概率论与数理统计

第16讲

大数定律、中心极限定理

一. 大数定律

概率论是研究随机现象统计规律性的学科。随机现象的统计规律性可以通过大量重复试验呈现出来。所以常常需要研究大量随机现象。

研究大量的随机现象，常常采用极限形式，由此导致了对极限定理的研究。极限定理的内容很广泛，其中最重要的有两种：**大数定律与中心极限定理。**

大数定律主要讲述随机变量前一些项的平均值在一定条件下的稳定性问题；**中心极限定理**主要讲述在一定条件下，大量随机变量之和的分布函数的极限逼近于正态分布。

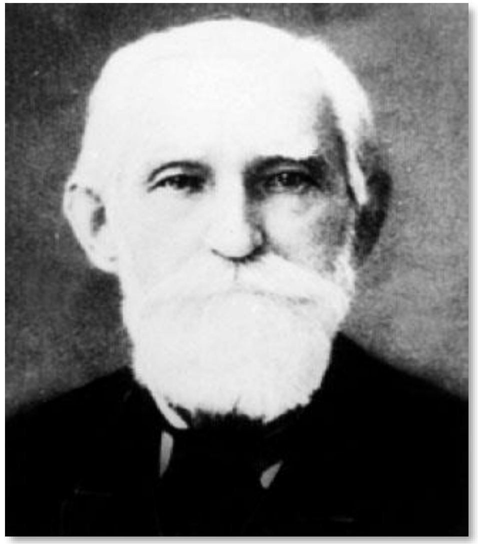
在给出大数定律之前，首先介绍概率论中的一个重要不等式——切比雪夫不等式以及用数学语言描述大数定律时用到的概率意义下的极限定义.

定理1(切比雪夫不等式) 设随机变量X的数学期望$E(X)$和方差$D(X)$都存在, 则对任意的$\varepsilon>0$, 均有

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

或等价地写成

$$P\{|X-E(X)|<\varepsilon\}\geq 1-\frac{D(X)}{\varepsilon^2}$$



切比雪夫
1821~1894

证明：设X为连续型随机变量，其概率密度为$f(x)$。

则有：

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

$$\leq\frac{1}{\varepsilon^2}\int_{-\infty}^{+\infty}(x-EX)^2f(x)dx=\frac{DX}{\varepsilon^2}$$

$$P\{|X-E(X)|\geq\varepsilon\}\leq\frac{D(X)}{\varepsilon^2}\qquad P\{|X-E(X)|<\varepsilon\}\geq1-\frac{D(X)}{\varepsilon^2}$$

切比雪夫不等式的重要性在于： 不管随机变量的分布类型是什么，且不管其分布是否已知，只要知道它的数学期望和方差，就可以对随机变量落入数学期望附近的区域（$EX-\varepsilon$，$EX+\varepsilon$）或（$-\infty\varepsilon$，$EX-\varepsilon$）$\cup$（$EX+\varepsilon$, $+\infty$）的概率给出一个下界或上界。

$$P\{|X-E(X)|\geq\varepsilon\}\leq\frac{D(X)}{\varepsilon^2}\qquad P\{|X-E(X)|<\varepsilon\}\geq 1-\frac{D(X)}{\varepsilon^2}$$

由切比雪夫不等式可以看出，对于同一个ε，方差DX越小，则事件$\{|X-E(X)|<\varepsilon\}$的概率越大，即随机变量X落入区间（$EX-\varepsilon$，$EX+\varepsilon$）的概率就越大，也就是说随机变量X集中在期望附近的可能性越大，这也就进一步说明了方差这个数字特征的确刻画了随机变量X与其数学期望的离散程度。

例 1. 设随机变量 X 服从 $N(\mu,\sigma^2)$，利用切比雪夫不等式估计 $P\{|X-\mu|\le 3\sigma\}$ 的下界。

解：易知 $E(X)=\mu$，$D(X)=\sigma^2$。 在切比雪夫不等式中取 $\varepsilon=3\sigma$，则有

$$P\{|X-\mu|\le 3\sigma\}\ge 1-\frac{\sigma^2}{(3\sigma)^2}=\frac{8}{9}\approx 0.8889$$

另外，由正态分布的 3σ 原则知， $P\{|X-\mu|\le 3\sigma\}\approx 0.9974$

因此，可以看出利用切比雪夫不等式估计的概率精度比较低。但这个不等式的理论意义远大于它的实际意义。例如，运用它可以证明大数定律。

例2.在每次试验中，事件A发生的概率为0.75，利用切比雪夫不等式求：n需要多么大时，才能使得在n次独立重复试验中，事件A出现的频率在0.74~0.76之间的概率至少为0.90?

