基础部分

第四课 概率统计

第 5 章 多元正态补充与极限定理

§5.1 多元正态重要性质补充与应用
§5.2 极限定理的概念和内容
§5.3 大数定理应用
§5.4 中心极限定理应用

本讲提要(略，请见大纲)

§5.1 多元正态分布的重要性质补充与应用

本节补充多元正态分布的数字特征的三个特别性质．在求联合分布及其参数、随机向量的函数变换、以及独立性等问题中，应用这些性质，常常使复杂的问题大大简化，并保证正确性，应该熟练掌握．

［ 内容精讲 ］

5.1.1 多元正态分布与重要性质补充

性质 1 $r\vec{v}$ $(X_1, X_2, \cdots, X_n)$服从n维正态分布的充要条件是$X_1, X_2, \cdots, X_n$的任意的线性组合 $l_1X_1+l_2X_2+\cdots+l_nX_n$ 服从一维正态分布，此处l_j不全为 0．

性质 2（正态变量的线性变换不变性） 若$(X_1, X_2, \cdots, X_n)$服从n维正态分布，设$Y_1, Y_2, \cdots, Y_k$是X_j ($j=1,2,\cdots,n$)的线性函数，相应系数矩阵的秩为k，则 $(Y_1, Y_2, \cdots, Y_k)$也服从k维正态分布．特别地，二维正态的满秩变换，仍为二元正态的．

性质 3 设$(X_1, X_2, \cdots, X_n)$服从n维正态分布，则“$X_1, X_2, \cdots, X_n$相互独立”与“$X_1, X_2, \cdots, X_n$两两不相关”等价．

5.1.2 二元正态重要补充性质的应用

利用上述性质，可以使许多问题大大简化．

［ 典型例题 ］

例 5.1.1* 设 $(X, Y)\sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$，问$U=X+Y$与$V=X-Y$能否独立？

例 5.1.2* 设X_1, X_2 iid, $\sim N(0, \sigma^2)$，令

$$Y_1=X_1-\tfrac{1}{2}X_2,\ Y_2=\tfrac{1}{2}X_1-X_2,$$

问Y_1和Y_2同分布吗？独立吗？

解 Y_1和Y_2都是正态变量．且 $EY_1=EY_2=0$，

$$DY_1=[1+(-1/2)^2]\sigma^2=\frac{5}{4}\sigma^2,$$

$$DY_2=[(1/2)^2+(-1)^2]\sigma^2=\frac{5}{4}\sigma^2,$$

故Y_1和Y_2 同分布，$N\left(0, \frac{5}{4}\sigma^2\right)$．

$$EY_1Y_2=E(X_1-\frac{1}{2}X_2)\ (\frac{1}{2}X_1-X_2)$$
$$=\frac{1}{2}E(X_1)^2+\frac{1}{2}E(X_2)^2=\sigma^2>0=EY_1EY_2$$

由不相关的等价命题，知Y_1与Y_2不是不相关的，因此知Y_1与Y_2不独立．

例 5.1.3 设二维随机变量(X, Y)服从二维正态分布，则随机变量$\xi=X+Y$与$\eta=X-Y$不相关的充要条件为

A) EX = EY;　　B) EX^2 - (EX)^2 = EY^2 - (EY)^2;
C) EX^2 = EY^2;　D) EX^2 + (EX)^2 = EY^2+ (EY)^2;　　【B】

例5.1.4* 设二维 rv (X,Y)的 pdf 为

$$f_{(X,Y)}(x,y)=[\ \varphi_1(x,y)+\varphi_2(x,y)\ ]\ /2\ ,$$

其中$\varphi_1(x,y)$和$\varphi_2(x,y)$都是二维正态pdf，且它们对应的二维rv 的相关系数分别为 1/3 和 1/3．它们的边缘pdf 所对应的rv 的数学期望都是 0，方差都是 1．

1) 求rv X和Y的pdf　$f_1(x)$ 和$f_2(x)$，及X和Y的相关系数ρ (可以直接利用二维正态密度的性质)

2) 问 X和 Y是否独立?为什么?

