第四章：概率极限理论

1. 经典概率极限定理简介

1.1 Bernoulli 大数律

1.2 de Moivre 中心极限定理

1.3 Poisson 极限定理

2. 经典极限定理的推广

2.1 Chebyshev和Khinchine 大数律

2.2 Levy-Feller 中心极限定理

3. 依概率收敛

4. 依分布收敛

5. 几乎处处收敛

随机现象规律的刻画:

(1) 随机现象具有可重复、可观察属性；任何有限次都只是真相的近似。

(2) 寻求极限是揭示随机现象规律的基本方法。

1. 经典极限定理

1.1 Bernoulli 大数律

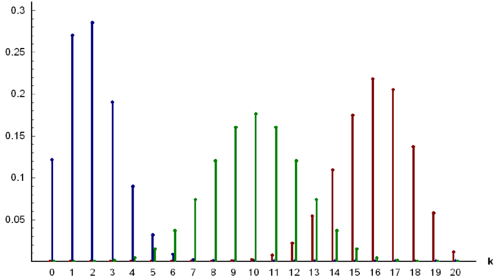
1.2 De Moivre-Laplace 中心极限定理

1.3 Poisson 极限定理

以上三个定理都只涉及Bernoulli试验（Bernoulli随机变量）$X \sim B(n, p)$:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \cdots, n$$

$$EX = np, \quad Var(X) = np(1-p)$$



Binomial distribution for $n = 20$

$p = 0.1$(blue), $p = 0.5$(green) and $p = 0.8$(red)

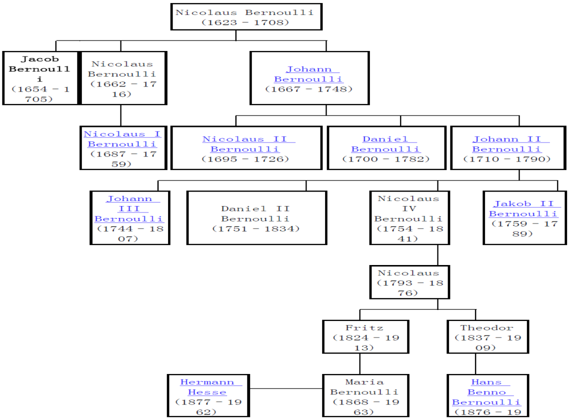
- Bernoulli 及其著作

Jacob Bernoulli

(1) Jacobi Bernoulli (1654—1705), 法国著名数学家；其家族在200年间先后出现8位数学家

(2) “The Art of Guessing”一书1713 出版，标志者概率论学科发展的正式开始

Bernoulli family tree



- 大数律

给定$0 < p < 1$, 假设$S_n \sim B(n,p)$, 那么

$$\frac{S_n}{n} \to p, \quad n \to \infty$$

“收敛”的含义

(1) 固定$0 < \varepsilon < \min\{p, 1-p\}$. 无论$n$有多大，总会发生

$$|\frac{S_n}{n} - p| > \varepsilon$$

即

$$P(\omega : |\frac{S_n(\omega)}{n} - p| > \varepsilon) > 0$$

(2) 事实上，

$$P(\omega : \left|\frac{S_n(\omega)}{n} - p\right| > \varepsilon) = \sum_{k:|\frac{k}{n}-p|>\varepsilon} \binom{n}{k} p^k (1-p)^{n-k} = ?$$

(3) Bernoulli的贡献

$$\sum_{k:|\frac{k}{n}-p|>\varepsilon} \binom{n}{k} p^k (1-p)^{n-k} \to 0, \quad n \to \infty$$

- Bernoulli大数律: 给定$0 < p < 1$, 假设$S_n \sim B(n,p)$, 那么

$$P\left(\omega : \left|\frac{S_n(\omega)}{n} - p\right| > \varepsilon\right) \to 0, \quad n \to \infty$$

Bernoulli大数律的意义

(1) 给出了“频率接近概率真值”的数学解释;

实验者	抛掷次数(n)	正面向上次数（频数m）	频率(m/n)
棣莫佛（英国 1667-1754）	2048	1061	0.5181
蒲丰（法国 1707-1788）	4040	2048	0.5069
费勒（美国 1906-1970）	10000	4979	0.4979
皮尔逊（英国 1857-1936）	12000	6019	0.5016
皮尔逊（英国 1857-1936）	24000	12012	0.5005
德-摩根（英国 1806-1871）	4092	2048	0.5005

对现代概率论的意义

(2) 引入了“依概率收敛”的概念

(Ω, Σ, P) 是一个概率空间，$X, X_n, n \geq 1$是一列随机变量，如果对任意$\varepsilon > 0$,

$$P(\omega : |X_n(\omega) - X(\omega)| > \varepsilon) \to 0, \qquad n \to \infty$$

称X_n依概率收敛到X，记做$X_n \xrightarrow{P} X$.

按此概念，Bernoulli大数律可写成

$$\frac{S_n}{n} \xrightarrow{P} p$$

(3) 问题: 如何推广和应用Bernoulli大数律？

BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 50, Number 3, July 2013, Pages 373–390
S 0273-0979(2013)01411-3
Article electronically published on March 28, 2013

TERCENTENNIAL ANNIVERSARY OF BERNOULLI'S LAW OF LARGE NUMBERS

MANFRED DENKER

0

The importance and value of Jacob Bernoulli's work was eloquently stated by Andreĭ Andreyevich Markov during a speech presented to the Russian Academy of Science[1] on December 1, 1913. Marking the bicentennial anniversary of the Law of Large Numbers, Markov's words remain pertinent one hundred years later:

> In concluding this speech, I return to Jacob Bernoulli. His biographers recall that, following the example of Archimedes he requested that on his tombstone the logarithmic spiral be inscribed with the epitaph *Eadem mutata resurgo*. This inscription refers, of course, to properties of the curve that he had found. But it also has a second meaning. It also expresses Bernoulli's hope for resurrection and eternal life.
>
> We can say that this hope is being realized. More than two hundred years have passed since Bernoulli's death but he lives and will live in his theorem.

Indeed, the ideas contained in Bernoulli's *Ars Conjectandi* have impacted many mathematicians since its posthumous publication in 1713. The twentieth century, in particular, has seen numerous advances in probability that can in some way be traced back to Bernoulli. It is impossible to survey the scope of Bernoulli's influence in the last one hundred years, let alone the preceding two hundred. It is perhaps more instructive to highlight a few beautiful results and avenues of research that demonstrate the lasting effect of his work.

