# 概率论与随机过程

授课教师：唐宏岩

# 前言

本讲义基于清华大学数学系唐宏岩老师于 2023 - 2024 学年秋季学期开设的《概率论与随机过程》课程，用于辅助同学们课后复习。

由于时间与能力所限，本讲义可能不会出现大段的文字论述（但会包含重要的定义、定理与公式等）。但是，对许多基本概念的深入理解是非常有必要的，同学们可以在浏览时检查自己是否能够回忆起课上的内容，对掌握不够扎实的地方，鼓励大家查阅参考书或在课程群提问以解决问题。

由于此为教学团队第一年尝试整理讲义，诸如格式编排、内容完整性方面可能存在许多不足，欢迎大家联系我提出宝贵的意见与建议。

曹子尧

2023 年 9 月

# 目录

# 第一部分

# 初等概率论

# 第一章　事件的概率

## 1.1　概率的发展史

赌博中的 de Méré's Problem：连续掷一个均匀六面骰 4 次，获得至少一次“6”的概率为 $1-(\frac{5}{6})^4 \approx 0.5177$；而连续掷两个均匀六面骰 24 次，获得至少一次“对 6”的概率为 $1-(35/36)^{24} \approx 0.4914$。

Pascal 和 Fermat 的通信中使用初等数学的方法，首创了概率论相当多的数学理论，虽然当时没有总结成通用的定理。

Laplace 创立了采用分析方法的分析概率论。

Kolmogorov 利用测度论方法发展了现代概率理论。

## 1.2　随机试验与事件

**定义 1.1.** 概率论中的随机试验指的是符合下面两个特点的试验：

1. 不能预先确知结果
2. 可以预测所有可能的结果

**定义 1.2.** 样本空间是指一个试验的所有可能结果的集合，常用 $\Omega$ 表示。

**定义 1.3.** 事件是样本空间的一个良定义的子集。

一次随机试验中，一个事件可能发生或不发生。

下面是一些常见的事件：

1. 全事件 $\Omega$（必然事件）
2. 空事件 $\varnothing$（不可能事件）
3. 基本事件 $\{a\}$，其中 $a \in \Omega$，即仅包含单一试验结果的事件

## 1.3 事件的运算

由于事件是集合，因此事件之间可以进行集合之间的运算，如：

1. 余 $A^c = \Omega \setminus A$
2. 和 $A + B = A \cup B = (A^c \cap B^c)^c$
3. 差 $A - B = A \setminus B$
4. 积 $AB = A \cap B = (A^c \cup B^c)^c$

集合的 De Morgan's laws 也适用于事件：$(\bigcup_n A_n)^c = \bigcap_n A_n^c$。

事件的运算像集合的运算一样，可以用 Venn 图来表示。

## 1.4 概率的几种解释

对于概率这一数学概念，人们形成了几种从不同角度出发的解释：

1. 古典解释：基于等可能性的解释
2. 频率解释：基于大量重复试验的解释（频率学派采用的解释）
3. 主观解释：概率是一种对确信程度的度量（Bayes 学派采用的解释）

## 1.5 概率的公理化定义

用 $2^\Omega$ 表示 $\Omega$ 的幂集，即 $\Omega$ 的所有子集组成的集合。

**定义 1.4.** 事件集类 $\mathscr{F} \subset 2^\Omega$ 必须满足所谓 $\sigma$-代数的性质：

1. $\Omega \in \mathscr{F}$
2. $A \in \mathscr{F} \Rightarrow A^c \in \mathscr{F}$（对补运算的封闭性）
3. $A_i \in \mathscr{F}, \forall i \in \mathbb{N}^* \Rightarrow \bigcup_{i=1}^\infty A_i \in \mathscr{F}$（对可列并的封闭性）

**例 1.1.** $\Omega = \{a, b, c, d\}$，以下是一些合法的事件集类：

1. $\mathscr{F}_1 = 2^\Omega$
2. $\mathscr{F}_2 = \{\Omega, \varnothing\}$
3. $\mathscr{F}_3 = \{\Omega, \varnothing, \{a, b\}, \{c, d\}\}$

**定义 1.5.** (Kolmogorov) 概率函数 $P : \mathscr{F} \to \mathbb{R}$ 是满足以下三条公理的映射：

1. $P(A) \geq 0, \forall A \in \mathscr{F}$
2. $P(\Omega) = 1$

3. $A_i \in \mathscr{F}, \forall i \in \mathbb{N}^*, A_iA_j = \varnothing, \forall i \neq j \Rightarrow P(\sum_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$（加法公理/可列可加性）

称 $(\Omega, \mathscr{F}, P)$ 是一个概率空间。

**命题 1.1.** 关于概率空间，有如下性质：

1. $P(A) \leq 1,\ \forall A \in \mathscr{F}$
2. $P(\varnothing) = 0$
3. $P(A) + P(A^c) = 1$
4. $A_i \in \mathscr{F}, \forall i \in \{1, 2, \cdots, n\},\ A_iA_j = \varnothing, \forall i \neq j \Rightarrow P(\sum_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$（有限可加性）
5. $A \subset B \Rightarrow P(A) \leq P(B)$（称事件 $A$ 蕴涵事件 $B$）
6. $$\begin{aligned} P(A_1 + \cdots + A_n) =& \sum_{i=1}^{n} P(A_i) - \sum_{i_1 < i_2} P(A_{i_1}A_{i_2}) \\ &+ \cdots + (-1)^{r+1} \sum_{i_1 < i_2 < \cdots < i_r} P(A_{i_1}A_{i_2} \cdots A_{i_r}) \\ &+ \cdots + (-1)^{n+1} P(A_1 \cdots A_n) \end{aligned}$$ （容斥公式）

   特别地，$P(A + B) = P(A) + P(B) - P(AB)$。

**例 1.2.** （配对问题）

有 $n$ 个人，每人有一顶帽子。现将所有帽子放到一起，再随机分配给每人一顶，考虑无人拿到自己的帽子的概率。

为此，设事件 $A_i$ 为“第 $i$ 个人拿到自己的帽子”，则 $P(A_i) = 1/n$。

利用容斥公式，至少一人拿到自己帽子的概率为

$$\begin{aligned} & P(A_1 + \cdots + A_n) \\ =& \sum_{i=1}^{n} P(A_i) - \sum_{i_1 < i_2} P(A_{i_1}A_{i_2}) \\ &+ \cdots + (-1)^{r+1} \sum_{i_1 < i_2 < \cdots < i_r} P(A_{i_1}A_{i_2} \cdots A_{i_r}) \\ &+ \cdots + (-1)^{n+1} P(A_1 \cdots A_n) \end{aligned}$$

其中 $\sum_{i_1 < i_2 < \cdots < i_r} P(A_{i_1}A_{i_2} \cdots A_{i_r}) = \frac{(n-r)!}{n!}\binom{n}{r} = \frac{1}{r!}$，即 $P(A_1 + \cdots + A_n) = 1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \cdots + (-1)^{r+1}\frac{1}{r!} + \cdots + (-1)^{n+1}\frac{1}{n!}$。

所求概率 $P_n = 1 - P(A_1 + \cdots + A_n) = 1 - (1 - \frac{1}{2!} + \cdots + (-1)^{n+1}\frac{1}{n!}) \to e^{-1} (n \to +\infty)$。

思考：恰有 $k$ 个人拿到自己的帽子的概率?

## 1.6 条件概率

**定义 1.6.** 若 $P(B)>0$，定义条件概率 $P(A|B)=\frac{P(AB)}{P(B)}$。

通常，计算条件概率的方法有两种：

1. 在缩小（受限）的样本空间（要求事件 $B$ 发生）上，考虑事件 $A$ 发生的概率
2. 根据定义计算

一种常用的形式是 $P(AB)=P(A|B)P(B)=P(B|A)P(A)$，这可以视作是求解两个事件的积的概率的方法（乘法法则）。

**例 1.3.** 掷一个均匀六面骰，$\Omega=\{1,2,3,4,5,6\}, A=\{2,3,4,5\}, B=\{1,3,5\}$，则 $P(A)=4/6, P(B)=3/6, P(AB)=2/6, P(A|B)=\frac{P(AB)}{P(B)}=2/3$。

**例 1.4.** 袋子中有 8 个红球和 4 个白球，无放回地取出两个球，利用组合数可知，两个都是红球的概率为 $\frac{\binom{8}{2}}{\binom{12}{2}}$。
用条件概率可以简化计算：$P(R_1R_2)=P(R_1)P(R_2|R_1)=\frac{8}{12}\times\frac{7}{11}$。

更一般地，有 $P(A_1A_2\cdots A_n)=P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1A_2\cdots A_{n-1})$，常用于序贯发生的一系列事件的积的概率求解。

**例 1.5.** 回忆上一节的“配对问题”。我们有

$$
\begin{aligned}
&P(A_{i_1}A_{i_2}\cdots A_{i_r})\\
=&P(A_{i_1})P(A_{i_2}|A_{i_1})\cdots P(A_{i_r}|A_{i_1}\cdots A_{i_{r-1}})\\
=&\frac{1}{n}\times\frac{1}{n-1}\times\cdots\times\frac{1}{n-(r-1)}\\
=&\frac{(n-r)!}{n!}.
\end{aligned}
$$

**命题 1.2.** 对于给定的事件 $B$，$P(\cdot|B):\mathscr{F}\to\mathbb{R}$ 是概率函数，即 $(\Omega,\mathscr{F},P(\cdot|B))$ 仍是概率空间。

对于上述命题的证明，只需验证 $P(\cdot|B)$ 满足概率的三条公理即可。

这提示我们，条件概率也是一种概率，如果将 $P(A)$ 称为观察到事件 $B$ 之前 $A$ 的“先验概率”，则 $P(A|B)$ 就是相应的“后验概率”。

一个常见的迷思是：观测到事件 $A$ 已经发生后，是否可以说事件 $A$ 发生的概率 $P(A)=1$? 学过条件概率之后，我们知道答案是否定的，实际上是后验概率 $P(A|A)=1$。

## 1.7 事件的独立性

**定义 1.7.** 若 $P(AB)=P(A)P(B)$，则称事件 $A,B$ 相互独立。

如果 $P(B)>0$，我们注意到 $A,B$ 独立等价于 $P(A|B)=P(A)$。

**命题 1.3.** 若 $A,B$ 独立，则 $A^c,B$ 独立。

**定义 1.8.** 若 $P(ABC)=P(A)P(B)P(C)$，且 $A,B,C$ 两两独立，则称事件 $A,B,C$ 独立。

注意，仅有 $A,B,C$ 两两独立，不能推出三者独立。

**定义 1.9.** 若对于事件列 $\{A_i\}_{i=1}^{\infty}$，任意取有限个事件 $A_{i_1},A_{i_2},\cdots,A_{i_r}$，都有 $P(A_{i_1}A_{i_2}\cdots A_{i_r})=P(A_{i_1})P(A_{i_2})\cdots P(A_{i_r})$，则称 $\{A_i\}_{i=1}^{\infty}$ 相互独立。

**例 1.6.** 每周开奖的彩票，各次中奖率均为 $10^{-5}$ 且独立，问连续十年（520 周）不中奖的概率？令事件 $A_i$ 为第 $i$ 周不中奖，则 $P(A_i)=1-10^{-5}$，故 $P(A_1\cdots A_{520})=(1-10^{-5})^{520}\approx 0.9948$。

**定义 1.10.** 若事件 $A,B,E$ 满足 $P(AB|E)=P(A|E)P(B|E)$，则称 $A,B$ 关于 $E$ 条件独立。

注意，条件独立性和独立性之间没有蕴涵关系。

## 1.8 Bayes 公式

**定理 1.1.** （全概率公式）
设 $\{B_i\}$ 是 $\Omega$ 的一个分割，即

1. $\sum_i B_i=\Omega$
2. $B_iB_j=\varnothing,\forall i\neq j$
3. $P(B_i)>0,\forall i$

则 $P(A)=P(\sum_i(AB_i))=\sum_i P(AB_i)=\sum_i P(A|B_i)P(B_i)$。
注：$\{B_i\}$ 可以是有限集合，或可数无穷集合。

**例 1.7.** 对于调查问卷中的敏感问题（如“你是否有过某病史”），被调查者可能会有所顾虑而做出虚假的回答。为保护被调查者的隐私，同时取得其信任，考虑引入一个“保护性问题”，即不具有敏感性的问题（如“你是否会游泳”），并让被调查者以抛硬币的方式，随机抽取一个问题回答。这样，抽到敏感问题的、确有过该病史的被调查者在回答“是”时也无须有病史暴露之虞。
设人群中，敏感问题答案为“是”的比例为 $p$（未知），保护性问题答案为“是”的比例为 $q$（假设已知），则若收集到 $n$ 个被调查者的结果，其中 $k$ 个为“是”，便有 $\frac{1}{2}p+\frac{1}{2}q\approx\frac{k}{n}$，可以据此得到 $p$ 的估计。

**定理 1.2.** （Bayes 公式 / Bayes 准则）
设 $\{B_i\}$ 是 $\Omega$ 的一个分割，则 $P(B_i|A)=\frac{P(B_i)P(A|B_i)}{\sum_j P(B_j)P(A|B_j)}$。

**例 1.8.** （假阳性悖论）
对于一种流行病，$A$ 表示一个人检查呈阳性，$B$ 表示此人确实患病。
设 $P(B)=10^{-4}, P(A|B)=0.99, P(A|B^c)=10^{-3}$，
则一个检查呈阳性的人真的患病的概率仅为 $P(B|A)=\frac{P(A|B)P(B)}{P(A|B)P(B)+P(A|B^c)P(B^c)}\approx 9\%$。
如果再次检测仍呈阳性，且两次检测效率不变，结果彼此独立，则此人真的患病的概率为 $P(B|A_1A_2)=\frac{P(A_1A_2|B)P(B)}{P(A_1A_2|B)P(B)+P(A_1A_2|B^c)P(B^c)}=\frac{P(A_1|B)P(A_2|B)P(B)}{P(A_1|B)P(A_2|B)P(B)+P(A_1|B^c)P(A_2|B^c)P(B^c)}\approx 99\%$。

# 第二章 随机变量

## 2.1 一维随机变量

**定义 2.1.** 随机变量是样本空间上的实值函数。

注意，上述定义是不严格的。

更严谨的定义：若对于可测空间 $(\Omega, \mathscr{F})$ 和函数 $X:\Omega \to \mathbb{R}$，有 $\forall x \in \mathbb{R}, \{\omega | X(\omega) \le x\} \in \mathscr{F}$，则称 $X$ 是 $(\Omega, \mathscr{F})$ 上的随机变量。其中“可测空间”是指 $\mathscr{F}$ 是样本空间 $\Omega$ 上的 $\sigma$-代数。此处不要求“概率空间”，即随机变量的定义并不依赖概率测度 $P$ 的存在。

**例 2.1.** 下表展示了两个随机变量。其中“像集”即 $\{X(\omega)|\omega \in \Omega\}$。

| 试验 | 样本空间 $\Omega$ | 随机变量 $X$ | 像集 |
|---|---|---|---|
| 随机调查 50 人对某议题支持与否 | $\Omega_1 = \{0,1\}^{50}$ | $X_1 =$ “1” 的个数 | $\{0,1,\cdots,50\}$ |
| 随机抽取一名北京成年市民 | $\Omega_2 =$ 所有北京成年市民之集 | $X_2 =$ 其年收入 | $\mathbb{R}$ |

注意，我们经常用“$X_1 = 20$”、“$X_2 > 100000$”等简化的记号来表示事件。例如，前者实际上指的是 $\{\omega \in \Omega_1 | X_1(\omega) = 20\}$。

诸如此类的试验结果集合需是事件，这体现出前述的随机变量严谨定义的意义。事实上，如果满足该严谨定义，则对于任意可测集 $I \subset \mathbb{R}$，都有 $\{\omega \in \Omega | X(\omega) \in I\} \in \mathscr{F}$。

随机变量是试验结果的数值摘要，起到一种概括的作用。随机变量的“随机”要素来自于样本点 $\omega \in \Omega$ 的随机选择。在实际应用中，随机变量常常比样本空间具有更直观的意义。

随机变量可以分为：

1. 离散型：至多可数多个取值
2. 连续型：区间型取值（非严格定义）
3. 其他

“其他”中的一个非常特殊的子类是所谓的混合型随机变量。

**定义 2.2.** 对于随机变量 $X$ 和 $\mathbb{R}$ 的可测子集 $I$（例如 $I=(a,b]$），令 $X^{-1}(I)=\{\omega\in\Omega|X(\omega)\in I\}\subset\Omega$ 为 $I$ 的原像集，定义记号 $P(X\in I)$ 表示"$X$ 的取值在 $I$ 中的概率"，其值为 $P(X^{-1}(I))$。

例如，$P(a<X\leq b)=P(\{\omega|X(\omega)\in(a,b]\})$。

**定义 2.3.** $F_X(x)=P(X\leq x),\forall x\in\mathbb{R}$ 称为随机变量 $X$ 的累积分布函数（Cumulative Distribution Function, CDF）。下标 $X$ 在无歧义时可省略。

我们有 $P(a<X\leq b)=F(b)-F(a)$。

**例 2.2.** 令 $X$ 表示掷两个均匀六面骰所得的点数和，则 $X$ 的分布表（详见 2.2 节）为

| $X$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $P$ | 1/36 | 2/36 | 3/36 | 4/36 | 5/36 | 6/36 | 5/36 | 4/36 | 3/36 | 2/36 | 1/36 |

相应的 CDF 见图 2.1。

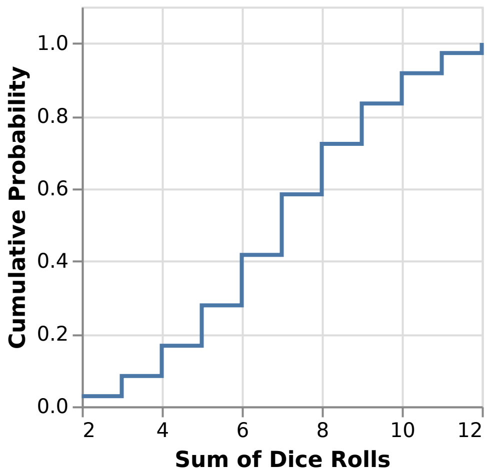



图 2.1: $X$ 的 CDF 图象

注：由于软件限制，各个阶跃点的绘制方式不太规范，实际上从其左侧逼近应该为一个空圈，例如 $F(3)=3/36$ 而不是 $1/36$。另外，$\forall x<2,F(x)=0;\forall x\geq 12,F(x)=1$。

**命题 2.1.** CDF 的性质：

1. $F$ 单调递增（未必严格单调递增）
2. $\lim_{x\to+\infty}F(x)=1,\lim_{x\to-\infty}F(x)=0$
3. $F$ 右连续

可以证明，上述三条性质是任意函数 $F:\mathbb{R}\to\mathbb{R}$ 成为 CDF 的充要条件。

思考：如果将 CDF 的定义改为 $P(X<x)$，上述性质会如何变化？

**命题 2.2.** 若 $X, Y$ 为随机变量，则 $aX + bY, XY, X/Y$ (需$Y \neq 0$) 都是随机变量。一般地，若 $g$ 为可测函数，则 $g(X, Y)$ 是随机变量。

**定义 2.4.** 设 $X_1, X_2$ 的 CDF 分别为 $F_1, F_2$，称 $X_1$ 与 $X_2$ 同分布，若 $\forall x \in \mathbb{R}, F_1(x) = F_2(x)$。

**命题 2.3.** 随机变量 $X_1$ 与 $X_2$ 同分布的一个充要条件是 $\forall$ 可测集$I \subset \mathbb{R}, P(X_1 \in I) = P(X_2 \in I)$。

注意，同分布不等价于"同变量"，即两个同分布的变量的取值不一定恒等。

**例 2.3.** 掷一次硬币，$X$ 表示正面向上次数，$Y$ 表示反面向上次数，显然 $X$ 与 $Y$ 同分布，但取值不等。

## 2.2 离散随机变量

**定义 2.5.** 离散随机变量 $X$ 的*概率质量函数*（Probability Mass Function, PMF）$f$ 是指该随机变量取各个可能值的概率，即 $f(x) = P(X = x), \forall x \in \mathbb{R}$。可以用*分布表*的形式展示各个可能取值与概率的对应关系。

