第五章 大数定律及中心极限定理

本章要解决的问题

ANSWER

1. 为何能以某事件发生的频率作为该事件的 概率的估计？
2. 为何能以样本均值作为总体期望的估计？

大数定律

3. 为何正态分布在概率论中占有极其重要的地位？
4. 大样本统计推断的理论基础是什么？

中心极限定理

- 大数定律和中心极限定理是现代概率论、统计学的重要基本理论，它们揭示了随机现象的重要统计规律，在概率论与数理统计的理论研究和实际应用中都具有重要的意义。
  - 大数定律与中心极限定理一起，成为理论科学和社会科学的基石
- 所谓大数定律，简单地说，就是大量的随机变量所呈现出的规律，这种规律一般用随机变量序列的某种收敛性来刻画。
  - 虽然大数定律的表述和证明都依赖现代数学知识，但其结论最早出现在微积分出现之前；而且在生活中，即使没有微积分的知识也可以应用。
  - 迄今为止,人们已发现很多大数定律(laws of large numbers)。本章仅介绍几个最基本的大数定律。

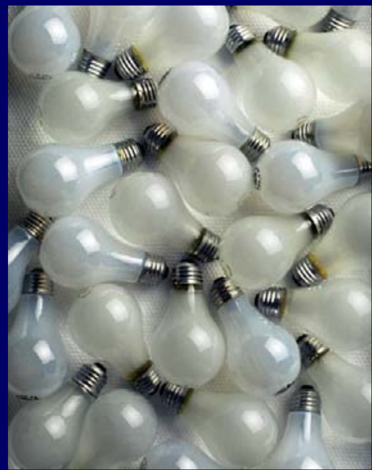
第一节 大数定律

- 大数定律
- 依概率收敛定义及性质
- 小结

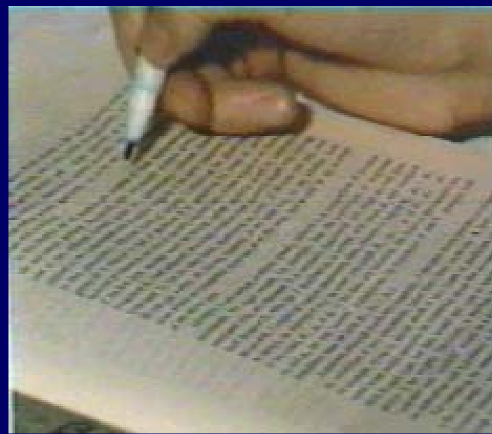
大数定律的客观背景

大量随机试验中 $\begin{cases} \text{事件发生的频率稳定于某一常数} \\ \text{测量值的算术平均值具有稳定性} \end{cases}$

大量抛掷硬币
正面出现频率



生产过程中的
废品率



字母使用频率

……

一、依概率收敛定义及性质

定义　设$Y_1,Y_2,\cdots Y_n,\cdots$是一个随机变量序列，$a$是一个常数.若对于任意正数$\varepsilon$，有

$$\lim_{n\to\infty} P\{|Y_n - a| < \varepsilon\} = 1$$

则称序列$Y_1,Y_2,\cdots Y_n,\cdots$依概率收敛于$a$.记为

$$Y_n \xrightarrow{P} a.$$

性质　设$X_n \xrightarrow{P} a$，$Y_n \xrightarrow{P} b$,又设函数$g(x,y)$在点(a,b)连续，则$g(X_n,Y_n) \xrightarrow{P} g(a,b)$.

请注意：

$\{X_n\}$依概率收敛于a，意味着对任意给定的$\varepsilon > 0$，当n充分大时，事件$|X_n - a| < \varepsilon$的概率很大，接近于1；并不排除事件$|X_n - a| \geq \varepsilon$的发生，而只是说他发生的可能性很小.

依概率收敛比高等数学中的普通意义下的收敛弱些，它具有某种不确定性.