1.2 De Moivre- Laplace 中心极限定理

(1) De Moivre 法国数学家棣莫弗(1667-1754年)

Abraham de Moivre

- 棣莫弗公式：

$$(\cos\theta + i\sin\theta)^n = \cos n\theta + i\sin n\theta$$

- 棣莫弗的重要著作《机会的学说》（The doctrine of chances）

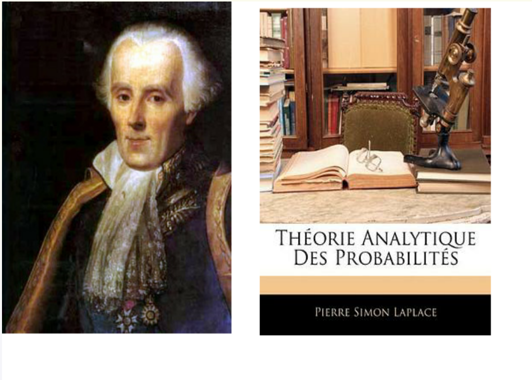
1697年，当选为英国皇家学会会员

1710年，被委派参与英国皇家学会调查牛顿-莱布尼茨关于微积分优先权的委员会

1735年，被选为柏林科学院院士

1754年，被选为法国的巴黎科学院会员

(2) Laplace (拉普拉斯1749－1827) 法国数学家、物理学家



1795年，担任巴黎综合工科学校，高等师范学校教授。

1799年，担任过法国经度局局长，并在拿破仑政府中任过6个星期的内政部长。

1816年，被选为法兰西学院院士，1817年任该院院长。

发表天文学、数学和物理学的论文270多篇，专著合计有4006多页。

其中最有代表性的专著

- 《天体力学》（1799～1825, 15卷16册）
- 《宇宙体系论》（1796，中译本1978年版）
- 《概率分析理论》（1812）. 在该书中总结了当时整个概率论的研究，论述了概率在选举审判调查、气象等方面的应用。

- 实际问题

(1) 通知200名学生参加某讲座，但每名学生出现讲座的概率仅为0.5，问应选择何种规模的报告厅？

(2) 计算

$$P(S_{200} \leq 80), \quad P(80 \leq S_{200} \leq 120), \quad P(S_{200} \geq 120)$$

(3)

$$P(80 \leq S_{200} \leq 120) = \sum_{80 \leq k \leq 120} \binom{200}{k} 0.5^{120} = ?$$

- De Moivre-Laplace 中心极限定理

假设$S_n \sim B(n,p)$，那么

$$P(\underbrace{\frac{S_n-np}{\sqrt{np(1-p)}}} \le x) \asymp \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

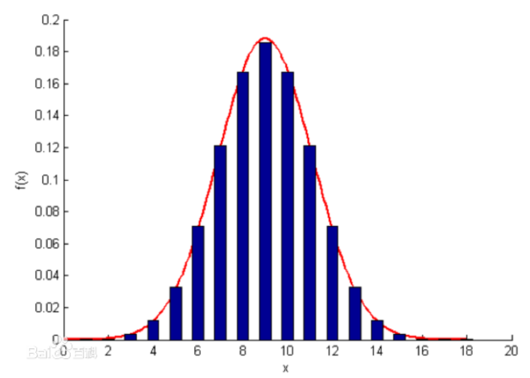
(1) 左边: 规范化$\frac{S_n-np}{\sqrt{np(1-p)}}$随机变量的分布函数

(2) 右边：正态分布函数

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

(3)

$$\begin{aligned} P(a \le S_n \le b) &= P(\frac{a-np}{\sqrt{np(1-p)}} \le \frac{S_n-np}{\sqrt{np(1-p)}} \le \frac{b-np}{\sqrt{np(1-p)}}) \\ &\asymp \Phi(\frac{b-np}{\sqrt{np(1-p)}}) - \Phi(\frac{a-np}{\sqrt{np(1-p)}}) \end{aligned}$$



$$
\begin{aligned}
P(80 \leq S_{200} \leq 120) &= P\left(\frac{80-100}{\sqrt{200 \times \frac{1}{4}}} \leq \frac{S_{200}-100}{\sqrt{200 \times \frac{1}{4}}} \leq \frac{120-100}{\sqrt{200 \times \frac{1}{4}}}\right) \\
&\asymp \Phi\left(\frac{20}{\sqrt{50}}\right) - \Phi\left(-\frac{20}{\sqrt{50}}\right) \asymp 2\Phi(2.828) - 1
\end{aligned}
$$

证明

(1)$p=\frac{1}{2}$ — De Moivre

(2) $p\neq\frac{1}{2}$ — Laplace (3) 基本出发点：

$$P\left(a\leq\frac{S_n-np}{\sqrt{np(1-p)}}\leq b\right)=\sum_{k:a\leq\frac{kn-np}{\sqrt{np(1-p)}}\leq b}\binom{n}{k}p^k(1-p)^{n-k}$$

de Moivre的基本思想（$p=\frac{1}{2}$）

$$P(S_n=\frac{n}{2}+k)=\frac{1}{2^n}\binom{n}{\frac{n}{2}+k},\quad \text{假设}\frac{n}{2}\text{ 是整数}$$

运用de Moivre 和他的朋友Stirling 1730年左右共同发展起来的公式：

$$n!=n^ne^{-n}\sqrt{2\pi n}(1+o(1))$$

$$n!=\sqrt{2\pi n}n^ne^{-n}e^{r_n},\quad \frac{1}{12n+1}<r_n\leq\frac{1}{12n}---Robbins(1955)$$

$$\frac{1}{2^n}\binom{n}{\frac{n}{2}+k} \approx \frac{2}{\sqrt{2\pi n}} e^{-\frac{2k^2}{n}}, \quad \frac{1}{2^n}\binom{n}{\frac{n}{2}} \approx \frac{2}{\sqrt{2\pi n}}$$

$$\begin{aligned} P(a \leq S_n - \frac{n}{2} \leq b) &= \sum_{a \leq k \leq b} P(S_n = \frac{n}{2} + k) \\ &\asymp \sum_{a \leq k \leq b} \frac{2}{\sqrt{2\pi n}} e^{-\frac{2k^2}{n}} \\ &\asymp \frac{2}{\sqrt{2\pi}} \int_{a/\sqrt{n}}^{b/\sqrt{n}} e^{-2y^2} dy \end{aligned}$$

(4) 将给出更现代的证明

- de Moivre - Laplace 中心极限定理的意义

(1) 给出近似计算公示

(2) 引入“依分布收敛”的概念

假设(Ω, Σ, P)是概率空间，X, X_n, $n \geq 1$ 是一列随机变量，F, F_n, $n \geq 1$是一列相应的分布函数，如果对$\underline{F\text{的任意连续点}x}$，

$$F_n(x) \to F(x), \quad n \to \infty$$

称F_n依分布收敛F，记$F_n \xrightarrow{d} F$ 或者$X_n \xrightarrow{d} X$.