解 1) 以$\varphi_i(x,y)$为pdf 的边际分布(分别记为$\varphi_{i1}(x)$和$\varphi_{i2}(y)$)都是$N(0,1)$，i =1,2，因此

$$\varphi_{11}(x)=\varphi_{21}(x),\ \ \varphi_{12}(y)=\varphi_{22}(y)$$

$$f_1(x)=\frac{1}{2}\int_{-\infty}^{\infty}\varphi_1(x,y)dy+\frac{1}{2}\int_{-\infty}^{\infty}\varphi_2(x,y)dy=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\ ,\quad f_2(y)=\frac{1}{\sqrt{2\pi}}e^{-y^2/2}\ .$$

由是 X ~ N(0, 1)，Y ~ N(0, 1)，故　$\rho=[E(XY)-EXEY]/\sqrt{DXDY}=E(XY)$.

$$E(XY)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xyf\ (x,y)dxdy:=\frac{1}{2}[r_1+r_2]=\frac{1}{2}[\frac{1}{3}-\frac{1}{3}]=0$$

2) $\varphi_1(x,y)\overset{r_1=1/3}{=\!=\!=\!=\!=}\dfrac{3}{4\pi\sqrt{2}}\ \exp\{-\dfrac{9}{16}[x^2+y^2-\dfrac{2}{3}xy]\}$

$$f(x,y)=\frac{1}{2}\varphi_1(x,y)+\frac{1}{2}\varphi_2(x,y)$$
$$=\frac{3}{8\pi\sqrt{2}}\left[e^{-\frac{9}{16}[x^2+y^2-\frac{2}{3}xy]}+e^{-\frac{9}{16}[x^2+y^2+\frac{2}{3}xy]}\right]$$

可见 $f_1(x)f_2(y)\neq f(x,y)$， 因此 X 和 Y不独立．

注意本题误区：误以为(X,Y)是二元正态．本题给出分量是正态而联合起来不是二元正态的例子，也给出不相关但是不独立的例子！本题说明：正确掌握概念何等重要！

§ 5.2　极限定理的概念和内容

本节研究 rv序列部分和的极限问题．极限定理是概率论基础研究的深入．从频率稳定

性引入大数定理，从Galton钉板的落点分布引入中心极限定理，正确了解大数定理和中心极限定理的概念.

在分析比较的基础上，掌握主要极限定理的条件和结论.

[内容精讲]

5.2.1 极限定理的概念和意义

问题的提出

定义 5.2.1 设*rv*列$\{X_n\}$期望存在，

$$\xi_n := \frac{1}{n}\sum_{j=1}^{n}(X_j - EX_j)$$

如ξ_n以概率收敛到0(记为$\xi_n \to 0$ (P)),即对任给 $\varepsilon>0$，

$$\lim_{n\to\infty} P(|\xi_n| \ge \varepsilon) = 0 \quad 或 \quad \lim_{n\to\infty} P(|\xi_n| < \varepsilon) = 1$$

则称$\{X_n\}$服从大数定律，记为$X_n \in$ LLN .

定义 5.2.2 设*rv*列$\{X_n\}$期望和方差存在，

$$\varsigma_n := \left(\sum_{j=1}^{n} X_j\right)^* = \frac{\sum_{j=1}^{n} X_j - E\sum_{j=1}^{n} X_j}{\sqrt{D\sum_{j=1}^{n} X_j}} .$$

如果它们依分布收敛到一个标准正态分布的变量，记为$\varsigma_n \xrightarrow{d} Z \sim N(0,1)$ ，

即 $$F_{\varsigma_n}(x) = P(\varsigma_n \le x) \to \Phi(x), \quad n\to\infty, \quad \forall x \in R_1 ,$$

则称$\{X_n\}$服从中心极限定理，记为$X_n \in$ CLT. □

当X_j们*iid* 时，记$EX_j=\mu$和$DX_j=\sigma^2$， ξ_n成为

$$\varsigma_n = \frac{\sum_{j=1}^{n}(X_j - \mu)}{\sqrt{n}\sigma} = \frac{1}{\sqrt{n}}\sum_{j=1}^{n} X_j^* ,$$

其中X_j^*是X_j的标准化.

