概率计算

本次课主要内容

- 古典概率（factorial）
- 随机变量、概率密度函数、分布函数的概念
  - 二项分布
  - 泊松分布
  - 均匀分布
  - 指数分布
  - 正态分布

概率计算

实验目的

学习计算机概率计算的基本过程与方法

实验内容

1. 计算古典概型、分布率、概率密度函数、分布函数和逆累计分布函数的命令。
2. 计算实例。

一　古典概率

1. 古典概型

若随机实验E满足：

① 样本空间S只含有有限个元素

$$S=\{\omega_1,\ldots,\omega_N\}$$

② 试验中，每个基本事件发生是等可能的，则称E为古典概型试验，简称古典概型。

一　古典概率

2. 事件概率的计算公式

设随机实验E的样本空间S含有n个样本点，事件A包含k个样本点，定义 $P(A)= k/n$

3. 排列组合

① 阶乘$n!$的计算函数 **factorial(n)**　　factorial(7)　　5040

一　古典概率

② 排列 $A_n^k = n \cdot (n-1) \cdots (n-k+1) = \dfrac{n!}{(n-k)!}$

编辑pailie.m文件

```
function y=pailie(n,k)
y=factorial(n)/factorial(n-k)
```

③ 组合 $C_n^k = \dfrac{n!}{k!(n-k)!} = \dfrac{A_n^k}{k!}$

编写pailie和zuhe的m文件，并计算：

pailie(10,5)

zuhe(10,4)

编辑zuhe.m文件

```
function y=zuhe(n,k)
y=pailie(n,k)/factorial(k)
```

一　古典概率

例1. 一盒中有10只产品，其中有7只正品，3只次品。任取3只，求恰好有1只次品的概率。

解：设事件A=“任取3只，恰有1只次品” $P(A)=\dfrac{C_7^2\cdot C_3^1}{C_{10}^3}$

p=zuhe(7,2)*zuhe(3,1)/zuhe(10,3)

$\boxed{p=0.5250}$

二　随机变量与概率分布

1. 随机变量的分布函数

(cdf: Cumulative Distribution Function)

定义　设X为随机变量，对每个实数x，$F(x)=P(X\le x)$则称$F(x)$为X的分布函数。

分布函数的性质

- $F(X)$单调不减，即 $\forall\ x_1 < x_2$，$F(x_1) \le F(x_2)$
- $0 \le F(x) \le 1$　且 $\lim\limits_{x\to+\infty} F(x) = 1$，$\lim\limits_{x\to-\infty} F(x) = 0$
- $F(x)$右连续，即

$$F(x+0) \overset{\Delta}{=} \lim_{t\to x+0} F(t) = F(x)$$

$$P(a<X \le b) = F(b) - F(a)$$

2. 离散型随机变量及其概率分布

定义　若随机变量X的可能取值是有限多个或无穷可列多个，则称X为离散型随机变量.

描述离散型随机变量的概率特性常用它的概率分布或分布律，即：$P(X = x_k) = p_k,\ k = 1,2,\ldots$

概率分布的性质　$p_k \geq 0,\ k = 1,2,\ldots$

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

二 随机变量与概率分布

离散型随机变量的分布函数

$$F(x)=P(X\le x)=\sum_{x_k\le x}P(X=x_k)=\sum_{x_k\le x}p_k$$

$$p_k=P\ (X\ =\ x_k)=F(x_k)\ -\ F(x_{k-1})$$

$F(x)$是分段阶梯函数，在X的可能取值x_k处发生间断，间断点为第一类跳跃间断点，在间断点处有跃度p_k.

3. 连续型随机变量

定义　设X是一随机变量，$F(x)$是它的分布函数．若存在一个非负可积函数$f(x)$，使得

$$F(X) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t \qquad -\infty < x < +\infty$$

则称X是连续型随机变量，$f(x)$是它的概率密度函数(p.d.f.)，简称为密度函数或概率密度．

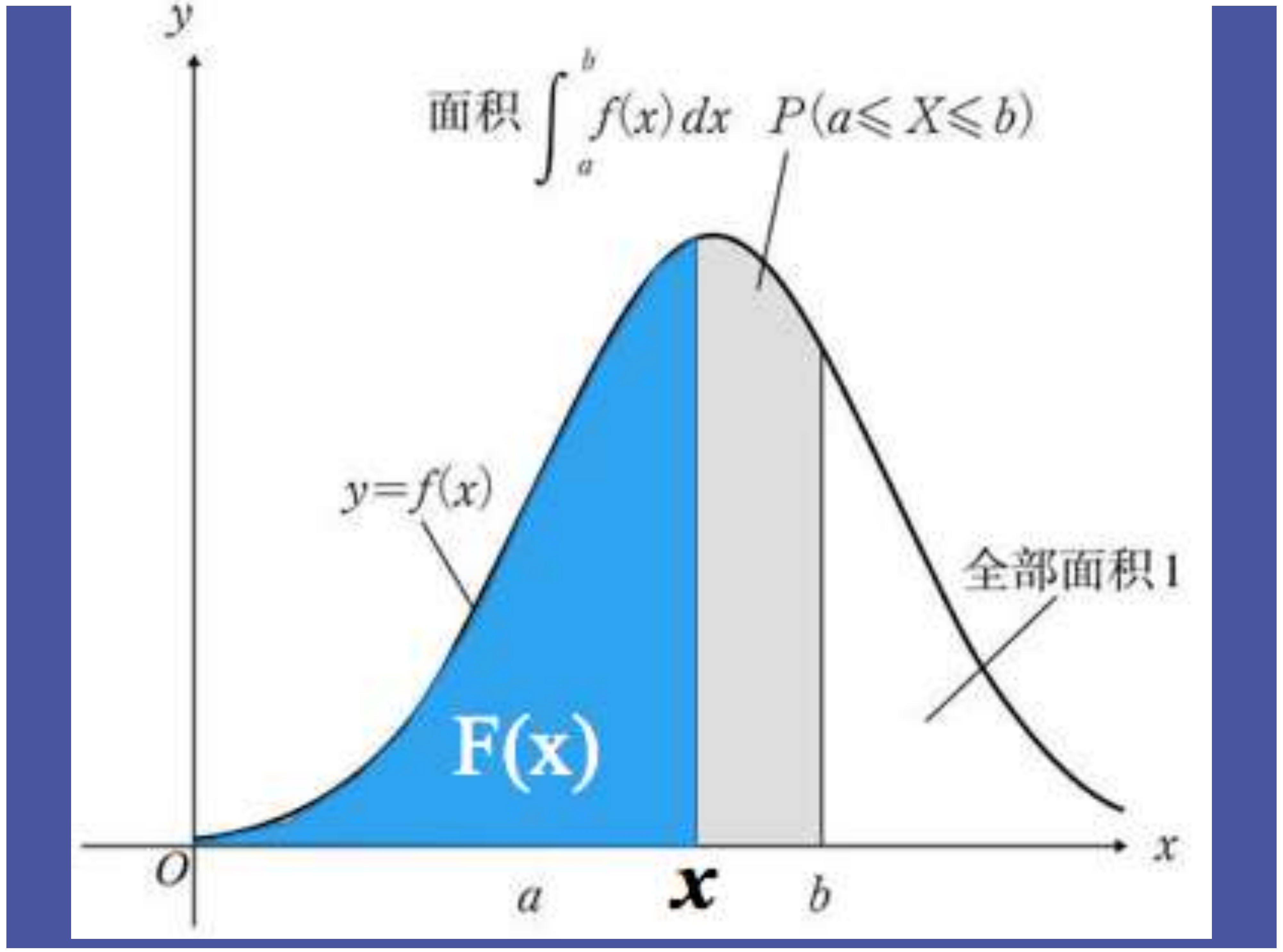
二 随机变量与概率分布

p.d.f. $f(x)$的性质

1. $f(x) \geq 0$

2. $\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = F(+\infty) = 1$

$$P(\mathrm{a} < X \leq b)$$
$$= P(\mathrm{a} < X < b)$$
$$= P(\mathrm{a} \leq X \leq b)$$
$$= P(\mathrm{a} \leq X < b)$$
$$= \int_{\mathrm{a}}^{\mathrm{b}} f(x)\,\mathrm{d}x$$



分布函数$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t$

4. 常见分布的计算

函数名	概率密度函数
binopdf	二项分布的概率密度函数
chi2pdf	卡方分布的概率密度函数
exppdf	指数分布的概率密度函数
fpdf	f分布的概率密度函数
gampdf	伽玛分布的概率密度函数
geopdf	几何分布的概率密度函数
hygepdf	超几何分布的概率密度函数
normpdf	正态分布的概率密度函数
poisspdf	泊松分布的概率密度函数
tpdf	学生氏t分布的概率密度函数
unidpdf	离散均匀分布的概率密度函数
unifpdf	连续均匀分布的概率密度函数

函数名	对应分布的分布函数
binocdf	二项分布的分布函数
chi2cdf	卡方分布的分布函数
expcdf	指数分布的分布函数
fcdf	f分布的分布函数
gamcdf	伽玛分布的分布函数
geocdf	几何分布的分布函数
hygecdf	超几何分布的分布函数
normcdf	正态分布的分布函数
poisscdf	泊松分布的分布函数
tcdf	学生氏t分布的分布函数
unidcdf	离散均匀分布的分布函数
unifcdf	连续均匀分布的分布函数

二　随机变量与概率分布

常见的离散型随机变量的分布

（1）二项分布　$B(n, p)$

背景：n重伯努利试验中，设每次试验感兴趣的事件A发生的概率为p, 则n次试验中事件A发生的次数X 是一离散型随机变量

若$P(A) = p$, $P_n(k) = P(X=k) = C_n^k p^k (1-p)^{n-k}$, $k = 0,1,\dots,n$

称X服从参数为n, p的二项分布，记作 $X \sim B(n, p)$

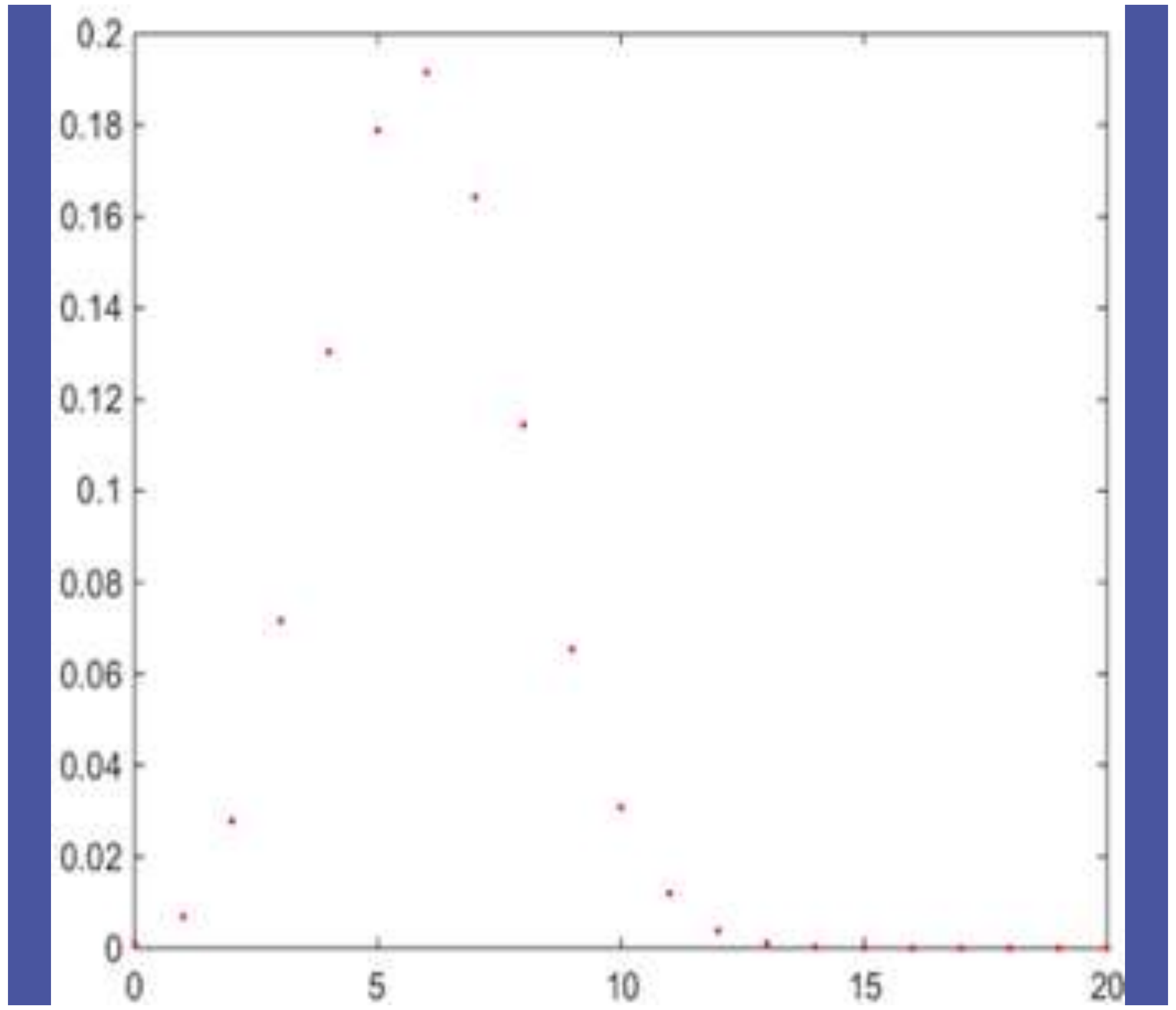
二　随机变量与概率分布

例2. 设X～B(20, 0.3)，求随机变量X取不同值的概率并作图展示.

