



概率论笔记

作者：肖程哲

时间：July 31, 2022



苟日新，日日新，又日新

目录

第 1 章 概率基础	1	3.4.2 预测	23
1.1 概率空间	1	3.5 熵与信息	23
1.2 古典概型与几何概型	3		
1.2.1 古典概型	3	第 4 章 常见分布	24
1.2.2 几何概型	4	4.1 离散分布	24
1.2.2.1 Buffon 投针	4	4.1.1 均匀分布	24
1.2.2.2 Bertrand 奇论	4	4.1.2 伯努利分布	24
1.3 条件概率	4	4.1.3 二项分布	25
		4.1.4 几何分布	26
第 2 章 随机变量	6	4.1.5 负二项分布	27
2.1 随机变量的分布	6	4.1.6 多项分布	28
2.2 多元随机变量	8	4.1.7 泊松分布	29
2.2.1 边际分布	9	4.1.8 超几何分布	31
2.2.2 条件分布	9	4.2 连续分布	32
2.2.3 独立	10	4.2.1 均匀分布	32
2.3 随机变量的函数	10	4.2.2 指数分布	32
2.3.1 分布函数法	11	4.2.3 伽马分布	34
2.3.2 Copula	12	4.2.4 贝塔分布	35
2.3.3 概率密度函数法	13	4.3 正态分布及其导出分布	37
2.3.4 矩母函数法	14	4.3.1 正态分布	37
2.3.5 次序统计量	14	4.3.2 卡方分布	38
		4.3.3 F 分布	39
第 3 章 随机变量的数值特征	16	4.3.4 t 分布	40
3.1 期望	16	4.3.5 柯西分布	41
3.1.1 均值	16	4.4 各分布间关系	43
3.1.2 方差	17		
3.1.3 协方差	18		
3.2 条件期望	19		
3.3 矩母函数与特征函数	20		
3.3.1 矩	20	第 5 章 概率极限	45
3.3.2 矩母函数	21	5.1 收敛	45
3.3.3 联合特征函数	22	5.2 大数定理	47
3.3.4 特征函数	22	5.3 中央极限定理	48
3.4 估计与预测	23		
3.4.1 δ 法	23		
第 A 章 基本数学工具	45		
A.1 集合论	45		
A.2 测度论	45		
A.3 排列与组合	45		

第5章 概率极限

5.1 收敛

定义 5.1

若随机变量序列 $\{X_n : \Omega \rightarrow \mathbb{R}\}$ 与随机变量 $X : \Omega \rightarrow \mathbb{R}$ 间存在以下关系：

$$\mathbb{P}\left\{\omega \in \Omega : \lim_{n \rightarrow \infty} |X_n(\omega) - X(\omega)| < \epsilon\right\} = 1, \forall \epsilon > 0$$

则称 X_n 几乎必然收敛 (converges almost surely) 于 X , 记为 $X_n \xrightarrow{\text{a.s.}} X$ 。



注 类似于微积分中逐点收敛的条件。考察满足条件 $\lim_{n \rightarrow \infty} |X_n(\omega) - X(\omega)| < \epsilon$ 的样本点，这些样本点组成事件的概率为 1。

定义 5.2

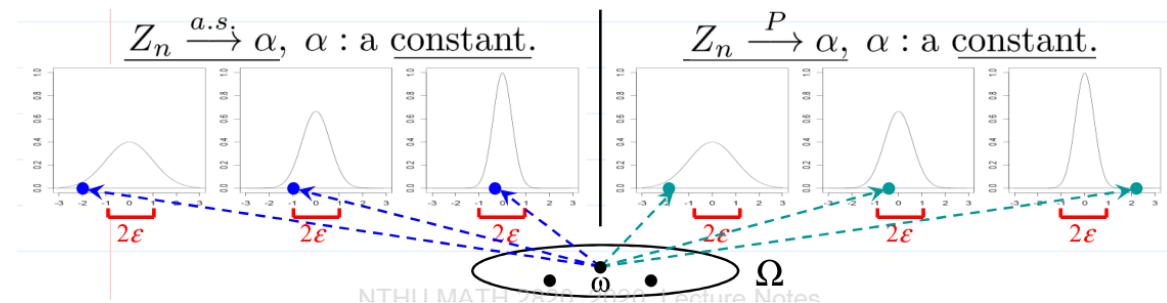
若随机变量序列 $\{X_n : \Omega \rightarrow \mathbb{R}\}$ 与随机变量 $X : \Omega \rightarrow \mathbb{R}$ 间存在以下关系：

