第五章 大数定律及中心极限定理

§1 大数定律

§1 大数定律

由频率到概率

讨论频率与概率间的关系

三人成虎——

庞葱与太子质于邯郸，谓魏王曰：“今一人言市有虎，王信之乎？”王曰：“否。”“二人言市有虎，王信之乎？”王曰：“寡人疑之矣。”“三人言市有虎，王信之乎？”王曰：“寡人信之矣。”庞葱曰：“夫市之无虎明矣，然而三人言而成虎。”

——《战国策·魏策二》

- 测量一个长度为a的物体，一次测量，结果未必等于a
- 测量多次，结果的计算平均值也未必等于a
- 测量次数很大时，算术平均值接近于a
- 这种现象为平均结果的稳定性
- 大量随机现象中的平均结果与个别随机现象无关，几乎不再随机。

思考：有一把最小刻度为毫米的直尺，如何测量一个直径小于大约为0.5毫米铜线的直径

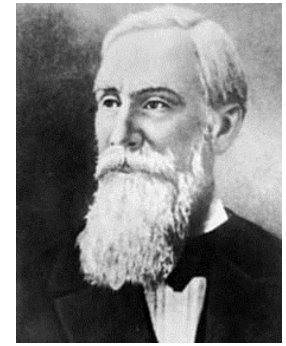
英语考试改革的数学思考

大数定律——频率逼近概率(频率稳定性)的理论支撑

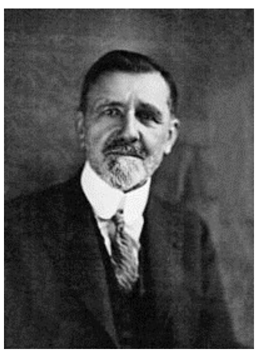
Gerolamo Cardano (1501-1576)未加证明地提出这一概念

Jacob Bernoulli (1654-1705)于1689年证明了二项分布的情况

Siméon Denis Poisson(1781-1840)，命名为"大数定律"

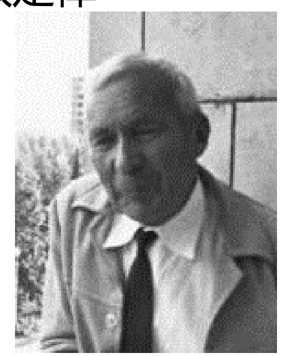


Pafnuty Lvovich Chebyshev (1821-1894)，用其不等式证明了这一定律



Émile Borel (1871-1956)

Andrey Andreyevich Markov (1856-1922)



Andrey Nikolaevich Kolmogorov (1903-1987)