调用binopdf命令即可.

```
syms x; x=0:1:20;
y=binopdf(x,20,0.3)
```

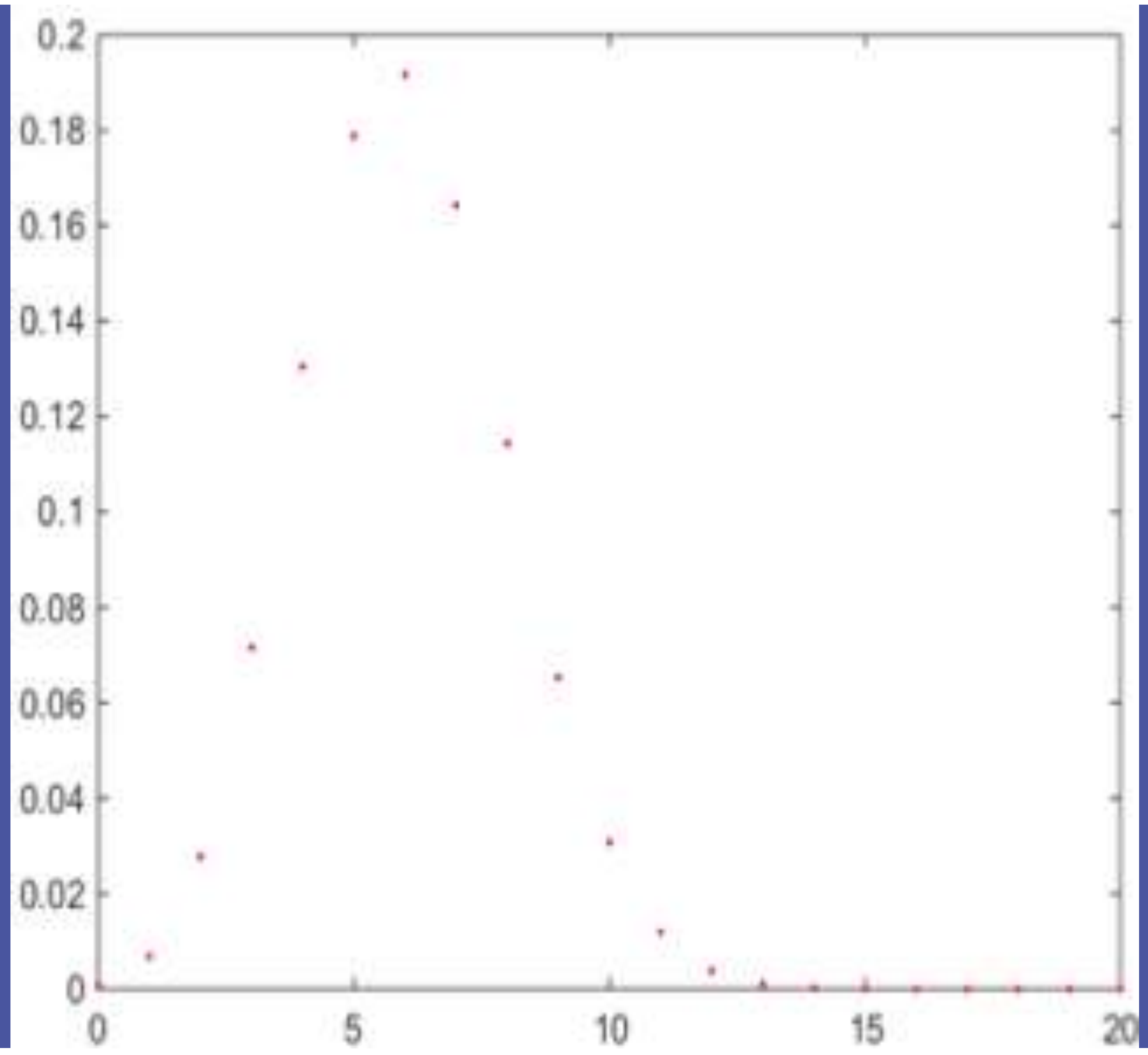
```
y = 0.0008    0.0068    0.0278    0.0716
    0.1304    0.1789    0.1916    0.1643
    0.1144    0.0654    0.0308    0.0120
    0.0039    0.0010    0.0002    0.0000
    0.0000    0.0000    0.0000    0.0000
    0.0000
```

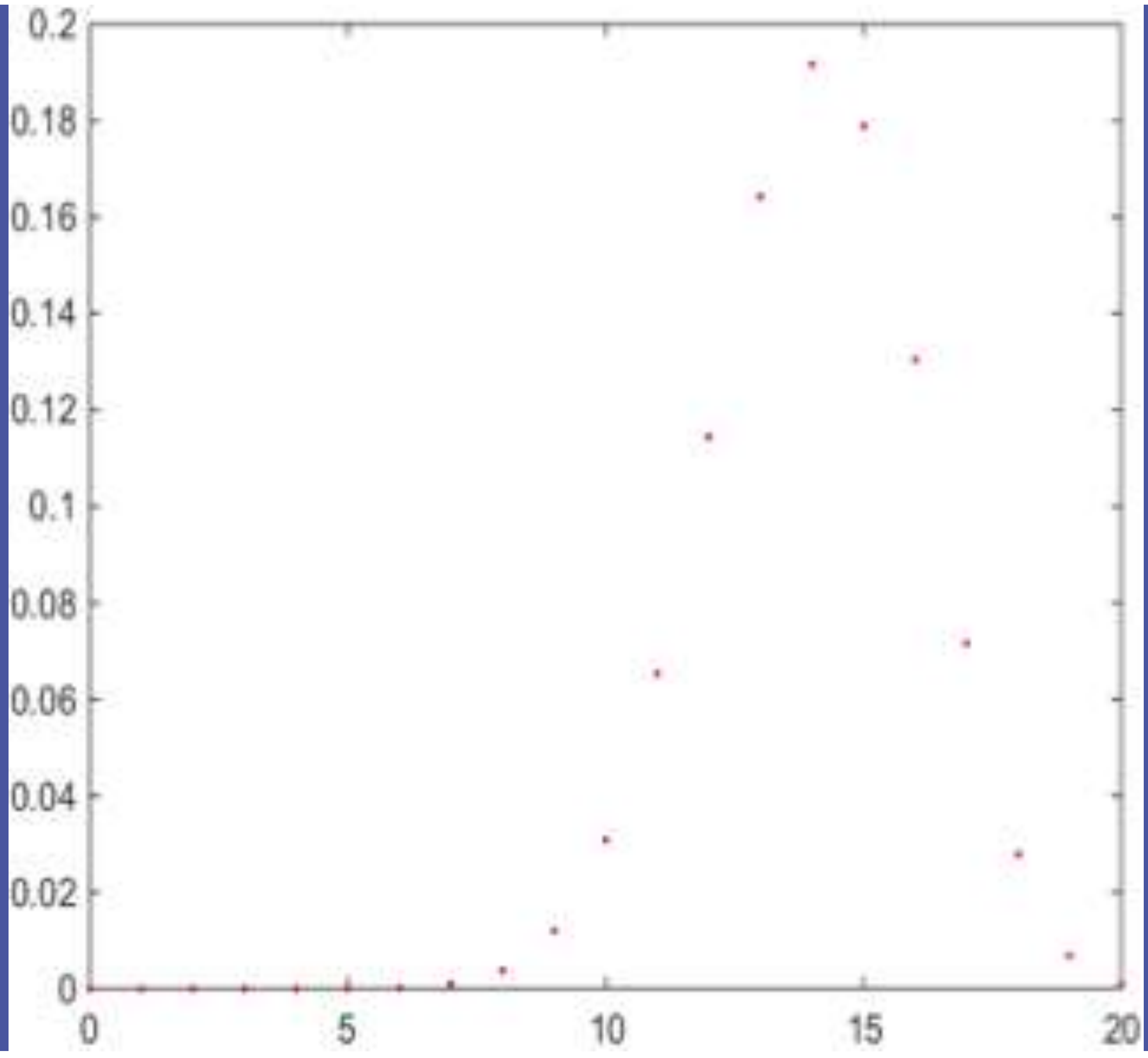


输入作图命令：plot(x, y,'r.')

练习：X～B(40, 0.3)，求随机变量X取不同值的概率并作图展示.

二　随机变量与概率分布





X～B(20, 0.3)　　　X～B(20, 0.7)

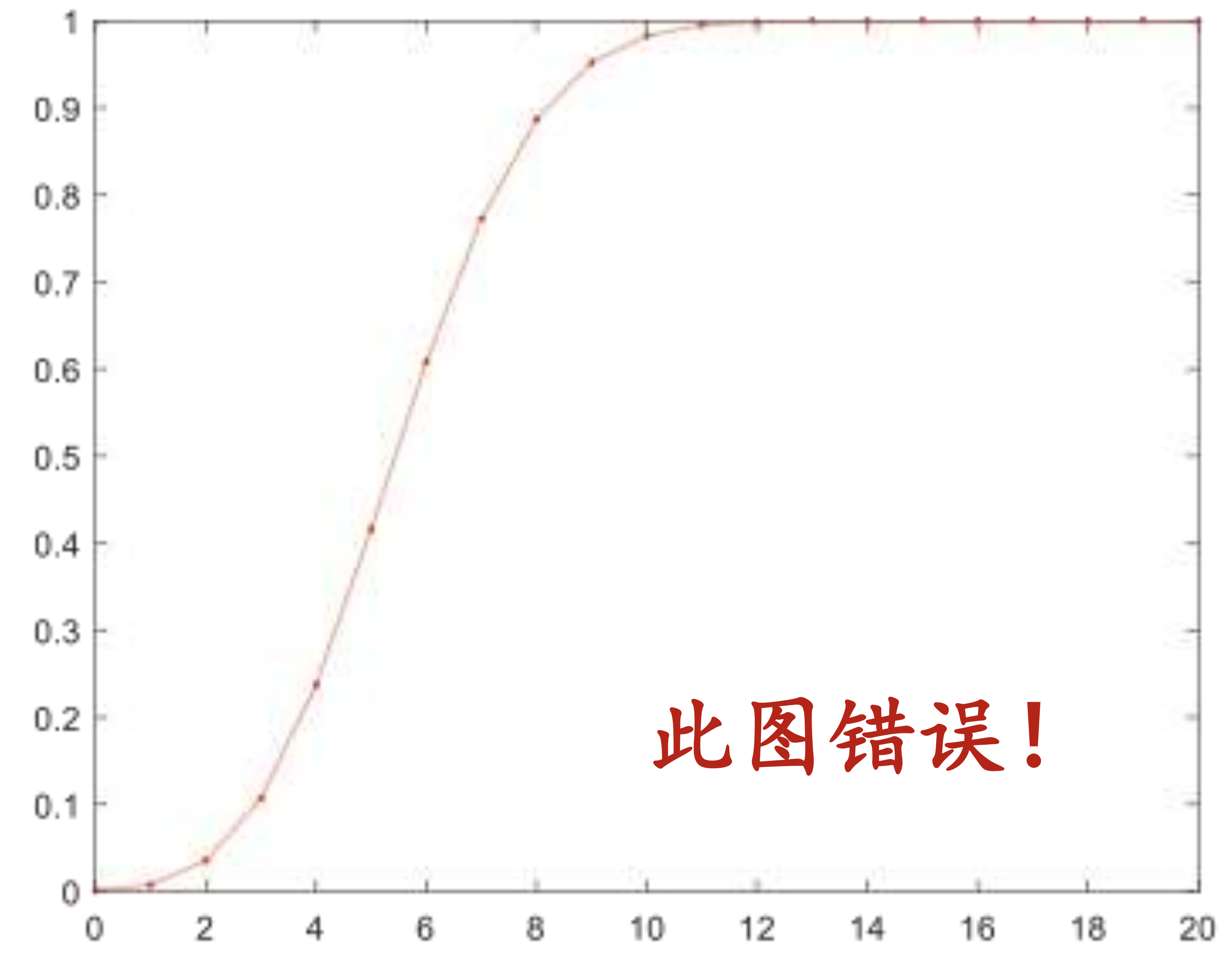
单峰的偏向性，　取决于B(n,p)中的p.