$$\lim_{n \rightarrow \infty} \mathbb{P}\{\omega \in \Omega : |X_n(\omega) - X(\omega)| < \epsilon\} = 1, \forall \epsilon > 0.$$

则称 X_n 依概率收敛 (converges in probability) 于 X , 记为 $X_n \xrightarrow{\mathbb{P}} X$ 。



注 即对于 Z_i , 事件 $A_i = \{\omega \in \Omega : |X_n(\omega) - X(\omega)| < \epsilon\}$, $P(A_i)$ 将逐渐变为 1。



注 对于几乎必然收敛的情况，每个样本点上的序列随机变量都在逐渐趋近收敛变量；而对于依概率收敛的情况，由于对某个样本点的概率为 0 (例如连续随机变量)，可能某一次样本点的序列随机变量接近收敛变量，下一次又离开收敛变量，只有保证总体概率趋近 1 即可。

易见 $X_n \xrightarrow{\text{a.s.}} X_\infty \implies X_n \xrightarrow{\mathbb{P}} X_\infty$, 但是反之不成立, 例如概率空间 $((0, 1], \mathcal{B}_{(0,1]}, \text{Uniform}((0, 1]))$ 上的打字机序列

$$\xi_n = \mathbb{1}_{(n/2^k - 1, (n+1)/2^k - 1]}, \quad 2^k \leq n < 2^{k+1}$$

适合 $\mathbb{P}\{|\xi_n| > \epsilon\} \leq 1/2^k \rightarrow 0$, 而 $\overline{\lim} \xi_n \equiv 1$ 且 $\underline{\lim} \xi_n \equiv 0$, 所以 $\lim \xi_n(\omega)$ 对 $\forall \omega \in (0, 1]$ 都不存在.

命题 5.1

依概率收敛是几乎必然收敛的必要条件, 但不是充分条件。



例题 5.1 依概率收敛而不几乎必然收敛的实例：设样本空间为 $\Omega = (0, 1]$, 概率均匀分布在样本空间上, 即 $P([a, b]) = b - a, 0 < a < b < 1$ 。接下来对于 $k \in \mathbb{N}_+$, 将区间 $(0, 1]$ 分为 2^k 个等长的子区间，并分别记为 $I_{k,j} = (\frac{j-1}{2^k}, \frac{j}{2^k}]$ 。

按以下方式定义随机变量序列 $\{X_n : \Omega \rightarrow \mathbb{R}\}$:

$$X_n(\omega) = \begin{cases} 1, \omega \in I_{k,j} \\ 0, \omega \notin I_{k,j} \end{cases}, \quad n = 2^k + j - 2$$

同时定义随机变量 $X : \Omega \rightarrow \mathbb{R}$ 为 $X(\omega) = 0, \forall \omega \in \Omega$

由于对任意 $0 < \epsilon < 1$

$$\mathbb{P}\{\omega \in \Omega : |X_n(\omega) - X(\omega)| < \epsilon\} = 1 - \frac{1}{2^K} \rightarrow 1$$

所以 $X_n \xrightarrow{\mathbb{P}} X$ 。然而对于任意 $\omega \in \Omega$, 无论 k 为何数, 此次分割中, 总有一个区间包含此样本点。即有无数个 $X_i(\omega) = 1$ 。所以

$$\mathbb{P}\left\{\omega \in \Omega : \lim_{n \rightarrow \infty} |X_n(\omega) - X(\omega)| < \epsilon\right\} = \mathbb{P}\{\emptyset\} = 0$$

