第五章　大数定律和中心极限定理

5.1　大数定律

1. **切比雪夫**(П. Л. Чебышев)**不等式**：若随机变量 ξ 的方差 $D(\xi)$ 存在，则对任意 $\varepsilon>0$，有

$$P(|\xi-E(\xi)|\geqslant\varepsilon)\leqslant\frac{D(\xi)}{\varepsilon^2},$$

或

$$P(|\xi-E(\xi)|<\varepsilon)\geqslant1-\frac{D(\xi)}{\varepsilon^2}.$$

推论：若 $D(\xi)=0$，则 ξ 以概率为 1 地等于它的数学期望 $E(\xi)$，即 $P(\xi=E(\xi))=1$。

2. 设 $\xi_1,\xi_2,\cdots,\xi_n,\cdots$ 是一随机变量序列，若存在随机变量 ξ，使对任意 $\varepsilon>0$，恒有

$$\lim_{n\to\infty}P(|\xi_n-\xi|\geqslant\varepsilon)=0,$$

或等价于

$$\lim_{n\to\infty}P(|\xi_n-\xi|<\varepsilon)=1,$$

则称随机变量序列 $\{\xi_n\}$ 依概率收敛于随机变量 ξ(ξ 也可以是一个常数).

3. **切比雪夫**(П. Л. Чебышев)**大数定律**：设 $\xi_1,\xi_2,\cdots,\xi_n,\cdots$ 是相互独立的随机变量序列，分别有有限的数学期望和方差，并且它们的方差有公共上界 c(常数)，$D(\xi_k)\leqslant c(k=1,2,\cdots,n,\cdots)$，则对任意 $\varepsilon>0$，恒有

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

推论：(切比雪夫大数定律的特殊情况)　设随机变量 $\xi_1,\xi_2,\cdots,\xi_n,\cdots$ 是相互独立的，服从同一分布且有相同的数学期望 μ 和方差 σ^2，则对任意 $\varepsilon>0$，恒有

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

4. **伯努利**(Bernoulli)**大数定律**：设 μ_n 是 n 重伯努利试验中事件 A 发生的次数，p 是事件 A 在每次试验中出现的概率，则对任意 $\varepsilon>0$，恒有

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

伯努利大数定律表明：一件事发生的频率 $\frac{\mu_n}{n}$ 依概率收敛于它的概率 p.

5. **泊松**(Poisson)**大数定律**：如果在一个独立试验序列中，事件 A 在第 k 次试验中出现的概率等于 p_k，以 u_k 记在前 n 次试验中事件 A 出现的次数，则对任意 $\varepsilon>0$，恒有

$$\lim_{n\to\infty}P\left(\left|\frac{u_n}{n}-\frac{p_1+p_2+\cdots+p_n}{n}\right|<\varepsilon\right)=1.$$

5.2　中心极限定理

1. **独立同分布的中心极限定理**：设随机变量 $\xi_1,\xi_2,\cdots,\xi_n,\cdots$ 是相互独立同分布的序列，且具有有限的数学期望和方差，$E(\xi_k)=\mu,D(\xi_k)=\sigma^2\neq0(k=1,2,\cdots)$，则随机变量

$$\eta_n=\frac{\sum_{k=1}^{n}\xi_k-n\mu}{\sqrt{n}\,\sigma}$$

的分布函数 $F_n(x)$，对任意 x，满足

$$\lim_{n\to\infty}F_n(x)=\lim_{n\to\infty}P(\eta_n\leqslant x)=\lim_{n\to\infty}P\left(\frac{\sum_{k=1}^{n}\xi_k-n\mu}{\sqrt{n}\,\sigma}\leqslant x\right)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt.$$

2. **德莫佛—拉普拉斯中心极限定理**：设随机变量 η_n 服从 $B(n,p)$，则对于区间 (a,b)，恒有

$$\lim_{n\to\infty}P\left(a<\frac{\eta_n-np}{\sqrt{np(1-p)}}\leqslant b\right)=\int_{a}^{b}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt.$$

问 5.1　切比雪夫不等式有什么作用？它的意义是什么？

答　切比雪夫不等式 $P(|\xi-E(\xi)|<\varepsilon)\geqslant1-\frac{D(\xi)}{\varepsilon^2}$，反映了随机变量 ξ 的取值落在数学期望 $E(\xi)=\mu$ 的 ε 领域内的概率不小于 $1-\frac{\sigma^2}{\varepsilon^2}$，它的意义在于：当知道随机变量 ξ 的数学期望和方差时，我们可以估计 ξ 落在以 $E(\xi)$ 为中心的某一区间的概率(至少能得出一个下限).

