第三章　小结

多维随机变量

二维随机变量

离散型-概率分布
边际分布
概率计算
条件分布

分布函数
概率计算

连续型-概率密度
边际密度
概率计算
条件密度

独立性

数字特征：协方差、相关系数等

随机变量函数的概率分布

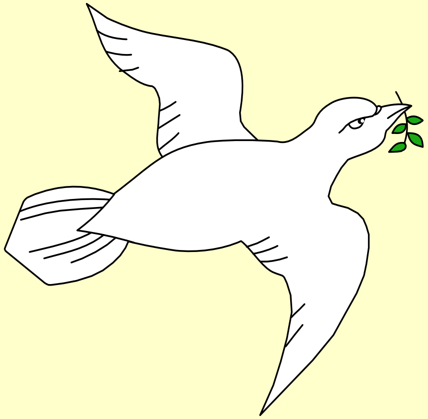
离散型-列表归纳
连续型-套用公式
由F求p

两个随机向量和、商、极值的分布

第四章

随机变量序列的极限分布

- **泊松定理与中心极限定理**
- 概率收敛与大数定律



一、泊松定理与中心极限定理

用泊松分布近似二项分布

定理 1.设 $\xi \sim \boldsymbol{B(n, p_n)}$，当 $n \to \infty$ 时，ξ 近似地服从 $P(\lambda)$，

即 $np_n \to \lambda$，

$$C_n^k p_n^{\ k} q_n^{\ n-k} \to \frac{\lambda^k}{k!} e^{-\lambda},$$

使用条件：n充分大，p_n很小(<0.1)，而np_n适当（可查表）

复习

用EXCEL计算的结果

k	B(10,0.1)	B(100,0.01)	B(1000,0.001)	B(10000,0.0001)	P(1)
0	0.348678	0.36603234	0.367695425	0.367861046	0.36787944
1	0.38742	0.36972964	0.368063488	0.367897836	0.36787944
2	0.19371	0.18486482	0.184031744	0.183948918	0.18393972
3	0.057396	0.06099917	0.061282509	0.061310174	0.06131324
4	0.01116	0.01494171	0.015289955	0.015324478	0.01532831
5	0.001488	0.00289779	0.003048808	0.003063976	0.00306566
6	0.000138	0.00046345	0.0005061	0.000510458	0.00051094
7	8.75E-06	6.2863E-05	7.19381E-05	7.28862E-05	7.2992E-05
8	3.65E-07	7.3817E-06	8.93826E-06	9.1053E-06	9.124E-06
9	9E-09	7.622E-07	9.86181E-07	1.01099E-06	1.0138E-06
10	1E-10	7.006E-08	9.78284E-08	1.01018E-07	1.0138E-07
11		5.7901E-09	8.81337E-09	9.17522E-09	9.2162E-09
12		4.3377E-10	7.27095E-10	7.63837E-10	7.6801E-10

例 1.某人进行射击，每次击中的概率为 0.02，独立射击 400 次，试求(1)至少击中两次的概率;(2) 击中次数不到两次的概率。

解：设击中的次数ξ，则$\xi \sim B(400,0.02)$

$$p = 0.02 < 0.1 \qquad \lambda = np = 400 \times 0.02 = 8$$

所以可用泊松分布近似二项分布，

(1) $$P(\xi \geq 2) = 1 - P(\xi = 0) - P(\xi = 1)$$

$$= 1 - (0.98)^{400} - C_{400}^{1}(0.98)^{399}(0.02)$$

$$\approx 1 - \frac{8^0}{0!}e^{-8} - \frac{8^1}{1!}e^{-8} = 0.997$$

即：至少击中两次的概率为0.997。

(2) $$P(\xi < 2) \approx 0.003.$$

即：击中次数不到两次的概率为0.003。

中心极限定理

正态分布为何如此广泛，从而在概率论中占如此重要的地位？

中心极限定理

概率论中，有关论证随机变量累加和的极限分布是正态分布的那些定理。

设独立同分布的随机变量序列$\xi_1,\xi_2,\cdots,\xi_n,\cdots$，期望和方差都存在，即$E\xi_i=\mu$，$D\xi_i=\sigma^2$，$i=1,2,\cdots,n,\cdots$，