Aleksandr Khinchin (1894-1959)最终给出完整的大数定律

随机事件A在n次试验中发生的频率

$$f_n(A) = \frac{f_n}{n}$$

当试验次数n无限增大时，一般来讲，它与事件A的概率$P(A)$是越来越接近的。

但是，这种“接近” 并不能直接用数学分析中的收敛性来表达

$$\cancel{\lim_{n\to\infty} f_n(A) = P(A)}$$

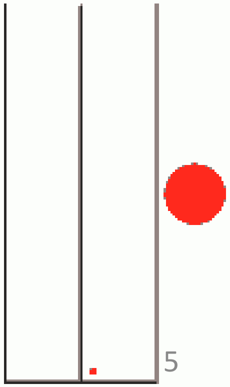
上面的公式等价于：

对任给$\epsilon > 0$，存在一个自然数N，使得n>N时有

$$|f_n(A) - P(A)| < \epsilon$$

或等价地有

$$P(A) - \epsilon < f_n(A) < P(A) + \epsilon$$



以均匀硬币为例，$P(A) = 0.5$。即使取$\epsilon = 0.1$，和很大的自然数N=10000（当然也可以更大），如果上述收敛性成立，则应该有

$$|f_n(A) - 0.5| < 0.1$$

但我们知道即使是10000次都是正面或反面的概率也不为零，此时则

$$f_n(A) = 1$$

或

$$f_n(A) = 0$$

因此上述假设不成立

依概率收敛

定义： 设$\{X_n, n = 1,2,\dots\}$是一随机变量序列，X是一随机变量，如果对于任意给定的$\epsilon > 0$ ，有

$$\lim_{n\to\infty} P\{|X_n - X| < \epsilon\} = 1$$

则称随机变量序列$\{X_n\}$依概率收敛于随机变量X，记作

$$X_n \xrightarrow{P} X$$

大数定律的一般形式

设$\{X_n, n = 1,2,\ldots\}$是一个随机变量序列，而且对每个n，$E(X_n)$存在，如果对于任意给定的$\epsilon > 0$ ，有

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

则称随机变量序列$\{X_n\}$服从大数定律．

等价形式：

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} E(X_k)\right| > \epsilon\right\} = 0$$

即$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{P} \frac{1}{n}\sum_{k=1}^{n} E(X_k)$

定理1 (Chebyshev大数定律)

设随机变量序列$\{X_n\}$满足如下条件：

(1) $X_1, X_2, \dots, X_n, \dots$两两相互独立；

(2) 对每一个$n(n=1,2,\dots)$，$D(X_n)$存在；

(3) 数列$\{D(X_n)\}$有界，即存在常数c，使得对于任意的$n(n=1,2,\dots)$，有$D(X_n)\le c$，则对任意$\epsilon>0$，有

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

即随机变量序列$\{X_n\}$ 服从大数定律

*Chebyshev大数定律揭示了关于算术平均值的稳定性

证明：由于$X_1, X_2, \dots, X_n, \dots$两两相互独立，故有

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

又因为

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

于是由切比雪夫不等式有

$$P\left\{\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} E(X_k)\right| \ge \epsilon\right\} \le \frac{1}{\epsilon^2} D\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) \le \frac{c}{n\epsilon^2}$$

即

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} E(X_k)\right| < \epsilon\right\} \ge \lim_{n\to\infty} 1 - \frac{c}{n\epsilon^2} = 1$$

推论1.1（ Chebyshev大数定律的特殊情况）

设随机变量$X_1, X_2, \ldots, X_n, \ldots$是独立同分布的随机变量，且$E(X_k) = \mu, D(X_k) = \sigma^2$ $(k = 1,2, \ldots)$，

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$\epsilon > 0$，有

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

推论1.2 (泊松大数定律)

如果在独立试验序列中，事件A在第n次试验中出现概率为p_n，设n_A是前n次试验中事件A出现的次数，则对任意$\epsilon > 0$，有

$$\lim_{n\to\infty} P\left\{\left|\frac{n_A}{n} - \frac{p_1 + p_2 + \cdots + p_n}{n}\right| < \epsilon\right\} = 1$$

推论1.3（Markov大数定律）

设随机变量$\{X_n\}$是随机变量序列，如果

$$\lim_{n\to\infty} D\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = 0 \text{ (称为Markov条件)},$$

则对任意$\epsilon > 0$，有

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

不要求随机变量的独立性, 因此提供了一种研究随机变量序列服从大数定律的方法

定理2（Khinchin大数定律）

设随机变量$X_1, X_2, \dots, X_n, \dots$是独立同分布的随机变量序列，且具有数学期望$E(X_k) = \mu,\ (k = 1,2,\dots)$，做前n个变量的算术平均

