05

大数定律与中心极限定理

《概率论与数理统计》数科学院统计系

目录/Contents

目录/Contents

5.1 大数定律

一、切比雪夫不等式

例1 设$X \sim N(\mu,\sigma^2)$　计算$P(|X-\mu| \geq 3\sigma)$。

解：因为$\frac{X-\mu}{\sigma} \sim N(0,1)$**，所以**

$$P\left(|X-\mu| \geq 3\sigma\right) = P\left(\left|\frac{X-\mu}{\sigma}\right| \geq 3\right)$$

$$\boxed{P\left(\left|\frac{X-\mu}{\sigma}\right| \geq 3\right) = 2-2\Phi(3) = 0.003}$$

一、切比雪夫不等式

定理1（切比雪夫不等式）

设随机变量$\boldsymbol{X}$的数学期望$E(X)$及方差$D(X)$存在，则对于任意的$\varepsilon > \mathbf{0}$，有

$$P(|X - E(X)| \geq \varepsilon) \leq \frac{D(X)}{\varepsilon^2}$$

证明：仅给出X为连续型随机变量的证明。

$$P\left(\left|X - E(X)\right| \geq \varepsilon\right) = \int_{|x-E(x)|\geq\varepsilon} f(x)dx \leq \int_{|x-E(x)|\geq\varepsilon} \frac{\left|X - E(X)\right|^2}{\varepsilon^2} f(x)dx$$

$$\leq \int_{-\infty}^{+\infty} \frac{\left|X - E(X)\right|^2}{\varepsilon^2} f(x)dx = \frac{D(X)}{\varepsilon^2}$$

例2 设$X \sim N(\mu,\sigma^2)$ 用切比雪夫不等式估计概率$P(|X-\mu| \geq 3\sigma)$。

解

因为$\varepsilon=3\sigma$，由切比雪夫不等式得

$$P(|X-E(X)| \geq \varepsilon) \leq \frac{D(X)}{\varepsilon^2}$$

$$P(|X-\mu| \geq 3\sigma) \leq \frac{D(X)}{(3\sigma)^2} = \frac{1}{9}$$

一、切比雪夫不等式

例3

设随机变量X 的方差 $D(X)=0$，求证，X 服从参数为c的退化分布。

证明　**利用切比雪夫不等式得，对任意的$\varepsilon>0$，有**

$$0 \le P\left(\left|X-E(X)\right| \ge \varepsilon\right) \le \frac{D(X)}{\varepsilon^2}=0$$

由ε的任意性知

$$P\left(X=E(X)\right)=1$$

二、依概率收敛

定义1 设 $X_1, X_2, \cdots, X_n, \cdots$ 是随机变量序列，如果存在一个常数 c，使得对任意一个$\varepsilon > 0$，

总有 $\lim\limits_{n\to\infty} P(|X_n - c| < \varepsilon) = 1$，

那么称$X_1, X_2, \cdots, X_n, \cdots$**依概率收敛于$c$**，记作 $X_n \xrightarrow{P} c$

当n充分大时 $\{X_n \in (c-\varepsilon, c+\varepsilon)\}$ 几乎总是发生

或等价地 $$\lim_{n\to\infty} P(|X_n - c| \geq \varepsilon) = 0.$$

依概率收敛性具有下列性质：

定理2

如果 $X_n \xrightarrow{P} c$，$Y_n \xrightarrow{P} b$，且函数$g(x,y)$ 在 (a,b) 处连续，则

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

例如 $X_n \xrightarrow{P} 1, Y_n \xrightarrow{P} 2$，则 $X_n+Y_n \xrightarrow{P} 3$

定理3 切比雪夫大数定律

设随机变量序列$X_1, X_2, \cdots, X_n, \cdots$两两不相关，若$E(X_i) < \infty$，$D(X_i) < \infty$，$i = 1, 2, \cdots$。则对任意$\varepsilon > 0$ 有

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)\right| \leq \varepsilon\right) \to 1$$

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\text{p}} \frac{1}{n}\sum_{i=1}^{n} E(X_i)\text{。}$$

证明 因为随机变量 $X_1, X_2, \cdots, X_n, \cdots$两两不相关，根据期望和方差的性质得

$$E\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n}\sum_{i=1}^{n}E(X_i),\quad D\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n^2}\sum_{i=1}^{n}D(X_i)\leq\frac{c}{n}$$

由切比雪夫不等式知，对任意$\varepsilon>0$，当$n\to\infty$时，

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\frac{1}{n}\sum_{i=1}^{n}E(X_i)\right|\geq\varepsilon\right)\leq\frac{1}{\varepsilon^2}D\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)\leq\frac{c}{n\varepsilon^2}\to 0。$$

