightarrow j$ , 则  $d_i = d_j$ .

**证明：** 要证明  $d_i = d_j$ ，只需证明  $d_i, d_j$  相互整除.

由于  $i \leftrightarrow j$ ，即  $\exists n > 0, m > 0$ ，使得  $p_{ij}^{(n)} > 0, p_{ji}^{(m)} > 0$ . 于是，有

$$p_{jj}^{(m+n)} \geq p_{ji}^{(m)} p_{ij}^{(n)} > 0.$$

根据状态  $j$  的周期性知， $m+n$  被  $d_j$  整除. 另外，任取  $l \in \mathcal{T}_i$ ，即  $p_{ii}^{(l)} > 0$ ，则有

$$p_{jj}^{(m+l+n)} \geq p_{ji}^{(m)} p_{ii}^{(l)} p_{ij}^{(n)} > 0.$$

可知， $m+l+n$  被  $d_j$  整除，从而  $l$  被  $d_j$  整除. 由  $l$  的任意性知， $d_i$  被  $d_j$  整除.

同理可证  $d_j$  被  $d_i$  整除. 定理得证.  $\square$

**例3.6.** 两状态周期链的一步转移矩阵为

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

**例3.7.** 周期为  $d$  的不可约周期链的一步转移矩阵为

$$P = \begin{pmatrix} \mathbf{0} & \mathbf{A}_{12} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{A}_{23} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A}_{d-1,d} \\ \mathbf{A}_{d1} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix}.$$

**例3.8.** 一维无限制随机徘徊，每个状态周期均为 2.

### 3.2.3 常返性

除了基于转移概率所定义的相通、周期等性质之外，对于齐次 Markov 链，人们往往还关注长时间之后的统计特性，比如哪些状态会被频繁访问，访问频率是多大，等等. 为此引入常返性的概念. 首先给出首达概率的定义，并借此定义常返.

**定义3.9.** 设  $i, j \in E$ ，经  $n$  步从  $i$  到达  $j$  的**首达概率**为

$$f_{ij}^{(n)} = P(X_n = j, X_{n-1} \neq j, \cdots, X_1 \neq j | X_0 = i).$$

根据定义，首达概率显然满足  $\sum_{n=1}^{\infty} f_{ij}^{(n)} \leq 1$ .

**定义3.10.** 如果

$$f_{ii} = \sum_{n=1}^{\infty} f_{ii}^{(n)} = 1,$$

则称  $i$  为**常返态 (recurrent)**，否则称为**瞬时态**，也称**滑过态 (transient)**或**非常返态 (non-recurrent)**.

对于常返态, 还可以进一步讨论它的平均返回时间. 为此, 引入首达时间的概念.

**定义3.11.** 设  $i \in \mathcal{E}$  为 Markov 链的一个状态, 到达  $i$  的首达时间为

$$\tau_i = \inf\{n \geq 1 : X_n = i\},$$

如果  $\{n \geq 1 : X_n = i\} = \emptyset$ , 则定义  $\tau_i = \infty$ .

**定义3.12.** 如果平均首返时间

$$\mu_i = E(\tau_i | X_0 = i) = \sum_{n=1}^{\infty} n f_{ii}^{(n)} < \infty,$$

则称  $i$  正常返 (positive recurrent); 否则称之为零常返 (null recurrent).

**例3.9.** 根据定义, 考察以下 Markov 链各状态的常返性

$$(1) \begin{pmatrix} \frac{1}{3} & \frac{2}{3} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & 0 & \frac{2}{3} \end{pmatrix}, \quad (2) \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & \frac{2}{3} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix}.$$

**解:** (1) 根据状态转移图 3.7, 计算状态 1 的  $n$  步首返概率

$$\begin{aligned} f_{11}^{(1)} &= \frac{1}{3}, \\ f_{11}^{(2)} &= \frac{2}{3} \cdot \frac{1}{2} = \frac{1}{3}, \\ f_{11}^{(3)} &= \frac{2}{3} \cdot \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{3^2}, \\ f_{11}^{(4)} &= \frac{2}{3} \cdot \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{1}{3} = \frac{1}{3^2} \cdot \frac{2}{3}, \\ &\dots \\ f_{11}^{(n)} &= \frac{1}{3^2} \left(\frac{2}{3}\right)^{n-3}, \\ &\dots \end{aligned}$$

由

$$\begin{aligned} f_{11} &= \sum_{n=1}^{\infty} f_{11}^{(n)} \\ &= \frac{1}{3} + \frac{1}{3} + \sum_{k=3}^{\infty} \frac{1}{3^2} \left(\frac{2}{3}\right)^{k-3} = 1 \end{aligned}$$

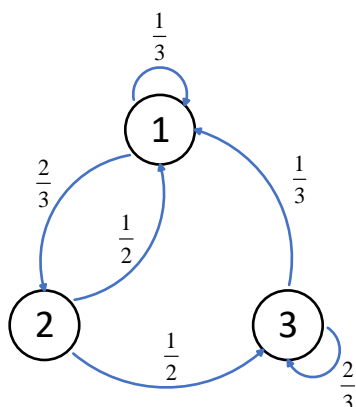


图 3.7 题 3.9 (1) 的状态转移图

知状态 1 常返. 又由

$$\sum_{n=1}^{\infty} n f_{11}^{(n)} = 1 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3} + \sum_{k=3}^{\infty} k \frac{1}{3^2} \left(\frac{2}{3}\right)^{k-3} = \frac{8}{3} < \infty$$

知状态 1 是正常返的, 平均返回时间为  $\frac{8}{3}$ . 状态 2 和 3 可类似的加以讨论.

(2) 根据状态转移图 3.8, 计算各个状态的首返概率

$$f_{11}^{(1)} = \frac{1}{2}, \quad f_{11}^{(2)} = \frac{1}{2} \cdot 1 = \frac{1}{2}, \quad f_{11}^{(n)} = 0, \quad n \geq 3.$$

易得

$$f_{11} = 1, \quad \sum_{n=1}^{\infty} n f_{11}^{(n)} = \frac{3}{2}.$$

因此, 状态 1 为正常返态, 平均返回时间为  $\frac{3}{2}$ .

$$f_{22}^{(1)} = 0, \quad f_{22}^{(2)} = \frac{1}{2} \cdot 1 = \frac{1}{2}, \quad f_{22}^{(n)} = \frac{1}{2} \cdot \left(\frac{1}{2}\right)^{n-2} \cdot 1 = \frac{1}{2^{n-1}}, \quad n \geq 3.$$

于是

$$f_{22} = 1, \quad \sum_{n=1}^{\infty} n f_{22}^{(n)} = 3.$$

则状态 2 为正常返态, 平均返回时间为 3.

$$f_{33}^{(1)} = \frac{2}{3}, \quad f_{33}^{(n)} = 0, \quad n \geq 2.$$

所以, 状态 3 为瞬时态.

$$f_{44}^{(n)} = 0, \quad n \geq 1.$$

所以, 状态 4 为瞬时态.

□

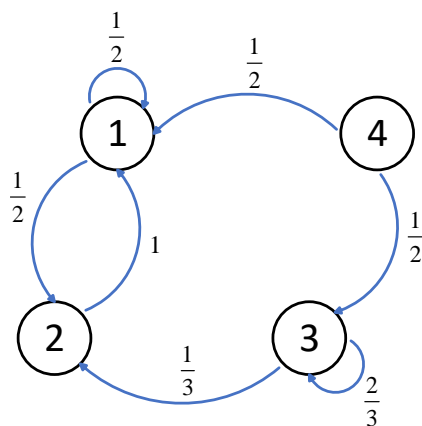


图 3.8 题 3.9 (2) 的状态转移图

**例3.10.** 计算一维平衡随机徘徊 0 状态的首返概率.

**解:** 显然,  $f_{00}^{(2k-1)} = 0, \forall k$ . 为了计算  $f_{00}^{(2k)}$ , 需要引入反射原理. 首先给出以下定义:

- $N(a, b, n)$ : 从  $a$  出发经过  $n$  步到达  $b$  的轨道数 (即可能的走法数). 设任一轨道中, 向右走的步数为  $k$ , 则向左走的步数为  $n - k$ , 满足

$$k - (n - k) = b - a,$$

可解得

$$k = \frac{n + b - a}{2}.$$

于是,

$$N(a, b, n) = \binom{n}{\frac{n+b-a}{2}}$$

- $N^0(a, b, n)$ : 从  $a$  出发经过  $n$  步到达  $b$ , 并且经过了 0 点的轨道数. 其数值可由反射原理求得.

**引理3.5. (反射原理)** 对任意  $a > 0, b > 0$ , 有

$$N^0(a, b, n) = N(-a, b, n).$$

**证明:** 对于任意一条从  $a$  出发经过  $n$  步到达  $b$ , 并且经过了 0 点的轨道, 将首次到达 0 点的时刻之前的路径沿时间轴翻折, 如图 3.9 所示, 得到一条从  $-a$  出发经过  $n$  步到达  $b$  的轨道. 因为任意一条从  $-a$  出发经过  $n$  步到达  $b$  的轨道必经过 0 点, 由于首达时刻的唯一性, 图中蓝色轨道与红色轨道一一对应, 于是有  $N^0(a, b, n) = N(-a, b, n)$ .  $\square$

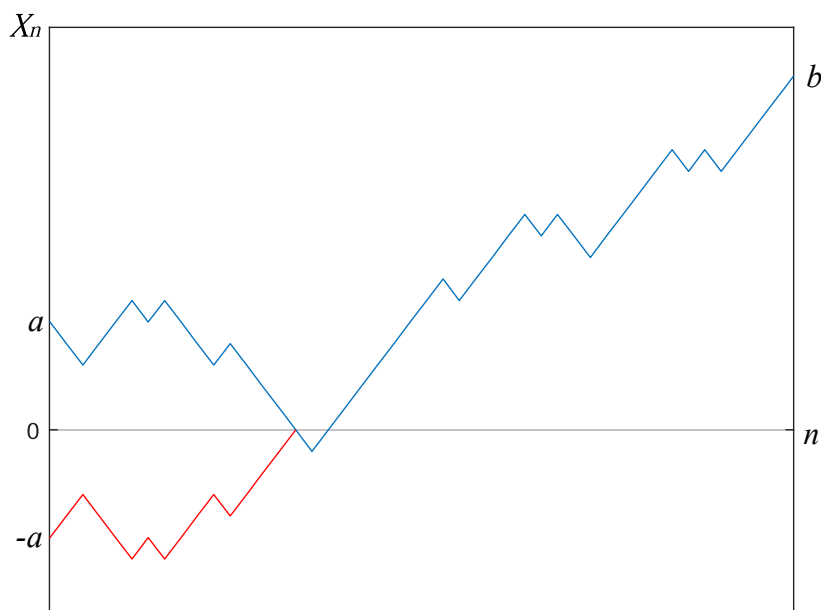


图 3.9 反射原理示意图

由于是平衡随机徘徊，从  $a$  出发经过  $n$  步到达  $b$  的每一条轨道都是等可能的，概率为  $(\frac{1}{2})^n$ 。于是，首返概率  $f_{00}^{(2k)}$  等于从 0 出发经过  $2k$  步首次返回 0 的轨道数乘以  $(\frac{1}{2})^{2k}$ ，即

$$\begin{aligned}
 f_{00}^{(2k)} &= 2(N(1, 1, 2k - 2) - N^0(1, 1, 2k - 2)) \left(\frac{1}{2}\right)^{2k} \\
 &= 2 \left( \binom{2k-2}{k-1} - \binom{2k-2}{k} \right) \left(\frac{1}{2}\right)^{2k} \\
 &= \frac{2}{k} \frac{(2k-2)!}{((k-1)!)^2} \left(\frac{1}{2}\right)^{2k}. \tag{3.2. 1}
 \end{aligned}$$

□

从上面的例子可以看到，当 Markov 链结构较简单时，可以从定义出发判断状态的常返性。但是对于一些复杂的 Markov 链，首返概率往往难以计算。即使求出了首返概率，用定义直接判断其常返性也不是十分容易。因此，需要寻找常返性的其他判定条件。

**定理3.6.** 状态  $i$  是常返态当且仅当  $n$  步转移概率满足

$$\sum_{n=0}^{\infty} p_{ii}^{(n)} = \infty.$$

状态  $i$  是瞬时态当且仅当

$$\sum_{n=0}^{\infty} p_{ii}^{(n)} < \infty.$$

**证明：** 首先考察  $n$  步首返概率与  $n$  步转移概率之间的关系. 令

$$A_k = \{X_k = j, X_{k-1} \neq j, \dots, X_1 \neq j\}$$

以状态  $i$  首达状态  $j$  的时间作为划分, 对转移概率作全概率展开, 有

$$\begin{aligned} p_{ij}^{(n)} &= P(X_n = j | X_0 = i) \\ &= \sum_{k=1}^n P(A_k | X_0 = i) P(X_n = j | A_k, X_0 = i) \\ &= \sum_{k=1}^n P(A_k | X_0 = i) P(X_n = j | X_k = j) \\ &= \sum_{k=1}^n f_{ij}^{(k)} p_{jj}^{(n-k)}. \end{aligned} \quad (3.2. 2)$$

定义概率母函数

$$P_{ij}(z) = \sum_{n=0}^{\infty} p_{ij}^{(n)} z^n, \quad F_{ij}(z) = \sum_{n=1}^{\infty} f_{ij}^{(n)} z^n,$$

其中

$$p_{ij}^{(0)} = \delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$



于是

$$\begin{aligned}
P_{ij}(z) &= \delta_{ij} + \sum_{n=1}^{\infty} p_{ij}^{(n)} z^n \\
&= \delta_{ij} + \sum_{n=1}^{\infty} \sum_{k=1}^n f_{ij}^{(k)} p_{jj}^{(n-k)} z^n \\
&= \delta_{ij} + \sum_{k=1}^{\infty} \sum_{n=k}^{\infty} f_{ij}^{(k)} z^k p_{jj}^{(n-k)} z^{n-k} \\
&= \delta_{ij} + \sum_{k=1}^{\infty} f_{ij}^{(k)} z^k \sum_{n=k}^{\infty} p_{jj}^{(n-k)} z^{n-k} \\
&= \delta_{ij} + \sum_{k=1}^{\infty} f_{ij}^{(k)} z^k \sum_{m=0}^{\infty} p_{jj}^{(m)} z^m \\
&= \delta_{ij} + F_{ij}(z) P_{jj}(z).
\end{aligned}$$

当  $i = j$  时, 有

$$P_{ii}(z) = 1 + F_{ii}(z) P_{ii}(z).$$

由此得到

$$P_{ii}(z) = \frac{1}{1 - F_{ii}(z)}.$$

令  $z \rightarrow 1^-$ , 有

$$\sum_{n=0}^{\infty} p_{ii}^{(n)} = \frac{1}{1 - \sum_{n=1}^{\infty} f_{ii}^{(n)}} = \frac{1}{1 - f_{ii}}.$$

由此易得定理的结论. □

判断级数的收敛性, 只需要对级数求和项的阶数有一个估计, 相比于判断首返概率的级数求和是否精确地等于 1, 往往要容易很多. 利用上述判据, 我们再来讨论随机徘徊的常返性.

**例3.11.** 讨论随机徘徊各状态的常返性.

- (1) 一维无限制随机徘徊, 质点向右和向左的概率分别为  $p$  和  $1 - p$ ;
- (2) 二维“平衡”随机徘徊, 质点向上、下、左、右四个方向运动的概率均为  $\frac{1}{4}$ ;
- (3) 三维“平衡”随机徘徊.

**解：** (1) 由于随机徘徊各个状态等价，只需要讨论 0 状态的常返性. 从 0 状态到 0 状态的  $n$  步转移概率为

$$p_{00}^{(n)} = \begin{cases} \binom{2k}{k} p^k (1-p)^k, & n = 2k, \\ 0, & n = 2k-1. \end{cases}$$

根据定理 3.6, 只需考察级数  $\sum_{k=1}^{\infty} p_{00}^{(2k)}$  的收敛性, 即通项的阶数. 由 Stirling 公式, 即

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n, \quad n \rightarrow \infty$$

可知, 当  $k \rightarrow \infty$  时, 有

$$\frac{(2k)!}{k!k!} p^k (1-p)^k \sim \frac{\sqrt{2\pi \cdot 2k} \left(\frac{2k}{e}\right)^{2k}}{\left(\sqrt{2\pi k} \left(\frac{k}{e}\right)^k\right)^2} p^k (1-p)^k = \frac{(4p(1-p))^k}{\sqrt{\pi k}}. \quad (3.2. 3)$$

当  $p \neq \frac{1}{2}$  时,  $4p(1-p) < 1$ , 通项按指数收敛, 于是级数收敛, 所有状态均为瞬时态; 当  $p = \frac{1}{2}$  时, 通项的阶数为  $\frac{1}{\sqrt{k}}$ , 则级数发散, 所有状态均为常返态.

(2) 考虑  $x-y$  直角坐标系, 要从 0 位置返回 0 位置, 首先步数必为偶数, 其次必有一半的步数沿  $x$  轴或  $y$  轴正方向走, 另一半沿  $x$  轴或  $y$  轴负方向走, 并且  $x$  轴或  $y$  轴正负方向的步数一致. 于是

$$\begin{aligned} p_{00}^{(2k)} &= \binom{2k}{k} \sum_{m=0}^k \binom{k}{m} \binom{k}{m} \left(\frac{1}{4}\right)^{2k} \\ &= \binom{2k}{k} \sum_{m=0}^k \binom{k}{m} \binom{k}{k-m} \left(\frac{1}{4}\right)^{2k} \\ &= \binom{2k}{k} \binom{2k}{k} \left(\frac{1}{4}\right)^{2k} \\ &= \left(\frac{(2k)!}{k!k!} \left(\frac{1}{2}\right)^{2k}\right)^2 \sim \frac{1}{\pi k}, \end{aligned}$$

其中, 最后一步由 (3.2. 3) 得到, 第三个等号是组合数公式\*. 由上式可知, 级数

---

\*公式  $\sum_{m=0}^k \binom{k}{m} \binom{k}{k-m} = \binom{2k}{k}$  可由以下方法证明: 箱子里有  $2k$  个球, 任取  $k$  个, 共有  $\binom{2k}{k}$  种取法. 也可以用另一种方法取出  $k$  个球, 即在箱子里放置一块木板隔成左右两半, 每一边各有  $k$  个球. 然后分别从两边取球, 如果从左边取出了  $m$  个球, 则从右边取出  $k-m$  个球, 总共刚好取出  $k$  个球. 取遍所有可能的  $m$  的取值, 共有  $\sum_{m=0}^k \binom{k}{m} \binom{k}{k-m}$  种取法, 同样得到  $2k$  个球中任取  $k$  个的所有取法.

$\sum_{k=1}^{\infty} p_{00}^{(2k)}$  发散, 则所有状态均常返.

(3) 对于三维随机徘徊, 类似的,

$$\begin{aligned} p_{00}^{(2k)} &= \binom{2k}{k} \sum_{n+m \leq k} \left( \binom{k}{n} \binom{k-n}{m} \right)^2 \left( \frac{1}{6} \right)^{2k} \\ &= \frac{(2k)!}{k!k!} \left( \frac{1}{2} \right)^{2k} \sum_{n+m+l=k} \frac{(k!)^2}{(n!)^2(m!)^2(l!)^2} \left( \frac{1}{3} \right)^{2k}. \end{aligned} \quad (3.2.4)$$

考察  $k$  为 3 的整数倍的情况  $k = 3j$ , 由于  $n + m + l = k$ , 不妨设  $n \leq m \leq j \leq l$ , 则有

$$\frac{(j!)^3}{n!m!l!} = \frac{(n+1) \cdots j \cdot (m+1) \cdots j}{(j+1) \cdots l} < 1.$$

所以

$$\frac{1}{n!m!l!} < \frac{1}{(j!)^3}.$$

代入式 (3.2.4), 得到

$$\begin{aligned} p_{00}^{(2k)} &< \frac{(2k)!}{k!k!} \left( \frac{1}{2} \right)^{2k} \sum_{n+m+l=k} \frac{k!}{n!m!l!} \frac{k!}{(j!)^3} \left( \frac{1}{3} \right)^{2k} \\ &= \frac{(2k)!}{k!k!} \left( \frac{1}{2} \right)^{2k} 3^k \frac{k!}{(j!)^3} \left( \frac{1}{3} \right)^{2k} \\ &\sim \frac{1}{\sqrt{\pi k}} \frac{\sqrt{2\pi k} \left( \frac{k}{e} \right)^k}{\left( \sqrt{2\pi \frac{k}{3}} \left( \frac{k}{3e} \right)^{\frac{k}{3}} \right)^3 3^k} = \frac{1}{2} \left( \frac{3}{\pi k} \right)^{\frac{3}{2}} \end{aligned}$$

当  $k$  不为 3 的整数倍时, 通过对  $\frac{k}{3}$  向下取整可作类似的估计. 综上可知,  $p_{00}^{(2k)}$  的阶数低于  $\frac{1}{k^{\frac{3}{2}}}$ , 而后者收敛, 故级数  $\sum_{k=1}^{\infty} p_{00}^{(2k)}$  收敛, 即所有状态非常返.  $\square$

对于一维平衡随机徘徊, 即  $p = \frac{1}{2}$  的情况, 进一步考虑其正常返性. 例 3.10 中已经求得了首返概率, 如式 (3.2.1) 所示. 则平均返回时间为

$$\mu_0 = \sum_{k=1}^{\infty} 2k f_{00}^{(2k)} = \sum_{k=1}^{\infty} \frac{(2k-2)!}{((k-1)!)^2} \left( \frac{1}{2} \right)^{2k-2}.$$

根据 (3.2.3) 可知, 级数求和通项的阶数为  $\frac{1}{\sqrt{k}}$ , 级数发散, 因此所有状态均为零常返.

瞬时态和常返态的转移概率具有如下性质:

**定理3.7.** 设  $i$  是 Markov 链的任一状态,

(1) 如果  $i$  是瞬时态或零常返态, 则

$$\lim_{n \rightarrow \infty} p_{ii}^{(n)} = 0;$$

(2) 如果  $i$  是非周期正常返态, 则

$$\lim_{n \rightarrow \infty} p_{ii}^{(n)} = \frac{1}{\mu_i};$$

(3) 如果  $i$  是周期为  $d_i$  的正常返态, 则

$$\lim_{n \rightarrow \infty} p_{ii}^{(nd_i)} = \frac{d_i}{\mu_i}.$$

其中,  $\mu_i$  是状态  $i$  的平均返回时间.

