孙志华 尹俊平 陈菲菲 叶雪 著

非参数与半参数统计

清华大学出版社

非参数与半参数统计

孙志华 尹俊平 陈菲菲 叶雪 著

清华大学出版社
北京

内 容 简 介

本书介绍了现代非参数和半参数统计的基于局部核方法的基本方法和基本理论，主要内容为密度函数以及相关函数的核估计、非参数局部回归方法、生存时间函数的非参数估计以及几类常见的半参数模型的估计和检验. 本书特点是力求把方法的直观背景以及来龙去脉介绍清楚，因而即使内容相对比较复杂，但仍然比较直观易懂.

本书可以作为高等院校数理统计专业、计量经济专业以及相关专业高年级本科生及研究生的教学用书，本书对高等院校和科研机构的研究人员、工程技术人员也具有参考价值.

图书在版编目(CIP)数据

非参数与半参数统计/孙志华等著. --北京：清华大学出版社，2016

ISBN 978-7-302-43343-9

Ⅰ. ①非…　Ⅱ. ①孙…　Ⅲ. ①非参数统计　Ⅳ. ①O212.7

中国版本图书馆 CIP 数据核字(2016)第 062868 号

责任编辑：汪　操　赵从棉
封面设计：常雪影
责任校对：赵丽敏
责任印制：刘海龙

出版发行：清华大学出版社
网　　址：http://www.tup.com.cn, http://www.wqbook.com
地　　址：北京清华大学学研大厦 A 座　　邮　　编：100084
社 总 机：010-62770175　　邮　　购：010-62786544
投稿与读者服务：010-62776969, c-service@tup.tsinghua.edu.cn
质量反馈：010-62772015, zhiliang@tup.tsinghua.edu.cn
印 装 者：北京国马印刷厂
经　　销：全国新华书店
开　　本：185mm×230mm　　印　张：10.75　　字　数：241 千字
版　　次：2016 年 6 月第 1 版　　印　次：2016 年 6 月第 1 次印刷
印　　数：1～2000
定　　价：29.00 元

产品编号：056541-01

前言

非参数统计与半参数统计方法在最近的 30~40 年来得到非常迅猛的发展，这和其自身的特点有密切的关系. 非参数方法因为其不需要模型的假定，具有稳健的特点. 半参数模型综合了参数模型和非参数模型的特点，具有灵活、容易解释的特点. 非参数统计方法和半参数统计方法不仅是目前统计研究的热点，同时，这些方法在很多实际应用领域也得到了广泛的应用. 本书介绍非参数与半参数模型的基于局部核的估计和检验方法，以及生存分析中常用的非参数方法和半参数模型.

本书第 1 章给出必要的一些概率论知识；第 2 章介绍用核方法估计密度函数及其相关函数；第 3 章介绍与密度函数有关的检验；第 4 章介绍非参数回归；第 5 章介绍生存时间的函数的非参数估计；第 6~8 章介绍部分线性回归模型、单指标模型和 Cox 模型. 本书第 1 章可以自学；第 2~4 章以及第 6、7 章可以作为一门 40 或者 56 学时的课程；也可以将第 2~8 章作为一个 72 或者 56 学时的课程. 鉴于本书所考虑的变量大部分都是随机向量，因而大部分记号均为向量或者矩阵. 因此，本书没有特别用黑体字标出向量或者矩阵.

我从 2007 年秋季开始在中科院研究生院讲授这门课程，其后基本每年都会讲一次. 刚开始讲课时，想找好一点的英文教材或者中文教材，但是都没有找到. 后来我就想基于讲义资料自己写一本书. 写书的另一个动因是想做电子课件，后来证明用电子课件的教学效果并不好，我又回到了板书加讲授的授课模式.

于是大约从 2008 年底开始写这本书，一直断断续续在写，在修改. 可以说这几年来一直在琢磨，一直在查找资料，如何使书的结构更为合理，如何使书的内容更加自然易懂. 写书实在是很耗费时间和精力的事情，中间也想放弃，幸运的是终于写完了.