令$\eta_n=\sum_{i=1}^{n}\xi_i$，

则$E\eta_n=\sum_{i=1}^{n}E\xi_i=n\mu$，　$D\eta_n=\sum_{i=1}^{n}D\xi_i=n\sigma^2\hat{=}s_n^2$

标准化$\varsigma_n\hat{=}\dfrac{\eta_n-E\eta_n}{\sqrt{D\eta_n}}=\dfrac{1}{s_n}\sum_{i=1}^{n}(\xi_i-\mu)$

则　$E\varsigma_n=0,D\varsigma_n=1$

林德伯格—列维定理（独立同分布中心极限定理）

设独立随机变量 $\xi_1, \xi_2, \cdots, \xi_n, \cdots$ ，服从相同分布，且

$E\xi_i = \mu$， $D\xi_i = \sigma^2$， $i = 1, 2, \cdots, n, \cdots,$

则当 $n \to \infty$ 时，有

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

其中 x 是任意实数。

意义：

假设被研究的随机变量可以表示为大量独立随机变量的和，其中每一个别随机变量对于总和只起微小的作用，则可以认为这个随机变量实际上是服从正态分布的。

当n充分大时，$\dfrac{\sum_{i=1}^{n}\xi_i - n\mu}{\sqrt{n}\sigma}$近似服从$N(0,1)$

$$\sum_{i=1}^{n}\xi_i \sim N(n\mu, n\sigma^2)$$

推论： 如果随机变量$\xi_1,\xi_2,\cdots,\xi_n,\cdots$ 独立，服从相同分布，且$E\xi_i=\mu$，$D\xi_i=\sigma^2$，$i=1,2,\cdots,n,\cdots$，

则n充分大时，有下面的近似公式：

$$P\left(x_1<\frac{\sum_{i=1}^{n}\xi_i-n\mu}{\sqrt{n}\sigma}\le x_2\right)\approx\Phi(x_2)-\Phi(x_1)$$

其中，x_1,x_2是任何实数。

例 2 计算机进行加法计算时，把每个加数取为最接近于它的整数来计算。设所有的取整误差是相互独立的随机变量，并且都在区间**[-0.5,0.5]**上服从均匀分布，求 **300** 个数相加时误差总和的绝对值小于 **10** 的概率。多少个数相加在一起能使误差总和小于**10**的概率为**0.9**?

解： 设 ξ_i 表示第 i 个加数的取整误差，则 ξ_i 在区间**[-0.5,0.5]**上服从均匀分布，并且有

$$E\xi_i = \frac{-0.5+0.5}{2} = 0\text{，即 } \mu = 0$$

$$D\xi_i = \frac{[0.5-(-0.5)]^2}{12} = \frac{1}{12}, \quad i=1,2,\cdots,n,\cdots\text{，即}$$

$\sigma^2 = \dfrac{1}{12}$，由列维定理的推论：

$$P\left(\left|\sum_{i=1}^{n}\xi_i\right|<10\right)=P\left(\frac{\left|\sum_{i=1}^{n}\xi_i-0\right|}{\sqrt{300\times\frac{1}{12}}}<\frac{10-0}{\sqrt{300\times\frac{1}{12}}}\right)$$

$$=P\left(\frac{\left|\sum_{i=1}^{n}\xi_i-0\right|}{\sqrt{300\times\frac{1}{12}}}<2\right)$$

$$=\Phi(2)-\Phi(-2)=2\Phi(2)-1$$

$$=2\times0.9772-1=0.9544$$

答： 300个数相加时误差总和的绝对值小于10的概率为0.9544。

设有n个加数，则 $P\{|\sum_{i=1}^{n}\xi_i|<10\}\geq 0.9$

$$P\{|\sum_{i=1}^{n}\xi_i|<10\}=2\Phi\left(\frac{10}{\sqrt{\frac{n}{12}}}\right)-1\geq 0.9$$

$$\Phi\left(\frac{10}{\sqrt{\frac{n}{12}}}\right)\geq 0.95, \qquad \frac{10}{\sqrt{\frac{n}{12}}}\geq 1.645$$

$$n\leq 443.5$$

所以$n=443$即可。

德莫威尔-拉普拉斯定理：设在独立试验序列中，事件 A 在各次试验中发生的概率为 $p(0 < p < 1)$，随机变量 η_n 表示事件 A 在 n 次试验中发生的次数，则有

$$\lim_{n\to\infty} P\left(\frac{\eta_n - np}{\sqrt{npq}} \le x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$$

