第五章
大数定律与中心极限定理

如何得到某人的投篮命中率？

如何得到某人的投篮命中率？

- 投篮n 次（ n很大），命中的频率可作为“投篮命中率”的估计值.

频率具有稳定性

试验者	抛掷次数 n	正面次数 μ_n	频 率 $\frac{\mu_n}{n}$
德·摩根	2048	1061	0.5181
浦 丰	4040	2048	0.5069
费 勒	10000	4979	0.4979
皮尔逊	12000	6019	0.5016
杰万斯	20480	10379	0.5068
皮尔逊	24000	12012	0.5005
罗曼诺夫斯基	80640	39699	0.4923

频率具有稳定性

设 f_A 是n次抛掷中正面出现（A）的次数，每次试验中 $P(A) = p=0.5$.

试验者	抛掷次数 n	正面次数 μ_n	频率 $\frac{\mu_n}{n}$
德·摩根	2048	1061	0.5181
浦丰	4040	2048	0.5069
费勒	10000	4979	0.4979
皮尔逊	12000	6019	0.5016
杰万斯	20480	10379	0.5068
皮尔逊	24000	12012	0.5005
罗曼诺夫斯基	80640	39699	0.4923

频率具有稳定性

设 f_A 是n重伯努利试验中事件A出现的次数，每次试验中 $P(A) = p$.

试验者	抛掷次数 n	正面次数 μ_n	频率 $\frac{\mu_n}{n}$
德·摩根	2048	1061	0.5181
浦丰	4040	2048	0.5069
费勒	10000	4979	0.4979
皮尔逊	12000	6019	0.5016
杰万斯	20480	10379	0.5068
皮尔逊	24000	12012	0.5005
罗曼诺夫斯基	80640	39699	0.4923

伯努利大数定理 设f_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

$$\lim_{n\to\infty} P\left\{\left|\frac{f_A}{n}-p\right|\geq\varepsilon\right\}=0 \qquad (1.2)$$

伯努利大数定理　设f_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

$$\lim_{n\to\infty} P\left\{\left|\frac{f_A}{n}-p\right|<\varepsilon\right\}=1 \qquad (1.2)$$

或

$$\lim_{n\to\infty} P\left\{\left|\frac{f_A}{n}-p\right|\geq\varepsilon\right\}=0 \qquad (1.2)$$

- 伯努利大数定律构成了概率的统计定义的理论基础，概率论理论体系由此获得了首尾呼应.

- 伯努利大数定律构成了概率的统计定义的理论基础，概率论理论体系由此获得了首尾呼应.
- 雅各布·伯努利一生最有创造的著作就是1713年出版的《猜度术》，在这部著作中，他提出了概率论中的“伯努利定理”，该定理是“大数定律”的最早形式.
- 在大数定律诞生两百周年的1913年，圣彼得堡科学院举行了隆重的纪念会.

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

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

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

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

依概率收敛的序列还有以下性质

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

$$g(X_n, Y_n) \xrightarrow{P} g(a,b).$$

在实践中，人们认识到

- 事件发生的频率具有稳定性，
- 大量测量值的算术平均也具有稳定性，

弱大数定理（辛钦大数定理）: 设 $X_1, X_2, \cdots$ 是相互独立，服从同一分布的随机变量序列，且具有数学期望 $E(X_k)=\mu, k=1,2,\cdots,$ 则对于任意 $\varepsilon > 0$ ，有

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

- 辛钦大数定律提供了求随机变量数学期望的近似值的方法；
- 设想对随机变量X独立重复地观察n次，第k次观察值为X_k，则$X_1, X_2, \ldots X_n$是相互独立的，且它们的分布与X的分布相同，按照辛钦大数定律，当n足够大时，

$$\frac{\sum_{k=1}^{n} X_k}{n}$$很可能接近EX.

小结

在实践中，人们认识到

- 事件发生的频率具有稳定性，
- 大量测量值的算术平均也具有稳定性，

这种稳定性就是大数定律的客观背景.

第二节 中心极限定理

◆误差的产生由大量微小的相互独立的随机因素叠加而成的.

第二节 中心极限定理

- 误差的产生由大量微小的相互独立的随机因素叠加而成的.
- 加工机械轴与规定要求有一定的误差，这是因为在加工时受到许多随机因素的影响：
  - 在机床方面有机床震动与转速的影响;
  - 在刀具方面有装配与磨损的影响;
  - 在操作方面有注意力集中程度、当天情绪的影响;
  - 在环境方面有车间的温度、湿度、照明等的影响
  - ……

设 $\{X_n\}$ 为独立随机变量序列，记其和为

$$\sum_{k=1}^{n} X_k$$

设 $\{X_n\}$ 为独立随机变量序列，记其和为

$$\sum_{k=1}^{n} X_k$$

- 讨论独立随机变量和的极限分布,
- 指出极限分布为正态分布.