按上述概念，中心极限定理可写成

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

(3) 问题：如何推广和应用中心极限定理?

20世纪初概率学家大都称呼该定理为极限定理(Limit Theorem)，由于该定理在概率论中处于如此重要的中心位置，于是数学家波利亚(G.Polya)于1920年在该定理前面冠以“中心”一词，由此后续人们都称之为“中心极限定理”。

1.3 Poisson 极限定理

令$0 < p_n < 1$，假设$S_n \sim B(n, p_n)$. 如果$np_n \to \lambda$，并且$0 < \lambda < 1$ 那么对任何$k = 0, 1, 2, \cdots$

$$P(S_n = k) \to \frac{\lambda^k e^{-\lambda}}{k!}, \quad n \to \infty$$

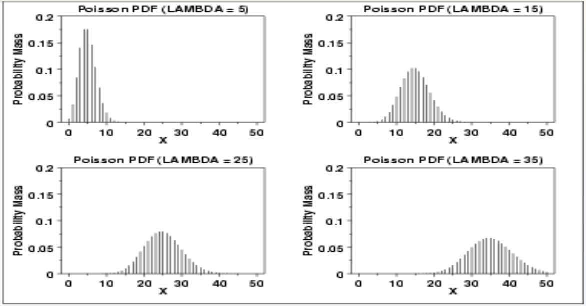
证明：由于$S_n \sim B(n, p_n)$

$$\begin{aligned} P(S_n = k) &= \frac{n!}{k!(n-k)!} p_n^k (1 - p_n)^{n-k} \\ &= \frac{1}{k!} \cdot n(n-1)\cdots(n-k+1) \cdot \frac{1}{n^k} \cdot (np_n)^k \\ &\quad \cdot (1 - \frac{\lambda}{n} + o(\frac{1}{n}))^{n-k} \\ &\to \frac{\lambda^k}{k!} e^{-\lambda}, \quad n \to \infty. \end{aligned}$$

- Poisson 分布

$$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}$$

The following is the plot of the Poisson probability density function for four values of λ.



2. 经典极限定理的推广

2.1 Chebyschev 大数律

2.2 Levy-Feller 中心极限定理

2.3 Lyapunov 中心极限定理

2.1 Chebyschev大数律

(1) Chebyschev 不等式: 对任意随机变量X，EX和EX^2存在有限，那么对任意$\varepsilon > 0$

$$P(|X - EX| > \varepsilon) \leq \frac{Var(X)}{\varepsilon^2}$$

(2) 应用Chebyschev 不等式证明Bernoulli大数律:

$$\begin{aligned} P(|\frac{S_n}{n} - p| > \varepsilon) &= P(|S_n - np| > n\varepsilon) \\ &\leq \frac{Var(S_n - np)}{n^2\varepsilon^2} \\ &= \frac{np(1-p)}{n^2\varepsilon^2} \to 0, \quad n \to \infty \end{aligned}$$

- Chebyschev 大数律

假设$\xi_k, k \geq 1$ 是一列随机变量，$E\xi_k = \mu$. 记$S_n = \sum_{k=1}^n \xi_k$，如果

$$\frac{Var(S_n)}{n^2} \to 0, \quad n \to \infty$$

那么

$$\frac{S_n}{n} \xrightarrow{P} \mu, \quad n \to \infty$$

更一般地, 假设$\xi_k, k \geq 1$ 是一列随机变量，$E\xi_k = \mu_k$. 如果

$$\frac{Var(S_n)}{n^2} \to 0, \quad n \to \infty$$

那么

$$\frac{S_n}{n} - \frac{\sum_{k=1}^n \mu_k}{n} \xrightarrow{P} 0, \quad n \to \infty$$

Chebyschev 大数律的证明: 对任意$\varepsilon > 0$,

$$P(\left|\frac{S_n}{n} - \frac{\sum_{k=1}^{n}\mu_k}{n})\right| > \varepsilon) \leq \frac{Var(S_n)}{n^2\varepsilon^2} \to 0$$

- Chebyschev 大数律的意义:

(1) 样本均值渐近总体均值

(2) 没有独立性要求

- Chebyschev 大数律的不足之处: <u>要求方差存在</u>

- Khinchine 大数律

假设$\xi_k, k \geq 1$ 是一列独立同分布的随机变量，$E\xi_k = \mu$.
记$S_n = \sum_{k=1}^{n} \xi_k$，那么

$$\frac{S_n}{n} \xrightarrow{P} \mu$$

Khinchine 大数律的证明

(1) 由于没有方差条件，因此不能直接使用Chebyschev 不等式

(2) 但可以先截尾，然后使用Chebyschev 不等式（略）

(3) 后面将给出另一种证明: 利用特征函数方法

应用

例1. 令$\xi_1, \xi_2, \cdots$,是一列独立同分布随机变量，服从参数为$\lambda > 0$的指数分布。记$S_n = \sum_{k=1}^{n} \xi_k$。证明：

$$\frac{S_n}{n} \xrightarrow{P} \frac{1}{\lambda}.$$

事实上，$E\xi_1 = \frac{1}{\lambda}$. 直接使用Khinchine 大数律即可。

例2. 令$\xi_1, \xi_2, \cdots$,是一列独立随机变量，ξ_k分布如下：$\xi_1 \equiv 0$; $k \geq 2$,

$$P(\xi_k = k) = P(\xi_k = -k) = \frac{1}{2k\log k}$$

$$P(\xi_k = 0) = 1 - \frac{1}{k\log k}$$

记$S_n = \sum_{k=2}^{n} \xi_k$。证明：

$$\frac{S_n}{n} \xrightarrow{P} 0$$

事实上，$E\xi_k = 0$，$Var(\xi_k) = \frac{k}{\log k}$

$$Var(S_n) = \sum_{k=1}^{n} Var(\xi_k) = \sum_{k=2}^{n} \frac{k}{\log k} \sim \frac{n^2}{\log n}$$