所以 X_n 不几乎必然收敛到 X 。

定义 5.3

设随机变量 X, X_1, X_2, \dots 的分布函数分别为 $F(x), F_1(x), F_2(x), \dots$ 。若对 $F(x)$ 的任一连续点 x , 都有

$$\lim_{n \rightarrow +\infty} F_n(x) = F(x)$$

则称 $\{F_n(x)\}$ 弱收敛于 $F(x)$, 记作 $F_n(x) \xrightarrow{W} F(x)$; 也称 $\{X_n\}$ 按分布收敛于 X , 记作 $X_n \xrightarrow{d} X$

定理 5.1

依分布收敛是依概率收敛的必要条件, 但不是充分条件。但若依概率收敛到一个常数随机变量, 则也可推出依分布收敛。

定理 5.2 (连续映射定理)

设 $g : \mathbb{R}^k \mapsto \mathbb{R}$ 为连续函数, 则随机向量序列通过次映射后收敛情况与原先相同:

$$\begin{aligned} X_n^{(i)} &\xrightarrow{\text{a.s.}} X^{(i)} \Rightarrow g(X_n^{(1)}, \dots, X_n^{(k)}) \xrightarrow{\text{a.s.}} g(X^{(1)}, \dots, X^{(k)}) \\ X_n^{(i)} &\xrightarrow{\mathbb{P}} X^{(i)} \Rightarrow g(X_n^{(1)}, \dots, X_n^{(k)}) \xrightarrow{\mathbb{P}} g(X^{(1)}, \dots, X^{(k)}) \\ (X_n^{(i)}) &\xrightarrow{d} (X^{(i)}) \Rightarrow g(X_n^{(1)}, \dots, X_n^{(k)}) \xrightarrow{d} g(X^{(1)}, \dots, X^{(k)}) \end{aligned}$$

注 随机向量依分布收敛是指联合分布函数收敛。

定理 5.3 (Slutsky 定理)

若 $X_n \xrightarrow{d} X, Y_n \xrightarrow{P} a$, 其中 a 是常数, 则:

- $Y_n X_n \xrightarrow{d} aX$
- $Y_n + X_n \xrightarrow{d} a + X$
- $\frac{X_n}{Y_n} \xrightarrow{d} \frac{X}{a}$, 若对所有的 n 都有 $P(Y_n \neq 0) = 1$, 并且 $a \neq 0$

证明 由条件可知, $(X_n, Y_n) \xrightarrow{d} (X, a)$, 且 $g_1(X_n, Y_n) = X_n Y_n, g_2(X_n, Y_n) = X_n + Y_n, g_3(X_n, Y_n) = \frac{X_n}{Y_n}$ 都是连续函数 (g_3 在定理中的限制下), 所以仍然保持依分布收敛。

注 当 $Y_n \xrightarrow{P} a$ 才能得出 $(X_n, Y_n) \xrightarrow{d} (X, a)$; 若 Y_n 依分布收敛到其他随机变量, 则未必成立。

定理 5.4 (theta 方法的极限定理)

设

$$\frac{\sqrt{n}(X_n - \theta)}{\sigma} \xrightarrow{d} N(0, 1)$$

那么对于连续函数 g , 若 $g'(\theta) \neq 0$ 存在, 则

$$\frac{\sqrt{n}(g(X_n) - g(\theta))}{\sigma |g'(\theta)|} \xrightarrow{d} N(0, 1)$$