二　随机变量与概率分布

例3.　设X～B(20, 0.3)，求X的分布函数并作图展示.

```
x=0:1:20;
y=binocdf(x,20,0.3)
```

y = 0.0008	0.0076	0.0355	0.1071
0.2375	0.4164	0.6080	0.7723
0.8867	0.9520	0.9829	0.9949
0.9987	0.9997	1.0000	1.0000
1.0000	1.0000	1.0000	1.0000
1.0000			



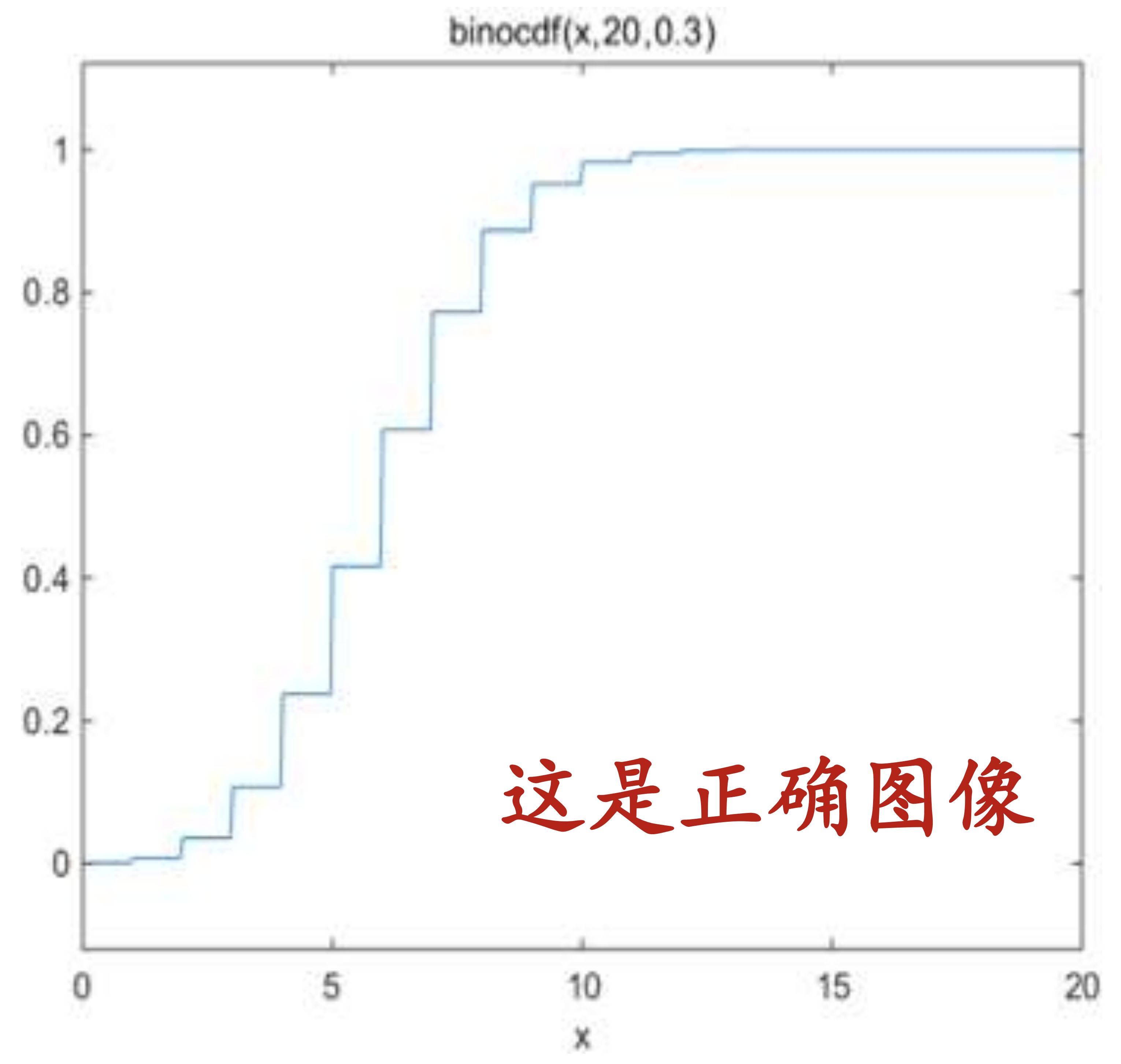
输入作图命令：plot(x,y)

二　随机变量与概率分布

例3. 设X～B(20, 0.3)，求X的分布函数并作图展示.

```
ezplot('binocdf(x,20,0.3)',[0,20])
```

使用ezplot命令直接绘制图形，可以看到离散型随机变量分布函数的典型特征：均为分段常值函数.



练习：X～B(20, 0.7)求分布函数并作图

（2）泊松分布　$\pi(\lambda)$

命令：

分布率：poisspdf(k,λ)

分布函数：poisscdf(k,λ)

定义：若离散型随机变量X的分布率为

$$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}$$

k=0, 1, 2,… 其中常数$\lambda>0$，则称X服从参数为λ的泊松分布，记为$X\sim\pi(\lambda)$。

一定时间里，进入超市购物的人数、电话总机接到的呼叫次数；一定空间内，微生物的个数等等．这些随机量通常都被认为服从泊松分布．

例4. 设电话总机1小时内收到的呼叫次数$X\sim\pi(3)$，求：

（1）1小时收到6次呼叫的概率；

（2）1小时收到呼叫次数不到5次的概率；

（3）绘出X的分布律、分布函数图像.

$X\sim\pi(\lambda)$，相应命令：

分布率：　poisspdf(k,λ)

分布函数：poisscdf(k,λ)

