随机过程 B
第 1 章 引论

殷哲 yinzhe@ustc.edu.cn

解扬洋 xieyclio@ustc.edu.cn

教材与参考书

- 教材:
  - 方兆本、缪柏其, 《随机过程》(第三版), 科学出版社, 2011 年.
- 参考书:
  - 更新过程:
    Ross, S.M. (1995). Stochastic Processes (2nd Edition), John Wiley & Sons.
    (龚光鲁译, 《随机过程》(第二版), 机械工业出版社, 2013)
  - Markov 链:
    李育强、姚强, 《应用随机过程》, 高等教育出版社, 2021 年.
  - 宽平稳过程:
    林元烈, 《应用随机过程》, 清华大学出版社, 2002 年.

成绩计算

平时成绩 (30%) $+$ 期末考试 (70%)

课程内容

- 第 1 章 引论
  - ▷ 基本概念
  - ▷ 条件期望和矩母函数
  - ▷ 收敛性
- 第 2 章 泊松过程
  - ▷ 定义
  - ▷ 间隔时间与到达时间
  - ▷ 泊松过程推广
- 第 3 章 马可夫过程
  - ▷ 定义
  - ▷ 马可夫链的状态分类
  - ▷ 极限定理与平稳分布
  - ▷ 分支过程
  - ▷ 连续时间马可夫链
- 第 4 章 平稳过程
  - ▷ 定义
  - ▷ 遍历性定理
  - ▷ 协方差函数与功率谱密度
- 第 5 章 布朗运动 *
  - ▷ 定义
  - ▷ 布朗运动的性质
  - ▷ (伊藤积分)

本章提纲

基本概念

定义 随机变量 (random variable)

随机变量是从样本空间(定义域) Ω 到可测空间(值域) E 的函数.

- 最常见的可测空间为 $E=\mathbb{R}$, 也即随机变量的取值均为实数
- 概率空间三元组 $(\Omega, \mathcal{F}, \mathbb{P})$
  - 样本空间: 集合 Ω, 包含了随机试验所有可能的结果 (基本事件)
  - 事件域: $\mathcal{F}$, 由 Ω 中元素的集合构成的域, 例如 $\mathcal{F}=2^{\Omega}$

定义 事件域 (集合代数)

我们称 $\mathcal{F}$ 是一个 (事件) 域, 若以下 2 个条件同时满足:

1. $\emptyset, \Omega \in \mathcal{F}$
2. 若 $A \in \mathcal{F}$ 且 $B \in \mathcal{F}$, 则 $A \cup B$, $A \cap B$, $A \setminus B \in \mathcal{F}$

- 注意: 集合域通常不是抽象代数中定义的域, 由于交运算和并运算没有逆元素
- 定义中 $A \setminus B \in \mathcal{F}$, 可替换为 $\bar{A} \in \mathcal{F}$, 为什么? 利用 $A \setminus B = A \cap \bar{B}$

基本概念 (续)

定义 概率函数 (概率测度)

概率函数 $\mathbb{P}:\mathcal{F}\mapsto[0,1]$ 将域中的每个事件对应一个概率, 且满足

- 非负性: $\mathbb{P}(\omega)\geq 0,\ \mathbb{P}(\omega)\in\mathbb{R}$
- 归一性(正则性): $\mathbb{P}(\Omega)=1$
- 可列可加性: $\mathbb{P}(\bigcup\limits_{i=1}^{\infty}\omega_i)=\sum\limits_{i=1}^{\infty}\mathbb{P}(\omega_i)$, 其中 ω_i 为不相交的事件

- 如何推导出 $\mathbb{P}(\emptyset)=0$? 利用可列可加性
- 如何由概率函数的可列可加性推导出有限可加性?

$$\mathbb{P}(\bigcup_{i=1}^{\infty}\omega_i)=\sum_{i=1}^{\infty}\mathbb{P}(\omega_i)\Rightarrow\mathbb{P}(\bigcup_{i=1}^{n}\omega_i)=\sum_{i=1}^{n}\mathbb{P}(\omega_i)$$

其中 ω_i 为不相交的事件

- 例: 随机变量 $X(\omega)$, $\omega \in \Omega$

ω	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6
$X(\omega)$	1	0	1	0	1	0
$\mathbb{P}$	0.1	0.2	0.3	0.1	0.2	0.1

  随机变量的概率分布为:
  $\mathbb{P}(X=1)=\mathbb{P}(\{\omega_1,\omega_3,\omega_5\})=0.6$
  $\mathbb{P}(X=0)=\mathbb{P}(\{\omega_2,\omega_4,\omega_6\})=0.4$

- 例子中随机变量 X 对应的 $(\Omega,\mathcal{F},\mathbb{P})$ 以及 E 分别是什么?

基本概念 (续)

定义 1.1 随机过程 (stochastic process)

随机过程是一族随机变量 $\{X(t), t \in T\}$, 其中 t 是参数, 它属于某个指标集 T, T 称为参数集

- 参数 t 通常代表时间, 可以离散或连续
  - 当时间 $T = \{0, 1, 2, \cdots\}$ 为离散时, 也称为随机序列
- 随机过程是随时间变化的随机变量

- 因此, 随机过程可以看做是关于时间 $t \in T$ 和样本 $\omega \in \Omega$ 的 (二元) 函数, 记作 $X(t, \omega)$
  - 给定 $\omega = \omega_0$ 时, $X(t, \omega_0)$ 是一条样本路径/轨道 (sample path), 是关于 t 的确定性函数
  - 给定 $t = t_0$ 时, $X(t_0, \omega)$ 是一个随机变量

- 随机过程在时刻 t 的取值 $X(t, \omega)$, 称为过程所处的状态
  - 状态的全体, 称为状态空间
  - 实值随机过程的状态是实数, 而非实数向量

- 依据参数 (时间) 集和状态空间的不同, 可将常见随机过程分为 4 类

	离散时间	连续时间
离散状态	随机徘徊 离散时间马可夫链	泊松过程 连续时间马可夫链
连续状态	离散时间马可夫过程	布朗运动 连续时间马可夫过程

- 当 T 是高维向量时, 则称 $X(t)$ 是随机场

- 例 1.1 布朗运动:$X(t)$
  - ▷ 粒子在时刻 t 的一维位置 $X(t)$, 服从正态分布 $N(0,t)$
  - ▷ 不确定性随时间增加
  - ▷ 一种定义: 轨道连续的平稳独立增量过程
- 例 1.2 随机徘徊:$X(t)$
  - ▷ 一醉汉每时刻以概率 p 前进一步, 概率 $1-p$ 后退一步
  - ▷ $X(t)$ 表示 t 时刻时, 醉汉在街上的位置
  - ▷ 用以刻画股票价格、质点运动、赌博问题等
- 例 1.3 神经细胞的位势:$X(t)$
  - ▷ 神经细胞在其细胞膜电位 $X(t)$ 达到某一临界值 C 时, 会产生兴奋
  - ▷ 刺激脉冲 (提高电位) 和抑制脉冲 (降低电位) 以一定速率 (例如 Poisson 过程) 抵达细胞
  - ▷ 每个脉冲带来的升降幅度, 服从相同的分布 $H(x)$
  - ▷ 兴奋过后, 细胞膜电位恢复到 0
  - ▷ 我们会关注两次兴奋 (更新点) 的间隔时间, 记为 T_i

- 例 1.5 流行病学模型
  - ▷ $X(t)$: t 时刻易感染人群的大小
  - ▷ $Y(t)$: t 时刻已被传染人数
  - ▷ 易感人群被传染的概率为 p
  - ▷ 每一时刻都有一部分易感染人群被传染

$$X(t) = X(t+1) + Y(t+1)$$

  - ▷ 对 $j \le i$ 时 ($j > i$ 时如何?), 有

$$\mathbb{P}(X(t+1) = j | X(t) = i) = C_i^{i-j} p^{i-j} (1-p)^j$$

  - ▷ 当 t 趋于无穷时, 感染和未感染的人群如何分布?
- 例 1.7 水库库容调度
  - ▷ $X(t)$: 第 t 年初的水库蓄水量
  - ▷ $Y(t)$: 水库在 $(t, t+1)$ 年间的蓄水量, 是随机值
  - ▷ K: 水库的设计库容
  - ▷ M: 每年年底固定的清库泄洪量, $M < K$

$$X(t+1) = \min\{X(t) + Y(t), K\} - \min\{X(t) + Y(t), M\}$$

  - ▷ 当年蓄水量较多或较少时, 来年的初始容量有什么规律?