解：设X为n次试验中事件A出现的次数。 则$X\sim b(n, 0.75)$

$$E(X)=0.75n,\quad D(X)=0.75\times0.25n=0.1875n$$

所求为满足 $P(0.74<\dfrac{X}{n}<0.76)\geq 0.90$ 的n。

$$P\{0.74<\frac{X}{n}<0.76\}=P\{0.74n<X<0.76n\}=P\{-0.01n<X-0.75n<0.01n\}$$

$$=P\{|X-E(X)|<0.01n\}$$

$$P\{0.74<\frac{X}{n}<0.76\}=P\{|X-E(X)|<0.01n\}$$

在切比雪夫不等式中取$\varepsilon=0.01n$，则

$$P\{0.74<\frac{X}{n}<0.76\}\geq 1-\frac{D(X)}{(0.01n)^2}=1-\frac{0.1875n}{0.0001n^2}=1-\frac{1875}{n}\geq 0.9$$

解得 $n\geq\dfrac{1875}{1-0.9}=18750$

即n取18750时，可以使得在n次独立重复试验中，事件A出现的频率在0.74~0.76之间的概率至少为0.90。

定义1 设$X_1, X_2, \ldots, X_n, \ldots$是随机变量序列，$a$是一个常数。若对任意的$\varepsilon>0$，有 $\lim\limits_{n\to+\infty} P\{|X_n-a|\geq\varepsilon\}=0$

则称序列$X_1, X_2, \ldots, X_n, \ldots$依概率收敛于常数$a$，记为 $X_n \xrightarrow{P} a$

或等价地写为 $\lim\limits_{n\to+\infty} P\{|X_n-a|<\varepsilon\}=1$

依概率收敛意味着，当n很大时，X_n与常数a的绝对偏差$|X_n-a|$不小于任一给定量的可能性将随着n的增大而越来越小；或者说绝对偏差$|X_n-a|$不小于任一给定量的可能性随着n的增大越来越接近于1，此为一个大概率事件。

请注意，这种收敛性是概率意义下的一种收敛，而不是数学意义下的一般的数列收敛。依概率收敛比数学意义下的数列收敛要弱一些。

以下用$\{X_n, n\geq1\}$表示随机变量序列$X_1, X_2, \ldots, X_n, \ldots$

定理2 $\{X_n, n\geq1\}$和$\{Y_n, n\geq1\}$是两个随机变量序列， a，b是两个常数，如果 $X_n \xrightarrow{P} a$，$Y_n \xrightarrow{P} b$

且$h(x)$在a点连续，$g(x, y)$在点(a, b)连续，则

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

推论1 $\{X_n, n\geq 1\}$和$\{Y_n, n\geq 1\}$是两个随机变量序列， a，b是两个常数，

如果 $X_n \xrightarrow{P} a$, $Y_n \xrightarrow{P} b$ 则有

(1) $X_n \pm Y_n \xrightarrow{P} a \pm b$

(2) $X_n \cdot Y_n \xrightarrow{P} a \cdot b$

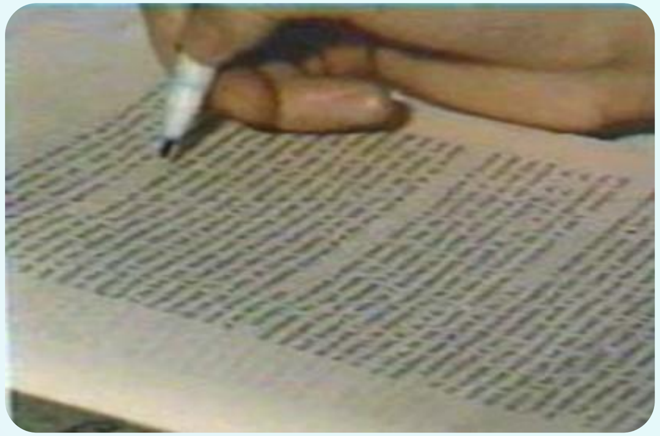
(3) $X_n / Y_n \xrightarrow{P} a / b \quad (b \neq 0)$

大数定律的客观背景

大量的随机现象中频率的稳定性

大量抛掷硬币正面出现频率

生产过程中的废品率



字母使用频率

定理3. （切比雪夫大数定律）设随机变量序列$\{X_i, i \geq 1\}$相互独立，它们的数学期望和方差都存在，且方差有共同的上界，即存在常数$M>0$，使得$D(X_i) \leq M$，$i=1,2,,\ldots$，则对任意给定的常数$\varepsilon>0$，有

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

Chebyshev
(1821-1894)

证明：由独立性以及数学期望和方差的性质可得

$$E(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n} EX_i, \quad D(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} DX_i$$

一. 大数定律

$$E(\frac{1}{n}\sum_{i=1}^{n}X_i)=\frac{1}{n}\sum_{i=1}^{n}EX_i,\quad D(\frac{1}{n}\sum_{i=1}^{n}X_i)=\frac{1}{n^2}\sum_{i=1}^{n}DX_i$$