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

则对任意$\epsilon > 0$，有

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

注意：与前面推论1.1 (Chebyshev大数定律的特殊情况) 相比，这里没有要求$D(X_n)$存在

证明：对于$D(X_n)$存在的情形，可参见前面的证明

考虑$D(X_n)$不存在的情形：

不失一般性，假设$E(X_k)=\mu=0$，采用截尾法证明

构造一个随机变量对

$$U_k=X_k, V_k=0, \quad \text{当}|X_k|\le\delta n$$

$$U_k=0, V_k=X_k, \quad \text{当}|X_k|>\delta n$$

于是，$X_k=U_k+V_k$

注意U_k,V_k对n的依赖性，只需要对任给$\epsilon>0$，证明存在常数δ，使得$n\to\infty$时有

$$P\left\{\left|\sum_{k=1}^{n}U_k\right|>\frac{\epsilon n}{2}\right\}\to 0, \quad (*1)$$

$$P\left\{\left|\sum_{k=1}^{n}V_k\right|>\frac{\epsilon n}{2}\right\}\to 0, \quad (*2)$$

令$a=E(|X_k|)$, 即$a=\int_{-\infty}^{\infty}|x_k|dF(x_k)$，$F(x_k)$为随机变量$X_k$的分布函数。

由U_k的定义，考虑U_k的上界为δn，于是

$$E\left(U_k^2\right)=\int_{-\delta n}^{\delta n}|u_k|^2dF(u_k)\le\delta n\int_{-\infty}^{\infty}|x_k|dF(x_k)=a\delta n$$

注意到$U_1,U_2,\dots,U_n$是独立同分布的，于是

$$D\left(\sum_{k=1}^{n}U_k\right)=nD(U_k)\le a\delta n^2$$

由Chebyshev不等式

$$P\left\{\left|\sum_{k=1}^{n}U_k\right|>\frac{\epsilon n}{2}\right\}\le\frac{4a\delta}{\epsilon^2}$$

于是可以取充分小的δ，使得(*1)成立

$$P\left\{\sum_{k=1}^{n} V_k \neq 0\right\} \leq nP\{V_1 \neq 0\}$$

注意到，

$$P\{V_1 \neq 0\} = P\{|X_k| > \delta n\}$$

$$= \int_{|x_k|>\delta n} dF(x_k) \leq \frac{1}{\delta n}\int_{|x_k|>\delta n} |x_k| dF(x_k)$$

$$= \frac{1}{\delta n}\left[E(|x_k|) - \int_{|x_k|\leq\delta n} |x_k| dF(x_k)\right]$$

于是，

$$P\left\{\sum_{k=1}^{n} V_k \neq 0\right\} \leq \frac{1}{\delta}\left[E(|x_k|) - \int_{|x_k|\leq\delta n} |x_k| dF(x_k)\right]$$

$$\lim_{n\to\infty} P\left\{\sum_{k=1}^{n} V_k \neq 0\right\} = 0$$

这一结论比(*2)还强，于是定理成立

Khinchin大数定律的基本理解——独立同分布场合下满足：

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu, \qquad (n \to \infty)$$

Bernoulli大数定律可以看作是Khinchin大数定律的特殊情况

定理3 (Bernoulli大数定律)

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$\epsilon > 0$，有

$$\lim_{n\to\infty} P\left\{\left|\frac{f_A}{n} - p\right| < \epsilon\right\} = 1$$

或

$$\lim_{n\to\infty} P\left\{\left|\frac{f_A}{n} - p\right| > \epsilon\right\} = 0$$

证明：引入随机变量

$$X_k = \begin{cases} 0, & \text{若在第}k\text{次实验中}A\text{不发生} \\ 1, & \text{若在第}k\text{次实验中}A\text{发生} \end{cases}, \qquad k = 1,2,\ldots$$

显然，

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

且X_k服从以参数p为参数的(0-1)分布，故$E(X_k) = p$

又因为$X_1, X_2, \dots, X_n$是相互独立的，由Khinchin大数定理有

$$\lim_{n\to\infty} P\left\{\left|\frac{f_A}{n} - p\right| < \epsilon\right\} = 1$$

即

$$\lim_{n\to\infty} P\left\{\left|\frac{f_A}{n} - p\right| > \epsilon\right\} = 0$$

