实践四 正态分布的综合性实践

一、实践问题

1. 问题背景

正态分布是工业、农业等生产活动中最常用的概率分布，许多实际问题都服从或近似服从正态分布．正是由于正态分布的普遍性，在概率论与数理统计的理论研究和实际应用中正态分布都具有十分重要的地位和作用，读者应熟练掌握和运用正态分布的知识．

2. 实践目的与要求

(1) 熟悉数学软件 MATLAB 的有关正态分布的命令;

(2) 熟悉在一个坐标系中同时产生几个图形的命令;

(3) 熟悉矩阵元素画图、画点的命令;

(4) 熟悉对图形区域填色的命令;

(5) 该实践题目意在加深对正态概率密度和正态分布函数等知识的全面理解，掌握观察实践现象或处理数据方面的方法.

二、 实践操作过程

服从参数为μ，σ^2的正态分布的概率密度是

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

其中μ，σ^2 $(\sigma>0)$是常数，记为$X\sim N(\mu,\ \sigma^2)$．正态分布的分布函数不能由初等函数表示.

在 MATLAB 中，用 normpdf 计算正态分布的概率密度的值．其基本调用格式是:

• y=normpdf(x, mu, sigma).

用 normcdf 计算正态分布的分布函数的值，其基本调用格式是:

• y=normcdf(x, mu, sigma).

1. 正态分布的概率密度图像关于 x=μ 对称

正态分布的概率密度函数图像见图 4-1.

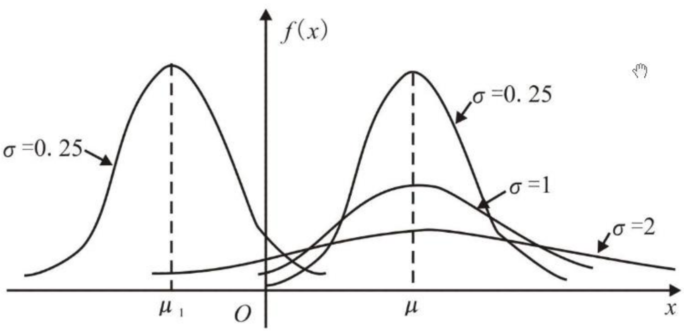

图 4-1 正态分布的概率密度图像

正态曲线关于 x=μ 对称.

例 4-1 验证正态分布概率密度图像(即正态曲线)关于 x=μ 对称.

解 在命令窗口中输入:

```
x=3;mu=2;sigma=5;
xl=normpdf(mu-x,mu,sigma), xr=normpdf(mu+x,mu,sigma)
```

回车后显示:

```
xl =
    0.0666
```

```
xr =
    0.0666
```

结果表明： 在 mu-x 与 mu+x 处的函数值是相等的. 也就是说，概率密度函数的图像关于 x=μ 对称.

也可以用分布函数来说明正态曲线的对称性. 正态分布的概率密度的图像关于 x=μ 对称，则 P{X≤μ-x}=P{X≥μ+x}，即 F(μ-x)=1-F(μ+x).

例 4-2 用分布函数性质验证正态分布概率密度的图像(即正态曲线)关于 x=μ 对称.

解 在命令窗口中输入:

```
x=3; mu=2; sigma=5;
pl=normcdf(mu-x,mu,sigma), pr=1-normcdf(mu+x,mu,sigma)
```

回车后显示:

```
pl =
    0.2743
pr =
    0.2743
```

从运行的结果可以看出，确实有 F(μ-x)=1-F(μ+x).

2. 正态分布概率密度在 x=μ 处取最大值

基本数学原理: 概率密度函数在 x=μ 处取最大值 $f(\mu)=\frac{1}{\sqrt{2\pi}\sigma}$.

例 4-3 验证正态分布概率密度在 x=μ 处取得最大值.

解 在命令窗口中输入:

```
mu=2;sigma=5;
P=normpdf(mu+(-1:0.5:1)*mu,mu,sigma), f=1/sqrt(2*pi)/sigma
```

回车后显示:

```
P =
    0.0737 0.0782  0.0798  0.0782  0.0737
f =
    0.0798
```

结果表明： 向量 P 的值是概率密度在 0, 0.5mu, mu, 1.5mu, 2mu 处的值. 从 运行的结果可以看出，在 x=μ 处的取值最大且正好是 $f(\mu)=\frac{1}{\sqrt{2\pi}\sigma}$.