且有 $D(\frac{1}{n}\sum_{i=1}^{n}X_i)\le\frac{1}{n^2}\sum_{i=1}^{n}M=\frac{M}{n}$

根据切比雪夫不等式 $P\{|\frac{1}{n}\sum_{i=1}^{n}X_i-\frac{1}{n}\sum_{i=1}^{n}EX_i|<\varepsilon\}\ge 1-D(\frac{1}{n}\sum_{i=1}^{n}X_i)\Big/\varepsilon^2\ge 1-\frac{M}{n\varepsilon^2}$

又由概率的性质得 $P\{|\frac{1}{n}\sum_{i=1}^{n}X_i-\frac{1}{n}\sum_{i=1}^{n}EX_i|<\varepsilon\}\le 1$

所以 $1-\frac{M}{n\varepsilon^2}\le P\{|\frac{1}{n}\sum_{i=1}^{n}X_i-\frac{1}{n}\sum_{i=1}^{n}EX_i|<\varepsilon\}\le 1$

一. 大数定律

$$1-\frac{M}{n\varepsilon^2}\le P\{|\frac{1}{n}\sum_{i=1}^{n}X_i-\frac{1}{n}\sum_{i=1}^{n}EX_i|<\varepsilon\}\le 1$$

由夹逼定理可得 $\lim\limits_{n\to\infty}P\{|\frac{1}{n}\sum_{i=1}^{n}X_i-\frac{1}{n}\sum_{i=1}^{n}EX_i|<\varepsilon\}=1$

注：（1）相互独立的条件可以改为两两不相关。

（2）按照依概率收敛的定义，此定理的结论可以表述为

$$\frac{1}{n}\sum_{i=1}^{n}X_i-\frac{1}{n}\sum_{i=1}^{n}EX_i\xrightarrow{P}0$$

（3）由 $D\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)\le\frac{M}{n}$ 知，当n越来越大时，随机变量$X_1, X_2, \dots, X_n$的平均值 $\frac{1}{n}\sum_{i=1}^{n}X_i$ 的方差越来越趋于0，即说明随着n的增大，大量随机变量的算术平均值的分散程度越来越小，呈现出一定的稳定性，且稳定于 $\frac{1}{n}\sum_{i=1}^{n}EX_i$ ，切比雪夫大数定律以数学的形式表达了这种算术平均值的稳定性。

推论2.（切比雪夫大数定律的特殊情况）

设 $X_1, X_2, \ldots$是相互独立的随机变量序列，它们具有相同的数学期望和方差，即$E(X_i)=\mu$， $D(X_i)=\sigma^2>0$， $i=1,2,\ldots$ 则对任意的$\varepsilon>0$，有

$$\lim_{n\to\infty} P\{|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu| < \varepsilon\} = 1 \quad 即 \quad \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu$$

定理3. 伯努利(Bernoulli)大数定律

设S_n是n重伯努利试验中事件A发生的次数，p是一次试验中事件A发生的概率$(0<p<1)$，则对任给的常数$\varepsilon>0$，有

$$\lim_{n\to\infty} P\{|\frac{S_n}{n}-p|<\varepsilon\}=1 \text{ 即 } \frac{S_n}{n}\xrightarrow{P} p$$

证明：设 $X_i=\begin{cases}1, & \text{第}i\text{次试验}A\text{发生}\\ 0, & \text{第}i\text{次试验}A\text{不发生}\end{cases}, i=1,\dots,n$

则有 $X_1, X_2, \dots, X_n$互相独立同分布，　易知 $S_n=\sum_{i=1}^{n} X_i$

且 $EX_i=p, DX_i=p(1-p), \quad i=1,2\dots n$

由推论2，有 $\lim\limits_{n\to\infty} P(|\frac{S_n}{n} - p| < \varepsilon) = \lim\limits_{n\to\infty} P(|\frac{1}{n}\sum_{i=1}^{n} X_i - p| < \varepsilon) = 1$

按照依概率收敛的定义，伯努利大数定律的结论可以表述为 $\frac{S_n}{n} \xrightarrow{P} p$

即在独立重复试验中，事件A发生的频率 $\frac{S_n}{n}$，当$n\to\infty$时依概率收敛于事件A在一次试验中发生的概率。伯努利大数定律给出了频率稳定性的严格的数学表述。

伯努利大数定律的结果表明，当重复试验次数n充分大时，事件A发生的频率S_n/n与事件A的概率p有较大偏差的概率很小。

因此在实际中，只要试验次数足够多，就可以用事件的频率估计事件的概率。

因此，伯努利大数定律提供了通过试验来确定事件概率的方法的理论依据。同时伯努利大数定律也解释了概率存在的客观意义，为什么“大数次”重复试验下，事件的概率是存在的。

正是因为频率的这种稳定性，我们才意识到概率的存在，才有了概率这门学课。

定理4 .辛钦（Khintchine）大数定律

设随机变量序列X_1, X_2, …独立同分布，且$E(X_i)=\mu$，$i=1, 2,\ldots$, 则对任意给定的常数$\varepsilon>0$，有

$$\lim_{n\to\infty} P\{|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu| < \varepsilon\} = 1 \text{ 即 } \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu$$

