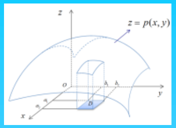


第六章 极限定理

即:大数定律和中心极限定理

杨勇制作

第一节　大数定律

n次独立试验中事件A出现m次，

$P(A)=p(0<p<1)$,

第一章提到的“频率具有稳定性”，
那是不是有

$$\lim_{n\to\infty}\frac{m}{n}=p \quad ?$$

如果成立,则有

$\forall \varepsilon > 0, \exists N,$当$n > N$时,有$\left|\dfrac{m}{n} - p\right| < \varepsilon$。

取 $\varepsilon_0 = \dfrac{1-p}{2}$,且 $m = n$,则

对特定的ε_0,存在N',当$n > N'$时,有$\left|\dfrac{n}{n} - p\right| < \dfrac{1-p}{2}$,

矛盾!

$m=n$的概率为

$$C_n^n p^n q^0 = p^n,$$

使$\left|\frac{m}{n}-p\right|<\varepsilon$不成立的概率为$p^n$,

随着$n\to\infty, p^n\to 0$,即 $n\to\infty$

使$\left|\frac{m}{n}-p\right|<\varepsilon$不成立的概率为趋向0,

使$\left|\frac{m}{n}-p\right|<\varepsilon$不成立的概率为趋向0，

即

使$\left|\frac{m}{n}-p\right|<\varepsilon$成立的概率为趋向1，

所以,应该是

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

“频率具有稳定性”应该这样描述:

$$\forall \varepsilon > 0, \qquad \lim_{n\to\infty} P\left(\left|\frac{m}{n} - p\right| < \varepsilon\right) = 1,$$

描述当n很大时,$\dfrac{m}{n}$与p 接近。

一. 切比雪夫大数定律

设$X_1, X_2, \cdots$是由相互独立的随机变量所构成的序列,各自有数学期望$EX_1, EX_2, \cdots$及有限方差$DX_1, DX_2, \cdots$,并且它们的方差有公共上界,即$DX_i \le C, i = 1, 2, \cdots$,其中$C$是与$i$无关的常数,那么对任意$\varepsilon > 0$,有

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

证明：令 $Y_n = \frac{1}{n}\sum_{k=1}^{n} X_k$， $EY_n = \frac{1}{n}\sum_{k=1}^{n} EX_k$，

$$DY_n = D\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{1}{n^2}\sum_{k=1}^{n} DX_k \le \frac{C}{n},$$

由切比雪夫不等式

$$P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} EX_k\right| < \varepsilon\right) = P\left(\left|Y_n - EY_n\right| < \varepsilon\right)$$

$$\ge 1 - \frac{DY_n}{\varepsilon^2} \ge 1 - \frac{C}{n\varepsilon^2},$$

$$1 \geq P\left(\left|Y_n - EY_n\right| < \varepsilon\right) \geq 1 - \frac{C}{n\varepsilon^2},$$

即
$$\lim_{n\to\infty} P\left(\left|Y_n - EY_n\right| < \varepsilon\right) = 1,$$

从而
$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} EX_k\right| < \varepsilon\right) = 1。$$

注:切比雪夫大数定律是弱大数定律,还有强大数定律。

切比雪夫大数定律的特例

推论(辛钦大数定律):

设$X_1, X_2, \cdots$是独立同分布的随机变量序列,且

$EX_k = a, DX_k = \sigma^2 < \infty, k = 1, 2, \cdots,$

那么对任意$\varepsilon > 0$,有

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

描述当n很大时，$\frac{1}{n}\sum_{k=1}^{n} X_k$与$a$接近。

从而得到“平均值具有稳定性”。

注：辛钦大数定律成立，也可以

称$\frac{1}{n}\sum_{k=1}^{n} X_k$依概率收敛于$a$，记$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{P} a$。

2017－2018年第二学期

2018年6月5日

概率论试卷

填空题13

13.将一枚骰子重复掷n次,则当$n \to \infty$时,n次掷出点数的算术平均值$\bar{X}_n$依概率收敛于________________。

解: X_k:表示第k次投掷出现的点数,$k = 1, 2, \cdots$,

$$X_k \sim \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix},$$

$$a = EX_k = \frac{1}{6}(1+2+3+4+5+6) = \frac{7}{2},$$

DX_k为一个常数,

$X_1, X_2, \cdots$独立同分布，再由辛钦大数定律知，

$$\bar{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{P} \frac{7}{2}。$$

即

$$\forall \varepsilon > 0, \lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{7}{2}\right| < \varepsilon\right) = 1。$$