3. 正态分布概率密度图像在 x=μ±σ 处出现拐点，且以 x 轴为渐近线

基本数学原理: 概率密度函数的图像在 x=μ±σ 处出现拐点，且以 x 轴为渐 近线，我们通过绘图来观察.

例 4-4 验证正态分布概率密度的图像(即正态曲线) 在 x=μ±σ 处出现拐点，且以 x 轴为渐近线.

解 在命令窗口中输入:

```
mu=2;sigma=5;
x=mu+sigma*(-4:0.1:4);x1=mu+[-1,1]*sigma;
y=normpdf(x,mu,sigma);y1=normpdf(x1,mu,sigma);
plot(x,y,x1,y1,'*')  %plot(x1,y1,'*')表示在点(x1,y1)处用“*”画点.
```

运行后生成下面的图像(见图 4-2).

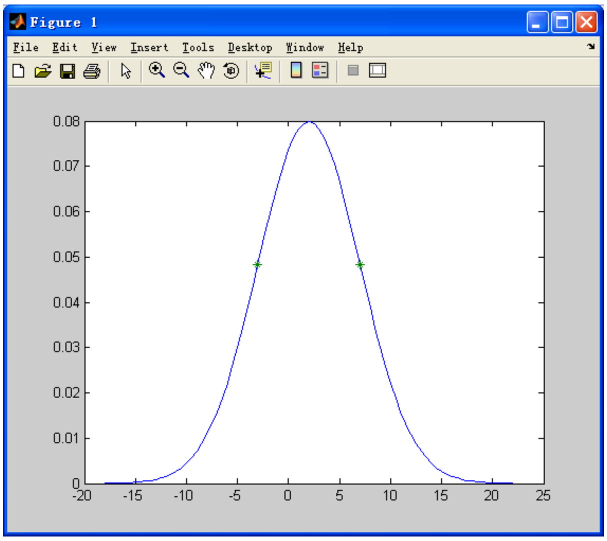

图 4-2 正态分布的概率密度图像及拐点

从图像中可以看出，正态分布概率密度以 x 轴为渐近线，其拐点标在星号处.

4. 固定 σ，改变 μ 时概率密度的图像沿 x 轴方向平移

基本数学原理：固定 σ，改变 μ 的值，概率密度的图像保持形状，沿 x 轴平移. μ 的值变大，则图像向右平移；μ 的值变小，则图像向左平移.

例 4-5 验证当固定 σ，改变 μ 的值时，正态分布概率密度的图像(即正态曲线) 保持形状，但是沿 x 轴平移. 取 σ =5, μ =-5, 0, 3, 10.

解 在命令窗口中输入：

```
mu=[-5,0,3,10];sigma=5;x0=sigma*(-4:0.1:4);
x1=mu(1)+x0;y1=normpdf(x1,mu(1),sigma);
x2=mu(2)+x0;y2=normpdf(x2,mu(2),sigma);
x3=mu(3)+x0;y3=normpdf(x3,mu(3),sigma);
x4=mu(4)+x0;y4=normpdf(x4,mu(4),sigma);
plot(x1,y1,':r',x2,y2,'--b',x3,y3,'-.k',x4,y4,'-c')          %
plot(x1,y1,':r')表示在点 (x1,y1)处用“:”画点, 填红色,其余类推. 其中: r-红色,b-蓝色,k-黑色,c-天蓝色.
```

运行后生成下面的图像(见图 4-3).

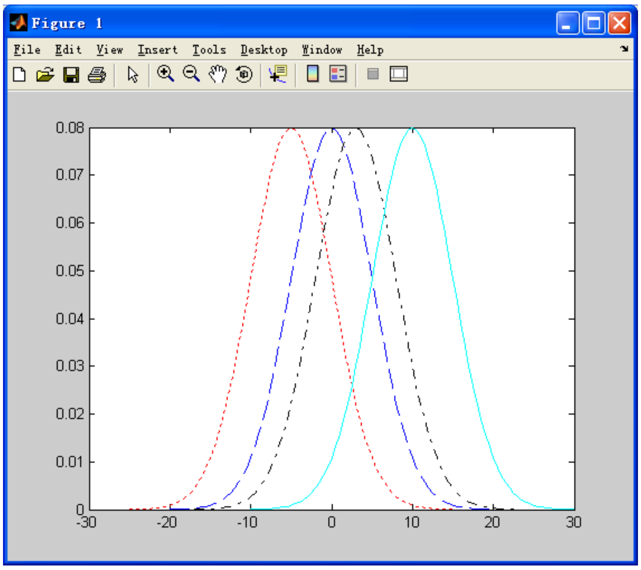

图 4-3 固定σ,改变μ时概率密度的图像沿 x 轴方向平移

这是固定σ=5, μ的值分别为-5,0,3,10 生成的正态分布的概率密度图像.