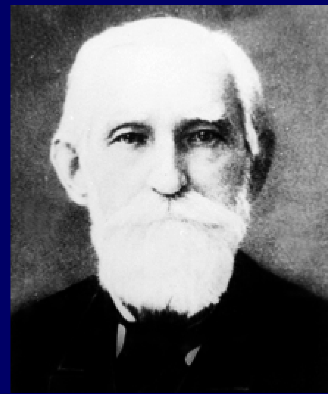
二、切比雪夫大数定理

【切比雪夫大数定理*】

设$x_1, x_2, \cdots, x_n$是一列两两不相关的随机变量，他们分别存在期望$E(x_k)$和方差$D(x_k)$。若存在常数C使得：$D(x_k) \le C \ (k=1,2,\cdots,n)$

则对任意小的正数ε，满足公式一：

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n} x_k - \frac{1}{n}\sum_{k=1}^{n} E(x_k)\right| < \varepsilon\right) = 1$$



切比雪夫

- 将该公式应用于抽样调查，就会有如下结论：随着样本容量n的增加，样本平均数将接近于总体平均数。从而为统计推断中依据样本平均数估计总体平均数提供了理论依据。

【注意】切比雪夫大数定理并未要求$x_1,x_2,\ldots,x_n$同分布，相较于后面介绍的伯努利大数定律和辛钦大数定律更具一般性。

定理1（切比雪夫大数定理的特殊情况）

设随机变量$X_1, X_2,\cdots, X_n,\cdots$相互独立，且具有相同的数学期望和方差：

$$E(X_k)=\mu, D(X_k)=\sigma^2 (k=1,2,\cdots).$$

做前 n 个随机变量的算术平均 $\bar{X}=\frac{1}{n}\sum_{k=1}^{n} X_k$

则对任意的$\varepsilon>0$，有

$$\lim_{n\to\infty} P\{|\bar{X}-\mu|<\varepsilon\}$$

$$=\lim_{n\to\infty} P\{|\frac{1}{n}\sum_{i=1}^{n} X_i-\mu|<\varepsilon\}=1$$

【证】由于

$$E\left(\frac{1}{n}\sum_{k=1}^{n}X_k\right)=\frac{1}{n}\sum_{k=1}^{n}E(X_k)=\frac{1}{n}\cdot n\mu=\mu$$

$$D\left(\frac{1}{n}\sum_{k=1}^{n}X_k\right)=\frac{1}{n^2}\sum_{k=1}^{n}D(X_k)=\frac{1}{n^2}\cdot n\sigma^2=\frac{\sigma^2}{n}$$

由切比雪夫不等式

$$P\left\{\left|\frac{1}{n}\sum_{k=1}^{n}X_k-\mu\right|<\varepsilon\right\}\geq 1-\frac{\sigma^2/n}{\varepsilon^2}$$

上式中令 $n\to\infty$ 得

$$\lim_{n\to\infty}P\{|\frac{1}{n}\sum_{i=1}^{n}X_i-\mu|<\varepsilon\}=1$$

【说明】

1、定理中$\{|\frac{1}{n}\sum_{i=1}^{n}X_i-\mu|<\varepsilon\}$是指一个随机事件，当$n\to\infty$时，这个事件的概率趋于1.

2、定理以数学形式证明了随机变量$X_1,\cdots X_n$的算术平均$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$接近数学期望E（$X_k$）$=\mu$（$k=1,2\cdots,n$），这种接近说明其具有的稳定性.

这种稳定性的含义说明算术平均值在依概率收敛的意义下逼近某一常数.

定理1的另一种叙述形式

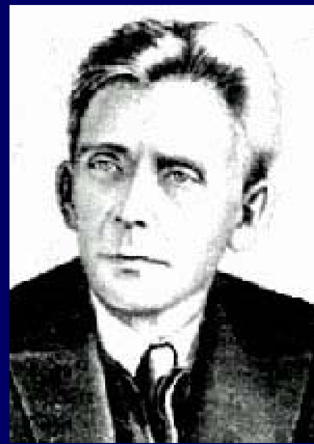
设随机变量X_1，$X_2, \cdots, X_n, \cdots$相互独立，且具有相同的数学期望和方差：$E(X_k) = \mu, D(X_k) = \sigma^2$ $(k = 1,2,\cdots)$，则序列$\overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$依概率收敛于$\mu$，即

$$\overline{X} \xrightarrow{P} \mu.$$

三、辛钦大数定律

下面给出的独立同分布下的大数定律，

不要求随机变量的方差存在.