辛钦大数定律表明当n充分大时，n个独立同分布的随机变量的算术平均值依概率收敛于它们共同的数学期望μ。因此，辛钦大数定律给出了平均值稳定性的科学描述；也为估计随机变量的期望值提供了实际可行的途径和理论依据。实际应用中，可以用多次观测的算术平均值近似期望值。

比如，为了精确称量某一物体的质量μ，我们可以在相同的条件下重复的称重n次，结果记为x_1，x_2，…，x_n，由于误差的存在，它们一般是不同的，可以看作n个相互独立同分布的随机变量X_1，X_2，…X_n的一次观测值。X_1，X_2，…X_n独立且同分布，它们共同的数学期望可以看作物体的真实质量μ，由辛钦大数定律知，

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu$$

意味着当n充分大时，$\frac{1}{n}\sum_{i=1}^{n} X_i$ 稳定于μ，因此可以用 $\frac{1}{n}\sum_{i=1}^{n} X_i$ 的观测值 $\frac{1}{n}\sum_{i=1}^{n} x_i$ 近似物体的真实质量μ。

例3.设$X_1, X_2, \ldots$ 独立同分布，且X_i的k阶矩$m_k=E(X_i^{\,k})$存在，证明

$$\frac{1}{n}\sum_{i=1}^{n} X_i^k \xrightarrow{P} m_k$$

证明： 令 $Y_i = X_i^k, i = 1, 2, \ldots$ 则有$Y_1, Y_2, \ldots$ 独立同分布

且 $EY_i = EX_i^k = m_k, i = 1, 2, \ldots$

所以由辛钦大数定律 $\frac{1}{n}\sum_{i=1}^{n} Y_i \xrightarrow{P} E(Y_1) = m_k$

也即 $\frac{1}{n}\sum_{i=1}^{n} X_i^k \xrightarrow{P} m_k$

例4. 已知随机变量序列$X_1, X_2, \ldots, X_n, \ldots$独立，都服从正态分布$N(\mu,\sigma^2)$，令

$$\bar{X}_n=\frac{1}{n}\sum_{i=1}^{n}X_i,\quad S_n^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$$

证明： $\bar{X}_n^2 \xrightarrow{P} \mu^2,\quad S_n^2 \xrightarrow{P} \sigma^2$

证明：由辛钦大数定律可得 $\bar{X}_n=\frac{1}{n}\sum_{i=1}^{n}X_i \xrightarrow{P} EX_i=\mu$

由定理2 可得 $\bar{X}_n^2 \xrightarrow{P} \mu^2$

又因为 $\sum_{i=1}^{n}(X_i-\bar{X}_n)^2=\sum_{i=1}^{n}(X_i^2-2X_i\bar{X}_n+\bar{X}_n^2)=\sum_{i=1}^{n}X_i^2-2\sum_{i=1}^{n}X_i\bar{X}_n+\sum_{i=1}^{n}\bar{X}_n^2$

$$=\sum_{i=1}^{n}X_i^2-2\bar{X}_n\sum_{i=1}^{n}X_i+\sum_{i=1}^{n}\bar{X}_n^2=\sum_{i=1}^{n}X_i^2-2n\bar{X}_n^2+n\bar{X}_n^2=\sum_{i=1}^{n}X_i^2-n\bar{X}_n^2$$

所以 $S_n^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2=\frac{1}{n-1}(\sum_{i=1}^{n}X_i^2-n\bar{X}_n^2)=\frac{n}{n-1}(\frac{1}{n}\sum_{i=1}^{n}X_i^2-\bar{X}_n^2)$

由例3知 $\frac{1}{n}\sum_{i=1}^{n}X_i^2\xrightarrow{P}EX_i^2=DX_i+(EX_i)^2=\sigma^2+\mu^2$

又因为 $\lim\limits_{n\to\infty}\dfrac{n}{n-1}=1$ $\bar{X}_n^2 \xrightarrow{P} \mu^2$

所以由定理2及其推论可得

$$S_n^2=\frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2-\bar{X}_n^2\right)\xrightarrow{P}1\times(\sigma^2+\mu^2-\mu^2)=\sigma^2$$

例5. 设随机变量序列$X_1, X_2, \dots, X_n, \dots$独立同分布，都服从均匀分布$U(0, 1)$。问当$n\to\infty$时，$\sqrt[n]{X_1X_2\cdots X_n}$ 依概率收敛吗？若收敛，请给出收敛的极限值，否则请说明理由。

解：令 $Y_n=\sqrt[n]{X_1X_2\cdots X_n}$ 则有 $\ln Y_n=\frac{1}{n}\sum_{i=1}^{n}\ln X_i$ 令 $Z_n=\ln Y_n$