**命题 2.4.** 如果离散随机变量 $X$ 的所有可能取值为 $\{x_i\}$，则 $X$ 的 PMF 具有如下性质:

1. $f(x_i) = p_i \geq 0, \forall i$
2. $\sum_i p_i = 1$
3. $F(x) = \sum_{x_i \leq x} f(x_i)$

**定义 2.6.** 离散随机变量 $X$ 的*期望*定义为 $\mathrm{E}(X) = \sum_i x_i p_i$。
称 $X$ 的期望*存在*，当且仅当 $\sum_i |x_i| p_i < +\infty$。
当期望存在时，其*方差*定义为 $\mathrm{Var}(X) = \sum_i (x_i - \mathrm{E}(X))^2 p_i = \mathrm{E}((X - \mathrm{E}(X))^2) = \mathrm{E}(X^2) - \mathrm{E}^2(X)$。
当方差有限时，称其算术平方根为 $X$ 的*标准差*，记作 $\mathrm{SD}(X)$。

注意，通常我们所说的一个随机变量的均值指的就是期望。

*标准化*指的是对 $X$ 作线性变换 $\frac{X-\mu}{\sigma}$，其中 $\mu$ 和 $\sigma$ 分别为 $X$ 的期望和标准差，得到均值为 0，标准差为 1 的随机变量。

对于可测函数 $g$，$g(X)$ 也是随机变量，其期望 $\mathrm{E}(g(X)) = \sum_i g(x_i) p_i$。

期望反映了随机变量的集中趋势，而方差反映了其分散程度。

# 2.3 常见离散分布

**定义 2.7.** 称一个随机变量 $X$ 服从 *Bernoulli* 分布，若 $\exists p \in (0,1)$，$X$ 的取值集合为 $\{0,1\}$，且 $P(X=1)=p, P(X=0)=1-p$。记作 $X \sim B(p)$。

$B(p)$ 中的 $p$ 称为该 Bernoulli 分布的参数。后续介绍的其他分布同理。

常将两种取值分别称为“成功”和“失败”。

计算可得，若 $X \sim B(p)$，则 $\mathrm{E}(X)=p, \mathrm{Var}(X)=p(1-p)$。

**定义 2.8.** 称一个随机变量 $X$ 服从二项分布，若 $\exists N \in \mathbb{N}^*$，$p \in (0,1)$，$X$ 的取值集合为 $\{0,1,\cdots,N\}$，且 $P(X=k)=\binom{N}{k}p^k(1-p)^{N-k}(k \in \{0,1,\cdots,N\})$。记作 $X \sim B(N,p)$。

常将 $k$ 理解为“$N$ 次独立 Bernoulli 试验中的成功次数”。

计算可得，若 $X \sim B(N,p)$，则 $\mathrm{E}(X)=Np, \mathrm{Var}(X)=Np(1-p)$。

**定义 2.9.** 称一个随机变量 $X$ 服从 *Poisson* 分布，若 $\exists \lambda > 0$，$X$ 的取值集合为 $\mathbb{N}$，且 $P(X=k)=\frac{\lambda^k e^{-\lambda}}{k!}(k \in \mathbb{N})$。记作 $X \sim P(\lambda)$。

计算可得，若 $X \sim P(\lambda)$，则 $\mathrm{E}(X)=\lambda, \mathrm{Var}(X)=\lambda$。

对 Poisson 分布的一种常见理解是“一段时间内某个小概率事件发生的次数”所服从的分布。例如，观察时间 $(0,1]$ 内某路口的交通事故数 $X$，将 $(0,1]$ 区间等分成 $n$ 个小区间，即 $l_i=(\frac{i-1}{n},\frac{i}{n}](i=1,2,\cdots,n)$。考虑到 $n$ 很大时，每个区间的长度很小，作如下假设：

1. 每段区间内，至多发生一次事故
2. $l_i$ 上发生一次事故的概率与区间长度（$1/n$）成正比，为 $p=\lambda/n$
3. 各区间内是否发生事故彼此独立

则 $P(X=k)=\binom{n}{k}p^k(1-p)^{n-k} \to \frac{\lambda^k e^{-\lambda}}{k!}(n \to +\infty)$，即 $X \sim P(\lambda)$。

**例 2.4.** 设某医院平均每天出生婴儿数为 $\lambda$，则接下来 $t$ 天内出生婴儿数服从参数为 $t\lambda$ 的 Poisson 分布。

对于一般的二项分布 $X \sim B(N,p)$，若 $p$ 很小，$N$ 很大，而 $\lambda=Np$ 不太大，则近似有 $X \sim P(\lambda)$，且近似误差不超过 $\min\{p, Np^2\}$。

进一步，若 $N$ 次 Bernoulli 试验并非严格独立，但满足弱相依条件，则 Poisson 分布仍为一种较好的近似。

**例 2.5.** （配对问题）

$A_i$ 表示第 $i$ 个人拿到自己的帽子，则 $P(A_i)=1/n, P(A_i|A_j)=\frac{1}{n-1}(j \neq i)$，当 $n$ 很大时，$1/n$

和 $\frac{1}{n-1}$ 很接近，可以认为满足弱相依条件。

记 $X$ 为拿到自己帽子的人数，则 $X$ 近似服从参数为 $\lambda = np = n \cdot \frac{1}{n} = 1$ 的 Poisson 分布，即 $P(X=k) \approx \frac{e^{-1}}{k!}$。

我们用常规做法检查这种近似是否合理。首先考虑指定的某 $k$ 人，记事件 $E$ 表示这 $k$ 人拿到自己的帽子，事件 $F$ 表示其余 $(n-k)$ 人未拿到自己的帽子，则 $P(EF) = P(E)P(F|E) = \frac{(n-k)!}{n!} \cdot P_{n-k}$，其中 $P_{n-k}$ 为 $(n-k)$ 人随机拿帽子时无人拿对的概率。那么 $P(X=k) = \binom{n}{k}P(EF) = \frac{1}{k!}P_{n-k} \to \frac{e^{-1}}{k!}(n \to +\infty)$。这说明前述的近似是较好的。

## 2.4 连续随机变量

**定义 2.10.** 对随机变量 $X$，若存在 $f : \mathbb{R} \to [0, +\infty)$，使得 $\forall$ 可测集 $I \subset \mathbb{R}$，都有 $P(X \in I) = \int_I f(x)\mathrm{d}x$，则称 $X$ 为*连续型随机变量*，$f$ 称为其*概率密度函数*（Probability Density Function, PDF）。

**命题 2.5.** 连续随机变量 $X$ 的 PDF 具有如下性质：

1. $\int_{-\infty}^{+\infty} f(x)\mathrm{d}x \equiv 1$
2. $P(a < X \le b) = \int_a^b f(x)\mathrm{d}x = P(a \le X \le b) = P(a \le X < b) = P(a < X < b)$
3. $P(X = a) \equiv 0, \forall a \in \mathbb{R}$
4. 若 $f$ 在 $x_0$ 处连续，则 $P(x_0 - \delta < X < x_0 + \delta) = \int_{x_0-\delta}^{x_0+\delta} f(t)\mathrm{d}t \approx f(x_0) \cdot 2\delta$
5. $F(x) = \int_{-\infty}^{x} f(t)\mathrm{d}t$ 连续，且若 $f$ 在 $x$ 处连续，有 $F'(x) = f(x)$
6. PDF 若存在，则不唯一（可以修改其在任意零测集上的值，得到不同的 PDF）

**定义 2.11.** 连续随机变量 $X$ 的期望定义为 $\mathrm{E}(X) = \int_{-\infty}^{+\infty} xf(x)\mathrm{d}x$。

称 $X$ 的期望存在，当且仅当 $\int_{-\infty}^{+\infty} |x|f(x)\mathrm{d}x < +\infty$。

当期望存在时，其*方差*定义为 $\mathrm{Var}(X) = \int_{-\infty}^{+\infty}(x - \mathrm{E}(x))^2 f(x)\mathrm{d}x = \mathrm{E}((X - \mathrm{E}(X))^2) = \mathrm{E}(X^2) - \mathrm{E}^2(X)$。

当方差有限时，称其算术平方根为 $X$ 的*标准差*，记作 $\mathrm{SD}(X)$。

对于可测函数 $g$，$g(X)$ 也是随机变量，其期望 $\mathrm{E}(g(X)) = \int_{-\infty}^{+\infty} g(x)f(x)\mathrm{d}x$。

## 2.5 常见连续分布

**定义 2.12.** 称一个连续型随机变量 $X$ 服从*均匀分布*，若其 PDF 为 $f(x) = \frac{1}{b-a}(x \in (a,b))$，$f$ 在其余各处取 0。记作 $X \sim U(a,b)$。

常将 $X \sim U(0,1)$ 称为随机数。

计算可得，若 $X \sim U(a,b)$，则 $\mathrm{E}(X)=\frac{a+b}{2}, \mathrm{Var}(X)=\frac{(b-a)^2}{12}$。

**定义 2.13.** 称一个连续型随机变量 $X$ 服从正态分布，若其 PDF 为 $f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}(\sigma>0)$。记作 $X \sim N(\mu,\sigma^2)$。

计算可得，若 $X \sim N(\mu,\sigma^2)$，则 $\mathrm{E}(X)=\mu, \mathrm{Var}(X)=\sigma^2$。

著名的“经验法则”见图 2.2。

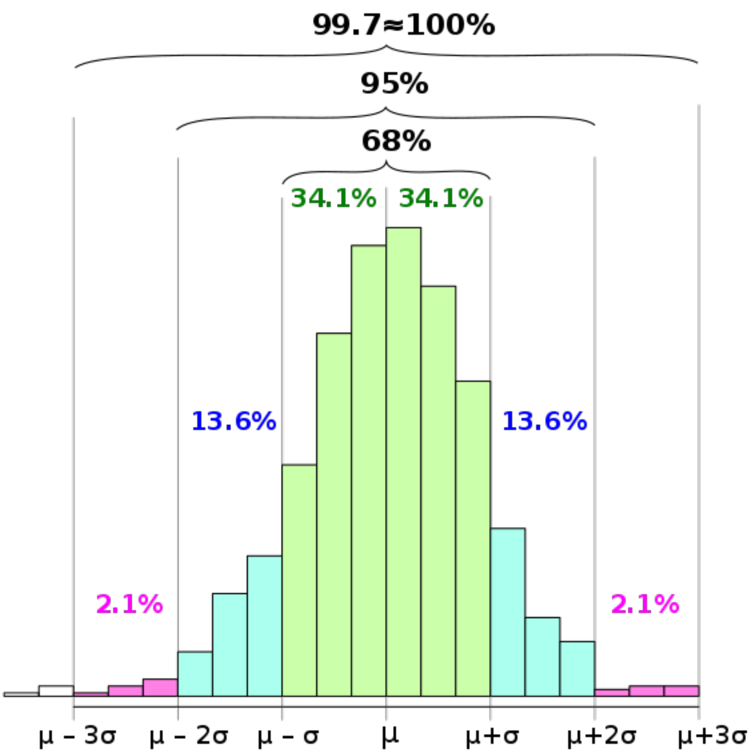



图 2.2: 经验法则

$X \sim N(\mu,\sigma^2)$ 的充要条件是 $Y=\frac{X-\mu}{\sigma} \sim N(0,1)$。$N(0,1)$ 称为标准正态分布。

**定义 2.14.** 称一个连续型随机变量 $X$ 服从指数分布，若其 PDF 为 $f(x)=\lambda e^{-\lambda x}(\lambda>0, x>0)$，$f$ 在其余各处取 0。记作 $X \sim Exp(\lambda)$。

指数分布常用于刻画等待时间、寿命等。

计算可得，若 $X \sim Exp(\lambda)$，则 $\mathrm{E}(X)=1/\lambda, \mathrm{Var}(X)=1/\lambda^2$。

指数分布有另一种符号约定，以 $\beta=1/\lambda$ 为参数，一些数学软件可能采用此种约定。

指数分布的 CDF 为 $F(x)=1-e^{-\lambda x}(x>0)$，所谓的“尾概率”为 $P(X>x)=1-F(x)=e^{-\lambda x}(x>0)$。

**例 2.6.** 设某医院平均每天出生婴儿数为 $\lambda$，现在观察到一名婴儿出生，则接下来 $t$ 天内有婴儿出生的概率为 $P(X \le t)$，其中 $X$ 表示到下一个婴儿出生所需等待的时间。
记 $N(t)$ 为 $t$ 天内出生婴儿数,我们已经知道 $N(t) \sim P(t\lambda)$,则 $P(X > t) = P(N(t) = 0) = e^{-\lambda t}$，故 $P(X \le t) = 1 - e^{-\lambda t}$，即 $X$ 服从参数为 $\lambda$ 的指数分布。

我们从另一个角度理解指数分布。

首先引入失效率或危险率的概念。设 $X$ 为连续型随机变量（表示某种零件的寿命），其 CDF 为 $F(x)$，且 $F(0) = 0$。考虑条件概率 $P(x < X < x + \mathrm{d}x | X > x) = \frac{P(x<X<x+\mathrm{d}x)}{P(X>x)} = \frac{F(x+\mathrm{d}x)-F(x)}{1-F(x)} \approx \frac{F'(x)}{1-F(x)}\mathrm{d}x$，即“年龄”为 $x$ 的零件不能继续工作的条件概率密度为 $\frac{F'(x)}{1-F(x)}$，称其为瞬时失效率 $\lambda(x)$，则 $F(x) = 1 - e^{-\int_0^x \lambda(t)\mathrm{d}t}$。

在“无老化”假设下，即 $\lambda(t) \equiv \lambda$ 不随时间变化，则 $F(x) = 1 - e^{-\lambda t}(x > 0)$，$X$ 服从指数分布。

指数分布有所谓“无记忆性”: $P(X > t+s|X > s) = \frac{P(X>t+s)}{P(X>s)} = e^{-\lambda t} = P(X > t)(t, s > 0)$。

“无老化”假设并不总是成立。为此，可以进行一定程度的改进，例如令 $\lambda(x) = \alpha\frac{x^{\alpha-1}}{\beta^\alpha}(x > 0, \alpha, \beta > 0$ 为常数)，则 $F(x) = 1 - e^{-(\frac{x}{\beta})^\alpha}(x > 0)$，称之为 *Weibull* 分布。当 $\alpha = 1$ 时，Weibull 分布退化为参数为 $1/\beta$ 的指数分布。

总览至此介绍过的各个分布的参数，可以将其大致分为以下几类:

1. 位置参数：决定了分布平移到的位置，通常在 PMF/PDF 中体现为 $f(x) = g(x - \cdot)$ 的形式，如正态分布的参数 $\mu$
2. 尺度参数：决定了分布伸缩的程度，通常在 PMF/PDF 中体现为 $f(x) = g(\frac{x}{\cdot})$ 的形式，如正态分布的参数 $\sigma$、Weibull 分布的参数 $\beta$
3. 形状参数：决定了分布的形状，如 Weibull 分布的参数 $\alpha$

## 2.6 随机变量的函数

对于随机变量 $X$ 和可测函数 $g$，$Y = g(X)$ 也是随机变量。特别地，若 $X$ 为离散型随机变量，则 $Y$ 也离散。但若 $X$ 为连续型随机变量，$Y$ 未必连续。

**例 2.7.** $X \sim Exp(\lambda)$，$Y = \begin{cases} 0, & X \le t_0, \\ 1, & X > t_0, \end{cases}$ 其中 $t_0 > 0$ 为常数，则 $Y \sim B(e^{-\lambda t_0})$。

**例 2.8.** 设 $X$ 为连续型随机变量，PDF 为 $f(x)$，考虑 $Y = X^2$。
从 CDF 入手，$\forall y > 0, P(Y \le y) = P(X^2 \le y) = P(-\sqrt{y} \le X \le \sqrt{y}) = \int_{-\sqrt{y}}^{\sqrt{y}} f(x)\mathrm{d}x$，有 $Y$ 的 PDF 为 $l(y) = \frac{\mathrm{d}}{\mathrm{d}y}P(Y \le y) = \frac{1}{2\sqrt{y}}(f(\sqrt{y}) + f(-\sqrt{y}))(y > 0)$。
特别地，若 $X \sim N(0, 1)$，称 $Y$ 服从自由度为 1 的 $\chi^2$-分布，读作“卡方分布”。

若 $Y = g(X)$ 为随机变量，可以计算 $Y$ 的分布如下：

- $P(Y = y) = P(g(X) = y) = P(X \in g^{-1}(y))$
- $P(Y \leq y) = P(g(X) \leq y) = P(X \in g^{-1}((-\infty, y]))$

# 第三章　联合分布

## 3.1　随机向量

**定义 3.1.** 称 $(X_1, X_2, \cdots, X_n) : \Omega \to \mathbb{R}^n$ 为（$n$ 维）随机向量，若 $\{X_i\}_{i=1}^n$ 均为随机变量。

**定义 3.2.** $n$ 维随机向量的（联合）（累积）分布函数（CDF）定义为 $F(x_1, \cdots, x_n) = P(X_1 \leq x_1, \cdots, X_n \leq x_n), \forall (x_1, \cdots, x_n) \in \mathbb{R}^n$。

对于 $n = 2$（二元分布）的情形，常用 $(X, Y)$ 来表示随机向量，对应的 CDF 为 $F(x, y)$。

## 3.2　离散分布

**定义 3.3.** 称 $n$ 维随机向量 $(X_1, \cdots, X_n)$ 是离散的，当且仅当 $\{X_i\}_{i=1}^n$ 均为离散随机变量。
离散随机向量 $(X_1, \cdots, X_n)$ 的（联合）概率质量函数（PMF）定义为 $f(x_1, \cdots, x_n) = P(X_1 = x_1, \cdots, X_n = x_n), \forall (x_1, \cdots, x_n) \in \mathbb{R}^n$。

**命题 3.1.** 离散随机向量 $(X_1, \cdots, X_n)$ 的 PMF 具有如下性质：

1. $f(x_1, \cdots, x_n) \geq 0, \forall (x_1, \cdots, x_n) \in \mathbb{R}^n$
2. $\sum_{x_i \in \{X_i(\omega) | \omega \in \Omega\}, \forall i \in \{1, \cdots, n\}} f(x_1, \cdots, x_n) \equiv 1$

注意第 2 条性质中求和的项数为至多可数，原因是有限个至多可数集的笛卡尔积仍是至多可数集。

**例 3.1.** 设 $\{B_i\}_{i=1}^n$ 为 $\Omega$ 的一个分割（分割的定义见 1.8 节），$P(B_i) = p_i \geq 0, \forall i \in \{1, \cdots, n\}$，$\sum_{i=1}^n p_i = 1$。
进行 $N$ 次独立试验，设 $\forall i \in \{1, \cdots, n\}$，有 $X_i$ 个试验结果落在 $B_i$ 中，则若 $k_1 + \cdots + k_n = N$，其中 $k_i$ 均为非负整数，有 $P(X_1 = k_1, \cdots, X_n = k_n) = \binom{N}{k_1, \cdots, k_n} p_1^{k_1} \cdots p_n^{k_n}$。其中 $\binom{N}{k_1, \cdots, k_n} = \frac{N!}{k_1! \cdots k_n!}$ 为多项式 $(a_1 + \cdots + a_n)^N$ 中 $a_1^{k_1} \cdots a_n^{k_n}$ 项的系数。
称 $(X_1, \cdots, X_n)$ 服从多项分布。

## 3.3 连续分布

**定义 3.4.** 对 $n$ 维随机向量 $(X_1,\cdots,X_n)$，若存在 $f:\mathbb{R}^n\to[0,+\infty)$，使得 $\forall$ 可测集 $Q\subset\mathbb{R}^n$，都有 $P((X_1,\cdots,X_n)\in Q)=\int_Q f(x_1,\cdots,x_n)\mathrm{d}x_1\cdots\mathrm{d}x_n$，则称 $(X_1,\cdots,X_n)$ 为连续型随机向量，$f$ 称为其（联合）概率密度函数（PDF）。

**命题 3.2.** 连续随机向量 $(X_1,\cdots,X_n)$ 的 PDF 具有如下性质：

1. $\int_{\mathbb{R}^n} f(x_1,\cdots,x_n)\mathrm{d}x_1\cdots\mathrm{d}x_n\equiv 1$
2. 以 $n=2$ 为例，$F(x,y)=\int_{-\infty}^x\int_{-\infty}^y f(t,s)\mathrm{d}s\mathrm{d}t, f(a,b)=\frac{\partial^2 F}{\partial x\partial y}(a,b)$, a.e.