辛钦

定理2　弱大数定律（辛钦大数定律）

设随机变量序列$X_1, X_2, \ldots$ 相互独立，服从同一分布，具有数学期$E(X_i)=\mu$，$i=1,2,\ldots$，

序列 $\overline{X}=\frac{1}{n}\sum_{k=1}^{n}X_k$ 依概率收敛于 μ，即 $\overline{X}\xrightarrow{P}\mu$.

【练习】设随机变量X_1，X_2，…，X_n独立同分布，

且 $E(X_i^k)=\mu_k$（$i=1$，2，…，n）存在，

令 $A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k$ （$k>1$）,则 $A_k \xrightarrow{P} \mu_k,(n\to\infty)$

证：因为X_1，X_2，…，X_n独立同分布，

所以 $X_1^k,X_2^k,\ldots,X_n^k$ 独立同分布。

又 $E(X_i^k)=\mu_k$存在，由辛钦大数定律：

$$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k \xrightarrow{P} \mu_k$$

注　1、辛钦大数定律为寻找随机变量的期望值提供了一条实际可行的途径.

2、辛钦定理具有广泛的适用性.

要估计某地区的平均亩产量，要收割某些有代表性块，例如n 块地．计算其平均亩产量，则当n 较大时，可用它作为整个地区平均亩产量的一个估计.

例　在一个罐子中,装有10个编号为0-9的同样的球,从罐中有放回地抽取若干次，每次抽一个，并记下号码.

设 $X_k = \begin{cases} 1 & \text{第}k\text{次取到号码}0 \\ 0 & \text{否则} \end{cases}$，$k=1,2,\ldots$

问对序列$\{X_k\}$能否应用大数定律?

解: $X_k \sim \begin{Bmatrix} 1 & 0 \\ 0.1 & 0.9 \end{Bmatrix}$, $E(X_k)=0.1$, $k=1,2,\ldots$

诸X_k 独立同分布，且期望存在，故能使用大数定律. 即对任意的$\varepsilon>0$,

$$\lim_{n\to\infty} P\{|\frac{1}{n}\sum_{k=1}^{n} X_k - 0.1| < \varepsilon\} = 1$$

例4 设$X_1, X_2, \cdots, X_n$独立同分布,且$EX_i = 0$,则
$\lim\limits_{n\to\infty} P(\sum_{i=1}^{n} X_i < n) =$ ______.

解: 由辛钦大数定律(取ε=1)有:

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - 0\right| < 1\right) = \lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right| < 1\right) = 1,$$

又显然有: $\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right| < 1\right) \subset \left(\sum_{i=1}^{n} X_i < n\right)$,故

$$\lim_{n\to\infty} P\left(\sum_{i=1}^{n} X_i < n\right) \geq \lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right| < 1\right) = 1.$$

从而有 $\lim\limits_{n\to\infty} P(\sum_{i=1}^{n} X_i < n) = 1$,即应填1.

四、伯努利大数定律

伯努利大数定律是切比雪夫大数定理的特殊情况.

伯努利

【问题】

设n_A是n重贝努里试验中事件A发生的次数，p是事件A发生的概率，

$\frac{n_A}{n}$ 是事件A发生的频率.

事件发生的频率能否代替事件的概率，频率是否具有稳定性呢？

抛硬币实验的数学意义

我们用 v_n 表示抛硬币 n 次中出现正面的次数，ε 是任意小的一个正数，譬如 $\varepsilon=0.001$，则

$$\lim_{n\to+\infty} P\left\{\left|\frac{v_n}{n}-\frac{1}{2}\right|\geq\varepsilon\right\}=0$$