它的作用大体有四个方面：

(1)估计概率，当 $E(\xi)$，$D(\xi)$ 及 ε 给定时，可以依公式直接计算.(2)当 $E(\xi)$，$D(\xi)$ 及概率确定时，估计所需要的区间长度(也就是确定 ε).(3)估计试验次数.在 n 重伯努利试验中，频率 n_a/n 与试验次数有关，在已知 $E(\xi)$，$D(\xi)$ 和 p 时，可用切比雪夫不等式确定 n.(4)它是导出其他定理的依据.

问 5.1　切比雪夫不等式有什么作用？它的意义是什么？

答　切比雪夫不等式 $P(|\xi-E(\xi)|<\varepsilon)\geqslant1-\frac{D(\xi)}{\varepsilon^2}$，反映了随机变量 ξ 的取值落在数学期望 $E(\xi)=\mu$ 的 ε 领域内的概率不小于 $1-\frac{\sigma^2}{\varepsilon^2}$，它的意义在于：当知道随机变量 ξ 的数学期望和方差时，我们可以估计 ξ 落在以 $E(\xi)$ 为中心的某一区间的概率(至少能得出一个下限).

它的作用大体有四个方面：

(1)估计概率，当 $E(\xi)$，$D(\xi)$ 及 ε 给定时，可以依公式直接计算.(2)当 $E(\xi)$，$D(\xi)$ 及概率确定时，估计所需要的区间长度(也就是确定 ε).(3)估计试验次数.在 n 重伯努利试验中，频率 n_a/n 与试验次数有关，在已知 $E(\xi)$，$D(\xi)$ 和 p 时，可用切比雪夫不等式确定 n.(4)它是导出其他定理的依据.

但是读者要清楚，切比雪夫不等式只能给出估计值，精确值要用别的更好的方法来计算.

问 5.2　大数定律的意义是什么？

答　大数定律深刻地揭示了随机事件的概率与频率之间的关系，因此是概率论的重要理论基础.大数定律从大量测量的平均值出发，讨论并反映了算术平均值及频率和稳定性.

教材讲述的大数定律是弱大数定律，它们的条件各不相同，但结论是一致的：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.

问 5.3　中心极限定理有什么实际意义？

答　正态分布是概率论中三个重要的分布之一，它是现实生活和科学技术中使用最多的一种分布，也是数理统计的最重要假设.许多随机变量本身并不属于正态分布，但在它们的共同作用下形成的随机变量的极限分布是正态分布，它们的概率如何计算是一个很重要的问题.中心极限定理阐明了在什么条件下、原本不属于正态分布的一些随机变量其总和分布渐近服从正态分布.

问 5.4 大数定律与中心极限定理有什么异同?

答 大数定律与中心极限定理都是通过极限理论来研究概率问题,研究对象都是随机变量序列,解决的都是概率论中的基本问题,因而大数定律与中心极限定理在概率论中的意义都十分重要.

它们的不同在于:大数定律给出的是当 $n\to\infty$ 时随机变量序列的函数(平均值或概率的极限;而中心极限定理则告诉我们随机变量序列总和的分布近似正态分布,总和的标准化随机变量服从渐近标准正态分布,而不论随机变量序列服从何种分布.它是近两个世纪概率论研究的中心问题,所以称为中心极限定理.

综合练习题五

$EX=\int_1^{+\infty} 2x^{-2}dx$

1. $\xi_1,\xi_2,\cdots$ 相互独立,$\xi_i\sim\varphi(x)$,$\varphi(x)=2x^{-3}(x\geqslant 1,i=1,2,\cdots)$,则有(　　).