其中 a.e. 表示 “almost everywhere”。

**例 3.2.** 矩形域上的均匀分布的 PDF：$f(x,y)=\begin{cases}\dfrac{1}{(b-a)(d-c)}, & (x,y)\in(a,b)\times(c,d),\\ 0, & \text{其他}.\end{cases}$

**例 3.3.** 二元正态分布 $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$ 的 PDF：
$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2}\frac{1}{\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}((\frac{x-\mu_1}{\sigma_1})^2+(\frac{y-\mu_2}{\sigma_2})^2-2\rho\frac{x-\mu_1}{\sigma_1}\frac{y-\mu_2}{\sigma_2})},\forall(x,y)\in\mathbb{R}^2,\sigma_1,\sigma_2>0,|\rho|<1$。
令 $\boldsymbol{x}=\begin{bmatrix}\frac{x-\mu_1}{\sigma_1}\\ \frac{y-\mu_2}{\sigma_2}\end{bmatrix}, W=\frac{1}{1-\rho^2}\begin{bmatrix}1 & -\rho\\ -\rho & 1\end{bmatrix}$，$W=A^{\mathrm{T}}A$ 为正定矩阵 $W$ 的 Cholesky 分解，则 $-\frac{1}{2(1-\rho^2)}((\frac{x-\mu_1}{\sigma_1})^2+(\frac{y-\mu_2}{\sigma_2})^2-2\rho\frac{x-\mu_1}{\sigma_1}\frac{y-\mu_2}{\sigma_2})=-\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}W\boldsymbol{x}=-\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}A^{\mathrm{T}}A\boldsymbol{x}=-\frac{1}{2}(A\boldsymbol{x})^{\mathrm{T}}(A\boldsymbol{x})$。
上述 Cholesky 分解的结果为 $A=\frac{1}{\sqrt{1-\rho^2}}\begin{bmatrix}1 & -\rho\\ 0 & \pm\sqrt{1-\rho^2}\end{bmatrix}$ 或 $A=\frac{1}{\sqrt{1-\rho^2}}\begin{bmatrix}-1 & \rho\\ 0 & \pm\sqrt{1-\rho^2}\end{bmatrix}$。

## 3.4 边际分布

对 $n$ 维随机向量 $(X_1,\cdots,X_n)$，称 $F_i(x)=P(X_i\le x)=P(X_i\le x,-\infty<X_j<+\infty,\forall j\neq i)$ 为 $X_i$ 的边际分布。

例如，若 $n=2$，随机向量 $(X,Y)$ 有 CDF $F(x,y)$，则 $X$ 的边际分布为 $F_X(x)=P(X\le x)=P(X\le x,Y\in\mathbb{R})=\lim\limits_{y\to+\infty}P(X\le x,-\infty<Y\le y)=\lim\limits_{y\to+\infty}F(x,y)$。

若 $n=3$，随机向量 $(X,Y,Z)$ 有 CDF $F(x,y,z)$，则 $F_X(x)=\lim\limits_{y,z\to+\infty}F(x,y,z)$，而 $(X,Y)$ 的边际分布为 $F_{X,Y}(x,y)=P(X\le x,Y\le y)=P(X\le x,Y\le y,-\infty<Z<+\infty)=\lim\limits_{z\to+\infty}F(x,y,z)$。

**例 3.4.** 设二维随机向量 $(X,Y)$ 的 CDF 为 $F(x,y)$，则 $\forall a,b\in\mathbb{R},P(X>a,Y>b)=1-F_X(a)-F_Y(b)+F(a,b)$。

对于离散型随机向量,以 $n=2$ 为例,定义边际 *PMF* 为 $P(X=x)=\sum\limits_{y}P(X=x,Y=y)$。

对于连续型随机向量,以 $n=2$ 为例,设联合 PDF 为 $f(x,y)$,则 $F_X(x)=P(X\le x,Y\in\mathbb{R})=\int_{-\infty}^{x}\int_{-\infty}^{+\infty}f(t,s)\mathrm{d}s\mathrm{d}t$,则 $X$ 的边际 *PDF* 为 $f_X(x)=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}y$。

**例 3.5.** $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$,则 $f_X(x)=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}y=\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}$,即 $X\sim N(\mu_1,\sigma_1^2)$。同理 $Y\sim N(\mu_2,\sigma_2^2)$。

## 3.5 条件分布

以 $n=2$ 为例说明条件分布的概念,考虑随机向量 $(X,Y)$。

对于离散型随机向量,设联合 PMF 为 $P(X=a_i,Y=b_j)=p_{ij}\ge 0,\sum_{i,j}p_{ij}\equiv 1$,则在 $Y=b_j$ 条件下的 $X$ 的条件 *PMF* 为 $P(X=a_i|Y=b_j)=\frac{P(X=a_i,Y=b_j)}{P(Y=b_j)}=\frac{p_{ij}}{\sum_k p_{kj}}$。条件 PMF 满足 $\sum_i P(X=a_i|Y=b_j)\equiv 1,\forall j$。

对于连续型随机向量,设联合 PDF 为 $f(x,y)$,首先考虑条件概率 $P(X\le x|y\le Y\le y+\mathrm{d}y)=\frac{P(X\le x,y\le Y\le y+\mathrm{d}y)}{P(y\le Y\le y+\mathrm{d}y)}=\frac{\int_{-\infty}^{x}\int_{y}^{y+\mathrm{d}y}f(t,s)\mathrm{d}s\mathrm{d}t}{\int_{y}^{y+\mathrm{d}y}f_Y(s)\mathrm{d}s}$,对 $x$ 求导得 $X$ 在 $y\le Y\le y+\mathrm{d}y$ 条件下的条件 PDF 为 $\frac{\int_{y}^{y+\mathrm{d}y}f(x,s)\mathrm{d}s}{\int_{y}^{y+\mathrm{d}y}f_Y(s)\mathrm{d}s}\to\frac{f(x,y)}{f_Y(y)}(\mathrm{d}y\to 0)$。

**定义 3.5.** 对于连续型随机向量 $(X,Y)$,设联合 PDF 为 $f(x,y)$,若 $f_Y(y)>0$,则称 $X$ 在 $Y=y$ 条件下的条件 *PDF* 为 $f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)}$。

可以验证 $f_{X|Y}(x|y)$ 满足 PDF 的各性质。

相应的条件 *CDF* 为 $F_{X|Y}(a|y)=P(X\le a|Y=y)=\int_{-\infty}^{a}f_{X|Y}(x|y)\mathrm{d}x$。

我们熟知的各个定理均有适用于连续型随机向量的版本:

1. $f(x,y)=f_{X|Y}(x|y)f_Y(y)=f_{Y|X}(y|x)f_X(x)$(乘法法则)
2. $f_X(x)=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}y=\int_{-\infty}^{+\infty}f_{X|Y}(x|y)f_Y(y)\mathrm{d}y$(全概率公式)
3. $f_{Y|X}(y|x)=\frac{f(x,y)}{f_X(x)}=\frac{f_{X|Y}(x|y)f_Y(y)}{\int_{-\infty}^{+\infty}f_{X|Y}(x|y)f_Y(y)\mathrm{d}y}$(Bayes 公式)

**例 3.6.** $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$,则 $f_{Y|X}(y|x)=\frac{f(x,y)}{f_X(x)}=\frac{1}{\sqrt{2\pi}\sigma_2}\frac{1}{\sqrt{1-\rho^2}}e^{-\frac{(y-(\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)))^2}{2(1-\rho^2)\sigma_2^2}}$,即 $Y|X=x\sim N(\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1),(1-\rho^2)\sigma_2^2)$。

## 3.6 独立性

**定义 3.6.** 设二维随机向量 $(X,Y)$ 的 CDF 为 $F(x,y)$,若 $F(x,y)=F_X(x)F_Y(y),\forall x,y\in\mathbb{R}$,则称 $X,Y$ 相互独立。

可以证明，对于二维离散型（或连续型）随机向量 $(X, Y)$，$X, Y$ 相互独立的充要条件是 $f(x, y) = f_X(x) f_Y(y), \forall x, y \in \mathbb{R}$，其中 $f(x, y)$ 为联合 PMF（或 PDF）。

**定义 3.7.** 设 $n$ 维随机向量 $(X_1, \cdots, X_n)$ 的 CDF 为 $F(x_1, \cdots, x_n)$，若 $F(x_1, \cdots, x_n) = F_1(x_1) \cdots F_n(x_n), \forall x_1, \cdots, x_n \in \mathbb{R}$，则称 $X_1, \cdots, X_n$ 相互独立。

可以证明，对于 $n$ 维离散型（或连续型）随机向量 $(X_1, \cdots, X_n)$，$X_1, \cdots, X_n$ 相互独立的充要条件是 $f(x_1, \cdots, x_n) = f_1(x_1) \cdots f_n(x_n), \forall x_1, \cdots, x_n \in \mathbb{R}$，其中 $f(x_1, \cdots, x_n)$ 为联合 PMF（或 PDF）。

**定理 3.1.**

1. 若 $X_1, \cdots, X_n$ 相互独立，则 $\forall m \in \{1, \cdots, n-1\}$，可测函数 $g_1, g_2$，有 $Y_1 = g_1(X_1, \cdots, X_m)$ 与 $Y_2 = g_2(X_{m+1}, \cdots, X_n)$ 相互独立。
2. 若 $n$ 维连续型随机向量 $(X_1, \cdots, X_n)$ 的联合 PDF 满足

$$f(x_1, \cdots, x_n) = g_1(x_1) \cdots g_n(x_n), \forall x_1, \cdots, x_n \in \mathbb{R},$$

   其中 $g_i : \mathbb{R} \to [0, +\infty), \forall i \in \{1, \cdots, n\}$，则 $X_1, \cdots, X_n$ 相互独立，且 $X_i$ 的边际 PDF $f_i$ 与 $g_i$ 相差常数因子，$\forall i \in \{1, \cdots, n\}$。

**例 3.7.** 设 $(X, Y)$ 服从如图 3.1 的三角形域 $D$ 上的均匀分布，即 $f(x, y) = \begin{cases} c, & (x, y) \in D, \\ 0, & \text{其他}, \end{cases}$ 则 $X, Y$ 不独立。

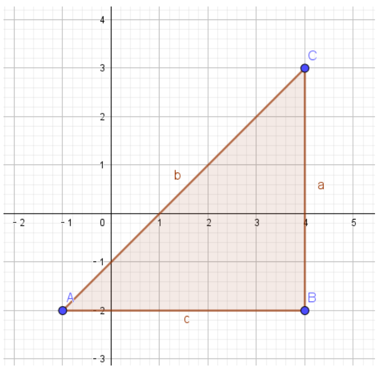



图 3.1: 三角形域上的均匀分布

## 3.7 随机向量的函数

本节中，考虑给定随机向量 $(X_1, \cdots, X_n)$ 和可测函数 $g$，如何求 $Y = g(X_1, \cdots, X_n)$ 的分布。

首先介绍“直接法”。

**例 3.8.** $X_i \sim B(n_i,p)(i=1,2)$ 独立，$Y=X_1+X_2$，则 $\forall k\in\{0,1,\cdots,n_1+n_2\}$，

$$
\begin{aligned}
&P(Y=k)\\
=&P(X_1+X_2=k)\\
=&\sum_{k_1=0}^{k}P(X_1=k_1,X_2=k-k_1)\\
=&\sum_{k_1=0}^{k}P(X_1=k_1)P(X_2=k-k_1)\\
=&\sum_{k_1=0}^{k}\tbinom{n_1}{k_1}p^{k_1}(1-p)^{n_1-k_1}\tbinom{n_2}{k-k_1}p^{k-k_1}(1-p)^{n_2-(k-k_1)}\\
=&\left(\sum_{k_1=0}^{k}\tbinom{n_1}{k_1}\tbinom{n_2}{k-k_1}\right)p^k(1-p)^{n_1+n_2-k}\\
=&\tbinom{n_1+n_2}{k}p^k(1-p)^{n_1+n_2-k}
\end{aligned}
$$

因此 $Y\sim B(n_1+n_2,p)$。

**例 3.9.** 随机向量 $(X_1,X_2)$ 有联合 PDF $f(x_1,x_2)$，且 $X_1>0$，考虑 $Y=X_2/X_1$，有 $\forall y\in\mathbb{R}, P(Y\le y)=P(\frac{X_2}{X_1}\le y)=P(X_2\le X_1y)=\int_D f(x_1,x_2)\mathrm{d}x_1\mathrm{d}x_2=\int_0^{+\infty}\int_{-\infty}^{yx_1}f(x_1,x_2)\mathrm{d}x_2\mathrm{d}x_1$，作 $x_2=x_1t$ 换元得 $P(Y\le y)=\int_0^{+\infty}\int_{-\infty}^{y}f(x_1,x_1t)x_1\mathrm{d}t\mathrm{d}x_1$，故 $Y$ 的 PDF 为 $l(y)=\int_0^{+\infty}x_1f(x_1,yx_1)\mathrm{d}x_1$。

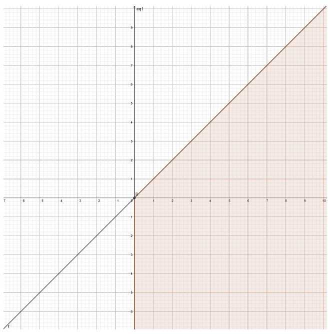


图 3.2: 区域 $D$ 的范围，其中边界线的斜率为 $y$

接下来介绍“密度函数变换法”。

设随机向量 $(X_1,X_2)$ 有联合 PDF $f(x_1,x_2)$，且有可逆可微的映射关系 $\begin{cases}Y_1= & g_1(X_1,X_2),\\ Y_2= & g_2(X_1,X_2),\end{cases}$ 据此解出逆映射 $\begin{cases}X_1= & h_1(Y_1,Y_2),\\ X_2= & h_2(Y_1,Y_2),\end{cases}$ 则对于任意可测集 $A$，若 $(h_1,h_2)$ 将 $A$ 映射到集合 $B$，

则由可逆性可知 $B$ 在 $(g_1, g_2)$ 的映射下的值域为 $A$。因此 $P((Y_1, Y_2) \in A) = P((X_1, X_2) \in B) = \int_B f(x_1, x_2)\mathrm{d}x_1\mathrm{d}x_2 = \int_A f(h_1(y_1, y_2), h_2(y_1, y_2))|J|\mathrm{d}y_1\mathrm{d}y_2$，其中 $J$ 为 Jacobi 行列式 $\det \begin{bmatrix} \frac{\partial h_1}{\partial y_1} & \frac{\partial h_1}{\partial y_2} \\ \frac{\partial h_2}{\partial y_1} & \frac{\partial h_2}{\partial y_2} \end{bmatrix}$，因此 $(Y_1, Y_2)$ 的联合 PDF 为 $l(y_1, y_2) = f(h_1(y_1, y_2), h_2(y_1, y_2))|J|$。

**例 3.10.** 随机向量 $(X_1, X_2)$ 有联合 PDF $f(x_1, x_2)$，为求 $Y = X_1 + X_2$ 的 PDF，引入 $Z = X_1$，则 $\begin{cases} X_1 = & Z, \\ X_2 = & Y - Z, \end{cases}$ Jacobi 行列式为 $\det \begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix} = -1$，故 $(Y, Z)$ 的联合 PDF 为 $f(z, y - z)|-1| = f(z, y - z)$，$Y$ 的边际 PDF 为 $l_Y(y) = \int_{-\infty}^{+\infty} f(z, y - z)\mathrm{d}z$。

上例中，若 $X_1, X_2$ 相互独立，则 $f(x_1, x_2) = f_1(x_1)f_2(x_2) \Rightarrow l_Y(y) = \int_{-\infty}^{+\infty} f_1(z)f_2(y-z)\mathrm{d}z$，这称之为 $f_1$ 和 $f_2$ 的卷积，记作 $f_1 * f_2$。

特别地，若 $(X_1, X_2) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$，则 $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2)$。

利用上述随机向量的函数的 PDF 求解方法，可以得到所谓卡方分布（$\chi^2$-分布）、$t$-分布和 $F$-分布的 PDF。这些分布的表达式较为复杂，在此不一一罗列。感兴趣的同学可以查阅资料，简单了解一下它们与标准正态分布的联系。

# 第四章　随机变量的数字特征

## 4.1　期望

离散型和连续型随机变量的期望分别参见定义 2.6 和定义 2.11。

对于随机向量，期望自然推广定义为 $\mathrm{E}((X_1,\cdots,X_n)) = (\mathrm{E}(X_1),\cdots,\mathrm{E}(X_n))$。

**命题 4.1.** 期望有如下性质：

1. 离散型和连续型随机向量的函数的期望 $\mathrm{E}(g(X_1,\cdots,X_n))$ 分别等于
   $\sum_{x_i\in\{X_i(\omega)|\omega\in\Omega\},\forall i\in\{1,\cdots,n\}} g(x_1,\cdots,x_n)f(x_1,\cdots,x_n)$
   和 $\int_{\mathbb{R}^n} g(x_1,\cdots,x_n)f(x_1,\cdots,x_n)\mathrm{d}x_1\cdots\mathrm{d}x_n$，
   其中 $g$ 为可测函数，$f$ 分别为联合 PMF 与联合 PDF
2. $\mathrm{E}(aX+bY) = a\mathrm{E}(X)+b\mathrm{E}(Y), \forall$ 常数 $a,b\in\mathbb{R}$
3. 若 $X_1,\cdots,X_n$ 相互独立，则 $\mathrm{E}(X_1\cdots X_n) = \mathrm{E}(X_1)\cdots\mathrm{E}(X_n)$

## 4.2　分位数

**定义 4.1.** 设 $X$ 为连续型随机变量，若 $P(X\le m) = F(m) = 1/2$，则称 $m$ 为 $X$ 的中位数。

和均值一样，中位数也是随机变量集中趋势的一种刻画。中位数不一定唯一。

若 $m$ 是连续型随机变量 $X$ 的中位数，则 $P(X<m) = P(X>m) = 1/2$。

以下给出更一般的中位数定义。

**定义 4.2.** 对随机变量 $X$，若 $P(X<m)\le 1/2$，且 $P(X>m)\le 1-1/2 = 1/2$，则称 $m$ 为 $X$ 的中位数。

**例 4.1.** 设离散型随机变量 $X$ 的分布表为

| $X$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $P$ | 1/3 | 1/2 | 1/12 | 1/12 |

则其中位数为 2。

**定义 4.3.** 对随机变量 $X$，$\forall\alpha\in(0,1)$，若 $P(X<a)\leq\alpha$ 且 $P(X>a)\leq 1-\alpha$，则称 $a$ 为 $X$ 的（下侧）$\alpha$-分位数。

上述定义的 $\alpha$-分位数是不唯一的。为了唯一性，考虑定义 $F^{-1}(\alpha)=\inf\{x|F(x)\geq\alpha\}$。

我们给出众数（mode）的方便定义：$f(x)$ 的最大值点，其中 $f(x)$ 为 PMF 或 PDF。由于 PDF 可在任意零测集上修改取值，故这一定义并非严谨的。

## 4.3 方差

离散型和连续型随机变量的方差分别参见定义 2.6 和定义 2.11。

方差的意义：若 $X$ 为收益率，则 $\mathrm{SD}(X)$ 称为波动率，刻画了风险的大小。定义变异系数 $\mathrm{CV}=\frac{\mathrm{SD}(X)}{\mu}$，其中 $\mu=\mathrm{E}(X)\neq 0$。

**命题 4.2.** 方差有如下性质：

1. $\mathrm{Var}(C)\equiv 0, C$ 为常数
2. $\mathrm{Var}(CX)=C^2\mathrm{Var}(X)$
3. $\mathrm{Var}(X+Y)=\mathrm{Var}(X)+\mathrm{Var}(Y)+2\mathrm{E}((X-\mathrm{E}(X))(Y-\mathrm{E}(Y)))$，且若 $X,Y$ 独立，则 $\mathrm{E}((X-\mathrm{E}(X))(Y-\mathrm{E}(Y)))=0$

## 4.4 协方差与相关系数

对随机变量 $X,Y$，设 $\mathrm{E}(X)=\mu_1,\mathrm{E}(Y)=\mu_2,\mathrm{Var}(X)=\sigma_1^2,\mathrm{Var}(Y)=\sigma_2^2$。

**定义 4.4.** 称 $X$ 与 $Y$ 的协方差 $\mathrm{Cov}(X,Y)=\mathrm{E}((X-\mu_1)(Y-\mu_2))$。

**命题 4.3.** 协方差有如下性质：

1. $\mathrm{Cov}(X,X)=\mathrm{Var}(X)$
2. $\mathrm{Cov}(X,Y)=\mathrm{Cov}(Y,X)$
3. $\mathrm{Cov}(X,Y)=\mathrm{E}(XY)-\mathrm{E}(X)\mathrm{E}(Y)$
4. $\mathrm{Cov}(aX_1+bX_2+c,Y)=a\mathrm{Cov}(X_1,Y)+b\mathrm{Cov}(X_2,Y),\forall$ 常数 $a,b,c\in\mathbb{R}$

**定义 4.5.** 称 $X$ 与 $Y$ 的（线性）相关系数 $\mathrm{Corr}(X,Y)=\frac{\mathrm{Cov}(X,Y)}{\sigma_1\sigma_2}=\mathrm{E}(\frac{X-\mu_1}{\sigma_1}\frac{Y-\mu_2}{\sigma_2})$。

若 $\mathrm{Corr}(X,Y)=0$，称 $X,Y$ 不相关。

**定理 4.1.** 相关系数有如下性质：

1. 若 $X,Y$ 相互独立，则 $X,Y$ 不相关（反之未必成立）

2. $|\mathrm{Corr}(X,Y)| \le 1$，且等号成立当且仅当 $\exists a,b, P(Y=aX+b)=1$，即 $Y=aX+b$, a.s.