频率不一定恰好就是 $\frac{1}{2}$，有细微偏差，

但是与 $\frac{1}{2}$ 的偏差超过 ε 的可能性趋于零。

定理3 贝努里大数定律

设 n_A 是n次独立重复试验中事件A发生的次数，p是事件A在每次试验中发生的概率，则对于任意正数$\varepsilon > 0$，有

$$\lim_{n\to\infty} P\{|\frac{n_A}{n} - p| < \varepsilon\} = 1$$

或

$$\lim_{n\to\infty} P\{|\frac{n_A}{n} - p| \geq \varepsilon\} = 0$$

证明　因为$n_A \sim b(n,p)$,由此可表示为

$$n_A = X_1 + X_2 + \cdots + X_n$$

其中$X_1, X_2, \cdots, X_n$相互独立，且都服从以p为参数的（0−1）分布。因而$E(X_k) = p$，$D(X_k) = p(1-p)$,

由定理1即得

$$\lim_{n\to\infty} P\{|\frac{n_A}{n}-p|<\varepsilon\}$$

$$=\lim_{n\to\infty} P\{|\frac{1}{n}(X_1+X_2+\cdots+X_n)-p|<\varepsilon\}=1$$

或 $$\lim_{n\to\infty} P\{|\frac{n_A}{n}-p|\geq\varepsilon\}=0$$ 证毕

注 贝努里大数定律表明，当重复试验次数n充分大时，事件A发生的频率n_A/n与事件A的概率p有较大偏差的概率很小.

事件发生的频率可以代替事件的概率.

说明:

(1) 切比雪夫弱大数定律和辛钦弱大数定律的条件是不同的，但它们都可以推导出伯努利大数定律.

- 切比雪夫弱大数定律里随机变量序列不要求是同分布的，但是要求它们的方差有一致的上界。

- 辛钦弱大数定律里随机变量序列是同分布的，但不要求它们的方差存在或有一致上界。

大数定律在计算机中的实际应用——蒙特卡罗方法(Monte Carlo method)

- 也称统计模拟方法，是二十世纪四十年代中期由于科学技术的发展和电子计算机的发明，而被提出的一类非常重要的数值计算方法。

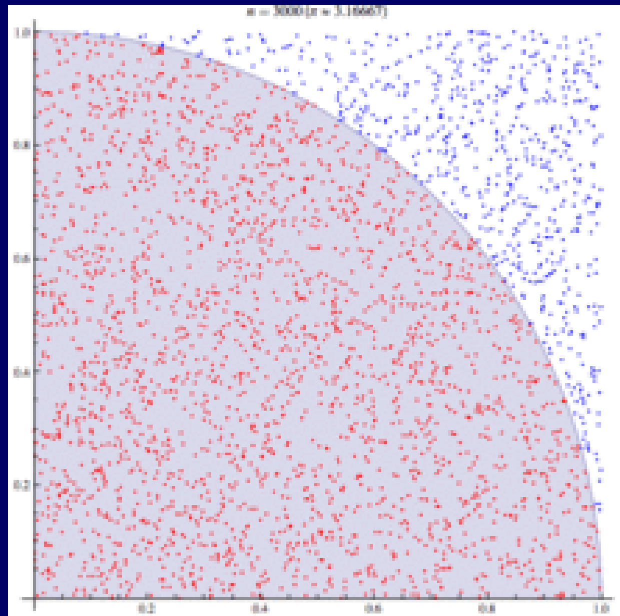
- 指构造符合一定规则的随机数（或更常见的伪随机数）来解决很多计算问题的一类方法。

- 对于那些由于计算过于复杂而难以得到解析解或者根本没有解析解的问题，蒙特卡罗方法是一种有效的求出数值解的方法。

【例】计算一个不规则图形的面积，图形的不规则程度和分析性计算（比如，积分）的复杂程度是成正比的。

- 蒙特卡罗方法的思想：假想你有一袋豆子，把豆子均匀地朝这个图形上撒，然后数这个图形之中有多少颗豆子，这个豆子的数目就是图形的面积。当你的豆子越小，撒的越多的时候，结果就越精确。

- 借助计算机程序可以生成大量均匀分布坐标点，然后统计出图形内的点数，通过它们占总点数的比例和坐标点生成范围的面积就可以求出图形面积。



使用蒙特卡罗方法估算π值. 放置30000个随机点后,π的估算值与真实值相差0.07%.