从 2014 年开始，北京应用物理与计算数学研究所的尹俊平副研究员加入帮助我完成书稿，我的两个学生陈菲菲和叶雪也加入到书的撰写之中，从而使书的完成速度大大加快. 本书的完成，尹俊平副研究员、陈菲菲与叶雪付出了很多心血. 我的另一个学生刘智凡帮助完成了部分图的编写以及部分内容的编写. 我的师弟胡大海、刘小惠，上这门课的学生王苗苗、华奕州等同学提供了很多修改意见. 我的博士生导师王启华研究员给我提供了很多有用的资料. 本书生存分析的很多地方借鉴了王启华老师的文章和专著的内容. 我最开始接触这个内容是我读博士时在北大光华管理学院旁听苏良军老师的课，受益很多. 再次对上述老师、朋友和学生表示真诚的感谢!

也感谢选修这门课的学生，这门课开课以后，收到很多来自学生的鼓励和肯定的意见，很多学生也提了很多很好的建议.

本书得到了国家自然科学基金（11571340，U1430103，10901162）、中国科学院大学校长基金和中国科学院大数据挖掘与知识管理重点试验室开放课题以及安徽省振兴计划团队项目（统计学前沿问题及应用）的资助.

由于时间仓促，作者的水平有限，书中的错误和缺点在所难免，希望广大读者给予批评指正.

孙志华

2016 年 3 月

目 录

第 1 章 预备知识

1.1 背景介绍

统计推断的一个基本任务是由样本观察值去了解总体. 若根据经验或某些理论, 能在进行统计推断之前对总体作一些假设, 然后基于这些假定进行统计推断, 这种统计方法称为参数统计方法. 如果所知甚少, 在进行统计推断之前不能对总体作任何假设, 或仅能作一些非常一般性 (例如分布是连续的、是对称的等) 的假设, 这时如果仍然使用参数统计方法, 其统计推断的结果显然是不可信的, 甚至有可能是错的. 在对总体的分布不作假设或仅作非常一般性的假设的条件下发展的统计方法称为非参数统计方法.

非参数统计方法是 19 世纪 40 年代以后兴起的. 1942 年, J.Wolfowitz 首先使用非参数统计一词, 早期的非参数统计主要是扩充参数检验的内容, 以使得传统的检验过程可以应用于小样本以及不同分布类型的数据. 比如常用的非参数检验有符号秩检验、双样本 Wilcoxon 检验、多样本 Kruskal-Wallis 检验等.

近年来, 由于统计理论的进一步发展与计算机收集和处理数据能力的提高, 使得发展随数据结构不同而灵活变化的模型的统计推断方法成为可能. 非参数密度估计、非参数回归等内容也成为新的研究和应用主题. 统计研究人员利用统计渐近理论突破了参数回归和模型估计的原有理论框架, 利用各种算法改进模型的计算过程, 通过调整预测偏差和方差的比例来发展适应性更强、解释更为精练、拟合优度更适中和计算更为有效的模型.

本书的主要内容之一是介绍这种研究数据结构的非参数方法, 包括非参数密度估计、非参数回归及其相关问题, 比如分布函数的估计、密度函数的导数的估计、条件密度函数的估计以及和密度函数有关的检验, 等等. 鉴于生存数据普遍存在于很多研究领域, 本书也给出了随机右删失模型下, 几种生存时间的函数的非参数估计方法.

非参数统计问题对总体分布的假定所要求的条件很宽. 因而针对这种问题而构造的非参数统计方法, 不致因为对总体分布的假定不当而导致重大错误, 所以它往往有较好的稳健性. 这是非参数统计方法的一个非常重要的特点. 但它也有以下缺点: 首先因为非参数方法基于更少的信息作出推断, 在模型假定正确的前提下, 非参数统计方法就会比参数统计方法的效果差一些. 例如, 在处理估计问题时, 估计的方差要大一些, 收敛速度要慢一些

(参数估计的速度一般为 $n^{-1/2}$，但非参数方法的收敛速度比 $n^{-1/2}$ 慢). 又例如，在给定的显著性水平下进行检验时，基于非参数估计方法构建的检验方法的第II类错误相比基于参数估计方法构建的检验方法要大些. 其次，非参数方法受数据维数的影响，存在维数祸根的问题. 具体表现为随着模型变量的维数增加，所需样本量成指数级增加. 这就导致数据的维数高于三维时，很多非参数方法的效果并不好. 发展克服或者部分克服维数问题的非参数和半参数方法是当前研究的热点之一.