其中 Z 是任何实数，$p + q = 1$。

（简称 **D-L** 定理）

D-L定理的应用

（1）设$\xi \sim B(n, p)$,当n充分大时，可认为$\xi \sim N(np, np(1-p))$

这时,$P\{a \le \xi \le b\} \approx \Phi(\frac{b-np}{\sqrt{np(1-p)}}) - \Phi(\frac{a-np}{\sqrt{np(1-p)}})$

说明：当 p值接近 0或1时，用 ***Poisson*** （泊松）分布近似较精　确。

例3. 某厂有 400台同类机器，每台机器　发生故障的概率都是 0.02，假设各台机器工作是　相互独立的，试分别用　二项分布、近似的泊松分布　和近似的正态分布计算　最多有 2 台机器发生故障的概率　。

解：设发生故障的机器 台数ξ，则$\xi \sim B(n, p), n = 400, p = 0.02$

(1)二项分布 (精确）

$$P\{0 \le \xi \le 2\} = P\{\xi = 0\} + P\{\xi = 1\} + P\{\xi = 2\} \approx 0.0131$$

(2)泊松分布 (近似） $\lambda = np = 400 \times 0.02 = 8$

$$P\{0 \le \xi \le 2\} \approx P\{\xi = 0\} + P\{\xi = 1\} + P\{\xi = 2\} \approx 0.0137$$

(3)正态分布 (近似） $np = 8, \sqrt{np(1-p)} = 2.8$

$$P\{0 \le \xi \le 2\} \approx \Phi\left(\frac{2-8}{2.8}\right) - \Phi\left(\frac{0-8}{2.8}\right) \approx \Phi(-2.14) - \Phi(-2.86) \approx 0.0141$$

（2）近似计算用频率估计概率是的误差

$$P\{|\frac{\mu_n}{n}-p|\le\varepsilon\}=P\{|\frac{\mu_n-np}{\sqrt{np(1-p)}}|\le\varepsilon\sqrt{\frac{n}{p(1-p)}}\}$$

$$\approx\Phi(\varepsilon\sqrt{\frac{n}{p(1-p)}})-\Phi(-\varepsilon\sqrt{\frac{n}{p(1-p)}})$$

$$=2\Phi(\varepsilon\sqrt{\frac{n}{p(1-p)}})-1$$

$$P\{|\frac{\mu_n}{n}-p|>\varepsilon\}\approx 2(1-\Phi(\varepsilon\sqrt{\frac{n}{p(1-p)}}))$$

解决三类问题

(1)已知 ε，n，p，求 $P\{|\frac{\mu_n}{n}-p|\le\varepsilon\}=2\Phi\left(\varepsilon\sqrt{\frac{n}{p(1-p)}}\right)-1$

(2)已知 $P\{|\frac{\mu_n}{n}-p|\le\varepsilon\}$，$\varepsilon$，$p$，求 n

(3)已知 $P\{|\frac{\mu_n}{n}-p|\le\varepsilon\}$，$n$，求 ε

a.已知 p　b.未知 p，利用 $p(1-p)\le\frac{1}{4}$.

例 4：设一批种子的良种率为 $\frac{1}{6}$，在其中任选 600 粒，求这 600 粒种子中良种所占的比例与 $\frac{1}{6}$ 之差的绝对值不超过 0.02 的概率。a) 用切比雪夫不等式估计；b) 用中心极限定理计算。

解：用 ξ 表示 600 粒种子中良种的粒数，则 $\xi \sim b(n, p)$，$n = 600$，$p = \frac{1}{6}$，故问题所求的概率为 $P\{|\frac{\xi}{600} - \frac{1}{6}| \le 0.02\}$。

$$E\xi = np = 600 \times \frac{1}{6} = 100, \quad D\xi = npq = 600 \times \frac{1}{6} \times (1 - \frac{1}{6}) = \frac{250}{3},$$

用切比雪夫不等式进行估计

$$P\{|\frac{\xi}{600} - \frac{1}{6}| \le 0.02\} \ge 1 - \frac{D(\frac{\xi}{600})}{0.02^2} = 1 - \frac{(\frac{1}{600})^2 \times \frac{250}{3}}{0.02^2} = 0.4213 \text{。}$$