(A)对每一个 $\xi_i(i=1,2,\cdots)$ 都满足切比雪夫不等式

$=-\frac{2}{x}\Big|_1^{+\infty}$

(B)$\xi_i(i=1,2,\cdots)$ 都不满足切比雪夫不等式

(C)$\xi_i(i=1,2,\cdots)$ 满足切比雪夫大数定律

$=2$

(D)$\xi_i(i=1,2,\cdots)$ 不满足切比雪夫大数定律条件

$DX=\int_1^{+\infty} 2x^{-1}dx=2\ln x\Big|_1^{+\infty}$

解 由切比雪夫不等式及切比雪夫大数定律可知选(B)、(D).

2. $\xi_1,\xi_2,\cdots,\xi_9$ 相互独立,$E(\xi_i)=1,D(\xi_i)=1(i=1,2,\cdots,9)$则对任意给定的 $\varepsilon>0$,有(　　).

(A)$P\left(\left|\sum\limits_{i=1}^{9}\xi_i-1\right|<\varepsilon\right)\geqslant 1-\varepsilon^{-2}$　(B)$\left(\frac{1}{9}\left|\sum\limits_{i=1}^{9}\xi_i-1\right|<\varepsilon\right)\geqslant 1-\varepsilon^{-2}$

(C)$P\left(\left|\sum\limits_{i=1}^{9}\xi_i-9\right|<\varepsilon\right)\geqslant 1-\varepsilon^{-2}$　(D)$P\left(\left|\sum\limits_{i=1}^{9}\xi_i-9\right|<\varepsilon\right)\geqslant 1-9\varepsilon^{-2}$

解 由题意得,$E(\xi)=\sum\limits_{i=1}^{9}E(\xi_i)=9,D(\xi)=D(\sum\limits_{i=1}^{9}\xi_i)=\sum\limits_{i=1}^{9}D(\xi_i)=9.$

由 $P\left(\left|\sum\limits_{i=1}^{9}\xi_i-E(\xi)\right|<\varepsilon\right)\geqslant 1-\frac{D(\xi)}{\varepsilon^2}$.　　故应选(D).

3. 设 $\xi_1,\xi_2,\cdots,\xi_i,\cdots$,为独立随机变量序列,且服从参数为 λ 的指数分布,则(　　).

(A)$\lim\limits_{n\to\infty}P\left\{\frac{\sum\limits_{i=1}^{n}\xi_i-\frac{n}{\lambda}}{n/\lambda^2}\leqslant x\right\}=\Phi(x)$　(B)$\lim\limits_{n\to\infty}P\left\{\frac{\sum\limits_{i=1}^{n}\xi_i-\frac{1}{\lambda}}{1/\lambda^2}\leqslant x\right\}=\Phi(x)$

3. 设 $\xi_1,\xi_2,\cdots,\xi_i,\cdots$,为独立随机变量序列,且服从参数为 λ 的指数分布,则(　　).

(A)$\lim\limits_{n\to\infty}P\left\{\frac{\sum\limits_{i=1}^{n}\xi_i-\frac{n}{\lambda}}{n/\lambda^2}\leqslant x\right\}=\Phi(x)$　(B)$\lim\limits_{n\to\infty}P\left\{\frac{\sum\limits_{i=1}^{n}\xi_i-\frac{1}{\lambda}}{1/\lambda^2}\leqslant x\right\}=\Phi(x)$

(C)$\lim\limits_{n\to\infty}P\left\{\frac{\lambda\sum\limits_{i=1}^{n}\xi_i-n}{\sqrt{n}}\leqslant x\right\}=\Phi(x)$　(D)$\lim\limits_{n\to\infty}P\left\{\frac{\lambda\sum\limits_{i=1}^{n}\xi_i-n}{n}\leqslant x\right\}=\Phi(x)$

解 由题意知,$E(\xi_i)=\frac{1}{\lambda},D(\xi_i)=\frac{1}{\lambda^2}$.

由独立同分布的中心极限定理

$$\lim_{n\to\infty}P\left\{\frac{\sum\limits_{i=1}^{n}\xi_i-n\mu}{\sqrt{n}\,\sigma}\leqslant x\right\}=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt=\Phi(x),$$

即$\lim\limits_{n\to\infty}P\left\{\frac{\lambda\sum\limits_{i=1}^{n}\xi_i-n}{\sqrt{n}}\leqslant x\right\}=\Phi(x)$.　故应选(C).

4. 设 $\xi_1,\xi_2,\cdots,\xi_n,\cdots$为独立随机变量序列,且 $\xi_i(i=1,2,\cdots)$服从参数为 λ 的泊松分布,则(　　)成立.

