现代数学
·专著版·

重庆理工大学优秀著作出版基金资助

缺失数据下的广义线性模型

肖枝洪 程新跃 著

WUHAN UNIVERSITY PRESS
武汉大学出版社

现代数学

重庆理工大学优秀著作出版基金资助

缺失数据下的广义线性模型

肖枝洪 程新跃 著

WUHAN UNIVERSITY PRESS
武汉大学出版社

图书在版编目(CIP)数据

缺失数据下的广义线性模型/肖枝洪,程新跃著.—武汉:武汉大学出版社,2013.6

ISBN 978-7-307-10411-2

Ⅰ.缺… Ⅱ.①肖… ②程… Ⅲ.线性模型—研究 Ⅳ.O212

中国版本图书馆 CIP 数据核字(2013)第 002776 号

责任编辑:顾素萍　　责任校对:王　建　　版式设计:詹锦玲

出版发行:武汉大学出版社　(430072　武昌　珞珈山)
(电子邮件:cbs22@whu.edu.cn　网址:www.wdp.com.cn)
印刷:武汉中远印务有限公司
开本:720×1000　1/16　印张:10.5　字数:166 千字　插页:1
版次:2013 年 6 月第 1 版　2013 年 6 月第 1 次印刷
ISBN 978-7-307-10411-2　定价:26.00 元

前 言

广义线性模型在自然科学、社会科学的各个领域有着广泛的用途，同时在数据观测或调查的过程中，有时难免出现缺失数据或者不完全信息. 在本著作中，我们主要研究自然联系函数情形下带有不完全信息随机截尾的广义线性模型参数的极大似然估计的存在性、相合性、渐近正态性、重对数律、Chung 型重对数律及中偏差. 这些优良性质是对所建模型能够得到合理运用的必要保障.

在实际应用中，特别是在医药学和社会科学中，广义线性模型的回归变量常常是随机的，所以，Fahrmei 考虑了回归变量 $X_1, X_2, \cdots$ 是某个随机矩阵的独立同分布的情况，在一定的条件下，不加证明地给出了此种模型的参数矩阵的极大似然估计的大样本性质. 然而，在实际中，独立同分布的情形有时也是不太贴切的. 例如，在医药学和社会科学的研究中，数据可能来自不同的人群、时间和地点，因而，数据的分布是不同的. 鉴于此，丁洁丽和陈希孺于 2005 年对上述情形进行了改进，得到并证明了回归变量不同分布情形下广义线性模型参数的极大似然估计的相合性和渐近正态性. 在本著作中，我们借鉴丁洁丽和陈希孺的思想和方法，得到并证明了回归变量不同分布情形下带有不完全信息随机截尾的广义线性模型参数的极大似然估计的相合性、渐近正态性和重对数律.

历史上还有许多研究者致力于拓展广义线性模型的应用，起初只局限于响应变量服从指数型分布族的情形. 1974 年，Wedderburn 提出拟似然函数的概念，可以在建模时只要求关于响应变量的数学期望函数和协方差函数的正确设定，而不必要求响应变量的分布为指数分布类型. 后续的研究表明，在方差函数不确知，但对期望有正确设定的情况下，这

种方法仍可适用. 这种方法称为拟似然法. 由其所得的估计, 则称为拟似然估计. 岳丽和陈希孺, 赵林城和尹长明分别于2004年与2005年在一定的条件下, 给出了广义线性模型的未知参数向量 β 的拟极大似然估计 $\hat{\beta}_n$ 及其强相合性和渐近正态性. 陈夏和陈希孺于 2005 年给出了广义线性模型参数的自适应拟似然估计. 在本著作中, 我们在岳丽和陈希孺工作的基础上, 讨论拟极大似然估计 $\hat{\beta}_n$ 的重对数律和 Chung 型重对数律.