这里随机变量序列$X_1, X_2, \cdots, X_n, \cdots$两两不相关指序列中的任意两个随机变量线性无关。

定理4（独立同分布大数定律）

设随机变量序列$X_1, X_2, \cdots, X_n, \cdots$独立同分布，若$E(X_i)=\mu<\infty$，$D(X_i)=\sigma^2<\infty$，$i=1,2,\cdots$。则对任意$\varepsilon>0$ 有

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

这里随机变量序列$X_1, X_2, \cdots, X_n, \cdots$独立同分布指随机变量序列相互独立，且序列中随机变量的分布类型及参数均相同。

定理5（伯努利大数定律）

设随机变量序列$X_1,X_2,\cdots$独立同分布，且$X_i \sim B(1,p)$，$i = 1,2,\cdots$.

则对任意$\varepsilon > 0$，有

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

显然伯努利大数定律是独立同分布大数定律的特例。这里

$$\frac{1}{n}\sum_{i=1}^{n} E(X_i) = p$$

频率的稳定性 在n次独立重复试验中，设随机变量

$$X_i=\begin{cases}1, \text{事件}A\text{在第}i\text{次试验中发生}, \\ 0, \text{事件}A\text{在第}i\text{次试验中不发生},\end{cases} \quad X_i \sim B(1,p),\ p=P(A)$$

那么n次重复试验中 A发生的频率为

$$f_n(A)=\frac{1}{n}\sum_{i=1}^{n}X_i=\bar{X}\xrightarrow{\ \mathrm{p}\ }p=P(A),$$

例4

设 $X_1, X_2, \cdots$ **是独立同分布的随机变量序列，在下列三种情况下，当** $n \to \infty$ **时**

试问$\bar{X}$, $\frac{1}{n}\sum_{i=1}^{n} X_i^2$分别依概率收敛于什么值?

1. $X_i \sim B(m, p), i = 1, 2, \cdots;$
2. $X_i \sim E(\lambda), i = 1, 2, \cdots;$
3. $X_i \sim N(\mu, \sigma^2), i = 1, 2, \cdots$。

解

三种情况下，$X_1, X_2, \cdots$均是独立同分布的随机序列，且X_i和X_i^2具有有限的数学期望和方差，对$X_1, X_2, \cdots$及$X_1^2, X_2^2, \cdots$分别使用独立同分布大数定律，得

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

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{P} \frac{1}{n}\sum_{i=1}^{n} E(X_i^2) = E(X_i^2) = D(X_i) + E^2(X_i)$$

例4续

01 OPTION

当$X_i \sim B(m,p)$时，$E\left(X_i\right)=mp, E\left(X_i^2\right)=mp\left(1-p\right)+m^2p^2$,有

$$\bar{X} \xrightarrow{P} mp,\quad \frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{P} mp\left(1-p\right)+m^2p^2$$

02 OPTION

当$X_i \sim E\left(\lambda\right)$时，$E\left(X_i\right)=\frac{1}{\lambda}, E\left(X_i^2\right)=\frac{2}{\lambda^2}$,有

$$\bar{X} \xrightarrow{P} \frac{1}{\lambda},\quad \frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{P} \frac{2}{\lambda^2}$$

03 OPTION

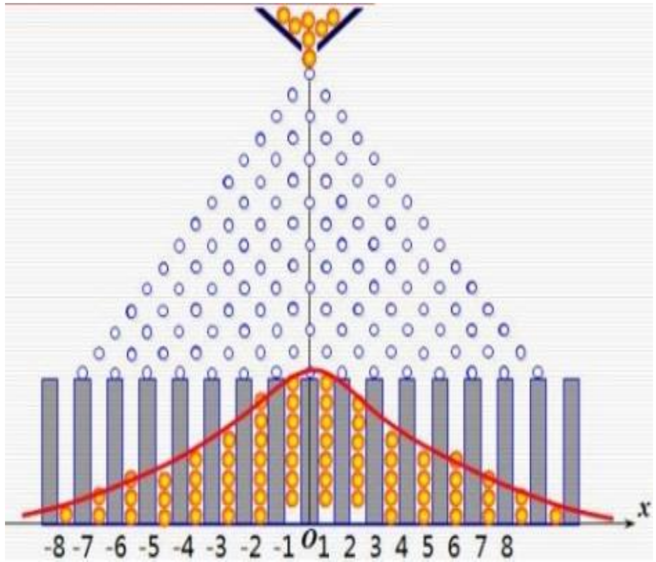
当$X_i \sim N\left(\mu,\sigma^2\right)$时，$E\left(X_i\right)=\mu,\ E\left(X_i^2\right)=\sigma^2+\mu^2$, 有

$$\bar{X} \xrightarrow{P} \mu,\quad \frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{P} \sigma^2+\mu^2$$