其中 a.s. 表示 “almost surely”。

为证明上述定理的 (2)，首先利用 Cauchy-Schwartz 不等式证明引理：对随机变量 $U,V$，有 $\mathrm{E}^2(UV) \le \mathrm{E}(U^2)\mathrm{E}(V^2)$，且等号成立当且仅当 $\exists t_0 \in \mathbb{R}, P(V=t_0U)=1$。接下来令 $U=\frac{X-\mu_1}{\sigma_1}, V=\frac{Y-\mu_2}{\sigma_2}$，即得。

当 $\mathrm{Corr}(X,Y)=\pm 1$，可以证明 $a=\pm\sigma_2/\sigma_1$。

**例 4.2.** $X \sim N(0,1), Y=X^2$，则 $X$ 与 $Y$ 不相关，但不独立。

**例 4.3.** $(X,Y) \sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，则

$$
\begin{aligned}
&\mathrm{Corr}(X,Y)\\
=&\mathrm{E}\left(\frac{X-\mu_1}{\sigma_1}\frac{Y-\mu_2}{\sigma_2}\right)\\
=&\int_{\mathbb{R}^2}\frac{x-\mu_1}{\sigma_1}\frac{y-\mu_2}{\sigma_2}\frac{1}{2\pi\sigma_1\sigma_2}\frac{1}{\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x-\mu_1}{\sigma_1}\right)^2+\left(\frac{y-\mu_2}{\sigma_2}\right)^2-2\rho\frac{x-\mu_1}{\sigma_1}\frac{y-\mu_2}{\sigma_2}\right)}\mathrm{d}x\mathrm{d}y
\end{aligned}
$$

进行换元 $(u,v)^{\mathrm{T}}=A(\frac{x-\mu_1}{\sigma_1},\frac{y-\mu_2}{\sigma_2})^{\mathrm{T}}$，其中 $A$ 的定义参见例 3.3，则指数上的项化为 $-\frac{1}{2}(u^2+v^2)$，这一步实质上是进行了二次型的标准化。后续过程留作习题，最终计算结果为 $\mathrm{Corr}(X,Y)=\rho$。

## 4.5 矩

**定义 4.6.** 对 $k=1,2,\cdots$，称 $\mathrm{E}((X-c)^k)$ 为 $X$ 关于 $c$ 点的 $k$ 阶矩。特别地，$c=0$ 的情况下称为 $k$ 阶原点矩，$c=\mathrm{E}(X)$ 的情况下称为 $k$ 阶中心矩。

根据定义，$\mathrm{E}(X)$ 为 1 阶原点矩，而 1 阶中心矩恒等于 0；$\mathrm{Var}(X)=\mathrm{E}(X^2)-\mathrm{E}^2(X)$ 为 2 阶中心矩。

若 $\mathrm{E}(X)=\mu, \mathrm{SD}(X)=\sigma$，称 $\mathrm{E}((\frac{X-\mu}{\sigma})^k)=\frac{\mathrm{E}((X-\mu)^k)}{\sigma^k}$ 为 $k$ 阶标准矩。

1 阶标准矩恒等于 0, 2 阶标准矩恒等于 1, 3 阶标准矩称为 $X$ 的偏度系数, 记作 $\mathrm{Skew}(X)$。

**例 4.4.** $X \sim N(0,1)$，则 $\mathrm{Skew}(X)=\int_{-\infty}^{+\infty}x^3f(x)\mathrm{d}x=0$，其中 $f$ 为 $X$ 的 PDF。

称偏度系数 $<0$ 的分布为“负偏”或“左偏”，如图 4.1。

5 阶以上的奇数阶标准矩计算更复杂，受噪声影响更大。

4 阶标准矩称为 $X$ 的峰度系数，记作 $\mathrm{Kurt}(X)$。由于正态分布的峰度系数恒等于 3，因此常定义超额峰度系数为 $\mathrm{Kurt}(X)-3$。

常将 $\mu\pm\sigma$ 以内的范围称为“峰”，范围在“峰”以外但在 $\mu\pm2\sigma$ 以内的范围称为“肩”，范围在“肩”以外的部分称为“尾”。

通常，峰度系数 $>3$ 表现为相对于正态分布“尖峰厚尾”，如图 4.2。

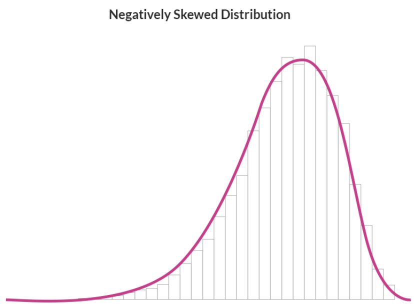



图 4.1: 负偏分布

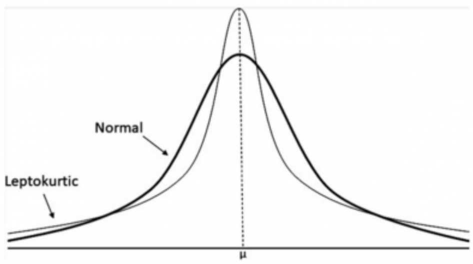



图 4.2: “Leptokurtic” 一词的含义即峰度系数 $> 3$

# 4.6 矩母函数

**定义 4.7.** 记 $M_X(t) = \mathrm{E}(e^{tX})$，若 $M_X(t)$ 在 $t = 0$ 的某邻域内存在，则称其为 $X$ 的矩母函数（Moment Generating Function, MGF），否则称 $X$ 的矩母函数不存在。

**例 4.5.** 若 $X \sim Exp(\lambda)$，则 $M_X(t) = \mathrm{E}(e^{tX}) = \int_0^{+\infty} e^{tx}\lambda e^{-\lambda x}\mathrm{d}x = \frac{\lambda}{\lambda-t}, t < \lambda$。

**例 4.6.** 若 $X \sim N(0,1)$，则 $M_X(t) = \mathrm{E}(e^{tX}) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{tx}e^{-\frac{x^2}{2}}\mathrm{d}x = e^{\frac{t^2}{2}}, t \in \mathbb{R}$。

**命题 4.4.** 矩母函数有如下性质：

1. $M_X(0) \equiv 1$
2. $Y = aX + b$，则 $M_Y(t) = \mathrm{E}(e^{tY}) = \mathrm{E}(e^{t(aX+b)}) = e^{tb}M_X(at)$

**例 4.7.** 若 $Y \sim N(\mu,\sigma^2)$，令 $Y = \sigma X + \mu$，则 $X \sim N(0,1)$，故 $M_Y(t) = e^{\mu t}M_X(\sigma t) = e^{\mu t}e^{\frac{(\sigma t)^2}{2}} = e^{\frac{\sigma^2 t^2}{2}+\mu t}, t \in \mathbb{R}$。

矩母函数可以用于确定矩。

**定理 4.2.** 随机变量 $X$ 的 $n$ 阶（原点）矩与其矩母函数有如下关系：$\mathrm{E}(X^n) = M_X^{(n)}(0)$。

*证明.* 由 Taylor 展开有 $M_X(t) = \sum_{n=0}^{+\infty} M_X^{(n)}(0)\frac{t^n}{n!}$，又 $M_X(t) = \mathrm{E}(e^{tX}) = \mathrm{E}(\sum_{n=0}^{+\infty} X^n\frac{t^n}{n!}) = \sum_{n=0}^{+\infty}\mathrm{E}(X^n)\frac{t^n}{n!}$，得到结论。 □

**例 4.8.** 若 $X\sim N(0,1)$，则 $M_X(t)=e^{\frac{t^2}{2}}=\sum_{n=0}^{+\infty}\frac{(\frac{t^2}{2})^n}{n!}=\sum_{n=0}^{+\infty}\frac{(2n)!}{2^n n!}\frac{t^{2n}}{(2n)!}$，因此 $\mathrm{E}(X^{2n})=\frac{(2n)!}{2^n n!},\mathrm{E}(X^{2n+1})\equiv 0\ (n=0,1,\cdots)$。
由此可以计算 $\mathrm{Var}(X)=\mathrm{E}(X^2)=1,\mathrm{Kurt}(X)=\mathrm{E}(X^4)=\frac{4!}{2^2\cdot 2!}=3$。

矩母函数还可以用于确定分布。

**定理 4.3.** 若存在 $a>0$，使得 $M_X(t)=M_Y(t),\forall t\in(-a,a)$，则 $X,Y$ 同分布。

**例 4.9.** 若随机变量 $X$ 的矩母函数 $M_X(t)=\frac{1}{2}e^{-t}+\frac{1}{4}+\frac{1}{8}e^{4t}+\frac{1}{8}e^{5t}$，则 $X$ 为离散型随机变量，分布表为

| $X$ | -1 | 0 | 4 | 5 |
|---|---|---|---|---|
| $P$ | 1/2 | 1/4 | 1/8 | 1/8 |

一般地，若离散型随机变量 $X$ 有 PMF $P(X=k)=p_k$ ($\sum_k p_k\equiv 1$)，则其 MGF 为 $M_X(t)=\mathrm{E}(e^{tX})=\sum_k e^{tk}p_k$。

注意，各阶矩均相同的随机变量未必同分布。

**例 4.10.** 设连续型随机变量 $X_1$ 和 $X_2$ 的 PDF 分别为 $f_1(x)=\frac{1}{\sqrt{2\pi}x}e^{-\frac{(\log x)^2}{2}},x>0$ 和 $f_2(x)=f_1(x)(1+\sin(2\pi\log x)),x>0$（$X_1$ 服从对数正态分布），则 $\mathrm{E}(X_2^n)=\mathrm{E}(X_1^n)+\int_0^{+\infty}x^n f_1(x)\sin(2\pi\log x)\mathrm{d}x$，其中后一项通过换元 $y=\log x-n$ 可以证明为 0，即 $X_1$ 和 $X_2$ 同矩但不同分布。

下面运用矩母函数，研究独立随机变量和的分布。

**定理 4.4.** 若随机变量 $X,Y$ 独立，$Z=X+Y$，则 $M_Z(t)=M_X(t)M_Y(t)$。

*证明.* $M_Z(t)=\mathrm{E}(e^{tZ})=\mathrm{E}(e^{t(X+Y)})=\mathrm{E}(e^{tX}e^{tY})=M_X(t)M_Y(t)$，其中最后一个等号利用了独立性。 □

推而广之，若 $\{X_i\}_{i=1}^n$ 相互独立，$Z=X_1+\cdots+X_n$，则 $M_Z(t)=\prod_{i=1}^n M_{X_i}(t)$。

**例 4.11.** 若 $\{X_i\}_{i=1}^n$ 相互独立且服从正态分布，则 $X_1+\cdots+X_n$ 也服从正态分布。
以 $n=2$ 为例说明。设 $X_i\sim N(\mu_i,\sigma_i^2)$ $(i=1,2)$，则 $M_{X_1+X_2}(t)=M_{X_1}(t)M_{X_2}(t)=e^{\frac{\sigma_1^2t^2}{2}+\mu_1 t}e^{\frac{\sigma_2^2t^2}{2}+\mu_2 t}=e^{\frac{1}{2}(\sigma_1^2+\sigma_2^2)+(\mu_1+\mu_2)t}$，对应 $N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$ 的 MGF，再由 MGF 确定分布可得结论。

定义随机向量 $(X_1,\cdots,X_n)$ 的 MGF 为 $M_{X_1,\cdots,X_n}(t_1,\cdots,t_n)=\mathrm{E}(e^{t_1X_1+\cdots+t_nX_n})$。
以下简介类似 MGF 的其他函数:

1. 概率母函数(Probability Generating Function, PGF), 仅针对非负整数取值的离散型随机变量 $X$, 设其 PMF 为 $P(X=k)=p_k$, 则其 PGF 定义为 $\mathrm{E}(t^X)=\sum_{k=0}^{+\infty}p_kt^k, t\in[-1,1]$, 或对于 $t\in(0,1]$, 等于 $\mathrm{E}(e^{X\log t})=M_X(\log t)$。
2. 特征函数, 定义为 $\mathrm{E}(e^{itX})$, 其中 $i^2=-1$。

## 4.7 条件期望

定义条件期望 $\mathrm{E}(Y|X\in A)=\begin{cases}\displaystyle\sum_i y_iP(Y=y_i|X\in A),\\ \displaystyle\int_{-\infty}^{+\infty}yf_{Y|X}(y|X\in A)\mathrm{d}y,\end{cases}$ 两种定义分别针对 $Y$ 为离散型和连续型随机变量。

进而, 定义 $\mathrm{E}(Y|x)=\mathrm{E}(Y|X=x)=\begin{cases}\displaystyle\sum_i y_iP(Y=y_i|X=x),\\ \displaystyle\int_{-\infty}^{+\infty}yf_{Y|X}(y|x)\mathrm{d}y,\end{cases}$ 注意到这是一个 $x$ 的函数, 记作 $h(x)$。将其作用在 $X$ 上, 得到 $h(X)=\mathrm{E}(Y|X)$, 这是一个 $X$ 的函数 (称为 $Y$ 对 $X$ 的回归函数), 因此是一个新的随机变量。

**例 4.12.** $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$, 则 $\mathrm{E}(Y|x)=\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)$。

**例 4.13.** 甲、乙两种同类产品, 平均使用寿命分别为 10 年和 15 年, 市场占有率分别为 60% 和 40%, 随机买一个, 则期望寿命是 $10\times 60\%+15\times 40\%=12$ 年, 我们发现这个计算过程可以表示为 $\mathrm{E}(Y)=\mathrm{E}(Y|X=1)P(X=1)+\mathrm{E}(Y|X=2)P(X=2)=h(1)P(X=1)+h(2)P(X=2)=\mathrm{E}(h(X))=\mathrm{E}(\mathrm{E}(Y|X))$, 其中 $X=1$ 表示抽到甲产品, $X=0$ 表示抽到乙产品, $Y$ 表示抽到的产品的寿命。

一般地, 有以下定理:

**定理 4.5.** (全期望公式)
对于随机向量 $(X,Y)$, 有 $\mathrm{E}(Y)=\mathrm{E}(\mathrm{E}(Y|X))$。

*证明.* 以连续型为例。设 $(X,Y)$ 的联合 PDF 为 $f(x,y)$, 有 $\mathrm{E}(Y|x)=\int_{-\infty}^{+\infty}yf_{Y|X}(y|x)\mathrm{d}y=\int_{-\infty}^{+\infty}y\frac{f(x,y)}{f_X(x)}\mathrm{d}y$, 故 $\mathrm{E}(Y)=\int_{-\infty}^{+\infty}yf_Y(y)\mathrm{d}y=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}yf(x,y)\mathrm{d}x\mathrm{d}y=\int_{-\infty}^{+\infty}\mathrm{E}(Y|x)f_X(x)\mathrm{d}x=\mathrm{E}(\mathrm{E}(Y|X))$。 □

一般地, 对于可测函数 $g$, 有 $\mathrm{E}(g(X,Y))=\mathrm{E}(\mathrm{E}(g(X,Y)|X))$。

**定理 4.6.** 对于随机向量 $(X,Y)$ 和任意可测函数 $g:\mathbb{R}\to\mathbb{R}$, 都有 $\mathrm{E}((Y-g(X))^2)\geq\mathrm{E}((Y-\mathrm{E}(Y|X))^2)$, 即条件期望是均方误差意义下的最优预测。

*证明.* 类比期望的性质 $E((Y-c)^2) \geq E((Y-E(Y))^2), \forall c \in \mathbb{R}$，有 $E((Y-g(X))^2|X) \geq E((Y-E(Y|X))^2|X), \forall g: \mathbb{R} \to \mathbb{R}$ 可测，两边对 $X$ 求期望即得。 □

我们经常用到最优线性预测，即 $\min\limits_{a,b} E((Y-(aX+b))^2)$，这种“均方意义上的最优”称之为最小二乘（least square）。

**命题 4.5.** 记 $\hat{Y} = E(Y|X)$ 为已知 $X$ 的条件下对 $Y$ 的最优估计，$\tilde{Y}$ 为估计误差 $\hat{Y}-Y$，则 $E(\tilde{Y}) = 0, E(\tilde{Y}\hat{Y}) = 0$，进而有 $\mathrm{Cov}(\hat{Y}, \tilde{Y}) = 0, \mathrm{Var}(Y) = \mathrm{Var}(\hat{Y}) + \mathrm{Var}(\tilde{Y})$。

# 第五章　不等式与极限定理

## 5.1　概率不等式

**定理 5.1.** （Markov 不等式）
若随机变量 $Y \geq 0$，则 $\forall a > 0$，有 $P(Y \geq a) \leq \frac{\mathrm{E}(Y)}{a}$。

*证明.* 取示性变量 $I = \begin{cases} 1, & Y \geq a, \\ 0, & Y < a, \end{cases}$ 则 $I \leq Y/a$，故 $P(Y \geq a) = \mathrm{E}(I) \leq \mathrm{E}(Y/a) = \mathrm{E}(Y)/a$。 □

**定理 5.2.** （Chebyshev 不等式）
若随机变量 $Y$ 的方差 $\mathrm{Var}(Y)$ 存在，则 $\forall a > 0$ 有 $P(|Y - \mathrm{E}(Y)| \geq a) \leq \frac{\mathrm{Var}(Y)}{a^2}$。

*证明.* $P(|Y - \mathrm{E}(Y)| \geq a) = P((Y - \mathrm{E}(Y))^2 \geq a^2) \leq \frac{\mathrm{E}((Y-\mathrm{E}(Y))^2)}{a^2} = \frac{\mathrm{Var}(Y)}{a^2}$. □

这告诉我们，如果 $\mathrm{Var}(Y) = 0$，则 $P(Y = \mathrm{E}(Y)) = 1$（即 a.s.）。

**定理 5.3.** （Chernoff 不等式）
对于任意随机变量 $Y$，$\forall a > 0, t > 0$，有 $P(Y \geq a) \leq \frac{\mathrm{E}(e^{tY})}{e^{ta}}$。

*证明.* $\forall t > 0, P(Y \geq a) = P(e^{tY} \geq e^{ta}) \leq \frac{\mathrm{E}(e^{tY})}{e^{ta}}$. □

**例 5.1.** 若 $X \sim N(0,1)$，则

1. 根据 Markov 不等式，$P(|X| \geq 3) \leq \frac{\mathrm{E}(|X|)}{3} = \frac{1}{3}\sqrt{\frac{2}{\pi}} \approx 0.27$；
2. 根据 Chebyshev 不等式，$P(|X| \geq 3) \leq \frac{\mathrm{Var}(X)}{3^2} = \frac{1}{9} \approx 0.11$；
3. 根据 Chernoff 不等式，$\forall t > 0, P(|X| \geq 3) = 2P(X \geq 3) \leq 2\frac{\mathrm{E}(e^{tX})}{e^{3t}} = 2e^{\frac{t^2}{2}-3t}$，取最小值点 $t = 3$，得 $P(|X| \geq 3) \leq 2e^{-\frac{9}{2}} \approx 0.022$；
4. 根据经验法则，$P(|X| \geq 3) \approx 0.003$。