对 *iid*的Bernoulli计数变量列，ξ_n和ς_n分别变为(注意$E\mu_n = np$，$D\mu_n = npq$)

$$\xi_n = \frac{\mu_n}{n} - p \quad 和 \quad \varsigma_n = \frac{\mu_n - np}{\sqrt{npq}} .$$

5.2.2 主要极限定理的条件和结论

主要极限定理的条件和结论，汇集于下表：

极限定理	条件:一般 *rv*列	条件:Ber－列
大数(切贝雪夫)	不相关，σ_n^2 <c ⇒	-
(辛钦)	*iid*时 μ存在⇔	-
中心(Levy－Lin)	*iid*时 0<σ^2 <∞ ⇔	-

一般 *rv*列		Bernoulli 列	
极限定理	结论	极限定理	结论

大数 (切一)	$\xi_n=\frac{1}{n}\sum_{j=1}^{n}(X_j-EX_j)$ $\xrightarrow{P}0$	大数 (贝努利)	$\xi_n=\frac{\mu_n}{n}-p$ $\xrightarrow{P}0$
大数 (辛钦)	$\xi_n=\frac{1}{n}\sum_{j=1}^{n}(X_j-\mu)$ $\xrightarrow{P}0$		
中心极限 (Levy—Lin)	$\varsigma_n=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}X_j{}^*$ $=\frac{1}{\sqrt{n}\sigma}\sum_{j=1}^{n}(X_j-\mu)$ $\xrightarrow{d}N(0,1)$	中心极限 (DeM — Laplace)	$\varsigma_n=\frac{\mu_n-np}{\sqrt{npq}}$ $\xrightarrow{d}N(0,1)$ $P(k_1<\mu_n\le k_2)\approx$ $\Phi(\frac{k_2-np}{\sqrt{npq}})-\Phi(\frac{k_1-np}{\sqrt{npq}})$ $P(\|\frac{\mu_n}{n}-p\|\le\varepsilon)\approx$ $\Phi\left(2\varepsilon\sqrt{\frac{n}{pq}}\right)-1$

§ 5.3 大数定理及其应用

5.3.1 大数定理

5.3.2 大数定理的应用

[典型例题]

- Чебыщев 不等式的应用

例 5.3.1（Чебыщев 不等式在二项分布计算中的应用）已知 n 重贝努利试验中参数 p=0.75，问至少应该做多少次试验，才能使试验成功的频率在 0.74 和 0.76 之间的概率不低于 0.90 ？ 【18750 】

解 求 n 使 $P(0.74<\frac{\mu_n}{n}<0.76)=P(|\frac{\mu_n}{n}-0.75|<0.01)\ge 0.90$.

$\mu_n\sim B(n,p)$，故 $E\mu_n=np$， $D\mu_n=npq$

$$P(0.74<\frac{\mu_n}{n}<0.76)\ge 1-\frac{D(\mu_n/n)}{\varepsilon^2}=1-\frac{pq}{n\varepsilon^2}$$

令 $1-\frac{pq}{n\varepsilon^2}\ge 0.90$， $n\ge\frac{pq}{\varepsilon^2(1-0.90)}=18750$

注：在二项分布中，Чебыщев 不等式

$$P(|\frac{\mu_n}{n}-p|\ge\varepsilon)\le\frac{pq}{n\varepsilon^2} \tag{5.3.2}$$

例 5.3.2 设 X_1，X_2，…，X_n iid，且 ~P(λ)，$\bar{X}$为 X_1，X_2，…，X_n的算术平均值．试用 Чебыщев不等式估计 $P(|\bar{X}-\lambda|<2\sqrt{\lambda})$ 的下界.