蒙特卡洛方法介绍

计算机产生的随机数$U \sim U[0,1]$,

$X = a+(b-a)U \sim U[a,b]$。

问题:设$f(x)$是连续函数,求$\int_a^b f(x)dx$。

由于

$$Ef(X)=\int_{-\infty}^{+\infty} f(x)p_X(x)dx=\frac{1}{b-a}\int_a^b f(x)dx$$

$$\int_a^b f(x)dx = (b-a)Ef(X)$$

设$X_k \sim U[a,b], k=1,2,\cdots,$ 且$X_1, X_2,\cdots$独立,则 $f(X_1), f(X_2),\cdots$独立同分布。

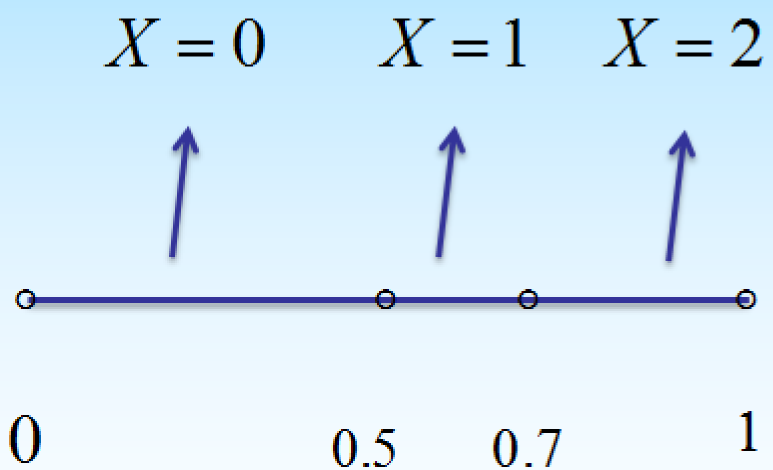
由切比雪夫大数定律的推论知:

当n很大时, $$\frac{1}{n}\sum_{k=1}^{n} f(X_k) \approx Ef(X)$$

即 $$\int_a^b f(x)dx \approx \frac{b-a}{n}\sum_{k=1}^{n} f(X_k)。$$

均匀分布怎么模拟离散型分布？

例如： $X \sim \begin{pmatrix} 0 & 1 & 2 \\ 0.5 & 0.2 & 0.3 \end{pmatrix}$



均匀分布怎么模拟连续型分布？

定理：设连续型随机变量X的分布函数为$F(x)$，$Y=F(X)$，则$Y\sim U[0,1]$。

$$X=F^{-1}(Y)$$

其中$F^{-1}(y)$为$F(x)$的广义反函数。

二. 贝努利大数定律

设m是n次重复独立试验中事件A出现的次数，$P(A)=p(0<p<1)$，那么对任意$\varepsilon>0$，有

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

证明：

$$X_k=\begin{cases}1, & 第k次试验中出现A \\ 0, & 第k次试验中出现\overline{A}\end{cases}, k=1,2,\cdots,$$

则X_k服从$0-1$分布，$k=1,2,\cdots,$

则$X_1, X_2, \cdots$独立同分布$, EX_1 = p, DX_1 = pq,$

且$\sum_{k=1}^{n} X_k = m,$由切比雪夫大数定律的推论

显然有

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

此定律说明频率具有稳定性。

注：　称$\frac{m}{n}$依概率收敛于p，记$\frac{m}{n} \xrightarrow{P} p$。

第二节 中心极限定理

林德贝格-勒维中心极限定理

设$X_1, X_2, \cdots$是相互独立同分布的随机变量所构成的序列，并且$EX_1=\mu, DX_1=\sigma^2<\infty$，那么对任意实数$x$，总有

$$\lim_{n\to\infty} P\left(\frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma} \le x\right) = \Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}dt。$$

通俗地说就是

$$\frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma} \overset{n\to\infty}{\sim} N(0,1)。$$

注1:此定理只要求独立同分布,分布可以是离散型,也可以是连续型。

注2: 只要独立同分布,已知数学期望和方差,具体分布不知道,也可以用中心极限定理,例如,习题六第10题（2018-2019第一学期考题）。

当$n(n \geq 100)$很大时,则

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

即 $$\sum_{k=1}^{n} X_k \overset{\text{近似}}{\sim} N(n\mu, n\sigma^2),$$

$$P\left(a \leq \sum_{k=1}^{n} X_k \leq b\right) \approx \Phi\left(\frac{b-n\mu}{\sqrt{n}\sigma}\right) - \Phi\left(\frac{a-n\mu}{\sqrt{n}\sigma}\right),$$

其中$\dfrac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma}$是$\sum_{k=1}^{n} X_k$标准化的随机变量。