## 5.2 大数定律

设随机变量 $X_1,\cdots,X_n$ 独立同分布，均值 $\mathrm{E}(X_i)=\mu$，方差 $\mathrm{Var}(X_i)=\sigma^2>0$，则样本均值 $\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$，其均值 $\mathrm{E}(\bar{X})=\mu$，方差 $\mathrm{Var}(\bar{X})=\frac{\sigma^2}{n}\to 0(n\to+\infty)$。

**定理 5.4.** （Khinchin 弱大数定律）
设随机变量 $X_1,\cdots,X_n$ 独立同分布，均值 $\mathrm{E}(X_i)=\mu$，方差 $\mathrm{Var}(X_i)=\sigma^2>0$，则 $\forall\epsilon>0$，有 $\lim_{n\to+\infty}P(|\bar{X}-\mu|\geq\epsilon)=0$，或等价地，$\lim_{n\to+\infty}P(|\bar{X}-\mu|<\epsilon)=1$。

*证明.* 由 Chebyshev 不等式，$P(|\bar{X}-\mu|\geq\epsilon)\leq\frac{\mathrm{Var}(\bar{X})}{\epsilon^2}=\frac{\sigma^2}{n}\frac{1}{\epsilon^2}\to 0(n\to+\infty)$。 □

$\forall\epsilon>0,\forall\alpha>0$，如果将 $\epsilon$ 和 $(1-\alpha)$ 分别称为精度和置信度，则根据 Khinchin 弱大数定律，$\exists N\in\mathbb{N}^+$，当 $n\geq N$ 时，$P(|\bar{X}-\mu|<\epsilon)\geq 1-\alpha$，即 $\bar{X}$ 至少以概率 $(1-\alpha)$ 落在区间 $(\mu-\epsilon,\mu+\epsilon)$ 内。

换句话说，当样本量足够大时，有很大的概率 $\bar{X}\approx\mu$，其中 $\mu$ 为未知的总体均值。

我们将 $X_i\sim B(p)$ 这一特例称之为 *Bernoulli 大数定律*。

通过更进一步的讨论可以证明，上述定理中关于方差的条件可以去掉，结论仍正确。

此外，还有对 Khinchin 弱大数定律的若干推广，如

1. 要求 $X_i$ 两两不相关，$\mathrm{Var}(X_i)$ 一致有界，我们就得到了 Chebyshev 大数定律；
2. 要求 $\mathrm{Var}(\bar{X})\to 0(n\to+\infty)$，我们就得到了 Markov 大数定律。

**定义 5.1.** 称 $Y_n$ 依概率收敛于 $Y$，记作 $Y_n\xrightarrow{P}Y$，如果 $\forall\epsilon>0$，有 $\lim_{n\to+\infty}P(|Y_n-Y|\geq\epsilon)=0$。

用上述定义，弱大数定律可以表述为 $\bar{X}\xrightarrow{P}\mu$。

**定理 5.5.** （Kolmogorov 强大数定律）
设随机变量 $X_1,\cdots,X_n$ 独立同分布，均值 $\mathrm{E}(X_i)=\mu$，则 $P(\lim_{n\to+\infty}\bar{X}=\mu)=1$。

考虑 $X_i\sim B(p)$ 的特殊情形，则 $\bar{X}$ 称之为频率，由强大数定律，$P(\lim_{n\to+\infty}\bar{X}=p)=1$，这说明概率的频率解释是合理的。

**定义 5.2.** 称 $Y_n$ 以概率 1 收敛于 $Y$，又称几乎必然收敛于 $Y$，记作 $Y_n\xrightarrow{\text{a.s.}}Y$，如果 $P(\lim_{n\to+\infty}Y_n=Y)=1$。

用上述定义，强大数定律可以表述为 $\bar{X}\xrightarrow{\text{a.s.}}\mu$。

**例 5.2.** （Monte Carlo 积分）
设要计算 $g(x)>0$ 在区间 $[a,b]$ 上的定积分，首先取一个适当的 $c>\sup\{g(x)|x\in[a,b]\}$，设

$(X_i, Y_i)$ 独立且服从区域 $[a,b]\times[0,c]$ 上的均匀分布，记 $I_i = \begin{cases} 1, & Y_i \le g(X_i), \\ 0, & Y_i > g(X_i), \end{cases}$ 则 $I_i \sim B(p)$，其中 $p = \frac{\int_a^b g(x)\mathrm{d}x}{c(b-a)}$，于是 $\bar{I} = \frac{1}{n}\sum_{i=1}^n I_i \approx p$，从而 $\int_a^b g(x)\mathrm{d}x \approx c(b-a)\bar{I}$。

**例 5.3.** 我们通过一个例子来考察一下上面介绍的两种收敛性的区别。

设概率空间 $(\Omega, \mathscr{F}, P)$，其中 $\Omega = [0,1]$，$\omega$ 在 $\Omega$ 上均匀分布。定义随机变量序列 $\forall \omega \in \Omega, Y_1(\omega) = \omega + I_{[0,1]}(\omega), Y_2(\omega) = \omega + I_{[0,1/2]}(\omega), Y_3(\omega) = \omega + I_{[1/2,1]}(\omega), Y_4(\omega) = \omega + I_{[0,1/3]}(\omega), Y_5(\omega) = \omega + I_{[1/3,2/3]}(\omega), Y_6(\omega) = \omega + I_{[2/3,1]}(\omega), \cdots$，则 $Y_n(\omega)$ 依概率收敛于 $Y(\omega) = \omega$，但不以概率 1 收敛于 $Y(\omega)$，因为 $\forall \omega_0 \in \Omega$，$Y_n(\omega_0)$ 无极限。

## 5.3 中心极限定理

**定理 5.6.** 设随机变量 $X_1, \cdots, X_n$ 独立同分布，均值 $\mathrm{E}(X_i) = \mu$，方差 $\mathrm{Var}(X_i) = \sigma^2 > 0$，则 $\forall x \in \mathbb{R}, \lim_{n\to+\infty} P\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le x\right) = \Phi(x)$，其中 $\Phi(x)$ 为标准正态分布的 CDF。或等价地，$\lim_{n\to+\infty} P\left(\frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}} \le x\right) = \Phi(x)$。

证明. 只对 $X_i$ 的 MGF 存在的情形给出证明。

不失一般性，假设 $\mu = 0, \sigma^2 = 1$，令 $M(t) = \mathrm{E}(e^{tX_i})$，则 $M(0) = 1, M'(0) = 0, M''(0) = 1$，于是 $\mathrm{E}(e^{t\frac{X_1+\cdots+X_n}{\sqrt{n}}}) = M^n\left(\frac{t}{\sqrt{n}}\right)$，而根据 Taylor 展开，$M\left(\frac{t}{\sqrt{n}}\right) = 1 + 0 + \frac{1}{2}\left(\frac{t}{\sqrt{n}}\right)^2 + o\left(\frac{t^2}{n}\right)$，故 $\mathrm{E}(e^{t\frac{X_1+\cdots+X_n}{\sqrt{n}}}) = (1 + \frac{t^2}{2n} + o(\frac{t^2}{n}))^n \to e^{t^2/2}(n\to+\infty)$，此为 $N(0,1)$ 的 MGF，这说明 $\frac{X_1+\cdots+X_n}{\sqrt{n}}$ 的分布趋近于 $N(0,1)$。 □

上述定理通常称为 Lindeberg-Lévy CLT，可推广至不同分布的情形。

如果将定理中的 $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ 理解为标准化的过程，则不难得出 $\bar{X}$ 近似服从 $N(\mu, \frac{\sigma^2}{n})$，$X_1 + \cdots + X_n$ 近似服从 $N(n\mu, n\sigma^2)$。

**例 5.4.** （De Moivre-Laplace CLT）

设 $X_i \sim B(p)$，则 $\sum_{i=1}^n X_i \sim B(n,p)$，当 $n$ 充分大时，可以近似地认为 $\sum_{i=1}^n X_i \sim N(np, np(1-p))$，于是可近似计算 $P(t_1 \le \sum_{i=1}^n X_i \le t_2) = P\left(\frac{t_1-np}{\sqrt{np(1-p)}} \le \frac{\sum_{i=1}^n X_i - np}{\sqrt{np(1-p)}} \le \frac{t_2-np}{\sqrt{np(1-p)}}\right) \approx \Phi(y_2) - \Phi(y_1)$，其中 $y_1 = \frac{t_1-np-\frac{1}{2}}{\sqrt{np(1-p)}}, y_2 = \frac{t_2-np+\frac{1}{2}}{\sqrt{np(1-p)}}$，其中 $\frac{1}{2}$ 是连续性修正项。

**定义 5.3.** （依分布收敛）

称 $Y_n$ 依分布收敛于 $Y$，记作 $Y_n \xrightarrow{d} Y$，如果 $\lim_{n\to+\infty} F_{Y_n}(x) = F_Y(x), \forall x \in \mathbb{R}$。

用上述定义，CLT 可以表述为 $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \xrightarrow{d} Z$，其中 $Z \sim N(0,1)$，或简记为 $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \to N(0,1)$。

**例 5.5.** （选举问题）

设 $p$ 为选民真实支持度（未知），随机抽样调查 $n$ 人（假设 $n$ 远远小于总人数 $N$，可以近似有放回抽样），样本支持比例 $P_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$，其中 $X_i \sim B(p)$ 且独立，表示第 $i$ 个人是否支持。

设置精度 $\epsilon = 0.03$，置信度 $1-\alpha = 95\%$，则至少需要 $n$ 为多少，才能保证 $P(|P_n-p| < \epsilon) \geq 1-\alpha$？根据 CLT 有 $P(|P_n - p| \geq \epsilon) \approx 2\left(1 - \Phi(\frac{\epsilon}{\sqrt{p(1-p)/n}})\right) \leq \alpha$，于是 $n \geq \frac{z_{\alpha/2}^2 p(1-p)}{\epsilon^2}$，其中 $z_{\alpha/2}$ 为标准正态分布的上 $\alpha/2$ 分位数，代入最大值点 $p = \frac{1}{2}$，得到 $n \geq \frac{z_{\alpha/2}^2}{4}\epsilon^2$，代入 $\epsilon = 0.03, \alpha = 0.05$，得到 $n \geq 1068$。这一结果与 $N$ 无关！

# 第二部分

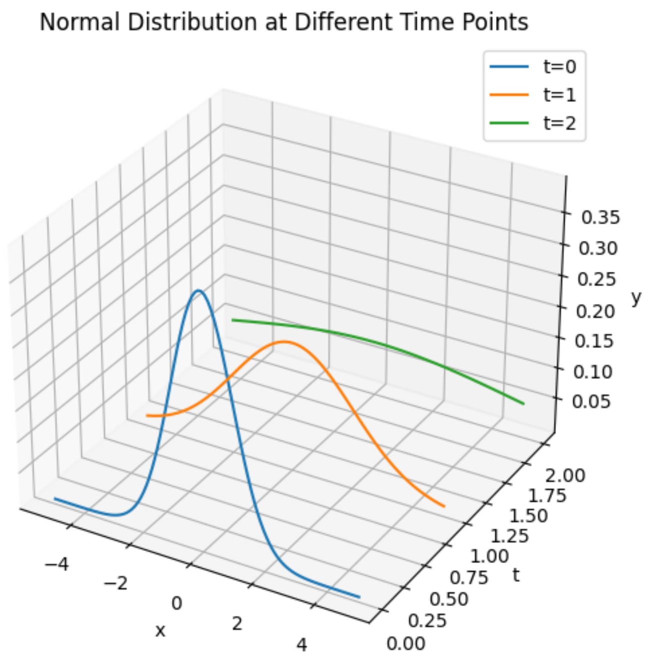
# 随机过程

# 随机过程引言

**定义.** 给定一个概率空间 $(\Omega, \mathscr{F}, P)$，若对于某个集合 $T$（称之为指标集），$\forall t \in T$，都有一个随机变量 $X_t$，则称 $\{X_t\}_{t\in T}$ 为一个随机过程（Stochastic Process），也记作 $\{X_t, t \in T\}$。称 $X_t$ 为该随机过程在 $t$ 时刻的状态，有时也记作 $X(t)$，其取值集合 $S$ 称为状态空间，即 $\forall t \in T, \forall \omega \in \Omega, X_t(\omega) \in S$。$\forall \omega_0 \in \Omega$，称 $\{X_t(\omega_0), t \in T\}$ 为该随机过程的一条样本轨道（Sample Path）。

随机变量是 $\Omega$ 上的函数，而随机过程是一列随机变量，因此 $X_t(\omega)$ 可以视为 $t, \omega$ 的函数。

**例.** 设 $X_0, U$ 独立且均服从 $N(0,1)$，令 $X_t = X_0 + tU$，则 $\{X_t, t \in \mathbb{R}\}$ 是一个随机过程，$\forall t \in \mathbb{R}, X_t \sim N(0, t^2+1)$，如下图所示。




$\{X_t, t \in \mathbb{R}\}$ 的分布

**例.** 设 $\{Y_i\}_{i=1}^{\infty}$ 独立同分布，且 $P(Y_i = 1) = P(Y_i = -1) = \frac{1}{2}$，令 $Y_0 = 0, X_n = \sum_{i=0}^{n} Y_i$，则 $\{X_n, n \in \mathbb{N}\}$ 是一个随机过程，称为随机游走（Random Walk）。

如果一个随机过程的 $T$ 是离散的，常称之为时间序列。而如果 $S$ 是离散的，有时用链来称呼这样的随机过程。

**定义.** 若 $\forall n, \forall t_1 < t_2 < \cdots < t_n \ (t_1, \cdots, t_n \in T)$，满足 $X_{t_2} - X_{t_1}, X_{t_3} - X_{t_2}, \cdots, X_{t_n} - X_{t_{n-1}}$ 相互独立，则称 $\{X_t, t \in T\}$ 为独立增量过程。若 $T$ 中包含最小指标 $t_0$，则要求 $X_{t_0}, X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \cdots, X_{t_n} - X_{t_{n-1}}$ 相互独立。

**定义.** 若 $\forall t \in T, \forall \tau \text{ s.t. } t + \tau \in T$，满足 $X_{t+\tau} - X_t$ 只依赖 $\tau$（不依赖 $t$），则称 $\{X_t, t \in T\}$ 为平稳增量过程。

一个随机过程 $\{X_t, t \in T\}$ 是平稳增量过程, 等价于 $\forall t_1, t_2 \in T, \forall \tau \text{ s.t. } t_i + \tau \in T \ (i = 1, 2)$，总有 $X_{t_1+\tau} - X_{t_1}$ 与 $X_{t_2+\tau} - X_{t_2}$ 同分布。

# 第六章　Poisson 过程

## 6.1　基本概念

**定义 6.1.** 设 $T=\mathbb{N}^*$，$\{X_n\}_{n=1}^{\infty}$ 是一列独立同分布的随机变量，且 $P(X_n=1)=p, P(X_n=0)=1-p$，则称 $\{X_n\}_{n=1}^{\infty}$ 为参数为 $p$ 的 *Bernoulli* 过程。

定义中提到的 $\{X_n\}_{n=1}^{\infty}$ 相互独立，指的是 $\forall n\in\mathbb{N}^*$，都有 $X_1,\cdots,X_n$ 相互独立。

如果在每个离散时刻 $n$，将事件 $X_n=1$ 即“第 $n$ 次试验成功”理解为该时刻有一个顾客到达某商店，则 Bernoulli 过程属于一种到达过程。

本章的主要内容 Poisson 过程也是一种到达过程。

**定义 6.2.** 称一个随机过程 $\{N(t),t\geq 0\}$ 为计数过程，若其满足：

1. $N(t)\in\mathbb{N}$
2. $\forall t>s\geq 0, N(t)\geq N(s)$
3. $N(t)-N(s)$ 为 $(s,t]$ 时间内发生的事件数

**定义 6.3.** 称一个计数过程 $\{N(t),t\geq 0\}$ 为 *Poisson* 过程，若其满足：

1. $N(0)=0$
2. $\{N(t),t\geq 0\}$ 有平稳增量性和独立增量性
3. $\exists\lambda>0$，当 $h\to 0$ 时，$P(N(h)=1)=\lambda h+o(h)$
4. 当 $h\to 0$ 时，$P(N(h)\geq 2)=o(h)$

利用平稳增量性可知，$\forall t\geq 0, P(N(t+h)-N(t)=1)=\lambda h+o(h), P(N(t+h)-N(t)\geq 2)=o(h)$。

考虑 $(0,t]$ 时间段，将其分成 $n$ 个长度为 $\frac{t}{n}$ 的子区间，当 $n\gg 1$ 时，每个小区间上发生 2 次及以上事件的概率趋于 0，而发生 1 次事件的概率 $p\approx\lambda\frac{t}{n}$，因此近似有 $N(t)\sim B(n,p)$，而当 $n\to\infty$ 时，$B(n,p)$ 会趋向于 Poisson 分布，即 $N(t)\sim P(\lambda t)$。稍后将严格证明这一结论。

称 $\lambda$ 为 Poisson 过程的强度或到达率。

**定理 6.1.** 若 $\{N(t), t \geq 0\}$ 是 Poisson 过程，则 $\forall t \geq 0, N(t) \sim P(\lambda t)$。

*证明.* 记 $P_n(t) = P(N(t) = n)$ $(n \in \mathbb{N})$，则 $P_0(t+h) = P(N(t+h) = 0) = P(N(t+h) - N(t) = 0, N(t) = 0) = P(N(t+h) - N(t) = 0)P(N(t) = 0) = (1 - \lambda h + o(h))P_0(t)$，即 $\frac{P_0(t+h)-P_0(t)}{h} = -\lambda P_0(t) + \frac{o(h)}{h}$。当 $h \to 0$ 时，$\frac{o(h)}{h} \to 0$，因此 $\frac{P_0(t+h)-P_0(t)}{h} \to -\lambda P_0(t)$，即 $P_0'(t) = -\lambda P_0(t)$，由边界条件 $P_0(0) = P(N(0) = 0) = 1$ 解得 $P_0(t) = e^{-\lambda t} = \frac{(\lambda t)^0}{0!} e^{-\lambda t}$。同理，

$$
\begin{aligned}
& P_n(t+h) \\
=& P(N(t+h) = n) \\
=& P(N(t+h) - N(t) = 0, N(t) = n) + P(N(t+h) - N(t) = 1, N(t) = n-1) \\
+& P(N(t+h) - N(t) \geq 2, N(t+h) = n) \\
=& P(N(t+h) - N(t) = 0)P(N(t) = n) + P(N(t+h) - N(t) = 1)P(N(t) = n-1) \\
+& P(N(t+h) - N(t) \geq 2, N(t+h) = n) \\
=& P(N(h) = 0)P_n(t) + P(N(h) = 1)P_{n-1}(t) + o(h) \\
& (\because P(N(t+h) - N(t) \geq 2, N(t+h) = n) \leq P(N(t+h) - N(t) \geq 2) \\
& = P(N(h) \geq 2) = o(h)) \\
=& (1 - \lambda h + o(h))P_n(t) + (\lambda h + o(h))P_{n-1}(t) + o(h) \\
=& P_n(t) + \lambda h(P_{n-1}(t) - P_n(t)) + o(h)
\end{aligned}
$$

因此 $\frac{P_n(t+h)-P_n(t)}{h} = \lambda(P_{n-1}(t) - P_n(t)) + \frac{o(h)}{h}$，当 $h \to 0$ 时，$\frac{P_n(t+h)-P_n(t)}{h} \to \lambda(P_{n-1}(t) - P_n(t))$，即 $P_n'(t) = \lambda(P_{n-1}(t) - P_n(t))$。由数学归纳法可证明 $P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}, \forall n \in \mathbb{N}$，即 $N(t) \sim P(\lambda t)$。 □

于是有 $\mathrm{E}(N(t)) = \lambda t, \mathrm{Var}(N(t)) = \lambda t$。

下面给出 Poisson 过程的一个等价定义。

**定义 6.4.** 称一个计数过程 $\{N(t), t \geq 0\}$ 为 *Poisson* 过程，若其满足：

1. $N(0) = 0$
2. $\{N(t), t \geq 0\}$ 有独立增量性
3. $\exists \lambda > 0, \forall t > s \geq 0, n \in \mathbb{N}, P(N(t+s) - N(s) = n) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$, 即 $N(t+s) - N(s) \sim P(\lambda t)$

## 6.2 到达时间与到达间隔

记 $T_1$ 为首次到达时刻，则 $P(T_1 \leq t) = 1 - P(T_1 > t) = 1 - P(N(t) = 0) = 1 - e^{-\lambda t}$，因此 $T_1 \sim Exp(\lambda)$。再记 $T_2$ 为第二次到达时刻，则在 $T_1 = t_0$ 条件下，$T_2 - T_1$ 的分布 $P(T_2 - T_1 \leq$