为了克服非参数方法的缺陷而发展起来的是所谓的半参数统计方法. 半参数统计方法是 20 世纪 70 年代以后发展起来的重要的统计方法. 它在参数模型的基础上引入非参数分量，从而使这种模型既含有参数分量又含有非参数分量，兼顾了参数模型的准确和非参数模型的稳健的优点，相比单纯的参数模型或非参数模型有更大的适应性，具有更强的解释能力，并且部分地克服了维数祸根的问题. 半参数模型吸引了很多理论研究领域和应用领域的关注. 本书的主要内容之一是介绍非常典型的并且应用非常广泛的几类半参数模型，包括部分线性模型、单指标模型等，以及研究生存数据时使用非常广泛的 Cox 模型.

1.2 收敛方式和极限分布

在介绍非参数和半参数方法之前，我们给出在后面的内容中经常需要用到的概率论基础知识，主要包括随机序列的几种收敛方式以及包括弱大数定律、强大数定律和中心极限定理在内的一些统计渐近理论.

1.2.1 依概率收敛

依概率收敛是用概率的方法刻画随机变量的极限.

定义 1.2.1 (依概率收敛) 对随机变量序列 $\{X_n, n=1,2,\cdots\}$ 和随机变量 X，若满足：$\forall\varepsilon>0, \lim_{n\to\infty} P(|X_n-X|\geqslant\varepsilon)=0$，则称随机变量序列 $\{X_n, n=1,2,\cdots\}$ 依概率收敛于随机变量 X，记为 $X_n \xrightarrow{P} X$.

举例：假设 $X_1, X_2, \cdots, X_n$ 是均值为 μ、方差为 σ^2 的独立同分布序列. $\bar{X}_n$ 为样本均值. 显然 $E(\bar{X}_n)=\mu$ 和 $\mathrm{var}(\bar{X}_n)=\sigma^2/n$. 由切比雪夫不等式，对于 $\forall\varepsilon>0$,

$$P(|\bar{X}_n-\mu|\geqslant\varepsilon)\leqslant\frac{\sigma^2}{n\varepsilon^2}.$$

所以 $\lim_{n\to\infty} P(|\bar{X}_n-\mu|\geqslant\varepsilon)\}\leqslant\lim_{n\to\infty}\dfrac{\sigma^2}{n\varepsilon^2}=0$, 即 $\bar{X}_n \xrightarrow{P} \mu$.

定理 1.2.1 (弱大数定律) 假设 $X_1, X_2, \cdots, X_n$ 是独立同分布随机变量，且 $E|X_1|<\infty$，则当 $n\to\infty$ 时有

$$\bar{X}_n=\frac{1}{n}\sum_{i=1}^{n}X_i \xrightarrow{P} E(X_1).$$

注：(1) 更一般的情况下，$\{X_n, n=1,2,\cdots\}$ 是独立随机变量序列，并且 $E(X_i)=\mu_i$，有

$$\frac{1}{n}\sum_{i=1}^{n}X_i-\frac{1}{n}\sum_{i=1}^{n}\mu_i\xrightarrow{P}0.$$

(2) 设 a_n 是 a 的估计，若 $a_n\xrightarrow{P}a$, 则称 a_n 是 a 的弱相合估计.

因此，定理 1.2.1 中，$\bar{X}_n$ 是 $E(X_1)$ 的弱相合估计.

(3) 大数定律 (law of large numbers，LLN) 说明当样本量足够大时，样本均值的随机性消失. 也就是说，从更多的数据，可以得到更多样本空间的信息.

下面给出需要经常使用的一个定理. 注意 $\{X_n, n=1,2,\cdots\}$ 和 $\{Y_n, n=1,2,\cdots\}$ 为随机变量序列.

定理 1.2.2 若 $X_n\xrightarrow{P}X$，$Y_n\xrightarrow{P}Y$，则有：

(1) $cX_n\xrightarrow{P}cX$，其中 c 为常数；

(2) $X_n+Y_n\xrightarrow{P}X+Y$；

(3) $X_nY_n\xrightarrow{P}XY$；

(4) 若 $Y\neq\mathbf{0}$，则有 $X_n/Y_n\xrightarrow{P}X/Y$.