**证明:** 略. □

由上述定理可得到如下推论:

**推论3.8.** 设  $i, j$  是 Markov 链的任意两个状态,

(1) 如果  $j$  是瞬时态或零常返态, 则

$$\lim_{n \rightarrow \infty} p_{ij}^{(n)} = 0, \quad \forall i;$$

(2) 如果  $j$  是非周期正常返态, 则

$$\lim_{n \rightarrow \infty} p_{ij}^{(n)} = \frac{f_{ij}}{\mu_j}, \quad \forall i.$$

其中,  $f_{ij} = \sum_{n=1}^{\infty} f_{ij}^{(n)}$ .

**证明:** 当  $i = j$  时, 由 3.7 知, 推论自然成立. 当  $i \neq j$  时, 由式 (3.2. 2), 对任意  $n > M > 1$ , 有

$$\begin{aligned} p_{ij}^{(n)} &= \sum_{k=1}^n f_{ij}^{(k)} p_{jj}^{(n-k)} \\ &= \sum_{k=1}^M f_{ij}^{(k)} p_{jj}^{(n-k)} + \sum_{k=M+1}^n f_{ij}^{(k)} p_{jj}^{(n-k)}. \end{aligned} \quad (3.2. 5)$$

由于首达概率满足  $\sum_{k=1}^{\infty} f_{ij}^{(k)} \leq 1$ , 则有

$$\sum_{k=M+1}^n f_{ij}^{(k)} p_{jj}^{(n-k)} \rightarrow 0, \quad n \rightarrow \infty, M \rightarrow \infty.$$

考察式 (3.2. 5) 中第一个求和项, 由定理 3.7 可知

$$\begin{aligned} \lim_{n \rightarrow \infty} \sum_{k=1}^M f_{ij}^{(k)} p_{jj}^{(n-k)} &= \sum_{k=1}^M f_{ij}^{(k)} \lim_{n \rightarrow \infty} p_{jj}^{(n-k)} \\ &= \begin{cases} 0, & j \text{ 是瞬时态或零常返态,} \\ \frac{1}{\mu_j} \sum_{k=1}^M f_{ij}^{(k)}, & j \text{ 是非周期正常返态.} \end{cases} \end{aligned}$$

由于  $M$  的任意性, 对式 (3.2. 5) 再令  $M \rightarrow \infty$ , 可得

$$\lim_{n \rightarrow \infty} p_{ij}^{(n)} = \begin{cases} 0, & j \text{ 是瞬时态或零常返态,} \\ \frac{f_{ij}}{\mu_j}, & j \text{ 是非周期正常返态.} \end{cases}$$

□

下面的定理表明, 常返性也是一种类性质.

**定理3.9.** 如果  $i \leftrightarrow j$ , 则  $i$  和  $j$  具有相同的常返性, 即

- (1)  $i$  是瞬时态当且仅当  $j$  是瞬时态;
- (2)  $i$  是常返态当且仅当  $j$  是常返态;
- (3)  $i$  是正常返态当且仅当  $j$  是正常返态.

**证明:** 略.

□

从随机徘徊的例子可以看到, 可数状态 Markov 链可以不存在正常返态, 甚至所有状态都非常返. 如果 Markov 链的状态有限, 则常返态一定存在.

**定理3.10.** 有限状态 Markov 链一定存在正常返态. 对于不可约的有限状态 Markov 链, 所有状态都是正常返态.

**证明:** 用反证法证明. 假设 Markov 链可有  $M$  个状态, 均为非常返或零常返. 则由推论 3.8 可知, 对任意  $i, j$ , 有

$$p_{ij}^{(n)} \rightarrow 0, \quad n \rightarrow \infty.$$

上式对  $j$  求和, 得到

$$\sum_{j=1}^M p_{ij}^{(n)} \rightarrow 0, \quad n \rightarrow \infty.$$

与条件分布律性质

$$\sum_{j=1}^M p_{ij}^{(n)} = 1, \quad \forall n$$

矛盾. 假设不成立, 因此一定存在正常返态.

对于不可约的有限状态 Markov 链, 结合定理 3.9 易知, 所有状态均为正常返态.  $\square$

根据上述定理, 例 3.9 中 Markov 链各状态的常返性几乎一望便知. 下面再看几个例子.

**例3.12.** 考察以下 Markov 链各状态的常返性

$$(1) \begin{pmatrix} 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \quad (2) \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{4} & 0 & 0 & \frac{1}{2} \end{pmatrix}.$$

**解:** (1) 容易验证, 该链各个状态都相通, 是不可约 Markov 链, 则所有状态均为正常返态.

(2) 容易验证,  $\{1, 2\}$  和  $\{3, 4\}$  是两个闭的真子集, 集合内状态相通, 均为正常返, 而状态 5 为瞬时态.  $\square$

### 3.2.4 非常返态分析

当时间足够长时, Markov 链的行为由常返态的行为决定, 非常返态只在链运行初期起作用. 在一些应用中, 非常返态分析 also 具有重要意义.

**例3.13. (赌徒输光问题)** 一个赌徒在赌场进行一系列赌博. 令  $i$  为其原始赌本, 每次赌博赌注为 1, 且以概率  $p$  赢, 收回赌注并获得 1 份收益. 如果输掉, 则失去赌注. 假设赌徒手中的赌本到达  $N$  ( $N > i$ ) 时, 则认为达到预期目标, 会离开赌场. 那么在他达到赚钱目的之前, 有多大可能会把手中的赌本全部输光?

**解:** 设  $X_n$  表示  $n$  时刻的剩余赌本, 根据题意  $X_n \in \{0, 1, \dots, N\}$ , 可以建模为带有两个吸收壁的随机徘徊. 定义

$$f_i = \lim_{n \rightarrow \infty} P(X_n = 0 | X_0 = i)$$

表示从状态  $i$  出发最终输光的概率. 根据题设条件, 有

$$f_0 = 1, \quad f_N = 0.$$

利用**单步递推法**进行求解. 具体来说, 由第一步的走法做全概率展开, 得到

$$\begin{aligned} P(X_n = 0 | X_0 = i) &= P(X_1 = i + 1 | X_0 = i)P(X_n = 0 | X_1 = i + 1, X_0 = i) \\ &\quad + P(X_1 = i - 1 | X_0 = i)P(X_n = 0 | X_1 = i - 1, X_0 = i) \\ &= pP(X_n = 0 | X_1 = i + 1) + qP(X_n = 0 | X_1 = i - 1), \end{aligned}$$

两边取极限  $n \rightarrow \infty$ , 得到

$$f_i = pf_{i+1} + qf_{i-1},$$

其中,  $q = 1 - p$ . 由上式得到

$$q(f_i - f_{i-1}) = p(f_{i+1} - f_i),$$

即  $\{f_i - f_{i-1}\}$  组成等比数列, 于是

$$f_i - f_{i-1} = \left(\frac{q}{p}\right)^{i-1} (f_1 - f_0).$$

当  $q \neq p$  时, 通过迭代计算可得

$$\begin{aligned} f_i &= \left(\frac{q}{p}\right)^{i-1} (f_1 - f_0) + \left(\frac{q}{p}\right)^{i-2} (f_1 - f_0) + (f_1 - f_0) + f_0 \\ &= \frac{1 - \left(\frac{q}{p}\right)^i}{1 - \frac{q}{p}} (f_1 - 1) + 1. \end{aligned}$$

代入边界条件

$$f_N = \frac{1 - \left(\frac{q}{p}\right)^N}{1 - \frac{q}{p}} (f_1 - 1) + 1 = 0$$

得

$$f_1 = 1 - \frac{1 - \frac{q}{p}}{1 - \left(\frac{q}{p}\right)^N}.$$

从而求得

$$f_i = \frac{\left(\frac{q}{p}\right)^i - \left(\frac{q}{p}\right)^N}{1 - \left(\frac{q}{p}\right)^N}.$$

当  $q = p$  时, 有

$$f_i = i(f_1 - 1) + 1.$$

代入边界条件

$$f_N = N(f_1 - 1) + 1 = 0$$

得

$$f_1 = 1 - \frac{1}{N}.$$

从而求得

$$f_i = \frac{N - i}{N}.$$

□

单步递推法是 Markov 性质的典型应用, 可以应用于非常返态分析的很多问题中. 下面简要介绍命中时间和吸收概率的概念.

**定义3.13.** 设  $X_n$  为离散时间 Markov 链, 状态空间为  $\mathcal{E}$ . 对于任意状态  $i$  和任意状态集合  $\mathcal{A}$ , 定义

$$T_i^{\mathcal{A}} = \min\{n : X_n \in \mathcal{A} | X_0 = i\}$$

为状态  $i$  到集合  $\mathcal{A}$  的命中时间 (Hitting Time).

**定义3.14.** 设  $X_n$  为离散时间 Markov 链, 状态空间为  $\mathcal{E}$ . 状态集合  $\mathcal{A}$  对状态  $i$  的吸收概率 (Absorbing Probability) 为

$$P_i^{\mathcal{A}} = P(T_i^{\mathcal{A}} < \infty)$$

很显然, 如果  $i \in \mathcal{A}$ , 则  $T_i^{\mathcal{A}} = 0, P_i^{\mathcal{A}} = 1$ . 下面求解集合  $\mathcal{A}$  对任意状态  $i \notin \mathcal{A}$  的吸收概率. 根据单步递推法, 有

$$P_i^{\mathcal{A}} = \sum_{j \in \mathcal{E}} p_{ij} P_j^{\mathcal{A}}. \quad \forall i \in \mathcal{E} \setminus \mathcal{A}.$$

再来计算平均命中时间. 同样利用单步递推法可得

$$E(T_i^{\mathcal{A}}) = \sum_{j \in \mathcal{E}} p_{ij} E(T_j^{\mathcal{A}}) + 1. \quad \forall i \in \mathcal{E} \setminus \mathcal{A}.$$

同样需要求解一个线性方程组.



### 3.3 极限行为与平稳分布

#### 3.3.1 Markov 链的极限行为

Markov 链可由转移概率  $p_{ij}^{(n)}$  表征, 其在时间足够长时的最终形态也可由转移概率的极限进行刻画. 这时要考虑的问题是, Markov 链状态概率分布的极限  $\lim_{n \rightarrow \infty} P(X_n = j)$  是否存在? 如果存在, 这个极限是多少?

**定义3.15.** 设 Markov 链  $X_n$  的状态空间为  $\mathcal{E}$ , 如果存在分布  $p_i$ , 即满足非负性  $p_i \geq 0, \forall i \in \mathcal{E}$  和归一性  $\sum_{i \in \mathcal{E}} p_i = 1$ , 使得

$$\lim_{n \rightarrow \infty} P(X_n = i) = p_i,$$

则称为该 Markov 链的极限分布.

由

$$P(X_n = j) = \sum_{i \in \mathcal{E}} P(X_0 = i) p_{ij}^{(n)}$$

可知, 极限分布依赖于初值和  $n$  步转移概率. 对于非常返态或零常返态  $j$ , 由于  $\lim_{n \rightarrow \infty} p_{ij}^{(n)} = 0, \forall i$ , 则  $\lim_{n \rightarrow \infty} P(X_n = j) = 0$ . 于是, Markov 链的最终形态由正常返态决定.

**例3.14.** 天气预报问题 (例 3.4) 中的转移概率为

$$\mathbf{P} = \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix},$$

$$\mathbf{P}^{(n)} = \frac{1}{\alpha + \beta} \begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix} + \frac{(1 - \alpha - \beta)^n}{\alpha + \beta} \begin{pmatrix} \alpha & -\alpha \\ -\beta & \beta \end{pmatrix},$$

讨论该 Markov 链的极限.

**解:** 当  $\alpha = \beta = 0$  时, 有

$$\mathbf{P} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

此时 Markov 链可约, 包含两个不可约的子链, 且两个状态均为吸收态, 极限由初值决定.

当  $\alpha = \beta = 1$  时, 有

$$\mathbf{P}^{(n)} = \frac{1}{2} \begin{pmatrix} 1 + (-1)^n & 1 - (-1)^n \\ 1 - (-1)^n & 1 + (-1)^n \end{pmatrix},$$

极限不存在. 特别的, 由

$$\mathbf{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

可知, 该链是周期的, 周期为 2.

当  $|1 - \alpha - \beta| < 1$  时, 有

$$\lim_{n \rightarrow \infty} \mathbf{P}^{(n)} = \frac{1}{\alpha + \beta} \begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix},$$

则

$$\lim_{n \rightarrow \infty} P(X_n = 1) = \lim_{n \rightarrow \infty} \sum_{i \in \{1, 2\}} P(X_0 = i) p_{i1}^{(n)} = \frac{\beta}{\alpha + \beta}.$$

类似的

$$\lim_{n \rightarrow \infty} P(X_n = 2) = \frac{\alpha}{\alpha + \beta}.$$

因此极限存在, 并且不依赖于初值. □

**例3.15.** 讨论下列 Markov 链的极限

$$(1) \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{4} & 0 & 0 & \frac{1}{2} \end{pmatrix}, \quad (2) \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

**解:** (1) 该 Markov 链可约, 存在  $\{1, 2\}$  和  $\{3, 4\}$  两个不可约的子链, 概率分布的极限由初值决定.

(2) 这是例 3.14 的特殊情形. □

从上面的例子可以看到, 概率分布的极限的存在性与不可约性和非周期性有紧密的联系. 如果 Markov 链可约, 则极限可能与初值有关; 如果存在周期性, 则极限可能不存在.

### 3.3.2 平稳分布

如果极限分布存在, 通过计算  $\mathbf{P}^n$  来求解极限分布比较繁琐. 为此引入平稳分布的概念.

**定义3.16.** 设  $\mathbf{P}$  为 Markov 链的一步转移概率矩阵, 如果概率分布  $\boldsymbol{\pi} = (\pi_0, \pi_1, \dots)$  满足平衡方程 (Equilibrium Equation)

$$\boldsymbol{\pi} = \boldsymbol{\pi} \mathbf{P},$$

则称其为该 Markov 链的**平稳分布 (Stationary Distribution)** 或**不变分布 (Invariant Distribution)**.

平稳分布中“不变”的物理意义非常直观: 若初始分布为  $\pi$ , 则任意时刻分布均为  $\pi$ . 通过求解平衡方程, 可以直接判断平稳分布是否存在.

**例3.16.** 讨论例 3.15 中 Markov 链的平稳分布.

**解:** (1) 直接求解平衡方程及  $\sum_n \pi_n = 1$  易知, 平稳分布存在, 且有无穷多个, 即

$$\pi = \left(\frac{\alpha}{2}, \frac{\alpha}{2}, \frac{\beta}{2}, \frac{\beta}{2}, 0\right), \quad \alpha \geq 0, \beta \geq 0, \alpha + \beta = 1.$$

(2) 求解得到

$$\pi = \left(\frac{1}{2}, \frac{1}{2}\right).$$

该链的平稳分布存在且唯一. □

从上述例子可以看到, 尽管极限分布不存在, 平稳分布仍有可能存在.

**例3.17.** 讨论带一个反射壁的一维随机徘徊 (例 3.1 (2)) 的平稳分布.

**解:** 带一个反射壁的一维随机徘徊的转移概率矩阵为

$$P = \begin{pmatrix} q & p & & \\ q & 0 & p & \\ & \ddots & \ddots & \ddots \end{pmatrix}.$$

由平衡方程  $\pi = \pi P$ , 即

$$\begin{aligned} \pi_0 &= q\pi_0 + q\pi_1, \\ \pi_1 &= p\pi_0 + q\pi_2, \\ &\dots \\ \pi_n &= p\pi_{n-1} + q\pi_{n+1}, \\ &\dots \end{aligned}$$

可得

$$\begin{aligned} p\pi_0 &= q\pi_1, \\ p\pi_1 &= q\pi_2, \\ &\dots \\ p\pi_n &= q\pi_{n+1}, \\ &\dots \end{aligned}$$

于是解得

$$\pi_n = \left(\frac{p}{q}\right)^n \pi_0.$$

当  $p \geq q$  时, 级数

$$\sum_{n=0}^{\infty} \pi_n = \sum_{n=0}^{\infty} \left(\frac{p}{q}\right)^n \pi_0$$

发散, 不存在分布解, 即平稳分布不存在.

当  $p < q$  时, 由

$$\sum_{n=0}^{\infty} \pi_n = \frac{1}{1 - \frac{p}{q}} \pi_0 = 1$$

解得

$$\pi_0 = 1 - \frac{p}{q}.$$

从而

$$\pi_n = \left(\frac{p}{q}\right)^n \left(1 - \frac{p}{q}\right).$$

□

上述例子的直观解释是, 当质点向右走的概率更大时, 会逐渐走向无穷远处, 任意位置都不会被频繁访问. 只有当质点向左走的概率较大时, 才会不断地“碰壁”反弹, 达到平衡.

**例3.18.** 讨论Ehrenfest 模型 (例 3.2) 的平稳分布.

**解:** Ehrenfest 模型的转移概率矩阵为

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & & & \\ \frac{1}{M} & 0 & \frac{M-1}{M} & & \\ & \ddots & \ddots & \ddots & \\ & & \frac{M-1}{M} & 0 & \frac{1}{M} \\ & & & 1 & 0 \end{pmatrix}.$$

由平衡方程  $\pi = \pi P$ , 即

$$\begin{aligned}\pi_0 &= \frac{1}{M}\pi_1, \\ \pi_1 &= \pi_0 + \frac{2}{M}\pi_2, \\ &\dots \\ \pi_n &= \frac{M-(n-1)}{M}\pi_{n-1} + \frac{n+1}{M}\pi_{n+1}, \\ &\dots \\ \pi_M &= \frac{1}{M}\pi_{M-1}.\end{aligned}$$

迭代关系为

$$\begin{aligned}M\pi_0 &= \pi_1, \\ (M-1)\pi_1 &= 2\pi_2, \\ &\dots \\ (M-n)\pi_n &= (n+1)\pi_{n+1}, \\ &\dots \\ \pi_{M-1} &= M\pi_M.\end{aligned}$$

求解得到

$$\pi_n = \frac{1}{2^M} \binom{M}{n}, \quad n = 0, 1, \dots, M.$$

该链的平稳分布存在且唯一. □

一般来说, 平稳分布与极限存在没有必然联系. 从上面的例子也可以看到:

- (1) 平稳分布可能不唯一;
- (2) 周期性 Markov 链没有极限, 但平稳分布可能存在;
- (3) 不可约零常返或瞬时 Markov 链, 极限为 0, 但没有平稳分布.

在满足不可约和非周期的条件下, 两者之间可以联系起来.

**定理3.11.** 一个不可约非周期的 Markov 链存在极限分布, 当且仅当存在平稳分布, 且两者相等.

当 Markov 链是不可约非周期的, 极限分布的存在性可以通过解平衡方程进行判断和求解. 进一步的, 平稳分布的存在性与正常返具有以下关系.

- 定理3.12.** (1) Markov 链平稳分布**存在**的充分必要条件是链中存在正常返态；
- (2) Markov 链平稳分布**存在且唯一**的充分必要条件是链中存在唯一的不可约正常返子集；
- (3) Markov 链平稳分布**存在、唯一且所有元素都为正**的充分必要条件是该链为不可约正常返的，且

$$\pi_i = \frac{1}{\mu_i},$$

其中  $\mu_i$  是平均返回时间.

### 3.3.3 应用举例——搜索引擎 PageRank 技术

PageRank 技术是 Google 搜索引擎的核心技术之一，以 Google 公司创办人拉里·佩奇 (Larry Page) 之姓来命名. 其基本原理是：将用户在不同网页间按照超链接跳转的行为看作 Markov 链，将 Markov 链的平稳分布作为网页排序标准. 这一思想的物理直观是，每个页面最初都有一个单位的重要性，每次链接跳转会在后继者中分享其重要性. 有大量传入链接的页面将最有较高的重要性.

考察图 3.10 所示的简单网络，该网络的转移概率矩阵为

$$\mathbf{P} = \begin{pmatrix} 0.5 & 0.5 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \end{pmatrix}.$$

求解平衡方程得到

$$\pi_A = \pi_B = \frac{2}{5}, \pi_C = \frac{1}{5}.$$

于是 A 和 B 具有较 C 更高的排序，直观的看，这些网页被更多的超链接指向.

然而，这种排序方法有一些缺陷，当某些网页只带有指向自身的超链接时，容易陷入死角或蜘蛛陷阱，如图 3.11 所示. 该网络中，由于 C 只具有指向自身的超链接，是吸收态，求解得到平稳分布为

$$\pi_A = \pi_B = 0, \pi_C = 1.$$

按此分布排序并不合理，易被恶意攻击. 更进一步的，如果存在多个蜘蛛陷阱，如图 3.12 所示，则平稳分布解不唯一，A 和 C 的排序具有一定的任意性.

为了解决上述问题，需要想办法跳出蜘蛛陷阱. 一种方案是对“重要性”进行征税，将一部分概率值平均分配给所有可能的状态转换. 在图 3.11 的例子中，将每个节点 30% 的转移概率平均分配给每个网页，得到新的转移概率矩阵为

$$\mathbf{P} = \begin{pmatrix} 0.45 & 0.45 & 0.1 \\ 0.45 & 0.1 & 0.45 \\ 0.1 & 0.1 & 0.8 \end{pmatrix}.$$

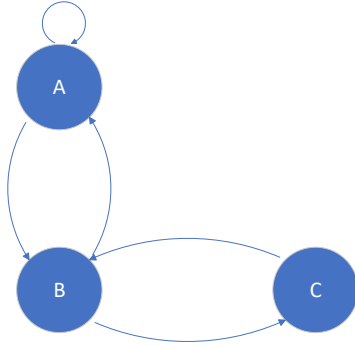


图 3.10 三个页面组成的网络

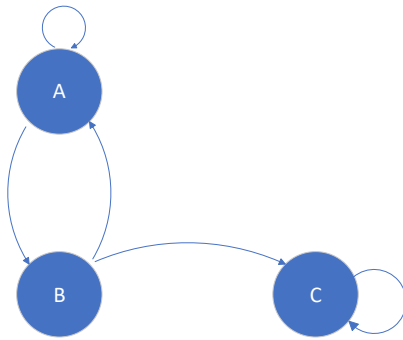


图 3.11 死角或蜘蛛陷阱

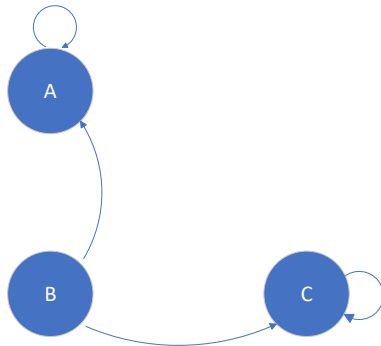


图 3.12 两个蜘蛛陷阱

上述方法避免了网络中出现吸收态，然而，通过求解会发现，低税收时蜘蛛陷阱的极限概率仍很高。如果采取高税收，又无法客观反映网页的真实重要性。为此，可以设计选择性的高税收策略：对于没有向外超链接的网页，如果人们不感兴趣，会通过退回操作离开该网页。因此，对频繁进行了退回操作的网页赋予较高的税收，可在降低蜘蛛陷阱的极限概率的同时，不影响其他网页的真实排序。当然，这一策略需要额外对用户访问网页的操作进行统计。

PageRank 技术是 Markov 链在实际中非常成功的应用范例之一。在实际应用中，还有很多工程问题需要解决。比如，由于全球网络规模巨大，需要求解一个超大的线性方程组，复杂度很高。平稳分布也可以通过求解转移概率矩阵的幂得到，而代表网页超链接关系的转移概率矩阵中，非零元素是相当稀疏的。因此，问题转化为高效求解超大型稀疏矩阵的幂的问题。另外，网络的规模和连接关系也在不断变化，导致 Markov 链动态变化。如何在发生变化时，高效地更新整个网络的平稳分布，也是一个重要问题。



## 第 4 章 Poisson 过程

Poisson 过程是典型的连续参数离散状态随机过程, 它刻画了“等待”和“计数”等行为中所蕴含的随机性, 在实际中有很多应用. 后面还会看到, Poisson 过程是最典型的 Markov 过程之一, 并且具有独特的性质, 内涵十分丰富. 本章将从 Poisson 过程的定义、性质、以及各种拓广等方面展开介绍.