$t) = P(T_2 - t_0 \le t) = 1 - P(T_2 - t_0 > t) = 1 - P(N(t_0+t) - N(t_0) = 0) = 1 - P(N(t) = 0) = 1 - e^{-\lambda t}$，因此 $T_2 - T_1 \sim Exp(\lambda)$，且 $T_2 - T_1$ 与 $T_1$ 相互独立。一般地，记 $T_i$ 为第 $i$ 次到达时刻，$W_i = T_i - T_{i-1}$ 为相邻两次到达间隔时间，约定 $T_0 = 0$，则 $\{W_i\}_{i=1}^{\infty}$ 相互独立且 $W_i \sim Exp(\lambda)$，而 $T_k = \sum_{i=1}^{k} W_i$，称 $T_k$ 服从参数为 $k$ 和 $\lambda$ 的 *Gamma* 分布，记作 $T_k \sim \Gamma(k,\lambda)$，$\mathrm{E}(T_k) = \frac{k}{\lambda}, \mathrm{Var}(T_k) = \frac{k}{\lambda^2}$。$T_k$ 服从的分布又称之为 *Erlang* 分布。事实上，Erlang 分布是 Gamma 分布在 $k \in \mathbb{N}^*$ 时的特例。$\Gamma(1,\lambda)$ 就是 $Exp(\lambda)$。

利用 $\forall t \ge 0, P(T_k \le t) = P(N(t) \ge k) = 1 - \sum_{n=0}^{k-1} \frac{(\lambda t)^n}{n!} e^{-\lambda t}$，求导可得 $T_k$ 的 PDF 为 $f_{T_k}(t) = \frac{\lambda^k t^{k-1}}{(k-1)!} e^{-\lambda t} (t \ge 0)$。

Poisson 过程还有一个等价定义如下。

**定义 6.5.** 称一个计数过程 $\{N(t), t \ge 0\}$ 为 *Poisson* 过程，若其满足：

1. $N(0) = 0$
2. $\exists \lambda > 0$，各相邻两次到达间隔时间 $\{W_i\}_{i=1}^{\infty}$ 相互独立且 $W_i \sim Exp(\lambda)$

这揭示了生成 Poisson 过程的一种方法便是从独立同分布的 $Exp(\lambda)$ 中抽取相邻两次到达间隔时间，据此给出各到达时刻，即可确定一 Poisson 过程。

**例 6.1.** 拨打服务热线时，被告知除了正在接受服务的人以外，前面还有 55 人在等待。假设呼叫者离开服从 Poisson 过程，$\lambda = 2$ 人/min，则平均等待时间为 $T_{56} = \sum_{i=1}^{56} W_i$，其中 $W_i \sim Exp(2)$ 且相互独立，因此平均等待时间为 $\mathrm{E}(T_{56}) = \frac{56}{2} = 28$ min，且 $\mathrm{Var}(T_{56}) = \frac{56}{4} = 14\ \mathrm{min}^2$。等待时间超过 30 分钟的概率为 $P(T_{56} > 30) = \int_{30}^{+\infty} f_{T_{56}}(t)\mathrm{d}t$。根据 CLT，近似有 $T_{56} \sim N(28, 14)$，故有 $P(T_{56} > 30) \approx P(Z > \frac{30-28}{\sqrt{14}}) = 1 - \Phi(\frac{2}{\sqrt{14}})$，其中 $Z \sim N(0,1)$。

## 6.3 进一步性质

首先介绍 Poisson 过程的分裂。

**定理 6.2.** 假设某 Poisson 过程每次发生的事件分为 I 类和 II 类，每个事件独立地以概率 $p$ 成为 I 类事件，以概率 $1-p$ 成为 II 类事件，记 $N_1(t)$ 和 $N_2(t)$ 分别为 $(0,t]$ 内 I 类和 II 类事件的个数，则

1. $N(t) = N_1(t) + N_2(t)$
2. $\{N_1(t), t \ge 0\}$ 和 $\{N_2(t), t \ge 0\}$ 均为 Poisson 过程，且到达率分别为 $\lambda p$ 和 $\lambda(1-p)$
3. 这两个过程相互独立

证明. 仅对第二条给出简要证明。$\forall h > 0, P(N_1(h) = 1) = P(N_1(h) = 1 | N(h) = 1) P(N(h) = 1) + P(N_1(h) = 1 | N(h) \ge 2) P(N(h) \ge 2) = p(\lambda h + o(h)) + o(h) = \lambda p h + o(h)$，而 $P(N_1(h) \ge$

$2) \le P(N(h) \ge 2) = o(h)$，因此 $\{N_1(t), t \ge 0\}$ 是 Poisson 过程，且到达率为 $\lambda p$。同理可证 $\{N_2(t), t \ge 0\}$ 是 Poisson 过程，且到达率为 $\lambda(1-p)$。 □

**例 6.2.** 设 $\{X_i\}_{i=1}^{\infty}$ 独立同分布且服从 $Exp(\lambda)$，$N$ 服从参数为 $p$ 的几何分布，且与 $\{X_i\}_{i=1}^{\infty}$ 相互独立，令 $Y = \sum_{i=1}^{N} X_i$。为求出 $Y$ 的分布，设每次事件都独立地以概率 $p$ 成为“特殊事件”，则 $N$ 可视为首次发生“特殊事件”时的事件总数，$Y$ 为特殊事件首次发生的时刻。由定理 6.2，特殊事件的发生是一个 Poisson 过程，到达率为 $\lambda p$，因此 $Y \sim Exp(\lambda p)$。

下面介绍 Poisson 过程的合并。

**定理 6.3.** 若 $\{N_1(t), t \ge 0\}$ 和 $\{N_2(t), t \ge 0\}$ 是两个独立的 Poisson 过程，且到达率分别为 $\lambda_1$ 和 $\lambda_2$，则 $\{N_1(t) + N_2(t), t \ge 0\}$ 也是 Poisson 过程，且到达率为 $\lambda_1 + \lambda_2$。

证明.

1. 显然有 $N(0) = 0$。
2. 由 $N(t+s) - N(t) = (N_1(t+s) - N_1(t)) + (N_2(t+s) - N_2(t))$ 易验证独立增量性。
3. 由于 $N_i(t+s) - N_i(s) \sim P(\lambda_i t), i = 1, 2$ 且独立，因此 $N(t+s) - N(s) \sim P((\lambda_1 + \lambda_2)t)$。

□

**定理 6.4.** 记定理 6.3 中的两类事件的首达时刻分别为 $T^{(1)}$ 和 $T^{(2)}$，则 $P(T^{(1)} < T^{(2)}) = \frac{\lambda_1}{\lambda_1 + \lambda_2}$。

**例 6.3.** 设 $X \sim Exp(\lambda)$，而在 $X = x$ 的条件下，$Y - 1 \sim P(x)$。为求出 $Y$ 的分布，首先考虑到达率为 $\lambda$ 的“成功”过程，则 $X$ 可视为首次成功的时刻。再考虑到达率为 1 的“失败”过程，则 $(0, x]$ 内“失败”的次数服从 $P(x)$，因此 $Y - 1$ 可视为首次“成功”之前“失败”的次数，即 $Y$ 为首次“成功”时的事件总数。合并两个过程，得到一个参数为 $\lambda + 1$ 的 Poisson 过程，且每个事件属于“成功”的概率为 $p = \frac{\lambda}{\lambda+1}$，因此 $Y$ 服从参数为 $p$ 的几何分布，即 $P(Y = k) = p(1-p)^{k-1}, k \in \mathbb{N}^*$。

最后介绍在 Poisson 过程在条件作用下的性质。

记 $T_i$ 为第 $i$ 次到达时刻，计算可得对于 $0 \le s \le t$，条件概率 $P(T_1 \le s | N(t) = 1) = \frac{P(T_1 \le s, N(t)=1)}{P(N(t)=1)} = \frac{P(N(s)=1, N(t)-N(s)=0)}{P(N(t)=1)} = \frac{P(N(s)=1)P(N(t)-N(s)=0)}{P(N(t)=1)} = \frac{\frac{(\lambda s)^1}{1!} e^{-\lambda s} \frac{(\lambda(t-s))^0}{0!} e^{-\lambda(t-s)}}{\frac{(\lambda t)^1}{1!} e^{-\lambda t}} = \frac{s}{t}$，即在 $N(t) = 1$ 的条件下，$T_1 \sim U(0, t)$。

一般地，有如下定理。

**定理 6.5.** 设 $\{N(t), t \ge 0\}$ 是 Poisson 过程，则 $\forall 0 \le t_1 \le t_2$，在 $N(t_2) = n$ 的条件下，有 $N(t_1) \sim B(n, \frac{t_1}{t_2})$。

*证明.* 由 Poisson 过程性质知，$N(t_1)\sim P(\lambda t_1)$，而 $N(t_2)-N(t_1)\sim P(\lambda(t_2-t_1))$，且 $N(t_1)$ 和 $N(t_2)-N(t_1)$ 相互独立。据此，$\forall 0\le k\le n$，有

$$\begin{aligned}&P(N(t_1)=k|N(t_2)=n)\\=&\frac{P(N(t_1)=k,N(t_2)=n)}{P(N(t_2)=n)}\\=&\frac{P(N(t_1)=k)P(N(t_2)-N(t_1)=n-k)}{P(N(t_2)=n)}\\=&\frac{\frac{(\lambda t_1)^k}{k!}e^{-\lambda t_1}\frac{(\lambda(t_2-t_1))^{n-k}}{(n-k)!}e^{-\lambda(t_2-t_1)}}{\frac{(\lambda t_2)^n}{n!}e^{-\lambda t_2}}\\=&\tbinom{n}{k}\left(\frac{t_1}{t_2}\right)^k\left(1-\frac{t_1}{t_2}\right)^{n-k}.\end{aligned}$$

□

事实上，上述讨论适用于任何长度为 $t_1$ 的子区间。因此，以 $N(t_2)=n$ 为条件，相当于在 $(0,t_2]$ 上以均匀分布随机放置 $n$ 个到达点，第 $i$ 次到达时刻就是这 $n$ 个独立且服从 $U(0,t_2)$ 的随机变量的第 $i$ 个次序统计量。

一般地，考虑随机样本 $X_1,\cdots,X_n$，将它们从小到大排序，记为 $X_{(1)}\le\cdots\le X_{(n)}$，则称 $X_{(i)}$ 为 $X_1,\cdots,X_n$ 的第 $i$ 个次序统计量，$X_{(1)}=\min\{X_1,\cdots,X_n\},X_{(n)}=\max\{X_1,\cdots,X_n\}$。严谨地说，$X_{(1)}$ 定义为 $\forall\omega\in\Omega,X_{(1)}(\omega)=\min\{X_1(\omega),\cdots,X_n(\omega)\}$，其余同理。

**命题 6.1.** 若连续型随机变量 $X_1,\cdots,X_n$ 独立同分布，且 CDF 为 $F(x)$，PDF 为 $f(x)$，则

1. $X_{(k)}$ 的 PDF 为 $f_{X_{(k)}}(x)=\frac{n!}{(k-1)!(n-k)!}F^{k-1}(x)(1-F(x))^{n-k}f(x)$
2. $(X_{(1)},\cdots,X_{(n)})$ 的联合 PDF 为 $f_{X_{(1)},\cdots,X_{(n)}}(x_1,\cdots,x_n)=n!f(x_1)\cdots f(x_n)(x_1\le\cdots\le x_n)$

*证明.* 考虑小区间 $(x,x+\mathrm{d}x]$，当 $\mathrm{d}x$ 充分小时，有两个及以上随机变量落入其中的概率极小，因此事件 $X_{(k)}\in(x,x+\mathrm{d}x]$ 的概率近似为 $\binom{n}{k-1,1,n-k}F^{k-1}(x)f(x)\mathrm{d}x(1-F(x))^{n-k}$，即有 $(k-1)$ 个随机变量落入 $(-\infty,x]$, 1 个随机变量落入 $(x,x+\mathrm{d}x]$, $(n-k)$ 个随机变量落入 $(x+\mathrm{d}x,+\infty)$ 的概率。于是 $X_{(k)}$ 的 PDF 为 $f_{X_{(k)}}(x)=\frac{n!}{(k-1)!(n-k)!}F^{k-1}(x)(1-F(x))^{n-k}f(x)$。类似讨论可知，$(X_{(1)},\cdots,X_{(n)})$ 的联合 PDF 为 $\binom{n}{1,\cdots,1}f(x_1)\cdots f(x_n)=n!f(x_1)\cdots f(x_n)(x_1\le\cdots\le x_n)$。 □

**定理 6.6.** 设 $\{N(t),t\ge 0\}$ 是 Poisson 过程, 则在 $N(t)=n$ 的条件下, 事件发生时刻 $T_1,\cdots,T_n$ 的联合 PDF 为 $f_{T_1,\cdots,T_n|N(t)}(t_1,\cdots,t_n|N(t)=n)=\frac{n!}{t^n}(0\le t_1\le\cdots\le t_n\le t)$。

*证明.* 直接应用命题 6.1 的结论即得。下面给出另一种证明方法的概要。

由于 $\{T_1=t_1,T_2=t_2,\cdots,T_n=t_n,N(t)=n\}$ 等价于 $\{W_1=t_1,W_2=t_2-t_1,\cdots,W_n=$

$t_n - t_{n-1}, W_{n+1} > t - t_n\}$，其中 $W_i$ 为相邻两次到达间隔时间，且 $W_1, \cdots, W_n, W_{n+1}$ 相互独立，且 $W_i \sim Exp(\lambda)$。因此

$$\begin{aligned}&f_{T_1,\cdots,T_n|N(t)}(t_1,\cdots,t_n|N(t)=n)\\=&\frac{\lambda e^{-\lambda t_1}\lambda e^{-\lambda(t_2-t_1)}\cdots\lambda e^{-\lambda(t_n-t_{n-1})}\cdot e^{-\lambda(t-t_n)}}{\frac{(\lambda t)^n}{n!}e^{-\lambda t}}\\=&\frac{n!}{t^n}(0\le t_1\le\cdots\le t_n\le t).\end{aligned}$$

需要指出，这里由 $T_1, \cdots, T_n$ 的分布过渡到 $W_1, \cdots, W_n$ 的分布时，严格的证明过程应当使用密度函数变换的方法。 □

这揭示了生成 Poisson 过程的另一种方法。首先从 $P(\lambda t)$ 中采样 $n$ 为 $(0, t]$ 内事件发生数，然后从 $(0, t]$ 中均匀采样 $n$ 个点，即取 $U_1, \cdots, U_n$ 独立同分布且服从 $U(0, t)$，令 $T_k = U_{(k)}$ 为各事件发生时刻，即可确定一 Poisson 过程。

**例 6.4.** 乘客到达火车站可视为一 Poisson 过程，设到达率为 $\lambda$。火车在 $t$ 时刻出发，则 $(0, t]$ 内到达的乘客总等待时间的期望为 $\mathrm{E}(\sum_{i=1}^{N(t)}(t-T_i)) = \mathrm{E}(\mathrm{E}(\sum_{i=1}^{N(t)}(t-T_i)|N(t)))$。而 $\mathrm{E}(\sum_{i=1}^{N(t)}(t-T_i)|N(t)=n) = nt - \mathrm{E}(\sum_{i=1}^{n}T_i|N(t)=n) = nt - \mathrm{E}(\sum_{i=1}^{n}U_i) = nt - n\cdot\frac{t}{2} = \frac{n}{2}t$，因此 $\mathrm{E}(\sum_{i=1}^{N(t)}(t-T_i)) = \mathrm{E}(\frac{N(t)}{2}t) = \frac{\lambda t^2}{2}$。

## 6.4 Poisson 过程推广

首先介绍非齐次 Poisson 过程。

**定义 6.6.** 称一个计数过程 $\{N(t), t \ge 0\}$ 为到达率 $\lambda(t) > 0$ 的*非齐次 Poisson 过程*，若其满足：

1. $N(0) = 0$
2. $\{N(t), t \ge 0\}$ 有独立增量性
3. $\forall t \ge 0, P(N(t+h) - N(t) = 1) = \lambda(t)h + o(h)$
4. $\forall t \ge 0, P(N(t+h) - N(t) \ge 2) = o(h)$

令 $m(t) = \int_0^t \lambda(s)\mathrm{d}s$，可以证明 $P(N(t+s) - N(s) = n) = \frac{(m(t+s)-m(s))^n}{n!}e^{-(m(t+s)-m(s))}$。

然后介绍复合 Poisson 过程。

**定义 6.7.** 设 $\{N(t), t \ge 0\}$ 是 Poisson 过程，$\{X_i\}_{i=1}^{\infty}$ 独立同分布且与 $\{N(t), t \ge 0\}$ 独立，记 $X(t) = \sum_{i=1}^{N(t)} X_i$，则称 $\{X(t), t \ge 0\}$ 为一*复合 Poisson 过程*。

最后简介*更新过程*。Poisson 过程的相邻两次到达间隔时间服从参数为 $\lambda$ 的指数分布，而如果将其推广为一般分布，就得到了更新过程。

# 第七章　离散时间 Markov 链

## 7.1　基本概念

本章讨论的离散时间 Markov 链是一种特殊的随机过程，其指标集 $T$ 和状态空间 $S$ 都是离散的，不妨记为 $T=\{0,1,\cdots\},S=\{0,1,\cdots\}$。

**定义 7.1.** 若随机过程 $\{X_n,n=0,1,\cdots\}$ 的状态空间为 $S$，满足 $\forall n\in\mathbb{N},\forall i,j,i_0,\cdots,i_{n-1}\in S,P(X_{n+1}=j|X_0=i_0,\cdots,X_{n-1}=i_{n-1},X_n=i)=P(X_{n+1}=j|X_n=i)$，则称 $\{X_n,n=0,1,\cdots\}$ 为（离散时间）*Markov* 链，上式称为 *Markov* 性，又称无后效性。

可以直接利用 Markov 性给出 $(X_0,\cdots,X_n)$ 的联合分布，即

$$
\begin{aligned}
&P(X_0=i_0,\cdots,X_n=i_n)\\
=&P(X_n=i_n|X_0=i_0,\cdots,X_{n-1}=i_{n-1})P(X_0=i_0,\cdots,X_{n-1}=i_{n-1})\\
=&P(X_n=i_n|X_{n-1}=i_{n-1})P(X_0=i_0,\cdots,X_{n-1}=i_{n-1})\\
=&\cdots\\
=&P(X_n=i_n|X_{n-1}=i_{n-1})P(X_{n-1}=i_{n-1}|X_{n-2}=i_{n-2})\cdots P(X_1=i_1|X_0=i_0)P(X_0=i_0)
\end{aligned}
$$

**定义 7.2.** 若 $\{X_n,n=0,1,\cdots\}$ 为离散时间 Markov 链，称 $P(X_{n+1}=j|X_n=i)$ 为其（一步）转移概率，若其与 $n$ 无关，则称该 Markov 链关于时间是齐次的，此时记 $p_{ij}=P(X_{n+1}=j|X_n=i)$，称 $P=(p_{ij})$ 为转移概率矩阵。

显然有 $\forall i,j,p_{ij}\geq 0,\sum_{j\in S}p_{ij}=1$。

状态空间有限时称该 Markov 链为有限链，否则称为无限链。多数情况下只讨论关于时间齐次的有限 Markov 链。

利用转移概率，容易写出 $P(X_0=i_0,\cdots,X_n=i_n)=P(X_0=i_0)p_{i_0i_1}\cdots p_{i_{n-1}i_n}$。

**例 7.1.** 假设每天的天气只与前一天的天气有关：

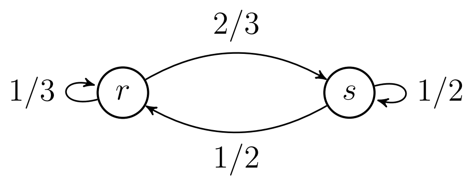
- 若前一天是雨天，则第二天是雨天的概率为 1/3，晴天的概率为 2/3

- 若前一天是晴天，则第二天是雨天的概率为 1/2，晴天的概率为 1/2

则各天的天气构成一个 Markov 链，其状态空间 $S=\{r,s\}$，其中 $r$ 和 $s$ 分别表示雨天和晴天。其转移概率矩阵为

$$\begin{array}{cc} & \begin{array}{cc} r & s \end{array} \\ \begin{array}{c} r \\ s \end{array} & \begin{pmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \end{array}$$