目录/Contents

例5（高尔顿钉板实验）

如图，有一排有一个板上面有 排钉子，每排相邻的两个钉子之间的距离均相等。上一排钉子的水平位置恰巧位于下一排紧邻的两个钉子水平位置的正中间。从上端入口处放入小球，在下落过程中小球碰到钉子后以相等的可能性向左或向右偏离，碰到下一排相邻的两个钉子中的一个。如此继续下去，直到落入底部隔板中的一格中。问当有大量的小球从上端依次放入，任其自由下落,问小球最终在底板中堆积的形态．设钉子有16排



高尔顿钉板

首先进行分析。小球堆积的形态取决于小球最终下落在底部隔板的位置的分布。设随机变量X为“小球最终下落在底部隔板中的位置”。又引入随机变量

$$X_i=\begin{cases}-1, & \text{小球碰到第}i\text{排钉子向左下落，}\\ 1, & \text{小球碰到第}i\text{排钉子向右下落。}\end{cases}\quad i=1,\cdots,n$$

显然$X=\sum_{i=1}^{n}X_i$和的分布计算是复杂的。有没有其他的方法呢？经过试验我们观察发现小球堆积形态呈现出中间高两边低的特点，能否认为X近似服从正态分布？

定理6（列维—林德伯格中心极限定理）

设随机变量序列$X_1, X_2, \cdots, X_n, \cdots$独立同分布，若$E(X_i)=\mu$，$D(X_i)=\sigma^2$，且$0<\sigma^2<\infty, i=1,2,\cdots,n,\cdots$。则对任意实数$x$，有

$$\lim_{n\to\infty} P\left(\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} \le x\right) = \Phi(x).$$

由于中心极限定理的证明需要使用其它的数学工具，因此这里不给出证明。

例5

已知某计算机程序进行加法运算时，要对每个加数四舍五入取整。假设所有取整的误差相互独立，并且均服从 $U(-0.5,0.5)$ 。（1）如果将1200个数相加，求误差总和的绝对值超过20的概率；（2）要使误差总和的绝对值不超过5的概率超过0.95，最多有多少个加数?

解 （1）设 X 为“对每个加数四舍五入，将1200 个数相加后的误差总和”，并设 X_i 为“第 i 个加数的四舍五入误差”，$i=1,2,\cdots,1200$ 。则 $X=\sum_{i=1}^{1200}X_i$，有

$$E(X_i)=\frac{-0.5+0.5}{2}=0,\quad D(X_i)=\frac{(0.5+0.5)^2}{12}=\frac{1}{12},\ i=1,\cdots,1200$$

$$E\left(\sum_{i=1}^{1200}X_i\right)=1200\times 0=0,\ D\left(\sum_{i=1}^{1200}X_i\right)=1200\times\frac{1}{12}=100。$$

由列维—林德伯格中心极限定理知，$X=\sum_{i=1}^{1200}X_i \overset{\text{近似}}{\sim} N(0,100)$ 。因此，

$$P\left(\left|\sum_{i=1}^{1200}X_i\right|>20\right)=P\left(\frac{\left|\sum_{i=1}^{1200}X_i-0\right|}{\sqrt{100}}>\frac{20-0}{\sqrt{100}}\right)\approx 2\big(1-\Phi(2)\big)=2\times 0.0228=0.0456$$

解 (2) 设加数最多有n 个才能使误差总和的绝对值不超过5的概率超过0.95。有

$$E\left(\sum_{i=1}^{n} X_i\right)=0, D\left(\sum_{i=1}^{n} X_i\right)=\frac{n}{12}$$

由列维—林德伯格中心极限定理知, $\sum_{i=1}^{n} X_i \overset{近似}{\sim} N\left(0,\frac{n}{12}\right)$. 因此,

$$P\left(\left|\sum_{i=1}^{n} X_i\right|\le 5\right)=P\left(\frac{\left|\sum_{i=1}^{n} X_i-0\right|}{\sqrt{\frac{n}{12}}}\le\frac{5}{\sqrt{\frac{n}{12}}}\right)=2\Phi\left(\frac{5}{\sqrt{\frac{n}{12}}}\right)-1\ge 0.95$$

即$\Phi\left(\frac{5\sqrt{12}}{\sqrt{n}}\right)\ge 0.975$,查表得 $\frac{5\sqrt{12}}{\sqrt{n}}\ge 1.96$,有$n\le 12\times\left(\frac{5}{1.96}\right)^2$ $=78.092$,取$n=78.$