定理 1.2.3 若 $X_n\xrightarrow{P}X$，且 $f(\cdot)$ 是连续函数，则有 $f(X_n)\xrightarrow{P}f(X)$.

定理 1.2.3 经常被称为 Slutsky 定理.

1.2.2 几乎必然收敛

几乎必然收敛又称为以概率 1 收敛.

定义 1.2.2 (几乎必然收敛) 随机变量序列 $\{X_n, n=1,2,\cdots\}$, 当 $P(\lim_{n\to\infty}X_n=X)=1$ 时，说它几乎必然 (以概率为 1) 收敛于一个随机变量 X，记为：$X_n\xrightarrow{\text{a.s.}}X$.

注：等价地，若对 $\forall\epsilon>0$，有 $P(\lim_{n\to\infty}|X_n-X|<\epsilon)=1$，则 $X_n\xrightarrow{\text{a.s.}}X$.

下面介绍另一个 a.s. 收敛的定义.

定理 1.2.4 $X_n\xrightarrow{\text{a.s.}}X$ 当且仅当对 $\forall\epsilon>0$，$\lim_{m\to\infty}P(\sup_{n\geqslant m}|X_n-X|\leqslant\epsilon)=1$.

注：若 $\forall\,\epsilon>0$，$\lim_{n\to\infty}P(|X_n-X|\leqslant\epsilon)=1$，则 $X_n\xrightarrow{P}X$. 由上面定理知几乎必然收敛强于依概率收敛.

定理 1.2.5 (强大数定律) 假设 $X_1,X_2,\cdots,X_n$ 是独立同分布的随机变量序列，且有 $E|X_1|<\infty$, 则当 $n\to\infty$ 时，有

$$\bar{X}_n=\frac{1}{n}\sum_{i=1}^{n}X_i\xrightarrow{\text{a.s.}}E(X_1).$$

注：设 a_n 是 a 的估计，若 $a_n\xrightarrow{\text{a.s.}}a$，则称 a_n 是 a 的强相合估计.

因此，定理 1.2.5 中，$\bar{X}_n$ 是 $E(X_1)$ 的强相合估计.

1.2.3 r 阶收敛

定义 1.2.3 (r 阶中心矩收敛) 对随机变量序列 $\{X_n, n=1,2,\cdots\}$，存在 $r>0$ 有 $E|X_n|^r<\infty$. 若存在一个随机变量 X，使得 $\lim_{n\to\infty}E(|X_n-X|^r)=0$，则称 X_n 依 r 阶中心矩收敛于 (在 L^r 空间)X，记为 $X_n\xrightarrow{\text{r.m.}}X$ 或 $X_n\xrightarrow{L^r}X$.

注: 一般在 $r=2$ 情况下讨论. 此时称其为均方收敛.

定义 1.2.4 (r 阶矩收敛) 对随机变量序列 $\{X_n, n=1,2,\cdots\}$，存在 $r>0$ 有 $E|X_n|^r<\infty$. 若 $\lim_{n\to\infty}E|X_n|^r=E|X|^r$，则 X_n 依 r 阶矩收敛于 X.

1.2.4 依分布收敛

定义 1.2.5 (依分布收敛) 设 $F_n(x)$ 和 $F(x)$ 分别是随机变量序列 X_n 和随机变量 X 的分布函数. 若 $\lim_{n\to\infty}F_n(x)=F(x)$ 对 $F(\cdot)$ 的定义域中的任意连续点都成立，则称随机变量序列 $\{X_n, n=1,2,\cdots\}$ 依分布收敛于分布函数为 $F(x)$ 的随机变量 X，记为 $X_n\xrightarrow{d}X$.

注: (1) 对于依分布收敛，$\{X_n, n=1,2,\cdots\}$ 不需要定义在相同的概率空间. 它不是 $\{X_i\}$ 的收敛，而是由 $\{X_n, n=1,2,\cdots\}$ 导出的概率分布 $\{F_n, n=1,2,\cdots\}$ 的收敛. 可以将其视为在一些概率测度下的弱拓扑的集合的收敛. 因此，文献中经常称依分布收敛为弱收敛. 它“弱”是因为它是可以由其他的一些收敛得到，例如依概率收敛和几乎必然收敛.