(A)$\lim\limits_{n\to\infty}P\left\{\frac{\sum\limits_{i=1}^{n}\xi_i-n\lambda}{\sqrt{n}}\leqslant x\right\}=\Phi(x)$

(B)当 n 充分大时,$\sum\limits_{i=1}^{n}\xi_i$ 近似服从 $N(0,1)$分布

(C)当 n 充分大时,$\sum\limits_{i=1}^{n}\xi_i$ 近似服从 $N(n\lambda,n\lambda)$分布

(D)当 n 充分大时,$P\left(\sum\limits_{i=1}^{n}\xi_i\leqslant x\right)=\Phi(x)$

解 因为 $\xi_i(i=1,2,\cdots)$服从参数为 λ 的泊松分布,故 $E(\xi_i)=\lambda,D(\xi_i)=\lambda$,由独立同分布的中心极限定理得

$$\lim_{n\to\infty}P\left\{\frac{\sum\limits_{i=1}^{n}\xi_i-n\lambda}{\sqrt{n\lambda}}\leqslant x\right\}=\Phi(x),$$

故(A)不对.由中心极限定理知,当 n 充分大时,$\sum\limits_{i=1}^{n}\xi_i$ 近似服从 $N(n\lambda,n\lambda)$分布,故选(C).

1. **填空题**

(1)设随机变量 ξ 和 η 的数学期望都是 2,方差分别为 1 和 4,而相关系数为 -0.5,则根据切比雪夫不等式 $P(|\xi-\eta|\geqslant 6)\leqslant$________.

因为　$E(\xi-\eta)=E(\xi)-E(\eta)=0,$

$D(\xi+\eta)=D(\xi)+D(\eta)+2\mathrm{cov}(\xi\eta)$

$=D(\xi)+D(\eta)+2\rho_{\xi\eta}\sqrt{D(\xi)D(\eta)}$

$=1+4-2\times 0.5\times 2=3.$

根据切比雪夫不等式

$$P(|\xi-E(\xi)|\geqslant\varepsilon)\leqslant\frac{D(\xi)}{2},$$

所以 $P\{|\xi-\eta|\geqslant 6\}\leqslant\frac{3}{36}=\frac{1}{12}$.

(2)设随机变量 $\xi_1,\xi_2,\cdots,\xi_n$,独立同分布,且 $E(\xi_i)=\mu,D(\xi_i)=\sigma^2>0,i=1,2,K$,那么$\lim\limits_{n\to\infty}P\left\{\frac{\sum\limits_{i=1}^{n}\xi_i-n\mu}{\sqrt{n}\,\sigma}\leqslant x\right\}=$________.

(2)设随机变量 $\xi_1,\xi_2,\cdots,\xi_n$,独立同分布,且 $E(\xi_i)=\mu,D(\xi_i)=\sigma^2>0,i=1,2,K$,那么$\lim\limits_{n\to\infty}P\left\{\frac{\sum\limits_{i=1}^{n}\xi_i-n\mu}{\sqrt{n}\,\sigma}\leqslant x\right\}=$________.

由独立同分布的中心极限定理

$$\lim_{n\to\infty}P\left\{\frac{\sum\limits_{k=1}^{n}\xi_k-n\mu}{\sqrt{n}\,\sigma}\leqslant x\right\}=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt.$$

(3)设随机变量 η_n 服从 $B(n,p)$,则对于区间(a,b),恒有

$\lim\limits_{n\to\infty}P\left(a<\frac{\eta_n-np}{\sqrt{np(1-p)}}\leqslant b\right)=$________.

根据德莫佛—拉普拉斯中心极限定理

$$\lim_{n\to\infty}P\left(a<\frac{\eta_n-np}{\sqrt{np(1-p)}}\leqslant b\right)=\int_a^b\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt.$$

(4)某计算机系统有 120 个终端，每个终端有 5%的时间在使用，若各个终端使用与否相互独立，则有 10 个或更多个终端在使用的概率是________.

