概率论笔记

作者：肖程哲

时间：August 2, 2022

苟日新，日日新，又日新

目录

第 5 章 概率极限

5.1 收敛

定义 5.1

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

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

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 $\lim_{n\to\infty}|X_n(\omega)-X(\omega)|<\epsilon$ 的样本点，这些样本点组成事件的概率为 1。

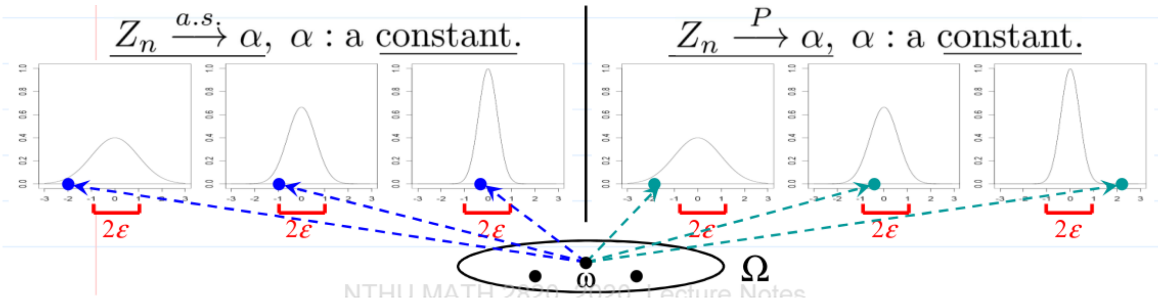
定义 5.2

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

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

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 $A_i=\{\omega\in\Omega:|X_n(\omega)-X(\omega)|<\varepsilon\}$，$P(A_i)$ 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 $((0,1],\mathscr{B}_{(0,1]},\text{Uniform}((0,1]))$ 上的打字机序列

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

适合 $\mathbb{P}\{|\xi_n|>\varepsilon\}\le 1/2^k\to 0$, 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 $\{X_n : \Omega \to \mathbb{R}\}$:

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

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

由于对任意 $0 < \epsilon < 1$

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

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

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

所以 X_n 不几乎必然收敛到 X。

定义 5.3

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

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

则称 $\{F_n(x)\}$ **弱收敛**于 $F(x)$, 记作 $F_n(x) \xrightarrow{W} \to F(x)$；也称 $\{X_n\}$ 按**分布收敛**于 X，记作 $X_n \xrightarrow{d} \to X$

命题 5.2

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 5.1 (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：

$$X_n^{(i)} \xrightarrow{\text{a.s.}} X^{(i)} \Rightarrow g(X_n^{(1)}, \cdots X_n^{(k)}) \xrightarrow{\text{a.s.}} g(X^{(1)}, \cdots X^{(k)})$$

$$X_n^{(i)} \xrightarrow{\mathbb{P}} X^{(i)} \Rightarrow g(X_n^{(1)}, \cdots X_n^{(k)}) \xrightarrow{\mathbb{P}} g(X^{(1)}, \cdots X^{(k)})$$

$$(X_n^{(i)}) \xrightarrow{d} (X^{(i)}) \Rightarrow g(X_n^{(1)}, \cdots X_n^{(k)}) \xrightarrow{d} g(X^{(1)}, \cdots X^{(k)})$$

注 随机向量依分布收敛是指联合 分布函数收敛。

定理 5.2 (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 5.3 (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 5.4 (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，

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

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 $P_n(x) \to 0, \forall x \in \mathbb{R}$（不满足质量函数定义），并且 $\lim_{n\to\infty} F_n(0) = \frac{1}{2}$（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 $\lim_{n\to\infty} \lambda_n = \infty$。将变量标准化：

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

那么

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

由于

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

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 5.5 (弱大数定理)

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

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

再通过切比雪夫不等式有：

$$P(\left|\overline{X_n}-\mu\right|>\epsilon)<\frac{\mathrm{Var}(\overline{X_n})}{\epsilon^2}=\frac{\sigma^2}{n\epsilon^2}\to 0$$

定理 5.6 (强大数定理)

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

$$\overline{X_n}=\frac{1}{n}\sum_{i=1}^{n}X_i\xrightarrow{\text{a.s.}}\mu$$

例题 5.4 蒙特卡洛积分 为计算积分 $I(f)=\int_0^1 f(x)\mathrm{d}x$，生成随机数 $X_1,\cdots,X_n$ i. i. d. $\sim U(0,1)$，并且计算 $\hat{I}(f)=\overline{Y_n},Y_i=f(X_n)$。由于 $X_1,\cdots,X_n$ i. i. d.，所以 $Y_1,\cdots,Y_n$ i. i. d.；并且 $E(Y_i)=E(f(X_i))=\int_0^1 f(x)\cdot 1\mathrm{d}x=I(f)$，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 $X_1,\cdots,X_n$ i. i. d.，并且均值和方差分别为 μ,σ。令：

$$S_n^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X_n})=(\frac{1}{n}\sum_{i=1}^{n}X_i^2)-\overline{X_n}^2+$$