(2) 此外，$X_n\xrightarrow{d}X$ 当且仅当对任意在紧集上有界连续的函数 f 有 $Ef(X_n)\to Ef(X)$. 进一步，φ_n 和 φ 分别为 X_n 和 X 的特征函数. $X_n\xrightarrow{d}X$ 当且仅当 $\varphi_n(t)\to\varphi(t)$. 这些结论的证明以及更多的弱收敛的等价定义可以参考文献 (Pollard，1984).

1.2.5 收敛方式间的关系

下面讨论随机变量序列的几种收敛方式之间的关系.

定理 1.2.6 (1) 若 $X_n\xrightarrow{\text{a.s.}}X$，则 $X_n\xrightarrow{P}X$.

(2) 若 $X_n\xrightarrow{r.m.}X$，则 $X_n\xrightarrow{P}X$.

(3) 若 $X_n\xrightarrow{P}X$, 则 $X_n\xrightarrow{d}X$.

证明: (1) 注意到 $\lim_{n\to\infty}P\{|X_n-X|>\epsilon\}\leqslant\lim_{n\to\infty}P\{\bigcup_{k\geqslant n}\{|X_k-X|>\epsilon\}\}=P\{\bigcap_{n=1}^{\infty}\bigcup_{k\geqslant n}\{|X_k-X|>\epsilon\}\}=0$.

(2) 因为对于充分大的 n 有 $\lim_{n\to\infty}E(|X_n-X|^r)=0$，$E(|X_n-X|^r)<\infty$，则由切比雪夫不等式得，对任意 $\epsilon>0$，有 $P(|X_n-X|\geqslant\epsilon)\leqslant E(|X_n-X|^r)/\epsilon^r\to0, n\to\infty$. 因此 $\lim_{n\to\infty}P(|X_n-X|<\epsilon)=1$.

(3) f 是任意有界且一致连续的函数. 令 $M=\sup_x|f(x)|$. 对任意 $\epsilon>0$，选择 δ 使得 $|X_n-X|\leqslant\delta$，有 $|f(X_n)-f(X)|\leqslant\epsilon$. 可以得到 $E|f(X_n)-f(X)|\leqslant\epsilon+2M\times P\{|X_n-X|>$

$\delta\}$. 这样就有 $|Ef(X_n) - Ef(X)| \leqslant E|f(X_n) - f(X)| \leqslant \epsilon + 2MP\{|X_n - X| > \delta\}$. 因为 $X_n \xrightarrow{P} X$，因此可证得 $Ef(X_n) \to Ef(X)$. 从而定理 (c) 结论得证.

1.3　中心极限定理和几个常用的定理

1.3.1　中心极限定理

下面介绍几个关于中心极限定理的著名结果.

定理 1.3.1 (Lindeberg-Levy 中心极限定理)　设 $\{X_i\}_{i=1}^n$ 是均值向量为有限向量 μ、协方差阵为正定阵 Σ 的独立同分布随机向量，则

$$Z_n \equiv \sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(\mathbf{0}, \Sigma),$$

其中 $\bar{X}_n = \dfrac{1}{n}\sum_{i=1}^n X_i$.

下面给出另一个常用的中心极限定理.

定理 1.3.2 (Liapounov 中心极限定理)　设 $\{X_{n,i}\}_{i=1}^n$ 是独立随机变量序列，$E(X_{n,i}) = \mu_{n,i}$ 且 $\mathrm{var}(X_{n,i}) = \sigma_{n,i}^2$. 假设存在 $\delta > 0$，有 $E|X_{n,i}|^{2+\delta} < \infty$. 令 $S_n = \sum_{i=1}^n (X_{n,i} - \mu_{n,i})$，$L_{n,i} = (X_{n,i} - \mu_{n,i})/\sigma_n$，其中 $\sigma_n^2 = \sum_{i=1}^n \sigma_{n,i}^2$. 若存在 $\delta > 0$，使得 $\lim_{n\to\infty}\sum_{i=1}^n E|L_{n,i}|^{2+\delta} = 0$，则有

$$\frac{S_n}{\sigma_n} = \sum_{i=1}^n L_{n,i} \xrightarrow{d} N(0,1).$$

注：上述定理的一个特殊情况是当 $\mu_{n,i} = \mathbf{0}$ 且 $\sigma_{n,i}^2$ 满足 $\lim_{n\to\infty}\dfrac{\sigma_n^2}{n} = \sigma^2$. 若存在 $\delta > 0$，使得 $\lim_{n\to\infty}\sum_{i=1}^n E\left|\dfrac{L_{n,i}}{\sqrt{n}}\right|^{2+\delta} = 0$，则有

