概率论与数理统计

主讲：四川大学 徐小湛教授

四川大学 徐小湛

我的QQ: 2243414853

微博: @川大徐小湛

微信: scuxuxz

Email: xuxzmail@163.com

§2.2 离散型随机变量及其分布律

第16讲 离散型随机变量及其分布律 (III)

几何分布与超几何分布

四川大学 徐小湛

前面我们讲了

二项分布和泊松分布

现在我们来讲

几何分布和超几何分布

（一）几何分布

几何分布与伯努利试验有关

设随机试验E只有两个可能的结果：A 和 $\overline{A}$

则称E为伯努利试验。

设 $P(A)=p$ $(0<p<1)$ 则 $P(\overline{A})=1-p$

将 E 独立重复进行n次，称为 n重伯努利试验。

设X表示n重伯努利试验中

事件A第一次发生时的试验次数，

则X是一个随机变量，

X可能的取值为$1, 2, \ldots, n$。

下面来求 X 的分布律。

设X表示n重伯努利试验中，事件A第一次发生时的试验次数，则X是一个随机变量，

X可能的取值为1, 2,…, n。下面来求 X 的分布律。

用A_i表示事件A在第i次试验中发生 (i =0,1, …, n)

假设第k次试验中A第一次发生，则

$$P\{X=k\}=P(\overline{A_1}\cdots\overline{A_{k-1}}\,A_k)$$

$$=P(\overline{A_1})\cdots P(\overline{A_{k-1}})P(A_k)=(1-p)\cdots(1-p)p$$

$$=(1-p)^{k-1}p$$

定义（几何分布）
设在多重伯努利试验中，事件A发生的概率为p（$0<p<1$），
记X为A第一次发生时的试验次数，则X取值k的概率

$$P\{X=k\}=(1-p)^{k-1}p \qquad (k=1,2,\ldots)$$

称X服从参数为p的几何分布，记作 $X\sim G(p)$

X	1	2	……	k	……
p_k	p	$(1-p)p$	……	$(1-p)^{k-1}p$	……

称X服从参数为p的几何分布，记作 $X \sim G(p)$

$$P\{X=k\}=(1-p)^{k-1}p \qquad (k=1,2,\ldots)$$

$$\sum_{k=1}^{\infty}(1-p)^{k-1}p = p\sum_{k=1}^{\infty}(1-p)^{k-1} = p\cdot\frac{1}{1-(1-p)}$$

$$= p\cdot\frac{1}{p} = 1$$

几何级数(等比级数)

例1　掷一颗色子，直到1点出现为止，求掷色子的次数*X*的分布律。

解　X 服从参数 p=1/6 的几何分布,

$$X \sim G(1/6)$$

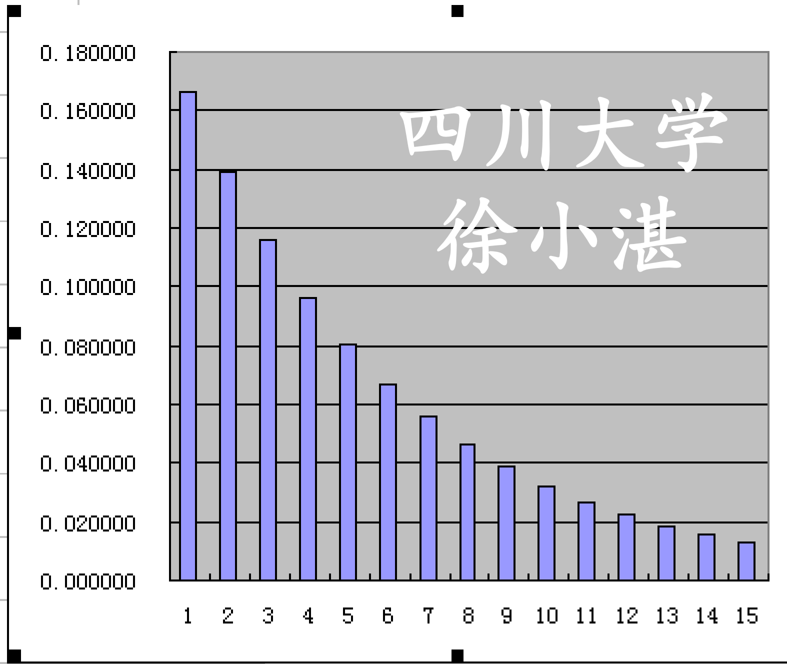
得　$$P\{X=k\}=(1-\frac{1}{6})^{k-1}\cdot\frac{1}{6}\ =\frac{1}{6}(\frac{5}{6})^{k-1}$$