历史上也有学者研究一些模型的参数的极大似然估计的中偏差原理, 这是极大似然估计的又一个很有意义的大样本性质. 令 $\{X_k, k \geq 1\}$ 是一个独立不必同分布的随机变量序列, 且在空间 $\mathbb{R}$ 中取值, 同时具有分布函数 $F_k(x,\theta)$, 其中 $\theta \in \Theta$, 且参数空间 $\Theta \subset \mathbb{R}$. θ 的极大似然估计 (MLE) 被记为 $\hat{\theta}_n \equiv \hat{\theta}_n(X_1, X_2, \cdots, X_n)$. 极大似然估计 $\hat{\theta}_n$ 的强相合性于1976年为Chao M. T. 所研究, 其渐近正态性于1971年为Hoadley所研究. 对于独立同分布情形的极大似然估计 $\hat{\theta}_n$ 的中偏差原理于 2001 年为高付清所研究. 在本著作中, 我们讨论独立不必同分布情形下单参数 θ 的极大似然估计 $\hat{\theta}_n$ 的中偏差原理. 我们的结果是高付清的结果的一个推广. 由于不完全信息随机截尾的广义线性模型是独立不必同分布的一个特殊情形, 在本著作中, 我们就将它作为一个例子给出.

在本著作的写作过程中, 我们尽量对建模所加的条件的合理性进行解释, 尽量给出我们提出问题的背景, 也尽量将我们用来解决问题的方法同相关文献解决问题的方法进行比较, 以体现我们所做工作的特色.

本著作的出版得到重庆理工大学优秀著作出版基金的资助, 也得到武汉大学出版社的大力支持, 在此致以衷心的感谢! 也对关心和支持我们出版此书的同行朋友致以衷心的谢意!

由于作者水平有限, 错谬之处在所难免, 恳请广大读者不吝赐教, 我们将致以非常诚挚的感谢!

肖枝洪　程新跃

于重庆理工大学花溪校区

2013 年 3 月

目录

第1章 准备工作

为了使本著作相对来说具有一定的封闭性，也便于读者阅读，我们将后续章节需要的概念和需要引证的基本结果集中放在本章.

1.1 大数定律和中心极限定理

因为本著作主要讨论极大似然估计的大样本性，作为本著作的准备工作，我们首先介绍概率论的若干极限理论，对于概率空间的若干概念不作介绍，有兴趣的读者可以在稍微深入一点的有关概率论书籍中查阅到.

定义 1.1 设样本空间为 $\Omega=\{\omega:\omega$ 为试验的基本可能结果或样本点$\}$，$\mathbb{P}$ 为定义在 Ω 上的概率测度，$X:\Omega\to\mathbb{R}$ 为随机变量. 定义**随机变量 X 的(累积)分布函数**为

$$F(x)=\mathbb{P}\{\omega:X(\omega)\leq x\}\equiv\mathbb{P}\{X\leq x\},$$

其中，x 为任意的实数.

有些教科书给出分布函数的定义为 $F(x)=\mathbb{P}\{X<x\}$，本书所采用的定义为定义 1.1 的形式.

类似于定义 1.1，我们可以给出多维随机向量的分布函数的定义：设随机向量 $\boldsymbol{X}(\omega)=\big(X_1(\omega),X_2(\omega),\cdots,X_p(\omega)\big)^{\mathrm{T}}\in\mathbb{R}^p$，对于任意的实向量 $\boldsymbol{x}=(x_1,x_2,\cdots,x_p)^{\mathrm{T}}\in\mathbb{R}^p$，**随机向量 $\boldsymbol{X}$ 的分布函数**为

$$F(\boldsymbol{x})=\mathbb{P}\{\omega:\boldsymbol{X}(\omega)\leq\boldsymbol{x}\}\equiv\mathbb{P}\{X_j\leq x_j,\ j=1,2,\cdots,p\}.$$

定义 1.2 对于分布函数列 $\{F_n(x)\}$，如果存在一个函数 $F(x)$ 使

$\lim\limits_{n\to\infty} F_n(x) = F(x)$ 在 $F(x)$ 的每一连续点上都成立，且

$$\lim_{n\to\infty} F_n(\pm\infty) = F(\pm\infty),$$

则称 $\{F_n(x)\}$ **弱收敛于**$F(x)$，并记为 $F_n(x) \xrightarrow{w} F(x)$.