Bernoulli大数定律的实用价值

1. 揭示了在n重Bernoulli试验中，f_A与事件A发生的概率p有较大偏差的概率$P\left\{\left|\frac{f_A}{n}-p\right|>\epsilon\right\}$，且此概率随着试验次数n的增大而趋于0。
2. 在大量重复独立试验中事件出现频率的稳定性。
3. 提供了通过试验来确定事件概率的方法，及这一方法的合理性

总之，大数定律从理论上确定了用算术平均值代替均值，以频率代替概率的合理性，它既验证了概率论中一些假设的合理性，又为数理统计中用样本推断总体提供了理论依据。

大数定律的关系

大数定律的一般形式

对随机变量序列$\{X_n\}$ $(n = 1,2,\dots)$，$E(X_n)$存在，

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{P} \frac{1}{n}\sum_{k=1}^{n} E(X_k)$$

不要求独立同分布

Khinchin大数定律

$X_1, X_2, \dots, X_n, \dots$ 独立同分布，

$E(X_k) = \mu,\ (k = 1,2,\dots)$

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

Markov大数定律

随机变量序列$\{X_n\}$，$E(X_n)$存在，且

$$\lim_{n\to\infty} D\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = 0 \ (\text{Markov条件})$$

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{P} \frac{1}{n}\sum_{k=1}^{n} E(X_k)$$

Bernoulli试验

Bernoulli大数定律

n次独立重复试验发生f_A次，p是每次试验中发生的概率

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

Bernoulli试验

Poisson大数定律

n 次独立试验发生 n_A 次，第k次发生概率为p_k

$$\frac{n_A}{n} \xrightarrow{P} \frac{1}{n}\sum_{k=1}^{n} p_k$$

不要求独立性，但对方差有约束

0-1分布

不要求方差的存在性

Chebyshev大数定律

$X_1, X_2, \dots, X_n, \dots$两两相互独立

$E(X_n)$、$D(X_k)$存在$(k = 1,2,\dots)$

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{P} \frac{1}{n}\sum_{k=1}^{n} E(X_k)$$

独立同分布

独立同分布下的Chebyshev大数定律

$E(X_k) = \mu, D(X_k)$存在

$(k = 1,2,\dots)$

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

历史上的抛硬币结果

实验者	总次数n	正面次数 f_A	f_A/n
G. Buffon	4040	2048	0.5069
A. De Morgan	4092	2048	0.5017
K. Pearson	24000	12012	0.5005

似乎实验次数越多,偏差越小

例1 抛掷一枚均匀的，若把这枚硬币连续抛掷10次，则因为n=10比较小，发生大偏差的可能性有时会大一些，有时会小一些。若把这枚硬币连续抛掷n次，当n很大时，由切比雪夫不等式知正面出现的频率与0.5的偏差大于预先给定的精度$\epsilon = 0.01$的可能性为

$$P\left\{\left|\frac{f_A}{n} - 0.5\right| > 0.01\right\} = P\{|f_A - 0.5n| > 0.01n\}$$

$$\leq \frac{0.5(1-0.5)n}{(0.01n)^2} = \frac{2500}{n}$$

于是，当$n = 10^5$时，偏差$>1\%$的可能性小于2.5%

当$n = 10^6$时，偏差$>1\%$的可能性小于0.25%

可见，实验次数越多，偏差发生的可能性越小。这与前面的实验也是吻合的

例2 用计算定积分

$$J=\int_a^b g(x)dx$$

采用下面的方法实现：

任取一列相互独立的随机变量$\{X_n\}$，服从$[a,b]$上的均匀分布，则$\{g(X_n)\}$也是独立同分布的随机变量，且

$$E[g(X_n)]=\frac{1}{b-a}\int_a^b g(x)dx=\frac{J}{b-a}$$

由大数定律只要生成一系列随机数$\{X_n\}$，进而有

$$\frac{g(X_1)+g(X_2)+\cdots+g(X_n)}{n}\xrightarrow{P}E[g(X_n)]$$

$$J \approx (b-a)\frac{g(X_1)+g(X_2)+\cdots+g(X_n)}{n}$$

这种通过概率论思想实现数值计算的方法就称为Monte Carlo方法，其理论基础就是大数定律

例3. 设随机变量$X_1, X_2, \dots, X_n, \dots$独立同分布，且$E(X_k)=0$，$D(X_k)=\sigma^2, k=1,2,\dots$, 证明对任意$\epsilon>0$，有