用Excel计算　　$(k=1,2,\ldots)$

$$P\{X=k\}=(1-\frac{1}{6})^{k-1}\cdot\frac{1}{6}\ =\frac{1}{6}(\frac{5}{6})^{k-1}\quad(k=1,2,\ldots)$$

A	B	C	D	E
k	P{X=k}	P{X<=k}	p=	1/6
1	0. 166667	0. 166667		
2	0. 138889	0. 305556		
3	0. 115741	0. 421296		
4	0. 096451	0. 517747		
5	0. 080376	0. 598122		
6	0. 066980	0. 665102		
7	0. 055816	0. 720918		
8	0. 046514	0. 767432		
9	0. 038761	0. 806193		
10	0. 032301	0. 838494		
11	0. 026918	0. 865412		
12	0. 022431	0. 887843		
13	0. 018693	0. 906536		
14	0. 015577	0. 922113		
15	0. 012981	0. 935095		

Excel程序

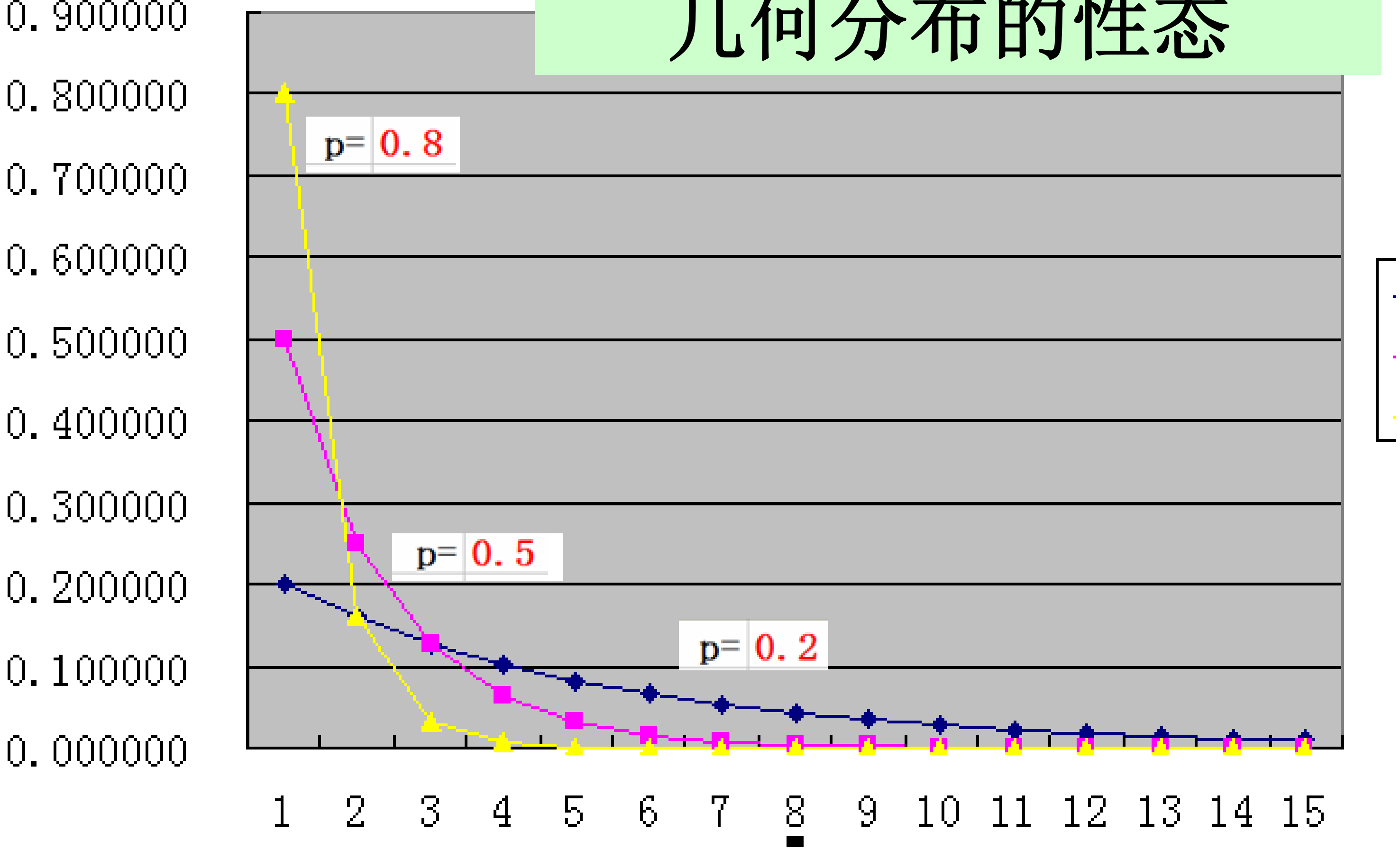


几何分布的性态

Excel程序

k	P{X=k}, p=0.2	P{X=k}, p=0.5	P{X=k}, p=0.8		
				p=	0.2
1	0.200000	0.500000	0.800000	p=	0.5
2	0.160000	0.250000	0.160000	p=	0.8
3	0.128000	0.125000	0.032000		
4	0.102400	0.062500	0.006400		
5	0.081920	0.031250	0.001280		
6	0.065536	0.015625	0.000256		
7	0.052429	0.007813	0.000051		
8	0.041943	0.003906	0.000010		
9	0.033554	0.001953	0.000002		
10	0.026844	0.000977	0.000000		
11	0.021475	0.000488	0.000000		
12	0.017180	0.000244	0.000000		