$$\frac{Var(S_n)}{n^2} \sim \frac{1}{\log n} \to 0, \quad n \to \infty$$

例3. 令$\xi_1, \xi_2, \cdots$,是一列具有相同分布的随机变量, $E\xi_k = \mu$,
$Var(\xi_k) = \sigma^2 < \infty$。假设$\xi_k$和$\xi_{k+1}$相关，$k \geq 1$
当$|k-l| \geq 2$时，ξ_k与ξ_l相互独立。记$S_n = \sum_{k=1}^n \xi_k$。证明：

$$\frac{S_n}{n} \xrightarrow{P} \mu$$

计算$Var(S_n) =?$，$\frac{Var(S_n)}{n^2} \xrightarrow{?} 0$

$$Var(S_n) = \sum_{k=1}^n Var(\xi_k) + 2\sum_{k<l} Cov(\xi_k, \xi_l)$$

$$|Cov(\xi_k, \xi_{k+1})| \leq (Var(\xi_k) \cdot Var(\xi_{k+1}))^{1/2} \leq \sigma^2$$

$$Cov(\xi_k, \xi_l) = 0, \quad |k-l| > 1$$

这样，

$$\frac{Var(S_n)}{n^2} \to 0, \quad n \to \infty$$

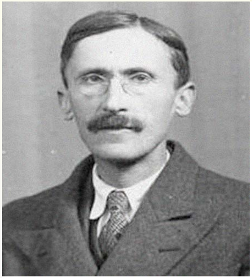
3. de Moivre-Laplace 中心极限定理的推广

3.1 Levy和Feller中心极限定理

(1) Paul Levy 莱维(1886-1971)，法国数学家，现代概率论开拓者之一. 1886年出生于巴黎，曾于1905年19岁时在Ecole理工学院发表了第一篇论文. 他的老师和顾问为Hadamard 阿达马。
毕业以后，法国炮兵队服兵役1年，此后3年就读于高等矿业，于1913年成为教授。
对极限理论和随机过程理论作出了重要贡献。概率论中的莱维过程(Lévy processes)，莱维测度（Lévy measure），莱维分布(Lévy distribution) 等都是以其命名。
《Theory of Sums of Independent Random Variables》

(2) William Feller 费勒(1906 – 1970)

1925年，Zagreb 大学毕业
1926年，Gottingen 大学博士毕业
1939年，移民美国，任Brown大学、Cornell大学、Princeton大学教授
现代概率论学科的奠基人之一
《An Introduction to Probability Theory and Its Applications》Vol. I, II ，20世纪最优秀的数学书籍之一。
1969年获美国国家奖（总统奖）
小行星编号：21276 Feller



Levy

Feller

- Levy-Feller 中心极限定理

假设$\xi_k, k \geq 1$ 是一列独立同分布随机变量，$E\xi_k = \mu, Var(\xi_k) = \sigma^2$.
记$S_n = \sum_{k=1}^n \xi_k$，那么对任意x,

$$P(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x) \to \Phi(x)$$

即

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

- Levy-Feller 中心极限定理的意义

(1) 应用于一般随机变量，推广了de Moivre-Laplace 中心极限定理
(2) 说明测量误差可用正态分布描述，即<u>正态分布无处不在</u>
随机测量值X_i, 真值μ. 每次误差为$X_i - \mu$，n次观测所得误差叠加，记为$\sum_{i=1}^n (X_i - mu)$。那么

$$\sum_{i=1}^n (X_i - mu) \sim N(0, n\sigma^2), \quad n \gg 1$$

应用（以下一些文字摘自百度百科，仅供参考，欢迎大家讨论）

一、教育统计学统计规律表明，学生的智力水平，包括学习能力，实际动手能力等呈正态分布。因而正常的考试成绩分布应基本服从正态分布。

考试分析要求绘制出学生成绩分布的直方图，以“中间高、两头低”来衡量成绩符合正态分布的程度。

其评价标准认为：考生成绩分布情况直方图，基本呈正态曲线状，属于好，如果略呈正（负）态状，属于中等，如果呈严重偏态或无规律，就是差的。

从概率统计规律看，“正常的考试成绩分布应基本服从正态分布”是正确的。

不同观点：大学成绩被正态分布支配，比内卷还可怕。

但是必须考虑人与物的本质不同，以及教育的有所作为可以使“随机”受到干预，用曲线或直方图的形状来评价考试成绩就有失偏颇。许多教育专家已经通过实践论证，教育是可以大有作为的，可以做到大多数学生及格，而且多数学生可以得高分，考试成绩曲线是偏正态分布的。但是长期受到“中间高、两头低”标准的影响，限制了教师的作为，抑制了多数学生能够学好的信心。这是很大的误会。通常正态曲线有一条对称轴。当某个分数（或分数段）的考生人数最多时，对应曲线的最高点，是曲线的顶点。该分数值在横轴上的对应点与顶点连接的线段就是该正态曲线的对称轴。考生人数最多的值是峰值。我们注意到，成绩曲线或直方图实际上很少对称的，称之为峰线更合适。

二、某些医学现象，如同质群体的身高、红细胞数、血红蛋白量，以及实验中的随机误差，呈现为正态或近似正态分布；
有些指标（变量）虽服从偏态分布，但经数据转换后的新变量可服从正态或近似正态分布，可按正态分布规律处理。其中经对数转换后服从正态分布的指标，被称为服从对数正态分布。
医学参考值范围亦称医学正常值范围。它是指所谓“正常人”的解剖、生理、生化等指标的波动范围。制定正常值范围时，首先要确定一批样本含量足够大的“正常人”，所谓“正常人”不是指“健康人”，而是指排除了影响所研究指标的疾病和有关因素的同质人群；其次需根据研究目的和使用要求选定适当的百分界值，常用95%.
根据指标的实际用途确定单侧或双侧界值，如白细胞计数过高过低皆属不正常须确定双侧界值，又如肝功中转氨酶过高属不正常须确定单侧上界，肺活量过低属不正常须确定单侧下界。另外，还要根据资料的分布特点，选用恰当的计算方法。

三、在生产条件不变的情况下，产品的强力、抗压强度、口径、长度等指标；同一种生物体的身长、体重等指标；同一种种子的重量；测量同一物体的误差；弹着点沿某一方向的偏差；某个地区的年降水量；以及理想气体分子的速度分量，等等。