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

实际上，对非参数和半参数模型的估计和检验问题，统计量经常表现为双求和的形式，这时经常要用到 U 统计量的中心极限定理，更多的关于 U 统计量的中心极限定理的内容可以参考文献 (Lee，1990).

1.3.2　几个常用的定理

下面列举一些在渐近分析中常用的定理. 下面定理中 $\{X_i\}_{i=1}^n, \{Y_i\}_{i=1}^n$ 为独立同分布随机变量序列.

定理 1.3.3　设 f 是一个连续函数，则有:

(1) 若 $X_n \xrightarrow{\text{a.s}} X$，则 $f(X_n) \xrightarrow{\text{a.s}} f(X)$.
(2) 若 $X_n \xrightarrow{P} X$，则 $f(X_n) \xrightarrow{P} f(X)$.
(3) 若 $X_n \xrightarrow{d} X$，则 $f(X_n) \xrightarrow{d} f(X)$.
特别地，(3) 被称为连续映射定理 (CMT).
定理 1.3.4 若 $X_n \xrightarrow{d} X$ 且 $Y_n \xrightarrow{P} a$，a 是常数，则
(1) $Y_nX_n \xrightarrow{d} aX$；
(2) $X_n + Y_n \xrightarrow{d} X + a$.

1.3.3 Delta 方法

下面介绍 Delta 方法.

定理 1.3.5 (Delta 方法) 映射 $\phi: \mathbf{R}^k \to \mathbf{R}$ 关于 θ 连续可微并且 $\phi'(\theta) \neq 0$. 若 $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, \Sigma)$，则 $\sqrt{n}(\phi(\hat{\theta}_n) - \phi(\theta)) \xrightarrow{d} N(0, [\phi'(\theta)]^{\mathrm{T}}\Sigma\phi'(\theta))$.

证明：由泰勒展开，

$$\phi(\hat{\theta}_n) = \phi(\theta) + [\phi'(\theta)]^{\mathrm{T}}(\hat{\theta}_n - \theta) + o(\|\hat{\theta}_n - \theta\|) = \phi(\theta) + [\phi'(\theta)]^{\mathrm{T}}(\hat{\theta}_n - \theta) + o_p(n^{-\frac{1}{2}}),$$

由此得到：

$$\sqrt{n}(\phi(\hat{\theta}_n) - \phi(\theta)) = [\phi'(\theta)]^{\mathrm{T}}\sqrt{n}(\hat{\theta}_n - \theta) + o_p(1) \xrightarrow{d} N(0, [\phi'(\theta)]^{\mathrm{T}}\Sigma\phi'(\theta)).$$

1.4 记号 $o_p(1)$ 和 $O_p(1)$

在这一节，介绍以后章节中经常要用到的两个十分重要的记号：$o_p(1)$ 和 $O_p(1)$. 这两个记号与数学分析中的 $o(1)$ 和 $O(1)$ 十分相似，它们的运算规律也十分相似. 需要注意的是 $o_p(1)$ 或 $O_p(1)$ 是具有某种大样本性质的随机变量序列.

令 ξ_n 为一个随机变量序列，又设 $n \to \infty$ 表示一个过程. 说 $\xi_n = o_p(1)$ 是指在 n 趋于无穷时 $\xi_n \xrightarrow{P} 0$，或对任意 $\varepsilon > 0$，当 $n \to \infty$ 时，$P\{|\xi_n| \geqslant \varepsilon\} \to 0$.