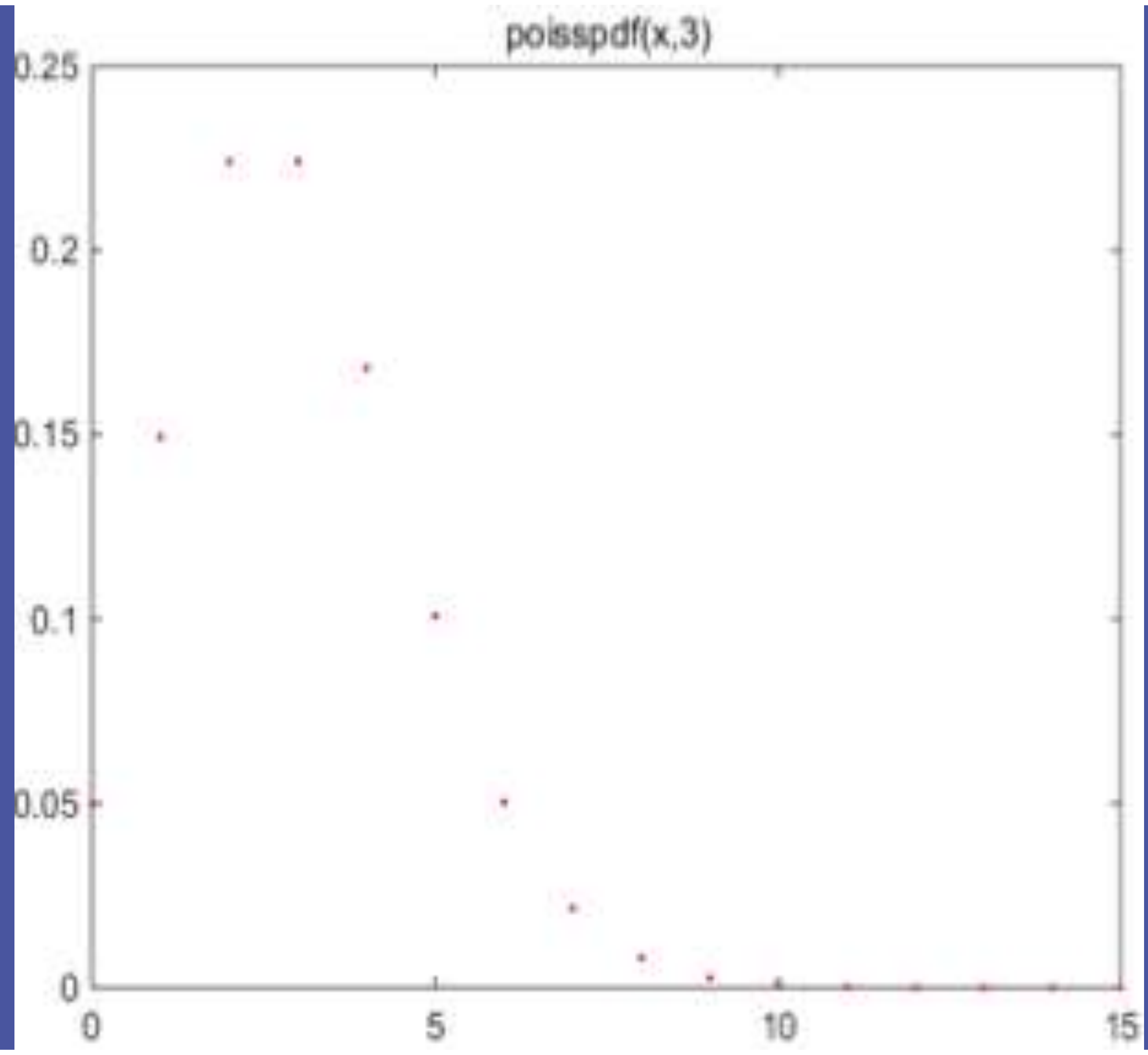
解：　$$P(X=k)=\frac{3^k}{k!}e^{-3}\ .$$

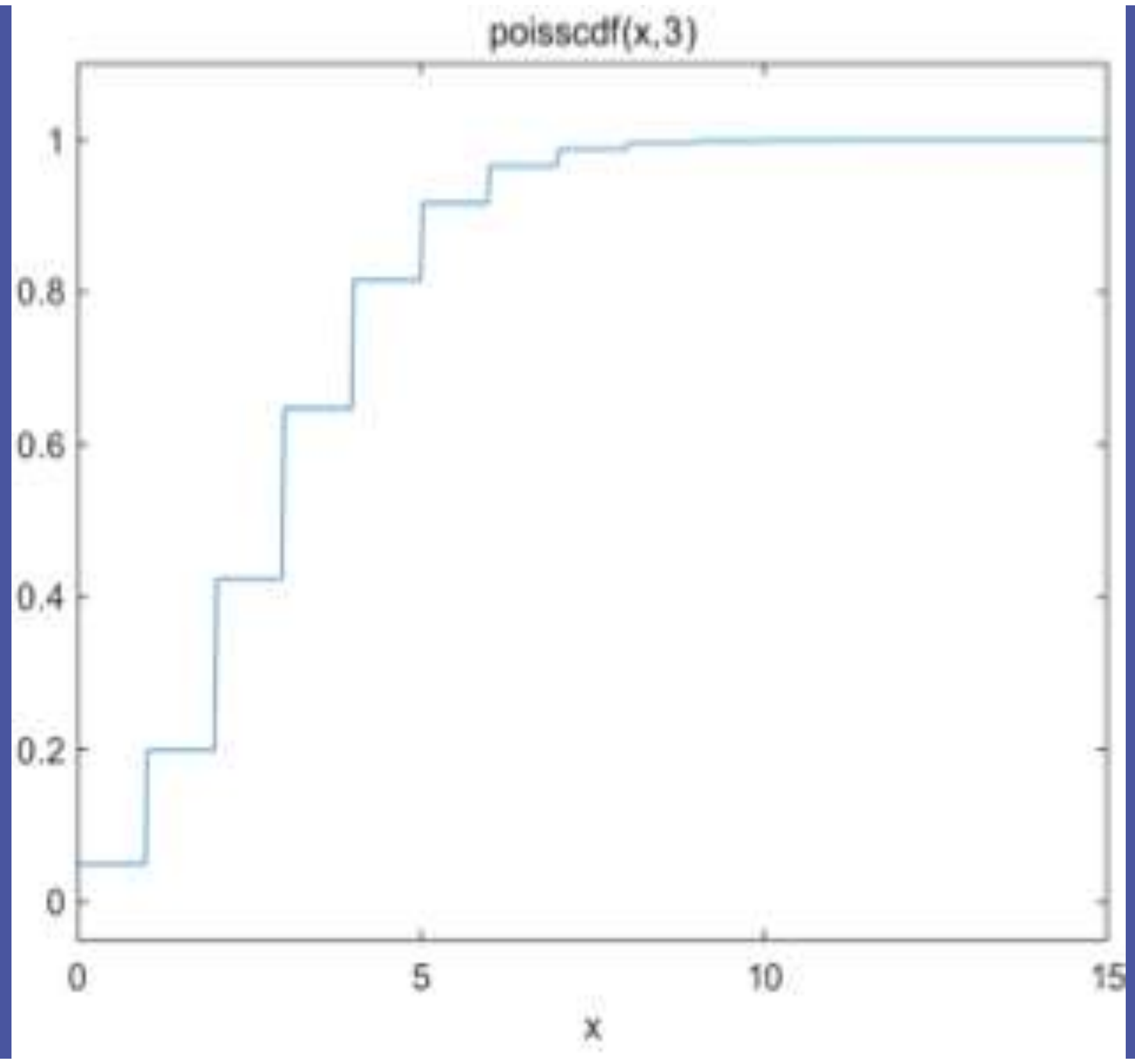
（1）输入：p=poisspdf(6,3)　　得到：p=0.0504.

（2）输入：p=poisscdf(4,3)　　得到：p=0.8153.

二　随机变量与概率分布

```
x=0:1:15;
y=poisspdf(x,3);
plot(x,y,'r.')
title('poisspdf(x,3)')
ezplot('poisscdf(x,3)',[0,15])
```





二　随机变量与概率分布

❖ 二项分布与泊松分布的关系

设有一列二项分布 $X_n \sim B(n, p_n)$, $n=1,2,\ldots$,如果 $\lim\limits_{n\to\infty} np_n = \lambda$

λ 是与n无关的正常数, 则对任意固定的非负整数k, 均有

$$\lim_{n\to\infty} P\{X_n = k\} = \lim_{n\to\infty} C_n^k p^k (1-p_n)^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda}$$

$$C_n^k p^k (1-p)^{n-k} \approx \frac{\lambda^k}{k!} e^{-\lambda}$$

例5　某人进行射击，设每次射击的命中率为0.001，他独立射击了5000次，试求他至少命中两次的概率。

解：设命中次数为X，则 $X \sim B(5000,0.001)$ 二项分布

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

$$= 1 - C_{5000}^{0}\, 0.001^{0}\, 0.999^{5000} - C_{5000}^{1}\, 0.001^{1}\, 0.999^{4999}$$

可用泊松分布进行近似计算，此时 $\lambda=5000\times0.001=5$

$$(1) = 1-\frac{5^0}{0!}e^{-5} - \frac{5^1}{1!}e^{-5} = 0.9596$$

二　随机变量与概率分布

精确值：

（1）=1-binopdf(0,5000,0.001)-binopdf(1,5000,0.001)

=0.959639689041808

近似值：

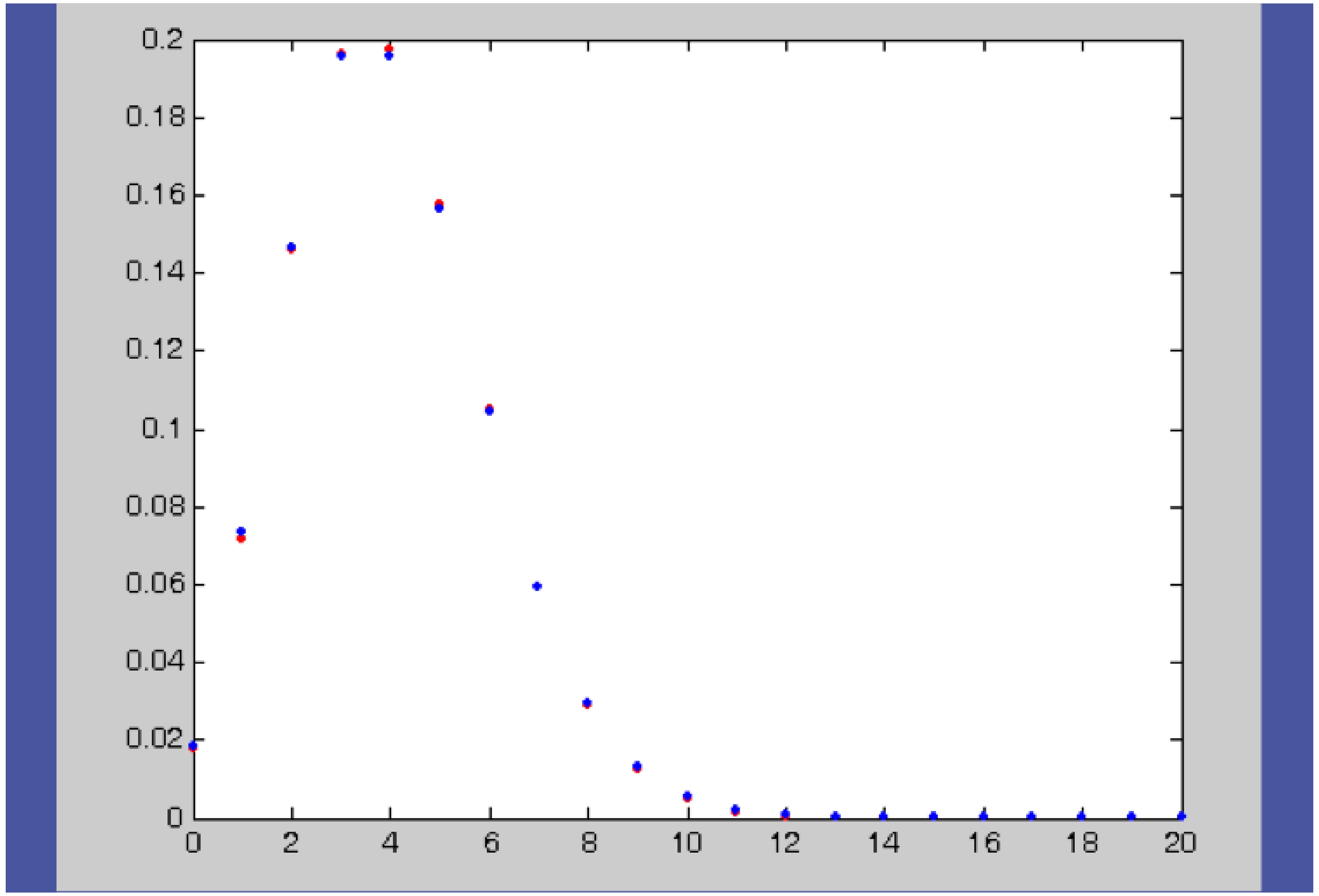
（1）=1-poisspdf(0,5)-poisspdf(1,5)

=0.959572318005487

二　随机变量与概率分布

例6：X～B(200,0.02)，*Y*服从参数为4的泊松分布，画出分布率图像

```
x=0:20;
y1=binopdf(x,200,0.02);
y2=poisspdf(x,4);
plot(x,y1,'r.',x,y2,'b.')
```



（3）离散均匀分布

命令：unidpdf(k,N)　　unidcdf(k,N)

随机变量X在1到N各自然数之间等可能取值

在Matlab中输入以下命令：x=1:1:10；y=unidpdf(x,10)（分布律）

结果：y = 0.1000　0.1000　0.1000　0.1000　0.1000　0.1000　0.1000　0.1000　0.1000　0.1000

在Matlab中输入以下命令：x=0:1:10；y=unidcdf(x,10)（分布函数）

结果：y = 0　0.1000　0.2000　0.3000　0.4000　0.5000　0.6000　0.7000　0.8000　0.9000　1.0000

二　随机变量与概率分布

常见的连续型随机变量的分布

（1）均匀分布

若X的密度函数为$f(x)$，则称X服从区间(a, b)上的均匀分布

记作$X \sim U(a,b)$

其中
$$f(x)=\begin{cases}\dfrac{1}{b-a}, & a<x<b \\ 0, & \text{其他}\end{cases}$$

$\forall\ (c,d) \subset (a,b)$

$$P(c<X<d) = \int_c^d \frac{1}{b-a}\,\mathrm{d}x = \frac{d-c}{b-a}$$

即X的取值在$(a,\ b)$内任何长为$d-c$的小区间的概率与小区间的位置无关，只与其长度成正比．这正是几何概型的情形．

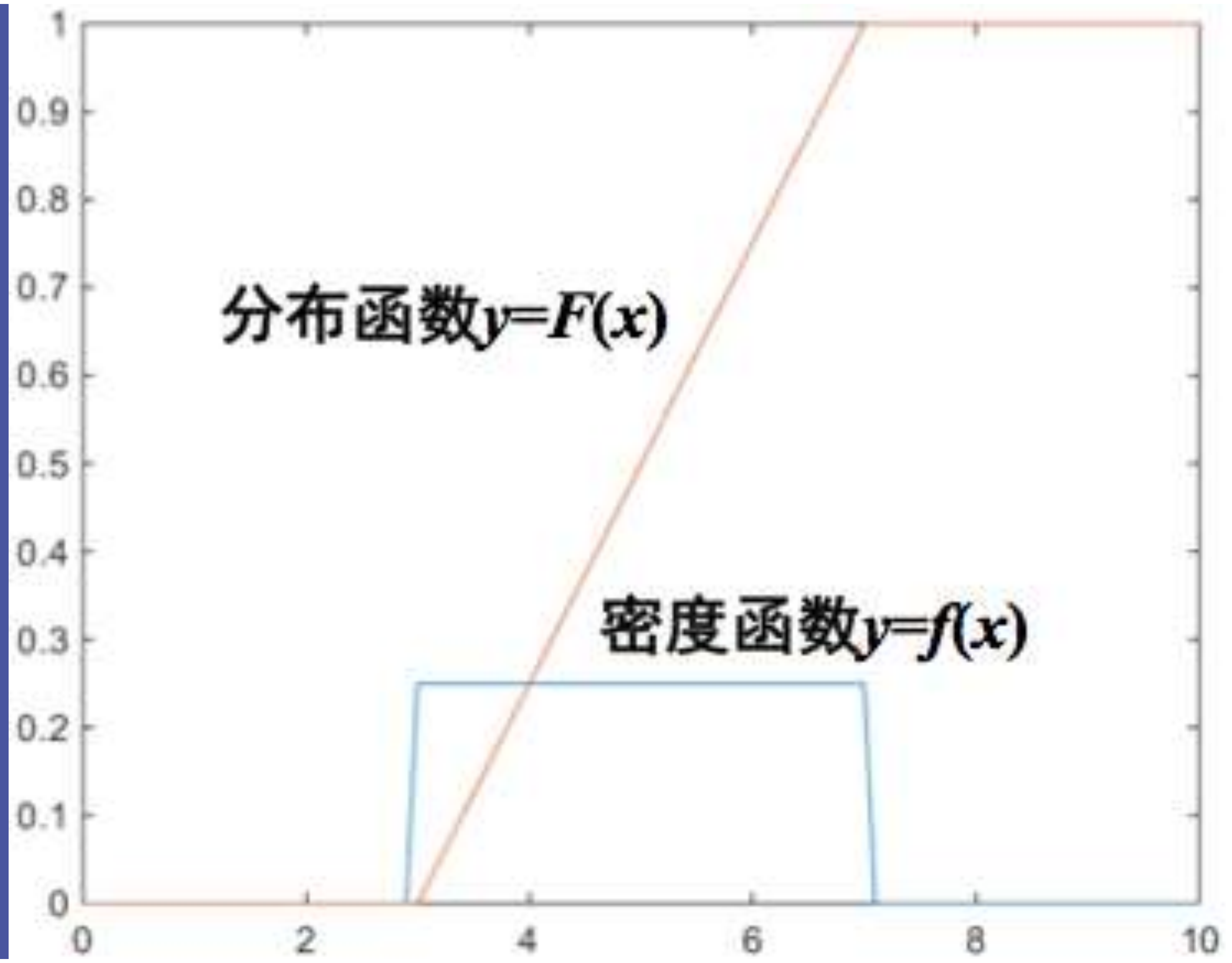
二　随机变量与概率分布

均匀分布：密度f=unifpdf(x,a,b) 分布函数：f=unifcdf(x,a,b)

例1．假设X～U(3,7)，画出X的概率密度函数和分布函数.