### 4.1 Poisson 过程的定义

**定义4.1.** 如果随机过程  $N(t), t \geq 0$  表示时间段  $[0, t]$  内发生的某种事件的总和, 则称其为**计数过程 (counting processes)**.

计数过程在现实生活中十分常见, 比如保险公司在时间段  $[0, t]$  内受理的保险理赔次数、 $[0, t]$  内到银行办理业务的顾客数、计算机网络中时间段  $[0, t]$  内到达路由器的数据包个数、等等. 计数过程具有以下典型特性:

- (1)  $N(t)$  取非负整数;
- (2)  $\forall s, t$ , 如果  $s < t$ , 则  $N(s) \leq N(t)$ , 即计数过程的样本轨道是非减函数;
- (3)  $N(t) - N(s)$  表示从时刻  $s$  到时刻  $t$  之间发生的事件次数.

Poisson 过程是一类特殊的计数过程, 定义如下:

**定义4.2.** 满足如下条件的计数过程  $N(t)$  称为**(标准) Poisson 过程**:

- (1)  $N(0) = 0$ ;
- (2)  $N(t)$  具有**平稳增量性**, 即  $N(t) - N(s)$  的分布仅依赖于  $t - s$ ;
- (3)  $N(t)$  具有**独立增量性**, 即  $\forall t_1 < t_2 \leq t_3 < t_4$ ,  $N(t_4) - N(t_3)$  和  $N(t_2) - N(t_1)$  独立;
- (4) 微元时间  $\Delta t$  内发生超过 1 次事件的概率是刚好发生 1 次事件的概率的高阶无穷小, 即

$$\lim_{\Delta t \rightarrow 0} \frac{P(N(t + \Delta t) - N(t) \geq 2)}{P(N(t + \Delta t) - N(t) = 1)} = 0.$$

由上述定义, 可以导出 Poisson 过程的概率分布.

**定理4.1.** Poisson 过程的概率分布为

$$P(N(t) = k) = \frac{(\lambda t)^k}{k!} \exp(-\lambda t), \quad k = 0, 1, \dots,$$

其中  $\lambda > 0$ .

**证明：** 令

$$p_k(t) = P(N(t) = k), \forall k = 0, 1, \dots, t \geq 0.$$

分以下三步进行计算.

(1) 首先计算  $p_0(t)$ . 注意到

$$\begin{aligned} p_0(t+s) &= P(N(t+s) = 0) \\ &= P(N(t) = 0, N(t+s) - N(t) = 0) \\ &= P(N(t) = 0)P(N(t+s) - N(t) = 0) \\ &= p_0(t)p_0(s), \end{aligned}$$

其中第三个等号用到了 Poisson 过程的独立增量性, 第四个等号用到了平稳增量性. 由  $t$  和  $s$  的任意性知, 必有

$$p_0(t) = a^t.$$

又由计数过程样本轨道的单调性知,  $0 < a < 1$ . 令  $a = e^{-\lambda}, \lambda > 0$ , 得到

$$p_0(t) = e^{-\lambda t}.$$

(2) 接下来计算  $\lim_{\Delta t \rightarrow 0} p_k(\Delta t), \forall k = 0, 1, \dots$ . 对于无穷小量  $\Delta t$ , 有

$$p_0(\Delta t) = P(N(\Delta t) = 0) = e^{-\lambda \Delta t} = 1 - \lambda \Delta t + o(\Delta t).$$

注意到

$$P(N(\Delta t) = 0) + P(N(\Delta t) = 1) + P(N(\Delta t) \geq 2) = 1,$$

从而

$$P(N(\Delta t) = 1) \left( 1 + \frac{P(N(\Delta t) \geq 2)}{P(N(\Delta t) = 1)} \right) = \lambda \Delta t + o(\Delta t).$$

由定义 4.2 的条件 (4) 可得

$$P(N(\Delta t) = 1) = \lambda \Delta t + o(\Delta t).$$

于是

$$P(N(\Delta t) = k) = o(\Delta t), \quad \forall k \geq 2.$$

(3) 最后计算  $p_k(t), \forall k, t$ . 利用概率母函数求解, 定义

$$G(z, t) = E(z^{N(t)}) = \sum_{n=0}^{\infty} z^n P(N(t) = n).$$

由于

$$\begin{aligned} & G(z, t + \Delta t) - G(z, t) \\ &= E(z^{N(t+\Delta t)}) - E(z^{N(t)}) \\ &= E(z^{N(t)}(z^{N(t+\Delta t)-N(t)} - 1)) \\ &= E(z^{N(t)})(E(z^{N(t+\Delta t)-N(t)}) - 1) \\ &= G(z, t) \left( \sum_{k=0}^{\infty} z^k P(N(t + \Delta t) - N(t) = k) - 1 \right) \\ &= G(z, t) \left( 1 - \lambda \Delta t + o(\Delta t) + z(\lambda \Delta t + o(\Delta t)) + \sum_{k \geq 2} z^k o(\Delta t) - 1 \right) \\ &= G(z, t)(\lambda(z - 1)\Delta t + o(\Delta t)), \end{aligned}$$

其中, 第三个等号是由于  $N(t)$  具有独立增量性, 第五个等号是由于  $N(t)$  的平稳增量性. 等式两边同时除以  $\Delta t$ , 然后令  $\Delta t \rightarrow 0$ , 得

$$\frac{\partial G(z, t)}{\partial t} = G(z, t)\lambda(z - 1).$$

考虑到概率母函数的初值  $G(z, 0) = E(z^{N(0)}) = 1$ , 可得上述微分方程的解为

$$G(z, t) = e^{\lambda t(z-1)} = e^{-\lambda t} \sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!} z^k. \quad (4.1. 1)$$

由母函数的定义可得

$$P(N(t) = k) = \frac{(\lambda t)^k}{k!} \exp(-\lambda t).$$

□

由定理 4.1 的证明过程, 可以直观理解  $\lambda$  的物理意义. 由

$$\begin{aligned} & \lim_{\Delta t \rightarrow 0} \frac{N(\Delta t)}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \frac{0 \cdot P(N(\Delta t) = 0) + 1 \cdot P(N(\Delta t) = 1) + \sum_{k \geq 2} k P(N(\Delta t) = k)}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \frac{P(N(\Delta t) = 1)}{\Delta t} = \lambda, \end{aligned}$$

可知,  $\lambda$  表示单位时间内事件发生的平均次数. 因此,  $\lambda$  通常称为**到达率 (arrival rate)** 或**强度 (intensity)**.

由 Poisson 过程的概率分布, 容易求得该过程的期望与方差为

$$E(N(t)) = D(N(t)) = \lambda t.$$

## 4.2 Poisson 过程的性质

### 4.2.1 联合分布及条件分布

根据 Poisson 过程的定义和概率分布, 可以得到其联合分布与条件分布.

**定理4.2. (联合分布)** 设  $N(t)$  是强度为  $\lambda$  的 Poisson 过程, 任意给定  $m > 0, 0 < t_1 < t_2 < \cdots < t_m, 0 \leq k_1 \leq k_2 \leq \cdots \leq k_m$ , 有

$$\begin{aligned} & P(N(t_1) = k_1, N(t_2) = k_2, \cdots, N(t_m) = k_m) \\ &= \frac{(\lambda t_1)^{k_1}}{k_1!} \frac{(\lambda(t_2 - t_1))^{k_2 - k_1}}{(k_2 - k_1)!} \cdots \frac{(\lambda(t_m - t_{m-1}))^{k_m - k_{m-1}}}{(k_m - k_{m-1})!} e^{-\lambda t_m} \end{aligned}$$

证明:

$$\begin{aligned} & P(N(t_1) = k_1, N(t_2) = k_2, \cdots, N(t_m) = k_m) \\ &= P(N(t_1) = k_1, N(t_2) - N(t_1) = k_2 - k_1, \cdots, N(t_m) - N(t_{m-1}) = k_m - k_{m-1}) \\ &= P(N(t_1) = k_1) P(N(t_2) - N(t_1) = k_2 - k_1) \cdots P(N(t_m) - N(t_{m-1}) = k_m - k_{m-1}) \\ &= \frac{(\lambda t_1)^{k_1}}{k_1!} \frac{(\lambda(t_2 - t_1))^{k_2 - k_1}}{(k_2 - k_1)!} \cdots \frac{(\lambda(t_m - t_{m-1}))^{k_m - k_{m-1}}}{(k_m - k_{m-1})!} e^{-\lambda t_m}, \end{aligned}$$

其中第二个等式成立是由于独立增量性, 第三个等式是由于平稳增量性.  $\square$

**定理4.3. (条件分布)** 设  $N(t)$  是强度为  $\lambda$  的 Poisson 过程, 当  $s < t$  时,

$$P(N(t) = m | N(s) = k) = \frac{(\lambda(t-s))^{m-k}}{(m-k)!} e^{-\lambda(t-s)}, \quad \forall m \geq k.$$

当  $s > t$  时,

$$P(N(t) = m | N(s) = k) = \frac{k!}{m!(k-m)!} \left(\frac{t}{s}\right)^m \left(1 - \frac{t}{s}\right)^{k-m}, \quad \forall m \leq k.$$

证明：当  $s < t$  时，

$$\begin{aligned}
 P(N(t) = m | N(s) = k) &= \frac{P(N(t) = m, N(s) = k)}{P(N(s) = k)} \\
 &= \frac{P(N(s) = k, N(t) - N(s) = m - k)}{P(N(s) = k)} \\
 &= \frac{P(N(s) = k)P(N(t) - N(s) = m - k)}{P(N(s) = k)} \\
 &= \frac{(\lambda(t-s))^{m-k}}{(m-k)!} e^{-\lambda(t-s)}, \quad \forall m \geq k.
 \end{aligned}$$

当  $s > t$  时，

$$\begin{aligned}
 P(N(t) = m | N(s) = k) &= \frac{P(N(t) = m, N(s) = k)}{P(N(s) = k)} \\
 &= \frac{P(N(t) = m, N(s) - N(t) = k - m)}{P(N(s) = k)} \\
 &= \frac{P(N(t) = m)P(N(s) - N(t) = k - m)}{P(N(s) = k)} \\
 &= \frac{\frac{(\lambda t)^m}{m!} e^{-\lambda t} \frac{(\lambda(s-t))^{k-m}}{(k-m)!} e^{-\lambda(s-t)}}{\frac{(\lambda s)^k}{k!} e^{-\lambda s}} \\
 &= \frac{k!}{m!(k-m)!} \left(\frac{t}{s}\right)^m \left(1 - \frac{t}{s}\right)^{k-m}, \quad \forall m \leq k.
 \end{aligned}$$

□

#### 4.2.2 到达时刻与事件间隔

到目前为止，我们仅仅讨论了给定时间下事件计数的随机特性. 事实上，事件发生的时刻也具有独特的性质.

**定理4.4.** 对于强度为  $\lambda$  的 Poisson 过程，发生第  $n$  次事件的到达时刻  $S_n$  (也称等待时间) 服从  $\Gamma$  分布，即概率密度为

$$f_{S_n}(t) = \lambda \frac{(\lambda t)^{n-1}}{(n-1)!} e^{-\lambda t}, \quad t \geq 0.$$

**证明：** 以下给出两种方法证明，核心思路都是将到达时刻有关的概率事件用计数事件的分布等价的表示出来.

**法一：** 先求解  $S_n$  的概率分布函数，有

$$F_{S_n}(t) = P(S_n \leq t) = P(N(t) \geq n) = \sum_{k=n}^{\infty} \frac{(\lambda t)^k}{k!} e^{-\lambda t}.$$

对分布函数求导，得到概率密度函数

$$f_{S_n}(t) = F'_{S_n}(t) = \sum_{k=n}^{\infty} \left( \frac{\lambda^k t^{k-1}}{(k-1)!} e^{-\lambda t} - \lambda \frac{(\lambda t)^k}{k!} e^{-\lambda t} \right) = \lambda \frac{(\lambda t)^{n-1}}{(n-1)!} e^{-\lambda t}.$$

法二：用微元分析法直接求解概率密度函数

$$\begin{aligned} f_{S_n}(t) &= \lim_{\Delta t \rightarrow 0} \frac{P(t \leq S_n \leq t + \Delta t)}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \frac{P(N(t) = n-1, N(t + \Delta t) - N(t) = 1)}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \frac{P(N(t) = n-1)P(N(t + \Delta t) - N(t) = 1)}{\Delta t} \\ &= \lambda \frac{(\lambda t)^{n-1}}{(n-1)!} e^{-\lambda t}. \end{aligned}$$

□

**定理4.5.** Poisson 过程 (强度为  $\lambda$ ) 的事件间隔  $T_k$  是独立同分布的随机变量，都服从参数为  $\lambda$  的指数分布，概率密度为

$$f_{T_k}(t) = \lambda e^{-\lambda t}, \quad t \geq 0.$$

**证明：** 我们仅对简单情形进行验证.

当  $k = 1$  时，有  $T_1 = S_1$ ，概率密度为

$$f_{T_1}(t) = f_{S_1}(t) = \lambda e^{-\lambda t}.$$

当  $k = 2$  时，有  $T_2 = S_2 - S_1$ ，考察条件概率

$$\begin{aligned} P(T_2 > t | T_1 = s) &= P(N(s+t) - N(s) = 0 | S_1 = s) \\ &= P(N(s+t) - N(s) = 0) \\ &= e^{-\lambda t}. \end{aligned}$$

可知， $T_2$  与  $T_1$  独立，且服从指数分布.

更一般的情形的证明略.

□

由概率论的知识可知，指数分布具有“无记忆性”，意味着 Poisson 过程的事件间隔与历史无关. 也就是说，从任意时刻重新开始观察这一过程，得到的新过程在概率意义上是一个与原过程“完全相同”的 Poisson 过程. 事实上，Poisson 过程的独立增量性和平稳增量性保证了这一性质.

**定理4.6.** Poisson 过程 (强度为  $\lambda$ ) 的任意  $n$  次事件的到达时刻  $S_1, S_2, \dots, S_n$  的联合概率密度函数为

$$f_{S_1, \dots, S_n}(t_1, \dots, t_n) = \lambda^n e^{-\lambda t_n}, \quad 0 \leq t_1 < \dots < t_n.$$

**证明:** 令  $\mathcal{L} = \{\Delta t_1 \rightarrow 0, \dots, \Delta t_n \rightarrow 0\}$ , 由定义

$$\begin{aligned} & f_{S_1 \dots S_n}(t_1, \dots, t_n) \\ &= \lim_{\mathcal{L}} \frac{P(t_1 \leq S_1 \leq t_1 + \Delta t_1, \dots, t_n \leq S_n \leq t_n + \Delta t_n)}{\Delta t_1 \dots \Delta t_n} \\ &= \lim_{\mathcal{L}} \frac{P(N(t_1) = 0, N(t_1 + \Delta t_1) - N(t_1) = 1, \dots, N(t_n + \Delta t_n) - N(t_n) = 1)}{\Delta t_1 \dots \Delta t_n} \\ &= \lim_{\mathcal{L}} \frac{e^{-\lambda t_1} (\lambda \Delta t_1 + o(\Delta t_1)) \dots e^{-\lambda(t_n - (t_{n-1} + \Delta t_{n-1}))} (\lambda \Delta t_n + o(\Delta t_n))}{\Delta t_1 \dots \Delta t_n} \\ &= \lambda^n e^{-\lambda t_n}. \end{aligned}$$

□

**定理4.7.** 设  $N(t)$  是强度为  $\lambda$  的 Poisson 过程. 已知  $[0, t]$  内恰好有  $n$  次事件发生, 事件到达时刻  $S_1, S_2, \dots, S_n$  的条件概率密度函数为

$$f_{S_1, \dots, S_n | N(t)}(t_1, \dots, t_n | N(t) = n) = \frac{n!}{t^n}, \quad 0 \leq t_1 < \dots < t_n \leq t.$$

**证明:** 令  $\mathcal{L} = \{\Delta t_1 \rightarrow 0, \dots, \Delta t_n \rightarrow 0\}$ , 由定义

$$\begin{aligned} & f_{S_1 \dots S_n | N(t)}(t_1, \dots, t_n | N(t) = n) \\ &= \lim_{\mathcal{L}} \frac{P(t_1 \leq S_1 \leq t_1 + \Delta t_1, \dots, t_n \leq S_n \leq t_n + \Delta t_n | N(t) = n)}{\Delta t_1 \dots \Delta t_n} \\ &= \lim_{\mathcal{L}} \frac{P(N(t_1) = 0, \dots, N(t_n + \Delta t_n) - N(t_n) = 1, N(t) = n)}{\Delta t_1 \dots \Delta t_n P(N(t) = n)} \\ &= \lim_{\mathcal{L}} \frac{P(N(t_1) = 0, \dots, N(t_n + \Delta t_n) - N(t_n) = 1, N(t) - N(t_n + \Delta t_n) = 0)}{\Delta t_1 \dots \Delta t_n P(N(t) = n)} \\ &= \lim_{\mathcal{L}} \frac{e^{-\lambda t_1} (\lambda \Delta t_1 + o(\Delta t_1)) \dots e^{-\lambda(t_n - (t_{n-1} + \Delta t_{n-1}))} (\lambda \Delta t_n + o(\Delta t_n)) e^{-\lambda(t - (t_n + \Delta t_n))}}{\Delta t_1 \dots \Delta t_n \frac{(\lambda t)^n}{n!} e^{-\lambda t}} \\ &= \frac{n!}{t^n}. \end{aligned}$$

□

到达时刻的条件概率密度与  $[0, t]$  内独立同均匀分布的随机变量序列  $U_1, \dots, U_n$  有密切的联系. 为了说明这一联系, 需要引入顺序统计量的概念.

**定义4.3.** 设  $X_1, X_2, \dots, X_n$  是  $n$  个相互独立的随机变量, 则

$$\begin{aligned} X_{(1)} &= \min\{X_1, X_2, \dots, X_n\} \\ X_{(k)} &= \min(\{X_1, X_2, \dots, X_n\} \setminus \{X_{(1)}, X_{(2)}, \dots, X_{(k-1)}\}) \\ X_{(n)} &= \max\{X_1, X_2, \dots, X_n\} \end{aligned}$$

是关于  $X_1, X_2, \dots, X_n$  的**顺序统计量 (order statistics)**.

下面推导顺序统计量的联合概率密度. 当  $x_1 \leq \dots \leq x_n$  时, 顺序统计量有非零的联合概率密度. 令  $\mathcal{L} = \{\Delta x_1 \rightarrow 0, \dots, \Delta x_n \rightarrow 0\}$ , 有

$$\begin{aligned} & f_{X_{(1)}, \dots, X_{(n)}}(x_1, \dots, x_n) \\ &= \lim_{\mathcal{L}} \frac{P(x_1 \leq X_{(1)} \leq x_1 + \Delta x_1, \dots, x_n \leq X_{(n)} \leq x_n + \Delta x_n)}{\Delta x_1 \cdots \Delta x_n} \quad (4.2. 1) \\ &= \lim_{\mathcal{L}} \frac{n! P(x_1 \leq X_1 \leq x_1 + \Delta x_1, \dots, x_n \leq X_n \leq x_n + \Delta x_n)}{\Delta x_1 \cdots \Delta x_n} \\ &= \lim_{\mathcal{L}} \frac{n! (f_{X_1}(x_1) \Delta x_1 + o(\Delta x_1)) \cdots (f_{X_n}(x_n) \Delta x_n + o(\Delta x_n))}{\Delta x_1 \cdots \Delta x_n} \\ &= n! f_{X_1}(x_1) \cdots f_{X_n}(x_n), \end{aligned}$$

其中, 式 (4.2. 1) 中分子表示  $X_1, \dots, X_n$  刚好分别落在  $n$  个小区间  $[x_1, x_1 + \Delta x_1], \dots, [x_n, x_n + \Delta x_n]$  内的概率. 由于  $X_1, \dots, X_n$  独立同分布, 它们落在  $n$  小区间的顺序有  $n!$  种可能, 且每种顺序都是等概的. 因此, 第二个等号成立.

当  $X_1, X_2, \dots, X_n$  独立同均匀分布时, 概率密度与式 4.7 完全一致. 我们得到如下结论.

**定理4.8.** 设  $N(t)$  是强度为  $\lambda$  的 Poisson 过程. 给定  $N(t) = n$ , 事件到达时刻  $S_1, S_2, \dots, S_n$  与  $n$  个独立的  $(0, t)$  上均匀分布随机变量的顺序统计量有相同的分布, 记为

$$(S_1, S_2, \dots, S_n | N(t) = n) \stackrel{d}{=} (U_{(1)}, U_{(2)}, \dots, U_{(n)}).$$

我们有如下推论.

**推论4.9.** 设  $S_1, S_2, \dots, S_n$  是 Poisson 过程的事件到达时刻, 对任意函数  $g(x)$ , 有

$$E \left( \sum_{k=1}^n g(S_k) \middle| N(t) = n \right) = n E g(U), \quad (4.2. 2)$$

其中,  $U \sim U(0, t)$ .