由于$X_1, X_2, \dots, X_n$独立同分布，所以$\ln X_1, \ln X_2, \dots, \ln X_n$独立同分布。

又因为$X_1\sim U(0, 1)$，易知其密度函数为 $f(x)=\begin{cases}1, & 0<x<1\\ 0, & \text{其它}\end{cases}$

所以 $E(\ln X_1)=\int_{-\infty}^{+\infty}(\ln x)f(x)dx=\int_0^1 \ln x\,dx=-1$

故由辛钦大数定律得 $$\frac{1}{n}\sum_{i=1}^{n}\ln X_i \xrightarrow{P} -1$$

即 $$Z_n \xrightarrow{P} -1$$

取 $h(x)=e^x$ 由定理2 可得 $$Y_n = e^{Z_n} \xrightarrow{P} e^{-1}$$

即 $$\sqrt[n]{X_1X_2\cdots X_n} \xrightarrow{P} e^{-1}$$

二. 中心极限定理

在客观实际中有许多随机变量，它们是由大量的相互独立的随机因素的综合影响所形成的。而其中每一因素在总的影响中所起的作用都是微小的。

这种随机变量往往近似地服从正态分布。

该结论得益于高斯对测量误差分布的研究。

Gauss
(1777-1855)

例如：考虑炮弹的射击误差．设靶心为坐标原点，弹着点的坐标为(X,Y)，X，Y分别表示弹着点与靶心的横向和纵向误差。我们来看造成误差的原因是什么?

炮身在每次射击后，因震动而造成微小的偏差;
每发炮弹外形上的细小差别引起空气阻力不同，由此出现的误差;
每发炮弹内炸药的数量和质量上的微小差异而引起的误差;
炮弹在前进时遇到的空气气流的微小扰动而造成的误差；等等

可以看到每种因素对误差的影响都是微小的，有的为正，有的为负，有的有时候大，有的有时候小，而且时有时无，都是随机的。

二. 中心极限定理

这些因素的综合影响最后使得射击出现了误差。设这个误差为随机变量 Y_n，可以将 Y_n 看成很多微小的相互独立的随机因素 $X_1, X_2, \ldots, X_n$ 的和，即

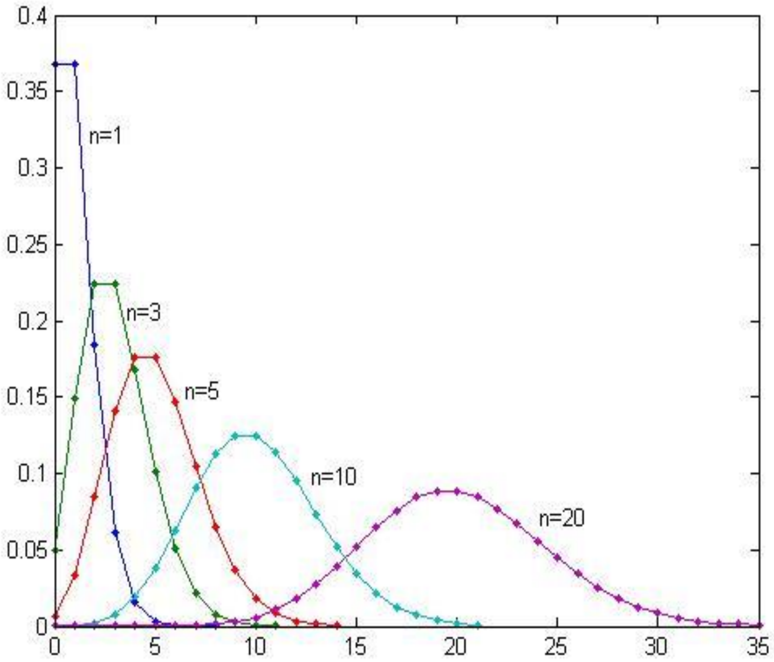
$$Y_n=\sum_{i=1}^{n}X_i$$

中心极限定理主要研究当 $n\to\infty$ 时，Y_n 的分布是什么?

当每个 X_i 的分布已知时，可以利用随机变量和的分布的求法，计算 Y_n 的分布，但是这个过程肯定是繁琐的，而且在大多数情况下，可能无法确定 X_i 的分布，从而也就无法确定 Y_n 的分布。这就迫使我们去寻求近似分布。若记 Y_n 的分布函数为 $F_n(x)$，此时也就是需要求当 $n\to\infty$ 时，$F_n(x)$ 的极限函数 $F(x)$。

例6. 设随机变量$X_1, X_2, \ldots\ldots, X_n$独立同分布，都服从参数为$\lambda$的泊松分布，即 $X_i \sim P(\lambda)$, 则 $S_n = \sum_{i=1}^{n} X_i \sim \pi(n\lambda)$

