§2 大数定律

2.1 贝努利(Bernoulli)大数定律

在 n次独立重复试验中, 设事件 A在每次试验中发生的概率为p. 当 A在第k次试验发生时令$X_k=1$; 当A在第k次试验不发生时令$X_k=0$. 于是

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

是n次试验中事件A发生的总次数, 而

$$\mu_n = S_n / n$$

就是A发生的频率. 按照概率的统计定义, 把事件A的概率定义当n无限增大时频率μ_n的稳定值, 在下面可以看到, 这一稳定值恰好就是p. 由于μ_n是随机变量, 我们不能在通常的数列的极限的意义下定义μ_n极限. 下面将证明频率μ_n"依概率收敛"于p.

为了考察频率μ_n与概率p相差有多大, 对任意固定的 $\varepsilon>0$, 估计μ_n 偏离p 不小于ε 的概率$P(|\mu_n - p| \ge \varepsilon)$. 由于$S_n$ 服从二项分布$B(n,p)$, 故$ES_n = np$, $DS_n = npq$.

$$P(|\mu_n - p| \ge \varepsilon) = P(|S_n - np| \ge n\varepsilon)$$

$$= \sum_{k:|k-np|\ge n\varepsilon} P(S_n = k) \le \frac{1}{n^2\varepsilon^2} \sum_{k:|k-np|\ge n\varepsilon} (k-np)^2 P(S_n = k)$$

$$\le \frac{1}{n^2\varepsilon^2} \sum_{0\le k\le n} (k - ES_n)^2 P(S_n = k) = \frac{DS_n}{n^2\varepsilon^2} = \frac{npq}{n^2\varepsilon^2} = \frac{pq}{n\varepsilon^2}.$$

于是推出

$$\lim_{n\to\infty} P(|S_n/n - p| \ge \varepsilon) = 0.$$

即 S_n/n 依概率收敛于 p（定义见后）. 上式称为**贝努利(Bernoulli)大数定律**, 它阐述了频率的极限与概率关系, 也阐明了概率的统计定义的数学意义.

上小节中, 证明了不等式 $P(|\mu_n - p| \geq \varepsilon) \leq \dfrac{pq}{n\varepsilon^2}$, 它是下面切比雪夫不等式的一个特例.

定理 2.1（切比雪夫不等式）

1）设 X 为非负随机变量, 且期望存在. 则对任意的 $\varepsilon > 0$ 都有

$$P(X \geq \varepsilon) \leq EX/\varepsilon.$$

2）设 X 为随机变量, 期望与方差都存在. 则对任意的 $\varepsilon > 0$ 都有

$$P(|X - EX| \geq \varepsilon) \leq DX/\varepsilon^2.$$

证　我们仅对连续性的情况证明切比雪夫不等式, 离散型的情况是类似的.

1）由 X 的非负性, 设当 $x < 0$ 时, x 的密度函数 $p(x) = 0$.

$$P(X \geq \varepsilon) = \int_{\varepsilon}^{+\infty} p(x)dx \leq \frac{1}{\varepsilon}\int_{\varepsilon}^{+\infty} xp(x)dx \leq \frac{1}{\varepsilon}\int_{-\infty}^{+\infty} xf(x)dx = \frac{EX}{\varepsilon}.$$

2）用非负随机变量 $Y = (X - EX)^2$ 代替 1) 中的 X 便得.

定理 2.2（贝努利大数定律） 设 μ_n 为 n 重贝努利试验中事件 A 发生的次数，p 为每次试验中 A 出现的概率，则对任意的 $\varepsilon > 0$ 有

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

证明 因为 $\mu_n \sim b(n,p)$, 且的数学期望和方差 $E(\frac{\mu_n}{n}) = p, Var(\frac{\mu_n}{n}) = \frac{p(1-p)}{n}$

所以由切比雪夫不等式得

$$1 \ge P\{|\frac{\mu_n}{n} - p| < \varepsilon\} \ge 1 - \frac{Var(\frac{\mu_n}{n})}{\varepsilon^2} = 1 - \frac{p(1-p)}{n\varepsilon^2}$$