证明:

$$E\left(\sum_{k=1}^n g(S_k) \middle| N(t) = n\right) = E\left(\sum_{k=1}^n g(U_{(k)})\right) = E\left(\sum_{k=1}^n g(U_k)\right) = nEg(U),$$

其中, 第一个等号成立是由于推论 4.9, 第二个等号成立是由于求和运算满足交换律.  $\square$

**例4.1.** 设乘客按照参数为  $\lambda$  的 Poisson 过程来到公交站台, 公交车于  $t$  时刻发出, 求  $(0, t)$  时间段内到达乘客的等待时间总和的期望.

**解:** 设  $(0, t)$  时间段内到达的乘客数为  $N(t)$ , 其中第  $k$  个乘客的到达时刻为  $S_k$ , 则其等待时间为  $t - S_k$ . 由全期望公式可得, 等待时间总和的期望为

$$E\left(\sum_{k=1}^{N(t)} (t - S_k)\right) = E\left(E\left(\sum_{k=1}^{N(t)} (t - S_k) \middle| N(t)\right)\right), \quad (4.2. 3)$$

由推论 4.9 可得

$$E\left(\sum_{k=1}^{N(t)} (t - S_k) \middle| N(t) = n\right) = nE(t - U) = \frac{nt}{2}.$$

代入式 (4.2. 3) 得

$$E\left(\sum_{k=1}^{N(t)} (t - S_k)\right) = E\left(\frac{N(t)t}{2}\right) = \frac{\lambda t^2}{2}.$$

$\square$

事实上, 如果式 (4.2. 2) 左边的求和式替换成任意  $n$  元对称函数, 即自变量任意交换次序时, 函数值不变, 则仍可应用定理 4.8 简化计算.

### 4.3 Poisson 过程的拓广

Poisson 过程的定义涉及四个条件, 其中初始条件是平凡的, 另外三个条件起到至关重要的作用. 如果将这三个条件进行适当的放松, 可得到 Poisson 过程的各种拓广.

### 4.3.1 非齐次 Poisson 过程

首先放松平稳增量的要求, 得到非齐次 Poisson 过程. 平稳增量意味着任意时间段内发生的事件数只与时间段长度有关, 而与该时间段所处的时间轴上位置无关. 与之对应的, 非齐次则意味着任意时间段内发生的事件数不仅与时间段长度有关, 还与该时间段所处的时间轴上位置有关, 即事件到达强度随时间变化, 不再是常数  $\lambda$ . 为此, 用

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(N(t + \Delta t) - N(t) = 1)}{\Delta t}$$

表示  $t$  时刻的到达率, 即

$$P(N(t + \Delta t) - N(t) = 1) = \lambda(t)\Delta t + o(\Delta t).$$

用上式对定理 4.1 的证明作相应的修改, 可得到

$$\frac{\partial G(z, t)}{\partial t} = G(z, t)\lambda(t)(z - 1).$$

由上式可得

$$P(N(t) = k) = \frac{\left(\int_0^t \lambda(\tau) d\tau\right)^k}{k!} \exp\left(-\int_0^t \lambda(\tau) d\tau\right).$$

同理可得

$$P(N(t) - N(s) = k) = \frac{\left(\int_s^t \lambda(\tau) d\tau\right)^k}{k!} \exp\left(-\int_s^t \lambda(\tau) d\tau\right).$$

可以看到, 当  $\lambda(t) \equiv \lambda$  时, 有  $\int_0^t \lambda(\tau) d\tau = \lambda t$ , 即退化为齐次 Poisson 过程.

**例4.2. (数值记录):** 设  $\{X_n, n \in N\}$  是一列独立同分布的非负随机变量,  $X_0 = 0$ . 定义其风险率函数  $\lambda(t)$  如下

$$\lambda(t) = \frac{f(t)}{1 - F(t)}.$$

其中  $f(t)$  和  $F(t)$  分别是  $X_k$  的概率密度和分布函数. 验证随机过程

$$N(t) = \#\{n : X_n > \max(X_{n-1}, \dots, X_0), X_n \leq t\}$$

是一个强度为  $\lambda(t)$  的非齐次 Poisson 过程, 其中  $\#\mathcal{S}$  表示集合  $\mathcal{S}$  的元素个数.

解：根据定义依次验证非齐次 Poisson 过程的条件.

(1) 显然,  $N(0) = 0$ .

(2) 由于

$$N(t_2) - N(t_1) = \#\{n : X_n > \max(X_{n-1}, \dots, X_0), t_1 < X_n \leq t_2\},$$

因为  $X_n$  相互独立, 而对任意  $t_1 < t_2 \leq t_3 < t_4$ ,  $N(t_2) - N(t_1)$  与  $N(t_4) - N(t_3)$  由各不相同的  $X_n$  决定, 所以相互独立.  $N(t)$  满足独立增量性.

(3) 考察微元时间  $\Delta t \rightarrow 0$ ,

$$\begin{aligned} & P(N(t + \Delta t) - N(t) = 1) \\ &= \sum_{n \geq 1} P(t < X_n \leq t + \Delta t, X_1 \leq t, \dots, X_{n-1} \leq t) \\ &= \sum_{n \geq 1} (f(t)\Delta t + o(\Delta t))F^{n-1}(t) \\ &= \frac{f(t)}{1 - F(t)}\Delta t + o(\Delta t) = \lambda(t)\Delta t + o(\Delta t). \end{aligned}$$

而

$$\begin{aligned} & P(N(t + \Delta t) - N(t) \geq 2) \\ &= \sum_{n \geq 1} \sum_{m > n} P(t < X_n, X_m \leq t + \Delta t, X_1 \leq t, \dots, X_{n-1} \leq t, X_{n+1} \leq t, \dots, X_{m-1} \leq t) \\ &= \sum_{n \geq 1} \sum_{m > n} (f(t)\Delta t + o(\Delta t))^2 F^{m-2}(t) \\ &= \sum_{n \geq 1} (f(t)\Delta t + o(\Delta t))^2 \frac{F^{n-1}(t)}{1 - F(t)} \\ &= \frac{(f(t)\Delta t + o(\Delta t))^2}{(1 - F(t))^2} = o(\Delta t), \end{aligned}$$

所以微元时间内最多发生 1 个记录, 满足微元条件, 且强度为  $\lambda(t)$ .

综上,  $N(t)$  是一个强度为  $\lambda(t)$  的非齐次 Poisson 过程. □

#### 4.3.2 复合 Poisson 过程

放松时间微元内出现事件次数的要求, 即微元时间内可能发生两次或两次以上的事件, 便得到复合 Poisson 过程. 事实上, 我们可以从“复合”的角度重新解释这一推广. 具体来说, 微元时间内仍然最多发生一次事件, 但每次事件发生时的计数值不再限定为 1, 而是随机的. 比如, 到达路由器的数据包个数服从标准

Poisson 过程, 但每个数据包 (同一时刻到达) 的字节数不同. 更一般的, 还可以将每次的计数拓广为连续型随机变量.

**定义4.4.**  $N(t)$  为标准 Poisson 过程,  $\{Y_k, k \in \mathbb{N}\}$  为独立同分布的随机变量, 且与  $N(t)$  独立,  $Y_0 = 0$ , 则随机过程

$$Y(t) = \sum_{k=1}^{N(t)} Y_k$$

称为**复合 Poisson 过程**.

**定理4.10.** 复合 Poisson 过程  $Y(t)$  的特征函数为

$$\phi_{Y(t)}(\omega) = \exp(\lambda t(\phi_{Y_1}(\omega) - 1)).$$

**证明:**

$$\begin{aligned} \phi_{Y(t)}(\omega) &= E(\exp(j\omega Y(t))) \\ &= E\left(\exp\left(j\omega \left(\sum_{k=1}^{N(t)} Y_k\right)\right)\right) \\ &= E\left(E\left(\exp\left(j\omega \left(\sum_{k=1}^{N(t)} Y_k\right)\right) \middle| N(t)\right)\right). \end{aligned} \quad (4.3. 1)$$

由于  $Y_k$  独立同分布, 且与  $N(t)$  相互独立, 有

$$\begin{aligned} &E\left(\exp\left(j\omega \left(\sum_{k=1}^{N(t)} Y_k\right)\right) \middle| N(t) = n\right) \\ &= E\left(\exp\left(j\omega \left(\sum_{k=1}^n Y_k\right)\right)\right) \\ &= E\left(\prod_{k=1}^n \exp(j\omega Y_k)\right) \\ &= \prod_{k=1}^n E(\exp(j\omega Y_k)) = \phi_{Y_1}^n(\omega). \end{aligned}$$

代入 (4.3. 1), 得

$$\phi_{Y(t)}(\omega) = E\left(\phi_{Y_1}^{N(t)}(\omega)\right) = E\left(z^{N(t)}\right) \Big|_{z=\phi_{Y_1}(\omega)} = e^{\lambda t(\phi_{Y_1}(\omega)-1)}.$$

其中, 最后一个等式是由标准 Poisson 过程的概率母函数 (式 (4.1. 1)) 的自变量取  $z = \phi_{Y_1}(\omega)$  得到.  $\square$

在复合 Poisson 过程的研究中, 利用特征函数或母函数 (当  $Y(t)$  取非负整数值时) 往往会带来方便.

**例4.3.** 考察两个独立的标准 Poisson 过程  $N_1(t), N_2(t)$ ,

(1) 试证明  $N_1(t) + N_2(t)$  是标准 Poisson 过程 (**Poisson 过程可加性**).

(2)  $N_1(t) - N_2(t)$  是否为标准 Poisson 过程?

**解:** (1) 首先, 由于  $N_1(t) + N_2(t) \in \{0, 1, \dots\}$ , 可利用母函数证明. 由

$$\begin{aligned} G_{N_1(t)+N_2(t)}(z, t) &= E(z^{N_1(t)+N_2(t)}) \\ &= E(z^{N_1(t)}) E(z^{N_2(t)}) \\ &= e^{\lambda_1 t(z-1)} e^{\lambda_2 t(z-1)} \\ &= e^{(\lambda_1+\lambda_2)t(z-1)} \end{aligned}$$

知,  $N_1(t) + N_2(t)$  是强度为  $\lambda_1 + \lambda_2$  的 Poisson 过程.

(2) 由于  $N_1(t) - N_2(t)$  可取负值, 显然不是标准 Poisson 过程. 进一步讨论其分布, 由于特征函数为

$$\begin{aligned} \phi_{N_1(t)-N_2(t)}(\omega) &= E(e^{j\omega(N_1(t)-N_2(t))}) \\ &= E(e^{j\omega N_1(t)}) E(e^{-j\omega N_2(t)}) \\ &= E(z^{N_1(t)}) \Big|_{z=e^{j\omega}} E(z^{N_2(t)}) \Big|_{z=e^{-j\omega}} \\ &= \exp(\lambda_1 t(e^{j\omega} - 1)) \exp(\lambda_2 t(e^{-j\omega} - 1)) \\ &= \exp\left((\lambda_1 + \lambda_2)t \left(\frac{\lambda_1}{\lambda_1 + \lambda_2} e^{j\omega} + \frac{\lambda_2}{\lambda_1 + \lambda_2} e^{-j\omega} - 1\right)\right). \end{aligned}$$

由定理 4.10 及特征函数的唯一性知,

$$N_1(t) - N_2(t) = \sum_{k=1}^{N(t)} Y_k$$

是复合 Poisson 过程, 其中  $N(t)$  是强度为  $\lambda_1 + \lambda_2$  的标准 Poisson 过程,  $Y_k$  为二值分布, 满足

$$\begin{aligned} P(Y_k = 1) &= \frac{\lambda_1}{\lambda_1 + \lambda_2}, \\ P(Y_k = -1) &= \frac{\lambda_2}{\lambda_1 + \lambda_2}. \end{aligned}$$

□

容易验证, Poisson 过程的可加性可推广到多个 Poisson 过程的和的情况.

**例4.4. (随机选取不变性):**  $[0, t]$  内进入商店的顾客人数服从标准 Poisson 过程. 每次进入商店的顾客中, 男顾客出现的概率为  $p$ , 女顾客出现的概率为  $q$ ,  $p + q = 1$ . 试证明男顾客人数  $N_m(t)$  是标准 Poisson 过程.

**解:** 令  $Y_k = 1$  表示到达的第  $k$  名顾客为男,  $Y_k = 0$  表示到达的第  $k$  名顾客为女, 则

$$N_m(t) = \sum_{k=1}^{N(t)} Y_k.$$

其母函数为

$$G_{N_m(t)}(z, t) = E\left(z^{\sum_{k=1}^{N(t)} Y_k}\right) = E\left(E\left(z^{\sum_{k=1}^{N(t)} Y_k} \middle| N(t)\right)\right). \quad (4.3. 2)$$

由于

$$E\left(z^{\sum_{k=1}^{N(t)} Y_k} \middle| N(t) = n\right) = E\left(z^{\sum_{k=1}^n Y_k}\right) = (Ez^{Y_1})^n = (pz + q)^n,$$

代入式 (4.3. 2) 得

$$G_{N_m(t)}(z, t) = E\left((pz + q)^{N(t)}\right) = e^{\lambda t(pz + q - 1)} = e^{\lambda pt(z - 1)}.$$

所以  $N_m(t)$  是参数为  $\lambda p$  的标准 Poisson 过程. □

容易验证, Poisson 过程的随机选取不变性也可推广到多个分类的情况.

### 4.3.3 条件 Poisson 过程

如果放松独立增量的要求, 得到的新过程与标准 Poisson 过程有比较大的差异. 独立增量性条件放宽的方式很多, 这里考虑参数  $\lambda$  的随机化.

**定义4.5.** 令  $\Lambda$  为非负随机变量, 如果给定  $\Lambda = \lambda$  的条件下,  $N(t)$  是参数为  $\lambda$  的 Poisson 过程, 则称  $N(t)$  为**条件 Poisson 过程** (也称**随机参数 Poisson 过程**).

条件 Poisson 过程的概率分布由下面定理给出.

**定理4.11.** 条件 Poisson 过程的参数  $\Lambda$  为连续随机变量时 (概率密度为  $f(\lambda)$ ),  $N(t)$  的概率分布为

$$P(N(t) = n) = \int_0^\infty \frac{(\lambda t)^n}{n!} \exp(-\lambda t) f(\lambda) d\lambda.$$

**证明：** 对  $\Lambda$  的所有可能取值做全概率展开，

$$\begin{aligned}
 P(N(t) = n) &= \lim_{\Delta\lambda_k \rightarrow 0} \sum_k P(\lambda_k < \Lambda < \lambda_k + \Delta\lambda_k) P(N(t) = n | \Lambda = \lambda_k) \\
 &= \lim_{\Delta\lambda_k \rightarrow 0} \sum_k (f(\lambda_k) \Delta\lambda_k + o(\Delta\lambda_k)) P(N(t) = n | \Lambda = \lambda_k) \\
 &= \int_0^\infty \frac{(\lambda t)^n}{n!} \exp(-\lambda t) f(\lambda) d\lambda. \quad \square
 \end{aligned}$$

由定理 4.10 容易验证条件 Poisson 过程不满足独立增量性. 由于

$$\begin{aligned}
 &P(N(s) = m, N(t+s) - N(s) = n) \\
 &= \int_0^\infty P(N(s) = m, N(t+s) - N(s) = n | \Lambda = \lambda) f(\lambda) d\lambda \\
 &= \int_0^\infty P(N(s) = m | \Lambda = \lambda) P(N(t+s) - N(s) = n | \Lambda = \lambda) f(\lambda) d\lambda \\
 &= \int_0^\infty \frac{(\lambda s)^m}{m!} \exp(-\lambda s) \frac{(\lambda t)^n}{n!} \exp(-\lambda t) f(\lambda) d\lambda.
 \end{aligned}$$

于是，

$$\begin{aligned}
 &P(N(s) = m) P(N(t+s) - N(s) = n) \\
 &= \int_0^\infty \frac{(\lambda s)^m}{m!} \exp(-\lambda s) f(\lambda) d\lambda \int_0^\infty \frac{(\lambda t)^n}{n!} \exp(-\lambda t) f(\lambda) d\lambda \\
 &\neq P(N(s) = m, N(t+s) - N(s) = n).
 \end{aligned}$$

**例4.5.** 某地区每年下雨次数服从 Poisson 过程，假定下雨的持续时间可以忽略. 由于一些未知因素的影响，下雨的强度年年不同，设为随机变量  $\Lambda$ . 希望通过对今年时刻  $t$  以前下雨次数的统计，对下雨强度作出推断，并预测何时再次下雨.

**解：** 令  $N(t)$  表示今年到时刻  $t$  以前的下雨次数. 给定  $N(t)$  的条件下， $\Lambda$  的概率密度为

$$\begin{aligned}
 f_{\Lambda|N(t)}(\lambda|n) &= \lim_{\Delta\lambda \rightarrow 0} \frac{P(\lambda \leq \Lambda \leq \lambda + \Delta\lambda | N(t) = n)}{\Delta\lambda} \\
 &= \lim_{\Delta\lambda \rightarrow 0} \frac{P(\lambda \leq \Lambda \leq \lambda + \Delta\lambda) P(N(t) = n | \Lambda = \lambda)}{\Delta\lambda P(N(t) = n)} \\
 &= f(\lambda) \frac{\frac{(\lambda t)^n}{n!} \exp(-\lambda t)}{\int_0^\infty \frac{(\lambda t)^n}{n!} \exp(-\lambda t) f(\lambda) d\lambda} \\
 &= \frac{\lambda^n \exp(-\lambda t) f(\lambda)}{\int_0^\infty \lambda^n \exp(-\lambda t) f(\lambda) d\lambda}.
 \end{aligned}$$

设  $T$  表示从时刻  $t$  开始到下次下雨的间隔时长. 给定  $N(t)$  的条件下,  $T$  的概率密度为

$$\begin{aligned}
 f_{T|N(t)}(s|n) &= \lim_{\Delta s \rightarrow 0} \frac{P(s < T \leq s + \Delta s | N(t) = n)}{\Delta s} \\
 &= \lim_{\Delta s \rightarrow 0} \frac{P(s < T \leq s + \Delta s, N(t) = n)}{\Delta s P(N(t) = n)} \\
 &= \lim_{\Delta s \rightarrow 0} \frac{\int_0^\infty P(s < T \leq s + \Delta s, N(t) = n | \Lambda = \lambda) f(\lambda) d\lambda}{\Delta s \int_0^\infty \frac{(\lambda t)^n}{n!} \exp(-\lambda t) f(\lambda) d\lambda} \\
 &= \lim_{\Delta s \rightarrow 0} \frac{\int_0^\infty P(s < T \leq s + \Delta s | \Lambda = \lambda) P(N(t) = n | \Lambda = \lambda) f(\lambda) d\lambda}{\Delta s \int_0^\infty \frac{(\lambda t)^n}{n!} \exp(-\lambda t) f(\lambda) d\lambda} \\
 &= \frac{\int_0^\infty \lambda \exp(-\lambda s) \frac{(\lambda t)^n}{n!} \exp(-\lambda t) f(\lambda) d\lambda}{\int_0^\infty \frac{(\lambda t)^n}{n!} \exp(-\lambda t) f(\lambda) d\lambda}. \quad \square
 \end{aligned}$$

#### 4.3.4 更新过程

Poisson 过程的一个重要性质是事件发生的时间间隔服从指数分布. 如果放宽这一条件, 即假设事件间隔独立同分布, 但不再限定是指数分布, 便得到更新过程.

**定义4.6.** 如果计数过程  $N(t), t \geq 0$  的事件间隔  $T_k, k \in \mathbb{N}$  为独立同分布的随机变量, 则称  $N(t)$  为**更新过程**.

前面提到过, 标准 Poisson 过程从任意时刻开始观察, 仍然是一个标准 Poisson 过程, 即从任意时刻重新开始. 这是由于指数分布的事件间隔满足无记忆性导致的. 更新过程则不然, 只保证该过程从各个事件发生时刻开始观察, 是概率意义上相同的随机过程, 即在各个事件发生时刻重新开始. 从下面的例子可以看到, 更新过程一般并不具有平稳增量性和独立增量性.

**例4.6.** 设一个更新过程  $N(t)$  的事件发生间隔服从  $T_k \sim U(1, 2)$ ,

(1) 计算  $P(N(1) - N(0) = 1)$  和  $P(N(2) - N(1) = 1)$ .

(2) 考虑  $N(1.1) - N(1)$  和  $N(2) - N(1.1)$  的独立性.

**解:** (1) 由定义易得

$$P(N(1) - N(0) = 1) = 0, \quad P(N(2) - N(1) = 1) = 1.$$

(2) 由于

$$P(N(2) - N(1.1) = 1 | N(1.1) - N(1) = 1) = 0,$$



而显然

$$P(N(2) - N(1.1) = 1) > 0,$$

于是

$$P(N(2) - N(1.1) = 1 | N(1.1) - N(1) = 1) \neq P(N(2) - N(1.1) = 1),$$

所以,  $N(1.1) - N(1)$  和  $N(2) - N(1.1)$  不独立. □

给定事件发生间隔的分布之后, 可以计算更新过程的分布.

**命题4.12.** 设更新过程时间间隔的概率分布函数和概率密度函数分别为  $F_T(t)$  和  $f_T(t)$ ,  $S_n = \sum_{k=1}^n T_k$  为第  $n$  次事件的发生时刻, 则有

$$P(N(t) = n) = F_{S_n}(t) - F_{S_{n+1}}(t),$$

$$E(N(t)) = m_N(t) = \sum_{n=1}^{\infty} F_{S_n}(t),$$

其中  $S_n$  的密度是  $f_T(x)$  的  $n$  重卷积.

**证明:** 直接计算可得

$$\begin{aligned} P(N(t) = n) &= P(N(t) \geq n) - P(N(t) \geq n+1) \\ &= P(S_n \leq t) - P(S_{n+1} \leq t) \\ &= F_{S_n}(t) - F_{S_{n+1}}(t). \end{aligned}$$

$$\begin{aligned} E(N(t)) &= \sum_{n=1}^{\infty} nP(N(t) = n) \\ &= P(N(t) = 1) + P(N(t) = 2) + P(N(t) = 3) + \cdots \\ &\quad + P(N(t) = 2) + P(N(t) = 3) + \cdots \\ &\quad + P(N(t) = 3) + \cdots \\ &= P(N(t) \geq 1) + P(N(t) \geq 2) + \cdots \\ &= P(S_1 \leq t) + P(S_2 \leq t) + \cdots \\ &= \sum_{n=1}^{\infty} F_{S_n}(t). \end{aligned} \quad \square$$

类似于非齐次 Poisson 过程, 对于更新过程可以定义更新强度为

$$\lambda_N(t) = \frac{d}{dt} m_N(t).$$

我们不加证明地指出, 更新强度满足以下更新方程

$$f_T(t) = \lambda_N(t) - \int_0^t \lambda_N(t-\tau) f_T(\tau) d\tau.$$

于是, 理论上可利用事件间隔的分布计算更新强度, 反之也.

**例4.7. (Poisson 过程)** 设更新过程的更新强度满足

$$\lambda_N(t) = \begin{cases} \lambda, & t \geq 0, \\ 0, & t < 0. \end{cases}$$

计算该过程的事件间隔分布.