一般来说，如果一个量是由许多微小的独立随机因素影响的结果，那么就可以认为这个量具有正态分布

- Levy-Feller 中心极限定理的证明

(1) 组合计算的方法不再适用

(2) 后面将用特征函数方法证明

例1. 假设$\xi_1, \xi_2, \cdots$,是一列独立同分布的随机变量，服从的分布如下：

$$P(\xi_k = 0) = P(\xi_k = -1) = P(\xi_k = 1) = \frac{1}{3}.$$

记$S_n = \sum_{k=1}^n \xi_k$。求$P(S_{10000} > 100) = ?$

$$E\xi_k = 0, \qquad Var(\xi_k) = \frac{2}{3}.$$

$$\begin{aligned} & P(S_{10000} > 100) \\ & = P\Big(\frac{S_{10000}}{\sqrt{10000 \times \frac{2}{3}}} > \frac{100}{\sqrt{10000 \times \frac{2}{3}}}\Big) \\ & = 1 - \Phi(\frac{100}{81.65}) \approx 0.11 \end{aligned}$$

例2. 某车间有200台车床，工作时每台车床60%的时间在开动，每台开动时耗电1千瓦。问应供给这个车间多少电力才能有0.999把握保证正常生产？

$$P(S_{200} < x) \approx 0.999$$

求$x =?$

$$\begin{aligned} P(S_{200} < x) &= P\left(\frac{S_{200} - 200 \times 0.6}{\sqrt{200 \times 0.6 \times 0.4}} < \frac{x - 200 \times 0.6}{\sqrt{200 \times 0.6 \times 0.4}}\right) \\ &\approx \Phi\left(\frac{x - 120}{\sqrt{48}}\right) = 0.999 \end{aligned}$$

$$\frac{x - 120}{\sqrt{48}} \asymp$$

3.2 Lyapunov 中心极限定理

假设$\xi_k, k \geq 1$ 是一列独立随机变量(不一定同分布)，$E\xi_k = \mu_k$, $Var(\xi_k) = \sigma_k^2$. 记$S_n = \sum_{k=1}^n \xi_k$，$B_n = \sum_{k=1}^n \sigma_k^2$. 如果

(1) $B_n \to \infty$

(2) $E|X_k|^3 < \infty$, 且

$$\frac{\sum_{k=1}^n E|X_k|^3}{B_n^{3/2}} \to 0, \quad n \to \infty$$

那么对任意x

$$P(\frac{\sum_{k=1}^n (\xi_k - \mu_k)}{\sqrt{B_n}} \leq x) \to \Phi(x)$$

即

$$\frac{\sum_{k=1}^n (\xi_k - \mu_k)}{\sqrt{B_n}} \xrightarrow{d} N(0,1)$$

例3. 假设$\xi_k, k \geq 1$是一列独立随机变量，

$P(\xi_k = 1) = p_k$，$P(\xi_k = 0) = 1 - p_k$，其中$0 < p_k < 1$.

如果

$$\sum_{k=1}^{n} p_k(1 - p_k) \to \infty$$

那么

$$\frac{\sum_{k=1}^{n}(\xi_k - p_k)}{\sqrt{\sum_{k=1}^{n} p_k(1 - p_k)}} \xrightarrow{d} N(0, 1)$$

4. 依概率收敛和依分布收敛

4.1 依概率收敛的概念

4.2 依概率收敛的基本性质

4.3 依分布收敛的概念

4.4 依分布收敛的基本性质

我们将给出Khinchine大数律和Levy-Feller中心极限定理的证明

4.1 依概率收敛的概念

假设(Ω, Σ, P)是一个概率空间，$X, X_n, n \geq 1$是一列随机变量，如果对任意$\varepsilon > 0$

$$P(\omega : |X_n(\omega) - X(\omega)| > \varepsilon) \to 0, \quad n \to \infty$$

我们称X_n依概率收敛到X，并记做$X_n \xrightarrow{P} X$.

X 可以是常数(退化随机变量)

- 极限唯一性

假设$X_n \xrightarrow{P} X, \quad X_n \xrightarrow{P} Y$，那么

$$P(X = Y) = 1$$

证明：只需证明$P(X \neq Y) = 0$。注意到,

$$P(X \neq Y) = P(|X - Y| > 0) = P(\cup_{m=1}^{\infty}\{|X - Y| > \frac{1}{m}\})$$

因此，需要证明对任意$\varepsilon > 0$，

$$P(|X - Y| > \varepsilon) = 0$$

给定$\varepsilon > 0$，对任意$n \geq 1$

$$\begin{aligned} P(|X - Y| > \varepsilon) &= P(|\underbrace{(X_n - X)} - \underbrace{(X_n - Y)}| > \varepsilon) \\ &\leq P(|X_n - X| + |(X_n - Y)| > \varepsilon) \\ &\leq P(|X_n - X| > \frac{\varepsilon}{2}) + P(|X_n - Y| > \frac{\varepsilon}{2}) \end{aligned}$$

令$n \to \infty$，

$$P(|X_n - X| > \frac{\varepsilon}{2}) \to 0, \quad P(|X_n - Y| > \frac{\varepsilon}{2}) \to 0$$

$$\Downarrow$$

$$P(|X - Y| > \varepsilon) = 0$$

4.2 依概率收敛的判别法则

如果存在某$r>0$, 成立

$$E|X_n - X|^r \to 0, \quad n \to \infty$$

那么

$$X_n \xrightarrow{P} X$$

证明：应用Markov不等式即可

- 依概率收敛的基本性质

(1) 运算性质

如果$X_n \xrightarrow{P} X, Y_n \xrightarrow{P} Y$，那么

(i) $X_n \pm Y_n \xrightarrow{P} X \pm Y$

(ii) $X_n \cdot Y_n \xrightarrow{P} X \cdot Y$

(iii) 如果$P(Y \neq 0) = 1$，那么$\frac{X_n}{Y_n} \xrightarrow{P} \frac{X}{Y}$

证明：以下仅证明(i)中加法成立，其余类似。定$\varepsilon > 0$，对任意$n \geq 1$

$$\begin{aligned} P(|X_n + Y_n - (X+Y)| > \varepsilon) &\leq P(|X_n - X| + |Y_n - Y| > \varepsilon) \\ &\leq P(|X_n - X| > \frac{\varepsilon}{2}) + P(|Y_n - Y| > \frac{\varepsilon}{2}) \end{aligned}$$

由于

$$P(|X_n - X| > \frac{\varepsilon}{2}) \to 0, \quad P(|Y_n - Y| > \frac{\varepsilon}{2}) \to 0$$