设 ξ 为使用的终端数，则 $\xi \sim B(120, 0.05)$.

$$P(\xi \geqslant 10) = 1 - P(\xi < 10) \approx 1 - \Phi\left(\frac{10-6}{\sqrt{0.95 \times 6}}\right)$$
$$= 1 - \Phi(1.675) = 1 - 0.952\ 5 = 0.047\ 5.$$

(5)一大批产品中优质品占一半，现每次抽取一次，看后放回再抽，问在 100 次抽取中取到优质品次数不超过 45 的概率等于________.

令 η 表示 100 次抽取中取到优质品的次数，

$$\xi_i = \begin{cases} 1, & \text{第 } i \text{ 次取到优质品;} \\ 0, & \text{第 } i \text{ 次没有取到优质品,} \end{cases} \quad i = 1, 2, \cdots, 100$$

则 $\eta = \sum_{i=1}^{100} \xi_i$，$\eta \sim B(100, 0.5)$.

那么 $E(\eta) = 100 \times 0.5 = 50$，$D(\eta) = 100 \times 0.5 \times 0.5 = 25$.

(5)一大批产品中优质品占一半，现每次抽取一次，看后放回再抽，问在 100 次抽取中取到优质品次数不超过 45 的概率等于________.

令 η 表示 100 次抽取中取到优质品的次数，

$$\xi_i = \begin{cases} 1, & \text{第 } i \text{ 次取到优质品;} \\ 0, & \text{第 } i \text{ 次没有取到优质品,} \end{cases} \quad i = 1, 2, \cdots, 100$$

则 $\eta = \sum_{i=1}^{100} \xi_i$，$\eta \sim B(100, 0.5)$.

那么 $E(\eta) = 100 \times 0.5 = 50$，$D(\eta) = 100 \times 0.5 \times 0.5 = 25$.

由德莫佛—拉普拉斯中心极限定理可得

$$P(\eta \leqslant 45) = P\left(\frac{\eta - 50}{\sqrt{D\eta}} \leqslant \frac{45 - 50}{\sqrt{D\eta}}\right) = P\left(\frac{\eta - 50}{5} \leqslant -1\right)$$
$$\approx \Phi(-1) = 1 - \Phi(1) = 1 - 0.841\ 3 = 0.158\ 7.$$

(1)据以往经验，某种电器元件寿命服从均值为 100 小时的指数分布，现随机地取 16 只，设它们的寿命是相互独立的，求这 16 只元件的寿命的总和大于 1 920 小时的概率.

解 记 16 只电器元件的寿命分别为 $\xi_1, \xi_2, \cdots, \xi_{16}$，则这 16 只电器元件的寿命之和为 $\xi = \sum_{i=1}^{16} \xi_i$，依题意，$E(\xi_i) = 100$，$D(\xi_i) = 100^2$，根据独立同分布的中心极限定理

$$\eta = \frac{\sum_{i=1}^{16} \xi_i - 16 \times 100}{4 \times 100} = \frac{\xi - 1600}{400},$$

近似地服从 $N(0,1)$，于是

$$P(\xi > 1\ 920) = 1 - P(\xi \leqslant 1\ 920)$$
$$= 1 - P\left(\frac{\xi - 1\ 600}{400} \leqslant \frac{1\ 920 - 1\ 600}{400}\right)$$
$$= 1 - \Phi(0.8) = 0.211\ 9.$$

(2)某校有 1 000 名学生，每人以 80%的概率去图书馆自习，问图书馆至少应设多少座位，才能以 99%的概率保证去上自习的同学都有座位？

解 设 ξ 表示同时去图书馆上自习的人数，并设图书馆至少设 n 个座位，才能以 99%的概率保证去上自习的同学都有座位.

即 n 满足 $P(\xi \leqslant n) \geqslant 0.99$.

因为 $\xi \sim B(1\ 000, 0.8)$，

所以 $P(\xi \leqslant n) \approx \Phi\left(\frac{n - 1\ 000 \times 0.8}{\sqrt{1\ 000 \times 0.8 \times 0.2}}\right) - \Phi\left(\frac{0 - 1\ 000 \times 0.8}{\sqrt{1\ 000 \times 0.8 \times 0.2}}\right)$

$$= \Phi\left(\frac{n - 800}{12.65}\right) \geqslant 0.99.$$

查表得 $\frac{n - 800}{12.65} \geqslant 2.33$，故 $n \geqslant 829.5$.

因此图书馆至少应该有 830 个座位.