随机序列的收敛性和中心极限定理

（2018.11.5，2018.11.7 庄逸在李启寨概统课及课下整理扩充）

随机序列的收敛性

- **几乎必然收敛、依概率收敛和依分布收敛**

$X_1, X_2, \dots$ 几乎必然收敛于$X \Leftrightarrow \lim\limits_{n\to\infty} P(X_n = X) = 1$，记为$X_n \xrightarrow{a.s.} X$

$X_1, X_2, \dots$ 依概率收敛于$X \Leftrightarrow \forall \varepsilon > 0, \lim\limits_{n\to\infty} P(|X_n - X| > \varepsilon) = 0$，记为$X_n \xrightarrow{P} X$

$X_1, X_2, \dots$ 依分布收敛于$X \Leftrightarrow$ 对分布函数$F(X)$的任何连续点x均成立

$\lim\limits_{n\to\infty} P(X_n \le x) = P(X \le x)$，记为$X_n \xrightarrow{d} X$

$X_n = X + \dfrac{1}{n}$依概率收敛于X但不几乎必然收敛于X

（条件不足，在假定四阶矩存在的情况下则几乎必然收敛）

几乎必然收敛和依概率收敛的差别在于**相差的概率为零**和**分布完全相同**不一样

考虑连续区间上的分布挖去一个点（甚至有限个点）。

更精细的考察推荐广泛查阅资料和例子（如维基等）。勿全信此处（因为我自己也还弄不太清楚）

推荐一篇博客

- **三种收敛的强弱和性质**

$X_n \xrightarrow{a.s.} X \Rightarrow X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{d} X$

几乎必然收敛最强，依概率收敛齐次，依分布收敛最弱

$X_n \xrightarrow{d} X, Y_n \xrightarrow{P} 0$ 则$X_n + Y_n \xrightarrow{d} X$

$X_n \xrightarrow{d} X, Y_n \xrightarrow{P} C$ 则$X_n Y_n \xrightarrow{d} CX$

- **平均收敛**

$X_1, X_2, \dots$ r次平均收敛于$X \Leftrightarrow \lim\limits_{n\to\infty} E|X_n - X|^r = 0$

$r=2$ 也称均方收敛

- **马尔科夫不等式**

若X只取非负值的随机变量，则对$\forall\varepsilon>0, P(X\ge\varepsilon)\le\frac{EX}{\varepsilon}$

证明：$EX-\varepsilon P(X\ge\varepsilon)=\int_0^{\infty}xf(x)dx-\int_{\varepsilon}^{\infty}\varepsilon f(x)dx=\int_{\varepsilon}^{\infty}(x-\varepsilon)f(x)dx+\int_0^{\varepsilon}xf(x)dx$

两项均不小于零，因此$EX-\varepsilon P(X\ge\varepsilon)\ge 0$

- **切比雪夫不等式**

设随机变量X的方差存在，则$\forall\varepsilon>0, P(|X-EX|\ge\varepsilon)\le\frac{DX}{\varepsilon^2}$

证明：利用马尔科夫不等式，

$P(|X-EX|\ge\varepsilon)=P((X-EX)^2\ge\varepsilon^2)\le\frac{E(X-EX)^2}{\varepsilon^2}=\frac{DX}{\varepsilon^2}$

- **切比雪夫弱大数律**

设$X_1,X_2,\ldots$是相互独立的随机变量序列，$EX_i=\mu_i$

$E(X_i-\mu_i)^2=\sigma_i^2(i\ge 1)$且$\{\sigma_i^2, i\ge 1\}$有界，$S_n=\sum_{i=1}^{n}X_i\,(n\ge 1)$

则当$n\to\infty$时，$\frac{S_n-ES_n}{n}\xrightarrow{P}0$

用切比雪夫不等式证明

$$D\left(\frac{S_n}{n}\right)=\frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2\in\left[\frac{\sigma_m^2}{n},\frac{\sigma_M^2}{n}\right]$$

$$P\left(\left|\frac{S_n-ES_n}{n}\right|\ge\varepsilon\right)=P\left(\left|\frac{S_n}{n}-E\left(\frac{S_n}{n}\right)\right|\ge\varepsilon\right)\le\frac{D\left(\frac{S_n}{n}\right)}{\varepsilon^2}\le\frac{\sigma_M^2}{n\varepsilon^2}\to 0,\text{当}n\to\infty$$

关于定理的意义

条件：方差均有限

$$\frac{S_n-ES_n}{n}=\frac{S_n}{n}-E\left(\frac{S_n}{n}\right)\coloneqq\bar{X}-E\bar{X}=\frac{\sum_{k=1}^{n}(X_k-EX_k)}{n}=\sum_{k=1}^{n}\left(\frac{X_k}{n}-E\left(\frac{X_k}{n}\right)\right)$$

注意此处$\frac{S_n}{n},\bar{X},\frac{X_k}{n}$等均为随机变量，$E\bar{X}\ne\frac{1}{n}\sum_{i=1}^{n}X_i$，而应该是一个数。

$$\frac{S_n - ES_n}{n} \xrightarrow{P} 0 \Leftrightarrow \bar{X} - E\bar{X} \xrightarrow{P} 0 \Leftrightarrow \forall \varepsilon > 0, \lim_{n\to\infty} P(|\bar{X} - E\bar{X}| > \varepsilon) = 0$$

大数定律是（随机变量序列的算术平均值）向（随机变量各数学期望的算术平均值）收敛的定律。

推论

设X_n是一列独立同分布$(\mathrm{i.i.d})$的随机变量序列，具有公共的数学期望μ和方差σ^2,

$S_n = \sum_{i=1}^{n} X_i\ (n \geq 1)$，则$\frac{S_n}{n} \xrightarrow{P} \mu\ (n \to \infty)$