同时用 $\xi_n = O_p(1)$ 表示 ξ_n 是依概率有界的量，即对任意 $\varepsilon > 0$，存在 $M > 0$，使得

$$P\{|\xi_n| \geqslant M\} < \varepsilon.$$

例：设 $\{X_i\}_{i=1}^n$ 为均值为 μ、方差为 σ^2 的独立随机变量，则：
(1) 若 $\mu = 0$，则 $\sum_{i=1}^n X_i$ 是 $O_p(n^{1/2})$；若 $\mu \neq 0$，则 $\sum_{i=1}^n X_i$ 是 $O_p(n)$.
(2) 若 $\mu = \sigma^2 = 0$ 不成立，则 $\sum_{i=1}^n X_i^2$ 是 $O_p(n)$.

注意：(1) 经常出现在渐近理论中. 若 $\mu = 0$，则由 $\sum_{i=1}^n X_i$ 的均值为 0 且方差为 $n\sigma^2$ 可以得到 $\sum_{i=1}^n \dfrac{X_i}{n^{1/2}}$ 的均值为 0 且方差为 σ^2. 所以若 $\mu = 0$，则 $\sum_{i=1}^n X_i$ 是 $O_p(n^{1/2})$. 然而，若 $\mu \neq 0$，则由 $\sum_{i=1}^n X_i$ 的均值为 $n\mu$，可以得到 $\sum_{i=1}^n X_i$ 是 $O_p(n)$，而不是 $O_p(n^{1/2})$.

另外，(2) 由以下事实得到：X_i^2 均值 $m_1 = \mu^2 + \sigma^2$ 有限且不为 0，由强大数定理可得结论.

注：文献中经常出现"$\sqrt{n}$ 相合"的概念，其定义为：若 $\hat{\theta} - \theta = O_p(n^{-1/2})$，则称 $\hat{\theta}$ 为 θ 的 $\sqrt{n}$ 相合估计. 若 $\hat{\theta}$ 有分解：$\hat{\theta} = \theta + n^{-1/2}\sum_{i=1}^n \xi_i + o_p(n^{-1/2})$，其中 $\xi_i, i = 1, 2, \cdots, n$ 独立同分布且 $E(\xi_1) = 0$. 上式第二个部分是 $O_p(n^{-1/2})$. 因此有 $\hat{\theta} - \theta = O_p(n^{-1/2})$，这时 $\hat{\theta}$ 为 θ 的 $\sqrt{n}$ 相合估计.

关于 $o_p(1)$ 和 $O_p(1)$，它们具有下列性质：

$$o_p(1) + o_p(1) = o_p(1),$$

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

$$o_p(1) \cdot o_p(1) = o_p(1),$$

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

$$o_p(1) + c = O_p(1)\ (c \neq 0).$$

另外，有两个类似的符号 $o(1)$ 和 $O(1)$. 其中 $o(1)$ 表示一个关于 1 的高阶无穷小量，也就是说表示一个极限为 0 的量；$O(1)$ 表示一个有界量.

举例：

(1) -4 是 $O(1)$.

(2) $6n^3$ 是 $O(n^3)$，$o(n^4)$，$o(n^5)$.

(3) $\dfrac{7}{n}$ 是 $O(n^{-1})$，同时也是 $o(1)$.

(4) $\dfrac{5}{n} - \dfrac{3}{n^{3/2}}$ 是 $O(n^{-1})$.

在运算中，$o_p(1)$ 和 $O_p(1)$ 与通常的 $o(1)$ 和 $O(1)$ 一起使用，具有下面的一些结果：

$$o_p(1) + o(1) = o_p(1),$$

$$o_p(1) + O(1) = O_p(1),$$

$$O_p(1) + o(1) = O_p(1),$$

等等. 但注意

$$o_p(1) + o(1) \neq o(1).$$

上面的等式之所以不成立，其原因是左边为随机变量序列，而右边为数列，两者性质不同.

$o_p(1)$ 和 $O_p(1)$ 还有下面的性质：

设 $\xi_n \xrightarrow{d} F$，则 $\xi_n = O_p(1)$，$\xi_n + o_p(1) \xrightarrow{d} F$，$\xi_n \cdot o_p(1) = o_p(1)$.

除了 $o_p(1)$ 和 $O_p(1)$ 外，有时还用记号 $o_p(\xi_n)$ 和 $O_p(\xi_n)$. $o_p(\xi_n)$ 表示随机变量序列 $\xi_n o_p(1)$，$O_p(\xi_n)$ 表示随机变量序列 $\xi_n O_p(1)$.