五、小结

大数定律

伯努利大数定律 $\lim_{n\to\infty} P\{|\frac{n_A}{n} - p| < \varepsilon\} = 1$ $n_A \sim b(n, p)$

切比雪夫大数定律 $\lim_{n\to\infty} P\{|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu| < \varepsilon\} = 1$ 独立 $\begin{cases} E(X_k) = \mu \\ D(X_k) = \sigma^2 \end{cases}$

辛钦大数定律 $\lim_{n\to\infty} P\{|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu| < \varepsilon\} = 1$ 独立同分布 $E(X_k) = \mu$

大数定律以严格的数学形式表达了随机现象最根本的性质之一：

平均结果的稳定性

在概率论发展初期，由于概率的数学定义尚未明确，所以缺乏理解概率收敛的理论基础，故把频率“趋于”概率视为经大量试验而得到的结果，就象物理学中的定律一样. 在概率论的公理化体系建立以后，大数定律可在理论上进行严格的证明而成为意义明确的定理，故现在教材上称为“大数定理”.

为什么叫“大数定律”
而不叫“大数定理”?

切比雪夫（1821 ~ 1894）

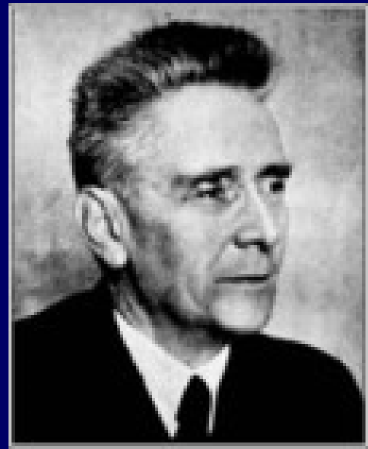
- 俄罗斯数学家、力学家，彼得堡数学学派的奠基人和领袖
  - 出身于贵族家庭，祖辈中许多人立过战功；排行第二，左脚生来有残疾，因而童年时代经常独坐家中，养成了在孤寂中思索的习惯。
  - 终生未娶，生活十分简朴，积蓄全部用来买书和制造机器。他最大的乐趣是与年轻人讨论数学问题。
  - 切比雪夫是土生土长的学者，而且以他自己的卓越才能和独特的魅力吸引了一批年轻的俄国数学家，形成了一个具有鲜明风格的数学学派，从而使俄罗斯数学摆脱了落后境地而开始走向世界前列。
    - 背景：19世纪以前俄罗斯没有自己的数学家，没有大学，甚至没有一部象样的初等数学教科书。彼得堡科学院中，早期数学方面的院士都是外国人。19世纪上半叶，俄国才开始出现了罗巴切夫斯基、布尼亚科夫斯基和奥斯特罗格拉茨基这样优秀的数学家，但是多数是在外国接受训练的。

【学术成就】

- 一生发表了70多篇科学论文，内容涉及**数论**、**概率论**、**函数逼近论**、**积分学**等方面。他证明了贝尔特兰公式，自然数列中素数分布的定理，大数定律的一般公式以及中心极限定理。他不仅重视纯数学，而且十分重视数学的应用。

返回

辛钦Khinchin（1894～1959）



- 前苏联数学家和数学教育家，现代概率论的奠基者之一， 莫斯科概率学派的开创者。

【学术成就】

- 辛钦在函数的度量理论、分析学、数论、概率论、信息论及对统计学力学的应用等方面都有重要贡献。
- 共发表150多种关于数学和数学史论著。在数学中以他的姓氏命名的有: 辛钦定理、辛钦不等式、辛钦积分、辛钦条件、辛钦可积函数、辛钦转换原理、辛钦单峰性准则等等，而其中以他的姓氏命名的定理有多种。
- 他十分重视数学教育和人才的培养，潜心的编著了多本思路清晰、引人入胜、突出论题本质风格的教材和专著，如：《数学分析八讲》、《数学分析简明教程》、《连分数》、《费马定理》、《公用事业理论的数学方法》等。