用中心极限定理计算：

利用公式 $P\{|\frac{\mu_n}{n}-p|<\varepsilon\}\cong\beta=2\Phi(\varepsilon\sqrt{\frac{n}{p(1-p)}})-1$，

注意到 μ_n 对应于此处的 ξ， $\varepsilon=0.02$ ，故

$$P\{|\frac{\xi}{600}-\frac{1}{6}|<0.02\}=2\Phi(0.02\times\sqrt{\frac{600}{\frac{1}{6}(1-\frac{1}{6})}})-1=2\Phi(1.3145)-1$$

$$=2\times0.9057-1=0.8114$$

例 5 某工厂有 **200** 台同类型的机器，每台机器工作时需要的电功率为 **Q** 千瓦。由于工艺等原因，每台机器的实际工作时间只占全部工作时间的 **75%**，各台机器是否工作是相互独立的。

求：**(1)**任一时刻有 **144** 至 **160** 台机器正在工作的概率；

(2)需要供应多少电功率可以保证所有机器正常工作的概率不小于 **0.99**？

解：已知 $n = 200, p = 0.75, q = 0.25$

所以有 $np = 150, npq = 37.5$

(1)设 η 表示任一时刻正在工作的机器的台数，则

$$P(144 \le \eta \le 160) \approx \Phi\left(\frac{160-150}{\sqrt{37.5}}\right) - \Phi\left(\frac{144-150}{\sqrt{37.5}}\right)$$

$= \Phi(1.63) - \Phi(-0.98) = \Phi(1.63) - [1 - \Phi(0.98)]$

$= 0.9484 - [1 - 0.8365] = 0.7849$

所以，任一时刻有 **144** 至 **160** 台机器正在工作的概率为 0.7849 。

“保证所有机器能正常工作”

=“所供电功率能使所有能够工作的机器都可以工作”

(2)设任一时刻正在工作的机器的台数不超过 m ，

则按题意有 $P(0 \le \eta \le m) \ge 0.99$

由 **D-L** 定理： $\Phi(\frac{m-150}{\sqrt{37.5}}) - \Phi(\frac{-150}{\sqrt{37.5}}) \ge 0.99$

因为 $\Phi(\frac{-150}{\sqrt{37.5}}) = \Phi(-24.5) \approx 0$

则 $\Phi(\frac{m-150}{\sqrt{37.5}}) \ge 0.99$，而 $\Phi(2.33) = 0.9901$

所以 $\dfrac{m-150}{\sqrt{37.5}} \geq 2.33$，

由此得：$m \geq 164.3$

即 $m = 165$

所以，需要供应165 Q 千瓦的电功率可以保证所有机器正常工作的概率不小于 0.99 。

例 1: 某民意调查公司受委托调查电视节目收视率 p，调查公司将所有调查对象中收看此节目的频率 $\frac{\mu_n}{n}$ 作为 p 的估计值，现在要以 90%把握保证“估计值与收视率 p 之间的差异不大于 5%”，问至少要调查多少对象?

分析: 由近似估计式 $P\{\left|\frac{\mu_n}{n}-p\right|\geq\varepsilon\}\approx 2\Phi(\varepsilon\sqrt{\frac{n}{pq}})-1\triangleq\beta$

即四个参数 $(n, p, \varepsilon, \beta)$ 间的等式关系，当知道了其中 3 个参数后利用等式求出余下的另一个参数值. 如果少掉的参数 p 可以用 $pq\leq\frac{1}{4}$ 的办法予以处理。。

解：这是个知道 ε，β欲求n的问题，

设被调查对象数为n，则$P\{\left|\frac{\mu_n}{n}-p\right|\le\varepsilon\}\approx 2\Phi(\varepsilon\sqrt{\frac{n}{pq}})-1=\beta$