输入命令：

```
x=0:0.1:10;
f=unifpdf(x,3,7);
F=unifcdf(x,3,7);
plot(x,f,x,F)
```



练习：X～U(10,200)画概率密度图、分布函数图

（2）指数分布

若X的密度函数为 $f(x)=\begin{cases}\dfrac{1}{\theta}e^{-x/\theta}, & x>0\\ 0, & 其他\end{cases}$ $\theta>0$为常数，

则称X服从参数为θ的指数分布. 记作$X\sim E(\theta)$

如果一个随机变量X可能在$(0,+\infty)$上取值，且取值比较大的可能性很小. 如动物的寿命，人们打电话的通话时间，银行接待一位客户花费的时间等. 这时可以认为X服从参数为$\theta>0$的指数分布E(θ).

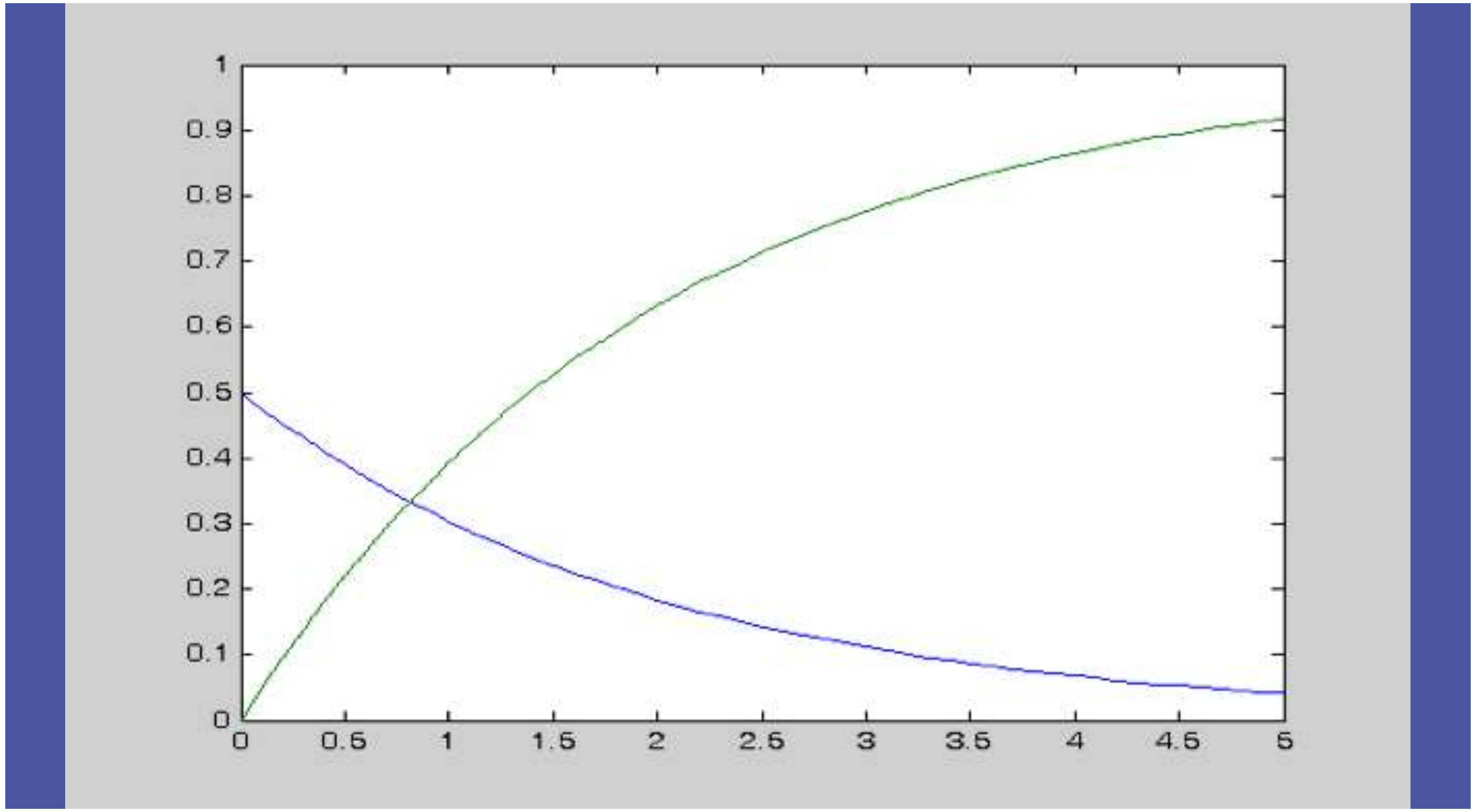
二　随机变量与概率分布

指数分布密度函数：f=exppdf(x,θ)　分布函数：F=expcdf(x,θ)

例2：画出指数分布$E(2)$的密度函数和分布函数图形．求 $P(0<X<5)$　$P(0<X<20)$. 在Matlab中输入以下命令：

```
x=0:0.1:5;
y=exppdf(x,2);
z=expcdf(x,2);
plot(x,y,x,z)        %密度函数和分布函数
result1=expcdf(5,2)-expcdf(0,2)
result2=expcdf(20,2)-expcdf(0,2)
```

二　随机变量与概率分布



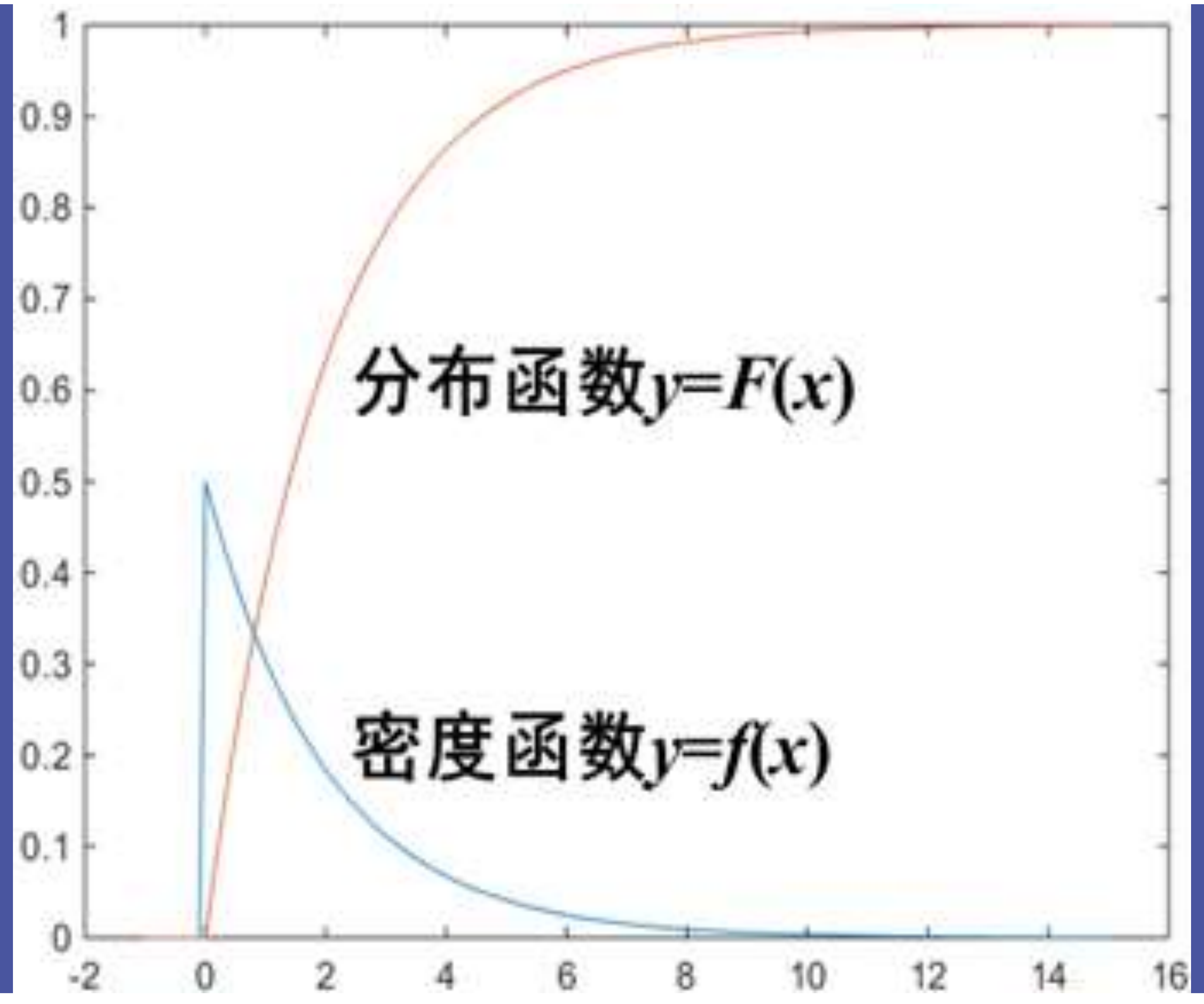
result1= 0.91791500137610

result2= 0.99995460007024

二　随机变量与概率分布

例3. 假设X～E(2)，画出X的概率密度函数和分布函数，并计算P(X>10)和P(2<X<5).

输入命令：
```
x=-1:0.1:30;
f=exppdf(x,2);
F=expcdf(x,2);
plot(x,f,x,F)
```



P(X>10)=1-expcdf(10,2)

P(2<X<5)=expcdf(5,2)-expcdf(2,2)

练习：X～E(0.1)画概率密度图、分布函数图

（3）正态分布

若X的密度函数为 $f(x)=\dfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ $-\infty<x<+\infty$

μ, σ为常数，$\sigma>0$则称X服从参数为μ, σ^2的正态分布，

记作$X\sim N(\mu,\ \sigma^2)$

❖ $f(x)$的两个参数：

μ—位置参数

固定μ，对于不同的σ，$f(x)$的形状不同.

σ—形状参数

即固定σ，对于不同的μ，对应的$f(X)$的形状不变化，只是位置不同

二　随机变量与概率分布

应用场合

若随机变量X受到众多相互独立的随机因素的影响，而每一个别因素的影响都是微小的，且这些影响可以叠加，则X服从正态分布.