证明：$\frac{S_n - ES_n}{n} = \frac{S_n}{n} - \frac{1}{n}ES_n = \frac{S_n}{n} - \frac{1}{n} \cdot n\mu = \frac{S_n}{n} - \mu$

$\frac{S_n}{n} - \mu \xrightarrow{P} 0 \Rightarrow \frac{S_n}{n} \xrightarrow{P} \mu$

贝努力大数律

设单次试验中事件A发生的概率为p,

在$n(\geq 2)$次独立试验中A发生了X_n次，则$\frac{X_n}{n} \xrightarrow{P} p \quad (n \to \infty)$

Cantelli **强大数律**

设$X_1, X_2, \ldots$是相互独立的随机变量序列，$EX_i = \mu_i$

$E(X_i - \mu_i)^4$均有限，$S_n = \sum_{i=1}^{n} X_i\ (n \geq 1)$

则当$n \to \infty$时，$\frac{S_n - ES_n}{n} \xrightarrow{a.s.} 0$

老师注：一般而言$E(X^4)$存在的话低阶的也都存在，也就是说$E(X^4)$存在的话四阶中心矩就存在。

推论

设X_n是一列独立同分布$(\mathrm{i.i.d})$的随机变量序列

$\mu = EX_i$和EX_i^4存在，$S_n = \sum_{i=1}^{n} X_i\ (n \geq 1)$，则

$\frac{S_n}{n} \xrightarrow{a.s.} \mu\ (n \to \infty)$

Borel **强大数律**

设单次试验中事件A发生的概率为p，在$n(\geq 2)$次独立试验中

A发生了X_n次，则$\frac{X_n}{n} \xrightarrow{a.s.} p \quad (n \to \infty)$

中心极限定理

棣莫弗-拉普拉斯中心极限定理

设$X_1,X_2,\dots,X_n$相互独立并且具有相同的分布，

$P(X_1=1)=1-P(X=0)=p,0<p<1$，则有

$$\frac{X_1+X_2+\cdots+X_n-np}{\sqrt{np(1-p)}}\xrightarrow{d}N(0,1)$$

二项分布经过标准化（减去均值，除以标准差）后分布可用标准正态分布估计。

即 $\lim\limits_{n\to\infty}P\left(\frac{X_1+X_2+\cdots+X_n-np}{\sqrt{np(1-p)}}\le x\right)=\Phi(x),\forall x\in\mathbb{R}$

棣莫弗-拉普拉斯中心极限定理是历史上最早的中心极限定理，它告诉我们可以用正态分布来近似二项分布。

林德伯格-莱维中心极限定理

设$X_1,X_2,\dots,X_n$为$i.i.d$的随机变量序列，具有公共的数学期望μ和方差σ^2

$$\frac{1}{\sqrt{n\sigma^2}}(X_1+X_2+\cdots+X_n-n\mu)\xrightarrow{d}N(0,1)$$

即$\forall x\in\mathbb{R}$, 有 $\lim\limits_{n\to\infty}F_n(x)=\Phi(x)$

其中$F_n(x)$为$\frac{1}{\sqrt{n\sigma^2}}(X_1+X_2+\cdots+X_n-n\mu)$的概率分布函数

$\Phi(x)$为标准正态分布$N(0,1)$的概率分布函数

波动是有规律的。

二项分布的估计方法

设$X_1,X_2,\dots,X_n$ $i.i.d$服从 $0-1$ 分布，成功的概率为p

则$X=X_1+X_2\dots+X_n\sim B(n,p)$，由中心极限定理

$$\frac{X_1+X_2+\cdots+X_n-np}{\sqrt{np(1-p)}}\xrightarrow{d}N(0,1)$$

对任意两个正整数$t_1<t_2$

$$P(t_1 \le X \le t_2) = P\left(\frac{t_1 - np}{\sqrt{np(1-p)}} \le \frac{X - np}{\sqrt{np(1-p)}} \le \frac{t_2 - np}{\sqrt{np(1-p)}}\right)$$

$$\approx \Phi\left(\frac{t_2 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{t_1 - np}{\sqrt{np(1-p)}}\right) \qquad \because \frac{X - np}{\sqrt{np(1-p)}} \sim N(0,1)$$

为了提高估计的精度，修正上式为

$$\Phi\left(\frac{t_2 - 0.5 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{t_1 - 0.5 - np}{\sqrt{np(1-p)}}\right)$$

例：设一个考生参加 100 道题的英语标准化考试，每道题均为 4 个备选答案的选择题，且只有一个答案是正确的。每道题他都随机选择一个答案，假设评分标准为：选对得 1 分，选错和不选得 0 分。问该考生最终得分大于等于 25 的概率。

$$P(X_1 + \cdots + X_{100} \ge 25) = P\left(\frac{\frac{1}{100}\sum_{i=1}^{100} X_i - 0.25}{\sqrt{0.25(1-0.25)/100}} \ge 0\right) = 1 - \Phi(0) = \frac{1}{2}$$

注：此时如果用-0.5 的修正会造成较大误差，问题在于标准正态分布 0 附近变化剧烈。

注：其实应该算$P(25 \le + \cdots + X_{100} \le 100)$，但是 100 算比较大了，其概率上面直接当成 1。精确计算可得到其概率为 0.（约 200 个 9）503……

注：用二项分布计算得到

$$\sum_{i=25}^{100} C_{100}^{i} \left(\frac{1}{4}\right)^{i} \left(\frac{3}{4}\right)^{100-i} \approx 0.5383288679185885682 6$$