**证明****定理 5.5 (连续性定理)**设 $F_n(x)$ 是一个累计函数序列, 并且分别对应矩母函数 $M_n(t)$; $F(x)$ 是一个累计函数, 并且对应矩母函数 $M(t)$ 。那么,

$$\forall t \in U(0, \delta), s.t. \lim_{n \rightarrow \infty} M_n(t) = M(t)$$

是在 $F(x)$ 的所有连续点上 $F_n(x) \xrightarrow{M} F(x)$ 的充要条件。**注** 若矩母函数不存在, 替换成特征函数, 定理也成立。但是, 不能替换成密度函数或质量函数**例题 5.2 离散情况:**设均匀随机变量 $X_n \sim U(-\frac{1}{2}, \frac{1}{2})$, 那么 $X_n \xrightarrow{d} 0$ 。然而 $P_n(x) \rightarrow 0, \forall x \in \mathbb{R}$ (不满足质量函数定义), 并且 $\lim_{n \rightarrow \infty} F_n(0) = \frac{1}{2}$ (不满足累积函数定义)**连续情况:**设累积函数为 $F_n(x) = x - \frac{\sin(2n\pi x)}{2n\pi}, 0 < x < 1$, 那么 $F_n \xrightarrow{d} U(0, 1)$ 。然而 $f_n(x)$ 无极限**例题 5.3 泊松分布收敛于正态分布** 令 $X_n \sim P(\lambda_n)$, 并且 $\lim_{n \rightarrow \infty} \lambda_n = \infty$ 。将变量标准化:

$$Z_n = \frac{X_n - \lambda_n}{\sqrt{\lambda_n}}$$

那么

$$M_{Z_n}(t) = e^{-t\sqrt{\lambda_n}} M_{X_n}(t) \left(\frac{t}{\sqrt{\lambda_n}} \right) = \exp(-t\sqrt{\lambda_n} + \lambda_n(e^{\frac{t}{\sqrt{\lambda_n}} - 1}))$$

由于

$$\lim_{n \rightarrow \infty} \ln(M_{Z_n}(t)) = \lim_{\lambda_n \rightarrow \infty} -t\sqrt{\lambda_n} + \lambda_n(e^{\frac{t}{\sqrt{\lambda_n}} - 1}) = \frac{t^2}{2}$$

所以 $Z_n \xrightarrow{d} N(0, 1)$, 即若 λ 很大, 可通过 $N(\lambda, \lambda)$ 估计 $P(\lambda)$

5.2 大数定理

定理 5.6 (弱大数定理)设随机变量 X_1, \dots, X_n 相互独立, 且 $E(X_i) = \mu, \text{Var}(X_i) = \sigma^2$, 那么

$$\overline{X_n} = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\mathbb{P}} \mu$$

**注** 柯西分布无均值、方差, 不能使用此定理。**证明** 易知

$$E(\overline{X_n}) = \mu, \text{Var}(\overline{X_n}) = \frac{\sigma^2}{n}$$

再通过切比雪夫不等式有：

$$P(|\bar{X}_n - \mu| > \epsilon) < \frac{\text{Var}(\bar{X}_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \rightarrow 0$$

定理 5.7 (强大数定理)

设随机变量 X_1, \dots, X_n 相互独立，且 $E(X_i) = \mu, \text{Var}(X_i) = \sigma^2$ ，那么

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\text{a.s.}} \mu$$



例题 5.4 蒙特卡洛积分 为计算积分 $I(f) = \int_0^1 f(x)dx$ ，生成随机数 X_1, \dots, X_n i.i.d. $\sim U(0, 1)$ ，并且计算 $\hat{I}(f) = \bar{Y}_n, Y_i = f(X_i)$ 。由于 X_1, \dots, X_n i.i.d.，所以 Y_1, \dots, Y_n i.i.d.；并且 $E(Y_i) = E(f(X_i)) = \int_0^1 f(x) \cdot 1 dx = I(f)$ ，所以 $\hat{I}(f) \xrightarrow{\mathbb{P}} E(Y_i) = I(f)$

例题 5.5 样本方差 设随机变量 X_1, \dots, X_n i.i.d.，并且均值和方差分别为 μ, σ^2 。令：

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

由于 $g(x) = x^2$ 是连续函数，所以 $\bar{X}_n^2 \xrightarrow{\mathbb{P}} \mu^2$ 。

5.3 中央极限定理