定义 1.3 设随机变量 X_n 和 X 的分布函数分别为 $F_n(x)$ 和 $F(x)$，如果 $F_n(x) \xrightarrow{w} F(x)$，则称随机变量序列 $\{X_n,\ n \geq 1\}$ **依分布收敛于** X，并记为 $X_n \xrightarrow{d} X$.

定义 1.4 对于随机变量序列 $\{X_n\}$ 和随机变量 X，如果对任意的 $\varepsilon > 0$，有

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

成立，则称 $\{X_n\}$ **依概率收敛于** X，记为 $X_n \xrightarrow{\mathbb{P}} X$.

定义 1.5 对于随机变量序列 $\{X_n\}$ 和随机变量 X，如果

$$\mathbb{P}\Big\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\Big\} \equiv \mathbb{P}\Big\{\lim_{n\to\infty} X_n = X\Big\} = 1$$

成立，或者

$$\lim_{n\to\infty} \mathbb{P}\Big\{\sup_{l>n} |X_l - X| \geq \varepsilon\Big\} = 0$$

成立，则称 $\{X_n\}$ **几乎处处收敛于** X，记为 $X_n \xrightarrow{\text{a.s.}} X$.

定义 1.6 设对于随机变量 X_n 和 X，有 $\mathbb{E}|X_n|^r < \infty$，$\mathbb{E}|X|^r < \infty$，其中 $r > 0$ 为常数. 如果

$$\lim_{n\to\infty} \mathbb{E}|X_n - X|^r = 0$$

成立，则称 $\{X_n\}$ **依 r 阶矩收敛于** X，记为 $X_n \xrightarrow{r} X$.

定义 1.7 设 $\{X_n,\ n \geq 1\}$ 是随机变量序列，令

$$S_n = X_1 + X_2 + \cdots + X_n.$$

如果存在常数序列 $\{a_n,\ n \geq 1\}$ 和 $\{b_n,\ n \geq 1\}$，且 $\lim\limits_{n\to\infty} a_n = \infty$，使得对于任意的 $\varepsilon > 0$，恒有

$$\lim_{n\to\infty} \mathbb{P}\left\{\left|\frac{S_n}{a_n} - b_n\right| < \varepsilon\right\} = 1,$$

则称随机变量序列$\{X_n,\ n \geq 1\}$ **服从大数定律(或大数法则)**. 此处的大

数定律也叫**弱大数定律**.

如果令常数序列 $\{b_n,\ n \geq 1\}$ 中每个 $b_n = \beta$, 把随机序列 $Y_n = \dfrac{S_n}{a_n}$ 看做 β 的估计序列, 且 Y_n 和 β 满足定义 1.7, 则称 Y_n 为 β 的**弱相合估计**, 有时简称为**相合估计**.

注　相合估计是统计推断中一个非常基本的准则. 其意思就是说, 如果从总体中合理地抽取了足够多个体, 得到的统计量还不能很好地刻画参数的真实值, 那么说明按照这个方法构造的统计量不合适.

定义 1.8　设 $\{X_n,\ n \geq 1\}$ 是随机变量序列, 令

$$S_n = X_1 + X_2 + \cdots + X_n.$$

如果存在常数序列 $\{a_n,\ n \geq 1\}$ 和 $\{b_n,\ n \geq 1\}$, 且 $\lim\limits_{n\to\infty} a_n = \infty$, 使得对于任意的 $\varepsilon > 0$, 恒有

$$\mathbb{P}\left\{\lim_{n\to\infty}\left|\frac{S_n}{a_n} - b_n\right| = 0\right\} = 1,$$

则称随机变量序列 $\{X_n,\ n \geq 1\}$ **服从强大数定律**(或**强大数法则**).