所以

$$P(|X_n + Y_n - (X+Y)| > \varepsilon) \to 0$$

(iv) 连续映射保持依概率收敛性

假设$f:\mathbb{R}\mapsto\mathbb{R}$是连续映射，如果$X_n\xrightarrow{P}X$，那么

$$f(X_n)\xrightarrow{P}f(X)$$

证明：对任意小的ε,η，需要证明：存在$N=N(\varepsilon,\eta)$，使得当$n\geq N$

$$P(|f(X_n)-f(X)|>\varepsilon)\leq\eta$$

注意到，$F_X(+\infty)=1,\ F_X(-\infty)=0,$ 那么存在$M>0$，

$$P(|X|>M)=\underbrace{1-F_X(M)}+\underbrace{F_X(-M)}\leq\frac{\eta}{8}$$

由于$P(|X_n-X|>1)\to 0,$ 那么存在N_1足够大，使得$n\geq N_1$

$$\begin{aligned}P(|X_n|>M+1) &\leq \underbrace{P(|X_n|>M+1,|X|\leq M)}\\ &\quad +\underline{P(|X_n|>M+1,|X|>M)}\\ &\leq \underbrace{P(|X_n-X|>1)}+\underline{P(|X|>M)}\leq\frac{\eta}{4}\end{aligned}$$

所以
$$P(|X|>M+1)\le\frac{\eta}{4},\quad P(|X_n|>M+1)\le\frac{\eta}{4}$$

- f是连续映射，因此f在$[-M-1,M+1]$上一致连续：对任意$\varepsilon>0$，存在$\delta>0$，如果

$$\{(x,y):|x-y|<\delta,\quad\forall x,y\in[-M-1,M+1]\}\subset\{(x,y):|f(x)-f(y)|<\varepsilon\}$$

$$\begin{aligned}
&P(|f(X_n)-f(X)|>\varepsilon)\\
&\le P(|f(X_n)-f(X)|>\varepsilon,\underbrace{X_n,X\in[-M-1,M+1]})\\
&+P(|f(X_n)-f(X)|>\varepsilon,\underbrace{X_n\mathrm{or}X\in[-M-1,M+1]^c})\\
&\le P(|X_n-X|>\delta)+P(|X_n|>M+1)+P(|X|>M+1)
\end{aligned}$$

对于上述$\delta > 0$，存在N_2使得$n \geq N_2$

$$P(|X_n - X| > \delta) \leq \frac{\eta}{2}$$

选取$N = \max\{N_1, N_2\}$，那么

$$P(|f(X_n) - f(X)| > \varepsilon) \leq \frac{\eta}{2} + \frac{\eta}{4} + \frac{\eta}{4} = \eta$$

证完

例. 假设ξ_k, $k \geq 1$ 独立同分布，服从$[0,1]$上的均匀分布，令

$$\eta_n = (\prod_{k=1}^{n} \xi_k)^{1/n}$$

求证：

$$\eta_n \xrightarrow{P} c, \quad n \to \infty$$

证：注意

$$\log \eta_n = \frac{1}{n} \sum_{k=1}^{n} \log \xi_k \xrightarrow{P} E \log \xi_1 = \int_0^1 \log x dx = -1$$

这样，

$$\eta_n = e^{\log \eta_n} \xrightarrow{P} e^{-1}$$

4.3 依分布收敛的概念

假设$X, X_n, n \geq 1$是一列随机变量，其分布函数分别为$F, F_n, n \geq 1$。如果对每个F的连续性点x，

$$F_n(x) \to F(x), \quad n \to \infty$$

那么称F_n依分布收敛到F，记做$F_n \xrightarrow{d} F$

也称X_n依分布收敛到X，记做$X_n \xrightarrow{d} X$

注:

(1) 如果F是在$\mathbb{R}$上连续，那么F_n处处收敛到F

(2) 一般地，F不是连续函数(左极限存在，右连续函数)。

(3) 既然F是单调有界函数，F的不连续性点集最多可数个：

$$D_F = \{x : F(x) - F(x-) > 0\} = \cup_{n=1}^{\infty} \underbrace{\{x : F(x) - F(x-) \geq \frac{1}{n}\}}$$

(4) F的连续性点集在$\mathbb{R}$上稠密。

例1. 考虑$X_n \equiv \frac{1}{n}, X \equiv 0.$

$$F_n(x) = \begin{cases} 1 & x \geq \frac{1}{n} \\ 0 & x < \frac{1}{n} \end{cases}, \quad F(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}$$

可以看出

(1) F_n 不是处处收敛到F

(2) 除0点外，F_n 处处收敛到F

$$F_n \xrightarrow{d} F, \quad X_n \xrightarrow{d} X$$

例2. 考虑$X_n \equiv n$

$$F_n(x) = \begin{cases} 1 & x \geq n \\ 0 & x < n \end{cases}$$

对每个x，

$$F_n(x) \to 0, \quad n \to \infty$$

但$F(x) \equiv 0$不是分布函数

• 依概率收敛意味着依分布收敛

假设$X_n \xrightarrow{P} X$，那么$X_n \xrightarrow{d} X$

证明：对任意$x \in \mathbb{R}$ 和$\varepsilon > 0$

$$\begin{aligned} P(X_n \le x) &= P(X_n \le x, \underbrace{X_n - X \ge -\varepsilon}) \\ &\quad + P(X_n \le x, \underbrace{X_n - X < -\varepsilon}) \end{aligned}$$

$$P(X_n \le x, X_n - X < -\varepsilon) \le P(X_n - X < -\varepsilon) \to 0$$

$$\limsup_{n\to\infty} P(X_n \le x) \le \underline{P(X \le x + \varepsilon) \to P(X \le x)}, \quad \varepsilon \to 0$$

类似地，

$$\liminf_{n\to\infty} P(X_n \le x) \ge \underline{P(X \le x - \varepsilon) \to P(X < x)}, \quad \varepsilon \to 0$$

如果x是F的连续性点，那么令$\varepsilon \to 0$得

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

结论成立

- 依分布收敛并不意味着依概率收敛

例3. 假设Y是非退化对称随机变量，令$X_n = Y, n \geq 1$，$X = -Y$. 那么$X, X_n, n \geq 1$的分布都相同，因此$X_n \xrightarrow{d} X$. 但

$$P(|X_n - X| > \varepsilon) = P(2|Y| > \varepsilon)$$

所以X_n不依概率收敛到X

- 如果$X_n \xrightarrow{d} c$，那么$X_n \xrightarrow{P} c$.

