第四章 随机变量的数字特征

（注：本章笔记不提供例题）

§1 数学期望

定义 设离散型随机变量 X 的分布律为

$$P\{X=x_k\}=p_k \quad k=1,2,\cdots$$

若级数

$$\sum_{k=1}^{\infty} x_k p_k$$

绝对收敛，则称级数 $\sum_{k=1}^{\infty} x_k p_k$ 的和为随机变量 X 的数学期望，记为 $E(X)$，即

$$E(X)=\sum_{k=1}^{\infty} x_k p_k$$

设连续型随机变量 X 的概率密度为 $f(x)$，若积分

$$\int_{-\infty}^{\infty} x f(x)\,dx$$

绝对收敛，则称积分 $\int_{-\infty}^{\infty} x f(x)\,dx$ 的值为随机变量 X 的数学期望，记为 $E(X)$，即

$$E(X)=\int_{-\infty}^{\infty} x f(x)\,dx$$

数学期望简称期望，又称均值

数学期望完全由随机变量的概率分布所确定，若 X 服从某一分布，也称 $E(X)$ 是这一分布的期望

期望的几个重要性质如下：

(1) 设 C 是常数，则有 $E(C)=C$

(2) 设 X 是一个随机变量，C 是常数，则 $E(CX)=CE(X)$

(3) 设 X,Y 是两个随机变量，则有

$$E(X+Y)=E(X)+E(Y)$$

这一性质可以推广到任意有限个随机变量之和的情况

(4) 设 X,Y 相互独立，则

$$E(XY)=E(X)E(Y)$$

§2 方差

定义 设 X 是随机变量，若 $E\{[X-E(X)]^2\}$ 存在，则称它为 X 的方差，记为 $D(X)$ 或 $Var(X)$，即

$$D(X)=Var(X)=E\{[X-E(X)]^2\}$$

在应用上还引入 $\sqrt{D(X)}$ 记为 $\sigma(X)$，称为标准差或均方差

方差的几个重要性质如下

(1) 设 C 是常数，则 $D(C)=0$

(2) 设 X 是随机变量，C 是常数，则有

$$D(CX)=C^2D(X) \qquad D(X+C)=D(X)$$

(3) 设 X,Y 是两个随机变量，则有

$$D(X+Y)=D(X)+D(Y)+2E\{[X-E(X)][Y-E(Y)]\}$$

特别，若 X,Y 相互独立，则有

$$D(X+Y)=D(X)+D(Y)$$

这个性质可以推广到任意有限个相互独立的随机变量的情况

(4) $D(X)=0$ 的充要条件是 X 以概率 1 取常数 $E(X)$，即

$$P\{X=E(X)\}=1$$

§3 协方差及相关系数

定义 量 $E\{[X-E(X)][Y-E(Y)]\}$ 称为随机变量 X 与 Y 的协方差，记为 $Cov(X,Y)$，即

$$Cov(X,Y)=E\{[X-E(X)][Y-E(Y)]\}$$

而

$$\rho_{XY}=\frac{Cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

称为随机变量 X 与 Y 的相关系数

由定义知，即知

$$Cov(X,Y)=Cov(Y,X) \qquad Cov(X,X)=D(X)$$

由上述方差定义知，对于任意两个随机变量 X 和 Y，下列等式

$$D(X+Y)=D(X)+D(Y)+2Cov(X,Y)$$

将 $Cov(X,Y)$ 的定义式展开，易得

$$Cov(X,Y)=E(XY)-E(X)E(Y)$$

协方差具有下述性质：

(1) $Cov(aX,bY)=abCov(X,Y)$，a,b 是常数

(2) $Cov(X_1+X_2,Y)=Cov(X_1,Y)+Cov(X_2,Y)$

相关系数有下述性质：

(1) $|\rho_{XY}|\le 1$

(2) $|\rho_{XY}|=1$ 的充要条件是，存在常数 a,b 使

$$P\{Y=a+bX\}=1$$

$|\rho_{XY}|=1$ 较大时，表明 X、Y（就线性关系来说）联系紧密，特别当 $|\rho_{XY}|=1$ 时，X,Y 之间以概率 1 存在着线性关系。于是 ρ_{XY} 是一个可以用来表征 X,Y 之间线性关系紧密程度的量。当 $|\rho_{XY}|$ 较大时，我们通常说，X,Y 线性相关的程度较好，反之较差。当 $\rho_{XY}=0$ 时，称 X 和 Y 不相关