可用正态变量描述的实例非常多：

各种测量的误差；	人的生理特征；
工厂产品的尺寸；	农作物的收获量；
海洋波浪的高度；	金属线的抗拉强度；
热噪声电流强度；	学生们的考试成绩；

二　随机变量与概率分布

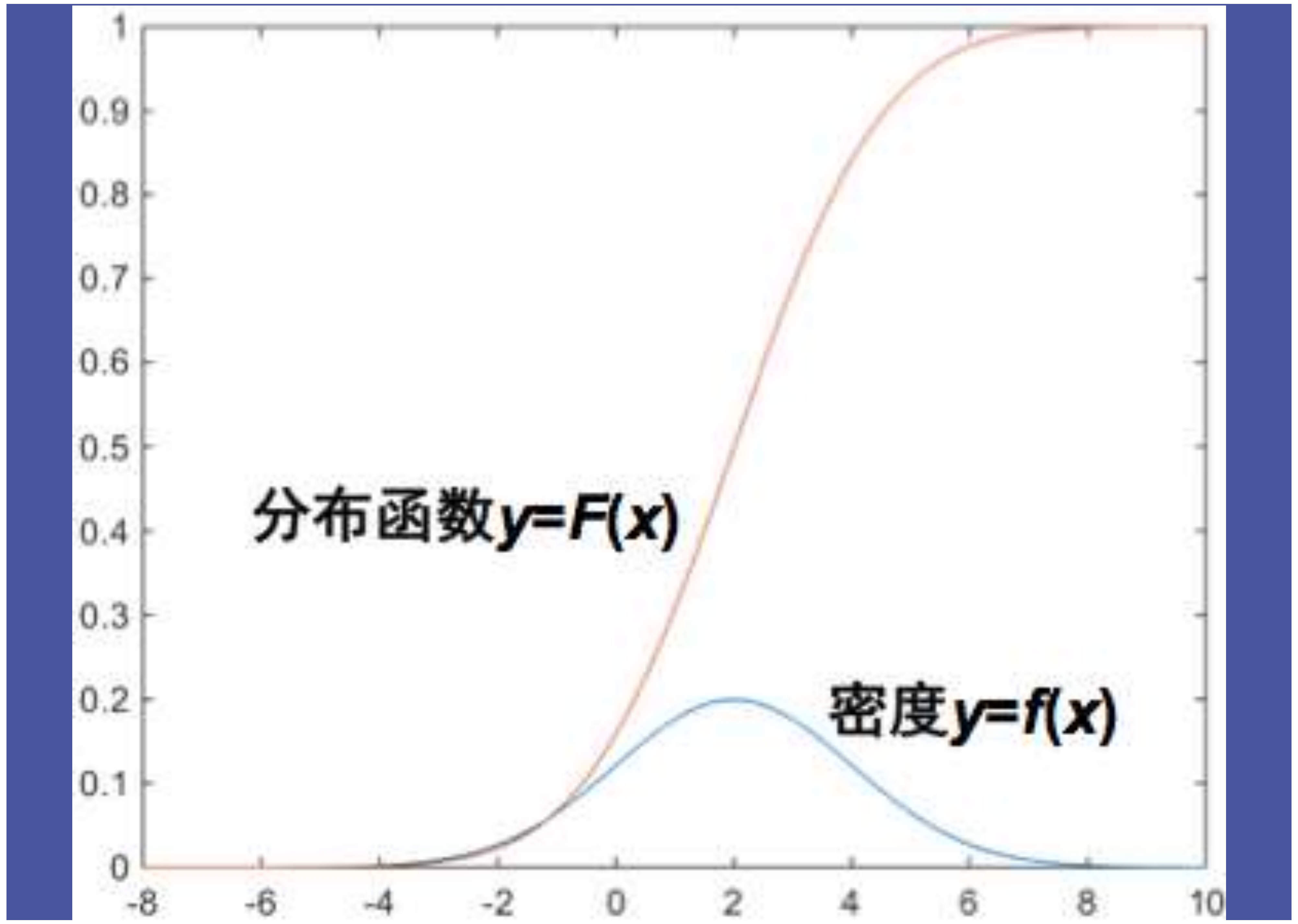
密度函数：$f(x)$=normpdf(x,μ,σ)

分布函数：$F(x)$=normcdf(x,μ,σ)

例4. 假设X～N(2,4)，画出X的概率密度函数和分布函数，并计算P(X>2).

输入命令：

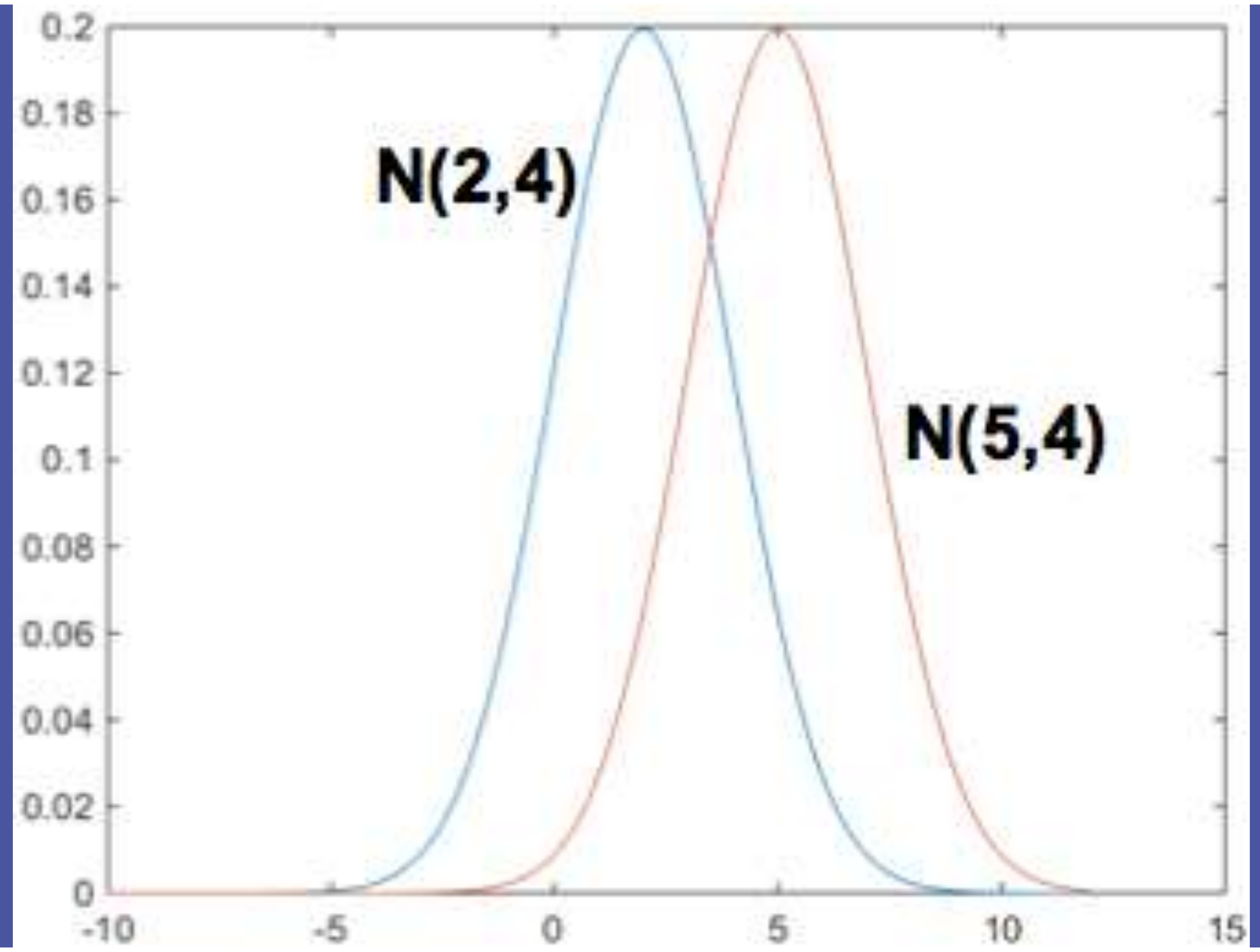
```
x=-8:0.1:10;
f=normpdf(x,2,2);
F=normcdf(x,2,2);
plot(x,f,x,F)
```