证明：

$$P(X_n \leq c + \varepsilon) \to 1 \quad \Rightarrow \quad P(X_n > c + \varepsilon) \to 0, \quad n \to \infty$$

类似地，

$$P(X_n < c - \varepsilon) \to 0, \quad n \to \infty$$

所以

$$P(|X_n - c| > \varepsilon) \longrightarrow 0$$

$$X_n \xrightarrow{P} c$$

4.4 依分布收敛的判别法则

- Levy连续性定理:

假设$X, X_n, n \geq 1$是一列随机变量，具有特征函数$\phi, \phi_n, n \geq 1$. 那么

$$X_n \xrightarrow{d} X \quad \Leftrightarrow \quad \phi_n(t) \to \phi(t), \quad t \in \mathbb{R}$$

- Levy连续性定理的另一种形式:

假设$X_n, n \geq 1$是一列随机变量，具有特征函数$\phi_n, n \geq 1$. 如果

$$\phi_n(t) \to \phi(t), \quad t \in \mathbb{R}$$

并且ϕ在0处连续，那么ϕ一定是特征函数. 记与ϕ相应的随机变量为X，那么

$$X_n \xrightarrow{d} X$$

- Levy 定理的应用

一、Khinchine大数律的证明

回忆$\xi_k, k \geq 1$独立同分布，$E\xi_k = \mu$，那么

$$\frac{1}{n}\sum_{k=1}^{n}\xi_k \xrightarrow{P} \mu$$

证明：只需证明

$$X_n =: \frac{1}{n}\sum_{k=1}^{n}\xi_k \xrightarrow{d} \mu$$

$$\Updownarrow$$

$$\phi_n(t) =: Ee^{itX_n} \to e^{it\mu}$$

$$\phi_n(t) = Ee^{itX_n} = [Ee^{i\frac{t}{n}\xi_1}]^n$$

在0处进行Taylor展开，

$$Ee^{i\frac{t}{n}\xi_1} = 1 + \frac{it\mu}{n} + o(\frac{1}{n}), \quad n \to \infty$$

所以，对每个$t \in \mathbb{R}$,

$$\phi_n(t) = [1 + \frac{it\mu}{n} + o(\frac{1}{n})]^n \to e^{it\mu}$$

二、Levy-Feller中心极限定理的证明

回忆: 如果$\xi_k, k \geq 1$独立同分布，$E\xi_k = \mu, Var(\xi_k) = \sigma^2$，那么

$$X_n = \frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{n} (\xi_k - \mu) \xrightarrow{d} N(0,1)$$

证明：只要证明

$$\phi_n(t) \doteq Ee^{itX_n} \to e^{-\frac{t^2}{2}}$$

注意到

$$Ee^{itX_n} = [Ee^{i\frac{t(\xi_k-\mu)}{\sigma\sqrt{n}}}]^n$$

在0处进行Taylor展开，

$$Ee^{i\frac{t(\xi_k-\mu)}{\sigma\sqrt{n}}} = 1 - \frac{t^2}{2n} + o(\frac{1}{n})$$

所以，对任意t，

$$Ee^{itX_n} = [1 - \frac{t^2}{2n} + o(\frac{1}{n})]^n \to e^{-\frac{t^2}{2}}$$

- 依分布收敛的运算性质

(1) 线性运算

(i) 假设$X_n \xrightarrow{d} X$，$b_n \to b$。那么

$$X_n + b_n \xrightarrow{d} X + b$$

证明：由于$X_n \xrightarrow{d} X$，所以

$$Ee^{itX_n} \to Ee^{itX}$$

这样，

$$\begin{aligned} Ee^{it(X_n+b_n)} &= Ee^{itX_n}e^{itb_n} \\ &\to Ee^{itX}e^{itb} \\ &= Ee^{it(X+b)} \end{aligned}$$

(ii) 假设$X_n \xrightarrow{d} X$，$a_n \to a$。那么

$$a_n X_n \xrightarrow{d} aX$$

证明：首先假设$a > 0$. 那么对任意$\varepsilon > 0$，当n充分大

$$\frac{a}{1+\varepsilon} < a_n < \frac{a}{1-\varepsilon}$$

$$\begin{aligned} P(a_n X_n \le x) &\le P(X_n \le \frac{(1+\varepsilon)x}{a}) \\ &\to P(X \le \frac{(1+\varepsilon)x}{a}) \\ &= P(aX \le (1+\varepsilon)x) \end{aligned} \tag{1}$$

类似地，

$$\begin{aligned} P(a_n X_n \le x) &\ge P(X_n \le \frac{(1-\varepsilon)x}{a}) \\ &\to P(X \le \frac{(1-\varepsilon)x}{a}) \\ &= P(aX \le (1-\varepsilon)x) \end{aligned}$$

综合起来，当x是aX的连续性点时，

$$P(a_nX_n \le x) \to P(aX \le x), \quad n \to \infty$$

假设$a < 0$，证明完全类似

假设$a = 0$. 对任意$\varepsilon, \eta > 0$，当n充分大时，$|a_n| < \frac{\varepsilon}{\eta}$.

取η使得$\pm\frac{1}{\eta}$是X的分布函数的连续性点，

$$\begin{aligned} P(|a_nX_n| > \varepsilon) &= P(|X_n| > \frac{1}{\eta}) \\ &\to P(|X| > \frac{1}{\eta}) \quad n \to \infty \end{aligned}$$

令$\eta \to 0$，

$$P(|a_n X_n| > \varepsilon) \to 0 \tag{2}$$

综合起来

(iii) 假设$X_n \xrightarrow{d} X$，$a_n \to a$，$b_n \to b$。那么

$$a_n X_n + b_n \xrightarrow{d} aX + b$$

(2)

假设$X_n \xrightarrow{d} X$，$Y_n \xrightarrow{P} c$ 。那么

$$Y_n \cdot X_n \xrightarrow{d} cX$$

证明：类似于(ii)

(3) 连续映射保持依分布收敛性

假设$f:\mathbb{R}\mapsto\mathbb{R}$是连续映射，$X_n \xrightarrow{d} X$。那么

$$f(X_n) \xrightarrow{d} f(X)$$

Helly 引理: 假设$F, F_n, n\geq 1$是一列分布函数，并且$F_n \xrightarrow{d} F$，那么对任意有界连续函数g

$$\int g(x)dF_n(x) \to \int g(x)dF(x)$$

这样，对任意$t\in\mathbb{R}$，

$$\begin{aligned} Ee^{itf(X_n)} &= \int e^{itf(x)}dF_n(x) \\ &\to \int e^{itf(x)}dF(x) \\ &= Ee^{itf(X)} \end{aligned}$$

结论成立

例. 假设$\xi_k, k \geq 1$是一列独立同分布随机变量，

$$P(\xi_k = \pm 1) = \frac{1}{2}$$

定义

$$U_n = \sum_{k=1}^{n} \frac{\xi_k}{2^k}$$

试证：$U_n \to U$，其中$U \sim U(-1, 1)$.