**解:** 由更新方程, 可得

$$f_T(t) = F'_T(t) = \lambda - \lambda F_T(t).$$

求解得到

$$F_T(t) = 1 - e^{-\lambda t},$$

即事件间隔服从指数分布. 所以, 更新强度为常数时, 更新过程即为 Poisson 过程.  $\square$

对于计数过程, 人们往往关注计数量的长时变化速率. 事实上, 由于

$$S_{N(t)} \leq t \leq S_{N(t)+1},$$

则有

$$\frac{S_{N(t)}}{N(t)} \leq \frac{t}{N(t)} \leq \frac{S_{N(t)+1}}{N(t)}.$$

从而

$$\frac{S_{N(t)}}{N(t)} \leq \frac{t}{N(t)} \leq \frac{N(t)+1}{N(t)} \frac{S_{N(t)+1}}{N(t)+1}. \quad (4.3. 3)$$

由大数定律可知

$$\frac{S_{N(t)}}{N(t)} \rightarrow \mu, \quad t \rightarrow \infty,$$

其中,  $\mu$  为事件间隔的期望. 上式代入 (4.3. 3), 令  $t \rightarrow \infty$ , 可得

$$\frac{N(t)}{t} \rightarrow \frac{1}{\mu}.$$

更进一步的, 我们有以下结论.

**定理4.13. (基本更新定理, elementary renewal theorem)** 设  $N(t)$  是更新过程, 事件间隔的期望为  $\mu$ , 则有

$$\lim_{t \rightarrow \infty} \frac{m_N(t)}{t} = \frac{1}{\mu}.$$

**证明:** 略. □

**定理4.14. (关键更新定理)** 设  $N(t)$  是更新过程, 事件间隔的期望为  $\mu$ , 则有

$$\lim_{t \rightarrow \infty} m_N(t+a) - m_N(t) = \frac{a}{\mu},$$

其中  $a$  为给定参数.

**证明:** 略. □

可以看到, 时间足够长时, 一段时间内发生的平均更新次数与时间长度成正比关系.

**例4.8.** 考虑一个服务员的银行柜台, 顾客的到达服从 Poisson 过程, 参数为  $\lambda$ , 如果顾客到达时柜台不空, 则顾客会立刻离开; 如果柜台空闲, 则顾客进入银行接受服务. 服务时间  $G$  服从任意给定分布, 均值为  $\mu_G$ . 现求进入银行的顾客的平均速率、以及进入银行的顾客占有所有到达银行顾客的平均比率.

**解:** 由于顾客到达的间隔服从指数分布, 具有无记忆性, 所以, 两个相邻进入银行的顾客的时间间隔包括前一个顾客的服务时间, 以及后一个顾客的到达间隔. 则平均间隔为

$$\mu = \mu_G + \frac{1}{\lambda}.$$

由基本更新定理可得, 进入银行的顾客的平均速率为

$$\frac{1}{\mu} = \frac{\lambda}{1 + \lambda\mu_G}.$$

由于顾客总到达率为  $\lambda$ , 则进入银行的顾客占有所有到达银行顾客的平均比率为

$$\frac{\frac{1}{\mu}}{\lambda} = \frac{1}{1 + \lambda\mu_G}.$$

□

## 第 5 章 连续时间 Markov 链

从概念上来说, 连续时间 Markov 链是离散时间 Markov 链从离散参数到连续参数的扩展, 但理论分析与实际应用中有很大差异. 本章重点讨论连续时间 Markov 链与离散时间 Markov 链的区别与联系, 并对其最典型的应用之一, 即排队理论加以介绍.

### 5.1 定义与例子

**定义5.1.** 设具有可数样本空间  $\mathcal{E}$  的连续时间随机过程  $X(t)$ ,  $\forall k \in \mathcal{N}$ ,  $t, s, s_1, s_2, \dots, s_k$ ,  $s > s_k \geq s_{k-1} \geq \dots \geq s_1 > 0$ , 满足

$$\begin{aligned} &P(X(t+s) = j | X(s) = i, X(s_k) = i_k, \dots, X(s_1) = i_1) \\ &= P(X(t+s) = j | X(s) = i), \end{aligned}$$

则称  $\{X(t)\}$  为**连续时间 Markov 链**.

**定义5.2.** 连续时间 Markov 链是**齐次的**, 如果对  $\forall s \geq 0$ , 都有

$$P(X(t+s) = j | X(s) = i) = P(X(t) = j | X(0) = i) = p_{ij}(t).$$

如不特别指出, 本章所讨论的均为齐次 Markov 链. 区别于离散时间 Markov 链状态间的跳转只发生在离散时刻, 连续时间 Markov 链的状态跳转可能发生在任意时刻. 图 5.1 展示了一个典型的连续时间 Markov 链的曲线. 对于连续时间 Markov 链的研究, 不仅需要关注于状态之间的**跳变**, 还需要关注在任一状态的**停留时间**, 这两者组成连续时间 Markov 链的特性行为.

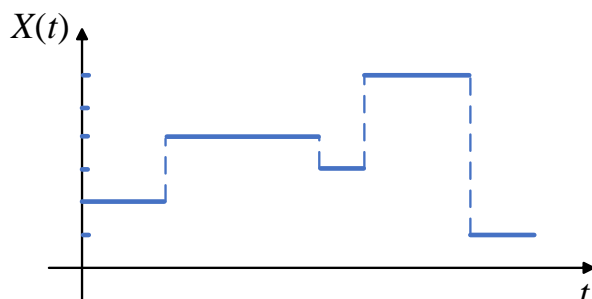


图 5.1 连续时间 Markov 链示例

**定义5.3.** 齐次 Markov 链的转移概率矩阵为

$$\mathbf{P}(t) = (p_{ij}(t))_{i,j \in \mathcal{E}} = \begin{pmatrix} p_{00}(t) & p_{01}(t) & \cdots \\ p_{10}(t) & p_{11}(t) & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix},$$

其中  $p_{ij}(t) \geq 0, \sum_j p_{ij}(t) = 1$ .

类似于离散的情况, 连续时间 Markov 链也满足 C-K 方程, 即

$$p_{ij}(t+s) = \sum_{k \in \mathcal{E}} p_{ik}(t)p_{kj}(s),$$

矩阵形式为

$$\mathbf{P}(t+s) = \mathbf{P}(t)\mathbf{P}(s).$$

**例5.1.** 验证 Poisson 过程  $N(t)$  是一个连续时间 Markov 链, 并求转移概率矩阵.

**解:** Markov 性由 Poisson 过程的平稳增量性和独立增量性可直接验证. 根据 Poisson 过程的条件分布, 即

$$p_{ij}(t) = P(N(s+t) = j | N(s) = i) = \frac{(\lambda t)^{j-i}}{(j-i)!} e^{-\lambda t}, \quad \forall j \geq i,$$

$$p_{ij}(t) = P(N(s+t) = j | N(s) = i) = 0, \quad \forall j < i,$$

可以写出转移概率矩阵为

$$\mathbf{P}(t) = \begin{pmatrix} e^{-\lambda t} & \lambda t e^{-\lambda t} & \frac{(\lambda t)^2}{2!} e^{-\lambda t} & \cdots \\ & e^{-\lambda t} & \lambda t e^{-\lambda t} & \cdots \\ & & e^{-\lambda t} & \cdots \\ & & & \ddots \end{pmatrix}. \quad \square$$

**例5.2.** 验证随机电报过程  $X(t) = X(0)(-1)^{N(t)}$  是一个连续时间 Markov 链, 并计算其转移概率矩阵. 其中  $X(0) \in \{-1, 1\}$  与 Poisson 过程  $N(t)$  独立.

**解:** 由于

$$\begin{aligned} X(t+s) &= X(0)(-1)^{N(t+s)} \\ &= X(0)(-1)^{N(t+s)-N(s)+N(s)} \\ &= X(0)(-1)^{N(t+s)-N(s)}(-1)^{N(s)} \\ &= X(s)(-1)^{N(t+s)-N(s)}, \end{aligned}$$

所以给定当前状态  $X(s)$  后,  $X(t+s)$  仅依赖于  $N(t+s) - N(s)$ , 而后者又仅依赖于区间长度  $t$ , 而与任意  $s' < s$  时刻的状态无关, 因此是连续时间 Markov 链. 由

$$\begin{aligned}
 P(X(t+s) = 1 | X(s) = -1) &= P(X(s)(-1)^{N(t+s)-N(s)} = 1 | X(s) = -1) \\
 &= P((-1)^{N(t+s)-N(s)} = -1) \\
 &= \sum_{k=1}^{\infty} P(N(t+s) - N(s) = 2k-1) \\
 &= \sum_{n=0}^{\infty} \frac{1 - (-1)^n}{2} P(N(t+s) - N(s) = n) \\
 &= \sum_{n=0}^{\infty} \frac{1}{2} \frac{(\lambda t)^n}{n!} e^{-\lambda t} - \sum_{n=0}^{\infty} \frac{1}{2} \frac{(-\lambda t)^n}{n!} e^{-\lambda t} \\
 &= \frac{1}{2} - \frac{1}{2} e^{-2\lambda t}.
 \end{aligned}$$

同理可得

$$P(X(t+s) = 1 | X(s) = 1) = \frac{1}{2} + \frac{1}{2} e^{-2\lambda t}.$$

于是转移概率矩阵为

$$\mathbf{P}(t) = \frac{1}{2} \begin{pmatrix} 1 + e^{-2\lambda t} & 1 - e^{-2\lambda t} \\ 1 - e^{-2\lambda t} & 1 + e^{-2\lambda t} \end{pmatrix}. \quad \square$$

区别于离散情况的是, 由于时间参数的连续性, 根据 C-K 方程并不能找到类似一步转移概率这样的基本单元, 可以用来计算所有的转移概率. 为此, 我们借助连续可微函数的导数, 引入  $\mathbf{Q}$  矩阵的概念.

## 5.2 $\mathbf{Q}$ 矩阵

### 5.2.1 $\mathbf{Q}$ 矩阵的物理意义

**定理5.1.** 设  $\mathbf{P}(t)$  所描述的连续时间 Markov 链是标准的, 即

$$\lim_{t \rightarrow 0} \mathbf{P}(t) = \mathbf{I},$$

则在广义意义下存在极限

$$\mathbf{Q} = (q_{ij}) = \lim_{t \rightarrow 0^+} \frac{\mathbf{P}(t) - \mathbf{I}}{t},$$

其中  $\mathbf{Q}$  为转移率矩阵, 元素  $q_{ij}$  为从状态  $i$  到  $j$  的转移率, 满足

$$0 \leq q_i = -q_{ii} \leq \infty, \quad q_{ij} \geq 0, \forall i \neq j, \quad \sum_j q_{ij} \leq 0.$$

证明：略。

□

**注5.1.** Markov 链的标准性意味着，在极短时间内不太可能发生状态跳转，这是符合物理系统的实际约束的。我们可以认为状态间的跳转是需要消耗一定能量的，极短时间内完成跳转意味着需要极大的功率，这是物理不可实现的。

**注5.2.** 简单来说，转移率矩阵就是转移概率矩阵在 0 点的导数，具有通常的“速率”的意义。根据  $Q$  矩阵的定义，当  $i \neq j$  时，

$$q_{ij} = \lim_{t \rightarrow 0^+} \frac{p_{ij}(t)}{t} = p'_{ij}(t)|_{t=0}$$

表示单位时间从状态  $i$  转移到状态  $j$  的概率，

$$q_i = -q_{ii} = \lim_{t \rightarrow 0^+} \frac{1 - p_{ii}(t)}{t} = -p'_{ii}(t)|_{t=0}$$

表示单位时间从状态  $i$  转出的概率。直观上，似乎应得到  $\sum_{i \neq j} q_{ij} = q_i$  这样的结论，但事实上并非如此。原因在于，由定义有

$$\sum_j q_{ij} = \sum_{j \neq i} \lim_{t \rightarrow 0^+} \frac{p_{ij}(t)}{t} + \lim_{t \rightarrow 0^+} \frac{p_{ii}(t) - 1}{t}. \quad (5.2. 1)$$

而转移概率满足行和为 1，即

$$\sum_j p_{ij}(t) = 1.$$

如果式 (5.2. 1) 中求和与取极限可交换次序，但得到与直观相符的结果。但这需要满足一定的条件，一般来说只能得到  $\sum_{i \neq j} q_{ij} \leq q_i$ 。当 Markov 链满足保守性时，上述式子取得等号。

**定义5.4.** 若  $Q$  矩阵满足  $\forall i \in E$ ，都有

$$\sum_{j \neq i} q_{ij} = -q_{ii} < \infty,$$

则称该链是保守的 (Conservative)。

显然，状态空间有限的 Markov 链必然是保守的，此时转移率矩阵满足行和为 0 这一性质。

**例5.3.** 求 Poisson 过程  $N(t)$  的转移率矩阵。

解：由定义及例 5.1 的结果可得

$$Q = P'(t)|_{t=0} = \begin{pmatrix} -\lambda & \lambda & & \\ & -\lambda & \lambda & \\ & & \ddots & \ddots \\ & & & \ddots \end{pmatrix}. \quad \square$$

例5.4. 求随机电报过程  $X(t)$  的转移率矩阵.

解：由定义及例 5.2 的结果可得

$$Q = P'(t)|_{t=0} = \begin{pmatrix} -\lambda & \lambda \\ \lambda & -\lambda \end{pmatrix}. \quad \square$$

在满足标准性和保守性的条件下，可以得到丰富的结论.

**定理5.2.** 设连续时间 Markov 链  $X(t)$  满足标准性和保守性， $\tau = \inf\{t : X(t) \neq X(0)\}$  为停留时间. 在  $X(0) = i$  的条件下，

- (1)  $\tau$  服从参数为  $q_i$  的指数分布;
- (2)  $X(\tau)$  的条件分布为  $P(X(\tau) = j | X(0) = i) = \frac{q_{ij}}{q_i}$ ;
- (3)  $\tau$  与  $X(\tau)$  条件独立.

**证明：** 这里仅证明第 (1) 条结论. 令

$$g_i(t) = P(\tau > t | X(0) = i).$$

由于

$$\begin{aligned} g_i(s+t) &= P(\tau > s+t | X(0) = i) \\ &= P(X(u) = i, \forall 0 < u \leq s+t | X(0) = i) \\ &= P(X(u) = i, \forall 0 < u \leq s, X(v) = i, \forall s < v \leq s+t | X(0) = i) \\ &= P(X(u) = i, \forall 0 < u \leq s | X(0) = i) \\ &\quad P(X(v) = i, \forall s < v \leq s+t | X(u) = i, \forall 0 < u \leq s, X(0) = i) \\ &= P(\tau > s | X(0) = i) P(X(v) = i, \forall s < v \leq s+t | X(s) = i) \\ &= P(\tau > s | X(0) = i) P(\tau > t | X(0) = i) \\ &= g_i(s) g_i(t), \end{aligned}$$

所以  $g_i(t)$  必为指数函数，设

$$g_i(t) = e^{-\alpha_i t},$$



当  $t \rightarrow 0^+$  时,

$$g_i(t) = p_{ii}(t) = 1 + q_{ii}t + o(t) = 1 - \alpha_i t + o(t).$$

于是,  $\alpha_i = -q_{ii} = q_i$ , 得证.  $\square$

**定理5.3.** 设  $\tau_n$  为  $X(t)$  的第  $n$  次跳转时刻, 则  $X_n = X(\tau_n)$  是离散时间 Markov 链, 称为  $X(t)$  的嵌入链, 转移概率为

$$p_{ij} = \begin{cases} \frac{q_{ij}}{q_i}, & j \neq i \\ 0, & j = i \end{cases}$$

证明: 略.  $\square$

## 5.2.2 Kolmogorov 向前-向后方程

实际问题中, 转移概率矩阵往往难以求得. 相对的, 转移率矩阵则可根据条件分析得到. 因此, 如何利用转移率矩阵求解转移概率矩阵, 是连续时间 Markov 链的一个基本问题. Kolmogorov 向前-向后方程就是解决这一问题的有力工具.

**定理5.4.** 设 Markov 链满足标准性和保守性, 则有

(1) Kolmogorov 向前方程

$$\frac{d}{dt}P(t) = P(t)Q.$$

(2) Kolmogorov 向后方程

$$\frac{d}{dt}P(t) = QP(t).$$

(3) 转移概率可表示为收敛级数

$$P(t) = e^{tQ} = I + \sum_{k=1}^{\infty} \frac{t^k}{k!} Q^k. \quad (5.2. 2)$$

(4) 瞬时分布  $p(t) = (p_0(t), p_1(t), \dots)$ ,  $p_i(t) = P(X(t) = i)$  满足 Fokker-Planck 方程

$$\frac{d}{dt}p(t) = p(t)Q.$$

证明: (1) 根据求导的定义, 有

$$\begin{aligned} \frac{d}{dt}P(t) &= \lim_{\Delta t \rightarrow 0} \frac{P(t + \Delta t) - P(t)}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \frac{P(t)P(\Delta t) - P(t)}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0} P(t) \frac{P(\Delta t) - I}{\Delta t} \\ &= P(t)Q, \end{aligned}$$

其中第二个等号是由 C-K 方程得到, 最后一个等号是  $Q$  矩阵的定义.

(2) 向后方程的证明与向对方程类似, 只需在应用 C-K 方程时, 将  $t$  和  $\Delta t$  互换位置即可.

(3) 这一结果是微分方程 (向前或向后方程) 解的矩阵形式.

(4) 类似的, 有

$$\begin{aligned}\frac{d}{dt}\mathbf{p}(t) &= \lim_{\Delta t \rightarrow 0} \frac{\mathbf{p}(t + \Delta t) - \mathbf{p}(t)}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \frac{\mathbf{p}(t)\mathbf{P}(\Delta t) - \mathbf{p}(t)}{\Delta t} \\ &= \mathbf{p}(t)\mathbf{Q}.\end{aligned}\quad \square$$

**例5.5. (机器维修问题)** 设机器的正常工作时间服从参数为  $\lambda$  的指数分布, 一旦机器损坏则立刻进行修理, 修理时间服从参数为  $\mu$  的指数分布. 如果在起始时刻 0 机器处于正常工作状态, 计算在时刻  $T$  机器处于工作状态的概率.

**解:** 设  $t$  时刻机器工作的状态为  $X_t$ ,  $X_t = 0$  表示正常工作,  $X_t = 1$  表示修理状态. 则所求概率为  $P(X(T) = 0 | X(0) = 0) = p_{00}(T)$ . 分析考察  $Q$  矩阵的各个元素,

$$\begin{aligned}q_{01} &= \lim_{t \rightarrow 0} \frac{P(X(t) = 1 | X(0) = 0)}{t} \\ &= \lim_{t \rightarrow 0} \frac{P(\tau_0 < t)}{t} \\ &= \lim_{t \rightarrow 0} \frac{1 - e^{-\lambda t}}{t} \\ &= \lim_{t \rightarrow 0} \frac{1 - (1 - \lambda t + o(t))}{t} \\ &= \lambda,\end{aligned}$$

其中,  $\tau_0$  表示在 0 状态停留的时长. 根据保守性, 可得  $q_{00} = -\lambda$ . 类似的, 有  $q_{10} = \mu, q_{11} = -\mu$ . 则转移率矩阵为

$$\mathbf{Q} = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}.$$

由  $|s\mathbf{I} - \mathbf{Q}| = 0$ , 解得  $Q$  矩阵的特征根为

$$s_1 = 0, \quad s_2 = -(\lambda + \mu).$$

再由  $\mathbf{Q}\mathbf{v} = s\mathbf{v}$ , 解得相应的特征向量分别为

$$\mathbf{v}_1 = (1, 1)^T, \quad \mathbf{v}_2 = (-1, \frac{\mu}{\lambda})^T.$$

于是,  $Q$  矩阵可作如下特征分解

$$Q = V \begin{pmatrix} 0 & 0 \\ 0 & -(\lambda + \mu) \end{pmatrix} V^{-1},$$

其中,  $V = (v_1, v_2)$ . 根据式 (5.2. 2) 可得

$$P(t) = V \begin{pmatrix} 1 & 0 \\ 0 & e^{-(\lambda+\mu)t} \end{pmatrix} V^{-1} = \frac{1}{\lambda + \mu} \begin{pmatrix} \mu + \lambda e^{-(\lambda+\mu)t} & \lambda - \lambda e^{-(\lambda+\mu)t} \\ \mu - \mu e^{-(\lambda+\mu)t} & \lambda + \mu e^{-(\lambda+\mu)t} \end{pmatrix}.$$

于是, 所求概率为  $p_{00}(T) = \frac{\mu}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu} e^{-(\lambda+\mu)t}$ .  $\square$

### 5.3 转移概率的极限行为

在讨论连续时间 Markov 链的极限行为时, 重点考虑与离散情形的不同之处. 根据定理 5.2, Markov 链在某一状态的“停留”行为与状态间的“跳转”行为是相互独立的. 由于停留时间服从指数分布, 那么如果对某一个  $t$ , 满足  $p_{ij}(t) > 0$ , 则  $p_{ij}(t) > 0$  对任意  $t$  成立. 因此, 连续时间 Markov 链不存在周期性, 其极限行为依赖于不可约性. 我们给出如下定义.

**定义5.5.**  $\forall i, j$ , 存在  $t$ , 使得  $P_{ij}(t) > 0$ , 那么称连续时间 Markov 链  $X(t)$  是不可约的.

根据定理 5.2 和定理 5.3 知, 连续时间 Markov 链不可约等价于其嵌入链不可约. 如果一个链是不可约的, 则其极限行为与平稳分布之间有着密切联系. 连续时间 Markov 链的平稳分布可类似于离散情形进行定义.

**定义5.6.** 连续时间 Markov 链转移概率矩阵为  $P(t)$ , 如果概率分布行向量  $\pi$  满足

$$\pi = \pi P(t), \quad (5.3. 1)$$

则称为 Markov 链的平稳分布 (不变分布).

区别于离散情形, 式 (5.3. 1) 含有时间变量  $t$ , 该方程组的解似乎也会依赖于  $t$ , 这与不变分布的定义相悖. 事实上, 这种情况并不会出现.