一. 当定理中的X_1服从0-1分布,参数是p,

此时$\mu=p, \sigma^2 = pq$,

$$P\left(a \le \sum_{k=1}^{n} X_k \le b\right) \overset{n\text{很大}}{\approx} \Phi\left(\frac{b-np}{\sqrt{npq}}\right) - \Phi\left(\frac{a-np}{\sqrt{npq}}\right)$$

注：　此时我们注意到$\sum_{k=1}^{n} X_k \sim B(n, p)$,

(1)当$np \le 10$时，用普阿松分布近似，

(2)当$np > 10$时，用中心极限定理近似。

二. 当定理中的X_1服从均匀分布$U[a,b]$,

此时$\mu=\dfrac{a+b}{2},\sigma^2=\dfrac{(b-a)^2}{12}$,

$$P\left(a\le\sum_{k=1}^{n}X_k\le b\right)\overset{n\text{很大}}{\approx}\Phi\left(\frac{b-n\frac{a+b}{2}}{\sqrt{n\frac{(b-a)^2}{12}}}\right)-\Phi\left(\frac{a-n\frac{a+b}{2}}{\sqrt{n\frac{(b-a)^2}{12}}}\right)。$$

杨勇制作

140页例6-2

例1: 某单位内部有260部电话分机, 每部分机有4%的时间要使用外线, 各分机是否使用外线相互独立。问总机需要多少条外线, 才能有95%的把握保证各分机使用外线时不必等候?

解: 设需要 a 条外线, 且令

$$X_k = \begin{cases} 1, \text{第}k\text{部分机使用外线} \\ 0, \text{第}k\text{部分机没使用外线} \end{cases}, k = 1, 2, \cdots, 260,$$

$X_1, X_2, \cdots, X_{260}$独立, 并且都服从参数为 $p=0.04$的0-1分布,

而$\sum_{k=1}^{260} X_k$表示同一时刻使用外线的分机数,

由中心极限定理,

$$P\left(\sum_{k=1}^{260} X_k \le a\right) \approx \Phi\left(\frac{a-260\times 0.04}{\sqrt{260\times 0.04\times 0.96}}\right) \ge 0.95,$$

查表得

$$\frac{a-260\times 0.04}{\sqrt{260\times 0.04\times 0.96}}\geq 1.65,$$

解得$a\geq 15.61$,

所以需要至少16条外线。

习题集149页例2

例2: 某电视机厂每月生产10000台电视机, 但是其显像管车间的正品率为0.8, 若以99.7%的概率保证出厂的电视机都装上正品的显像管, 该车间每月至少应该生产多少显像管?

解: 设每月生产n只显像管,

且令

$$X_k=\begin{cases}1, 第k只显像管为正品\\0, 第k只显像管为次品\end{cases}, k=1,2,\cdots,n,$$

此时$X_1, X_2, \cdots, X_n$独立同分布，且均服从参数为$p=0.8$的0-1分布。

又$\sum_{k=1}^{n} X_k$表示 n 个显像管中正品的数量，

依题设条件，n 应当满足

$$P\left(\sum_{k=1}^{n} X_k \geq 10000\right)=0.997,$$

由中心极限定理,

$$P\left(\sum_{k=1}^{n} X_k \geq 10000\right) \approx 1 - \Phi\left(\frac{10000 - n \times 0.8}{\sqrt{n \times 0.8 \times 0.2}}\right)$$

$$= \Phi\left(\frac{0.8n - 10000}{0.4\sqrt{n}}\right) = 0.997,$$

查表得

$$\frac{0.8n - 10000}{0.4\sqrt{n}} = 2.75,$$

解之得

$$\sqrt{n}=112.49\text{(负的舍去)},$$

$$n=12654.0001,$$

故此车间每月至少应当生产12655只显像管。

习题集150页例4

例3:某人要测量甲和乙两地之间的距离,限于测量工具,他分成1200段进行测量,每段测量误差(单位:千米)相互独立,且都服从均匀分布$U[-0.5,0.5]$。试求总距离测量误差的绝对值不超过20千米的概率。

解: 设X_k表示第k段的测量误差,

$k=1,2,\cdots,1200,$

则$X_1, X_2, \cdots, X_{1200}$独立同分布,

且$X_1 \sim U[-0.5, 0.5], EX_1 = 0, DX_1 = \frac{1}{12}$,

显然总误差为$X = \sum_{k=1}^{1200} X_k$，由中心极限定理

$$P(|X| \le 20) = P\left(\left|\sum_{k=1}^{1200} X_k\right| \le 20\right)$$

$$= P\left(-20 \le \sum_{k=1}^{1200} X_k \le 20\right)$$

$$\approx \Phi\left(\frac{20-1200\times 0}{\sqrt{1200\times\frac{1}{12}}}\right)-\Phi\left(\frac{-20-1200\times 0}{\sqrt{1200\times\frac{1}{12}}}\right)$$

$$=2\Phi(2)-1\approx 0.9544\text{。}$$

概率论试卷

2018-2019学年第1学期

2019年1月3日 计算题5

教材142页习题10

5.某生产线上生产的产品成箱包装,每箱的重量是随机的,每箱的平均重量为50千克,标准差为5千克。现用最大载重量为5吨的汽车来运载,试用中心极限定理说明每辆车最多可以装载多少箱产品才能保障不超载的概率大于0.977?$\left(\Phi(2)=0.977\right)$