当 X,Y 独立时，$Cov(X,Y)=0$ 从而 $\rho_{XY}=0$，即 X,Y 不相关。反之，若 X,Y 不相关，X 和 Y 却不一定独立

这是因为不相关只是就线性关系来说的，而相互独立一般是就一般关系而言的。

§4 矩、协方差矩阵

定义 设 X 和 Y 是随机变量，若

$$E(X^k) \qquad k=1,2,\cdots$$

存在，称它为 X 的 k 阶原点矩，简称 k 阶矩

若

$$E\{[X-E(X)]^k\} \quad k=2,3,\cdots$$

存在，称它为 X 的 k 阶中心矩

若

$$E(X^kY^l) \qquad k,l=1,2,\cdots$$

存在，称它为 X 和 Y 的 $k+l$ 阶混合矩

若

$$E\{[X-E(X)]^k[Y-E(Y)]^l\} \quad k,l=1,2,\cdots$$

存在，称它为 X 和 Y 的 $k+l$ 阶混合中心矩

显然，X 的数学期望 $E(X)$ 是 X 的一阶原点矩，方差 $D(X)$ 是 X 的二阶中心矩，协方差 $Cov(X,Y)$ 是 X 和 Y 的二阶混合中心矩

下面介绍 n 维随机变量的协方差矩阵，先从二维变量讲起

二维随机变量 (X_1,X_2) 有四个二阶中心矩（设它们都存在），分别记为

$$C_{11}=E\{[X_1-E(X_1)]^2\}$$
$$C_{12}=E\{[X_1-E(X_1)][X_2-E(X_2)]\}$$
$$C_{21}=E\{[X_2-E(X_2)][X_1-E(X_1)]\}$$
$$C_{22}=E\{[X_2-E(X_2)]^2\}$$

将它们排成矩阵的形式

$$\begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}$$

这矩阵称为随机变量 (X_1,X_2) 的协方差矩阵

一般地，设 n 维随机变量 $(X_1,X_2,\cdots,X_n)$ 的二阶混合中心矩

$$C_{ij}=Cov(X_i,X_j)=E\{[X_i-E(X_i)][X_j-E(X_j)]\} \quad i,j=1,2,\cdots,n$$

都存在，则称矩阵

$$C=\begin{pmatrix} C_{11} & C_{12} & \cdots & C_{1n} \\ C_{21} & C_{22} & \cdots & C_{2n} \\ \vdots & \vdots & & \vdots \\ C_{n1} & C_{n2} & \cdots & C_{nn} \end{pmatrix}$$