【$1-\frac{1}{4n}$】

例5.3.3 设随机变量 X和 Y的数学期望分别为-2 和 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 $P(\ |X+Y|\ge 6\ )\le$ ________

【1/12 】

- **在数理统计中的应用**

例 5.3.4 设 *rv* X_1，X_2，…，X_n *iid*，且 X_1 的 k 阶矩存在，值为 μ_k．试证明 $M_k(n):=\frac{1}{n}\sum_{i=1}^{n}X_i^k \xrightarrow{P} \mu_k$．

证明 $EM_k=\frac{1}{n}\sum_{i=1}^{n}EX_i^k=\mu_k$；由 X_1，X_2，…，X_n *iid* 知 X_1^k，X_2^k，…，X_n^k *iid*. 对 $\{X_i^k\}$ 利用大数定理可得关于收敛性的结论．

注 这个结论实际是第 6 讲中矩估计的一致性；

【(2003-3-1(6)[4]) 设总体 X 服从参数为 2 的指数分布，X_1，X_2，…，X_n 为来自总体 X 的简单随机样本，则当 $n\to\infty$ 时，$Y_n=\frac{1}{n}\sum_{i=1}^{n}X_i^2$ 依概率收敛于______．　】

【1/2】

§ 5.4　中心极限定理及其应用

5.4.1 中心极限定理

主要中心极限定理．

5.4.2 中心极限定理的应用

[**典型例题**]

- **用正态分布作为独立和的近似分布**

例 5.4.1 某生产线生产的产品成箱包装，每箱的重量是随机的．假设每箱平均重 50 千克，标准差为 5 千克．若用最大载重量为 5 吨的汽车承运，试利用中心极限定理说明每辆车最多可以装多少，才能保障不超载的概率大于 0.977．($\Phi(2)=0.977$，其中 $\Phi(x)$ 是标准正态分布函数.)

【$P(X\le 5000)\approx\Phi(\frac{1000-10n}{\sqrt{n}})>0.977=\Phi(2)$，$n$ =98.0199，　最多 98】

例 5.4.2 设 X_1，X_2，…，X_n 为 *iid*，算术均值 $\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$．请完成下表：

X_i 的分布	和 $\sum_{i=1}^{n}X_i$ 的分布为	n 足够大时算术均值 $\bar{X}$ 的近似分布
设为 0-1 分布		
设 P(λ)		

【…；$N(p,\ p(1-p)/n)$，　$N(\lambda,\ \lambda/n)$ 】

【+(02-4-2(5)) 设 *rv* X_1，X_2，…，X_n 相互独立，　$S_n=X_1+X_2+\cdots+X_n$，根据列维-林德伯格(Levy-Lindeberg)中心极限定理，当 n 充分大时，S_n 近似服从正态分布，只要 X_1，X_2，…，X_n

(A)有相同的数学期望；　　(B)有相同的方差；

(C)服从同一指数分布；　　(D)服从同一离散型分布．【C】

- **Galton 钉板试验的落点分布**

例 5.4.3 求 Galton 钉板试验小球落点分布．

解　建立坐标系．　　令

$$X_j = \begin{cases} 1, & \text{第 } j \text{ 次 碰钉后小球向右落下} \\ -1, & \text{反之} \end{cases}$$

于是 $\eta_n = \sum_{k=1}^{n} X_k$ 表示 n 次碰钉后小球落点离原点的距离．我们证明当 n 足够大时 η_n 有近似正态分布．

注意诸 X_j iid， $X_j \sim \begin{pmatrix} -1 & 1 \\ 1/2 & 1/2 \end{pmatrix}$，

故 $EX_j = 0$， $DX_j == EX_i^2 = 1$．由中心极限定理，

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{n} X_j^* \xrightarrow{d} Z \sim N(0,1) .$$