取$\lambda=1$，当$n=1, 3, 5, 10, 20$时，S_n的分布律的图像如右图。



从图中可以看出，当n越来越大时，Y_n的分布律$p_n(k)$的图像越来越光滑，且形状越来越接近正态分布的密度曲线。

通过例6可以看出在满足一定的条件下，这种和的分布近似地服从正态分布。但是我们知道，$EY_n=DY_n=n\lambda$，也就是说随着n的增大，Y_n的分布中心会趋于∞，其方差也趋于∞，分布极不稳定，为了克服这个缺点，在中心极限定理的研究中，对Y_n进行标准化

$$Z_n=\frac{Y_n-EY_n}{\sqrt{DY_n}}=\frac{\sum_{i=1}^{n}X_i-E(\sum_{i=1}^{n}X_i)}{\sqrt{D(\sum_{i=1}^{n}X_i)}}$$

这时需要研究Z_n的极限分布是否是标准正态分布$N(0,1)$。

定理6. 林德伯格-莱维中心极限定理或者独立同分布的中心极限定理

设X_1，X_2，…是独立同分布的随机变量序列，且$E(X_i)=\mu$，$D(X_i)=\sigma^2>0$，$i=1,2,\dots$，则对任意的实数x，有

$$\lim_{n\to\infty} P\left\{\frac{\sum_{i=1}^{n} X_i - E(\sum_{i=1}^{n} X_i)}{\sqrt{D(\sum_{i=1}^{n} X_i)}} \le x\right\} = \lim_{n\to\infty} P\left\{\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \le x\right\} = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = \Phi(x)$$

当n充分大时 $\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \overset{近似地}{\sim} N(0,1)$ $\frac{\frac{1}{n}\sum_{i=1}^{n} X_i - \mu}{\sigma/\sqrt{n}} \overset{近似地}{\sim} N(0,1)$

此时，对任意的实数$a, b, a<b$，有以下近似公式

$$P\{a<\sum_{i=1}^{n}X_i<b\}=P\{\frac{a-n\mu}{\sigma\sqrt{n}}<\frac{\sum_{i=1}^{n}X_i-n\mu}{\sigma\sqrt{n}}<\frac{b-n\mu}{\sigma\sqrt{n}}\}\approx\Phi(\frac{b-n\mu}{\sigma\sqrt{n}})-\Phi(\frac{a-n\mu}{\sigma\sqrt{n}})$$

$$P\{\sum_{i=1}^{n}X_i<b\}=P\{\frac{\sum_{i=1}^{n}X_i-n\mu}{\sigma\sqrt{n}}<\frac{b-n\mu}{\sigma\sqrt{n}}\}\approx\Phi(\frac{b-n\mu}{\sigma\sqrt{n}})$$

$$P\{\sum_{i=1}^{n}X_i>a\}=P\{\frac{\sum_{i=1}^{n}X_i-n\mu}{\sigma\sqrt{n}}>\frac{a-n\mu}{\sigma\sqrt{n}}\}\approx1-\Phi(\frac{a-n\mu}{\sigma\sqrt{n}})$$

例7.当辐射的强度超过每小时0.5毫伦琴（mr）时，辐射会对人的健康造成伤害。设一台彩电工作时的平均辐射强度是0.036（mr/h），方差是0.0081。则家庭中一台彩电的辐射一般不会对人造成健康伤害。但是彩电销售店同时有多台彩电工作时，辐射可能对人造成健康伤害.现在有16台彩电同时工作，问这16台彩电的辐射量可以对人造成健康伤害的概率。

解：设X_i为第i台彩电的辐射量，则$EX_i=0.036$，$DX_i=0.0081$，$i=1,2,\ldots,16$。

则$S=X_1+X_2+\ldots+X_{16}$是16台彩电的总辐射量。$ES=16\times0.036$，$DS=16\times0.0081$

由独立同分布的中心极限定理 $\dfrac{S-0.576}{\sqrt{0.1296}} \overset{\text{近似}}{\sim} N(0,1)$

$$\frac{S-0.576}{\sqrt{0.1296}} \overset{近似}{\sim} N(0,1)$$

所以

$$P\{S>0.5\}=1-P\{S\le 0.5\}=1-P\{\frac{S-0.576}{\sqrt{0.1296}}\le\frac{0.5-0.576}{\sqrt{0.1296}}\}$$

$$\approx 1-\Phi(\frac{0.5-0.576}{\sqrt{0.1296}})=1-\Phi(-0.21)$$

$$=\Phi(0.21)=0.5832$$

例8. 某快餐店午餐有三种套餐出售，其价格分别为20元、25元、30元，售出概率分别为0.6、0.2、0.2，若已知某日售出了100件套餐，求总收入超过2400元的概率。

解：设X_i表示售出的第i件套餐的价格，$i=1,2,\dots,100$. 易知X_i的分布律为

X_i	20	25	30
P	0.6	0.2	0.2

计算得 $E(X_i)=20\times0.6+25\times0.2+30\times0.2=23$

$E(X_i^2)=20^2\times0.6+25^2\times0.2+30^2\times0.2=545$ $\quad DX_i=EX_i^2-(EX_i)^2=16$