为 n 维随机变量 $(X_1,X_2,\cdots,X_n)$ 的协方差矩阵

由于 $C_{ji}=C_{ij}$ $(i\ne j,\ i,j=1,2,\cdots,n)$，因而上述矩阵是一个对称矩阵

一般，n 维随机变量的分布是不知道的，或是太复杂，以致在数学上不易处理，因此在实际应用中协方差矩阵就显得重要了。

增补

下面增补一个重要的不等式

定理 设随机变量 X 具有数学期望 $E(X)=\mu$，方差 $D(X)=\sigma^2$，则对于任意正数 ε，不等式

$$P\{|X-\mu|\ge\varepsilon\}\le\frac{\sigma^2}{\varepsilon^2}$$

这一不等式称为切比雪夫（Chebyshev）不等式

证：我们只就连续型随机变量的情况说明

$$P\{|X-\mu|\ge\varepsilon\}=\int_{|x-\mu|\ge\varepsilon} f(x)\,dx$$
$$\le\int_{|x-\mu|\ge\varepsilon}\frac{|x-\mu|^2}{\varepsilon^2}f(x)\,dx \quad \left(\varepsilon>0 \text{ 时有 } \frac{|x-\mu|^2}{\varepsilon^2}\ge 1\right)$$
$$\le\frac{1}{\varepsilon^2}\int_{-\infty}^{\infty}(x-\mu)^2f(x)\,dx$$
$$=\frac{\sigma^2}{\varepsilon^2}$$

切比雪夫不等式也可以写成如下形式：

$$P\{|X-\mu|<\varepsilon\}\ge 1-\frac{\sigma^2}{\varepsilon^2}$$

切比雪夫不等式给出了在随机变量的分布未知，而只知道 $E(X)$ 和 $D(X)$ 的情况下估计概率 $P\{|X-E(X)|<\varepsilon\}$ 的界限

这个估计是比较粗糙的，如果已知随机变量的分布，那么所需求的概率可以确切的计算出来，也就没有利用这一不等式来作估计了。

第五章 大数定律及中心极限定理

§1 大数定律

第一章曾讲过，大量试验证实，随机事件 A 的频率 $f_n(A)$ 当重复试验的次数 n 增大时总呈现出稳定性，稳定在某一个常数的附近。频率的稳定性是概率定义的客观基础

弱大数定律（辛钦大数定律） 设 $X_1,X_2,\cdots$ 是相互独立，服从同一分布的随机变量序列，且具有数学期望 $E(X_k)=\mu\ (k=1,2,\cdots)$。作前 n 个变量的算术平均 $\frac{1}{n}\sum_{k=1}^{n}X_k$，则对于任意 $\varepsilon>0$，有

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

证：我们只在随机变量方差 $D(X_k)=\sigma^2\ (k=1,2,\cdots)$ 存在这一条件下证明

$$\because E\left(\frac{1}{n}\sum_{k=1}^{n}X_k\right)=\frac{1}{n}\sum_{k=1}^{n}E(X_k)=\frac{1}{n}n\mu=\mu$$

又由独立性得

$$D\left(\frac{1}{n}\sum_{k=1}^{n}X_k\right)=\frac{1}{n^2}\sum_{k=1}^{n}D(X_k)=\frac{1}{n^2}n\sigma^2=\frac{\sigma^2}{n}$$

由切比雪夫不等式

$$1\ge P\left\{\left|\frac{1}{n}\sum_{k=1}^{n}X_k-\mu\right|<\varepsilon\right\}\ge 1-\frac{\sigma^2/n}{\varepsilon^2}$$

取 $n\to\infty$ 得证

通俗地说，辛钦大数定律是说，对于独立同分布且有均值 μ 的随机变量 $X_1\sim X_n$，当 n 很大时他们的算术平均很可能接近 μ

设 $Y_1\sim Y_n$ 是一个随机变量的序列，a 是一个常数，若对于任意常数 ε，有

$$\lim_{n\to\infty}\{P|Y-a|<\varepsilon\}=1$$

则称序列 $Y_1\sim Y_n$ 依概率收敛于 a，记为

$$Y_n\xrightarrow{P}a$$

依概率收敛的序列有以下的性质

设 $X_n\xrightarrow{P}a$，$Y_n\xrightarrow{P}b$，又设函数 $g(x,y)$ 在 (a,b) 连续，则

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

这样，辛钦大数定律又可叙述为

弱大数定律（辛钦大数定律） 设随机变量 $X_1,\cdots,X_n$ 相互独立，服从同一分布且有数学期望 $E(X_k)=\mu\ (k=1,2,\cdots)$，则序列 $\bar{X}=\frac{1}{n}\sum_{k=1}^{n}X_k$ 依概率收敛于 μ