因此当 n 足够大时， $\eta_n = \sum_{j=1}^{n} X_j \approx \sqrt{n} Z \sim N(0,n)$ ． □

- **在二项分布中的应用**

例 5.4.4 已知 n 重贝努利试验中参数 p =0.75，问至少应该做多少次试验，才能使试验成功的频率在 0.74 和 0.76 之间的概率不低于 0.95 ？

【7203】

解 求 n 使 $P(0.74 < \frac{\mu_n}{n} < 0.76) = P(\left|\frac{\mu_n}{n} - 0.75\right| < 0.01)$

$$\approx 2\Phi\left(0.01\sqrt{\frac{n}{0.75(0.25)}}\right) - 1 \ge 0.95 \cdot$$

即求 n 使 $\Phi\left(0.01\sqrt{\frac{n}{0.1875}}\right) \ge 0.975$， $0.01 \times \sqrt{\frac{n}{0.75(0.25)}} \approx 1.96$。

至少应取 $n = 1.96^2 \times 0.1875 \approx 7203$．

例 5.4.5 设某车间有同型号车床 200 台，独立工作，开工率 0.8，开工时每台车床耗电 1kw（千瓦）．问应该至少供多少电，可以 99.9%的概率，保证该车间不因供电不足而影响生产？

解

$$\begin{aligned} P(0 \le \mu_n \le r) &\approx \Phi\left(\frac{r - 200(0.8)}{\sqrt{200(0.8)0.2}}\right) - \Phi\left(\frac{0 - 200(0.8)}{\sqrt{200(0.8)0.2}}\right) \\ &= \Phi\left(\frac{r-160}{\sqrt{32}}\right) - \Phi\left(-\frac{160}{\sqrt{32}}\right) \approx \Phi\left(\frac{r-160}{\sqrt{32}}\right) \\ &\ge 0.999 \end{aligned}$$

令等号成立，查表知 $\frac{r-160}{\sqrt{32}} = 31$，故 r = 177.5，取 r = 178.

例 5.4.6 经以往检验已确认某公司组装 PC 机的次品率为 0.04，现对该公司所组装的 PC 机 100 台逐个独立地测试，

1）求不少于 4 台次品的概率（写出精确计算的表达式）；

2）利用极限定理和 Poisson 定理给出此概率两个近似值.

【 0.5705； 0.5，0.5669】

- **近似数定点运算的误差分析**

例 5.4.7 设对十进制的x_j的小数点后第 6 位作四舍五入，得到x_j的近似数y_j，误差为 $\varepsilon_j = x_j - y_j \in (-0.5\times10^{-5},\ 0.5\times10^{-5})$. 试求n个数作舍入处理的累积误差估计.

解 I 计算方法中，累积误差估计

$$\eta = \sum_{j=1}^{n}\varepsilon_j \text{ 故 } |\eta| \le \sum_{j=1}^{n}|\varepsilon_j| \le n\times0.5\times10^{-5} .$$

当 n = 10,000 时，$|\eta| \le 0.05$.

解 II 认为舍入误差 ε_j *iid* $\varepsilon_j \sim U_{(-0.5\times10^{-5},\ 0.5\times10^{-5})}$. 故 $E\varepsilon_j = 0$ ，$\sigma^2 = D\varepsilon_j = (0.5\times10^{-5})^2/3$.

由 Lin-Levy 定理 $P(|\eta/\sqrt{n}\sigma| < k) = P(|\sum_{j=1}^{n}\varepsilon_j| < k\sqrt{n}\sigma) \approx 2\Phi(k) - 1$.

取 k = 3，则此概率近似为 0.997，故有 99.7%的把握断言

$$|\eta| = |\sum_{j=1}^{n}\varepsilon_j| \le 3\times\sqrt{n}\times0.5\times10^{-5}/\sqrt{3} .$$

当 n = 10,000 ，$|\eta| = |\sum_{j=1}^{n}\varepsilon_j| \le \sqrt{3}\times.50\times10^{-3} \approx 0.00087$