显然, 由定义 1.8 的条件可以推出定义 1.7 的条件, 当然可以得到定义 1.7 的结论, 因此定义 1.8 的结论比定义 1.7 的结论更深刻. 这也是将定义 1.8 称为强大数定律的缘由.

类似地, 对应强大数定律, 如果令常数序列 $\{b_n,\ n \geq 1\}$ 中每个 $b_n = \beta$, 把随机变量序列 $Y_n = \dfrac{S_n}{a_n}$ 看做 β 的估计序列, 则称 Y_n 为 β 的**强相合估计**.

下面给出一个验证大数定律的方法.

Markov 大数定律　对于随机变量序列 $\{X_n,\ n \geq 1\}$, 若有

$$\frac{1}{n^2}\mathbb{V}\mathrm{ar}\Big(\sum_{k=1}^{n} X_k\Big) \to 0 \tag{1.1.1}$$

成立, 则对任意的 $\varepsilon > 0$, 均有

$$\lim_{n\to\infty}\mathbb{P}\left\{\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}X_k\right| < \varepsilon\right\} = 1.$$

证 参见《概率论基础》(复旦大学编) 第 5 章中极限定理. □

在第 7 章介绍的引理 7.2.2 比此处的 Markov 大数定律更一般而且结论更强!

定义 1.9 设$\{X_n,\ n\geq 1\}$是独立随机变量序列, 且 $\mathbb{E}X_n$ 和 $\mathbb{D}X_n$ 均存在. 令

$$Z_n=\frac{\sum_{k=1}^{n}X_k-\sum_{k=1}^{n}\mathbb{E}X_k}{\sqrt{\sum_{k=1}^{n}\mathbb{V}\mathrm{ar}(X_k)}}.$$

若对任意的 $x\in\mathbb{R}$, 有

$$\lim_{n\to\infty}\mathbb{P}\{Z_n\leq x\}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\exp\left\{-\frac{t^2}{2}\right\}\mathrm{d}t,$$

则称 $\{X_n,\ n\geq 1\}$ **服从中心极限定理**. $\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\exp\left\{-\frac{t^2}{2}\right\}\mathrm{d}t$ 为标准正态分布的分布函数, 记为 $\Phi(x)$.

以下定理给出了中心极限定理成立的充要条件.

Lindeberg-Feller 中心极限定理 设$\{X_n,\ n\geq 1\}$是独立随机变量序列, 且设 $\mathbb{E}X_n=a_n$ 和 $\mathbb{D}X_n=b_n^2$. 令 $B_n^2=\sum_{k=1}^{n}b_k^2$, 记

$$Z_n=\frac{1}{B_n}\sum_{k=1}^{n}(X_k-a_k),\quad G_n(x)=\mathbb{P}\{Z_n\leq x\}.$$

则使得当 $n\to\infty$ 时,

$$\frac{1}{B_n^2}\max_{1\leq k\leq n}b_k^2\to 0,\quad G_n(x)\xrightarrow{w}\Phi(x)$$

成立的充要条件是 Lingdeberg 条件成立: 对任意给定的 $\varepsilon>0$,

$$\Lambda_n(\varepsilon)=\frac{1}{B_n}\sum_{k=1}^{n}\int_{|x-a_k|\geq\varepsilon\sqrt{B_n}}(x-a_k)^2\mathrm{d}F_k(x)\to 0,$$

其中 $F_k(x)$ 是 X_k 的分布函数.

证　参见《概率论基础》(复旦大学编) 第 5 章中极限定理. □

由 Lindeberg-Feller 中心极限定理推广可以得到下面的中心极限定理.