**定理5.5.** 连续时间 Markov 链满足标准性和保守性, 转移率矩阵为  $Q$ . 概率分布行向量  $\pi$  是该链的平稳分布的充要条件是

$$\pi Q = 0. \quad (5.3. 2)$$

证明：略。

□

由于连续时间 Markov 链不存在周期性，其转移概率的极限与平稳分布的关系相对于离散情形变得简单了。其结果总结如下。

**定理5.6.** 设不可约连续时间 Markov 链  $X(t)$  满足标准性和保守性，

(1) 如果存在平稳分布  $\pi$ ，那么

$$\lim_{t \rightarrow \infty} P_{ij}(t) = \pi_j, \quad \forall i, j$$

(2) 如果不存在平稳分布，那么

$$\lim_{t \rightarrow \infty} P_{ij}(t) = 0, \quad \forall i, j$$

(3) 瞬时分布  $\lim_{t \rightarrow \infty} p_i(t)$  与转移概率具有相同的极限。

证明：略。

□

根据以上定理，连续时间 Markov 链的极限问题归结为线性方程组是否有解的问题，这提供了一条实际可行的判定思路。

**例5.6.** 求机器维修问题 (例 5.5) 的转移概率极限。

**解：** 由于转移概率已求出，直接取极限得

$$\lim_{t \rightarrow \infty} \mathbf{P}(t) = \begin{pmatrix} \frac{\mu}{\lambda + \mu} & \frac{\lambda}{\lambda + \mu} \\ \frac{\mu}{\lambda + \mu} & \frac{\lambda}{\lambda + \mu} \end{pmatrix}.$$

上述结果亦可通过求解平衡方程 (5.3. 2) 得到。

□

**例5.7. (纯生过程)：** 连续时间 Markov 链  $N(t)$  的状态空间为  $\mathcal{E} = \{0, 1, 2, \dots\}$ ，在时间段  $[t, t + \Delta t]$  内的转移概率满足

$$P(N(t + \Delta t) = j | N(t) = i) = \begin{cases} \lambda_i \Delta t + o(\Delta t), & j = i + 1 \\ o(\Delta t), & j \geq i + 2 \\ 0, & j < i \end{cases}$$

讨论其极限分布。

解：首先分析该链的  $Q$  矩阵. 由定义, 当  $i \neq j$  时,

$$q_{ij} = \lim_{\Delta t \rightarrow 0} \frac{P(N(t + \Delta t) = j | N(t) = i)}{\Delta t} = \begin{cases} \lambda_i, & j = i + 1 \\ 0, & \text{其他} \end{cases}$$

当  $i = j$  时, 由保守性得

$$q_{ii} = -\lambda_i.$$

于是转移率矩阵为

$$Q = \begin{pmatrix} -\lambda_0 & \lambda_0 & & \\ & -\lambda_1 & \lambda_1 & \\ & & \ddots & \ddots \end{pmatrix}.$$

易验证平衡方程 (5.3. 2) 无解, 因此极限分布为 0.  $\square$

与例 5.3 的结果进行比较, 不难发现, 纯生过程是 Poisson 过程的某种拓广, 其强度随着状态发生变化.

**例5.8. (生灭过程):** 连续时间 Markov 链  $N(t)$  的状态空间为  $\mathcal{E} = \{0, 1, 2, \dots\}$ , 在时间段  $[t, t + \Delta t]$  内的转移概率满足

$$P(N(t + \Delta t) = j | N(t) = i) = \begin{cases} \lambda_i \Delta t + o(\Delta t), & j = i + 1 \\ \mu_i \Delta t + o(\Delta t), & j = i - 1 \geq 0 \\ o(\Delta t), & |j - i| \geq 2 \end{cases}$$

讨论其极限分布.

解：与例 5.7 类似进行分析, 可以得到转移率矩阵

$$Q = \begin{pmatrix} -\lambda_0 & \lambda_0 & & \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & \\ & \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 \\ & & \ddots & \ddots & \ddots \end{pmatrix}.$$

由平衡方程 (5.3. 2) 得

$$\begin{aligned} -\lambda_0 \pi_0 + \mu_1 \pi_1 &= 0, \\ \lambda_0 \pi_0 - (\lambda_1 + \mu_1) \pi_1 + \mu_2 \pi_2 &= 0, \\ \dots \\ \lambda_{k-1} \pi_{k-1} - (\lambda_k + \mu_k) \pi_k + \mu_{k+1} \pi_{k+1} &= 0, \\ \dots \end{aligned}$$

解得

$$\begin{aligned}\pi_1 &= \frac{\lambda_0}{\mu_1} \pi_0, \\ \pi_2 &= \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2} \pi_0, \\ &\dots \\ \pi_k &= \frac{\lambda_0 \lambda_1 \cdots \lambda_{k-1}}{\mu_1 \mu_2 \cdots \mu_k} \pi_0, \\ &\dots\end{aligned}$$

如果

$$1 + \sum_{k=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{k-1}}{\mu_1 \mu_2 \cdots \mu_k} < \infty,$$

则存在平稳分布

$$\pi_k = \frac{\lambda_0 \lambda_1 \cdots \lambda_{k-1}}{\mu_1 \mu_2 \cdots \mu_k} \left( 1 + \sum_{k=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{k-1}}{\mu_1 \mu_2 \cdots \mu_k} \right)^{-1}, \quad (5.3.3)$$

此即该链的极限分布. 否则, 极限分布为 0.  $\square$

生灭过程的应用十分广泛, 包括生物种群繁衍、排队理论、流行病传染学 (如新冠病毒) 等. 下一节将以排队论为例, 简要介绍连续时间 Markov 链在实际中的应用.

## 5.4 排队论简介

排队现象是实际生活中经常遇到的情况, 如去银行排号办理业务、去食堂排队打饭. 网络中排队问题更是彼彼皆是, 数据包在每一级网络设备 (如交换机、路由器等) 都需要排队等待转发. 本节将对排队系统进行数学建模, 并介绍一些典型排队系统进行分析.

### 5.4.1 排队模型基本概念

基本的排队模型如图 5.2 所示, 由以下 3 个基本过程组成:

- (1) 到达过程: 描述顾客到达排队系统的规律, 包括到达的时间间隔的分布、每次到达的数目 (单个到达、还是成批到达) 等;
- (2) 等待过程: 刻画等待队列的容量 (有限或是无限)、以及顾客进入排队系统中排队等待的规则 (先到先服务、或是具有优先级队列等). 主要的性能指标包括顾客排队的等待时间、等待队列长度等;

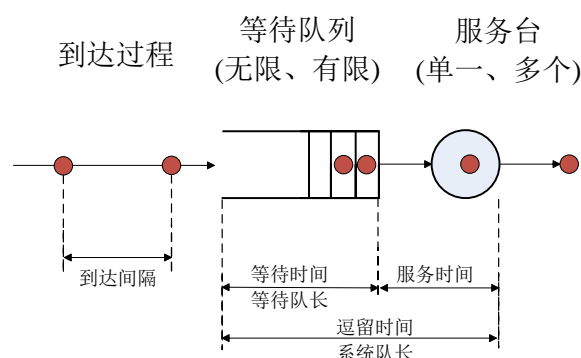


图 5.2 排队模型

- (3) 服务过程：服务过程由服务台对顾客提供服务，顾客在完成服务后离开排队系统. 服务过程需要描述服务台的数量 (单个或是多个) 以及服务的时间分布规律. 将等待过程和服务过程看作整体，相应的性能指标有顾客从进入排队系统到离开的逗留时间、排队系统中的总共逗留的顾客数 (包括排队等待的和正在接受服务的)，即系统队列长度.

顾客的**到达率**是指单位时间内到达排队系统的平均顾客数，通常用  $\lambda$  表示. 相应的，平均到达间隔为  $\lambda^{-1}$ . 服务台的**服务率**是指单位时间内完成服务的平均顾客数，通常用  $\mu$  表示. 相应的，平均服务时间为  $\lambda^{-1}$ . 由到达率和服务率可以计算排队系统的**资源利用率 (服务器占用率)**，即

$$\rho = \frac{\lambda}{\mu}.$$

排队系统的模型描述通常使用 Kendall 记号体系. 一个标准的排队模型可以用符号  $A/B/s(k)Z$  表示，其中，

- A 表示到达过程. 常见的过程包括：按 Poisson 过程到达，用 M 表示；按固定间隔到达，用 D 表示；一般性到达过程，用 G 表示.
- B 表示服务过程. 常用的记号与到达过程类似，M 表示指数分布的服务时间；D 表示固定的服务时长；G 表示一般的服务时间分布.
- s 表示服务台的数目.
- k 表示等待队列可容纳的顾客数. 这一项可以缺省不写出来，缺省时则默认为无穷大.
- Z 表示排队规则. 也可以缺省不写，默认为先入先出 (First-in-first-out, FIFO) 的排队方式.

举例来说, M/M/1 表示排队模型的到达过程为 Poisson, 服务时间服从指数分布, 单个服务台, 无穷等待空间, 先入先出的排队规则. M/M/1 是最基本的排队模型, 我们将从这个模型开始, 分析各类排队系统的性能.

#### 5.4.2 基本 Markov 排队模型: M/M/1

设  $t$  时刻 M/M/1 的系统队列长度为  $X(t)$ , 则  $X(t)$  可以用生灭过程来描述, 其中,  $\lambda_k = \lambda, \mu_k = \mu$ . 代入式 (5.3. 3), 得到当  $\lambda < \mu$  时,  $X(t)$  的平稳分布为

$$\pi_n = \rho^n (1 - \rho),$$

其中,  $\rho = \frac{\lambda}{\mu}$ . 此时,  $X(t)$  的极限分布存在, 且与平稳分布相同. 当  $\lambda \geq \mu$  时,  $X(t)$  不存在平稳分布. 直观来看, 即到达比服务更快时, 队列将无限制地增长下去, 系统无法进入平稳状态.

当  $\lambda < \mu$  时, 我们可以计算  $X(t)$  的期望, 即系统中的平均顾客数

$$\bar{L} = \sum_{n=0}^{\infty} n\pi_n = \sum_{n=0}^{\infty} n\rho^n (1 - \rho) = \frac{\rho}{1 - \rho}.$$

如果只考虑等待的平均顾客数, 则  $n$  个顾客中有  $n - 1$  人在排队等候, 则排队等待的平均顾客数为

$$\bar{L}_Q = \sum_{n=1}^{\infty} (n - 1)\pi_n = \frac{\rho^2}{1 - \rho}.$$

为了计算顾客的平均等待时间, 需要引入下面的重要定理.

**定理5.7. PASTA (Poisson Arrivals See Time Averages) 定理:** 对于任意一个排队系统, 其  $t$  时刻的队列长度为  $X(t)$ . 如果其到达过程为 Poisson, 到达时刻为  $S_1, S_2, \dots$ , 令  $Y_k = X(S_k^-)$ , 其中  $S_k^-$  为到达时刻的前一瞬间, 则有  $Y_k$  与  $X(t)$  同分布.

**证明:** 我们从  $Y_k$  的分布出发直接计算

$$\begin{aligned} P(Y_k = n) &= P(X(t) = n | S_k^- = t) \\ &= \lim_{\Delta t \rightarrow 0} P(X(t) = n | X(t + \Delta t) - X(t) = 1) \\ &= \lim_{\Delta t \rightarrow 0} \frac{P(X(t + \Delta t) - X(t) = 1 | X(t) = n) P(X(t) = n)}{P(X(t + \Delta t) - X(t) = 1)} \\ &= P(X(t) = n), \end{aligned}$$

其中, 最后一个等式是由于当  $\Delta t$  足够小时,  $X(t + \Delta t) - X(t)$  等价于到达过程增量, 即服从 Poisson 分布. 由于 Poisson 过程满足独立增量性, 其分布与  $t$  时刻及



以前的系统状态无关, 即有  $P(X(t + \Delta t) - X(t) = 1 | X(t) = n)P(X(t) = n) = P(X(t + \Delta t) - X(t) = 1)$ .  $\square$

根据 PASTA 定理, 任意顾客到达排队系统时, 看到的队列长度即为  $X(t)$ . 那么该顾客的逗留时间应为队列中所有顾客及其自身服务时间的总和. 值得注意的是, 其中有一个顾客正在被服务, 但由于服务时间服从指数分布, 具有无记忆性, 其在顾客到达后的剩余服务时间仍然服从同样的指数分布. 由于平均服务时间为  $\frac{1}{\mu}$ , 则顾客的平均逗留时间为

$$\bar{W} = \sum_{n=0}^{\infty} \frac{n+1}{\mu} \pi_n = \frac{1}{\mu(1-\rho)}.$$

排队等待的平均时间为

$$\bar{W}_Q = \sum_{n=0}^{\infty} \frac{n}{\mu} \pi_n = \frac{\rho}{\mu(1-\rho)}. \quad (5.4.1)$$

平均逗留时间与平均等待时间也可以通过计算分布求得. 以平均等待时间为例, 设  $x > 0$ ,  $W_Q$  为任意顾客的等待时间, 根据该顾客到达时刻前  $\tau^-$  系统中的顾客数进行全概率展开, 得到

$$\begin{aligned} P(W_Q \leq x) &= \sum_{n=0}^{\infty} P(X(\tau^-) = n)P(W_Q \leq x | X(\tau^-) = n) \\ &= \pi_0 + \sum_{n=1}^{\infty} \pi_n P(T_1 + \cdots + T_n \leq x) \\ &= 1 - \rho + \sum_{n=1}^{\infty} \rho^n (1 - \rho) \int_0^x \frac{\mu(\mu y)^{n-1}}{(n-1)!} e^{-\mu y} dy \\ &= 1 - \rho + \rho(1 - \rho) \int_0^x \sum_{n=1}^{\infty} \frac{\mu(\mu y \rho)^{n-1}}{(n-1)!} e^{-\mu y} dy \\ &= 1 - \rho + \rho(1 - \rho) \int_0^x \mu e^{-\mu y(1-\rho)} dy \\ &= 1 - \rho e^{-\mu(1-\rho)x}, \end{aligned}$$

其中, 第三个等号是由于  $T_k$  是独立同指数分布的随机变量, 其服从  $\Gamma$  分布. 根据这个分布, 同样可以求得式 (5.4.1) 所示的结果.

### 5.4.3 Little 定理

前面已经计算了平稳状态下排队系统的平均队列长度和平均逗留时间, 从概念上来说, 两者刻画的是不同层面的统计特性. 具体来说, 平均队列长度定义为

$$\bar{L} = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t N_\tau d\tau,$$

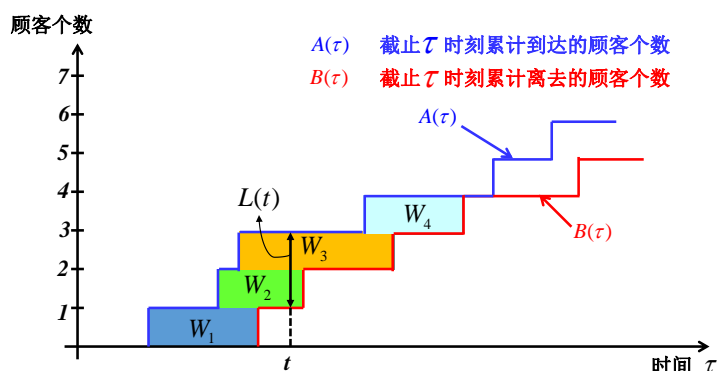


图 5.3 Little 定理的图示

其中,  $N_\tau$  表示时刻  $\tau$  的系统队列长度. 可以看到, 平均队列长度是时间平均的概念, 是对排队系统性能的一种刻画. 另一方面, 平均逗留时间定义为

$$\bar{W} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n W_k,$$

其中,  $W_k$  表示第  $k$  个顾客的逗留时间. 因此, 平均逗留时间是顾客平均的概念, 是顾客服务质量的一种评价指标. 两者的物理意义有很大差别, 他们之间的联系由 Little 定理给出.

**定理5.8.** 对于任意一个排队系统, 在平稳状态下有

$$\bar{L} = \lambda \bar{W}$$

其中  $\lambda$  为顾客平均到达率,  $\bar{L}$  为平均系统队列长度,  $\bar{W}$  为平均逗留时间.

**证明:** 这里我们避开严格的数学推导, 通过图示的方法直观的给出 Little 定理的证明. 如图 5.3 所示,  $A(\tau)$  表示截止  $\tau$  时刻累计到达的顾客个数,  $B(\tau)$  表示截止  $\tau$  时刻累计离去的顾客个数, 两条曲线都是“阶梯”状的, 并且由于顾客必然先到达, 之后才可能离去,  $B(\tau)$  一定在  $A(\tau)$  的右侧. 考察两条曲线围成的区域的面积, 有以下两种计算方法:

- (1) 横向积分. 此时, 面积的计算分解为若干个矩形块的面积  $W_k$  之和. 由于每个矩形块的宽对应着一个顾客的到达和离去时刻, 高度均为 1, 于是,  $W_k$  即为顾客的系统逗留时间. 总面积为  $\sum_{k=1}^{A(t)} W_k$ .
- (2) 纵向积分. 设任意  $t$  时刻所考察区域的高度为  $L(t)$ , 由于  $L(t) = A(t) - B(t)$ , 其物理意义即为该时刻系统中的总顾客数. 按照此定义, 总面积可写为  $\int_0^t L(\tau) d\tau$ .

综上所述，我们从两个不同角度计算了同一块区域的面积，两者应该相等，即

$$\int_0^t L(\tau) d\tau = \sum_{k=1}^{A(t)} W_k.$$

两边同时取时间平均，并令  $t \rightarrow \infty$ ，得到

$$\begin{aligned} \bar{L} &= \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t L(\tau) d\tau = \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{k=1}^{A(t)} W_k \\ &= \lim_{t \rightarrow \infty} \frac{A(t)}{t} \frac{\sum_{k=1}^{A(t)} W_k}{A(t)} \\ &= \lambda \bar{W}. \end{aligned} \quad \square$$

**注5.3.** Little 定理的直观解释是自然且合理的：某一顾客结束服务即将离开排队系统时、回头看到的平均队列长度，等于该顾客在系统中逗留的过程中、进入排队系统的平均顾客数。

**注5.4.** Little 定理具有很强的普适性，体现在以下两个方面：

- (1) Little 定理适用于任何达到平衡状态的排队系统。事实上，Little 定理没有对顾客到达过程、排队规则、服务过程等作任何假定，只要达到平稳状态，定理就一定成立。
- (2) 尽管我们在上面的讨论中关注的是系统逗留时间和系统顾客数，但是 Little 定理应用到排队系统的任意子系统也是成立的，只要公式中的三个统计平均量是针对同一个顾客群定义的即可。比如，对于等待队列长度和顾客的排队等待时间，根据前面的计算，容易验证  $\bar{L}_Q = \lambda \bar{W}_Q$ ，也是满足 Little 定理的。

关于 Little 定理的第二点普适性，可以通过下面的两个例子进一步体现。

**例5.9.** 设有一个等待空间数有限的排队系统，顾客到达率为  $\lambda$ 。由于系统等待空间数有限，当系统满负荷时，新到达的顾客无法进入排队系统而直接离去，这部分顾客的比例为  $P_B$  (称为阻塞率)。

针对这个系统，应用 Little 定理时有两种方法：

- (1) 只考虑进入排队系统的顾客，其平均队列长度为  $\bar{L}$ ，平均逗留时间为  $\bar{W}$ ，实际进入排队系统的顾客的到达率为  $\lambda(1 - P_B)$ 。于是，Little 公式应写为

$$\bar{L} = \lambda(1 - P_B) \bar{W}$$

(2) 如果考虑全体顾客, 其到达率为  $\lambda$ . 这些顾客分成两部分, 一部分进入队列排队, 另一部分直接离开. 后者对平均队列长度没有贡献, 仍然为  $\bar{L}$ . 但因为其逗留时间为 0, 会拉低平均逗留时间, 即  $(1 - P_B) \cdot \bar{W} + P_B \cdot 0 = (1 - P_B)\bar{W}$ . 代入 Little 公式, 得到的结果是一样的.

**例5.10.** 考虑一个排队系统只有一个服务台, 等待空间无穷大. 只关注于该系统的服务过程, 设服务台占用率为  $\rho$ , 即服务台有顾客正在被服务的概率. 于是, 服务台平均队列长度为

$$\bar{L} = 1 \cdot \rho + 0 \cdot (1 - \rho) = \rho.$$

在服务台平均逗留时间即平均服务时间

$$\bar{W} = \frac{1}{\mu}.$$

又由于等待空间无穷大, 所有到达的顾客都会被服务, 所以服务台看到的到达率为  $\lambda$ . 从而,

$$\rho = \frac{\lambda}{\mu}.$$

这便是服务台占用率  $\rho$  的计算公式.

#### 5.4.4 基本 Markov 排队模型的扩展

##### 5.4.4.1 M/M/s

考虑有多个服务台 ( $s > 1$ ) 的 Markov 排队系统, 到达率为  $\lambda$ , 服务率为  $\mu$ . 当顾客数不多于服务台数目, 即  $k \leq s$  时, 所有的顾客都不需要排队而直接进入服务, 此时实际工作的服务台数为  $k$ , 等效服务率为  $k\mu$ . 顾客的离去过程是多个 Poisson 的合流, 仍为 Poisson, 故 M/M/s 排队系统仍可用生灭过程来建模, 其中,  $\lambda_k = \lambda$ , 当  $k \leq s$  时, 有  $\mu_k = k\mu$ , 当  $k > s$  时, 有  $\mu_k = s\mu$ . 类似可解得队列长度稳态分布

$$\pi_0 = \left( \sum_{j=0}^{s-1} \frac{a^j}{j!} + \frac{a^s}{s!} \frac{s}{s-a} \right)^{-1},$$

$$\pi_j = \frac{a^j}{j!} \pi_0, \quad j \leq s,$$

$$\pi_j = \frac{a^s}{s!} \rho^{j-s} \pi_0, \quad j > s,$$

其中  $a = \frac{\lambda}{\mu}$ ,  $\rho = \frac{a}{s} < 1$ .

利用上述平稳分布, 读者可以自行计算平均队列长度和平均逗留时间.

#### 5.4.4.2 M/M/s(k)

当等待空间数  $k$  有限时, 相当于状态空间在  $s + k$  处被截断的 M/M/s 排队系统, 即状态空间变为有限  $\{0, 1, \dots, s + k\}$ . 转移率矩阵为:

$$Q = \begin{pmatrix} -\lambda & \lambda & & & & \\ \mu & -(\lambda + \mu) & \lambda_1 & & & \\ & \ddots & \ddots & \ddots & & \\ & & s\mu & -(\lambda + s\mu) & \lambda & \\ & & & \ddots & \ddots & \ddots \\ & & & & s\mu & -s\mu \end{pmatrix}.$$

由于状态空间有限, 此时没有收敛问题, 队列总能达到平稳. 队列长度稳态分布为

$$\begin{aligned} \pi_0 &= \left( \sum_{j=0}^{s-1} \frac{a^j}{j!} + \frac{a^s}{s!} \frac{1 - \rho^{k+1}}{1 - \rho} \right)^{-1}, \\ \pi_j &= \frac{a^j}{j!} \pi_0, \quad j \leq s, \\ \pi_j &= \frac{a^s}{s!} \rho^{j-s} \pi_0, \quad j > s. \end{aligned}$$

特别的, 当  $k = 0$ , 即无等待空间时, 顾客要么被即时服务, 要么因服务台被全部占用而被阻塞. 此时有

$$\pi_j = \frac{\frac{a^j}{j!}}{\sum_{j=0}^s \frac{a^j}{j!}}, \quad j \leq s.$$

由于服务器满 ( $j = s$ ) 时, 新到的顾客会直接离开系统, 则系统的阻塞率即服务台全被占用的概率:

$$P_{\text{block}} = \pi_s = \frac{\frac{a^s}{s!}}{\sum_{j=0}^s \frac{a^j}{j!}}.$$

上式称为爱尔兰损失公式 (Erlang-B 公式). 由于 M/M/s(0) 能很好地建模电话网的即时服务特性, Erlang-B 公式是通信话务理论中最重要的公式之一, 被广泛应用于电话网和数据网的业务量设计中.