定理一 (独立同分布的中心极限定理)

设随机变量$X_1,X_2,\dots,X_n,\dots$相互独立, 服从同一分布, 且具有数学期望和方差$E(X_k)=\mu$, $D(X_k)=\sigma^2$, (k=1,2,…). 则随机变量之和$X_1+X_2+\dots+X_n$的标准化变量设为Y_n,

$$Y_n=\frac{\sum_{k=1}^{n}X_k-E\left(\sum_{k=1}^{n}X_k\right)}{\sqrt{D\left(\sum_{k=1}^{n}X_k\right)}}=\frac{\sum_{k=1}^{n}X_k-n\mu}{\sqrt{n}\sigma}$$

2,…),则随机变量之和 $\sum_{k=1}^{n} X_k$ 的标准化变量

$$Y_n = \frac{\sum_{k=1}^{n} X_k - E\left(\sum_{k=1}^{n} X_k\right)}{\sqrt{D\left(\sum_{k=1}^{n} X_k\right)}} = \frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma}$$

的分布函数 $F_n(x)$ 对于任意 x 满足

$$\lim_{n\to\infty} F_n(x) = \lim_{n\to\infty} P\left\{\frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma} \leqslant x\right\}$$

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

$$Y_n = \frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma} \overset{\text{近似地}}{\sim} N(0,1). \quad (2.2)$$

如果一个随机变量是由大量相互独立的随机因素的综合影响所造成，而每一个别因素对这种综合影响中所起的作用不大．则这种随机变量一般都服从或近似服从正态分布．

例1　一加法器同时收到20个噪声电压 $V_k(k=1,2,\cdots,20)$, 设它们是相互独立的随机变量, 且都在区间(0,10)上服从均匀分布, 记 $V=V_1+V_2+\cdots+V_{20}$, 求 $P(V>105)$的近似值.

例1 一加法器同时收到20个噪声电压 $V_k(k=1,2,\cdots,20)$, 设它们是相互独立的随机变量, 且都在区间(0,10)上服从均匀分布, 记 $V=V_1+V_2+\cdots+V_{20}$, 求$P(V>105)$的近似值.

易知$E(V_k)=5, D(V_k)=100/12 \quad (k=1,2,\cdots 20)$

由中心极限定理知,

$$V=\frac{\sum_{k=1}^{20} V_k - 20\times 5}{\sqrt{\frac{100}{12}\times 20}} \overset{\text{近似地}}{\sim} N(0,1)$$

$$P\{V>105\}=p\left\{\frac{V-20\times 5}{\left(\sqrt{100/12}\right)\sqrt{20}}>\frac{105-20\times 5}{\left(\sqrt{100/12}\right)\sqrt{20}}\right\}$$

$$=p\left\{\frac{V-20\times 5}{\left(\sqrt{100/12}\right)\sqrt{20}}>0.387\right\}$$

$$=1-p\left\{\frac{V-20\times 5}{\left(\sqrt{100/12}\right)\sqrt{20}}\le 0.387\right\}$$

$$\approx 1-\Phi(0.387)=0.348$$

定理一 (独立同分布的中心极限定理)

设随机变量$X_1,X_2,\dots,X_n,\dots$相互独立, 服从同一分布, 且具有数学期望和方差$E(X_k)=\mu$, $D(X_k)=\sigma^2$, (k=1,2,...). 则随机变量之和$X_1+X_2+\dots+X_n$的标准化变量设为Y_n,

$$Y_n=\frac{\sum_{k=1}^{n}X_k-E\left(\sum_{k=1}^{n}X_k\right)}{\sqrt{D\left(\sum_{k=1}^{n}X_k\right)}}=\frac{\sum_{k=1}^{n}X_k-n\mu}{\sqrt{n}\sigma}$$

定理三(棣莫弗-拉普拉斯定理)

设随机变量η_n(n=1,2,...)服从参数为n,p(0<p<1)的二项分布, 则有

$$Y_n=\frac{\eta_n-np}{\sqrt{np(1-p)}}\overset{\text{近似地}}{\sim}N(0,1).\quad(2.5)$$

例2 一船舶在某海区航行, 已知每遭受一次波浪的冲击, 纵摇角大于3^o的概率为$p=1/3$, 若船舶遭受了90000次波浪冲击, 问其中有29500~30500次纵摇角大于3^o的概率是多少?

例2 一船舶在某海区航行, 已知每遭受一次波浪的冲击, 纵摇角大于3^{o}的概率为p=1/3, 若船舶遭受了90000次波浪冲击, 问其中有29500~30500次纵摇角大于3^{o}的概率是多少?

解 将船舶每遭受一次波浪冲击看作是一次试验, 并假定各次试验是独立的.

在90000次波浪冲击中纵摇角度大于3^{o}的次数记为X,

$$X \sim b(90000, 1/3).$$

小结

小结

（1）在实践中，人们认识到

- 事件发生的频率具有稳定性，
- 大量测量值的算术平均也具有稳定性，

这种稳定性就是大数定律的客观背景.

(2) 中心极限定理讨论独立随机变量和的极限分布, 并指出极限分布为正态分布.

- **126页**

  4 ； 8