所以最多有78个加数，才能使误差总和的绝对值不超过5的概率超过0.95。

例6

在街头赌博中,庄家在高尔顿钉板的底板两端距离原点超出8格的位置放置了值钱的东西来吸引顾客，试用中心极限定理来揭穿这个街头赌博中的骗术。

解 设X为“小球在底板中的位置”，$X_i=\begin{cases}-1, & \text{小球碰到第}i\text{排钉子向左下落,}\\ 1, & \text{小球碰到第}i\text{排钉子向右下落。}\end{cases}$

$i=1,2,\cdots,16$，显然 $X_1,\cdots,X_{16}$ 相互独立且同分布，且 $X=\sum_{i=1}^{16}X_i$。X_i 的概率函数如表

X_i	-1	1
P_r	0.5	0.5

所以，有
$$E(X_i)=-1\times0.5+1\times0.5=0,$$
$$E(X_i^2)=(-1)^2\times0.5+1^2\times0.5=1,$$
$$D(X_i)=1, i=1,\cdots,16,$$
$$E\left(\sum_{i=1}^{16}X_i\right)=16\times0=0, D\left(\sum_{i=1}^{16}X_i\right)=16\times1=16,$$

由列维-林德伯格中心极限定理知 $X=\sum_{i=1}^{16}X_i \overset{\text{近似}}{\sim} N(0,16)$.因此

$$P(|X|>8)=P(X>8)+P(X<-8)$$
$$\approx 1-\Phi\left(\frac{8-0}{\sqrt{16}}\right)+\Phi\left(\frac{-8-0}{\sqrt{16}}\right)=2\left[1-\Phi(2)\right]=0.0456$$

说明顾客中奖的可能性微乎其微。

定理7（德莫弗—拉普拉斯中心极限定理）

设随机变量序列$X_1, X_2, \cdots, X_n, \cdots$ 独立同分布，且$X_i \sim B(1,p)$，$i = 1,2,\cdots$

则对任何实数x，有

$$\lim_{n\to\infty} P\left(\frac{\sum_{i=1}^{n} X_i - np}{\sqrt{np(1-p)}} \le x \right) = \Phi(x)$$

例7

某单位的局域网有100个终端,每个终端有 10%的时间在使用,如果各个终端使用与否是相互独立的.（1）计算在任何时刻同时最多有15个个终端在使用的概率；（2）用中心极限定理计算在任何时刻同时最多有15个个终端在使用的概率的近似值；（3）用泊松定理计算在任何时刻同时最多有15个终端在使用的概率近似值。

解 设随机变量$X_i=\begin{cases}1,\text{第}i\text{个终端在使用；}\\0,\text{否则。}\end{cases}$ $i=1,2,\cdots,100.$ 由已知得$X_1,\cdots,X_{100}$ 独立同分布

且$X_i\sim B(1,p)$，其中$p=0.1$.同时使用的终端数$\sum_{i=1}^{100}X_i \sim B(100,0.1)$.

解　(1) 借助于计算机得

$$P\left(\sum_{i=1}^{100} X_i \le 15\right)=\sum_{k=0}^{15}\binom{100}{k}0.1^k \times 0.9^{100-k}=0.9601$$

即在任何时刻同时最多有15个终端在使用的概率为0.9601。

(2) 因为$E\left(\sum_{i=1}^{100} X_i\right)=100\times 0.1=10, D\left(\sum_{i=1}^{100} X_i\right)=10\times 0.9=9$。运用德莫弗—拉普拉斯中心极限定理$\sum_{i=1}^{100} X_i \overset{\text{近似}}{\sim} N(10,9)$. 因此

$$P\left(\sum_{i=1}^{100} X_i \le 15\right)\approx 1-\Phi\left(\frac{15-10}{\sqrt{9}}\right)=\Phi\left(\frac{5}{3}\right)=0.9522$$

即在任何时刻同时最多有15个终端在使用的概率近似值为0.9522.

解

（3）因为$n>10, p<0.1$，所以$\sum_{i=1}^{100} X_i \overset{\text{近似}}{\sim} P(10)$. 有

$$P\left(\sum_{i=1}^{100} X_i \le 15\right) \approx \sum_{k=0}^{15} e^{-10}\frac{10^k}{k!}=0.9513$$

即在任何时刻同时最多有15个终端在使用的概率近似值为0.9513.

总结/summary

切比雪夫不等式	理解切比雪夫不等式的定义， 掌握用切比雪夫不等式求解概率上界
大数定律	理解依概率收敛的定义 了解切比雪夫大数定律 了解伯努利大数定律 了解辛钦大数定律。
中心极限定理	掌握运用列维-林德伯格中心定理和棣莫弗-拉普拉斯中心极限定理求解独立随机变量之和的近似概率值