13	0.013744	0.000122	0.000000		
14	0.010995	0.000061	0.000000		
15	0.008796	0.000031	0.000000		

几何分布的性态



概率随 k 的增加单调减少。

p越大，概率减少越快。

（二）超几何分布

在第一章第4节例4（第6讲），

我们讲了这样一个例子：

设有N件产品，其中有D件次品。

今从中任取n件，

问其中恰有k件次品（$k\leqslant D$）的概率是多少？

现在来复习一下这个例子。

设有N件产品，其中有D件次品。今从中任取n件，问其中恰有k件次品($k \leqslant D$)的概率是多少？

k件次品	$n-k$件正品

取 n 件产品

Step 1 C_D^k 种取法　　Step 2 C_{N-D}^{n-k} 种取法　　C_N^n 种取法

D件次品	$N-D$件正品

N件产品

设有N件产品，其中有D件次品。今从中任取n件，问其中恰有k件次品($k \leqslant D$)的概率是多少？

在 N 件产品中取 n 件(不放回)的取法有 C_N^n 种

在D件次品中取k件的取法有 C_D^k 种

$$N(S) = C_N^n$$

在$N-D$件正品中取$n-k$件的取法有 C_{N-D}^{n-k} 种

由乘法原理，在N件产品中取n件，其中恰有k件次品的取法有 $C_D^k \cdot C_{N-D}^{n-k}$ 种

$$N(A) = C_D^k \cdot C_{N-D}^{n-k}$$

设有N件产品，其中有D件次品。今从中任取n件，问其中恰有k件次品($k \leqslant D$)的概率是多少？

在 N 件产品中取 n 件(不放回)的取法有 C_N^n 种

在D件次品中取k件的取法有 C_D^k 种 $N(S)=C_N^n$

在$N-D$件正品中取$n-k$件的取法有 C_{N-D}^{n-k} 种

由乘法原理，在N件产品中取n件，其中恰有k件次品的取法有 $C_D^k \cdot C_{N-D}^{n-k}$ 种 $N(A)=C_D^k \cdot C_{N-D}^{n-k}$

$$P(A)=\frac{N(A)}{N(S)}=\frac{C_D^k \cdot C_{N-D}^{n-k}}{C_N^n} \quad (k=0,1,2,\ldots,n)$$