Lyapunov 中心极限定理　设 $\{X_n, n \geq 1\}$ 是独立随机变量序列, 其中至少有一个随机变量服从非退化分布. 设对某个 $\delta > 0$, $\mathbb{E}|X_n|^{2+\delta} < \infty$, $n \geq 1$. 令 $\mathbb{E}X_n = a_n$, $\mathbb{D}X_n = b_n^2$, $B_n^2 = \sum_{k=1}^{n} b_k^2$,

$$F_n(x) = \mathbb{P}\left\{\frac{1}{B_n}\sum_{k=1}^{n}(X_k - a_k) \leq x\right\}.$$

如果

$$\frac{1}{B_n^{2+\delta}}\sum_{k=1}^{n}\mathbb{E}|X_k - a_k|^{2+\delta} \to 0,$$

那么 $F_n(x) \xrightarrow{w} \Phi(x)$.

1.2　重对数律与中偏差

下面给出概率极限理论中极为深刻的结果——重对数律, 它是强大数律的精细化.

定义 1.10　设 $\{X_n, n \geq 1\}$ 是独立随机变量序列, $\mathbb{E}X_n = 0$, $\mathbb{D}X_n = \sigma_n^2$. 记 $D_n = \sum_{k=1}^{n}\sigma_k^2$, $S_n = \sum_{k=1}^{n} X_k$, 若

$$\limsup_{n\to\infty}\frac{S_n}{\sqrt{2D_n \log\log D_n}} = 1 \quad \text{a.s.}, \tag{1.2.1}$$

则称随机变量序列 $\{X_n, n \geq 1\}$ **服从重对数律**.

显然, 若 $\{X_n, n \geq 1\}$ 满足 (1.2.1), 则 $\{-X_n, n \geq 1\}$ 满足 (1.2.1), 从而有

$$\liminf_{n\to\infty}\frac{S_n}{\sqrt{2D_n \log\log D_n}} = -1 \quad \text{a.s.} \tag{1.2.2}$$

下面给出判定重对数律的准则.

Kolmogorov 重对数律 设$\{X_n,\ n\geq 1\}$是独立随机变量序列, $\mathbb{E}X_n=0$, $\mathbb{V}\mathrm{ar}(X_n)=\sigma_n^2$. 记 $D_n=\sum_{k=1}^n\sigma_k^2$, $S_n=\sum_{k=1}^n X_k$. 若$D_n\to\infty$, 且存在常数序列$\{M_n\}$满足

$$M_n=O\left(\sqrt{\frac{D_n}{\log\log D_n}}\right),\quad |X_n|\leq M_n\quad \text{a.s.},$$

则(1.2.1)成立.

Hartman-Wintner 重对数律 设 $\{X_n,\ n\geq 1\}$ 是独立随机变量序列, $\mathbb{E}X_n=0$, $\mathbb{D}X_n=1$, 则

$$\limsup_{n\to\infty}\frac{S_n}{\sqrt{2n\log\log n}}=1\quad \text{a.s.}$$

以上两个定律的证明参见[49] p. 106, p. 115.

定义 1.11 设 $\{X_n,\ n\geq 1\}$ 是独立随机变量序列, $\mathbb{E}X_n=0$, $\mathbb{D}X_n=\sigma_n^2$. 记 $D_n=\sum_{k=1}^n\sigma_k^2$, $S_n=\sum_{k=1}^n X_k$. 若

$$\liminf_{n\to\infty}\sqrt{\frac{\log\log D_n}{D_n}}\max_{1\leq i\leq n}\left|\sum_{k=1}^i X_k\right|=\frac{\pi}{\sqrt{8}}\quad \text{a.s.},\tag{1.2.3}$$

则称随机变量序列$\{X_n,\ n\geq 1\}$**服从 Chung 型重对数律.**

注 如果假设 $\mathbb{E}X_n=\beta$, 把 S_n 看做统计量, 那么从统计推断的观点看, 第一个重对数律给出了未知参数 β 的渐近意义上的最小 100% 置信区间, 而第二个重对数律几乎必然给出了估计量能够达到的精确下界.