当 $n \to +\infty$ 时，上式右端趋于 1，因此

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

结论得证。

贝努利大数定律说明：随着 n 的增大，事件 A 发生的频率 $\frac{\mu_n}{n}$ 与其频率 p 的偏差 $|\frac{\mu_n}{n} - p|$ 大于预先给定的精度 ε 的可能性愈来愈小，小到可以忽略不计。这就是频率稳定与概率的含义，或者说频率依概率收敛与概率。

譬如，抛一枚硬币出现正面的概率 p＝0.5。若把这枚硬币连抛 10 次，则因为 n 较小，发生大偏差的可能性有时会大一些，有时会小一些。若把这枚硬币连抛 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 ε（若取精度 ε=0.01）的可能性

$$P\{|\frac{\mu_n}{n} - 0.5| > 0.01\} \le \frac{0.5 \times 0.5}{n0.01^2} = \frac{10^4}{4n}$$

当 n＝10^5 时，大偏差发生的可能性小于 1/40＝2.5％。当 n=10^6 时，大偏差发生的可能性小于 1/400＝0.25％。

例 2.1 **（用蒙特卡洛方法计算定积分（随机投点法））** 设 $0 \le f(x) \le 1$，求 $f(x)$在区间[0,1]上的积分值：

$$J=\int_0^1 f(x)dx$$

设(X,Y)服从正方形$\{0\le x\le 1, 0\le y\le 1\}$上的均匀分布，则可知$X$服从[0,1]上的均匀分布，$Y$也服从[0,1]上的均匀分布，且$X$与$Y$独立。又记事件$A=\{Y\le f(x)\}$

则A的概率为$p=P(Y\le f(X))=\int_0^1\int_0^{f(x)}dydx=\int_0^1 f(x)dx=J$

即定积分的值J就是事件A的概率p。由贝努利大数定律，我们可以用重复试验中A出现的频率作为p的估计值。这种求定积分的方法也称为随机投点法，即将(X,Y)看成是向正方形$\{0\le x\le 1, 0\le y\le 1\}$内的随机投点，用随机点落在区域$\{y\le f(x)\}$中的频率作为定积分的近似值。

下面用蒙特卡洛方法，来得到A出现的频率：

(1) 先用计算机产生(0,1)上的均匀分布的$2n$个随机数；x_i, y_i，$i=1,2,\ldots,n$. 这里n可以很大，譬如$n=10^4$，甚至$n=10^5$。

(2) 对 n 对数据(x_i,y_i)，$i=1,2,\ldots,n$.记录满足如下不等式$y_i\le f(x_i)$的次数，这就是事件A发生的频数μ_n。由此可得事件A发生的频率$\frac{\mu_n}{n}$，则$J\approx\frac{\mu_n}{n}$

2.2 常用的几个大数定律

一、大数定律的一般形式

定义 2.1 设有一随机变量序列$\{X_n\}$,假如它具有

$$\lim_{n\to+\infty}P\{|\frac{1}{n}\sum_{i=1}^{n}X_i-\frac{1}{n}\sum_{i=1}^{n}E(X_i)|<\varepsilon\}=1,\forall\varepsilon>0$$

形式的性质，则称该随机变量序列$\{X_n\}$服从**大数定律**。

二、切比雪夫大数定律

定理 2.3 （切比雪夫大数定律）设$\{X_n\}$为一列两两不相关的随机变量序列，若每个X_i的方差存在，且有共同的上界，即$Var(X_i) \le c, i = 1,2,\cdots$，则$\{X_n\}$服从大数定律，即对任意的$\varepsilon > 0$，定义 2.1 中的式子都成立。

证明 因为$\{X_n\}$两两不相关，故

$$Var(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) \le \frac{c}{n},$$

再由切比雪夫不等式得到：对任意的$\varepsilon > 0$，有

$$\lim_{n\to+\infty} P\{|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)| < \varepsilon\} \ge 1 - \frac{Var(\frac{1}{n}\sum_{i=1}^{n} X_i)}{\varepsilon^2} \ge 1 - \frac{c}{n\varepsilon^2}$$

于是当$n \to +\infty$时，有

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

注意，切比雪夫大数定律只要求$\{X_n\}$互不相关，并不要求它们是同分布的。假如$\{X_n\}$是独立同分布的随机变量序列，且方差有限，则$\{X_n\}$必定服从大数定律。