5. 固定 μ,改变 σ时正态分布概率密度的图像陡峭程度改变

基本数学原理: 固定μ, 改变σ, 概率密度的图像对称轴保持不变, 图像的陡峭程度发生了改变. σ的值变大, 则图像变得平缓; σ的值变小, 则图像变得陡峭.

例 4-6 验证当固定μ, 改变σ的值时, 正态分布概率密度的图像(即正态曲线)陡峭程度发生改变. 取 μ=10, σ=2,3,5,9.

解 在命令窗口中输入:

```
mu=10; sigma=[2,3,5,9]; x0=(-3:0.1:3);
x1=mu+sigma(1)*x0;  y1=normpdf(x1,mu,sigma(1));
x2=mu+sigma(2)*x0;  y2=normpdf(x2,mu,sigma(2));
x3=mu+sigma(3)*x0;  y3=normpdf(x3,mu,sigma(3));
x4=mu+sigma(4)*x0;  y4=normpdf(x4,mu,sigma(4));
plot(x1,y1,':r',x2,y2,'--b',x3,y3,'-.k',x4,y4,'-c')   %见例 4-5
解释.
```

运行后生成下面的图像(见图 4-4):

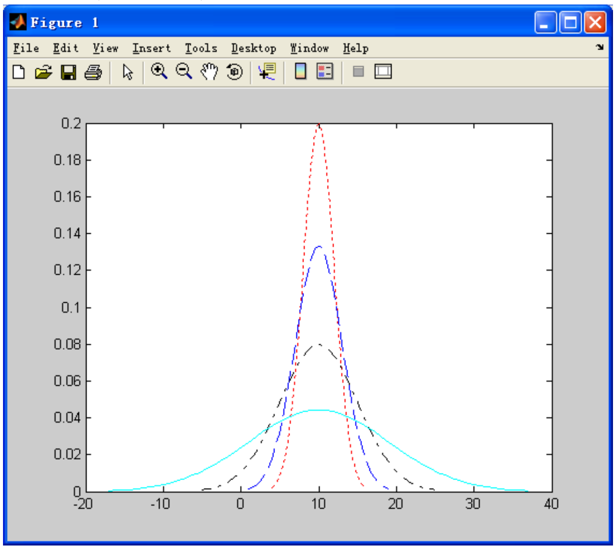

图 4-4 固定μ,改变 σ时概率密度图像变化

这是固定 μ=5, σ的值分别为 2,3,5,9 生成的正态分布概率密度图像.

6. 正态分布的分布函数图像

基本数学原理: 设$X \sim N(\mu,\sigma^2)$, 则随机变量 X 的分布函数是

$$F(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx,\quad -\infty<x<\infty$$

由于正态分布的概率密度的图像关于 x=μ对称, 所以当 x=μ时, 分布函数的值是 0.5, 也就是说正态分布的分布函数的图像过(μ,0.5)点.

例 4-7 画出μ=10, σ=5 的正态分布的分布函数的图像.

解在命令窗口中输入:

```
mu=10; sigma=5; x0=(-3:0.1:3);
x=mu+sigma*x0;  y=normcdf(x,mu,sigma);
plot(x,y,mu,0.5,'*',[mu, mu, min(x)],[0,0.5,0.5],':k')
```

在点(mu,0.5)画点“*” ; [mu, mu, min(x),[0,0.5,0.5]矩阵对应元素构成坐标, 这里

三点坐标分别是(mu,0), (mu,0.5), (min(x),0.5); plot([mu, mu, min(x)],[0,0.5,0.5],':k')表示在三点之间依次以“:”画直线.