定义 1.12 设随机变量 X_n 和 X 的期望均为 0, 方差分别为 σ_n^2 和 σ^2, 且 $X_n\xrightarrow{d}X$. 令 $n\to\infty$ 时, $\lambda(n)\to\infty$, $\frac{\lambda(n)}{\sqrt{n}}\to 0$, $\frac{1}{n}\sum_{k=1}^n\sigma_k^2\to\sigma^2$. 如果有

$$\liminf_{n\longrightarrow\infty}\frac{\lambda^2(n)}{n}\log\mathbb{P}\Big\{\lambda(n)\big(\max_{1\leq k\leq n}X_k-X\big)\geq\varepsilon\Big\}\geq-\frac{1}{2}\sigma^2\varepsilon^2,$$

以及

$$\limsup_{n\longrightarrow\infty}\frac{\lambda^2(n)}{n}\log\mathbb{P}_\theta\Big\{\lambda(n)\big(\min_{1\le k\le n}X_k-X\big)\ge\varepsilon\Big\}\le-\frac{1}{2}\sigma^2\varepsilon^2$$

成立，就称随机变量序列 $\{X_n,\ n\ge 1\}$ **具有中偏差性**. 这类问题统称为**中偏差极限理论**.

中偏差给出了统计量依分布收敛的性质，它比中心极限定理更加精细地刻画了统计量依分布收敛的情形.

1.3　基本工具

本节将后面所需要的基本工具罗列在此，以凸显后面主要结果的论证.

定义 1.13　设 $\xi_1,\xi_2,\cdots,\xi_n,\cdots$ 为实值随机变量序列. 若 $\xi_n\xrightarrow{\mathbb{P}}0$，则称 $\xi_n=o_p(1)$；若 $\forall\varepsilon>0$，$\exists M_\varepsilon>0$，$N>0$，使得 $n\ge N$ 时，

$$\mathbb{P}\big\{|\xi_n|\le M_\varepsilon\big\}\ge 1-\varepsilon,$$

则称 $\xi_n=O_p(1)$.

又设 $\eta_1,\eta_2,\cdots,\eta_n,\cdots$ 为另一实值随机变量序列. 若 $\dfrac{\xi_n}{\eta_n}=o_p(1)$，则称 $\xi_n=o_p(\eta_n)$；若 $\dfrac{\xi_n}{\eta_n}=O_p(1)$，则称 $\xi_n=O_p(\eta_n)$.

对于上述定义有如下性质：

性质　(1) 设 ξ 为实值随机变量. 若 $\xi_n\xrightarrow{d}\xi$，则 $\xi_n=O_p(1)$.

(2) 若 $\xi_n=o_p(1)$，则 $\xi_n=O_p(1)$.

(3) $o_p(\xi_n)=\xi_n o_p(1)$，$O_p(\xi_n)=\xi_n O_p(1)$.

(4) $o_p(1)+o_p(1)=o_p(1)$，$O_p(1)+O_p(1)=O_p(1)$.

(5) $o_p(1)\cdot o_p(1)=o_p(1)$，$O_p(1)\cdot O_p(1)=O_p(1)$.

(6) $c\cdot o_p(1)=o_p(1)$，$c\cdot O_p(1)=O_p(1)$，其中 c 为常数.

(7) $o_p(1)+O_p(1)=O_p(1)$，$O_p(1)\cdot o_p(1)=o_p(1)$.

(8) $o_p\big(O_p(1)\big)=o_p(1)$，$O_p\big(o_p(1)\big)=o_p(1)$，$O_p\big(O_p(1)\big)=O_p(1)$.

定义 1.14 设 $\boldsymbol{A}_n = \big(a_{ij}(n)\big)_{r\times s}$ 和 $\boldsymbol{A} = \big(a_{ij}\big)_{r\times s}$ 均为随机矩阵. 若 $\|\boldsymbol{A}_n - \boldsymbol{A}\| \xrightarrow{\mathbb{P}} 0$, 即 $\forall\varepsilon > 0$, 当 $n \to \infty$ 时, $\mathbb{P}\big\{\|\boldsymbol{A}_n - \boldsymbol{A}\| \geq \varepsilon\big\} \to 0$, 则称 $\boldsymbol{A}_n \xrightarrow{\mathbb{P}} \boldsymbol{A}$, 其中 $\|\boldsymbol{A}\|$ 指矩阵 $\boldsymbol{A}$ 的范数.