设有N件产品，其中有M件次品。从中任取n件，则其中恰有k件次品($k\leqslant M$)的概率是

把D改成M

$$\frac{C_M^k \cdot C_{N-M}^{n-k}}{C_N^n}$$

设随机变量X表示取出n件产品中的次品数，

则 $$P\{X=k\}=\frac{C_M^k \cdot C_{N-M}^{n-k}}{C_N^n} \qquad k=0,1,\dots,n$$

称X服从参数为n, M, N的超几何分布，

Hypergeometric distribution

记为 $X \sim H(n,M,N)$

例2 从合格率为90%的100件产品中抽出30个检查，求其中不合格产品个数*X*的分布律。

解 X服从参数为 $n=30\quad M=10\quad N=100$

的超几何分布：$X\sim H(30,10,100)$

$$P\{X=k\}=\frac{C_{10}^{k}\cdot C_{90}^{30-k}}{C_{100}^{30}}\qquad k=0,1,2,\ldots,10$$

$$\boxed{P\{X=k\}=\frac{C_{M}^{k}\cdot C_{N-M}^{n-k}}{C_{N}^{n}}}$$

从合格率为90%的100件产品中抽出30个检查，求其中不合格产品个数X的分布律。

$$P\{X=k\}=\frac{C_{10}^{k}\cdot C_{90}^{30-k}}{C_{100}^{30}}\qquad k=0,1,2,\ldots,10$$

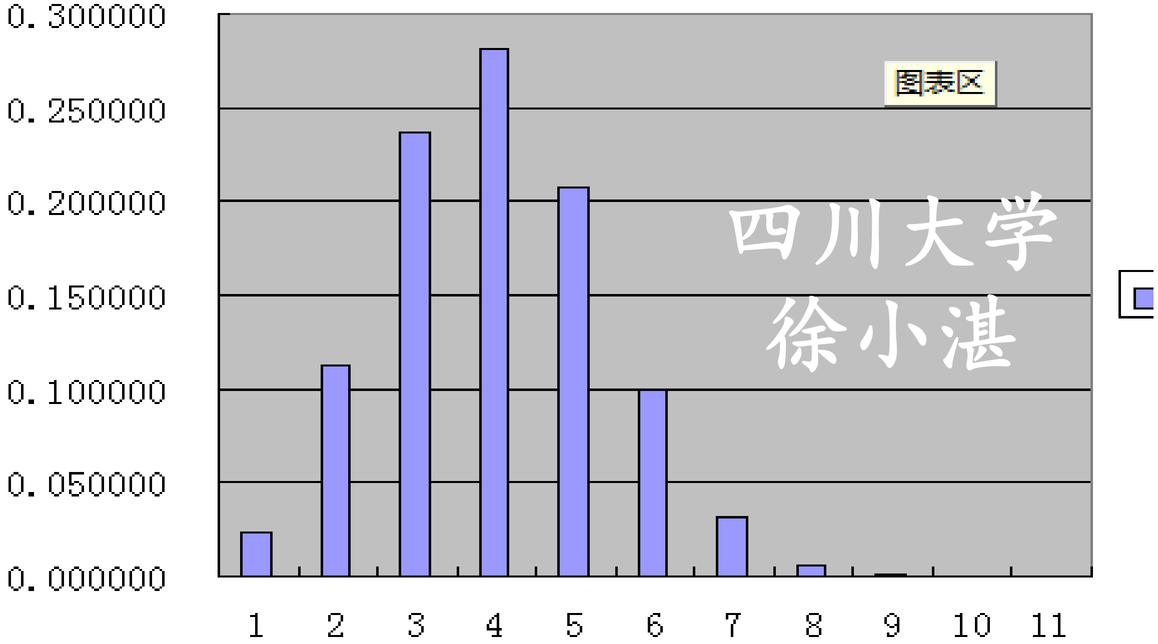
我们用Excel来计算

$$P\{X=k\}=\frac{C_{M}^{k}\cdot C_{N-M}^{n-k}}{C_{N}^{n}}$$ = **Hypgeomdist(k, n, M, N)**

$$P\{X=k\}=\frac{C_{10}^{k}\cdot C_{90}^{30-k}}{C_{100}^{30}}\qquad k=0,1,2,\dots,10$$