则$\sum_{i=1}^{100} X_i$ 表示售出100件套餐的总收入。且 $E(\sum_{i=1}^{100} X_i)=2300, D(\sum_{i=1}^{100} X_i)=1600$

由独立同分布的中心极限定理 $\dfrac{\sum_{i=1}^{100} X_i - 2300}{\sqrt{1600}} \overset{近似}{\sim} N(0,1)$

故有 $P\{\sum_{i=1}^{100} X_i > 2400\} = P\{\dfrac{\sum_{i=1}^{100} X_i - 2300}{\sqrt{1600}} > \dfrac{2400-2300}{\sqrt{1600}}\}$

$\approx 1-\Phi(2.5) = 1-0.9938 = 0.0062$

例9. 已知随机变量$X_1,\dots, X_{100}$独立同分布且均服从$U(0, 1)$，令$Y=X_1X_2\dots X_{100}$。求$Y<e^{-80}$的概率的近似值。

解：易知X_i的概率密度函数为 $f(x)=\begin{cases}1, & 0<x<1\\ 0, & \text{其他}\end{cases}$

令$Y_i=\ln X_i$，$i=1,2,\dots,100$，则Y_i，$i=1,2,\dots,100$独立同分布，且

$$EY_i=E(\ln X_i)=\int_{-\infty}^{+\infty}\ln x f(x)dx=\int_0^1 \ln x dx=-1$$

$$EY_i^2=E(\ln X_i)^2=\int_{-\infty}^{+\infty}(\ln x)^2 f(x)dx=\int_0^1(\ln x)^2 dx=2 \quad DY_i=EY_i^2-(EY_i)^2=1$$

二. 中心极限定理

则有 $E(\sum_{i=1}^{100} Y_i) = -100$，$D(\sum_{i=1}^{100} Y_i) = 100$ 由**中心极限定理**得：$\dfrac{\sum_{i=1}^{100} Y_i - (-100)}{\sqrt{100}} \overset{近似}{\sim} N(0,1)$

所以 $P\{Y < e^{-80}\} = P\{X_1 X_2 \cdots X_{100} < e^{-80}\} = P\{\ln(X_1 X_2 \cdots X_{100}) < \ln e^{-80}\}$

$$= P\{\sum_{i=1}^{100} \ln X_i < -80\} = P\{\sum_{i=1}^{100} Y_i < -80\}$$

$$= P\left(\frac{\sum_{i=1}^{100} Y_i + 100}{\sqrt{100}} < \frac{-80+100}{\sqrt{100}}\right) \approx \Phi(2) = 0.9772$$

定理7. 棣莫佛(De Moivre) - 拉普拉斯(Laplace)中心极限定理

在n重伯努利试验中，事件A在每次试验中发生的概率为$p(0<p<1)$，设Y_n表示n次试验中事件A发生的次数，则对于任意的x，有

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

定理表明，当n较大，$0<p<1$时，$\frac{Y_n-np}{\sqrt{np(1-p)}}$ 近似于标准正态分布$N(0, 1)$。

于是，当n较大，$0<p<1$时，Y_n近似服从$N(np, np(1-p))$。

即，若$Y\sim b(n, p)$，则：$Y\sim N(np, np(1-p))$ 近似

二. 中心极限定理

当Y服从二项分布$b(n,\ p)$时，对任意的实数$a,\ b$，且$a<b$，有如下的近似计算公式：

$$P\{a<Y<b\}=P\{\frac{a-np}{\sqrt{np(1-p)}}<\frac{Y-np}{\sqrt{np(1-p)}}<\frac{b-np}{\sqrt{np(1-p)}}\}\approx\Phi(\frac{b-np}{\sqrt{np(1-p)}})-\Phi(\frac{a-np}{\sqrt{np(1-p)}})$$

$$P\{Y<b\}=P\{\frac{Y-np}{\sqrt{np(1-p)}}<\frac{b-np}{\sqrt{np(1-p)}}\}\approx\Phi(\frac{b-np}{\sqrt{np(1-p)}})$$

$$P\{Y>a\}=P\{\frac{Y-np}{\sqrt{np(1-p)}}>\frac{a-np}{\sqrt{np(1-p)}}\}\approx 1-\Phi(\frac{a-np}{\sqrt{np(1-p)}})$$

例10.某保险公司有10000个同龄又同阶层的人参加某种医疗保险。已知该类人在一年内生该种病的概率为0.006。每个参加保险的人在年初付12元保险费，而在生病时可从公司领得1000元。问在此项业务活动中，（1）保险公司赔钱的概率是多少？（2）保险公司获得利润（暂不计管理费）不少于40000的概率是多少？