类似地, 有如下定义:

定义 1.15 设 $\boldsymbol{A}_n = \big(a_{ij}(n)\big)_{r\times s}$ 为随机矩阵. 若 $\boldsymbol{A}_n \xrightarrow{\mathbb{P}} \boldsymbol{O}$, $\boldsymbol{O}$ 为相应的零矩阵, 则称 $\boldsymbol{A}_n = o_p(\mathbf{1})$; 若 $\forall\varepsilon > 0$, $\exists M_\varepsilon > 0$, $N > 0$, 使得 $n \geq N$ 时,

$$\mathbb{P}\big\{\|\boldsymbol{A}_n\| \leq M_\varepsilon\big\} \geq 1 - \varepsilon,$$

则称 $\boldsymbol{A}_n = O_p(\mathbf{1})$, 其中 $\|\boldsymbol{A}\|$ 指矩阵 $\boldsymbol{A}$ 的范数.

对于随机矩阵序列, 也有上述性质, 在此不作叙述, 有兴趣的读者可以参看文献 [63].

下面给出本书后面将要用到的引理.

引理 1.3.1 (Borel-Cantelli 引理) 设事件列 $A_1, A_2, \cdots \in Æ$(事件类).

(i) 如果

$$\sum_n \mathbb{P}(A_n) < \infty,$$

则 $\mathbb{P}(A_n, \text{i.o.}) = 0$ 成立, 其中 $\{A_n, \text{i.o.}\}$ 意指有无穷多个 A_n 发生; 或者 $\mathbb{P}\Big(\varlimsup\limits_{n\to\infty} A_n\Big) = 0$ 成立, 其中 $\varlimsup\limits_{n\to\infty} A_n$ 意指有无穷多个 A_n 发生.

(ii) 如果

$$\sum_n \mathbb{P}(A_n) = \infty,$$

且 $A_1, A_2, \cdots$ 相互独立, 则 $\mathbb{P}\Big(\varlimsup\limits_{n\to\infty} A_n\Big) = 1$ 成立.

证 参见 Olva Kallenberg 所著的 *Foundations of Modern Probability*. □

对任意的随机变量 X, 记 $X^c = X \cdot I\big(|X| \leq c\big)$, 其中 $I\big(|X| \leq c\big)$ 满足: 若 $|X| \leq c$, 则 $I\big(|X| \leq c\big) = 1$; 否则 $I\big(|X| \leq c\big) = 0$.

引理 1.3.2 (三级数定理) 设 $\{X_n\}$ 是独立随机变量序列，那么使得级数 $\sum_{k=1}^{\infty} X_k$ a.s. 收敛的必要条件是：对每一 $c \in (0, \infty)$，有

(i) $\sum_{k=1}^{\infty} \mathbb{P}\{|X_k| > c\} < \infty$；

(ii) $\sum_{k=1}^{\infty} \mathbb{E}X_k^c$ 收敛；

(iii) $\sum_{k=1}^{\infty} \mathbb{V}\mathrm{ar}X_k^c < \infty$.

充分条件是对某一 $c \in (0, \infty)$，上述三级数收敛.

证 参见[49] p. 89 定理 2. 3. □

引理 1.3.3 设 $\xi_1, \xi_2, \cdots, \xi_n$ 为相互独立随机变量，$\mathbb{E}\xi_i = 0$, $1 \le i \le n$. 记 $B_n = \sum_{i=1}^{n} \mathbb{V}\mathrm{ar}(\xi_i)$，$D_n = \sum_{i=1}^{n} \mathbb{E}|\xi_i|^{2+\epsilon}$，$\epsilon \in [0, 1]$. 记 ξ_n 的分布函数为 $F_n(x)$，标准正态分布 $N(0, 1)$ 的分布函数为 $\Phi(x)$. 若 $B_n > 0$, $D_n \le \infty$，则有

$$\sup_x \left|F_n(x) - \Phi(x)\right| \le \frac{CD_n}{B_n^{\epsilon/2}},$$

其中，C 不依赖于 n.