思考与讨论

- 什么是随机过程？离散时间随机过程和离散状态随机过程有什么关系？
- 随机过程和随机向量的区别是什么？

对于随机变量 X:

- 其概率特征反映在分布函数中

$$F(x) = \mathbb{P}(X \leq x)$$

对于随机过程 $\{X(t),\ t \in T\}$:

- 定义过程的一维分布为

$$F_t(x) = \mathbb{P}(X(t) \leq x)$$

- 过程一维分布的数字特征有
  - ▷ 均值函数: $\mathbb{E}X(t)$, 也记作 $\mu_X(t)$
  - ▷ 方差函数: $\mathrm{Var}[X(t)]$

- 然而, 一维分布不足以完全描述随机过程的概率特征

- 还需了解随机过程在不同时刻对应的随机变量的联合分布 (关联性), 也即过程的有限维分布

例如, $X(t_1)$ 与 $X(t_2)$ 的二维联合分布, 相应的数字特征有

- 自相关函数: $\mathbb{E}[X(t_1)X(t_2)]$, 也记作 $r_X(t_1, t_2)$
  - ▷ 与二阶矩的关系: $r_X(t, t) = \mathbb{E}[X(t)]^2$
  - ▷ 对称性: $r_X(t_1, t_2) = \mathbb{E}[X(t_1)X(t_2)] = \mathbb{E}[X(t_2)X(t_1)] = r_X(t_2, t_1)$
  - ▷ 非负定性: 对任意 $t_1, t_2, \cdots, t_n \in T$ 及任意实数 $b_1, b_2, \cdots, b_n$, 总有

$$\mathbf{b}^T \left(r_X(t_i, t_j)\right)_{n\times n} \mathbf{b} = \sum_{i=1}^n \sum_{j=1}^n b_i b_j r_X(t_i, t_j) \geq 0$$

证明: 基于 $\mathbb{E}[X(t_i)X(t_j)]$ 的关于 $X(t_i)$ 和 $X(t_j)$ 的双线性性 (关于两个自变量均有线性性: $f(ax + by) = af(x) + bf(y)$), 有

$$\begin{aligned}
\sum_{i=1}^n \sum_{j=1}^n b_i b_j r_X(t_i, t_j) &= \sum_{i=1}^n \sum_{j=1}^n b_i b_j \mathbb{E}[X(t_i)X(t_j)] \\
&= \mathbb{E}\left[\sum_{i=1}^n \sum_{j=1}^n b_i X(t_i) b_j X(t_j)\right] = \mathbb{E}\left[\sum_{i=1}^n b_i X(t_i)\right]\left[\sum_{j=1}^n b_j X(t_j)\right] \\
&= \mathbb{E}\left[\sum_{i=1}^n b_i X(t_i)\right]^2 \geq 0
\end{aligned}$$

- 协方差函数:

$$
\begin{aligned}
& R_X(t_1, t_2) \triangleq \mathrm{Cov}(X(t_1), X(t_2)) \\
= \; & \mathbb{E}\{[X(t_1) - \mathbb{E}X(t_1)][X(t_2) - \mathbb{E}X(t_2)]\} \\
= \; & \mathbb{E}\{X(t_1)X(t_2) - X(t_1)\mathbb{E}X(t_2) - X(t_2)\mathbb{E}X(t_1) + \mathbb{E}X(t_1)\mathbb{E}X(t_2)]\} \\
= \; & \mathbb{E}[X(t_1)X(t_2)] - \mathbb{E}X(t_1)\mathbb{E}X(t_2)
\end{aligned}
$$

- 与方差的关系: $R_X(t,t) = \mathrm{Var}[X(t)] \geq 0$
- 协方差函数有对称性、非负定性
- 非负定性证明: 协方差函数有双线性性

$$
\mathrm{Cov}(b_1X(t_1)+b_2X(t_2), X(s)) = b_1\mathrm{Cov}(X(t_1), X(s))+b_2\mathrm{Cov}(X(t_2), X(s))
$$

- 再利用方差的非负性

$$
\begin{aligned}
0 \quad \leq \quad & \mathrm{Var}\left[\sum_{i=1}^n b_iX(t_i)\right] = \mathrm{Cov}\left(\sum_{i=1}^n b_iX(t_i), \sum_{j=1}^n b_jX(t_j)\right) \\
= \quad & \sum_{i=1}^n b_i\mathrm{Cov}\left(X(t_i), \sum_{j=1}^n b_jX(t_j)\right) = \sum_{i=1}^n\sum_{j=1}^n b_ib_jR_X(t_i,t_j) = \mathbf{b}^T\left(R_X(t_i,t_j)\right)_{n\times n}\mathbf{b}
\end{aligned}
$$

定义 有限维分布族

对所有正整数 n, 有限维分布函数

$$F_{t_1,t_2,\cdots,t_n}(x_1,x_2,\cdots,x_n)=\mathbb{P}\left(\{X(t_1)\le x_1,\cdots,X(t_n)\le x_n\}\right),\ \forall t_1,\cdots,t_n\in T$$

组成的集合, 称为随机过程 $\{X(t),\ t\in T\}$ 的有限维分布族.

- 知道了有限维分布族, 就知道了过程任意 n 个时刻 (n 也是任意的) 对应随机变量的联合分布
- 对称性: 对 $(1,\cdots,n)$ 的任一置换 $(i_1,\cdots,i_n)$ 有

$$F_{t_1,t_2,\cdots,t_n}(x_1,x_2,\cdots,x_n)=F_{t_{i_1},t_{i_2},\cdots,t_{i_n}}(x_{i_1},x_{i_2},\cdots,x_{i_n})$$

  证明直接根据定义
- 相容性: 对 $m<n$

$$F_{t_1,\cdots,t_m,t_{m+1},\cdots,t_n}(x_1,\cdots,x_m,\infty,\cdots,\infty)=F_{t_1,\cdots,t_m}(x_1,\cdots,x_m)$$

  必然事件不影响其它事件的分布

有限维分布和数字特征 (续)

- 例子: 设 X_t 为第 t 次独立地扔一均匀六面骰子得到的点数
- 则 $\{X_t,\ t \geq 1\}$ 是一个随机过程
- 参数 (时间指标) 集: $T = \{1, 2, \cdots\}$
- 状态空间: $\{1, 2, 3, 4, 5, 6\}$
- 均值:
$$\mathbb{E}X_t = \frac{1+2+3+4+5+6}{6} = 3.5,\ \forall t$$
- 方差:
$$\mathrm{Var}X_t = \mathbb{E}X_t^2 - (\mathbb{E}X_t)^2 = \frac{35}{12}$$
- 协方差: $\mathrm{Cov}(X_t, X_s) = \mathbb{E}X_tX_s - \mathbb{E}X_t\mathbb{E}X_s = 0,\ \forall t \neq s$
- 有限维分布:
$$F_{t_1,t_2,\cdots,t_n}(x_1, x_2, \cdots, x_n) = F(x_1)F(x_2)\cdots F(x_n),$$
其中 $F(x)$ 为 X_1 的分布函数

思考与讨论

- 为什么说一维分布不足以刻画随机过程的概率特征?

- 协方差函数和自相关函数的区别是什么?