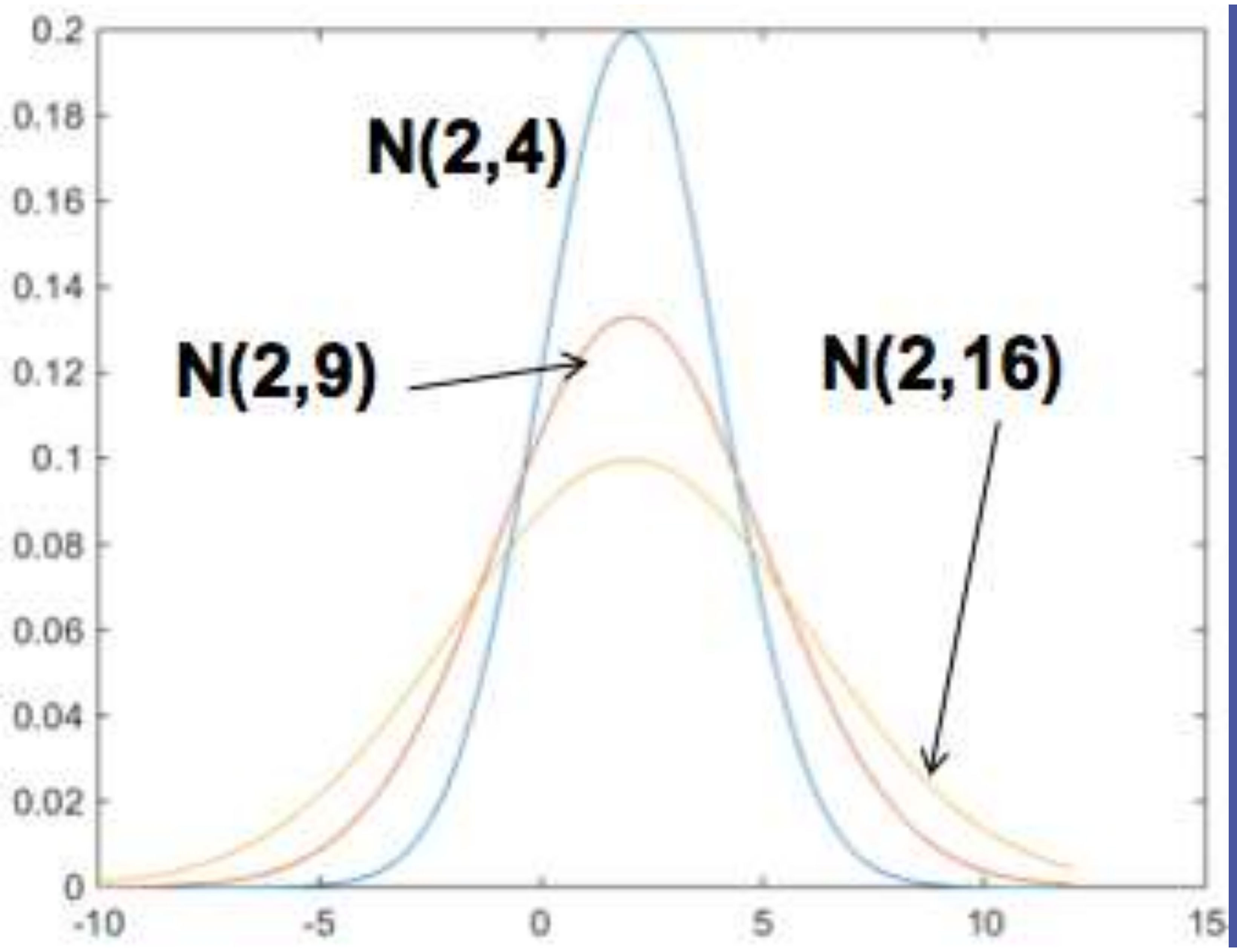


P(X>2)=1-normcdf(2,2,2)=0.5000

二　随机变量与概率分布

如果X～$N(\mu,\sigma^2)$，则X的数学期望为μ，方差为σ^2，当μ和σ变化时，X的概率密度函数会如何变化？





二　随机变量与概率分布

在同一坐标下，画下列正态分布的密度函数图像

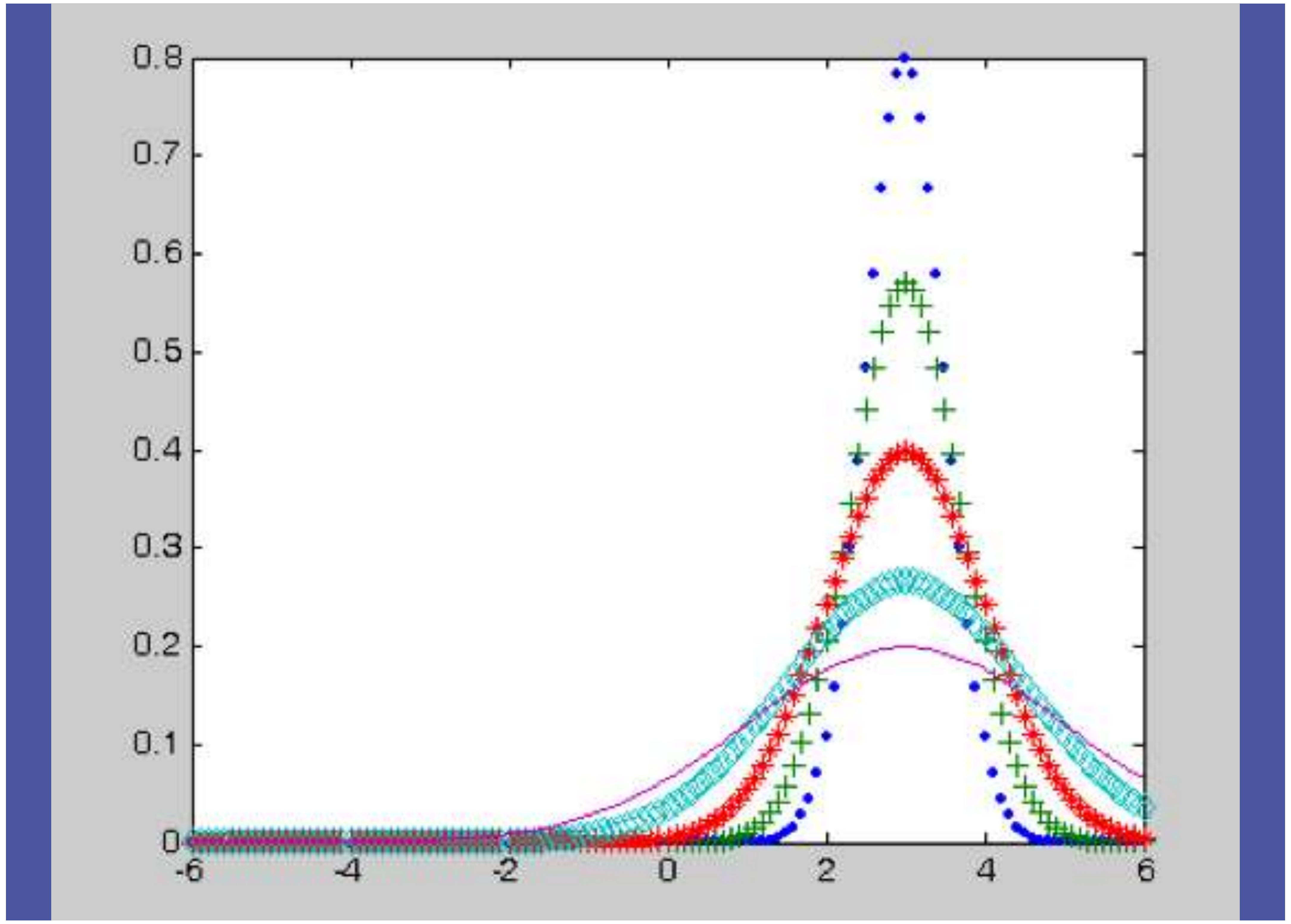
（1）$\mu=3$，$\sigma=0.5$，0.7，1，1.5，2

（2）$\sigma=0.5$，$\mu=1,2,3,4$

二　随机变量与概率分布

（1）命令：

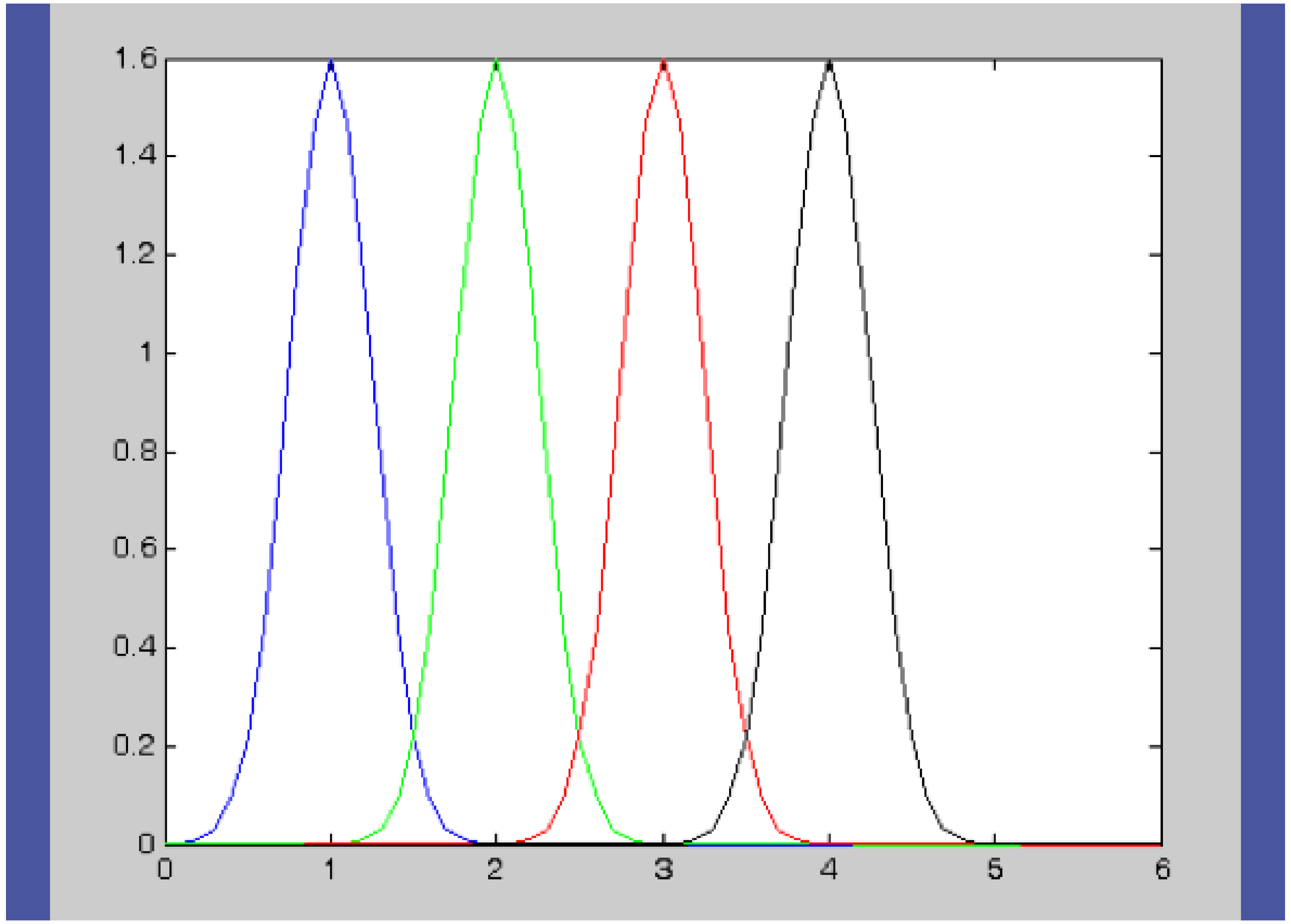
```
x=-6:0.1:6;
y1=normpdf(x,3,0.5);
y2=normpdf(x,3,0.7);
y3=normpdf(x,3,1);
y4=normpdf(x,3,1.5);
y5=normpdf(x,3,2);
plot(x,y1,'.',x,y2,'+',x,y3,'*', x,y4,'d',x,y5)
```



二　随机变量与概率分布

（2）命令：

```
x=0:0.1:6;
y1=normpdf(x,1,0.25);
y2=normpdf(x,2,0.25);
y3=normpdf(x,3,0.25);
y4=normpdf(x,4,0.25);
plot(x,y1,'b',x,y2,'g',x,y3,'r', x,y4,'k')
```



例5. 某行星周围有大量小卫星，质量X服从E(5000)分布，温度Y服从N(0,40000)分布. 假设小卫星的质量与温度相互独立. 随机选择一颗小卫星，问其质量高于8000单位且温度在-30到40单位之间的概率.

要求的是：

$P(X>8000$ 且 $-30\le Y\le 40)$.

$X \sim E(5000)$，$Y \sim N(0,40000)$，求$P(X>8000$且$-30 \le Y \le 40)$.

$P(X>8000)=$1-expcdf(8000,5000)

$P(-30 \le Y \le 40)=$normcdf(40,0,200)-normcdf(-30,0,200)

$P = P(X>8000) \cdot P(-30 \le Y \le 40)=0.0280$.

三　逆累积分布函数icdf

(Inverse Cumulative Distribution Function)

$p(a<x\leq b)=F(b)-F(a)$.　其中$F(x)=P(X\leq x)$.

如果已知$P(X\leq x)=p$，能否反过来求x?

逆累积分布函数 就是返回给定概率条件下的自变量的临界值，实际上是分布函数的逆函数。

已知p，求分布函数$F(x)=p$相对应的x，$\mathrm{inv}(p)=\inf\{x\colon F(x)\geq p\}$

三　逆累积分布函数icdf

逆累计分布函数的命令：

函数名	对应的逆累积分布函数
X=binoinv(y,n,p)	二项分布的逆累积分布函数
X=chi2inv(p,v)	卡方分布的逆累积分布函数
X=expinv(p,mu)	指数分布的逆累积分布函数
X=finv(p,v1,v2)	f分布的逆累积分布函数
X=gaminv(p,a,b)	伽玛分布的逆累积分布函数
X=geoinv(y,p)	几何分布的逆累积分布函数
X=hygeinv(p,m,k,n)	超几何分布的逆累积分布函数
X=norminv(p,mu,sgm)	正态分布的逆累积分布函数
X=poissinv(p,lambda)	泊松分布的逆累积分布函数
X=tinv(p,v)	学生氏t分布的逆累积分布函数
X=unidinv(p,n)	离散均匀分布的逆累积分布函数
X=unifinv(p,a,b)	连续均匀分布的逆累积分布函数

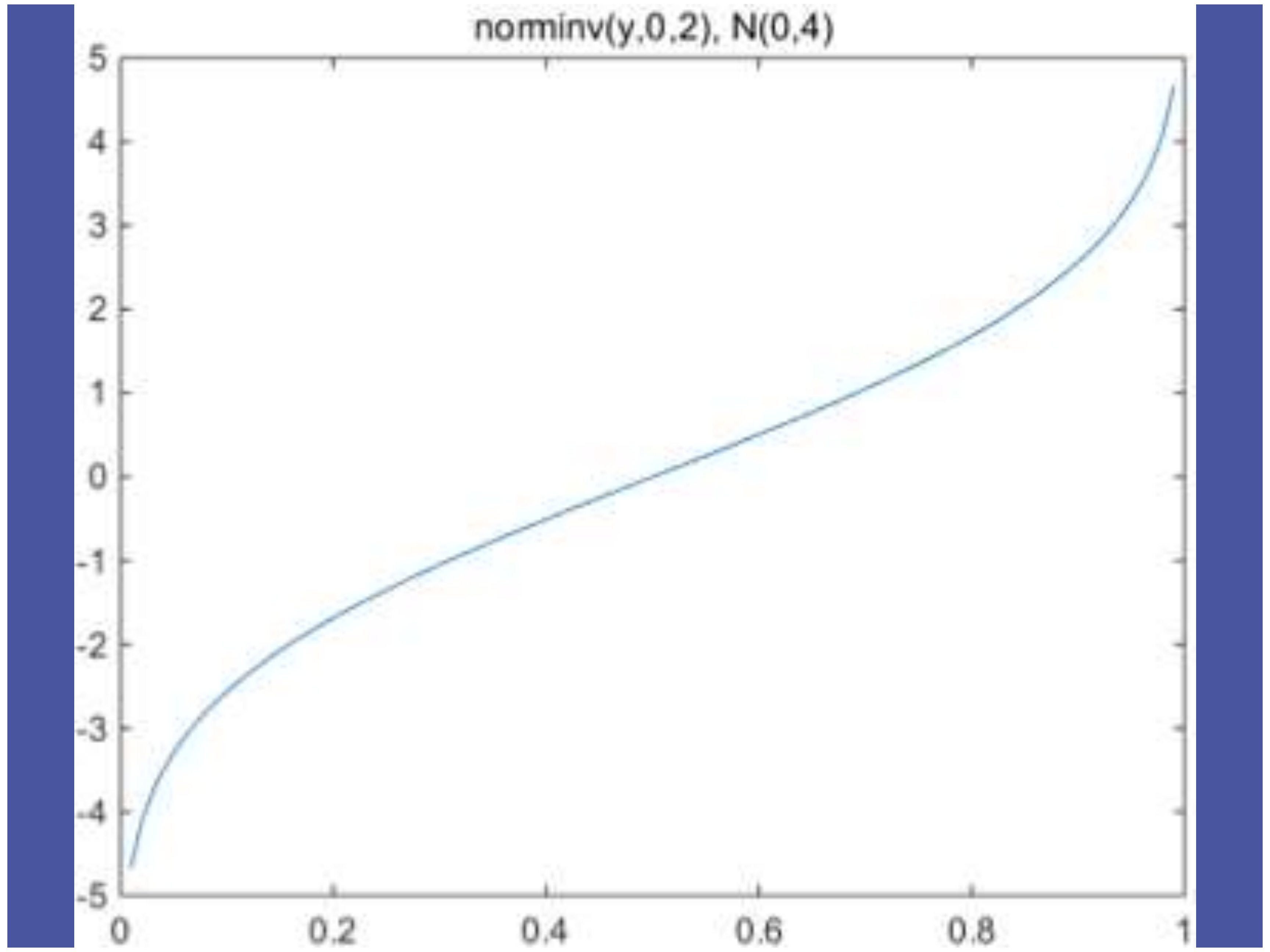
三　逆累积分布函数icdf

调用格式：X=norminv(p,mu,sgm)

例1．设X～N(0,4)，若有分布函数$F(x)$=0.1、0.3、0.5、0.9，求对应的最小的x值.