证明：首先，分别计算U_n和U的特征函数$\phi_n(t)$ 和$\phi(t)$:

$$\begin{aligned}
\phi_n(t) &= Ee^{itU_n} = \prod_{k=1}^{n} Ee^{it\frac{\xi_k}{2^k}} \\
&= \prod_{k=1}^{n} (\frac{1}{2}(e^{\frac{it}{2^k}} + e^{-\frac{it}{2^k}})) \\
&= \prod_{k=1}^{n} \cos(\frac{t}{2^k})
\end{aligned}$$

$$\phi(t) \quad = \quad Ee^{itU} = \frac{1}{2}\int_{-1}^{1} e^{itx}dx = \frac{\sin t}{t}$$

下面证明

$$\prod_{k=1}^{n} \cos(\frac{t}{2^k}) \longrightarrow \frac{\sin t}{t}$$

事实上，注意到$2\sin\theta\cos\theta = \sin 2\theta$, 和$\frac{\sin\theta}{\theta} \to 1 \ (\theta \to 0)$

$$\begin{aligned}
\prod_{k=1}^{n} \cos(\frac{t}{2^k}) \quad &= \quad \frac{1}{\sin\frac{t}{2^n}} \prod_{k=1}^{n} \cos(\frac{t}{2^k}) \sin\frac{t}{2^n} \\
&= \quad \frac{\sin t}{2^n \sin\frac{t}{2^n}} \longrightarrow \frac{\sin t}{t}
\end{aligned}$$

这样，

$$\phi_n(t) \to \phi(t) \quad U_n \xrightarrow{d} U.$$

5. 几乎处处收敛

5.1 几乎处处收敛的概念

5.2 Borel-Cantelli 引理

5.3 Borel 大数律, Kolmogorov 强大数律

注：

(1) 几乎处处收敛比依概率收敛，依概率收敛比依分布收敛强

(2) Borel 大数律比Bernoulli 大数律强

(3) Kolmogorov 大数律比Khinchine 大数律强

5.1 几乎处处收敛的概念

(1) 处处收敛：假设$(\Omega, \mathcal{A}, P)$是一个概率空间，$X, X_n, n \geq 1$是一列随机变量，如果对每个$\omega \in \Omega$，

$$X_n(\omega) \to X(\omega), \quad n \to \infty$$

那么称X_n处处收敛于X

(2) 几乎处处收敛: 假设$(\Omega, \mathcal{A}, P)$是一个概率空间，$X, X_n, n \geq 1$是一列随机变量，如果存在Ω_0使得

(i) $P(\Omega_0) = 0$

(ii) 对每个$\omega \in \Omega \setminus \Omega_0$，

$$X_n(\omega) \to X(\omega), \quad n \to \infty$$

那么称X_n几乎处处收敛于X，记做$X_n \to X, a.s.$

即除一个零概率事件外，X_n 处处收敛于X

- 几乎处处收敛的判别法则

$$X_n \to X, \quad a.s.$$

当且仅当对任意$\varepsilon > 0$,

$$P(\bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} \{|X_n(\omega) - X(\omega)| > \varepsilon\}) = 0$$

或者说

$$P(\{|X_n(\omega) - X(\omega)| > \varepsilon\}, i.o.) = 0$$

等价地, 对任意$\varepsilon > 0$,

$$\lim_{N\to\infty} P(\bigcup_{n=N}^{\infty} \{|X_n(\omega) - X(\omega)| > \varepsilon\}) = 0$$

由此，也可看出几乎处处收敛比依概率收敛强。

5.2 Borel-Cantelli 引理

(1) 假设$A_n, n \geq 1$是一列事件，如果

$$\sum_{n=1}^{\infty} P(A_n) < \infty$$

那么

$$P(A_n, i.o.) = 0$$

证明：

$$P(A_n, i.o.) = \lim_{N \to \infty} P(\bigcup_{n=N}^{\infty} A_n)$$

$$P(\bigcup_{n=N}^{\infty} A_n) \leq \sum_{n=N}^{\infty} P(A_n) \to 0$$

(2) 假设$A_n, n \geq 1$是一列独立事件，如果

$$\sum_{n=1}^{\infty} P(A_n) = \infty$$

那么

$$P(A_n, i.o.) = 1$$

证明：由于$A_n, n \geq 1$是一列独立事件，那么

$$\begin{aligned} P(\bigcap_{n=N}^{M} A_n^c) &= \prod_{n=N}^{M} P(A_n^c) = \prod_{n=N}^{M} \underbrace{(1 - P(A_n))} \\ &\leq \prod_{n=N}^{M} \underbrace{e^{-P(A_n)}} = e^{-\sum_{n=N}^{M} P(A_n)} \end{aligned}$$

$$\lim_{N\to\infty} \lim_{M\to\infty} P(\bigcap_{n=N}^{M} A_n^c) = 0$$

Borel-Cantelli 引理也被称为Borel-Cantelli 0-1律

- Borel 大数律

假设$(\Omega, \mathcal{A}, P)$是一个概率空间，$\xi_k, k \geq 1$是一列独立同分布随机变量，

$$P(\xi_k = 1) = p, \quad P(\xi_k = 0) = 1 - p$$

记$S_n = \sum_{k=1}^{n} \xi_k$，那么$\frac{S_n}{n} \to p, \quad a.s.$

证明: 对任意$\varepsilon > 0$

$$P(|\frac{S_n}{n} - p| > \varepsilon, i.o.) = 0$$

由Borel-Cantelli引理，只需验证

$$\sum_{n=1}^{\infty} P(|\frac{S_n}{n} - p| > \varepsilon) < \infty$$

由Markov不等式，

$$P(|\frac{S_n}{n} - p| > \varepsilon) \leq \frac{E|S_n - np|^4}{n^4 \varepsilon^4}$$

容易计算得，

$$E|S_n - np|^4 = np(1-p)[p^3 + (1-p)^3] + n(n-1)p^2(1-p)^2$$

$$\sum_{n=1}^{\infty} P(|\frac{S_n}{n} - p| > \varepsilon) \leq K(\varepsilon, p) \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty$$

5.3 Kolmogorov 强大数律

假设$(\Omega, \mathcal{A}, P)$是一个概率空间，$\xi_k, k \geq 1$是一列独立同分布随机变量. 如果$E\xi_k = \mu$，那么

$$\frac{S_n}{n} \to \mu, \quad a.s.$$

(1) Kolmogorov 强大数律推广了Borel 强大数律

(2) Kolmogorov 强大数律推广了Khinchine 大数律