即 $\bar{X}\xrightarrow{P}\mu$

下面介绍辛钦大数定律的一个重要推论

伯努利大数定律 设 f_A 是 n 次独立重复实验中事件 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 $\varepsilon>0$，有

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

或

$$\lim_{n\to\infty}P\left\{\left|\frac{f_A}{n}-p\right|\ge\varepsilon\right\}=0$$

证：因为 $f_A\sim b(n,p)$，有

$$f_A=X_1+\cdots+X_n$$

其中 $X_1,\cdots,X_n$ 相互独立且都服从以 p 为参数的 $(0-1)$ 分布，因而 $E(X_k)=p\ (k=1,2,\cdots,n)$

故得结果

§2 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，而其中每一个别因素在总的影响中所起的作用是微小的，这种随机变量往往近似地服从正态分布，这种现象就是中心极限定理的客观背景

本节介绍三个常用的中心极限定理

定理1（独立同分布的中心极限定理） 设随机变量 $X_1,X_2,\cdots,X_n,\cdots$ 相互独立，且服从同一分布，且具有数学期望和方差：$E(X_k)=\mu$，$D(X_k)=\sigma^2>0\ (k=1,2,\cdots)$，则随机变量之和 $\sum_{k=1}^{n}X_k$ 的标准化变量

$$Y_n=\frac{\sum_{k=1}^{n}X_k-E\left(\sum_{k=1}^{n}X_k\right)}{\sqrt{D\left(\sum_{k=1}^{n}X_k\right)}}=\frac{\sum_{k=1}^{n}X_k-n\mu}{\sqrt{n}\sigma}$$

的分布函数 $F_n(x)$ 对于任意 x 满足

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

一般情况下，很难求出 n 个随机变量之和的分布函数，上式表明，当 n 充分大时，可以通过 $\Phi(x)$ 给出其近似的分布

$$\frac{\sum_{k=1}^{n}X_k-n\mu}{\sqrt{n}\sigma}\overset{\text{近似地}}{\sim}N(0,1)$$

将上式左端改写成 $\dfrac{\frac{1}{n}\sum_{k=1}^{n}X_k-\mu}{\sigma/\sqrt{n}}=\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$，这样，上述结果可写成，当 n 充分大时，

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\overset{\text{近似地}}{\sim}N(0,1) \quad \text{或} \quad \bar{X}\overset{\text{近似地}}{\sim}N\left(\mu,\frac{\sigma^2}{n}\right)$$

这是独立同分布的中心极限定理结果的另一个形式，这就是说，均值为 μ，方差为 $\sigma^2>0$ 的独立同分布的随机变量 $X_1,\cdots,X_n$ 的算术平均 $\bar{X}=\frac{1}{n}\sum_{k=1}^{n}X_k$ 当 n 充分大时近似地服从均值为 μ，方差为 $\frac{\sigma^2}{n}$ 的正态分布，这是大样本统计推断的基础

定理2（李雅普诺夫（Lyapunov）定理） 设随机变量 $X_1,X_2,\cdots,X_n,\cdots$ 相互独立，它们具有数学期望和方差

$$E(X_k)=\mu_k \quad D(X_k)=\sigma_k^2>0 \quad k=1,2,\cdots$$

记

$$B_n^2=\sum_{k=1}^{n}\sigma_k^2$$

若存在正数 δ，使得当 $n\to\infty$ 时，

$$\frac{1}{B_n^{2+\delta}}\sum_{k=1}^{n}E\{|X_k-\mu_k|^{2+\delta}\}\to 0$$

则随机变量之和 $\sum_{k=1}^{n}X_k$ 的标准化变量

$$Z_n=\frac{\sum_{k=1}^{n}X_k-E\left(\sum_{k=1}^{n}X_k\right)}{\sqrt{D\left(\sum_{k=1}^{n}X_k\right)}}=\frac{\sum_{k=1}^{n}X_k-\sum_{k=1}^{n}\mu_k}{B_n}$$