输入命令：

```
y=[0.1 0.3 0.5 0.9];
norminv(y,0,2)
```



得到对应的x值分别是：

-2.5631, -1.0488, 0, 2.5631

三　逆累积分布函数icdf

例2. 计算标准正态分布$N(0,1)$概率值0.1, 0.3, 0.5, 0.7, 0.9, 所对应的x的值.　　X=norminv(p,mu,sgm)

命令：

```
y=[0.1,0.3,0.5,0.7,0.9];
x=norminv(y,0,1)
```

结果：

x=-1.2816　-0.5244

0　0.5244　1.2816

检验：y1=normcdf(x,0,1);

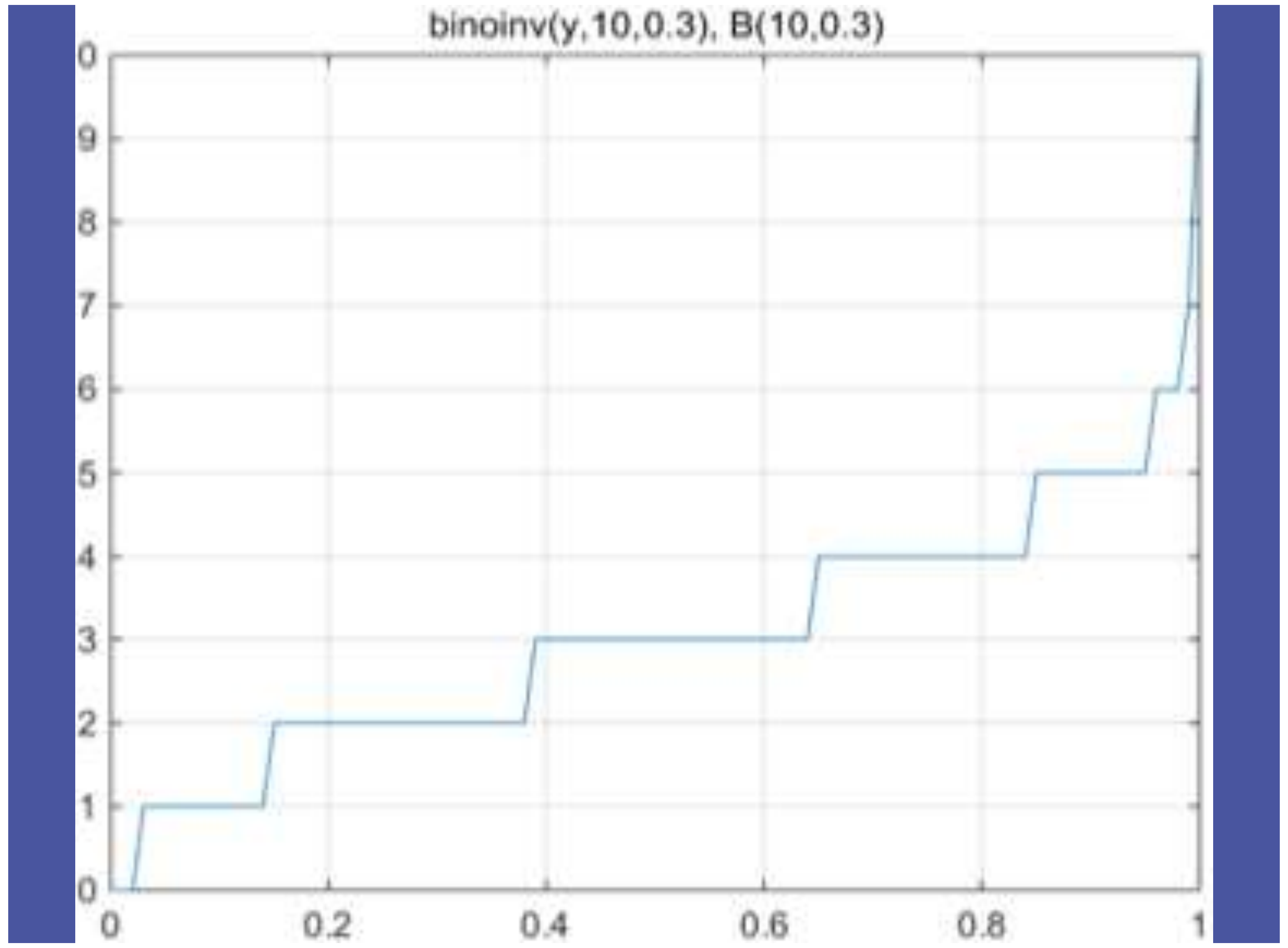
y1=0.1000　0.3000　0.5000　0.7000　0.9000

三 逆累积分布函数icdf

例3. 设X～B(10,0.3)，分布函数$F(x)$=0.1、0.2、0.3、0.9，求对应的最小的x值. **binoinv(y,n,p)**

输入命令：

```
y=[0.1 0.2 0.3 0.9];
binoinv(y,10,0.3)
```



得到对应的x值分别是：1，2，2，5.

三　逆累积分布函数icdf

例4 X~B(10,0.5)，求F(x)=0.1，0.3，0.5，0.7，0.9对应的x值

```
p=[0.1 0.3 0.5 0.7 0.9];
x=binoinv(p,10,0.5)
```

结果：x=3　4　5　6　7

检验：`y1=binocdf(x,10,0.5);`

结果：y1=0.1719　0.3770　0.6230　0.8281　0.9453

上例中，对p=0.1，对应的$cdf(x)\geq 0.1$的第一个值为3.

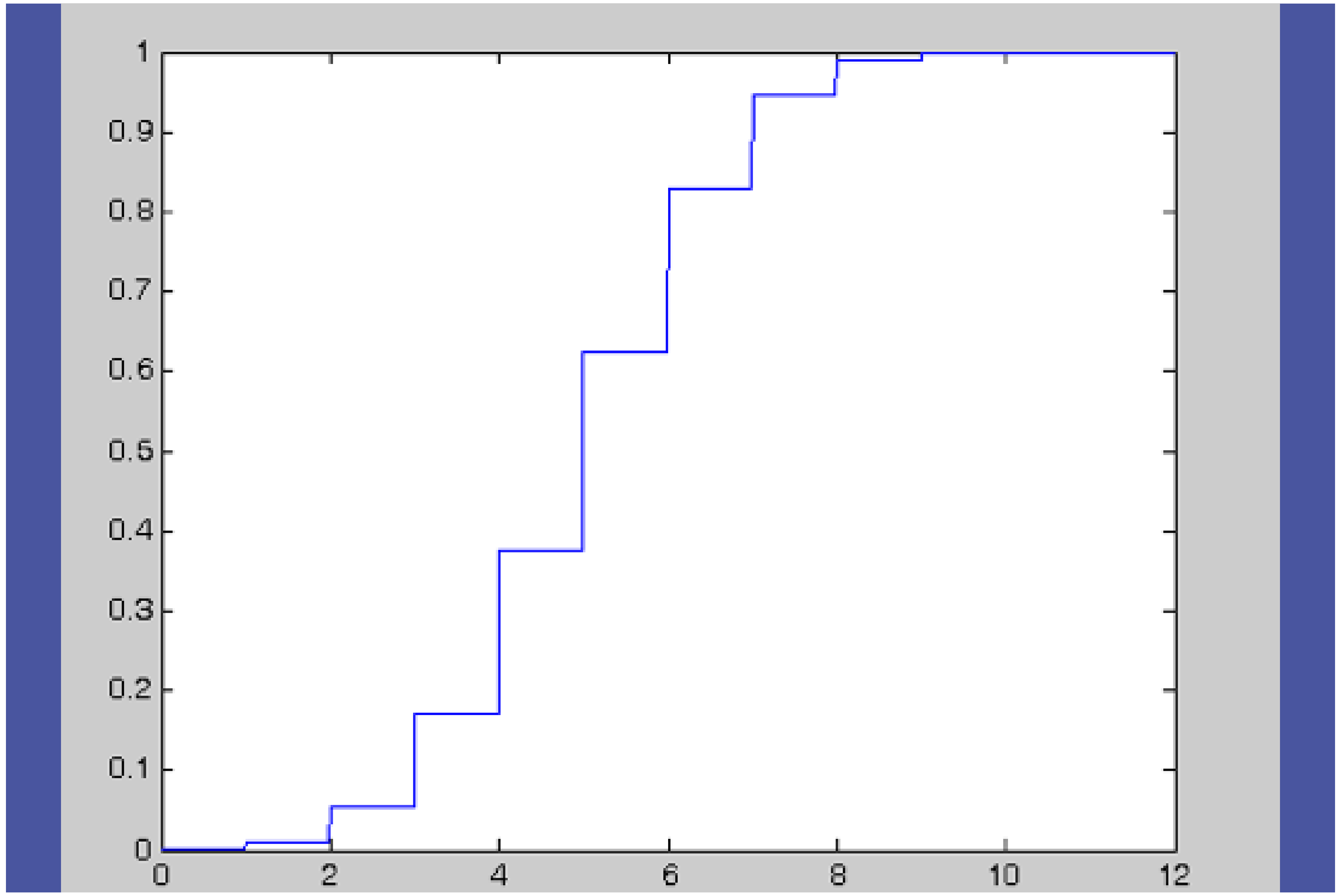
三　逆累积分布函数icdf

$B(10,0.5)$的分布函数图像

p=[0.1 0.3 0.5 0.7 0.9];

x=binoinv(p,10,0.5)

x=3　4　5　6　7



三　逆累积分布函数icdf

例5. 一家机电厂准备批量采购发动机，要求无故障运行时间不低于2000的概率达到90%. 某款发动机的无故障运行时间X服从参数为20000的指数分布，即X～E(20000)，该款发动机是否符合机电厂的要求？

需要检验的是：$P(X\geq 2000)=0.9$　　设X的分布函数为$F(X)$.

需要验证满足$P(X<x)=F(x)=0.1$的x，是否满足要求$x\geq 2000$.

三 逆累积分布函数icdf

X~E(20000)，求$P(X<x)=F(x)=0.1$的x，验证是否满足$x\geq 2000$.

输入命令：>>expinv(0.1,20000)

输出结果为2.1072e+003，即2107.2.

因此该款发动机完全满足机电厂的要求.

作 业

1. 在同一坐标下，画下列正态分布的密度函数图像

   $\mu=5$，$\sigma=0.05$，0.75，1.5，2.5, 4

2. 设$Y\sim B(10,0.25)$，求Y分布函数的值并画出函数图像

作 业

3. 生成两个10-20之间的随机数μ_1,μ_2；两个1-5之间的随机数σ_1,σ_2.

（1）在同一坐标下画$N(\mu_1,\sigma_1^2)$、$N(\mu_2,\sigma_2^2)$的密度函数、分布函数图像.

（2）设$X\sim N(\mu_1,\sigma_1^2)$，$Y\sim N(\mu_2,\sigma_2^2)$，问$P(X>5)$与$P(Y<4)$哪个更大？

（3）$P(X\leq x)=P(Y\leq y)=0.1$，对应的x，y分别是多少？