## 第 6 章 Gauss 过程

Gauss 过程是典型的连续参数连续状态随机过程. 由中心极限定理我们知道, 大量的、微小的随机因素共同作用后的宏观表现近似为 Gauss 分布. 因此, Gauss 过程在实际中的应用十分广泛, 例如能很好地建模背景噪声等. 同时, Gauss 过程有良好的闭合表达式, 便于进行解析计算, 在理论研究中同样举足轻重.

### 6.1 Gauss 过程的定义

**定义6.1.** 设有随机过程  $X(t)$ ,  $t$  为实参, 如果  $\forall n, \forall t_1, t_2, \dots, t_n$ , 随机向量  $\mathbf{X} = (X(t_1), X(t_2), \dots, X(t_n))^T$  都服从  $n$  元 Gauss 分布, 则称  $X(t)$  为 **Gauss 过程**.

可见, Gauss 过程的定义完全依赖于  $n$  元 Gauss 分布. 我们接下来给出  $n$  元 Gauss 分布的定义.

#### 6.1.1 多元 Gauss 分布的定义

**定义6.2.** 如果  $n$  元实随机向量  $\mathbf{X} = (X_1, X_2, \dots, X_n)^T$  的联合概率密度为

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{|\boldsymbol{\Sigma}|}} \exp \left( -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right), \quad (6.1. 1)$$

其中  $\boldsymbol{\mu} = E\mathbf{X} = (EX_1, EX_2, \dots, EX_n)^T$  为均值向量,  $\boldsymbol{\Sigma} = \text{Cov}(\mathbf{X}, \mathbf{X}) = (\text{Cov}(X_i, X_j))_{i,j \in \{1, \dots, N\}}$  为  $n$  元对称正定协方差矩阵, 则称  $\mathbf{X}$  服从  $n$  元 **Gauss 分布** (也称  $n$  元正态分布), 记为  $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ .

**注6.1.** 根据定理 2.3 可知, 协方差矩阵一定是非负定的. 这里, 我们进一步假定  $\boldsymbol{\Sigma}$  是正定的, 以保证其可逆性. 后续再讨论非正定的一般情形.

当  $n = 1$  时, 上述分布退化为一维的情形. 其中, 均值为  $\mu = EX$ , 协方差矩阵退化为标量方差  $\Sigma = \sigma^2$ . 于是, 概率密度函数退化为一维正态分布概率密度形式:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left( -\frac{(x - \mu)^2}{2\sigma^2} \right).$$

接下来验证式 (6.1. 1) 是一个概率密度. 显然,  $f_{\mathbf{X}}(\mathbf{x}) \geq 0$ . 因此, 重点是验证该函数在  $n$  维空间中的积分等于 1. 注意到  $\boldsymbol{\Sigma}$  是对称正定的, 则一定可对角化为  $\boldsymbol{\Sigma} = \mathbf{U}^T \boldsymbol{\Lambda} \mathbf{U}$ , 其中  $\mathbf{U}$  是由  $\boldsymbol{\Sigma}$  的特征向量组成的方阵,  $\boldsymbol{\Lambda} = \text{diag}\{\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2\}$  是一个对角阵, 对角线元素为  $\boldsymbol{\Sigma}$  的特征值. 相应的,  $\boldsymbol{\Sigma}^{-1} = \mathbf{U}^T \boldsymbol{\Lambda}^{-1} \mathbf{U}$ , 其中

$\Lambda^{-1} = \text{diag}\{\sigma_1^{-2}, \sigma_2^{-2}, \dots, \sigma_n^{-2}\}$ . 于是

$$\begin{aligned}
\int_{\mathbb{R}^n} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} &= \int_{\mathbb{R}^n} \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{|\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right) d\mathbf{x} \\
&= \int_{\mathbb{R}^n} \frac{1}{(2\pi)^{\frac{n}{2}} \sigma_1 \sigma_2 \cdots \sigma_n} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{U}^T \Lambda^{-1} \mathbf{U}(\mathbf{x} - \boldsymbol{\mu})\right) d\mathbf{x} \\
&= \int_{\mathbb{R}^n} \frac{1}{(2\pi)^{\frac{n}{2}} \sigma_1 \sigma_2 \cdots \sigma_n} \exp\left(-\frac{1}{2} \mathbf{y}^T \Lambda^{-1} \mathbf{y}\right) d\mathbf{y} \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi} \sigma_1} \exp\left(-\frac{y_1^2}{2\sigma_1^2}\right) dy_1 \cdots \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi} \sigma_n} \exp\left(-\frac{y_n^2}{2\sigma_n^2}\right) dy_n \\
&= 1.
\end{aligned}$$

其中,  $d\mathbf{x} = dx_1 \cdots dx_n$ ,  $d\mathbf{y} = dy_1 \cdots dy_n$ , 第三个等号是做积分换元  $\mathbf{y} = \mathbf{U}(\mathbf{x} - \boldsymbol{\mu})$ . 可见, 式 (6.1. 1) 定义的  $n$  元函数的确是一个概率密度.

### 6.1.2 多元 Gauss 分布的特征函数

我们知道, 特征函数与概率分布是一一对应的关系. 由于多元 Gauss 的概率密度函数涉及到协方差矩阵的逆, 无论是分析还是计算, 都有诸多不便. 为此, 需要考虑其特征函数. 后面即将看到, 多元 Gauss 分布的特征函数对讨论其性质有很大帮助.

根据随机向量的特征函数的定义, 有

$$\begin{aligned}
\phi_{\mathbf{X}}(\boldsymbol{\omega}) &= E(\exp(j\boldsymbol{\omega}^T \mathbf{X})) \\
&= \int_{\mathbb{R}^n} \exp(j\boldsymbol{\omega}^T \mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \\
&= \int_{\mathbb{R}^n} \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{|\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) + j\boldsymbol{\omega}^T \mathbf{x}\right) d\mathbf{x} \\
&= \int_{\mathbb{R}^n} \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{|\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu} - j\boldsymbol{\Sigma}\boldsymbol{\omega})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu} - j\boldsymbol{\Sigma}\boldsymbol{\omega}) \right. \\
&\quad \left. + j\boldsymbol{\omega}^T \boldsymbol{\mu} - \frac{1}{2}\boldsymbol{\omega}^T \boldsymbol{\Sigma} \boldsymbol{\omega}\right) d\mathbf{x} \\
&= \exp\left(j\boldsymbol{\omega}^T \boldsymbol{\mu} - \frac{1}{2}\boldsymbol{\omega}^T \boldsymbol{\Sigma} \boldsymbol{\omega}\right).
\end{aligned}$$

其中, 第四个等号是高维向量“配方”的结果, 可类比于例 2.5 中一维正态分布的特征函数的计算过程. 将指数上第一个二次型展开后, 易验证等式成立.

关于多元 Gauss 分布的特征函数的推导, 亦可以类比于概率密度的验证, 通过积分换元得到同样的结果. 读者可自行完成.

**注6.2.** 多元 Gauss 分布特征函数的表达式中, 不涉及到协方差矩阵取逆的运算. 因此, 当  $\Sigma$  不可逆, 即不满秩时, 概率密度无法写出, 但仍可以写出其特征函数. 事实上, 协方差矩阵不满秩时, 意味着  $\mathbf{X}$  中某些元素可被其他元素线性表出 (线性相关). 此时随机向量仍然是 Gauss 的, 但其概率密度只集中在  $\mathbb{R}^n$  的某个低维子空间中.

## 6.2 多元 Gauss 分布的性质

多元 Gauss 分布具有很多良好的性质, 本节将对这些性质逐一进行探讨. 首先, Gauss 分布在线性变换下仍然服从 Gauss 分布, 只是参数发生变化. 这一性质是多元 Gauss 分布最重要的性质之一.

### 6.2.1 线性性质

**定理6.1. (线性变换不变性)** 设  $\mathbf{X} = (X_1, X_2, \dots, X_n)^T$  服从  $n$  元 Gauss 分布  $N(\boldsymbol{\mu}, \Sigma)$ ,  $\mathbf{C}$  为任意  $m \times n$  维矩阵, 则随机向量  $\mathbf{Y} = \mathbf{C}\mathbf{X}$  仍然服从 Gauss 分布, 均值为  $\mathbf{C}\boldsymbol{\mu}$ , 协方差矩阵为  $\mathbf{C}\Sigma\mathbf{C}^T$ .

**证明:** 我们通过计算  $\mathbf{Y}$  的特征函数加以证明. 由特征函数的定义,

$$\begin{aligned}\phi_{\mathbf{Y}}(\boldsymbol{\omega}) &= E(\exp(j\boldsymbol{\omega}^T \mathbf{Y})) \\ &= E(\exp(j\boldsymbol{\omega}^T \mathbf{C}\mathbf{X})) \\ &= E(\exp(j(\mathbf{C}^T \boldsymbol{\omega})^T \mathbf{X})) \\ &= \phi_{\mathbf{X}}(\mathbf{C}^T \boldsymbol{\omega}) \\ &= \exp\left(j(\mathbf{C}^T \boldsymbol{\omega})^T \boldsymbol{\mu} - \frac{1}{2}(\mathbf{C}^T \boldsymbol{\omega})^T \Sigma (\mathbf{C}^T \boldsymbol{\omega})\right) \\ &= \exp\left(j\boldsymbol{\omega}^T \mathbf{C}\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{\omega}^T \mathbf{C}\Sigma\mathbf{C}^T \boldsymbol{\omega}\right),\end{aligned}$$

根据特征函数与概率分布的一一对应性可知,  $\mathbf{Y} \sim N(\mathbf{C}\boldsymbol{\mu}, \mathbf{C}\Sigma\mathbf{C}^T)$ . □

上述定理的反命题也成立, 即可以通过判定线性变换不变性来判断随机向量是否服从多元 Gauss 分布. 事实上, 这一判定方法只需要对  $\mathbf{C}$  退化为向量的情形成立即可. 由此, 可以得到下面的定理.

**定理6.2.** 随机向量  $\mathbf{X} = (X_1, X_2, \dots, X_n)^T$  服从  $n$  元 Gauss 分布的充要条件是: 任取  $\mathbf{c} = (c_1, c_2, \dots, c_n)^T \in \mathbb{R}^n$ , 线性组合  $\mathbf{c}^T \mathbf{X} = c_1 X_1 + c_2 X_2 + \dots + c_n X_n$  都服从一元 Gauss 分布.



**证明：** (1) 必要性：由线性变换不变性直接得到必要性.

(2) 充分性：设  $E\mathbf{X} = \boldsymbol{\mu}$ ,  $\text{Cov}(\mathbf{X}, \mathbf{X}) = \boldsymbol{\Sigma}$ , 令  $Y = \mathbf{c}^T \mathbf{X}$ , 则

$$\begin{aligned} EY &= E(\mathbf{c}^T \mathbf{X}) = \mathbf{c}^T \boldsymbol{\mu}, \\ DY &= E((\mathbf{c}^T \mathbf{X} - \mathbf{c}^T \boldsymbol{\mu})^2) \\ &= E((\mathbf{c}^T (\mathbf{X} - \boldsymbol{\mu}))^2) \\ &= E(\mathbf{c}^T (\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T \mathbf{c}) \\ &= \mathbf{c}^T E((\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T) \mathbf{c} \\ &= \mathbf{c}^T \boldsymbol{\Sigma} \mathbf{c}. \end{aligned}$$

由于  $Y$  服从一元 Gauss 分布, 所以

$$\phi_Y(\omega) = E(\exp(j\omega Y)) = \exp\left(j(\mathbf{c}^T \boldsymbol{\mu})\omega - \frac{1}{2}(\mathbf{c}^T \boldsymbol{\Sigma} \mathbf{c})\omega^2\right).$$

而  $\mathbf{X}$  的特征函数在自变量取  $\mathbf{c}$  时为

$$\begin{aligned} \phi_{\mathbf{X}}(\mathbf{c}) &= E(\exp(j\mathbf{c}^T \mathbf{X})) \\ &= E(\exp(jY)) \\ &= \phi_Y(\omega)|_{\omega=1} \\ &= \exp\left(j\mathbf{c}^T \boldsymbol{\mu} - \frac{1}{2}\mathbf{c}^T \boldsymbol{\Sigma} \mathbf{c}\right). \end{aligned}$$

由于  $\mathbf{c}$  是任取的, 所以  $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ . □

利用线性变换不变性, 我们可以很容易地得知, Gauss 分布的任意维边缘分布仍然服从 Gauss 分布.

**定理6.3.** 如果  $\mathbf{X} = (X_1, X_2, \dots, X_n)^T$  服从  $n$  元 Gauss 分布, 则  $\mathbf{X}$  的任何一个子向量  $\tilde{\mathbf{X}} = (X_{k_1}, X_{k_2}, \dots, X_{k_m})^T$  ( $m < n$ ) 均服从  $m$  元 Gauss 分布.

**证明：**  $\tilde{\mathbf{X}}$  可以通过对  $\mathbf{X}$  作线性变换得到, 即

$$\tilde{\mathbf{X}} = \mathbf{A}\mathbf{X},$$

其中,  $\mathbf{A}$  的元素为

$$A_{ij} = \begin{cases} 1, & j = k_i, \\ 0, & \text{其他.} \end{cases}$$

由线性变换不变性可知,  $\tilde{\mathbf{X}}$  服从  $m$  元 Gauss 分布. □

上述定理的反命题不成立, 即边缘分布服从 Gauss 分布, 不能保证联合分布服从 Gauss 分布. 从定理 6.2 也可以看到, 联合 Gauss 分布需要满足的条件并不能由边缘的 Gauss 性直接得到.

### 6.2.2 独立性

通常情况下, 随机变(向)量之间的独立性与不相关性是两个不同的概念. 独立性可以得到不相关性, 但反过来则不一定成立. 而服从 Gauss 分布的随机向量, 其独立性和不相关性是等价的.

**定理6.4.** 设  $n$  元随机向量  $\mathbf{X}_1$  与  $m$  元随机向量  $\mathbf{X}_2$  服从联合 Gauss 分布, 均值向量分别为  $\boldsymbol{\mu}_1 \in \mathbb{R}^n$  与  $\boldsymbol{\mu}_2 \in \mathbb{R}^m$ .  $\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}$  的协方差矩阵为

$$\boldsymbol{\Sigma}_{\mathbf{X}} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix},$$

其中

$$\begin{aligned} \boldsymbol{\Sigma}_{11} &= \text{Cov}(\mathbf{X}_1, \mathbf{X}_1) = E((\mathbf{X}_1 - \boldsymbol{\mu}_1)(\mathbf{X}_1 - \boldsymbol{\mu}_1)^T), \\ \boldsymbol{\Sigma}_{22} &= \text{Cov}(\mathbf{X}_2, \mathbf{X}_2) = E((\mathbf{X}_2 - \boldsymbol{\mu}_2)(\mathbf{X}_2 - \boldsymbol{\mu}_2)^T), \\ \boldsymbol{\Sigma}_{12} &= \text{Cov}(\mathbf{X}_1, \mathbf{X}_2) = \boldsymbol{\Sigma}_{21}^T = E((\mathbf{X}_1 - \boldsymbol{\mu}_1)(\mathbf{X}_2 - \boldsymbol{\mu}_2)^T). \end{aligned}$$

则  $\mathbf{X}_1$  和  $\mathbf{X}_2$  相互独立的充要条件是

$$\boldsymbol{\Sigma}_{12} = \mathbf{0}.$$

**证明:** (1) 必要性: 如果  $\mathbf{X}_1$  和  $\mathbf{X}_2$  相互独立, 则

$$\begin{aligned} \boldsymbol{\Sigma}_{12} &= E((\mathbf{X}_1 - \boldsymbol{\mu}_1)(\mathbf{X}_2 - \boldsymbol{\mu}_2)^T) \\ &= E(\mathbf{X}_1 - \boldsymbol{\mu}_1)E(\mathbf{X}_2 - \boldsymbol{\mu}_2)^T \\ &= (E\mathbf{X}_1 - \boldsymbol{\mu}_1)(E\mathbf{X}_2 - \boldsymbol{\mu}_2)^T = \mathbf{0}. \end{aligned}$$

(2) 充分性: 如果  $\boldsymbol{\Sigma}_{12} = \mathbf{0}$ , 则协方差矩阵是分块对角阵

$$\boldsymbol{\Sigma}_{\mathbf{X}} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{22} \end{pmatrix},$$

$\mathbf{X}$  的概率密度函数为

$$\begin{aligned}
& f_{\mathbf{X}}(\mathbf{x}_1, \mathbf{x}_2) \\
&= \frac{1}{(2\pi)^{\frac{n+m}{2}} \sqrt{|\boldsymbol{\Sigma}_{\mathbf{X}}|}} \exp \left( -\frac{1}{2} \begin{pmatrix} \mathbf{x}_1 - \boldsymbol{\mu}_1 \\ \mathbf{x}_2 - \boldsymbol{\mu}_2 \end{pmatrix}^T \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{22} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{x}_1 - \boldsymbol{\mu}_1 \\ \mathbf{x}_2 - \boldsymbol{\mu}_2 \end{pmatrix} \right) \\
&= \frac{1}{(2\pi)^{\frac{n+m}{2}} \sqrt{|\boldsymbol{\Sigma}_{11} \boldsymbol{\Sigma}_{22}|}} \exp \left( -\frac{1}{2} (\mathbf{x}_1 - \boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}_{11}^{-1} (\mathbf{x}_1 - \boldsymbol{\mu}_1) \right. \\
&\quad \left. - \frac{1}{2} (\mathbf{x}_2 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}_{22}^{-1} (\mathbf{x}_2 - \boldsymbol{\mu}_2) \right) \\
&= \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{|\boldsymbol{\Sigma}_{11}|}} \exp \left( -\frac{1}{2} (\mathbf{x}_1 - \boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}_{11}^{-1} (\mathbf{x}_1 - \boldsymbol{\mu}_1) \right) \\
&\quad \cdot \frac{1}{(2\pi)^{\frac{m}{2}} \sqrt{|\boldsymbol{\Sigma}_{22}|}} \exp \left( -\frac{1}{2} (\mathbf{x}_2 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}_{22}^{-1} (\mathbf{x}_2 - \boldsymbol{\mu}_2) \right) \\
&= f_{\mathbf{X}_1}(\mathbf{x}_1) f_{\mathbf{X}_2}(\mathbf{x}_2),
\end{aligned}$$

所以  $\mathbf{X}_1$  和  $\mathbf{X}_2$  相互独立. □

**例6.1.** 设  $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)^T$  是服从 Gauss 分布的随机向量, 设有变换

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{A} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix},$$

试确定矩阵  $\mathbf{A}$ , 使得  $\mathbf{Y}_1, \mathbf{Y}_2$  不相关.

**解:** 根据变换关系, 有

$$\mathbf{Y}_1 = \mathbf{X}_1 + \mathbf{A}\mathbf{X}_2,$$

$$\mathbf{Y}_2 = \mathbf{X}_2.$$

根据协方差矩阵的性质 (定理2.3), 可得  $\mathbf{Y}_1$  和  $\mathbf{Y}_2$  的协方差为

$$\begin{aligned}
\text{Cov}(\mathbf{Y}_1, \mathbf{Y}_2) &= \text{Cov}(\mathbf{X}_1 + \mathbf{A}\mathbf{X}_2, \mathbf{X}_2) \\
&= \text{Cov}(\mathbf{X}_1, \mathbf{X}_2) + \mathbf{A}\text{Cov}(\mathbf{X}_2, \mathbf{X}_2) \\
&= \boldsymbol{\Sigma}_{12} + \mathbf{A}\boldsymbol{\Sigma}_{22}.
\end{aligned}$$

为了使得  $\mathbf{Y}_1, \mathbf{Y}_2$  不相关, 上式应满足  $\text{Cov}(\mathbf{Y}_1, \mathbf{Y}_2) = \mathbf{0}$ . 于是

$$\mathbf{A} = -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}.$$

此时,

$$\begin{aligned}
\text{Cov}(\mathbf{Y}_1, \mathbf{Y}_1) &= \text{Cov}(\mathbf{X}_1 + \mathbf{A}\mathbf{X}_2, \mathbf{X}_1 + \mathbf{A}\mathbf{X}_2) \\
&= \text{Cov}(\mathbf{X}_1, \mathbf{X}_1) + \mathbf{A}\text{Cov}(\mathbf{X}_2, \mathbf{X}_1) \\
&\quad + \text{Cov}(\mathbf{X}_1, \mathbf{X}_2)\mathbf{A}^\text{T} + \mathbf{A}\text{Cov}(\mathbf{X}_2, \mathbf{X}_2)\mathbf{A}^\text{T} \\
&= \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}, \\
\text{Cov}(\mathbf{Y}_2, \mathbf{Y}_2) &= \text{Cov}(\mathbf{X}_2, \mathbf{X}_2) = \boldsymbol{\Sigma}_{22}.
\end{aligned}$$

□

### 6.2.3 条件分布

利用例 6.1 的去相关操作, 可以计算多元 Gauss 的条件分布. 设  $n$  元随机向量  $\mathbf{X}_1$  与  $m$  元随机向量  $\mathbf{X}_2$  服从联合 Gauss 分布, 均值向量分别为  $\boldsymbol{\mu}_1 \in \mathbb{R}^n$  与  $\boldsymbol{\mu}_2 \in \mathbb{R}^m$ .  $\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}$  的协方差矩阵为

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix},$$

则  $\mathbf{X}_1, \mathbf{X}_2$  的联合概率密度为

$$\begin{aligned}
&f_{\mathbf{X}_1, \mathbf{X}_2}(\mathbf{x}_1, \mathbf{x}_2) \\
&= \frac{1}{(2\pi)^{\frac{n+m}{2}} \sqrt{|\boldsymbol{\Sigma}|}} \exp \left( -\frac{1}{2} \begin{pmatrix} \mathbf{x}_1 - \boldsymbol{\mu}_1 \\ \mathbf{x}_2 - \boldsymbol{\mu}_2 \end{pmatrix}^\text{T} \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{x}_1 - \boldsymbol{\mu}_1 \\ \mathbf{x}_2 - \boldsymbol{\mu}_2 \end{pmatrix} \right).
\end{aligned}$$

而  $\mathbf{X}_2$  的概率密度为

$$f_{\mathbf{X}_2}(\mathbf{x}_2) = \frac{1}{(2\pi)^{\frac{m}{2}} \sqrt{|\boldsymbol{\Sigma}_{22}|}} \exp \left( -\frac{1}{2} (\mathbf{x}_2 - \boldsymbol{\mu}_2)^\text{T} \boldsymbol{\Sigma}_{22}^{-1} (\mathbf{x}_2 - \boldsymbol{\mu}_2) \right).$$

为了计算条件概率密度

$$f_{\mathbf{X}_1|\mathbf{X}_2}(\mathbf{x}_1|\mathbf{x}_2) = \frac{f_{\mathbf{X}_1, \mathbf{X}_2}(\mathbf{x}_1, \mathbf{x}_2)}{f_{\mathbf{X}_2}(\mathbf{x}_2)}, \quad (6.2. 1)$$

我们考虑将  $\boldsymbol{\Sigma}$  对角化. 设有可逆线性变换  $\mathbf{Y} = \mathbf{C}\mathbf{X}$ , 则

$$\boldsymbol{\Sigma}_\mathbf{Y} = \text{Cov}(\mathbf{Y}, \mathbf{Y}) = \text{Cov}(\mathbf{C}\mathbf{X}, \mathbf{C}\mathbf{X}) = \mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^\text{T},$$

进而有

$$\boldsymbol{\Sigma}^{-1} = \mathbf{C}^\text{T} \boldsymbol{\Sigma}_\mathbf{Y}^{-1} \mathbf{C}. \quad (6.2. 2)$$

根据例 6.1 的结果, 取

$$\mathbf{C} = \begin{pmatrix} \mathbf{I} & -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \\ \mathbf{0} & \mathbf{I} \end{pmatrix},$$

则

$$\boldsymbol{\Sigma}_Y = \begin{pmatrix} \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{22} \end{pmatrix}.$$

利用式 (6.2. 2) 及上述表达式, 由条件概率密度的定义式 (6.2. 1) 可直接计算得到

$$f_{\mathbf{X}_1|\mathbf{X}_2}(\mathbf{x}_1|\mathbf{x}_2) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{|\tilde{\boldsymbol{\Sigma}}_{11}|}} \exp\left(-\frac{1}{2}(\mathbf{x}_1 - \tilde{\boldsymbol{\mu}}_1)^T \tilde{\boldsymbol{\Sigma}}_{11}^{-1}(\mathbf{x}_1 - \tilde{\boldsymbol{\mu}}_1)\right),$$

其中,

$$\begin{aligned} \tilde{\boldsymbol{\Sigma}}_{11} &= \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}, \\ \tilde{\boldsymbol{\mu}}_1 &= \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2). \end{aligned}$$

#### 6.2.4 应用举例——生成式人工智能 DDPM

去噪扩散概率模型 (Denoising Diffusion Probabilistic Models, DDPM) 是一种生成式模型, 在图像生成、视频生成等领域有广泛应用. DDPM 基于扩散过程, 对数据 (如图像) 逐渐添加 Gauss 噪声, 直至变成 Gauss 分布的纯噪声, 这个过程是正向扩散 (Direct Diffusion) 过程. 然后, 通过学习一个反向过程 (Reverse Inference), 从纯噪声中逐步恢复出原始数据, 实现生成新的数据样本.