定义 同分布

若两个随机变量 X_1, X_2 的分布函数 $F_{X_1}(x)$ 与 $F_{X_2}(x)$ 对任意 x 都是相等的, 则称它们为同分布的, 记作 $X_1 \stackrel{d}{=} X_2$.

- 注意区别 $X_1 \stackrel{d}{=} X_2$ 和 $X_1 = X_2$
- 类似地, 如果一个随机向量 $\mathbf{X} = (X_1, \cdots, X_n)$ 与另一随机向量 $\mathbf{Y} = (Y_1, \cdots, Y_n)$ 有相同的联合分布, 则也称其为同分布的, 记作 $\mathbf{X} \stackrel{d}{=} \mathbf{Y}$

定义 1.2 严平稳过程

若随机过程 $\{X(t),\ t \in T\}$, 对任意的 $t_1, \cdots, t_n \in T$ 和任意 h 有

$$(X(t_1+h), \cdots, X(t_n+h)) \stackrel{d}{=} (X(t_1), \cdots, X(t_n)),$$

则称过程为严平稳的.

- 任意 n 维的概率特性与时间起点无关
- 平稳即指概率特性 (例如有限维分布) 关于时间的不变性

定义 1.3 宽平稳过程

若随机过程 $\{X(t),\ t \in T\}$ (i) 在所有时刻的二阶矩存在 ($\mathbb{E}X^2(t) < \infty$, $\forall t$, 也称二阶矩过程), (ii) $\mathbb{E}X(t) = m$, $\forall t$, 且 (iii) 协方差函数 $R_X(t,s)$ 只与时间差 $t-s$ 有关, 则称过程为宽平稳的或二阶矩平稳的.

- 宽平稳过程实际上只保证了一、二阶矩是平稳的
- 一个二阶矩过程 (所有时刻二阶矩 $\mathbb{E}X^2(t)$ 存在), 一定是一阶矩过程 (所有时刻一阶矩 $\mathbb{E}X(t)$ 存在) 吗?
- 若一个随机变量的高阶矩存在, 则低阶矩一定存在

$$\mathbb{E}X^{n+1} = \int_{-\infty}^{-1} x^{n+1}f(x)dx + \int_{-1}^{1} x^{n+1}f(x)dx + \int_{1}^{\infty} x^{n+1}f(x)dx$$

$$\mathbb{E}X^{n} = \int_{-\infty}^{-1} x^{n}f(x)dx + \int_{-1}^{1} x^{n}f(x)dx + \int_{1}^{\infty} x^{n}f(x)dx$$

其中 $$\int_{1}^{\infty} x^{n}f(x)dx \le \int_{1}^{\infty} x^{n+1}f(x)dx < \infty, \quad \left|\int_{-1}^{1} x^{n}f(x)dx\right| \le \int_{-1}^{1} f(x)dx$$

$$\int_{-\infty}^{-1} |x^{n}|f(x)dx \le \int_{-\infty}^{-1} |x^{n+1}|f(x)dx = \left|\int_{-\infty}^{-1} x^{n+1}f(x)dx\right| < \infty$$

平稳过程 (续)

- 严平稳过程未必是宽平稳的, 例如二阶矩不存在的情形
- 二阶矩存在的严平稳过程是宽平稳的, 证明如下:
  - ▷ 由严平稳过程定义 $(X(t_1+h),\cdots,X(t_n+h)) \overset{d}{=} (X(t_1),\cdots,X(t_n))$, 可知 $X(t+h) \overset{d}{=} X(t)$, $\forall t,h$, 从而 (二阶矩存在保障一阶矩存在)
    $$\mathbb{E}X(t+h) = \mathbb{E}X(t),\ \forall t,h \qquad (\text{一阶矩为常数})$$
  - ▷ 关于协方差函数, 二阶矩存在, 意味着协方差 (自相关) 函数存在
    $[\mathbb{E}X(t)X(s)]^2 \le \mathbb{E}X^2(t)\mathbb{E}X^2(s) < \infty$ (Cauchy–Schwarz 不等式, 见讨论)
  - ▷ 利用 $(X(t),X(s)) \overset{d}{=} (X(t-s),X(0))$, 有
    $$\begin{aligned} R_X(t,s) &= \mathrm{Cov}(X(t),X(s)) \\ &= \mathrm{Cov}(X(t-s),X(0)) \\ &= R_X(t-s,0) \qquad (\text{协方差函数只与时间差 } t-s \text{ 相关}) \end{aligned}$$
- 对于宽平稳过程, 总有 $R_X(t,s) = R_X(t-s,0)$
  - ▷ 故可将协方差函数记为 $R_X(t-s)$
  - ▷ 由对称性, 得 $R_X(t) = R_X(t,0) = R_X(0,t) = R_X(-t)$ 为偶函数
  - ▷ $R_X(0) = R_X(t,t) = \mathrm{Var}X(t),\ \forall t$

思考与讨论

- 平稳性的直观含义是什么?
- 宽平稳过程和严平稳过程的联系和区别是什么?
- 如何证明 Cauchy–Schwarz 不等式?

$$[\mathbb{E}XY]^2 \leq \mathbb{E}X^2\mathbb{E}Y^2$$

  考虑 $\mathbb{E}(X+tY)^2 \geq 0$

定义 1.4 独立增量过程

若对任意 $t_1 < t_2 < \cdots < t_n$, $t_1, t_2, \cdots, t_n \in T$, 随机变量 $X(t_2) - X(t_1)$, $X(t_3) - X(t_2)$, $\cdots$, $X(t_n) - X(t_{n-1})$ 是相互独立的, 则称 $X(t)$ 为独立增量过程.

- 例 1.10:
  - ▷ 设 Z_i, $i = 0, 1, 2, \cdots$, 是一系列独立的随机变量
  - ▷ 定义 $X_n = \sum_{i=0}^{n} Z_i$
  - ▷ 则过程 $\{X_n, \ n \geq 0\}$ 是独立增量过程, 也称 X_n 为独立和

定义 平稳增量过程

若过程 $X(t)$ 满足对任意的 t_1, t_2 和 h, 有 $X(t_1 + h) - X(t_1) \stackrel{d}{=} X(t_2 + h) - X(t_2)$, 则称 $X(t)$ 为平稳增量过程.

- $X(t_1 + h) - X(t_1)$ 与 $X(t_2 + h) - X(t_2)$ 未必独立

定义 平稳独立增量过程

若独立增量过程 $X(t)$ 满足对任意的 t_1, t_2 和 h, 有 $X(t_1+h)-X(t_1) \overset{d}{=} X(t_2+h)-X(t_2)$, 则称 $X(t)$ 为平稳独立增量过程.

- 例 1.10 中, 若 Z_i 是独立同分布的, 则 $X_n = \sum_{i=0}^{n} Z_i$ 是平稳独立增量过程
- 平稳独立增量过程不一定是平稳过程, 仅指增量是平稳的
  - 例如:

$$X_n = \sum_{i=0}^{n} Z_i,$$

    其中 Z_i 独立同分布, $\mathbb{E}Z_i = m \neq 0$
  - 过程的均值不是常数:

$$\mathbb{E}X_{n+1} = (n+2)m \neq (n+1)m = \mathbb{E}X_n,$$

    因此, 过程不是宽平稳或严平稳的

独立增量过程 (续)

- 平稳独立增量过程的均值 (如存在) 一定是 t 的线性函数, 证明如下
  - ▷ 设 $\mathbb{E}X(t) = f(t)$
  - ▷ 由平稳独立增量过程定义 $X(t_1+h) - X(t_1) \stackrel{d}{=} X(t_2+h) - X(t_2)$ 知, 对任意 t_1, t_2 有

$$\frac{f(t_1+h) - f(t_1)}{h} = \frac{f(t_2+h) - f(t_2)}{h}, \quad \forall h$$

  - ▷ 令 $h \to 0$, 可得

$$f'(t_1) = f'(t_2) = \text{constant}$$

- Poisson 过程和 Brown 运动都是平稳独立增量过程

思考与讨论

- 平稳过程、平稳增量过程、平稳独立增量过程的联系和区别是什么?
- 平稳增量过程一定是平稳过程吗? 平稳过程一定是平稳增量过程吗?