解：令X表示10000个参加该种医疗保险的人中，一年内生该病的人数，易知 $X \sim b(10000, 0.006)$

$$EX=10000\times0.006=60 \qquad DX=10000\times0.006\times0.994=59.64$$

由棣莫佛-拉普拉斯定理，有 $\dfrac{X-60}{\sqrt{59.64}} \overset{\text{近似}}{\sim} N(0,1)$

$$\frac{X-60}{\sqrt{59.64}} \overset{\text{近似}}{\sim} N(0,1)$$

（1）$P\{10000\times 12 - X\times 1000 \le 0\} = P\{X \ge 120\} \approx 1 - F(120)$

$$\approx 1-\Phi\left(\frac{120-60}{\sqrt{59.64}}\right) = 1-\Phi(7.769) \approx 0$$

（2）$P\{10000\times 12 - X\times 1000 \ge 40000\} = P\{X \le 80\}$

$$\approx \Phi\left(\frac{80-60}{\sqrt{59.64}}\right) = \Phi(2.59) = 0.9952$$

例11.(供电问题)某车间有200台车床，在生产期间由于需要检修、调换刀具、变换位置及调换工件等常需停车。设开工率为0.6，并设每台车床的工作是独立的，且在开工时需电力1千瓦。问应供应多少千瓦电力就能以99.9%的概率保证该车间不会因供电不足而影响生产?

解：设需供应x千瓦电力。用X 表示在某时刻工作着的车床数。

求满足$P\{X \le x\} \ge 0.999$的最小的x 。 易知 $X \sim b(200,0.6)$

$EX=200\times0.6=120$ $DX=200\times0.6\times0.4=48$

由棣莫佛 - 拉普拉斯定理，有 $\dfrac{X-120}{\sqrt{48}} \overset{\text{近似}}{\sim} N(0,1)$

$$\frac{X-120}{\sqrt{48}} \overset{\text{近似}}{\sim} N(0,1)$$ 求满足$P\{X \le x\} \ge 0.999$的最小的x。

$$P\{X \le x\} \approx \Phi\left(\frac{x-120}{\sqrt{48}}\right) \ge 0.999$$

查正态分布函数表得 $\Phi(3.1) = 0.999$

故 $\frac{x-120}{\sqrt{48}} \ge 3.1$ 从中解得 $x \ge 141.5$

定理7给出了独立同分布的情况下，随机变量和的极限分布的问题。实际问题中，诸X_i的独立性是容易满足的，但是同分布性就很难说了。因此需要在独立不同分布下研究和的极限分布问题。

定理8. 李雅普诺夫(Lyapunov)中心极限定理

设$X_1, X_2, \ldots X_n, \ldots$是相互独立的随机变量序列，它们的数学期望和方差都是存在的，且其数学期望$E(X_k)=\mu_k$，方差$D(X_k)=\sigma_k^2>0$存在，$k=1,2,\ldots$，记

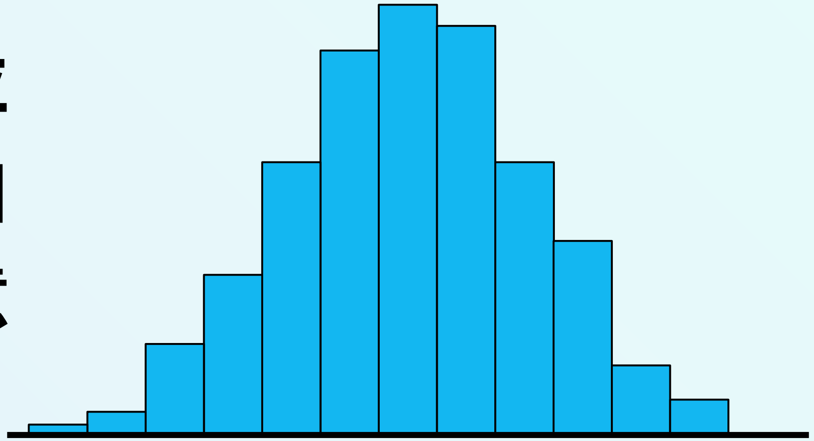
$B_n=\sqrt{\sum\limits_{k=1}^{n}\sigma_k^2}$ 若存在正数δ，使得 $\lim\limits_{n\to\infty}\dfrac{1}{B_n^{2+\delta}}\sum\limits_{k=1}^{n}E\{|X_k-\mu_k|^{2+\delta}\}=0$

则对任意x，有 $\lim\limits_{n\to\infty}P\{\dfrac{\sum\limits_{k=1}^{n}X_k-\sum\limits_{k=1}^{n}\mu_k}{B_n}\le x\}=\int_{-\infty}^{x}\dfrac{1}{\sqrt{2\pi}}e^{-t^2/2}dt=\Phi(x)$

二. 中心极限定理

这就是说，无论各个随机变量X_k服从什么分布，只要满足定理的条件，那么当n充分大时，他们的和$\sum_{k=1}^{n} X_k$近似地服从正态分布。

这就是正态分布在概率论中占有重要地位的一个基本原因，也有助于解释为什么很多自然群体的经验频率呈现钟形曲线这一值得注意的事实。



作业： 2,4,6,9,10

第16讲

谢谢聆听