k	P{X=k}	P{X<=k}			
0	0.022917	0.022917	M=	10	次品数
1	0.112708	0.135625	N=	100	产品数
2	0.237232	0.372857			
3	0.281163				
4	0.207578				
5	0.099637				
6	0.031451				
7	0.006438				
8	0.000817				
9	0.000058				
10	0.000002				
次品数	单个概率				

n= 30 抽取数



超几何分布与二项分布的关系

当产品总数N很大时，采用不放回抽样，次品率几乎没有变化，因此不放回抽取n件产品近似于有放回抽取n件产品，从而近似于n重伯努利试验。

因此当N很大时，超几何分布与二项分布近似：

$$\frac{C_M^k \cdot C_{N-M}^{n-k}}{C_N^n} \approx C_n^k p^k (1-p)^{n-k} \qquad p=\frac{M}{N}$$

例3 500件产品的合格率为90%，(1) 不放回地抽出30个检查，求其中不合格产品个数X的分布律。(2) 有放回抽出30个检查，求其中不合格产品个数Y的分布律。

解 (1) $X \sim H(30,50,500)$

$$P\{X=k\}=\frac{C_M^k \cdot C_{N-M}^{n-k}}{C_N^n}$$

$$P\{X=k\}=\frac{C_{50}^k \cdot C_{450}^{30-k}}{C_{500}^{30}}$$

(2) Y 服从参数为 $n=30$ $p=0.1$ 的二项分布：

$Y \sim b(30,0.1)$

$$P\{Y=k\}=C_{10}^k(0.1)^k(0.9)^{10-k} \qquad k=0,1,2,\ldots,10$$

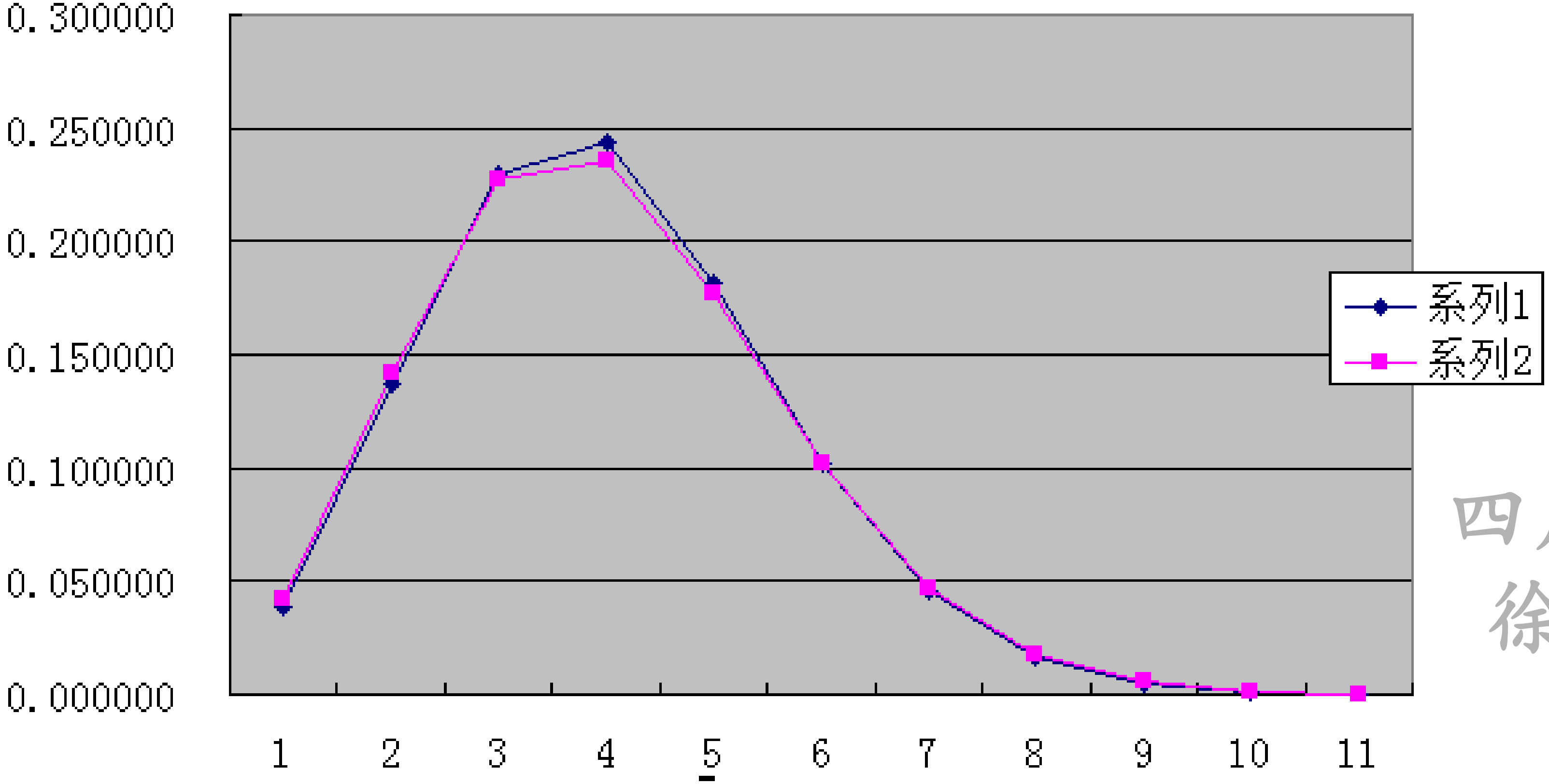
$$P\{X=k\}=\frac{C_{50}^{k}\cdot C_{450}^{30-k}}{C_{500}^{30}} \qquad P\{Y=k\}=C_{10}^{k}(0.1)^{k}(0.9)^{10-k}$$