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{k=1}^{n} X_k^2-\sigma^2\right|<\epsilon\right\}=1$$

证明：由于$X_1, X_2, \dots, X_n, \dots$独立同分布，故$X_1^2, X_2^2, \dots, X_n^2, \dots$也是相互独立的

注意到

$$E\left(X_k^2\right)=D(X_k)+[E(X_k)]^2=\sigma^2$$

于是由Khinchin定理有对任意$\epsilon>0$，下式成立

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{k=1}^{n} X_k^2-\sigma^2\right|<\epsilon\right\}=1$$

大数定律与公平博弈

基本假设：

1. 赌徒的赌本无限，不考虑破产问题
2. 赌徒无权随意终止，试验次数n事先固定

对赌模型：

1. X_k是第k次试验获得的盈利，总盈利为 $S_n = X_1 + X_2 + \cdots + X_n$
2. 每次试验需要支付的入场费为μ'，总的入场费为$n\mu'$

累计的盈利或损失预期为$S_n - n\mu'$，当期望 $\mu = E(X_k)$存在时，若$\mu = \mu'$，在古典理论中就称为公平的博弈

St. Petersburg悖论

抛掷一枚均匀的硬币，直至出现正面为止，如果是第 k 次出现正面，则可以赢得 2^k 元，同时游戏结束。在这样的规则下，公平的入场费应该是多少？

假设p_k为前$k-1$次为反面，第k次出现正面的概率，显然，$p_k=2^{-k}$，预期的盈利为

$$S_n=\sum_{i=1}^{+\infty}2^k p_k=+\infty$$

古典意义下，入场费+∞似乎是合理的

一些理解：

1. 效用角度

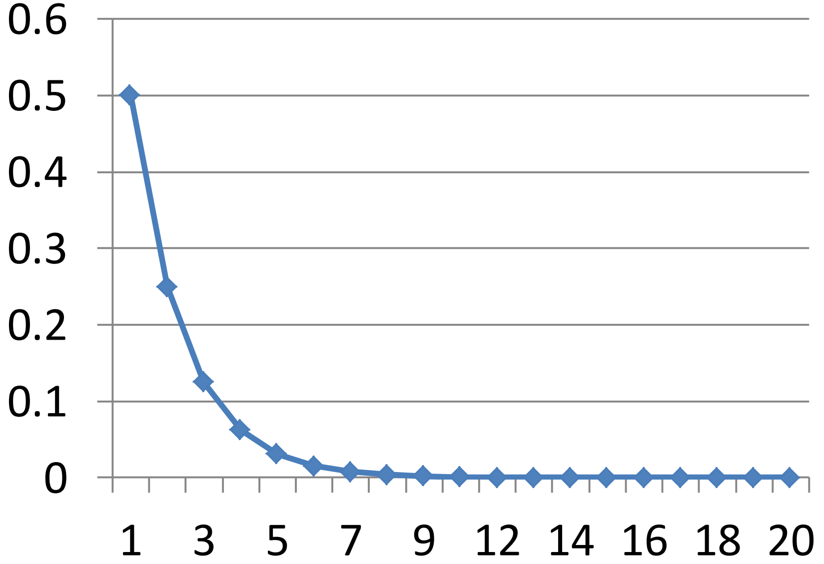
n	p_n	V_n	期望获利 $(P_n V_n)$	效用 $(\log_2 V_n)$	期望效用 $(P_n \log_2 V_n)$
1	1/2	\$2	1	1	0.5000
2	1/4	\$4	1	2	0.5000
3	1/8	\$8	1	3	0.3750
4	1/16	\$16	1	4	0.2500
5	1/32	\$32	1	5	0.1563
6	1/64	\$64	1	6	0.0938
7	1/128	\$128	1	7	0.0547
8	1/256	\$256	1	8	0.0313
9	1/512	\$512	1	9	0.0176
⋮	⋮	⋮	⋮	⋮	⋮
Σ	1		∞		2