运行后生成下面的图像(见图 4-5):

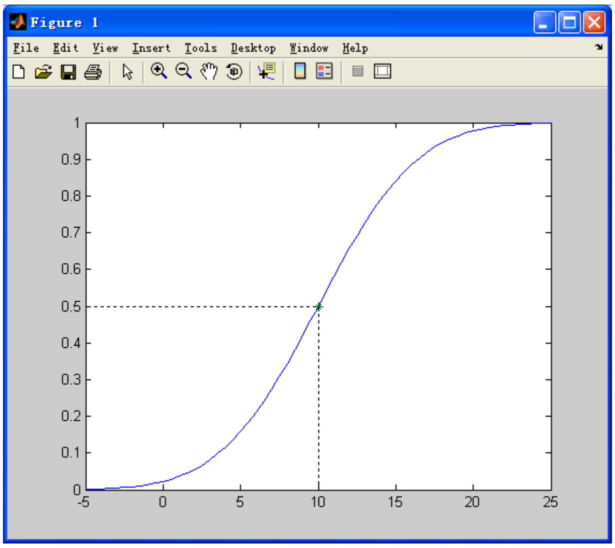

图 4-6 正态分布的分布函数图像

从图像还可以看出，正态分布的分布函数图像关于(μ,0.5)点对称.

7. **固定σ，改变μ时正态分布的分布函数的图像沿 x 轴平移**

基本数学原理：固定σ，改变μ的值，正态分布的分布函数图像保持形状不变，沿 x 轴平移．μ的值变大，则图像向右平移；μ的值变小，则图像向左平移.

例 4-8 验证当固定σ，改变μ的值时，正态分布的分布函数的图像保持形状，但是沿 x 轴平移．取σ=5, μ=-5, 0,3, 10.

解 在命令窗口中输入:

```
mu=[-5,0,3,10]; sigma=5; x0=sigma*(-4:0.1:4);
x1=mu(1)+x0;  y1=normcdf(x1, mu(1), sigma);
x2=mu(2)+x0;  y2=normcdf(x2, mu(2), sigma);
x3=mu(3)+x0;  y3=normcdf(x3, mu(3), sigma);
x4=mu(4)+x0;  y4=normcdf(x4, mu(4), sigma);
plot(x1, y1, ':r', x2, y2, '--b', x3, y3, '-.k', x4, y4, '-c')   %
见例 4-5 解释.
```

运行后生成下面的图像(见图 4-6).

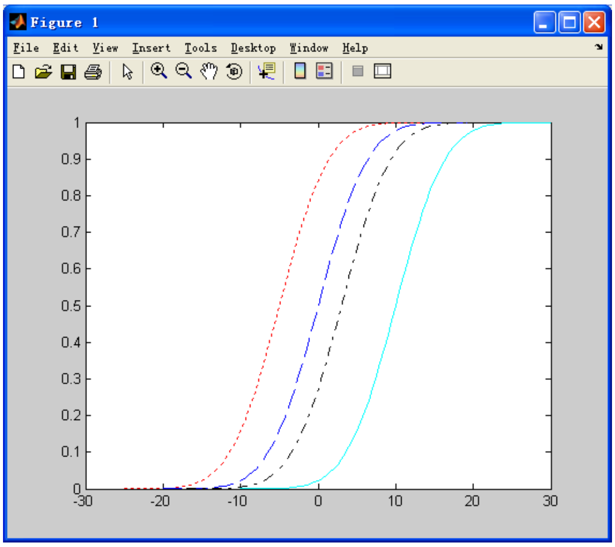

图 4-6 固定σ，改变μ时分布函数的图像沿 x 轴平移

8. 固定μ，改变σ时分布函数的图像陡峭程度改变

基本数学原理：固定μ，改变σ，正态分布的分布函数的图像，都过点(μ，0.5)，图像的陡峭程度发生了改变．σ的值变大，则图像变得平缓；σ的值变小，则图像变得陡峭．

例 4-9 验证当固定μ，改变σ的值时，正态分布的分布函数的图像陡峭程度发生改变．取μ=10, σ=2, 3, 5, 9.

解 在命令窗口中输入:

```
mu=10;sigma=[2,3,5,9];x0=(-3:0.1:3);
x1=mu+sigma(1)*x0;  y1=normcdf(x1,mu,sigma(1));
x2=mu+sigma(2)*x0;  y2=normcdf(x2,mu,sigma(2));
x3=mu+sigma(3)*x0;  y3=normcdf(x3,mu,sigma(3));
x4=mu+sigma(4)*x0;  y4=normcdf(x4,mu,sigma(4));
plot(x1,y1,':r',x2,y2,'--b',x3,y3,'-.k',x4,y4,'-c')   % 见例 4-5
解释.
```

运行后生成下面的图像(见图 4-7):