条件期望

- 给定事件 B 发生时, 事件 A 发生的条件概率

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

  - ▷ 若 $\mathbb{P}(B) = 0$, 则条件概率无法用上式定义
  - ▷ 此时, $\mathbb{P}(A|B)$ 可以定义为任意 (符合概率公理的) 的值
  - ▷ 每一个 $\mathbb{P}(A|B)$ 的取值, 都是条件概率的一种形式 (version)
- 对于离散型随机变量, 给定事件 $\{\omega : Y(\omega) = y\}$ (简记为 $\{Y = y\}$) 且 $\mathbb{P}(Y = y) > 0$ 时, X 取值为 x 的条件概率为

$$\mathbb{P}(X = x|Y = y) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)}$$

- 相应的条件分布函数定义为

$$F(x|y) = \mathbb{P}(X \leq x|Y = y)$$

- 条件期望定义为

$$\mathbb{E}(X|Y = y) = \sum_x x\mathbb{P}(X = x|Y = y)$$

- 对于连续型随机变量, $\mathbb{P}(Y=y)$ 通常为 0, 此时

$$\mathbb{P}(X=x|Y=y)=\frac{\mathbb{P}(X=x,Y=y)}{\mathbb{P}(Y=y)}$$

没有意义

- 此时, 为定义条件概率, 我们取一个包含 y 的小区间 Δ_y
  - ▷ 若 $\mathbb{P}(Y\in\Delta_y)=0$, 则定义相应的条件概率 (的一种形式) 为

$$\mathbb{P}(X\in A|Y=y)=0$$

  - ▷ 若 $\mathbb{P}(Y\in\Delta_y)>0$, 则定义

$$\mathbb{P}(X\in A|Y=y)=\lim_{\Delta_y\downarrow 0}\mathbb{P}(X\in A|Y\in\Delta_y)$$

其中 $\Delta_y\downarrow 0$ 表示使 Δ_y 的区间长度趋于 0

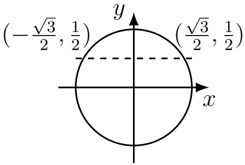
- 例如, 假设 (X,Y) 是均匀分布在单位圆内的随机点, 概率密度为 $\frac{1}{\pi}$
- 则 $\mathbb{P}(X=x)=0,\ \forall x$, $\mathbb{P}(Y=y)=0,\ \forall y$
- 考虑 $\mathbb{P}(X\in A|Y=0.5)$, 其中 $A=[-a,a]$ 且 $a\le\frac{\sqrt{3}}{2}$
- 包含 $y=0.5$ 的小区间为 $\Delta_y=(0.5-h,0.5+h)$
- 由图可知

$$\mathbb{P}(Y\in\Delta_y)=\left[2\cdot\frac{\sqrt{3}}{2}\cdot 2h+o(h)\right]\frac{1}{\pi}=\frac{2\sqrt{3}}{\pi}h+o(h)\qquad (\text{类梯形区域})$$

$$\mathbb{P}(Y\in\Delta_y,X\in A)=2a\cdot 2h\cdot\frac{1}{\pi}=\frac{4}{\pi}ah\qquad (\text{矩形区域})$$

- 从而有

$$\begin{aligned}\mathbb{P}(X\in A|Y=y)&=\lim_{\Delta_y\downarrow 0}\mathbb{P}(X\in A|Y\in\Delta_y)\\ &=\lim_{\Delta_y\downarrow 0}\frac{\mathbb{P}(X\in A,\ Y\in\Delta_y)}{\mathbb{P}(Y\in\Delta_y)}\\ &=\lim_{h\to 0+}\frac{\frac{4}{\pi}ah}{\frac{2\sqrt{3}}{\pi}h+o(h)}=\frac{2}{3}\sqrt{3}a,\quad a\le\frac{\sqrt{3}}{2}\end{aligned}$$



条件期望 (续)

- 条件分布函数: 在条件概率的定义中, 取 $A=(-\infty,x)$

$$F(x|y)=\mathbb{P}(X\le x|Y=y)=\lim_{\Delta_y\downarrow 0}\mathbb{P}(X\le x|Y\in\Delta_y)$$

- 条件密度函数: 若存在 $f(x|y)$, 对任意集合 A 有

$$\mathbb{P}(X\in A|Y=y)=\int_A f(x|y)dy$$

且 $\int_{-\infty}^{\infty} f(x|y)dx=1$, 则称 $f(x|y)$ 为给定 $Y=y$ 时, X 的条件密度

- 条件分布函数与条件密度函数的关系:

$$F(x|y)=\int_{-\infty}^{x} f(x|y)dx$$

条件期望 (续)

- 对随机变量 X 与 Y, 若存在非负函数 $f(x,y)$, 使得

$$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}f(u,v)dvdu$$

  则称 $f(x,y)$ 为随机变量 X 与 Y 的联合概率密度

- 记 $f(y)=\int_{-\infty}^{+\infty}f(x,y)dx$ 为 Y 的边缘概率密度, 则有

$$f(x,y)=f(x|y)f(y)\qquad (\text{联合}=\text{条件}\times\text{边缘})$$

- 条件期望: 给定 $Y=y$, X 的条件期望为

$$\mathbb{E}(X|Y=y)=\int xf(x|y)dx=\int xdF(x|y)$$

  ▷ $\mathbb{E}(X|Y=y)$ 是一个关于 y 的确定性函数
  ▷ $\mathbb{E}(X|Y)$ 是一个随机变量

例 1.11: 投掷一枚硬币, 出现正面的概率为 p, 独立地进行 n 次投掷

- 记 S 为 n 次试验中出现正面的次数: $S \in \{0, 1, \cdots, n\}$
- T 为首次出现正面的试验次数: T 的取值可能有哪些?
- 求给定 n 次试验中仅出现 1 次正面时, 随机变量 T 的条件分布, 也即 $\mathbb{P}(T = k|S = 1)$

解: 注意到 $\{T = k, S = 1\}$ 实际上只包含了一种可能的试验结果序列

$$\text{反}\cdots\text{反}\ \underbrace{\text{正}}_{\text{第 } k \text{ 个}}\ \text{反}\cdots\text{反}$$

- $\mathbb{P}(T = k, S = 1) = p(1-p)^{n-1}$, 与 k 无关
- $\mathbb{P}(S = 1) = C_n^1 p(1-p)^{n-1}$
- $\mathbb{P}(T = k|S = 1) = \frac{\mathbb{P}(T=k, S=1)}{\mathbb{P}(S=1)} = \frac{1}{n}, \quad \forall k$
- 唯一的一次正面, 将以相同的机会在 n 次试验中出现

命题 1.1 条件期望的性质

(a) 若 X 与 Y 独立, 则

$$\mathbb{E}(X|Y=y)=\mathbb{E}X$$

(b) 平滑性 (全期望公式):

$$\mathbb{E}X=\int\mathbb{E}(X|Y=y)dF_Y(y)=\mathbb{E}[\mathbb{E}(X|Y)]$$

(c) 对随机变量 X, Y 的函数 $\phi(X,Y)$ 有

$$\mathbb{E}[\phi(X,Y)|Y=y]=\mathbb{E}[\phi(X,y)|Y=y]$$

- (a) 独立时, 条件期望与无条件期望相同
- (b) 期望的计算可以两步走, 先在给定条件下求条件期望, 再对条件期望做概率加权平均
- (c) 可以将条件期望中取条件的随机变量, 直接替换成其条件取值
- 3 条性质分别证明如下

证明思路:

(a) 独立时, 条件分布与无条件分布相同

$$\mathbb{P}(X=x|Y=y)=\frac{\mathbb{P}(X=x,Y=y)}{\mathbb{P}(Y=y)}=\mathbb{P}(X=x)$$

$$\mathbb{E}(X|Y=y)=\sum_x x\mathbb{P}(X=x|Y=y)=\sum_x x\mathbb{P}(X=x)=\mathbb{E}X$$