的分布函数 $F_n(x)$ 对于任意 x，满足

$$\lim_{n\to\infty}F_n(x)=\lim_{n\to\infty}P\left\{\frac{\sum_{k=1}^{n}X_k-\sum_{k=1}^{n}\mu_k}{B_n}\le x\right\}$$
$$=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt$$
$$=\Phi(x)$$

定理2表明，在定理的条件下，随机变量

$$Z_n=\frac{\sum_{k=1}^{n}X_k-\sum_{k=1}^{n}\mu_k}{B_n}$$

当 n 很大时，近似地服从正态分布 $N(0,1)$，由此，当 n 很大时，$\sum_{k=1}^{n}X_k=B_nZ_n+\sum_{k=1}^{n}\mu_k$ 近似地服从正态分布 $N\left(\sum_{k=1}^{n}\mu_k,B_n^2\right)$。这就是说，无论各个随机变量 $X_k\ (k=1,2,\cdots)$ 服从什么分布，只要满足定理的条件，那么它们的和 $\sum_{k=1}^{n}X_k$ 当 n 很大时，就近似地服从正态分布，这就是随机变量在概率论中占有重要地位的一个基本原因。

定理3（棣莫弗－拉普拉斯（De Moivre－Laplace）定理） 设随机变量 $\eta_n\ (n=1,2,\cdots)$ 服从参数为 $n,p\ (0<p<1)$ 的二项分布，则对于任意 x，有

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

证：将二项分布 η_n 分解为 n 个独立同 $(0-1)$ 分布的随机变量 $X_1\sim X_n$，即有

$$\eta_n=\sum_{k=1}^{n}X_k \qquad E(X_k)=p \quad D(X_k)=p(1-p)$$

由定理1得结果

这定理表明，正态分布是二项分布的极限分布，当 n 足够大时，我们可以利用上式来计算二项分布的概率

注：与泊松近似（适用于 $n\to\infty$ $p\to 0$ λ 一般小于等于10）相比，正态二项分布近似在 $n\to\infty$，p 适中（p 趋近0.5）时近似效果较好

e.g. 对于一个学生而言，来参加家长会的家长人数是一个随机变量，设一个学生无家长，有1名家长，有2名家长来参加会议的概率分别为0.05、0.8、0.15，若学校共有400名学生，设学生参加会议的家长人数相互独立，且服从同一分布。

(1) 求参加会议的家长人数 X 超过450的概率

(2) 求有1名家长来参加会议的学生人数不多于340的概率

(1) 以 $X_k\ (k=1,2,\cdots,400)$ 记第 k 个学生来参加会议的家长人数，则 X_k 的分布律为

X_k	0	1	2
p_k	0.05	0.8	0.15

易知 $E(X_k)=1.1$ $D(X_k)=0.19$ $k=1,2,\cdots,400$ 而 $X=\sum_{k=1}^{400}X_k$，由定理1，

$$\frac{\sum_{k=1}^{400}X_k-400\times1.1}{\sqrt{400}\sqrt{0.19}}=\frac{X-400\times1.1}{\sqrt{400}\sqrt{0.19}}$$

近似服从正态分布 $N(0,1)$，于是

$$P\{X>450\}=P\left\{\frac{X-400\times1.1}{\sqrt{400}\sqrt{0.19}}>\frac{450-400\times1.1}{\sqrt{400}\sqrt{0.19}}\right\}$$
$$=1-\Phi(1.147)$$
$$\approx 0.1251$$

(2) 以 Y 记有1名家长参加会议的学生人数，则 $Y\sim b(400,0.8)$，由定理3

$$P\{Y\le 340\}=P\left\{\frac{Y-400\times0.8}{\sqrt{400\times0.8\times0.2}}\le\frac{340-400\times0.8}{\sqrt{400\times0.8\times0.2}}\right\}$$
$$\approx\Phi(2.5)$$
$$=0.9938$$