一些理解：

1. 效用角度

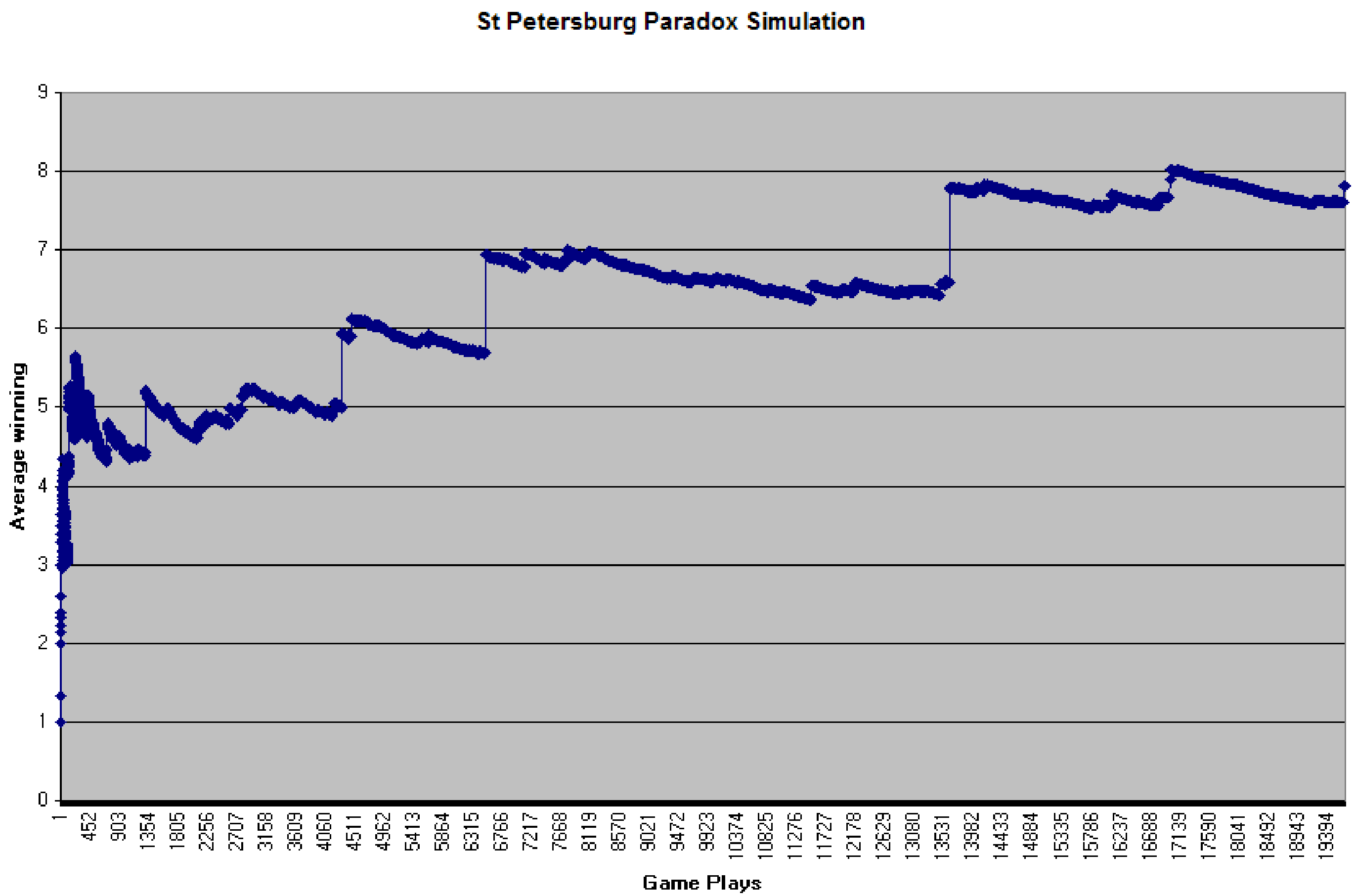
2. 规避风险角度

　　获得高回报的机会非常之少，很少有人愿意承担这样的风险



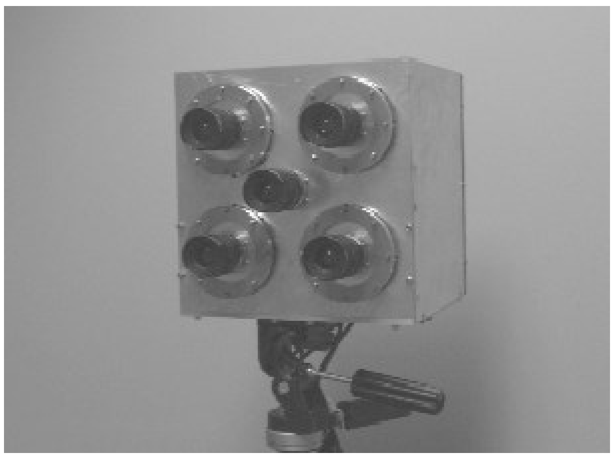
一些理解：

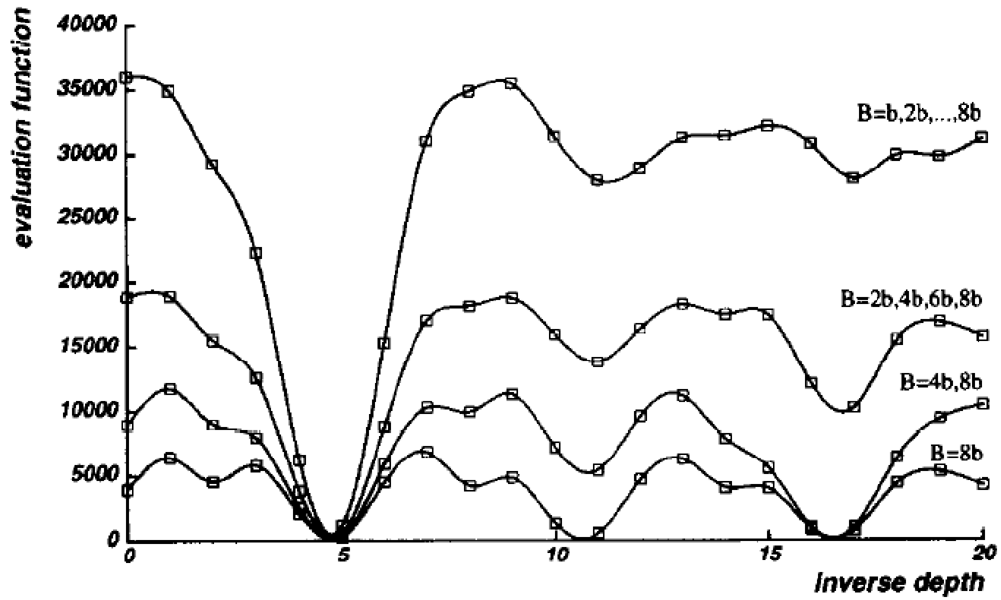
1. 效用角度
2. 规避风险角度
3. 平均获利



https://commons.wikimedia.org/wiki/File:Petersburg-Paradox.png

大数定律在实际系统中的应用





Masatoshi Okutomi, Takeo Kanade, A Multiple-Baseline Stereo, IEEE Transactions on Pattern Analysis and Machine Intelligence, 15(4), 1993

作业

概率论及其应用

p. 201 #7