该 Markov 链的转移概率图如下。




**例 7.2.** 设 $\{X_i\}_{i=1}^{\infty}$ 独立同分布，且 $P(X_i=1)=1-P(X_i=-1)=p$，令 $X_0=0, Y_n=\sum_{i=0}^{n}X_i$，则 $\{Y_n, n\in\mathbb{N}\}$ 是一维随机游走，状态空间 $S=\mathbb{Z}$，且其为 Markov 链，转移概率为 $p_{i,i+1}=p, p_{i,i-1}=1-p$。

**例 7.3.** 状态空间为 $S=\{0,\cdots,N\}$ 的 Markov 链，对于 $i=1,\cdots,N-1$ 其转移概率同上例，而 $p_{00}=1, p_{01}=0, p_{NN}=1, p_{N,N-1}=0$，则称为具有吸收壁的随机游走，0 和 $N$ 称为吸收态。其转移概率矩阵为

$$\begin{array}{cc} & \begin{array}{cccccccc} 0 & 1 & 2 & 3 & \cdots & N-2 & N-1 & N \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ \vdots \\ N-1 \\ N \end{array} & \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 1-p & 0 & p & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1-p & 0 & p & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1-p & 0 & p \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{pmatrix} \end{array}$$

**例 7.4.** 现假设每天的天气只与前两天的天气有关：

- 若前两天是雨天，则第三天是雨天的概率为 0.7，晴天的概率为 0.3
- 若前两天分别是晴天和雨天，则第三天是雨天的概率为 0.5，晴天的概率为 0.5
- 若前两天分别是雨天和晴天，则第三天是雨天的概率为 0.4，晴天的概率为 0.6
- 若前两天是晴天，则第三天是雨天的概率为 0.2，晴天的概率为 0.8

则仍可构造 Markov 链，其状态空间 $S=\{rr,sr,rs,ss\}$，各状态表示近两天的天气，则转移概率矩阵为

$$\begin{array}{c} \\ rr \\ sr \\ rs \\ ss \end{array}\begin{array}{c} \begin{array}{cccc} rr & sr & rs & ss \end{array} \\ \begin{pmatrix} 0.7 & 0 & 0.3 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.4 & 0 & 0.6 \\ 0 & 0.2 & 0 & 0.8 \end{pmatrix} \end{array}$$

## 7.2 Chapman-Kolmogorov 方程

**定义 7.3.** 若 $\{X_n, n=0,1,\cdots\}$ 为关于时间齐次的 Markov 链，对 $n\in\mathbb{N}^*$，称 $P(X_n=j|X_0=i)$ 为其 $n$ 步转移概率，记为 $p_{ij}^{(n)}$。

利用时间齐次性，可得 $P(X_{m+n}=j|X_m=i)=P(X_n=j|X_0=i), \forall m,n\in\mathbb{N}$。约定 $p_{ij}^{(0)}=\delta_{ij}=\begin{cases}1, & i=j,\\ 0, & i\neq j.\end{cases}$

**定理 7.1.** $\forall m,n\in\mathbb{N}, \forall i,j\in S$，有 $p_{ij}^{(m+n)}=\sum_{k\in S}p_{ik}^{(m)}p_{kj}^{(n)}$。

*证明.*

$$\begin{aligned} & p_{ij}^{(m+n)} \\ =& P(X_{m+n}=j|X_0=i) \\ =& \sum_{k\in S}P(X_{m+n}=j|X_m=k)P(X_m=k|X_0=i) \\ =& \sum_{k\in S}p_{ik}^{(m)}p_{kj}^{(n)}. \end{aligned}$$

□

定理 7.1 称之为 *Chapman-Kolmogorov* 方程。

记 $P^{(n)}=(p_{ij}^{(n)})$ 为 $n$ 步转移概率矩阵，则 C-K 方程可写为 $P^{(m+n)}=P^{(m)}P^{(n)}$，由此可得 $P^{(n)}=P^n$。

根据 C-K 方程可给出 $X_n$ 的边际分布，即 $P(X_n=j)=\sum_{i\in S}P(X_n=j|X_0=i)P(X_0=i)=\sum_{i\in S}p_{ij}^{(n)}P(X_0=i), \forall j\in S$。若记 $\boldsymbol{\beta}_n=(\beta_{n0},\beta_{n1},\cdots)$，其中 $\beta_{ni}=P(X_n=i)$，则有 $\boldsymbol{\beta}_n=\boldsymbol{\beta}_0P^{(n)}=\boldsymbol{\beta}_0P^n$。

## 7.3　状态分类

**定义 7.4.** $\forall i,j\in S$，称状态 $i$ 可达状态 $j$，若 $\exists n\geq 0, p_{ij}^{(n)}>0$，记作 $i\to j$。若 $i\to j$ 且 $j\to i$，则称 $i$ 和 $j$ 是互通的，记作 $i\leftrightarrow j$。

**命题 7.1.** $\leftrightarrow$ 是 $S$ 上的等价关系，即其具有

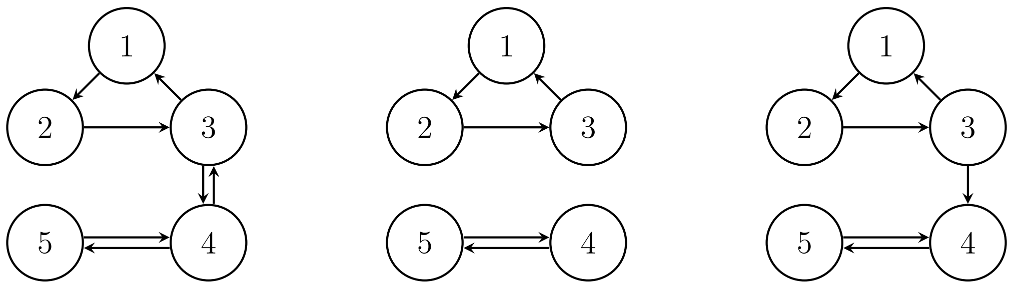
1. 自反性：$i\leftrightarrow i,\forall i\in S$
2. 对称性：$i\leftrightarrow j\Rightarrow j\leftrightarrow i,\forall i,j\in S$
3. 传递性：$i\leftrightarrow j,j\leftrightarrow k\Rightarrow i\leftrightarrow k,\forall i,j,k\in S$

*证明.* 仅证明传递性。若 $i\to j,j\to k$，则 $\exists m,n\geq 0,p_{ij}^{(m)}>0,p_{jk}^{(n)}>0$，由 C-K 方程可得 $p_{ik}^{(m+n)}=\sum_{l\in S}p_{il}^{(m)}p_{lk}^{(n)}\geq p_{ij}^{(m)}p_{jk}^{(n)}>0$，故 $i\to k$。同理可证 $k\to i$，故 $i\leftrightarrow k$。 □

若 $i\leftrightarrow j$，则称 $i,j$ 归属一类。

**定义 7.5.** 若某 Markov 链的所有状态只有一类，则称该链不可约，否则称为可约。

**例 7.5.** 下图中的三个 Markov 链分别是不可约、可约、可约的。




**定义 7.6.** 记 $f_{ij}^{(n)}$ 为从 $i$ 出发，经 $n$ 步首达 $j$ 的概率，即 $f_{ij}^{(n)}=P(X_n=j,X_k\neq j,k=1,\cdots,n-1|X_0=i)$。记 $f_{ij}=\sum_{n=1}^{\infty}f_{ij}^{(n)}$ 为从 $i$ 出发经有限步到达 $j$ 的概率。约定 $f_{ij}^{(0)}=\delta_{ij}$。

**定义 7.7.** 若 $f_{ii}=1$，则称状态 $i$ 为常返（recurrent）的。若 $f_{ii}<1$，则称状态 $i$ 为非常返的或瞬态（transient）的。

**定理 7.2.** $i$ 常返等价于 $\sum_{n=0}^{\infty}p_{ii}^{(n)}=+\infty$。$i$ 非常返等价于 $\sum_{n=0}^{\infty}p_{ii}^{(n)}=\frac{1}{1-f_{ii}}$。

先证明一个引理。

**引理 7.1.** $\forall i,j\in S,n\in\mathbb{N}^*$，有 $p_{ij}^{(n)}=\sum_{k=1}^{n}f_{ij}^{(k)}p_{jj}^{(n-k)}$。

证明. 用数学归纳法。对 $n=1$，由 $p_{ij}^{(1)}=f_{ij}^{(1)}, p_{jj}^{(0)}=1$ 显然成立。假设对 $n-1$ 成立，即 $p_{ij}^{(n-1)}=\sum_{k=1}^{n-1} f_{ij}^{(k)} p_{jj}^{(n-1-k)}$，则

$$
\begin{aligned}
& p_{ij}^{(n)} \\
= & \sum_{l\in S} p_{il}^{(1)} p_{lj}^{(n-1)} \\
= & p_{ij}^{(1)} p_{jj}^{(n-1)} + \sum_{l\neq j} p_{il}^{(1)} p_{lj}^{(n-1)} \\
= & p_{ij}^{(1)} p_{jj}^{(n-1)} + \sum_{l\neq j} p_{il}^{(1)} (\sum_{k=1}^{n-1} f_{lj}^{(k)} p_{jj}^{(n-1-k)}) \\
= & p_{ij}^{(1)} p_{jj}^{(n-1)} + \sum_{k=1}^{n-1} (\sum_{l\neq j} p_{il}^{(1)} f_{lj}^{(k)}) p_{jj}^{(n-1-k)} \\
= & p_{ij}^{(1)} p_{jj}^{(n-1)} + \sum_{k=1}^{n-1} f_{ij}^{(k+1)} p_{jj}^{(n-1-k)} \\
= & p_{ij}^{(1)} p_{jj}^{(n-1)} + \sum_{k=2}^{n} f_{ij}^{(k)} p_{jj}^{(n-k)} \\
= & \sum_{k=1}^{n} f_{ij}^{(k)} p_{jj}^{(n-k)}.
\end{aligned}
$$

□

下面证明定理 7.2。

证明. $\sum_{n=0}^{\infty} p_{ii}^{(n)} = 1 + \sum_{n=1}^{\infty} p_{ii}^{(n)} = 1 + \sum_{n=1}^{\infty}\sum_{k=1}^{n} f_{ii}^{(k)} p_{ii}^{(n-k)} = 1 + \sum_{k=1}^{\infty} f_{ii}^{(k)} \sum_{n=k}^{\infty} p_{ii}^{(n-k)} = 1 + \sum_{k=1}^{\infty} f_{ii}^{(k)} \sum_{n=0}^{\infty} p_{ii}^{(n)} = 1 + f_{ii}\sum_{n=0}^{\infty} p_{ii}^{(n)}$，故若 $i$ 非常返，则 $\sum_{n=0}^{\infty} p_{ii}^{(n)} = \frac{1}{1-f_{ii}} < +\infty$。而若 $i$ 常返，则 $\sum_{n=0}^{\infty} p_{ii}^{(n)} = +\infty$。 □

若令 $I_n = \begin{cases} 1, & X_n = i, \\ 0, & X_n \neq i, \end{cases}$ 则 $\sum_{n=0}^{\infty} I_n$ 为经过状态 $i$ 的次数，$\mathrm{E}(\sum_{n=0}^{\infty} I_n | X_0 = i) = \sum_{n=0}^{\infty} \mathrm{E}(I_n | X_0 = i) = \sum_{n=0}^{\infty} p_{ii}^{(n)}$ 为链从 $i$ 出发经过 $i$ 的期望次数。若 $i$ 常返，从 $i$ 出发以概率 1 经过 $i$ 无穷多次。若 $i$ 非常返，从 $i$ 出发有概率 $(1-f_{ii})$ 回不到 $i$（逃离 $i$），经过 $i$ 的次数为 $k$ 的概率为 $f_{ii}^{k-1}(1-f_{ii})$，即经过 $i$ 的次数服从几何分布，链以概率 1 最终逃离 $i$。

**定理 7.3.** 关于常返性，有如下性质：

1. 若 $i \leftrightarrow j$，则 $i, j$ 同为常返态或同为非常返态
2. 有限链至少存在一个常返态
3. 若链有限且不可约，则所有状态均为常返态

证明. 仅证明性质 1。由 $i \leftrightarrow j$，知 $\exists m,n,p_{ij}^{(m)} > 0, p_{ji}^{(n)} > 0$，且由 C-K 方程知 $p_{ii}^{(m+n+k)} \geq p_{ij}^{(m)}p_{jj}^{(k)}p_{ji}^{(n)}$，故 $\sum_{k=0}^{\infty} p_{ii}^{(k)} \geq \sum_{k=0}^{\infty} p_{ii}^{(m+n+k)} \geq p_{ij}^{(m)}p_{ji}^{(n)} \sum_{k=0}^{\infty} p_{jj}^{(k)}$。交换 $i$ 与 $j$ 可得类似不等式，故 $\sum_{k=0}^{\infty} p_{ii}^{(k)}$ 与 $\sum_{k=0}^{\infty} p_{jj}^{(k)}$ 相互控制，二者同为发散或同为收敛，故 $i,j$ 同为常返态或同为非常返态。 □

**例 7.6.** 一维随机游走，状态空间 $S=\mathbb{Z}$，转移概率为 $p_{i,i+1} = p \in (0,1), p_{i,i-1} = q = 1-p$，则显然其不可约。可以证明，若 $p=\frac{1}{2}$，则该链常返，否则非常返。

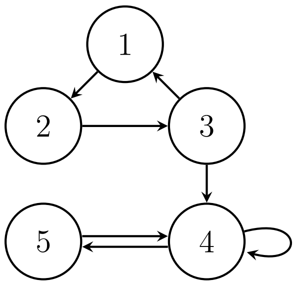
**命题 7.2.** 若 $i \to j$ 且 $i$ 常返，则 $f_{ji} = 1$。

证明. 若 $f_{ji} < 1$，则从 $j$ 出发以概率 $1 - f_{ji} > 0$ 不能到 $i$。又 $i \to j$，则从 $i$ 出发有正概率回不到 $i$，与 $i$ 常返矛盾。 □

上述命题说明，从常返态出发不能到达非常返态。对于有限链，若从非常返态出发，最终一定会到达某个常返态。

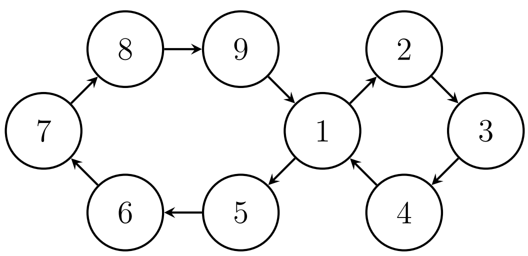
**定义 7.8.** 若集合 $\{n|p_{ii}^{(n)} > 0\}$ 非空，则其所有元素的最大公约数 $d(i)$ 称为 $i$ 的周期。若 $d(i) > 1$，则称 $i$ 为周期的，否则称为非周期的。若链的所有状态都是非周期的，则称链是非周期的，否则称链是周期的。

**例 7.7.** 下图中的 Markov 链满足 $d(1) = d(2) = d(3) = 3, d(4) = d(5) = 1$，因此是周期的。




需注意，一般情况下 $p_{ii}^{(nd(i))}(n \in \mathbb{N}^*)$ 不一定大于 0，如下例。

**例 7.8.** 下图中的 Markov 链满足 $p_{11}^{(n)} > 0, n = 4,6,8,10,12,\cdots$，故 $d(1) = 2$，但 $p_{11}^{(2)} = 0$。



但可以证明，$\forall i \in S, \exists N \in \mathbb{N}^*, \forall n \geq N, p_{ii}^{(nd(i))} > 0$。

**定理 7.4.** 若 $i \leftrightarrow j$，则 $d(i) = d(j)$。

证明. 由 $i \leftrightarrow j$，知 $\exists m, n, p_{ij}^{(m)} > 0, p_{ji}^{(n)} > 0$，且由 C-K 方程知 $p_{ii}^{(m+n)} \geq p_{ij}^{(m)} p_{ji}^{(n)} > 0$。对于所有满足 $p_{jj}^{(s)} > 0$ 的 $s \in \mathbb{N}^*$，都有 $p_{ii}^{(m+n+s)} \geq p_{ij}^{(m)} p_{jj}^{(s)} p_{ji}^{(n)} > 0$，则 $d(i) | m+n+s$ 且 $d(i) | m+n$，故 $d(i) | s$，于是 $d(i) | d(j)$。同理可证 $d(j) | d(i)$，故 $d(i) = d(j)$。 □

一般地，可根据状态的分类对状态重新编号，实现状态空间的分解，形如 $S = C_1 \cup \cdots \cup C_r \cup C_0$，其中 $C_i \cap C_j = \varnothing, \forall i \neq j$，且 $C_0$ 为非常返状态集，而 $C_1, \cdots, C_r$ 为常返互通类，此时转移概率矩阵可写为 $\begin{bmatrix} P_1 & \cdots & O & \\ \vdots & \ddots & \vdots & O \\ O & \cdots & P_r & \\ & W & & Q \end{bmatrix}$。其中，$P_1, \cdots, P_r, Q$ 分别为 $C_1, \cdots, C_r, C_0$ 内部的转移概率矩阵。

## 7.4 稳态性质

**定义 7.9.** 若 $\boldsymbol{\beta} = (\beta_0, \cdots, \beta_j, \cdots)$ 为状态空间上的概率分布，即 $\beta_j \geq 0, \forall j \in S, \sum_{j \in S} \beta_j \equiv 1$，且 $\boldsymbol{\beta} P = \boldsymbol{\beta}$，则称 $\boldsymbol{\beta}$ 为该 Markov 链的平稳分布。

若 $\boldsymbol{\beta}$ 是平稳分布且 $X_0$ 的分布为 $\boldsymbol{\beta}$，即 $P(X_0 = j) = \beta_j, \forall j \in S$，则 $\forall n \in \mathbb{N}^*$，都有 $X_n$ 的边际分布是 $\boldsymbol{\beta}$。

**定义 7.10.** 设 $i$ 为常返态，定义平均首返时间 $u_i = \sum_{n=1}^{\infty} n f_{ii}^{(n)}$，若 $u_i < \infty$，则称 $i$ 为正常返的，否则称 $i$ 为零常返的。

**定理 7.5.** 若 $i$ 常返，则 $\lim_{n \to \infty} p_{ii}^{(nd(i))} = \frac{d(i)}{u_i}$。特别地，若 $u_i = +\infty$，则 $\frac{d(i)}{u_i} = 0$。

**定理 7.6.** 若 $i$ 为零常返或非常返，则 $\lim_{n \to \infty} p_{ii}^{(n)} = 0$。

证明.