解: X_k :第k箱的重量, $k=1,2,\cdots,n$,

$X_1,X_2,\cdots,X_n$独立, $EX_k=50, DX_k=5^2=25$,

$$P\left(\sum_{k=1}^{n} X_k \le 5000\right) > 0.977,$$

$$P\left(\sum_{k=1}^{n} X_k \le 5000\right) \approx \Phi\left(\frac{5000-50n}{5\sqrt{n}}\right) > 0.977,$$

$$\frac{5000-50n}{5\sqrt{n}} > 2,$$

$$5n+\sqrt{n}-500<0,$$

$$n<98.01,$$

所以,最多可以装载98箱。

2017－2018年第二学期

2018年6月5日

概率论试卷

计算题5

5.设随机变量$X_1,\cdots,X_n$相互独立,均服从参数为1的泊松分布,定义随机变量

$$若Y_k=\begin{cases}1, X_k=0\\0,否则\end{cases},(1\le k\le n)。$$

(1)记$\bar{Y}=\dfrac{Y_1+\cdots+Y_n}{n}$,求当$n$足够大时$\bar{Y}$的近似分布;

(2)利用切比雪夫不等式估计,当n至少取多大时,可使$P(|\bar{Y}-e^{-1}|<0.1)\ge 0.8$?

（注:e^{-1}可用0.4近似）。

解：　$X_k \sim P(1)$,

$$P(Y_k=1)=P(X_k=0)=\frac{1^0}{0!}e^{-1}=e^{-1},$$

$$Y_k \sim \begin{pmatrix} 0 & 1 \\ 1-e^{-1} & e^{-1} \end{pmatrix},\quad EY_k=e^{-1}, DY_k=e^{-1}(1-e^{-1}),$$

显然$Y_1,\cdots,Y_n$独立同分布。

(1)　由中心极限定理知,

$$\bar{Y} \overset{\bullet}{\sim} N\left(EY_1,\frac{DY_1}{n}\right)=N\left(e^{-1},\frac{e^{-1}-e^{-2}}{n}\right)。$$

(2)　由切比谢夫不等式,

$$P\left(\left|\overline{Y}-e^{-1}\right|<0.1\right)=P\left(\left|\overline{Y}-E\overline{Y}\right|<0.1\right)\geq 1-\frac{D\overline{Y}}{0.1^2}\geq 0.8,$$

$$1-\frac{0.4\times 0.6}{0.01n}\geq 0.8,\quad n\geq 120。$$

2017－2018年第一学期

2017年12月5日

概率论试卷　计算题5

5.假设某种型号的螺丝钉的重量是随机变量,期望值为50克,标准差为5克。求

(1) 100个螺丝钉一袋的重量超过5. 1千克的概率;

(2) 每箱螺丝钉装有500袋，500袋中最多有4%的重量超过5. 1千克的概率。

解： (1)

X_i :第i个螺丝钉的重量，$i=1,2,\cdots,100,$

可以认为$X_1, X_2, \cdots, X_{100}$独立,

$$p = P\left(\sum_{i=1}^{100} X_i > 5.1 \times 1000\right)$$

$$\approx 1 - \Phi\left(\frac{5100 - 5000}{\sqrt{100 \times 5^2}}\right) = 1 - \Phi(2)$$

$$= 1 - 0.97725 = 0.02275;$$

(2) $Y:500$袋中重量超过5.1千克的袋数,

$$Y \sim B(500, 0.02275),$$

$$Y_k=\begin{cases}1, \text{第}k\text{袋重量超过5.1千克} \\ 0, \text{第}k\text{袋重量不超过5.1千克}\end{cases}, k=1,2,\cdots,500,$$

$Y_1,Y_2,\cdots,Y_{500}$相互独立,

$$P\left(Y=\sum_{k=1}^{500}Y_k\le 20\right)\approx\Phi\left(\frac{20-500\times 0.02275}{\sqrt{500\times 0.02275\times 0.97725}}\right)$$

$$=\Phi(2.58698)=0.995。$$

祝大家都考一个好成绩

心想事成

谢谢大家上我的课