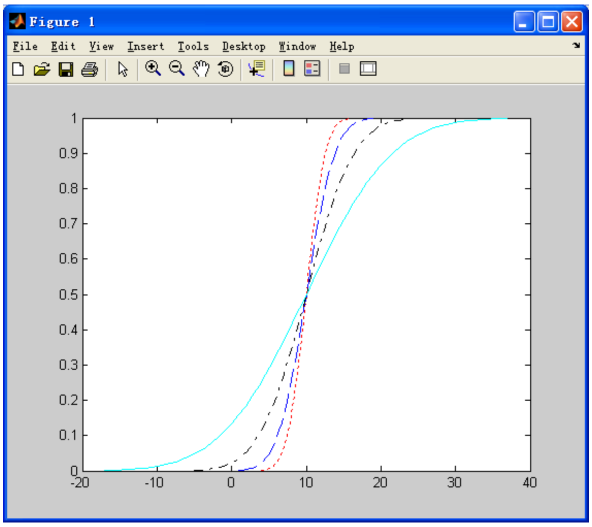

图 4-7 固定μ,改变σ时分布函数的图像变化

9. 一般正态分布与标准正态分布的关系

基本数学原理:标准正态分布 N(0, 1)的概率密度和分布函数常用φ(x)和Φ(x)表示.

$$\phi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}},\quad -\infty<x<\infty$$

$$\Phi(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx,\quad -\infty<x<\infty$$

关于φ(x)的图像和Φ(x)的意义见图 4-8.

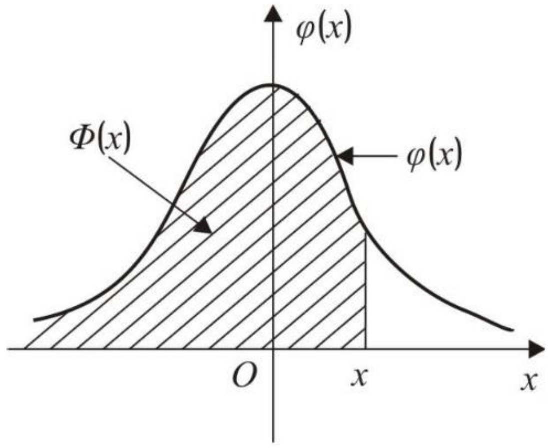

图 4-8 标准正态分布概率密度φ(x)和分布函数Φ(x)关系

标准正态分布 N(0, 1)是最重要的一种正态分布. 对于一般正态分布 N(μ, σ^2), 都可以通过作标准化变换$Z=\frac{X-\mu}{\sigma}$化为标准正态分布.

例 4-10 正态分布的概率与标准正态分布的概率计算比较.

解 在命令窗口中输入:

```
mu=2;  sigma=7;  x=4;
p1=normcdf(4, mu, sigma)      %对N(2,7)计算概率F(4)=P{X≤4}.
p2=normcdf((x-mu)/sigma,0,1)   % 对 N(0,1) 计算概率 Φ((4-mu)/sigma)=P{Z≤(4-mu)/sigma}.
```

回车后显示:

```
p1 =
    0.6125
p2 =
    0.6125
```

结果表明：直接用 N(μ,σ^2)的分布函数计算的概率 F(4)=P{X≤4}与标准化变换后用标准正态分布计算的概率Φ((4-mu)/sigma)=P{Z≤(4-mu)/sigma}相同.

10. 正态分布的 3σ准则

基本数学原理:设 X～N(μ,σ^2), 我们得到下列结论:

P{μ-σ≤X≤μ+σ}≈0.6827;

P{μ-2σ≤X≤μ+2σ}≈0.9545;

P{μ-3σ≤X≤μ+3σ}≈0.9973.

可见, 尽管正态分布的取值范围是(-∞,+∞), 但它的值落在(μ-3σ, μ+3σ)内几乎是肯定的事, 发生的概率为 99.7%. 这就是“3σ准则”.

例 4-11 验证正态分布的“3σ准则”.

解 在命令窗口中输入:

```
mu=10; sigma=5;
p1=normcdf(mu+sigma,mu,sigma)-normcdf(mu-sigma,mu,sigma)
p2=normcdf(mu+2*sigma,mu,sigma)-normcdf(mu-2*sigma,mu,sigma)
p3=normcdf(mu+3*sigma,mu,sigma)-normcdf(mu-3*sigma,mu,sigma)
```

运行后在命令窗口中显示:

```
p1 =
    0.6827
p2 =
    0.9545
p3 =
    0.9973
```

11. 正态分布的随机数

在 MATLAB 中, 用函数 randn 产生正态分布的随机数. 函数 randn 的调用格式是:

• y=randn(m, n)　　%生成 m 行 n 列的正态分布的随机数矩阵.