雅各布·伯努利Bernoulli （1654～1705）

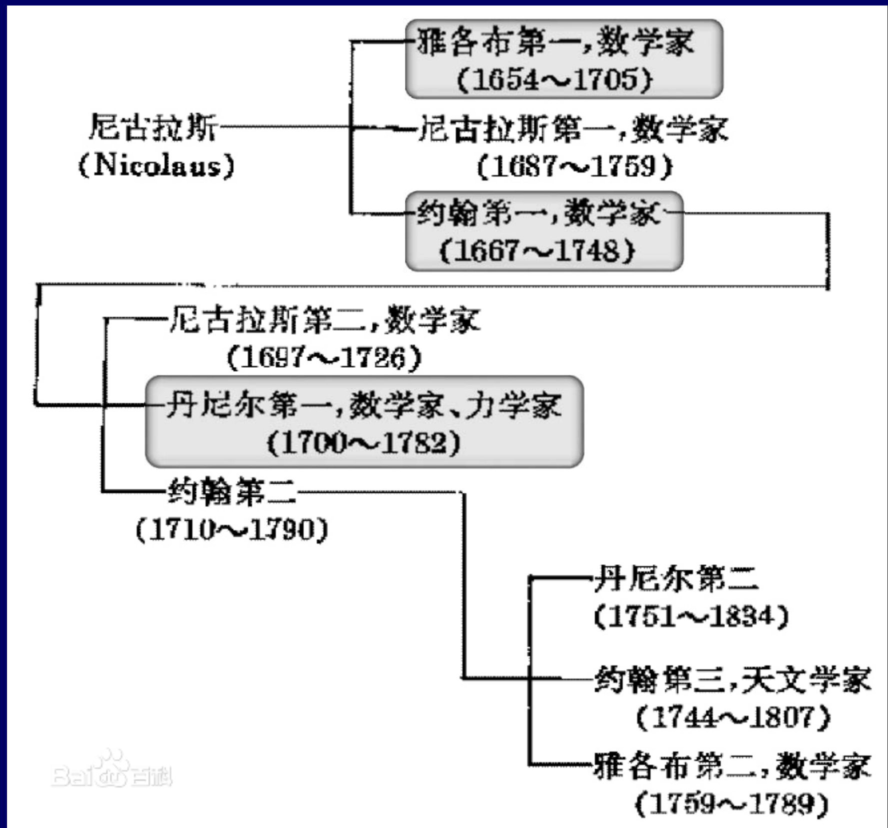
- 伯努利家族代表人物之一，瑞士数学家，被公认的概率论的先驱之一。
  - 他是最早使用“积分”这个术语的人，也是较早使用极坐标系的数学家之一。
  - 出身于一个商人世家，获艺术硕士学位后遵照父亲的意愿又取得神学硕士学位，但他不顾父亲的反对，自学了数学和天文学，成为巴塞尔大学的数学教授。

【学术成就】

- 在数学上的贡献涉及微积分、微分方程、无穷级数求和、解析几何、概率论以及变分法等领域。
- 对数学的最突出的贡献是在概率论和变分法这两个领域中。 他在概率论方面的工作成果包含在他的论文《推测的艺术》之中。
- 一生最有创造力的著作就是 1713年出版的《猜度术》，是组合数学及概率论史的一件大事，他在这部著作中给出的伯努利数有很多应用；提出了概率论中的“伯努利定理”，这是大数定律的最早形式。

伯努利Bernoulli家族

- 瑞士的伯努利家族3代人中产生了8位科学家，出类拔萃的至少有3位；而在他们一代又一代的众多子孙中，至少有一半相继成为杰出人物。



欧拉和高斯之间差了几个伯努利?

- 莱昂哈德·欧拉，生于1707年4月15日；
  约翰·卡尔·弗里德里希·高斯，生于1777年4月30日；
- **雅各布·伯努利**，生于1654年12月27日；
  约翰·伯努利，生于1667年7月27日；
  尼古拉一世·伯努利，生于1687年10月21日；
  小尼古拉·伯努利，生于1695年；
  丹尼尔·伯努利，生于1700年2月9日；

• 故欧拉跟高斯之间一个出名的伯努利都没有。但是，他俩之间差着五个不出名的伯努利

- 分别是Johann III.(1744-1807)、Daniel II.(1751-1834) 、 Nikolaus III.(1754-1781) 、Jakob II.(1759-1809) 、Emanuel IV.(1776-1844)。