例 2.2 设$\{X_n\}$是相互独立同分布的随机变量序列，$E(X_n^4) < +\infty$，若令$E(X_n) = \mu, Var(X_n) = \sigma^2$，考察$Y_n = (X_n - \mu)^2, n = 1,2,\cdots$**则随机变量序列**$\{Y_n\}$服从大数定律，即对任意的$\varepsilon > 0$，有

$$\lim_{n\to+\infty} P\{|\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 - \sigma^2| \ge \varepsilon\} = 0$$

证明 显然$\{Y_n\}$是独立同分布随机变量序列，其方差

$$Var(Y_n) = Var(X_n - \mu)^2 = E(X_n - \mu)^4 - \sigma^4$$

由于$E(X_n^4)$存在，故$E(X_n^2)$也存在，从而$E(X_n-\mu)^4$也存在。由切比雪夫大数定律知$\lim\limits_{n\to+\infty}P\{|\frac{1}{n}\sum_{i=1}^{n}Y_i-\frac{1}{n}\sum_{i=1}^{n}E(Y_i)|\geq\varepsilon\}=0$

其中$\frac{1}{n}\sum_{i=1}^{n}Y_i=\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2,-\frac{1}{n}\sum_{i=1}^{n}E(Y_i)=\sigma^2$，故$\{Y_n\}$服从大数定律。

三、马尔可夫大数定律

对随机变量序列$\{X_n\}$，若$\frac{1}{n^2}Var(\sum_{i=1}^{n}X_i)\xrightarrow{n\to\infty}0,$成立，则$\{X_n\}$服从大数定律，即对任意的$\varepsilon>0$，都成立。

例 2.3 设$\{X_n\}$为一同分布、方差存在的随机变量序列，且X_n仅与X_{n-1}和X_{n+1}相关，而与其他X_i不相关，试问该随机变量序列$\{X_n\}$是否服从大数定律？

解 $\{X_n\}$为相依随机变量序列，考虑其马尔可夫条件

$$\frac{1}{n^2}Var(\sum_{i=1}^{n}X_i)=\frac{1}{n^2}[\sum_{i=1}^{n}Var(X_i)+2\sum_{i=1}^{n-1}Cov(X_i,X_{i+1})]$$

记$Var(X_n)=\sigma^2$，则$|Cov(X_i,X_j)|\leq\sigma^2$于是有

$$\frac{1}{n^2}Var(\sum_{i=1}^{n}X_i)\leq\frac{1}{n^2}[n\sigma^2+2(n-1)\sigma^2]\to 0,(n\to+\infty)$$

即马尔可夫条件成立，故$\{X_n\}$服从大数定律。

四、辛钦大数定律

定理 2.4 (辛钦大数定律) 设$\{X_n\}$为一独立同分布的随机变量序列，若X_i的数学期望存在，则服从大数定律，即

$$\lim_{n\to+\infty}P\{|\frac{1}{n}\sum_{i=1}^{n}X_i-\frac{1}{n}\sum_{i=1}^{n}E(X_i)|<\varepsilon\}=1,\forall\varepsilon>0$$

成立。

辛钦大数定律提供了求随机变量数学期望 $E(X)$的近似值的方法，设想对随机变量 X 独立重复的观察 n 次，第 k 次观察值为 X_k，则 $X_1,X_2,\ldots,X_n$ 应该是相互独立的，且它们的分布应该与 X 的分布相同。所以，在 $E(X)$存在的条件下，按照辛钦大数定律，当 n 足够大时，可以把平均观察值 $\frac{1}{n}\sum_{i=1}^{n}X_i$ 作为 $E(X)$的近似值。

例 2.4　（用蒙特卡洛方法计算定积分（平均值法）） 为计算定积分

$$J=\int_0^1 f(x)dx$$

设随机变量 X 服从(0,1)上的均匀分布，则 $Y=f(X)$的数学期望为

$$E(f(X))=\int_0^1 f(x)dx=J$$

所以估计 J 的值就是估计 $f(X)$的数学期望的值。由辛钦大数定律，可以用 $f(X)$的观察值的平均去估计 $f(X)$的数学期望的值。具体做法如下：先用计算机产生 n 个(0,1)上的均匀分布的随机数：x_i，$i=1,2,\ldots,n$.然后对每个 x_i 计算 $f(x_i)$,最后得 J 的估计值为

$$J\approx\frac{1}{n}\sum_{i=1}^{n}f(x_i)。$$