A	B	C	D	E	F
k	P{X=k}	P{Y=k}	n=	30	抽取数
0	0.038323	0.042391	M=	50	次品数
1	0.136544	0.141304	N=	500	产品数
2	0.229893	0.227656	p=	0.1	次品率
3	0.243480	0.236088			
4	0.182179	0.177066			
5	0.102535	0.102305			
6	0.045130	0.047363			
7	0.015944	0.018043			
8	0.004605	0.005764			
9	0.001102	0.001565			
10	0.000221	0.000365			
次品数	超几何分布	二项分布			

Excel程序

$$P\{X=k\}=\frac{C_{50}^{k}\cdot C_{450}^{30-k}}{C_{500}^{30}}$$

$$P\{Y=k\}=C_{10}^{k}(0.1)^{k}(0.9)^{10-k}$$



我们学了以下离散型随机变量及其分布律

0-1分布（两点分布） $X \sim b(1,p)$

$$P\{X=k\}=p^k(1-p)^{1-k} \quad (k=0,1) \quad (0<p<1)$$

二项分布 $X \sim b(n,p)$ $(0<p<1)$

$$P\{X=k\}=C_n^k p^k(1-p)^{n-k} \quad (k=0,1,2,\dots)$$

泊松分布 $X \sim \pi(\lambda)$

$$P\{X=k\}=\frac{\lambda^k}{k!}e^{-\lambda} \quad (k=0,1,2,\dots) \quad (\lambda>0)$$

几何分布 $X \sim G(p)$

$$P\{X=k\}=(1-p)^{k-1}p \quad (k=1,2,\ldots) \quad (0<p<1)$$

超几何分布 $X \sim H(n,M,N)$

$$P\{X=k\}=\frac{C_M^k \cdot C_{N-M}^{n-k}}{C_N^n} \quad (k=0,1,2,\ldots n)$$

它们之间的关系

超几何分布 **二项分布** **泊松分布**

$$\frac{C_M^k \cdot C_{N-M}^{n-k}}{C_N^n} \approx C_n^k p^k (1-p)^{n-k} \approx \frac{\lambda^k}{k!} e^{-\lambda}$$

$$(p = M/N) \qquad (\lambda = np)$$

N和M比n大得多 **N较大，p较小**

请继续看下一讲

第17讲 随机变量的分布函数

本课程的视频在百度传课
http://www.chuanke.com
在百度传课搜：徐小湛

本课程主要参考教材

浙江大学

第四版

四川大学 徐小湛

我的QQ: 2243414853

微博: @川大徐小湛

微信: scuxuxz

Email: xuxzmail@163.com