首先考察正向扩散过程. 设  $\mathbf{X}_0$  是一个  $n$  维随机向量, 按如下方式构造一个随机向量序列:

$$\mathbf{X}_t = \sqrt{\alpha_t}\mathbf{X}_{t-1} + \sqrt{1 - \alpha_t}\mathbf{Z}_t, \quad t = 1, 2, \dots, T, \quad (6.2. 3)$$

其中,  $0 < \alpha_t < 1$  是给定确定性实数,  $\mathbf{Z}_t \sim N(0, \mathbf{I})$  服从  $n$  维标准正态分布, 且相互独立. 迭代计算可得

$$\begin{aligned} \mathbf{X}_t &= \sqrt{\alpha_t}(\sqrt{\alpha_{t-1}}\mathbf{X}_{t-2} + \sqrt{1 - \alpha_{t-1}}\mathbf{Z}_{t-1}) + \sqrt{1 - \alpha_t}\mathbf{Z}_t \\ &= \sqrt{\alpha_t\alpha_{t-1}}\mathbf{X}_{t-2} + \sqrt{\alpha_t(1 - \alpha_{t-1})}\mathbf{Z}_{t-1} + \sqrt{1 - \alpha_t}\mathbf{Z}_t \\ &= \sqrt{\alpha_t\alpha_{t-1}}\mathbf{X}_{t-2} + \sqrt{1 - \alpha_t\alpha_{t-1}}\hat{\mathbf{Z}}_t, \end{aligned}$$

其中,  $\hat{\mathbf{Z}}_t \sim N(0, \mathbf{I})$ , 第三个等号成立是由于相互独立的 Gauss 分布的线性组合仍然服从 Gauss 分布. 不断迭代下去可得,

$$\mathbf{X}_t = \sqrt{\bar{\alpha}_t}\mathbf{X}_0 + \sqrt{1 - \bar{\alpha}_t}\tilde{\mathbf{Z}}_t, \quad (6.2. 4)$$

其中,  $\bar{\alpha}_t = \alpha_t \alpha_{t-1} \cdots \alpha_1$ ,  $\tilde{\mathbf{Z}}_t \sim N(0, \mathbf{I})$ . 当  $t \rightarrow \infty$  时, 有  $\bar{\alpha}_t \rightarrow 0$ , 于是  $\mathbf{X}_t \rightarrow N(0, \mathbf{I})$ , 即服从标准正态分布的纯噪声.

再看反向过程. 考察给定  $\mathbf{X}_0$  和  $\mathbf{X}_t$  时,  $\mathbf{X}_{t-1}$  的条件概率密度, 由 Bayes 公式可得

$$q(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{x}_0) = \frac{q(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{x}_0) q(\mathbf{x}_{t-1} | \mathbf{x}_0)}{q(\mathbf{x}_t | \mathbf{x}_0)}. \quad (6.2.5)$$

根据式 (6.2.3) 可知, 给定  $\mathbf{X}_{t-1}$  时,  $X_t$  服从 Gauss 分布, 且与  $\mathbf{X}_0$  独立 (事实上, 序列  $\mathbf{X}_t$  具有 Markov 性), 即

$$\mathbf{X}_t | \mathbf{X}_{t-1} \sim N(\sqrt{\alpha_t} \mathbf{X}_{t-1}, (1 - \alpha_t) \mathbf{I}),$$

概率密度为

$$q(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{x}_0) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{1 - \alpha_t}} \exp \left( -\frac{(\mathbf{x}_t - \sqrt{\alpha_t} \mathbf{x}_{t-1})^T (\mathbf{x}_t - \sqrt{\alpha_t} \mathbf{x}_{t-1})}{2(1 - \alpha_t)} \right).$$

根据式 (6.2.4) 可知, 给定  $\mathbf{X}_0$  时,  $X_t$  也服从 Gauss 分布, 即

$$\mathbf{X}_t | \mathbf{X}_0 \sim N(\sqrt{\bar{\alpha}_t} \mathbf{X}_0, (1 - \bar{\alpha}_t) \mathbf{I}),$$

概率密度为

$$q(\mathbf{x}_t | \mathbf{x}_0) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{1 - \bar{\alpha}_t}} \exp \left( -\frac{(\mathbf{x}_t - \sqrt{\bar{\alpha}_t} \mathbf{x}_0)^T (\mathbf{x}_t - \sqrt{\bar{\alpha}_t} \mathbf{x}_0)}{2(1 - \bar{\alpha}_t)} \right).$$

根据 Gauss 分布的性质易知, 式 (6.2.5) 所对应的条件分布也是 Gauss 的. 因此, 只需要确定其期望和方差即可. 于是, 我们只关注指数上与自变量  $\mathbf{x}_{t-1}$  相关的部分, 即

$$\begin{aligned} & -\frac{(\mathbf{x}_t - \sqrt{\alpha_t} \mathbf{x}_{t-1})^T (\mathbf{x}_t - \sqrt{\alpha_t} \mathbf{x}_{t-1})}{2(1 - \alpha_t)} - \frac{(\mathbf{x}_{t-1} - \sqrt{\bar{\alpha}_{t-1}} \mathbf{x}_0)^T (\mathbf{x}_{t-1} - \sqrt{\bar{\alpha}_{t-1}} \mathbf{x}_0)}{2(1 - \bar{\alpha}_{t-1})} \\ &= -\left( \frac{\alpha_t}{2(1 - \alpha_t)} + \frac{1}{2(1 - \bar{\alpha}_{t-1})} \right) \mathbf{x}_{t-1}^T \mathbf{x}_{t-1} + \left( \frac{\sqrt{\alpha_t}}{1 - \alpha_t} \mathbf{x}_t^T + \frac{\sqrt{\bar{\alpha}_{t-1}}}{1 - \bar{\alpha}_{t-1}} \mathbf{x}_0^T \right) \mathbf{x}_{t-1} + c, \end{aligned}$$

于是有

$$\mathbf{X}_{t-1} | \mathbf{X}_t, \mathbf{X}_0 \sim N(\boldsymbol{\mu}, \sigma^2 \mathbf{I}),$$

其中

$$\begin{aligned} \sigma^2 &= \left( \frac{\alpha_t}{1 - \alpha_t} + \frac{1}{1 - \bar{\alpha}_{t-1}} \right)^{-1} \\ &= \frac{(1 - \alpha_t)(1 - \bar{\alpha}_{t-1})}{\alpha_t(1 - \bar{\alpha}_{t-1}) + (1 - \alpha_t)} \\ &= \frac{(1 - \alpha_t)(1 - \bar{\alpha}_{t-1})}{1 - \bar{\alpha}_t}, \end{aligned}$$

$$\begin{aligned}
\mu &= \sigma^2 \left( \frac{\sqrt{\alpha_t}}{1 - \alpha_t} \mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}}{1 - \bar{\alpha}_{t-1}} \mathbf{x}_0 \right) \\
&= \frac{\sqrt{\alpha_t}(1 - \bar{\alpha}_{t-1})}{1 - \bar{\alpha}_t} \mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}(1 - \alpha_t)}{1 - \bar{\alpha}_t} \mathbf{x}_0 \\
&= \frac{\sqrt{\alpha_t}(1 - \bar{\alpha}_{t-1})}{1 - \bar{\alpha}_t} \mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}(1 - \alpha_t)}{\sqrt{\bar{\alpha}_t}(1 - \bar{\alpha}_t)} (\mathbf{x}_t - \sqrt{1 - \bar{\alpha}_t} \tilde{\mathbf{z}}_t) \\
&= \frac{1}{\sqrt{\alpha_t}} \mathbf{x}_t + \frac{1 - \alpha_t}{\sqrt{\alpha_t}(1 - \bar{\alpha}_t)} \tilde{\mathbf{z}}_t,
\end{aligned}$$

其中, 第三个等号是利用式 (6.2. 4) 消去  $\mathbf{x}_0$ ,  $\tilde{\mathbf{z}}_t$  即在正向扩散过程中加上的噪声项. 对于图像生成任务, 实际上就是通过深度神经网络, 参数化拟合出反向条件概率  $q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) = q_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)$ , 亦即参数化拟合出噪声项  $\tilde{\mathbf{z}}_t = \tilde{\mathbf{z}}_{t,\theta}$ , 从而从一幅纯噪声的图像  $\mathbf{x}_T$ , 逐步生成所需要的图像  $\mathbf{x}_0$ .

## 6.3 Brown 运动

Brown 运动是英国生物学家 Robert Brown 于 1827 年通过显微镜研究花粉在液体中的运动时发现的, 因此而得名. Brown 观察到花粉颗粒在液体中作无规则运动, 认为是大量液体分子和原子无序运动、与颗粒碰撞造成的. 根据上述观点, 由中心极限定理可知, Brown 运动与 Gauss 分布有着紧密的联系. 关于 Brown 运动的研究在 20 世纪初迅速发展, 法国人 Bachelier 于 1900 年在其博士论文中研究期权定价问题, 就运用了 Brown 运动为股票价格建模. 随后, Einstein 于 1905 年发表了一篇研究 Brown 运动的论文, 按照流体分子动力学理论, 推导出热扩散方程并由此推定分子的大小. 物理学定 Perrin 通过实验证实了 Einstein 的理论成果. Brown 运动的严格数学定义和构造直到 1923 年才由 Wiener 给出, 因而 Brown 运动也称为 Wiener 过程. Brown 运动成为 20 世纪概率论的主要研究对象之一, 大量知名学者推动了相关理论的研究与应用.

### 6.3.1 定义与基本性质

**定义 6.3.** 实值随机过程  $B(t), t \geq 0$  如果满足以下三个条件, 则称其为**标准 Brown 运动**:

- (1)  $B(t)$  具有平稳增量性和独立增量性, 且  $B(0) = 0$ ;
- (2)  $B(t)$  的每一条样本轨道都是连续的;
- (3)  $\forall t, B(t)$  服从 Gauss 分布,  $B(t) \sim N(0, t)$ .

以下从 3 个方面讨论 Brown 运动的基本性质: (1) 数字特征; (2) 联合分布; (3) 样本曲线性质. 具体总结如下.

**定理6.5.** 设  $B(t)$  为标准 Brown 运动, 则

$$(1) EB(t) = 0, DB(t) = t, \text{Cov}(B(s), B(t)) = \min(s, t).$$

(2)  $\forall n, \forall 0 \leq t_1 \leq t_2 \leq \cdots \leq t_n$ , 标准 Brown 运动的任意  $n$  维联合分布为

$$f_{B(t_1), \dots, B(t_n)}(x_1, \dots, x_n) = \frac{1}{\sqrt{(2\pi)^n \prod_{k=1}^n (t_k - t_{k-1})}} \exp \left( -\frac{1}{2} \sum_{k=1}^n \frac{(x_k - x_{k-1})^2}{t_k - t_{k-1}} \right).$$

(3) Brown 运动的样本轨道几乎处处连续, 但几乎处处不可微.

**证明:** (1)  $B(t)$  的均值和方差由 Gauss 分布直接得到. 为了计算协方差, 不妨设  $s < t$ , 则

$$\begin{aligned} \text{Cov}(B(s), B(t)) &= E(B(s)B(t)) \\ &= E(B(s)(B(t) - B(s) + B(s))) \\ &= E(B(s)(B(t) - B(s))) + E(B(s)^2) \\ &= EB(s)E((B(t) - B(s)) + s) = s \end{aligned}$$

其中, 第 4 个等号是由于独立增量性, 最后一个等号是由于均值为 0.

(2) 根据平稳增量性知, 对  $\forall t < s$ ,

$$B(s) - B(t) \stackrel{d}{=} B(s - t) - B(0) = B(s - t) \sim N(0, s - t).$$

又根据独立增量性知,  $B(t_1), B(t_2) - B(t_1), \dots, B(t_n) - B(t_{n-1})$  相互独立, 因此  $n$  维随机向量  $\mathbf{B} = (B(t_1), B(t_2) - B(t_1), \dots, B(t_n) - B(t_{n-1}))^T$  服从  $n$  维 Gauss 分布, 概率密度为

$$f_{\mathbf{B}}(y_1, \dots, y_n) = \frac{1}{\sqrt{(2\pi)^n \prod_{k=1}^n (t_k - t_{k-1})}} \exp \left( -\frac{1}{2} \sum_{k=1}^n \frac{y_k^2}{t_k - t_{k-1}} \right).$$

而

$$\begin{pmatrix} B(t_1) \\ B(t_2) \\ \vdots \\ B(t_n) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 & 0 \\ 1 & 1 & \cdots & \cdots & 1 \end{pmatrix} \begin{pmatrix} B(t_1) \\ B(t_2) - B(t_1) \\ \vdots \\ B(t_n) - B(t_{n-1}) \end{pmatrix}$$

为线性变换, 由 Gauss 分布的线性变换不变性知,  $B(t_1), \dots, B(t_n)$  也服从联合 Gauss 分布. 由于上述线性变换的 Jacobi 行列式为 1, 便可得到定理中的联合概率密度.



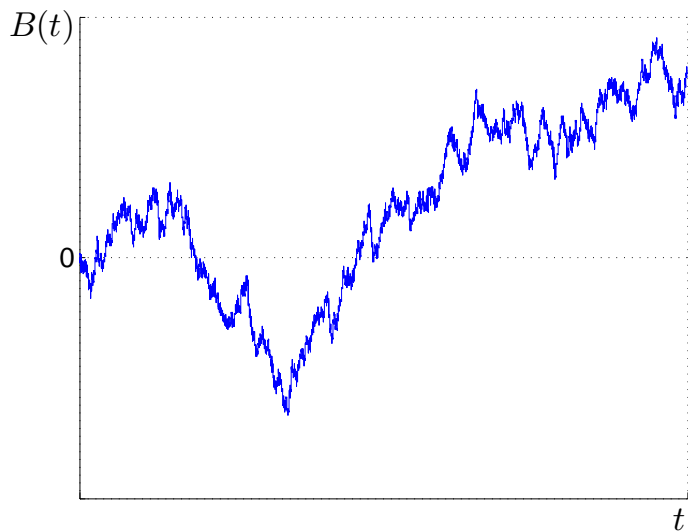


图 6.1 Brown 运动示意图

(3) 这里不给出严格证明, 仅给出粗略分析. 由于  $E(B(t) - B(s))^2 = |t - s|$ , 根据 Chebyshev 不等式, 对  $\forall \epsilon > 0$ , 有

$$P(|B(t) - B(s)| > \epsilon) \leq \frac{E|B(t) - B(s)|^2}{\epsilon^2} = \frac{|t - s|}{\epsilon^2}.$$

于是, 当  $t \rightarrow s$  时,  $P(|B(t) - B(s)| > \epsilon) \rightarrow 0$ , 即  $B(t)$  依概率连续. 几乎处处连续的证明略.

另一方面, 设  $t > s$ , 则

$$\frac{B(t) - B(s)}{t - s} \sim N\left(0, \frac{1}{t - s}\right),$$

当  $t \rightarrow s$  时, 方差  $\frac{1}{t-s} \rightarrow \infty$ , 则  $\frac{B(t)-B(s)}{t-s}$  不会收敛到有限的极限. 因此, 曲线  $B(t)$  不可微.  $\square$

如何画出 Brown 运动的曲线呢? 下面的定理提供了一种计算机模拟 Brown 运动的方法.

**定理6.6.** 设  $X_1, X_2, \dots, X_k, \dots$  为一列独立同分布随机变量, 其分布为  $P(X_k = 1) = P(X_k = -1) = 0.5$ , 定义随机过程

$$S_n(t) = \sum_{k=1}^{\lfloor nt \rfloor} X_k + (nt - \lfloor nt \rfloor)X_{\lfloor nt \rfloor + 1},$$

其中  $\lfloor x \rfloor$  表示不大于  $x$  的最大整数, 则在任意有限维联合分布的意义下有

$$\frac{S_n(t)}{\sqrt{n}} \rightarrow B(t), \quad n \rightarrow +\infty,$$

其中  $B(t)$  是标准 Brown 运动.

证明: 略. □

定理 6.6 刻画了 Brown 运动与随机徘徊之间的关联, 随机徘徊的状态跳转间隔趋于零的极限就是 Brown 运动. 图 6.1 给出了 Brown 运动的典型曲线, 从中可以体会 Brown 运动的连续性和不可微性.

### 6.3.2 最大值分布

对于 Brown 运动, 比较值得关注的是, 一定时间后能走到多远. 为此, 引入首达时间和最大值的概念.

**定义 6.4.** Brown 运动  $B(t)$  关于给定位置  $a$  的首达时间定义为

$$T_a = \inf\{t : t \geq 0, B(t) = a\}.$$

**定义 6.5.** Brown 运动  $B(t)$  在  $[0, t]$  中的最大值定义为

$$M_t = \max_{0 \leq s \leq t} B(s).$$

为了讨论最大值分布, 首先给出 Brown 运动的反射原理.

**定理 6.7. (反射原理)** 设  $B(t)$  是标准 Brown 运动, 令

$$\hat{B}(t) = \begin{cases} B(t), & t < T_a, \\ 2a - B(t), & t \geq T_a. \end{cases}$$

则随机过程  $\hat{B}(t)$  仍是一个标准 Brown 运动, 与  $B(t)$  同分布.

**证明:** 上述定理是随机徘徊的反射原理 (引理 3.5) 在连续空间中的推广, 可仿照引理 3.5 进行类似的证明. □

根据反射原理, 可以计算 Brown 运动的首达时间和最大值分布.

**定理 6.8.** 当  $a \geq 0$  时, Brown 运动  $B(t)$  的最大值满足

$$P(M_t \geq a) = \sqrt{\frac{2}{\pi}} \int_{\frac{a}{\sqrt{t}}}^{\infty} e^{-\frac{x^2}{2}} dx.$$

**证明：** 直接计算可得

$$\begin{aligned}
 P(M_t \geq a) &= P(M_t \geq a, B(t) \geq a) + P(M_t \geq a, B(t) < a) \\
 &= P(B(t) \geq a) + P(M_t \geq a, \hat{B}(t) > a) \\
 &= P(B(t) \geq a) + P(M_t \geq a, B(t) > a) \\
 &= 2P(B(t) \geq a) \\
 &= 2 \int_a^\infty \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}} dx \\
 &= 2 \int_{\frac{a}{\sqrt{t}}}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy,
 \end{aligned}$$

其中，第二个等号是由  $\hat{B}(t)$  的定义得到，并且  $\{B(t) \geq a\} \subset \{M_t \geq a\}$ ，第三个等号是由于  $\hat{B}(t)$  与  $B(t)$  同分布，最后一个等号是作积分换元  $y = \frac{x}{\sqrt{t}}$ .  $\square$

由定理 6.8 可以计算首达时间的分布.

**推论6.9.** Brown 运动  $B(t)$  的首达时间  $T_a$  的概率密度为

$$f_a(t) = \frac{|a|}{\sqrt{2\pi t^3/2}} e^{-a^2/2t}.$$

**证明：** 首先根据定理 6.8，计算首达时间的分布. 当  $a \geq 0$  时，显然， $\{T_a \leq t\}$  与  $\{M_t \geq a\}$  是等价事件，即

$$P(T_a \leq t) = P(M_t \geq a) = \sqrt{\frac{2}{\pi}} \int_{\frac{a}{\sqrt{t}}}^\infty e^{-\frac{x^2}{2}} dx.$$

当  $a < 0$  时，考察随机过程  $-B(t)$ ，显然也是标准 Brown 运动，其最大值记为  $\hat{M}_t$ ，则有

$$P(T_a \leq t) = P(\hat{M}_t \geq -a) = \sqrt{\frac{2}{\pi}} \int_{\frac{-a}{\sqrt{t}}}^\infty e^{-\frac{x^2}{2}} dx.$$

综上所述，可得  $T_a$  的概率分布函数为

$$F_a(t) = P(T_a \leq t) = \sqrt{\frac{2}{\pi}} \int_{\frac{|a|}{\sqrt{t}}}^\infty e^{-\frac{x^2}{2}} dx.$$

于是，

$$f_a(t) = F'_a(t) = \frac{|a|}{\sqrt{2\pi t^3/2}} e^{-a^2/2t}.$$

$\square$