- 若 $i$ 为零常返，则 $u_i = +\infty$，由定理 7.5 知 $\lim_{n \to \infty} p_{ii}^{(nd(i))} = 0$，而对于不被 $d(i)$ 整除的 $m \in \mathbb{N}^*$，恒有 $p_{ii}^{(m)} = 0$。故 $\lim_{n \to \infty} p_{ii}^{(n)} = 0$。
- 若 $i$ 为非常返，由 $\sum_{n=0}^{\infty} p_{ii}^{(n)} < +\infty$ 知 $\lim_{n \to \infty} p_{ii}^{(n)} = 0$。

□

**定理 7.7.** 若 $i \leftrightarrow j$ 且同常返，则 $i, j$ 同为正常返或同为零常返。

**定理 7.8.** 若 $j$ 为零常返或非常返，则 $\forall i \in S, \lim_{n\to\infty} p_{ij}^{(n)} = 0$。

*证明.*

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

由于 $\sum_{k=1}^{\infty} f_{ij}^{(k)} = f_{ij} \leq 1$，故 $\forall \epsilon > 0, \exists M \in \mathbb{N}^*, \sum_{k=M+1}^{\infty} f_{ij}^{(k)} < \epsilon$，如此选取上式中的 $M$，有 $\lim_{n\to\infty} p_{ij}^{(n)} \leq \lim_{n\to\infty} \sum_{k=1}^{M} f_{ij}^{(k)} p_{jj}^{(n-k)} + \epsilon$。由定理 7.6 知 $\lim_{n\to\infty} p_{jj}^{(n-k)} = 0, k = 1, \cdots, M$，故 $\lim_{n\to\infty} p_{ij}^{(n)} \leq \epsilon$，由 $\epsilon$ 的任意性知 $\lim_{n\to\infty} p_{ij}^{(n)} = 0$。 □

定理 7.8 的直观理解为，若 $j$ 不是正常返的，则从 $i$ 出发经足够长时间后处于 $j$ 的概率很小。

可以证明，有限链不可能有零常返态，故不可约有限链的所有状态均为正常返的。并且若链有零常返态，则必有无穷多个零常返态。

**定义 7.11.** 若 $i$ 为正常返的和非周期的，则称 $i$ 为*遍历*（ergodic）的。若链不可约，且所有状态都是遍历的，则称链为*遍历的*。

**定理 7.9.** 对不可约、非周期的 Markov 链，有以下结论：

1. 若所有状态均为正常返，则 $\pi_j = \lim_{n\to\infty} p_{ij}^{(n)} (j \in S)$ 是唯一的平稳分布
2. 若所有状态均为零常返或非常返的，则平稳分布不存在

*证明.* 仅证明结论 1。不妨设 $S = \{0, 1, \cdots\}$，则 $\forall M \in \mathbb{N}^*, \sum_{j=0}^{M} p_{ij}^{(n)} \leq \sum_{j=0}^{\infty} p_{ij}^{(n)} = 1$，令 $n \to \infty$ 有 $\sum_{j=0}^{M} \pi_j \leq 1$，再令 $M \to \infty$ 有 $\sum_{j=0}^{\infty} \pi_j \leq 1$。又由 $\forall M \in \mathbb{N}^*, p_{ij}^{(n+1)} = \sum_{k=0}^{\infty} p_{ik}^{(n)} p_{kj} \geq \sum_{k=0}^{M} p_{ik}^{(n)} p_{kj}$，令 $n \to \infty$ 有 $\pi_j \geq \sum_{k=0}^{M} \pi_k p_{kj}$，再令 $M \to \infty$ 有 $\pi_j \geq \sum_{k=0}^{\infty} \pi_k p_{kj}$。若此不等式对某个 $j$ 取严格大于号，则 $\sum_{j=0}^{\infty} \pi_j > \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \pi_k p_{kj} = \sum_{k=0}^{\infty} \pi_k \sum_{j=0}^{\infty} p_{kj} = \sum_{k=0}^{\infty} \pi_k$，矛盾。故 $\pi_j = \sum_{k=0}^{\infty} \pi_k p_{kj}, \forall j \in S$，即 $\boldsymbol{\pi} = \boldsymbol{\pi} P$。进而，有 $\boldsymbol{\pi} = \boldsymbol{\pi} P^n$，即 $\pi_j = \sum_{k=0}^{\infty} \pi_k p_{kj}^{(n)}, \forall j \in S$，令 $n \to \infty$ 有 $\pi_j = \pi_j \sum_{k=0}^{\infty} \pi_k, \forall j \in S$，故 $\sum_{k=0}^{\infty} \pi_k = 1$，至此证明了 $\boldsymbol{\pi}$ 是平稳分布。
下证唯一性。若 $\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots)$ 为任一平稳分布，有 $\boldsymbol{\beta} = \boldsymbol{\beta} P = \cdots = \boldsymbol{\beta} P^n$，即 $\beta_j = \sum_{k=0}^{\infty} \beta_k p_{kj}^{(n)}, \forall j \in S$，令 $n \to \infty$ 有 $\beta_j = \sum_{k=0}^{\infty} \beta_k \pi_j = \pi_j, \forall j \in S$，故 $\boldsymbol{\beta} = \boldsymbol{\pi}$。 □

**命题 7.3.** 若 $i \leftrightarrow j$ 且 $j$ 为遍历的，则 $\lim_{n\to\infty} p_{ij}^{(n)} = \frac{1}{u_j}$。

**定义 7.12.** 若 $\pi_j = \lim_{n\to\infty} p_{ij}^{(n)}$ 存在，则 $\boldsymbol{\pi} = (\pi_0, \cdots)$ 称为该链的极限分布。

于是定理 7.9 的结论 1 可以表述为：不可约遍历链的极限分布存在，且为其唯一平稳分布。

若极限分布 $\boldsymbol{\pi}$ 存在，则对于任意函数 $g$，有 $\frac{1}{N}\sum_{n=0}^{N-1} g(X_n) \overset{\text{a.s.}}{\to} \mathrm{E}_{\boldsymbol{\pi}}(g(X)) = \sum_{j=0}^{\infty} g(j)\pi_j$。

由于 $\lim_{n\to\infty} P^{(n)} = \lim_{n\to\infty} P^n = \begin{bmatrix} \pi_0 & \pi_1 & \cdots & \pi_j & \cdots \\ \pi_0 & \pi_1 & \cdots & \pi_j & \cdots \\ \vdots & \vdots & \ddots & \vdots & \ddots \end{bmatrix} = \begin{bmatrix} \boldsymbol{\pi} \\ \boldsymbol{\pi} \\ \vdots \end{bmatrix}$，故对任意初始分布 $\boldsymbol{\beta}$，有 $\lim_{n\to\infty} \boldsymbol{\beta} P^{(n)} = \boldsymbol{\beta} \begin{bmatrix} \boldsymbol{\pi} \\ \boldsymbol{\pi} \\ \vdots \end{bmatrix} = \boldsymbol{\pi}$。

需要强调，一个链有平稳分布并不意味着其极限分布存在，例如 $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$，其平稳分布为 $\boldsymbol{\pi} = (\frac{1}{2}, \frac{1}{2})$，但 $\lim_{n\to\infty} P^{(n)}$ 不存在。

可以证明：

1. 有限链总存在平稳分布
2. 不可约链的平稳分布若存在则唯一

**例 7.9.** **TODO: Google PageRank**

## 7.5 可逆性

**TODO: 当状态空间 $S$ 很大时，平稳分布难计算。**

**定义 7.13.** 若 $\pi_j \geq 0, \forall j \in S, \sum_{j\in S} \pi_j = 1$，且 $\forall i, j \in S, \pi_i p_{ij} = \pi_j p_{ji}$，则称链关于 $\boldsymbol{\pi}$ 可逆。

**定理 7.10.** 若 $P$ 关于 $\boldsymbol{\pi}$ 可逆，则 $\boldsymbol{\pi}$ 为其平稳分布。

*证明.* $\forall j \in S$，有 $\sum_{i\in S} \pi_i p_{ij} = \sum_{i\in S} \pi_j p_{ji} = \pi_j \sum_{i\in S} p_{ji} = \pi_j$，故 $\boldsymbol{\pi} P = \boldsymbol{\pi}$。又 $\sum_{j\in S} \pi_j = 1$，故 $\boldsymbol{\pi}$ 为平稳分布。 □

**例 7.10.** （无向图上的 Random Walk）

设有一无向图，其结点集为 $S = \{1, \cdots, n\}$，边集为 $E$，各结点的度分别为 $d_1, \cdots, d_n$，且 $p_{ij} = \begin{cases} \dfrac{1}{d_i}, & (i,j) \in E, \\ 0, & 其他, \end{cases}$ 并记 $\pi_j = \frac{d_j}{\sum_{i=1}^n d_i}$，则对 $(i,j) \in E$ 有 $\pi_i p_{ij} = \frac{1}{\sum_{i=1}^n d_i} = \pi_j p_{ji}$，对 $(i,j) \notin E$ 有 $\pi_i p_{ij} = 0 = \pi_j p_{ji}$，即 $P$ 关于 $\boldsymbol{\pi}$ 可逆，故 $\boldsymbol{\pi}$ 为其平稳分布。

# 7.6 MCMC

MCMC 的全称是 Markov Chain Monte Carlo，是一类模拟复杂分布的高效算法，其基本思想是从状态空间上已知的任意不可约 Markov 链出发，经过调整，构建一个不可约 Markov 链，使其收敛到给定的平稳分布。

下面着重介绍其中的 Metropolis-Hastings 算法。

设 $\boldsymbol{\pi} = (\pi_1, \cdots, \pi_n)$ 为 $S = \{1, \cdots, M\}$ 上的目标分布，$\pi_j > 0, \forall j \in S$，否则可以删掉 $\pi_j = 0$ 的状态 $j$。现在已有 $S$ 上的某 Markov 链，转移概率为 $p_{ij}$，初始状态 $X_0$ 为随机或确定的。若当前状态 $X_n = i$，则遵循以下步骤调整链的下一次转移:

1. 根据 $p_{ij}$ 给出转移到 $j$ 的概率
2. 计算接受建议的概率 $a_{ij} = \begin{cases} \min\{\dfrac{\pi_j p_{ji}}{\pi_i p_{ij}}, 1\}, & p_{ij} > 0 \\ 1, & p_{ij} = 0 \end{cases}$
3. 若 $j \neq i$，调整后，以概率 $q_{ij} = p_{ij} a_{ij}$ 转入 $j$，即 $P(X_{n+1} = j | X_n = i) = q_{ij}$；若 $j = i$，则 $a_{ii} = 1$，有 $q_{ii} = p_{ii} + \sum_{j \neq i} (1 - a_{ij}) p_{ij} = 1 - \sum_{j \neq i} a_{ij} p_{ij} = 1 - \sum_{j \neq i} q_{ij}$ 的概率留在 $i$，即 $P(X_{n+1} = i | X_n = i) = q_{ii}$。

可以证明 $Q = (q_{ij})$ 为概率转移矩阵且关于 $\boldsymbol{\pi}$ 可逆，故 $\boldsymbol{\pi}$ 为其平稳分布。

**例 7.11.** 用 Metropolis-Hastings 算法模拟 Zipf 分布。对于 $a > 0$，有 $\pi_k = P(X = k) = \frac{\frac{1}{k^a}}{\sum_{j=1}^M \frac{1}{j^a}}, k = 1, \cdots, M$，状态空间 $S = \{1, \cdots, M\}$。建议选择原始链为 $S$ 上的 Random Walk，即

$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & \cdots & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \cdots & 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}, \text{而 } a_{ij} = \begin{cases} \min\{\dfrac{\pi_j p_{ji}}{\pi_i p_{ij}}, 1\}, & |i - j| = 1, \\ 1, & \text{其他}, \end{cases}$$

据此给出 $Q = (q_{ij})$，可以验证 $q_{ii} > 0, \forall i \in S$，故 $Q$ 不可约且所有状态遍历，故 $\boldsymbol{\pi}$ 为其唯一平稳分布。

# 第八章　Brown 运动

## 8.1　基本概念

本章的主题 Brown 运动是一种随机连续运动，其指标集 $T=[0,+\infty)$，状态空间 $S=\mathbb{R}$。

为理解 Brown 运动，先来考虑熟悉的随机游走。假设 $X(0)=0$，各 $X_i$ 独立同分布且 $P(X_i=1)=P(X_i=-1)=\frac{1}{2}$，以 $\Delta x$ 为步长，$\Delta t$ 为时间跨度，令 $t$ 时刻的粒子位置为 $X(t)=\Delta x\sum_{i=1}^{\lfloor t/\Delta t\rfloor}X_i$，则 $\mathrm{E}(X(t))=0,\mathrm{Var}(X(t))=(\Delta x)^2\lfloor t/\Delta t\rfloor$。令 $(\Delta x)^2=\sigma^2\Delta t$，则 $\mathrm{Var}(X(t))\to\sigma^2 t(\Delta t\to 0)$，且 $\{X(t)\}$ 具有平稳增量性和独立增量性。

**定义 8.1.** 称一个随机过程 $\{X(t),t\geq 0\}$ 为 *Brown 运动*，若其满足：

1. $X(0)=0$（这一要求并不本质）
2. $\{X(t),t\geq 0\}$ 有平稳增量性和独立增量性
3. $\forall t>0$，$X(t)\sim N(0,\sigma^2 t)$

若 $\sigma^2=1$，称之为标准 *Brown 运动*，记为 $\{B(t),t\geq 0\}$。

历史上，Brown 运动是由 Brown 于 1827 年发现的，Einstein 于 1905 年给出了其数学解释，而 Wiener 于 1918 年给出了精确的数学刻画。

根据平稳增量性，$\forall 0<s<t,B(t+s)-B(s)\sim N(0,\sigma^2 t)$。而根据独立增量性，$\forall s,t>0,a,x\in\mathbb{R},P(B(t+s)\leq x|B(s)=a,B(u)(0\leq u<s))=P(B(t+s)-B(s)\leq x-a|B(s)=a,B(u)(0\leq u<s))=P(B(t+s)-B(s)\leq x-a)=P(B(t+s)\leq x|B(s)=a)$，这说明 Brown 运动有 Markov 性（是 Markov 过程）。

可以证明，$B(t)$ 以概率 1 关于 $t$ 连续且处处不可微。直观地，$\mathrm{E}((B(t+\Delta t)-B(t))^2)=\Delta t$，故 $B(t+\Delta t)-B(t)\sim\sqrt{\Delta t}$，因此 $\lim_{\Delta t\to 0}\frac{B(t+\Delta t)-B(t)}{\Delta t}$ 不存在。

## 8.2　简单性质

由 $B(t)\sim N(0,t)$ 知其 PDF 为 $f_t(x)=\frac{1}{\sqrt{2\pi t}}e^{-\frac{x^2}{2t}},x\in\mathbb{R}$。

可以证明，$f_t(x)$ 满足 $\frac{\partial f}{\partial t}=\frac{1}{2}\frac{\partial^2 f}{\partial x^2}$（称为热传导方程），且是其在定解条件 $\int_{-\infty}^{+\infty} f(x,t)\equiv 1, \lim_{t\to 0} f(x,t)=0, \forall x\neq 0$ 下的唯一解。

由 $\forall t_1<t_2<\cdots<t_n, B(t_1)=x_1,\cdots,B(t_n)=x_n$ 等价于 $B(t_1)=x_1, B(t_2)-B(t_1)=x_2-x_1,\cdots,B(t_n)-B(t_{n-1})=x_n-x_{n-1}$，可得 $B(t_1),\cdots,B(t_n)$ 的联合 PDF 为 $f(x_1,\cdots,x_n)=f_{t_1}(x_1)f_{t_2-t_1}(x_2-x_1)\cdots f_{t_n-t_{n-1}}(x_n-x_{n-1})$。需要指出，这里由 $B(t_1), B(t_2)-B(t_1),\cdots,B(t_n)-B(t_{n-1})$ 的分布过渡到 $B(t_1),\cdots,B(t_n)$ 的分布时，严格的证明过程应当使用密度函数变换的方法。

由于 Brown 运动的良好性质，原则上我们可以计算任何想要的概率分布。

**例 8.1.** $\forall 0<s<t, a\in\mathbb{R}$，求在 $B(t)=a$ 下 $B(s)$ 的条件分布。
由 $B(t)=a, B(s)=x$ 等价于 $B(s)=x, B(t)-B(s)=a-x$，故所求条件 PDF 为 $f_{B(s)|B(t)}(x|B(t)=a)=\frac{f_s(x)f_{t-s}(a-x)}{f_t(a)}\propto e^{-\frac{x^2}{2s}-\frac{(x-a)^2}{2(t-s)}}\propto e^{-\frac{(x-\frac{as}{t})^2}{2\frac{s(t-s)}{t}}}$，故 $B(s)|B(t)=a\sim N(\frac{as}{t},\frac{s(t-s)}{t})$。若记 $\alpha=\frac{s}{t}$，则 $\mathrm{E}(B(s)|B(t)=a)=\alpha a, \mathrm{Var}(B(s)|B(t)=a)=\alpha(1-\alpha)t$，即条件方差与 $a$ 无关。

**命题 8.1.** Brown 运动具有对称性，即 $P(B(t_0+t)\geq x_0|B(t_0)=x_0)=P(B(t_0+t)\leq x_0|B(t_0)=x_0)=\frac{1}{2}, \forall t_0\geq 0, t>0, x_0\in\mathbb{R}$。

**定义 8.2.** 若对于 $\forall n\geq 1, \forall 0\leq t_1<t_2<\cdots<t_n$，有 $X(t_1),\cdots,X(t_n)$ 服从多元正态分布，则称 $\{X(t), t\geq 0\}$ 为 *Gauss* 过程。

Brown 运动是 Gauss 过程。

**定理 8.1.** Gauss 过程 $\{X(t), t\geq 0\}$ 是标准 Brown 运动的充要条件为

1. $X(0)=0$
2. $\mathrm{E}(X(t))=0$
3. $\mathrm{E}(X(t)X(s))=s, \forall 0\leq s<t$

*证明.* 必要性是显然的。下证充分性。
$\forall 0\leq t_1<t_2<t_3<t_4$，有 $\mathrm{E}((X(t_2)-X(t_1))(X(t_4)-X(t_3)))=t_2-t_2-t_1+t_1=0$，故 $X(t_2)-X(t_1)$ 与 $X(t_4)-X(t_3)$ 不相关，又由多元正态分布性质知 $X(t_2)-X(t_1)$ 与 $X(t_4)-X(t_3)$ 相互独立，即证明了独立增量性。又由 $\mathrm{E}((X(t)-X(s))^2)=t+s-2s=t-s, \forall 0\leq s<t$，得 $X(t)-X(s)\sim N(0,t-s)$，即证明了平稳增量性，以及 $X(t)\sim N(0,t)$。 □

**命题 8.2.** 如下定义的 $\{W(t), t\geq 0\}$ 均为标准 Brown 运动。

1. $W(t)=-B(t), t\geq 0$（反射，reflection）

2. $W(t) = B(t+h) - B(h), t \geq 0$，其中 $h > 0$ 固定（平移，translation）
3. $W(t) = \frac{B(at)}{\sqrt{a}}, t \geq 0$，其中 $a > 0$ 固定（伸缩，scaling）
4. $W(t) = tB(\frac{1}{t}), t > 0, W(0) = 0$（反演，inversion）
5. $W(t) = B(T-t) - B(T), t \in [0, T]$，其中 $T > 0$ 固定

## 8.3 首中时与最大值

Brown 运动首次到达 $a$ 处的时刻 $T_a = \inf\{t | t \geq 0, B(t) = a\}$ 称为 $a$ 的首中时，而 $M_t = \sup_{0 \leq s \leq t} B(s)$ 为 $B(t)$ 在 $[0, t)$ 上的最大值。

显然 $\forall a \geq 0$，事件 $\{T_a \leq t\}$ 与 $\{M_t \geq a\}$ 等价。注意到对称性（如图 8.1），有 $P(B(t) \geq a) = P(T_a \leq t)P(B(t) \geq a | T_a \leq t) + P(T_a \geq t)P(B(t) \geq a | T_a \geq t) = P(T_a \leq t)P(B(t) \geq a | T_a \leq t) = \frac{1}{2}P(T_a \leq t)$，因此 $P(T_a \leq t) = 2P(B(t) \geq a) = 2(1 - \Phi(\frac{a}{\sqrt{t}}))$。

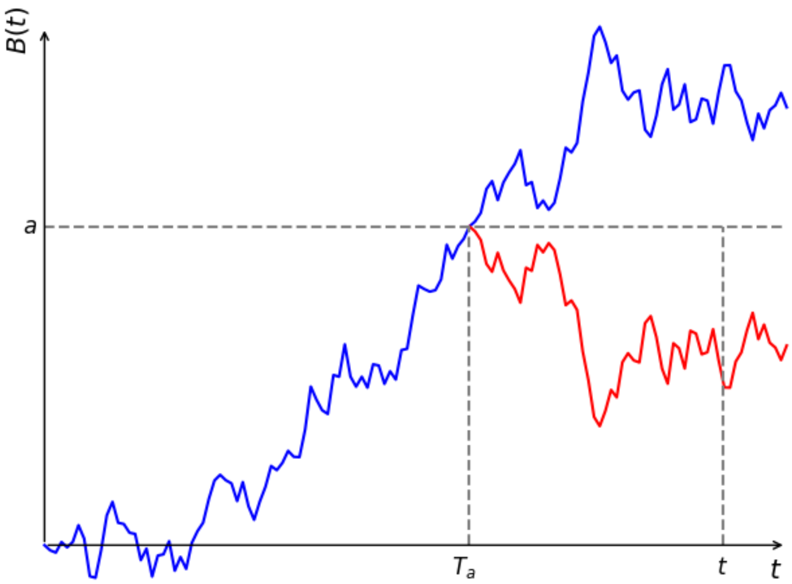



图 8.1: Brown 运动在首中时后的对称性

**命题 8.3.** 粒子从 0 出发，在有限时刻到达 $a$ 的概率为 1，即 $P(T_a < +\infty) = 1$（具有常返性）；但首次到达 $a$ 的平均用时 $\mathrm{E}(T_a) = +\infty$（具有零常返性）。

## 8.4 变形与推广

## 8.5 Ito 积分