证 见文献 [58] p. 155. □

引理 1.3.4 (Bernstein 不等式) 设 $\xi_1, \xi_2, \cdots, \xi_n$ 为相互独立随机变量，$\mathbb{E}\xi_i = 0$, $1 \le i \le n$. 若存在常数 b 使得 $|\xi_i| \le b$, $1 \le i \le n$，则对任何 $\varepsilon > 0$，有

$$\mathbb{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n} \xi_i\right| \ge \varepsilon\right\} \le 2\exp\left\{-\frac{n\varepsilon^2}{2b\varepsilon + 2\overline{\sigma}^2}\right\},$$

其中，$\overline{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} \mathbb{V}\mathrm{ar}(\xi_i)$.

证 见文献[4]. □

引理 1.3.5 设 $\xi_1, \xi_2, \cdots, \xi_n$ 为相互独立随机变量，$\mathbb{E}\xi_i = 0$, $1 \le i \le n$, $\overline{p} > 2$. 则有

$$\mathbb{E}\Big|\sum_{i=1}^{n}\xi_i\Big|^{\overline{p}} \le Cn^{\overline{p}/2-1}\sum_{i=1}^{n}\mathbb{E}|\xi_i|^{\overline{p}},$$

其中，C 和 n 与 ξ_i, $1 \le i \le n$ 的分布无关.

特别，若 $\sup\limits_{i\ge 1}\mathbb{E}|\xi_i|^{\overline{p}} \le Cn^{\overline{p}/2}$，则 $\mathbb{E}\Big|\sum\limits_{i=1}^{n}\xi_i\Big|^{\overline{p}} \le Cn^{\overline{p}/2}$.

证 见文献 [67] p. 154. □

引理 1.3.6 (Kronecker 引理) 设 $\{a_n\}$ 和 $\{x_n\}$ 是两实数序列，$a_n > 0$ 且 $\lim\limits_{n\to\infty} a_n = \infty$, $\sum\limits_{n=1}^{\infty}\dfrac{x_n}{a_n}$ 收敛，则 $\lim\limits_{n\to\infty}\dfrac{1}{a_n}\sum\limits_{k=1}^{n}x_k = 0$.

证 记 $a_0 = 0$，定义 $y_1 = 0$, $y_n = \sum\limits_{i=1}^{n-1}\dfrac{x_i}{a_i}$, $n \ge 2$. 则

$$y_n \to y = \sum_{i=1}^{\infty}\frac{x_i}{a_i}. \tag{1.3.1}$$

而

$$\frac{1}{a_n}\sum_{k=1}^{n}x_k = \frac{1}{a_n}\sum_{k=1}^{n}a_k(y_{k+1}-y_k) = y_{n+1} - \frac{1}{a_n}\sum_{k=1}^{n-1}(a_{k+1}-a_k)y_{k+1}. \tag{1.3.2}$$

由 (1.3.1) 知，对任意给定的 $\varepsilon > 0$，存在 n_0，当 $n \ge n_0$ 时，有 $|y_n - y| < \varepsilon$，从而有

$$\begin{aligned}\left|\frac{1}{a_n}\sum_{k=1}^{n}(a_{k+1}-a_k)y_k - y\right| &= \left|\frac{1}{a_n}\sum_{k=1}^{n}(a_{k+1}-a_k)(y_k-y)\right| \\ &\le \left|\frac{1}{a_n}\sum_{k=1}^{n_0}(a_{k+1}-a_k)(y_k-y)\right| \\ &\quad + \frac{a_n + a_{n_0-1}}{a_n}\varepsilon.\end{aligned}$$

由上式易知，