中$\varepsilon=0.05$，$\beta\ge 0.90$。

查正态分布表，由$\Phi(x)$的单调性以及$\Phi(1.645)=\frac{1+0.90}{2}=0.95$，

得到

$$0.05\sqrt{\frac{n}{pq}}\ge 1.645$$

从中解出 $n\ge pq(\frac{1.645}{0.05})^2=1082.41pq$。

利用$pq\le\frac{1}{4}=0.25$，故$n\ge 0.25\times 1082.41=270.6$。

二、概率收敛
与
大数定律

复习

概率收敛

问题：频率 $\frac{\mu_n}{n}$ 的极限是否为概率 p ？

概率收敛定义：设 $\{\xi_n\}$ 为一随机变量序列，ξ 为一随机变量，若对任意的 $\varepsilon>0$，成立

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

则称 $\{\xi_n\}$ 按概率收敛于 ξ，记作 $\xi_n \xrightarrow{P} \xi$。

（*）式等价于 $\lim_{n\to\infty} P\{|\xi_n-\xi|\geq\varepsilon\}=0$

注意:概率收敛这一极限概念，与我们在高等数学中所的极限不同，在定义时要兼顾随机变量的“取值”与“概率”两个特性，又要强调例外情况为小概率这一事实。

伯努利大数定律： 设 μ_n 为 n 重伯努利试验中事件 A 发生的次数， p 为每次试验中 A 发生的概率，则对任意 $\varepsilon > 0$，成立

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

即 $\frac{\mu_n}{n} \xrightarrow{P} p$ 。

意义: 频率的稳定性：当试验在不变的条件下重复进行很多次时，随机事件的<u>频率在</u>它们的<u>概率附近摆动</u>。

它是“概率论” 的理论基础.

注意: 伯努利大数定律仅指出极限，而二项分布的中心极限定理不仅指出了极限，而且指出了极限附近的情况

定理1 马尔可夫大数定律

若随机变量序列 $\xi_1, \xi_2, \cdots, \xi_n, \cdots$，其方差存在，且满足 $\frac{1}{n^2} D\left(\sum_{i=1}^{n} \xi_i\right) \to 0 \quad (n \to \infty)$，则对任意的 $\varepsilon > 0$，有

$$\lim_{n\to\infty} P\{|\frac{1}{n}\sum_{i=1}^{n}\xi_i - \frac{1}{n}\sum_{i=1}^{n}E\xi_i| < \varepsilon\} = 1。$$

定理2 切比雪夫大数定律

若 $\xi_1, \xi_2, \cdots, \xi_n, \cdots$，是两两不相关的随机 变量序列，方差存在且一致有界， 即存在常数 C，使得 $D\xi_i \le C$ $(i = 1,2,\cdots)$，则对任意的 $\varepsilon > 0$，有

$$\lim_{n\to\infty} P\{|\frac{1}{n}\sum_{i=1}^{n}\xi_i - \frac{1}{n}\sum_{i=1}^{n}E\xi_i| < \varepsilon\} = 1。$$

定理 3　辛钦大数定律

设 ξ_1，$\xi_2,\cdots$，$\xi_n,\cdots$，是满足相互独立同分　布的随机变量序列，且具有有限　的数学期望 $E\xi_i = \mu(i = 1,2,\cdots)$，则对任意的 $\varepsilon > 0$，

有 $$\lim_{n\to\infty} P\{|\frac{1}{n}\sum_{i=1}^{n}\xi_i - \mu| < \varepsilon\} = 1。$$

定理 4　泊松大数定理

如在一个独立试验序列　中，事件 A 在第 k 次试验中发生的概率为 p_k，以 μ_n 记在前 n 次试验中事件 A 发生的次数，则对任意的 $\varepsilon > 0$，有

$$\lim_{n\to\infty} P\{|\frac{\mu_n}{n} - \frac{p_1 + p_2 + \cdots + p_n}{n}| < \varepsilon\} = 1。$$

用泊松分布近似二项分布

设 $X \sim B(n,p)$，当 $n \to \infty$ 时， X 近似地服从 $P(\lambda)$，其中 $\lambda = np$．即

$$C_n^m p^m q^{n-m} \approx \frac{\lambda^m}{m!} e^{-\lambda},$$

中心极限定理

设独立随机变量 $\xi_1, \xi_2, \cdots, \xi_n, \cdots$ ，服从相同分布，且

$E\xi_i = \mu \quad D\xi_i = \sigma^2$， $i = 1,2,\cdots,n,\cdots$，

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

德莫威尔-拉普拉斯定理：设在独立试验序列中，事件 A 在各次试验中发生的概率为 $p(0<p<1)$，随机变量 η_n 表示事件 A 在 n 次试验中发生的次数，则有

$$\lim_{n\to\infty} P\left(\frac{\eta_n - np}{\sqrt{npq}} \le x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$$

其中 z 是任何实数， $p+q=1$。