(b) 以离散情形为例

$$\begin{aligned}
\mathbb{E}(X) &= \sum_x x\mathbb{P}(X=x) \\
&= \sum_x x\left[\sum_y \mathbb{P}(X=x,Y=y)\right] \qquad \text{(将事件关于 } Y \text{ 的取值进行拆分)} \\
&= \sum_x\sum_y \mathbb{P}(Y=y)\left[x\frac{\mathbb{P}(X=x,Y=y)}{\mathbb{P}(Y=y)}\right] \\
&= \sum_y \mathbb{P}(Y=y)\left[\sum_x x\frac{\mathbb{P}(X=x,Y=y)}{\mathbb{P}(Y=y)}\right] \qquad \text{(交换求和顺序)} \\
&= \sum_y \mathbb{P}(Y=y)\mathbb{E}(X|Y=y)=\mathbb{E}[\mathbb{E}(X|Y)]
\end{aligned}$$

证明思路:

(c) 把 $Z=\phi(X,Y)$ 看作一个随机变量, 其是随机向量 (X,Y) 的函数

$$
\begin{aligned}
& \mathbb{E}[\phi(X,Y)|Y=y] \\
= & \sum_{(x_i,y_j)} \phi(x_i,y_j)\mathbb{P}((X,Y)=(x_i,y_j)|Y=y) \quad \text{(按 } (X,Y) \text{ 取值计算条件期望)} \\
= & \sum_{(x_i,y_j)} \phi(x_i,y_j)\frac{\mathbb{P}((X,Y)=(x_i,y_j),Y=y)}{\mathbb{P}(Y=y)} \quad \text{(由条件概率定义)} \\
= & \sum_{x_i} \phi(x_i,y)\frac{\mathbb{P}((X,Y)=(x_i,y),Y=y)}{\mathbb{P}(Y=y)} \quad \text{(互斥事件同时发生的概率为 0)} \\
= & \sum_{x_i} \phi(x_i,y)\frac{\mathbb{P}(X=x_i,Y=y)}{\mathbb{P}(Y=y)} \quad \text{(用等概率事件替换)} \\
= & \sum_{x_i} \phi(x_i,y)\mathbb{P}(X=x_i|Y=y) \\
= & \mathbb{E}[\phi(X,y)|Y=y] \quad \text{(按 } X \text{ 取值计算条件期望)}
\end{aligned}
$$

思考与讨论

- $\mathbb{P}(A|B)$, $\mathbb{P}(X=x|Y)$, $\mathbb{P}(X=x|Y=y)$ 的联系和区别是什么？其中 A 和 B 是随机事件，X 和 Y 是随机变量
- $\mathbb{E}(X|Y)$, $\mathbb{E}(X|Y=y)$ 的联系和区别呢？

定义 1.5 矩母函数 (也称矩生成函数)

随机变量 X 的矩母函数定义为 $\mathbb{E}e^{tX}$, 记作

$$g_X(t) = \mathbb{E}e^{tX} = \int e^{tx} dF_X(x), \ t \in \mathbb{R}.$$

- 当矩母函数存在时, 它唯一确定 X 的分布
- 通过 $g_X(t)$, 可以计算 X 的各阶矩

$$\mathbb{E}X^n = \int x^n dF_X(x) = \int x^n e^{tx} dF_X(x) \Big|_{t=0} = g_X^{(n)}(0), \ n \geq 1,$$

其中

$$g_X^{(n)}(t) = \frac{d^n}{dt^n} \int e^{tx} dF_X(x) = \int x^n e^{tx} dF_X(x)$$

- 注意, 交换求导和广义积分需要条件, 被积函数一致收敛是一个充分条件

矩母函数 (续)

- 对于相互独立的随机变量 X 与 Y, 有

$$g_{X+Y}(t) = g_X(t)g_Y(t)$$

证明如下:

$$g_{X+Y}(t) = \mathbb{E}e^{t(X+Y)} = \mathbb{E}e^{tX}e^{tY} = \mathbb{E}e^{tX}\mathbb{E}e^{tY} = g_X(t)g_Y(t)$$

- 常见分布的矩母函数 (见教材后附表)

分布	矩生成函数 $M_X(t)$	特征函数 $\varphi(t)$
退化 δ_a	e^{ta}	e^{ita}
伯努利 $P(X=1)=p$	$1-p+pe^t$	$1-p+pe^{it}$
几何 $(1-p)^{k-1}p$	$\frac{pe^t}{1-(1-p)e^t}$ $\forall t < -\ln(1-p)$	$\frac{pe^{it}}{1-(1-p)e^{it}}$
二项式 $B(n,p)$	$(1-p+pe^t)^n$	$(1-p+pe^{it})^n$
负二项 $\mathrm{NB}(r,p)$	$\left(\frac{p}{1-e^t+pe^t}\right)^r, t < -\log(1-p)$	$\left(\frac{p}{1-e^{it}+pe^{it}}\right)^r$
泊松 $\mathrm{Pois}(\lambda)$	$e^{\lambda(e^t-1)}$	$e^{\lambda(e^{it}-1)}$
均匀(连续型) $\mathrm{U}(a,b)$	$\frac{e^{tb}-e^{ta}}{t(b-a)}$	$\frac{e^{itb}-e^{ita}}{it(b-a)}$
均匀(离散型) $\mathrm{DU}(a,b)$	$\frac{e^{at}-e^{(b+1)t}}{(b-a+1)(1-e^t)}$	$\frac{e^{ait}-e^{(b+1)it}}{(b-a+1)(1-e^{it})}$
拉普拉斯 $L(\mu,b)$	$\frac{e^{t\mu}}{1-b^2t^2}, \ \lvert t\rvert < 1/b$	$\frac{e^{it\mu}}{1+b^2t^2}$
正态 $N(\mu,\sigma^2)$	$e^{t\mu+\frac{1}{2}\sigma^2t^2}$	$e^{it\mu-\frac{1}{2}\sigma^2t^2}$
卡方(Chi-squared) χ_k^2	$(1-2t)^{-\frac{k}{2}}$	$(1-2it)^{-\frac{k}{2}}$
Noncentral chi-squared $\chi_k^2(\lambda)$	$e^{\lambda t/(1-2t)}(1-2t)^{-\frac{k}{2}}$	$e^{i\lambda t/(1-2it)}(1-2it)^{-\frac{k}{2}}$
伽玛(Gamma) $\Gamma(k,\theta)$	$(1-t\theta)^{-k}, \ \forall t < \frac{1}{\theta}$	$(1-it\theta)^{-k}$
指数(Exponential) $\mathrm{Exp}(\lambda)$	$\left(1-t\lambda^{-1}\right)^{-1}, \ t < \lambda$	$\left(1-it\lambda^{-1}\right)^{-1}$
多元正态 $N(\mu,\Sigma)$	$e^{t^T\left(\mu+\frac{1}{2}\Sigma t\right)}$	$e^{t^T\left(i\mu-\frac{1}{2}\Sigma t\right)}$
柯西(Cauchy) $\mathrm{Cauchy}(\mu,\theta)$	不存在	$e^{it\mu-\theta\lvert t\rvert}$
Multivariate Cauchy $\mathrm{MultiCauchy}(\mu,\Sigma)$[1]	不存在	$e^{it^T\mu-\sqrt{t^T\Sigma t}}$

例 1.12 随机和的矩母函数

- 设 $X_1, X_2, \cdots$ 为一系列独立同分布的随机变量
- N 为非负整数值随机变量, 且与 $\{X_i\}$ 序列独立
- $Y = \sum_{i=1}^{N} X_i$, 称为随机和
- 求 Y 的矩母函数与 X_i 矩母函数的关系

解:

- 在给定 $N = n$ 的条件下, 利用独立随机变量矩母函数性质, 得

$$\begin{aligned}
\mathbb{E}(e^{tY}|N=n) &= \mathbb{E}(e^{t(\sum_{i=1}^{N} X_i)}|N=n) \\
&= \mathbb{E}(e^{t(\sum_{i=1}^{n} X_i)}|N=n) \\
&= \mathbb{E}(e^{t(\sum_{i=1}^{n} X_i)}) \qquad (N \text{ 与 } \{X_i\} \text{ 序列独立}) \\
&= \mathbb{E}\prod_{i=1}^{n} e^{tX_i} = \prod_{i=1}^{n} \mathbb{E}e^{tX_i} = [g_{X_1}(t)]^n \quad (X_i \text{ 与 } X_1 \text{ 同分布})
\end{aligned}$$

- 再利用两步走求期望

$$g_Y(t) = \mathbb{E}e^{tY} = \mathbb{E}[\mathbb{E}(e^{tY}|N)] = \mathbb{E}[g_{X_1}(t)]^N$$

矩母函数 (续)

- 利用随机和 Y 的矩母函数 $g_Y(t)=\mathbb{E}[g_{X_1}(t)]^N$ 及 $\mathbb{E}Y^n=g_Y^{(n)}(0)$, 可求出 Y 的各阶矩与 X_1 及 N 的关系
- 以均值、方差为例, 我们先求 $g_Y'(t)$ 和 $\mathbb{E}Y$

$$g_Y'(t)=\mathbb{E}\{N[g_{X_1}(t)]^{N-1}g_{X_1}'(t)\}$$ (为什么? 对条件期望求导)

- 注意到 $g_{X_1}(0)=\int e^{0\cdot x}dF_{X_1}(x)=1$, 得

$$\mathbb{E}Y = g_Y'(t)\big|_{t=0}=\mathbb{E}\{N[g_{X_1}(0)]^{N-1}g_{X_1}'(0)\}=\mathbb{E}[N\mathbb{E}X_1]=\mathbb{E}N\cdot\mathbb{E}X_1$$

- 类似地, 利用 $g_{X_1}(0)=1$, $g_{X_1}'(0)=\mathbb{E}X_1$ 和 $g_{X_1}''(0)=\mathbb{E}X_1^2$, 有

$$g_Y''(t)=\mathbb{E}\{N(N-1)[g_{X_1}(t)]^{N-2}[g_{X_1}'(t)]^2+N[g_{X_1}(t)]^{N-1}g_{X_1}''(t)\}$$

$$\begin{aligned}\mathbb{E}Y^2 &= g_Y''(t)\big|_{t=0}=\mathbb{E}N(N-1)\cdot(\mathbb{E}X_1)^2+\mathbb{E}N\cdot\mathbb{E}X_1^2\\ &= \mathbb{E}N[\mathbb{E}X_1^2-(\mathbb{E}X_1)^2]+\mathbb{E}N^2\cdot\mathbb{E}X_1^2=\mathbb{E}N\cdot\mathrm{Var}X_1+\mathbb{E}N^2\cdot(\mathbb{E}X_1)^2\end{aligned}$$

$$\begin{aligned}\mathrm{Var}Y &= \mathbb{E}Y^2-(\mathbb{E}Y)^2=\mathbb{E}N\cdot\mathrm{Var}X_1+\mathbb{E}N^2\cdot(\mathbb{E}X_1)^2-(\mathbb{E}N\cdot\mathbb{E}X_1)^2\\ &= \mathbb{E}N\cdot\mathrm{Var}X_1+(\mathbb{E}X_1)^2\cdot\mathrm{Var}N\end{aligned}$$

定义 1.6 概率母函数 (也称概率生成函数)

若 X 为取值在非负整数域上的离散型随机变量, 则定义 $\mathbb{E}s^X$ 为其概率母函数, 记作 $\phi_X(s)$. 若 $\mathbb{P}(X=k)=p_k$, $k=0,1,2,\cdots$, 则 $\phi_X(s)=\mathbb{E}s^X=\sum_{k=0}^{\infty}p_k s^k$, $s\in\mathbb{C}$.

- 概率母函数是以 p_k 为系数的幂级数, 且有

$$p_0=\phi_X(0),\qquad p_k=\frac{1}{k!}\frac{d^k}{ds^k}\phi_X(s)\bigg|_{s=0},\ k=1,2,\cdots$$

- $\phi_X(s)$ (对 $|s|\le 1$ 必存在) 和 X 的分布是一一对应的
- 若 X 与 Y 相互独立, 则 $\phi_{X+Y}(s)=\phi_X(s)\phi_Y(s)$
- 高阶矩满足: $\mathbb{E}[X(X+1)\cdots(X-r+1)]=\frac{d^r}{ds^r}\phi_X(s)|_{s=1}$
- 对随机和 $Y=\sum_{i=1}^{N}X_i$, 其中 X_i 同分布, Y 的概率母函数是随机变量 N 与 X_i 母函数的复合

$$\phi_Y(s)=\mathbb{E}\left[\mathbb{E}(s^{\sum_{i=1}^{N}X_i}|N)\right]=\mathbb{E}\left[\mathbb{E}(\prod_{i=1}^{N}s^{X_i}|N)\right]=\mathbb{E}[\phi_{X_1}(s)]^N=\phi_N\left(\phi_{X_1}(s)\right)$$

特征函数

定义 特征函数

随机变量 X 的特征函数定义为 $\mathbb{E}e^{jtX}$ ，记作

$$\varphi_X(t) = \mathbb{E}e^{jtX} = \int e^{jtx} dF_X(x),\ t \in \mathbb{R}$$

- 其中 $j = \sqrt{-1}$, 是单位虚数
- 类似矩母函数, 特征函数和概率分布也是一一对应的
- 特征函数与矩母函数的关系: $\varphi_X(t) = g_X(jt)$
- 在后续的功率谱密度计算和 Fourier 变换中会经常用到
- 求特征函数的过程, 实际上是对复值随机变量求期望
- 实值随机变量特征函数的性质: 有限性(复可积)、共轭对称性

$$|\varphi_X(t)| = \left|\int_{-\infty}^{\infty} e^{jtx} f_X(x) dx\right| \le \int_{-\infty}^{\infty} |e^{jtx}| f_X(x) dx \le \int_{-\infty}^{\infty} e^{0} f_X(x) dx = |\varphi_X(0)| = 1$$

$$\varphi_X(-t) = \mathbb{E}e^{-jtX} = \overline{\mathbb{E}e^{jtX}} = \overline{\varphi_X(t)}$$

思考与讨论

- 矩母函数、概率母函数、特征函数有哪些区别？(随机变量、自变量、收敛性)

$$g_X(t) = \mathbb{E}e^{tX} = \int e^{tx} dF_X(x)$$

$$\phi_X(s) = \mathbb{E}s^X = \sum_{k=0}^{\infty} p_k s^k$$

$$\varphi_X(t) = \mathbb{E}e^{jtX} = \int e^{jtx} dF_X(x)$$

定义 1.7 依概率收敛

设 $\{X_n, n \geq 1\}$ 为一随机变量序列, 若存在随机变量 X, 使得对 $\forall \varepsilon > 0$, 有

$$\lim_{n\to\infty} \mathbb{P}\left(\omega : |X_n(\omega) - X(\omega)| \geq \varepsilon\right) = 0,$$

则称序列 $\{X_n, n \geq 1\}$ 依概率收敛于 X, 记为 $X_n \xrightarrow{p} X$.

- 简记为 $\lim\limits_{n\to\infty} \mathbb{P}\left(|X_n - X| \geq \varepsilon\right) = 0$, 其中 $|X_n - X|$ 是一个随机变量
- 使 $X_n(\omega)$ 与 $X(\omega)$ 差值超过 ε 的所有基本事件 ω 的概率和趋于 0 (概率的极限)

定义 几乎必然收敛 (也称 以概率 1 收敛 或 几乎处处收敛)

若

$$\mathbb{P}\left(\omega : \lim_{n\to\infty} [X_n(\omega) - X(\omega)] = 0\right) = 1,$$

则称序列 $\{X_n, n \geq 1\}$ 几乎必然 (almost surely) 收敛于 X, 记为 $X_n \xrightarrow{a.s.} X$.

