第十五课

5．1 大数定律

定义：称随机变量序列 $X_1,X_2,\ldots,X_n,\ldots$是相互独立的，又若 X_k (k=1,2,3,…)有同一的分布函数，则称 $X_1,X_2,\ldots,X_n,\ldots$为独立同分布的随机变量序列。

定理 1：(契比雪夫大数定律)
设 $X_1,X_2,\ldots,X_n,\ldots$是独立同分布的随机变量序列，并且 $E(X_k)=\mu$,$D(X)=\sigma^2$ (k=1,2,…)则对任意正数 ε，有

$$\lim_{n\to+\infty} P\{|\frac{1}{n}\sum_{k=1}^{n} X_k-\mu|\ge\varepsilon\}=0$$

证：

因为 $E(\frac{1}{n}\sum_{k=1}^{n}X_k)=\frac{1}{n}\sum_{k=1}^{n}E(X_k)=\frac{1}{n}n\mu=\mu$， $D(\frac{1}{n}\sum_{k=1}^{n}X_k)=\frac{1}{n^2}\sum_{k=1}^{n}D(X_k)=\frac{1}{n^2}n\sigma^2=\frac{\sigma^2}{n}$

由契比雪夫不等式有 $P\{|\frac{1}{n}\sum_{k=1}^{n}X_k-\mu|<\varepsilon\}\ge 1-\frac{\frac{\sigma^2}{n}}{\varepsilon^2}$

取极限（注意概率不能大于 1），所以 $\lim_{n\to+\infty} P\{|\frac{1}{n}\sum_{k=1}^{n}X_k-\mu|<\varepsilon\}=1$

说明：平均值与数学期望的差大于某正数 ε 的概率当 n→+∞时为 0，即当 n 较大时，$\frac{1}{n}\sum_{k=1}^{n}X_k$ 在概率条件下等于数学期望 μ。以后在统计中就是用 $\frac{1}{n}\sum_{k=1}^{n}X_k$ 来代替均值 μ

定理 2：

(贝努利大数定律)
设 n_A 为 n 次独立试验中事件 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 ε，总有 $\lim_{n\to+\infty} P\{|\frac{n_A}{n}-p|<\varepsilon\}=1$

说明：当 n→+∞时，事件 A 发生的频率 $\frac{n_A}{n}$ 与概率 p，在概率条件下相差无几。

5．2 中心极限定理

在什么样的条件下，随机变量的和标准化后的分布以正态分布为极限。

定理 1：

(中心极限定理)
设 $X_1,X_2,\dots,X_n,\dots$是独立同分布的随机变量序列，并且有数学期望和方差 $E(X_k)=\mu$,$D(X)=\sigma^2\neq 0$ $(k=1,2,\dots)$，

令 $Y_n=\dfrac{\sum_{k=1}^{n}X_k-E(\sum_{k=1}^{n}X_k)}{\sqrt{D(\sum_{k=1}^{n}X_k)}}=\dfrac{\sum_{k=1}^{n}X_k-n\mu}{\sqrt{n}\sigma}$，其分布函数 $F_n(x)=P\{Y_n\leqslant x\}$，则对任意 x 有

$$\lim_{n\to+\infty}F_n(x)=\lim_{n\to+\infty}P\{Y_n\le x\}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}dt=\Phi(x)$$

推论：

(德莫佛-拉普拉斯定理)
设随机变量 μ_n 服从二项分布 B(n,p)，则对任意区间 (x_1,x_2)有

$$\lim_{b\to+\infty}P\{x_1<\frac{\mu_n-np}{\sqrt{np(1-p)}}\le x_2\}=\int_{x_1}^{x_2}e^{-\frac{t^2}{2}}dt=\Phi(x_2)-\Phi(x_1)$$

说明：当 μ_n 服从二项分布 B(n,p)，当 n 充分大时，

$$P\{x_1<\mu_n\le x_2\}\approx\Phi(\frac{x_2-np}{\sqrt{np(1-p)}})-\Phi(\frac{x_1-np}{\sqrt{np(1-p)}})$$

例 1：一加法器同时收到 20 个噪声电压 V_k (k=1,2,…,20)，设它们是相互独立的随机变量，且都在(0,10)上服从均匀分布，记 $V=\sum_{k=1}^{20}V_k$，求 P{V>105}。

这是一个电子随机设计中的处理问题的方法。即在 20 个噪声电压加起来的和电压大于 105 伏的概率。

解：因为 V_k 服从(0,10)上的均匀分布，

所以数学期望 $E(V_k)=\frac{0+10}{2}=5$，方差 $D(V_k)=\frac{(10-0)^2}{12}=\frac{25}{3}$ （k=1,2,…,20）

由中心极限定理，随机变量函数 $Z=\frac{\sum_{k=1}^{20}V_k-nE(V_k)}{\sqrt{nD(V_k)}}=\frac{V-20\times5}{\sqrt{20\times\frac{25}{3}}}$ 近似服从正态分布 N(0,1)，于是

$$P\{V>105)=P\{\frac{V-20\times5}{\sqrt{20\times\frac{25}{3}}}>\frac{105-20\times5}{\sqrt{20\times\frac{25}{3}}}\}$$

$$=P\{\frac{V-20\times5}{\sqrt{20\times\frac{25}{3}}}>0.387\}$$

$$=1-P\{\frac{V-20\times5}{\sqrt{20\times\frac{25}{3}}}\le0.387\}\approx1-\Phi(0.387)=0.348$$

两边同时进行整理，目的是为了凑出 $\frac{V-20\times5}{\sqrt{20\times\frac{25}{3}}}$ 形式的函数来，因为它是服从标准正态分布的。

例 2：某电站供应 10000 户居民用电，设在高峰时每户用电的概率为 0.8，且各户用电量多少是相互独立的。求：

1、 同一时刻有 8100 户以上用电的概率；

2、 若每户用电功率为 100W，则电站至少需要多少电功率才能保证以 0.975 的概率供应居民用电？

解：（1）设随机变量 Y_n 表示 10000 户中在同一时刻用电的户数，则 $Y_n \sim B(10000,0.8)$，于是

np=10000X0.8=8000， $\sqrt{np(1-p)} = \sqrt{10000\times 0.8\times 0.2} = 40$

所以概率为 $P\{8100 \le Y_n \le 10000\} = P\{\frac{8100-np}{\sqrt{np(1-p)}} \le \frac{Y_n - np}{\sqrt{np(1-p)}} \le \frac{10000-np}{\sqrt{np(1-p)}}\}$

$$= P\{2.5 \le \frac{Y_n - np}{\sqrt{np(1-p)}} \le 50\} \approx \Phi(50) - \Phi(2.5) = 1 - 0.9938 = 0.0062$$

(2)若每户用电功率为 100W，则 Y_n 户用是功率为 $100Y_n$W，设电站供电功率为 QW，则按题意有

$$P\{100Y_n \le Q\} = P\{Y_n \le \frac{Q}{100}\} = P\{\frac{Y_n - np}{\sqrt{np(1-p)}} \le \frac{\frac{Q}{100} - 8000}{40}\} \approx \Phi(\frac{\frac{Q}{100} - 8000}{40}) = 0.975$$

查正态分布表得 $\Phi(1.96)=0.975$，所以 $\dfrac{\dfrac{Q}{100}-8000}{40}=1.96$，解得 Q=807840

所以，电站供电功率应不少于 807.84 kW.

通过上面的例子的学习，我们可以看到用概率的方法来解决问题是非常服合实际的。这正是概率方法被广泛应用的原因。