由于 $g(x)=x^2$ 是连续函数，所以 $\overline{X_n}^2\xrightarrow{\mathbb{P}}\mu^2$。由于 $X_1^2,\cdots X_n^2$ i. i. d.，且 $E(X_i^2)=\sigma^2+\mu^2$，所以 $\overline{X_n^2}\xrightarrow{\mathbb{P}}\mu^2+\sigma^2$。因此：

$$S_n^2\xrightarrow{\mathbb{P}}\mu^2+\sigma^2-\mu^2=\sigma^2$$

例题 5.6 若 $X_n\sim t_n$，则 $X_n\xrightarrow{d}N(0,1)$。令 $Z\sim N(0,1),U_1,\cdots,U_n\sim\chi_1^2$ 相互独立，则 $\frac{Z}{\sqrt{(U_1+\cdots+U_n)/n}}\sim t_n$。由于 $E(U_i)=1$，所以 $\overline{U_n}\xrightarrow{\mathbb{P}}1$，所以 $\sqrt{\overline{U_n}}\xrightarrow{\mathbb{P}}1$，所以 $t_n=\frac{Z}{\sqrt{\overline{U_n}}}\xrightarrow{d}N(0,1)$（定理5.2）。

例题 5.7 若 $X_n\sim F_{m,n}$，则 $mX_n\xrightarrow{d}\chi_m^2$。令 $U\sim\chi_m^2,V_1,\cdots,V_n\sim\chi_1^2$ 相互独立，则 $\frac{U/m}{(V_1+\cdots+V_n)/n}\sim F_{m,n}$。由于 $E(V_i)=1$，所以 $\overline{V_n}\xrightarrow{\mathbb{P}}1$，所以 $mF_{m,n}=\frac{U}{\overline{V_n}}\xrightarrow{d}\chi_m^2$（定理5.2）。

5.3 中央极限定理

定理 5.7 (中央极限定理)

设随机变量 $X_1,\cdots,X_n$, i. i. d.，且 $E(X_i)=\mu,\mathrm{Var}(X_i)=\sigma^2$，令

$$\overline{X_n}=\frac{1}{n}\sum_{i=1}^{n}X_i,T_n=\sum_{i=1}^{n}X_i$$

则：

$$\lim_{n\to\infty}P(\frac{T_n-n\mu}{\sigma\sqrt{n}}\le x)=\lim_{n\to\infty}P(\frac{\sqrt{n}(\overline{X_n}-\mu)}{\sigma}\le x)=\Phi(x)$$

其中 $\Phi(x)$ 是 $N(0,1)$ 的累积函数，即：

$$\frac{T_n-n\mu}{\sigma\sqrt{n}}=\frac{\sqrt{n}(\overline{X_n}-\mu)}{\sigma}\xrightarrow{d}N(0,1)$$

证明 令 $W_i = \frac{X_i-\mu}{\sigma}$，则 $E(W_i)=0, \mathrm{Var}(W_i)=1$。再令 $Z_n = \frac{T_n - n\mu}{\sigma\sqrt{n}} = \frac{1}{\sqrt{n}}\sum_{i=1}^n W_i$。则

$$\begin{aligned}M_{Z_n}(t) &= \left[M_W(\frac{t}{\sqrt{n}})\right]^n \\ &= \left[M_W(0) + M_W'(0)\frac{t}{\sqrt{n}} + \frac{M_W''(0)}{2}(\frac{t}{\sqrt{n}})^2 + o(\frac{1}{n})\right]^n \\ &= \left[1 + \frac{t^2}{2n} + o(\frac{1}{n})\right]^n \\ &\to e^{\frac{t^2}{2}}\end{aligned}$$

根据定理5.4，$Z_n \xrightarrow{d} N(0,1)$。

注 若矩母函数不存在，可替换为特征函数。

例题 5.8 二项分布近似为正态分布 若随机变量 $X_1,\cdots,X_n$ i.i.d. $\sim B(1,p)$, 则 $T_n \sim B(n,p)$. 其中 $E(T_n) = nE_(X_i) = np, \mathrm{Var}(T_n) = n\,\mathrm{Var}(X_i) = np(1-p)$. 根据中央极限定理有:

$$\frac{T_n - np}{\sqrt{np(1-p)}} \xrightarrow{d} N(0,1)$$

即当 n 很大时, $B(n,p)$ 可用 $N(np, np(1-p))$ 近似.

注 其他与二项分布一样, 可以通过独立同分布的随机变量相加得到的分布, 也可在一定条件下近似为正态分布. 例如 Gamma 分布 (可通过指数分布相加得到), 泊松分布 (泊松分布相加), 负二项分布 (几何分布相加), 参见图**??**

注 由定理4.1可知, 当 $B(n,p)$ 中 n 很大 p 很小时, 二项分布可近似为泊松分布; 而泊松分布 λ 很大时, 又可近似为正态分布.

例题 5.9 测量误差估计 假设每次测量结果为独立同分布的随机变量 $X_1,\cdots,X_n$, 其均值与方差分别为 μ,σ^2. 由中央极限定理可知 $\frac{\sqrt{n}(\overline{X_n}-\mu)}{\sigma} \xrightarrow{d} N(0,1)$. 由例5.5可知 $S_n^2 \xrightarrow{\mathbb{P}} \sigma^2$, 即 $\frac{\sigma}{S_n} \xrightarrow{\mathbb{P}} 1$(定理5.2). 所以:

$$\frac{\sqrt{n}(\overline{X_n}-\mu)}{S_n} = \frac{\sqrt{n}(\overline{X_n}-\mu)}{\sigma}\frac{\sigma}{S_n} \xrightarrow{d} N(0,1)\cdot 1 = N(0,1)$$

即可以通过分布 $N(0,\frac{S_n^2}{n})$ 估计测量误差 $\overline{X_n}-\mu$

注 由定理**??**可知 $\frac{\sqrt{n}(\overline{X_n}-\mu)}{S_n} \sim t_{n-1}$. 但由例5.6可知, $n\to\infty$ 时, t 分布收敛于正态分布, 所以不冲突.