- $X_n(\omega)$ 的极限收敛到 $X(\omega)$ 的所有基本事件 ω 的概率和为 1 (极限的概率)
- 也即几乎每个样本轨道都收敛

随机变量序列的收敛性 (续)

以概率 1 收敛 $\Rightarrow$ 依概率收敛 (证明见下页), 反之则未必

- 反例: 考虑独立随机变量序列 $\{X_n\}$, 满足

$$\mathbb{P}(X_n = 1) = \frac{1}{n}, \quad \mathbb{P}(X_n = 0) = 1 - \frac{1}{n}$$

- 令 $X = 0$, 对 $0 < \epsilon < 1$, 有

$$\mathbb{P}(|X_n - X| \geq \varepsilon) = \frac{1}{n}$$

- 故而 $\lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = 0$, 也即 $X_n \xrightarrow{p} X$ 成立
- 由于 $\sum_{n=1}^{\infty} \mathbb{P}(X_n = 1) = +\infty$ 且事件 $\{X_n = 1\}$ 相互独立, 根据 Borel-Cantelli 第二引理 (点击参考), 有

$$\mathbb{P}\left(\bigcap_{N=1}^{\infty} \bigcup_{m=N}^{\infty} \{X_m = 1\}\right) = \mathbb{P}\left(\lim_{N\to\infty} \bigcup_{m=N}^{\infty} \{X_m = 1\}\right) = 1 > 0$$

- 即对任意 N 和几乎所有样本轨道, 总存在 $n > N$, 使得 $X_n(\omega) = 1$
- 而 $\mathbb{P}\left(\omega : \lim_{n\to\infty} X_n(\omega) = 0\right) = 1$ 要求, 对任意 $\varepsilon > 0$, 存在 N, 当 $n > N$ 时, 在几乎所有样本轨道下, $|X_n(\omega) - 0| \leq \varepsilon$, 故 $X_n \xrightarrow{a.s.} X$ 不成立

证明: 以概率 1 收敛 $\Rightarrow$ 依概率收敛

- 以概率 1 收敛: $\mathbb{P}\left(\lim\limits_{n\to\infty}[X_n - X] = 0\right) = 1$
- 也即, 事件 $\left\{\omega : \lim\limits_{n\to\infty}[X_n(\omega) - X(\omega)] = 0\right\}$ 的概率为 1
  - ▷ 其中 $\lim\limits_{n\to\infty}[X_n - X]$ 是一个随机变量 (随机变量序列 $[X_n - X]$ 的极限), 其等于 0 的概率为 1
- 样本 ω 满足事件 $\left\{\omega : \lim\limits_{v\to\infty}[X_v(\omega) - X(\omega)] = 0\right\}$, 等价于 ω 满足
  - ▷ 对任意 $\varepsilon_k > 0$ (即 $\bigcap_{k=1}^{\infty}$)
  - ▷ 存在(依赖于 ε_k 的) n (即 $\bigcup_{n=1}^{\infty}$)
  - ▷ 使对任意$v \geq n$ (即 $\bigcap_{v=n}^{\infty}$)
  - ▷ 有 $|X_v(\omega) - X(\omega)| < \varepsilon_k$
- 也即, 事件 $\left\{\omega : \lim\limits_{n\to\infty}[X_n(\omega) - X(\omega)] = 0\right\}$ 等价于, 对 $\varepsilon_k \downarrow 0$, 有

$$\bigcap_{k=1}^{\infty}\bigcup_{n=1}^{\infty}\bigcap_{v=n}^{\infty}\{\omega : |X_v(\omega) - X(\omega)| < \varepsilon_k\}$$

- 基于上述表示, 以概率 1 收敛等价于

$$
\begin{aligned}
& X_n \xrightarrow{a.s.} X \\
\Leftrightarrow\quad & \mathbb{P}\left(\bigcap_{k=1}^{\infty}\bigcup_{n=1}^{\infty}\bigcap_{v=n}^{\infty}\{\omega : |X_v(\omega) - X(\omega)| < \varepsilon_k\}\right) = 1 \\
\Leftrightarrow\quad & \mathbb{P}\left(\bigcup_{k=1}^{\infty}\bigcap_{n=1}^{\infty}\bigcup_{v=n}^{\infty}\{\omega : |X_v(\omega) - X(\omega)| \geq \varepsilon_k\}\right) = 0
\end{aligned}
$$

(利用 $\mathbb{P}(A \cup B) = 0 \;\Leftrightarrow \mathbb{P}(A) = 0$ 且 $\mathbb{P}(B) = 0$)

$$
\Leftrightarrow\quad \mathbb{P}\left(\bigcap_{n=1}^{\infty}\bigcup_{v=n}^{\infty}\{\omega : |X_v(\omega) - X(\omega)| \geq \varepsilon_k\}\right) = 0, \;\; \forall k
$$

(利用递减事件序列 A_n 的极限满足 $\lim\limits_{n\to\infty} A_n = \bigcap\limits_{i=1}^{\infty} A_i$

及 $\mathbb{P}\left(\lim\limits_{n\to\infty} A_n\right) = \lim\limits_{n\to\infty}\mathbb{P}(A_n)$, 证明见下页备注)

$$
\Leftrightarrow\quad \lim_{n\to\infty}\mathbb{P}\left(\bigcup_{v=n}^{\infty}\{\omega : |X_v(\omega) - X(\omega)| \geq \varepsilon\}\right) = 0, \;\; \forall \varepsilon > 0
$$

备注

- 递减事件序列 A_n 的概率极限满足 $\mathbb{P}(\bigcap_{n=1}^{\infty} A_n) = \lim_{n\to\infty} \mathbb{P}(A_n)$，证明如下

$$
\begin{aligned}
& \mathbb{P}(\bigcap_{n=1}^{\infty} A_n) = \mathbb{P}(\overline{\bigcup_{n=1}^{\infty} \overline{A_n}}) = 1 - \mathbb{P}(\bigcup_{n=1}^{\infty} \overline{A_n}) \quad \text{(德·摩根定律)} \\
= \quad & 1 - \mathbb{P}\left(\bigcup_{n=1}^{\infty}\left(\overline{A_n}(\bigcap_{i=1}^{n-1} A_i)\right)\right) \quad \text{(所有事件中有至少一个不发生的等价表示)} \\
= \quad & 1 - \sum_{n=1}^{\infty} \mathbb{P}\left(\overline{A_n}(\bigcap_{i=1}^{n-1} A_i)\right) \quad \text{(互斥事件概率和)} \\
= \quad & 1 - \lim_{k\to\infty} \sum_{n=1}^{k} \mathbb{P}\left(\overline{A_n}(\bigcap_{i=1}^{n-1} A_i)\right) = 1 - \lim_{k\to\infty} \mathbb{P}\left(\bigcup_{n=1}^{k}\left(\overline{A_n}(\bigcap_{i=1}^{n-1} A_i)\right)\right) \\
= \quad & 1 - \lim_{k\to\infty} \mathbb{P}(\bigcup_{n=1}^{k} \overline{A_n}) = \lim_{k\to\infty} \mathbb{P}(\bigcap_{n=1}^{k} A_k) = \lim_{n\to\infty} \mathbb{P}(A_n) \quad \text{(德·摩根定律及事件递减性)}
\end{aligned}
$$

$$X_n \xrightarrow{a.s.} X \Leftrightarrow \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{v=n}^{\infty} \{\omega : |X_v(\omega) - X(\omega)| \geq \varepsilon\}\right) = 0, \ \forall \varepsilon > 0$$

- 由于依概率收敛 $\lim\limits_{n\to\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = 0$ 中, 事件

$$\{|X_n - X| \geq \varepsilon\} \subset \left\{\bigcup_{v=n}^{\infty} \{|X_v - X| \geq \varepsilon\}\right\}$$

- 故

$$\begin{aligned}
& X_n \xrightarrow{a.s.} X \\
\Rightarrow\ & \lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) \\
& \leq \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{v=n}^{\infty} \{\omega : |X_v(\omega) - X(\omega)| \geq \varepsilon\}\right) = 0, \ \forall \varepsilon > 0 \\
\Rightarrow\ & \lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = 0, \ \forall \varepsilon > 0 \\
\Leftrightarrow\ & X_n \xrightarrow{p} X
\end{aligned}$$

例 1.13

- Bernoulli 试验中, 每次成功概率为 p
- S_n 表示 n 次试验中成功的次数
- 求证 $\frac{S_n}{n} \xrightarrow{p} p$
  - ▷ 此时随机变量 $\frac{S_n}{n}$ 收敛于确定的值, 是收敛于随机变量的特例

解: 利用 Markov 不等式

- 对于非负随机变量,

$$\begin{aligned}
\mathbb{E}X &= \int_0^{+\infty} x d\mathbb{P}(X \geq x) = \int_0^{a} x d\mathbb{P}(X \geq x) + \int_a^{+\infty} x d\mathbb{P}(X \geq x) \\
&\geq \int_a^{+\infty} x d\mathbb{P}(X \geq x) \geq \int_a^{+\infty} a d\mathbb{P}(X \geq x) \\
&= a\mathbb{P}(X \geq a)
\end{aligned}$$

得到 Markov 不等式:

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}X}{a}$$

例 1.13 (续)

- 从 $\frac{S_n}{n} \xrightarrow{p} p$ 定义出发, 需证 $\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \geq \varepsilon\right) = 0, \ \forall \varepsilon > 0$
- 由于 $S_n \sim B(n,p)$, 有 $\mathbb{E}S_n = np$ 和 $\mathrm{Var}S_n = np(1-p)$, 故对 $\forall \varepsilon > 0$

$$
\begin{aligned}
& \mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \geq \varepsilon\right) = \mathbb{P}\left(|S_n - np| \geq n\varepsilon\right) \\
= \ & \mathbb{P}\left((S_n - np)^2 \geq (n\varepsilon)^2\right) \qquad \text{(等价事件替换)} \\
\leq \ & \frac{\mathbb{E}(S_n - np)^2}{(n\varepsilon)^2} = \frac{np(1-p)}{(n\varepsilon)^2} \qquad \left(\text{Markov 不等式 } \mathbb{P}(X \geq a) \leq \frac{\mathbb{E}X}{a}\right) \\
= \ & \frac{p(1-p)}{n\varepsilon^2} \to 0, \ \text{as } n \to \infty
\end{aligned}
$$

得证 $\frac{S_n}{n} \xrightarrow{p} p$

- 成功的**频率** $\frac{S_n}{n}$(为随机变量), 趋近于成功的**概率** p(为确定值)
  - ▷ 大数定律的一个应用

定义 1.8 均方收敛

设随机变量 X 和随机序列 $\{X_n,\ n \geq 1\}$, 都有有限二阶矩, 若

$$\lim_{n\to\infty} \mathbb{E}(X_n - X)^2 = 0,$$

则称 X_n 均方收敛于 X, 记作 $X_n \xrightarrow{L_2} X$

- 在例 1.13 中, 由于 $\mathbb{E}(\frac{S_n}{n} - p)^2 = \frac{\mathrm{Var} S_n}{n^2} = \frac{p(1-p)}{n} \xrightarrow{n\to\infty} 0$, 故 $\frac{S_n}{n} \xrightarrow{L_2} p$
- 均方收敛 $\Rightarrow$ 依概率收敛:
  - 对 $\forall \varepsilon > 0$, 由 Markov 不等式有
  $$\mathbb{P}\left(|X_n - X| \geq \varepsilon\right) = \mathbb{P}\left(|X_n - X|^2 \geq \varepsilon^2\right) \leq \frac{\mathbb{E}(X_n - X)^2}{\varepsilon^2}$$
  - 再对两边取 $n \to \infty$, 得
  $$\lim_{n\to\infty} \mathbb{P}\left(|X_n - X| \geq \varepsilon\right) \leq 0$$
  - 由概率非负性, 得 $\lim\limits_{n\to\infty} \mathbb{P}\left(|X_n - X| \geq \varepsilon\right) = 0$, 即依概率收敛

定义 依分布收敛

对随机变量 X 和随机序列 $\{X_n,\ n \geq 1\}$, 若对所有的 x, 有

$$\lim_{n\to\infty} \mathbb{P}(X_n \leq x) = \mathbb{P}(X \leq x)$$

则称 X_n 依分布收敛于 X, 记作 $X_n \xrightarrow{d} X$

- 另一种等价表示为: $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$
- 依概率收敛 $\Rightarrow$ 依分布收敛, 证明如下: 对任意的 $\varepsilon > 0$,

$$\begin{aligned}
\mathbb{P}(X_n \leq x) &\geq \mathbb{P}(X_n \leq x,\ X \leq x-\varepsilon) \\
&= \mathbb{P}(X \leq x-\varepsilon) - \mathbb{P}(X_n > x,\ X \leq x-\varepsilon) \\
&\geq \mathbb{P}(X \leq x-\varepsilon) - \mathbb{P}(|X_n - X| > \varepsilon)
\end{aligned}$$

- 两边取 $n \to \infty$, 利用依概率收敛定义 $\lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = 0$, 有

$$\lim_{n\to\infty} \mathbb{P}(X_n \leq x) \geq \mathbb{P}(X \leq x - \varepsilon)$$

$\lim_{n\to\infty} \mathbb{P}(X_n \le x) \ge \mathbb{P}(X \le x - \varepsilon)$

- 再令 $\varepsilon \to 0^+$, 得 $\lim\limits_{n\to\infty} \mathbb{P}(X_n \le x) \ge \mathbb{P}(X \le x)$
- 下面证 $\lim\limits_{n\to\infty} \mathbb{P}(X_n \le x) \le \mathbb{P}(X \le x)$

$$
\begin{aligned}
\mathbb{P}(X \le x + \varepsilon) &\ge \mathbb{P}(X \le x + \varepsilon, X_n \le x) \\
&= \mathbb{P}(X_n \le x) - \mathbb{P}(X_n \le x, X > x + \varepsilon) \\
&\ge \mathbb{P}(X_n \le x) - \mathbb{P}(|X_n - X| > \varepsilon)
\end{aligned}
$$

- 同样, 两边取 $n \to \infty$, 再取 $\varepsilon \to 0^+$, 得

$$\lim_{n\to\infty} \mathbb{P}(X_n \le x) \le \mathbb{P}(X \le x)$$

- 综上, 有 $\lim\limits_{n\to\infty} \mathbb{P}(X_n \le x) = \mathbb{P}(X \le x)$, 依分布收敛得证
- 几种收敛性的关系

均方收敛
(互不包含) $\Rightarrow$ 依概率收敛 $\Rightarrow$ 依分布收敛
几乎必然收敛

思考与讨论

- 极限运算在概率函数内和函数外时，其区别是什么?
- 试简述四种收敛性之间的关系

$$\lim_{n\to\infty} \mathbb{P}\left(|X_n - X| \geq \varepsilon\right) = 0$$

$$\mathbb{P}\left(\lim_{n\to\infty} X_n = X\right) = 1$$

$$\lim_{n\to\infty} \mathbb{E}(X_n - X)^2 = 0$$

$$\lim_{n\to\infty} \mathbb{P}(X_n \leq x) = \mathbb{P}(X \leq x)$$

第 1 章 习题

习题 1:

- 1.1 节作业: 1, 2, 8
- 1.2 节作业: 9, 10, 17