例 4-12 生成 100 个服从标准正态分布的随机数, 并画出直方图.

解 在命令窗口中输入:

```
y=randn(100, 1); hist(y)
```

运行后生成下面的图像(见图 4-9):

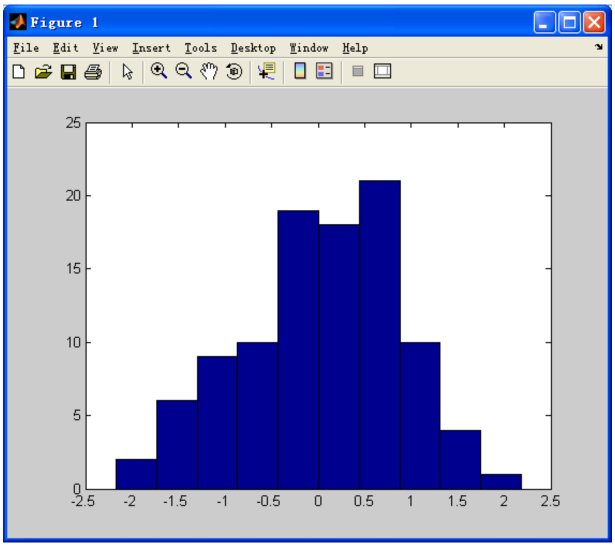

图 4-9 标准正态分布的随机数的直方图

由 randn 生成的随机数服从标准正态分布, 如果想要生成一般正态分布的随机数, 需要作变换$Y=\mu+\sigma X$. 也可以用 R = normrnd(MU, SIGMA, m, n)生成, 见实践三.

例 4-13 生成 10000 个服从 N(6, 1)的正态分布随机数, 并以 0.5 为组距绘出直方图.

解 在命令窗口中输入:

```
x=randn(10000, 1);  % 生成标准正态分布的随机数.
mu=6; sigma=1;
y=mu+sigma*x;    %生成正态分布 N(6, 1)的随机数.
hist(y, (mu-3*sigma):0.5:(mu+3*sigma))
```

运行后生成下面的图像(见图 4-10):

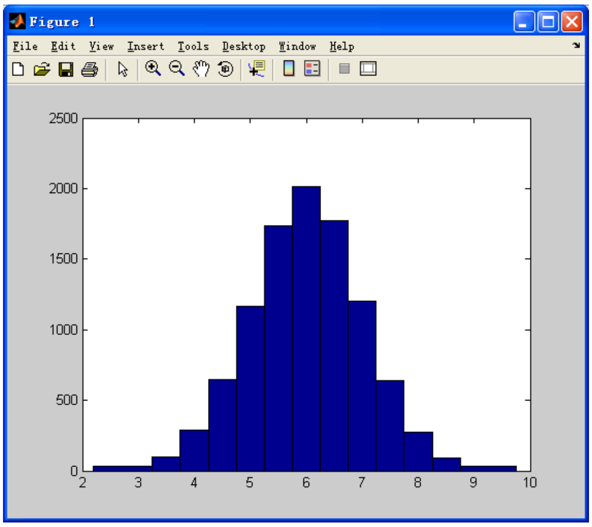

图 4-10 组距为 0.5 的正态分布 N(6,1)的随机数直方图

例 4-14 生成 10000 个服从 N(6, 1)的正态分布随机数，并以 0.5 为组距绘出累积百分比曲线图.

解 在命令窗口中输入:

```
x=randn(10000, 1); mu=6; sigma=1;
y=mu+sigma*x;
intval=(mu-3*sigma):0.5:(mu+3*sigma);    % 以组距 0.5 划分区间((mu-3*sigma), (mu+3*sigma)).
N=histc(y, intval);  %计算 y 值位于每个小区间上的个数.
X=cumsum(N)/sum(N);  % cumsum(N)由上到下顺序逐次计算一列矩阵元素的各个和, 显然, cumsum(N)的第一个元素就是 N 的第一个元素, cumsum(N)的最后一个元素等于 sum(N); sum(N)表示对 N 的所有元素求和; X 为累积百分比.
plot(intval, X)
```

运行后生成下面的图像(见图 4-11):

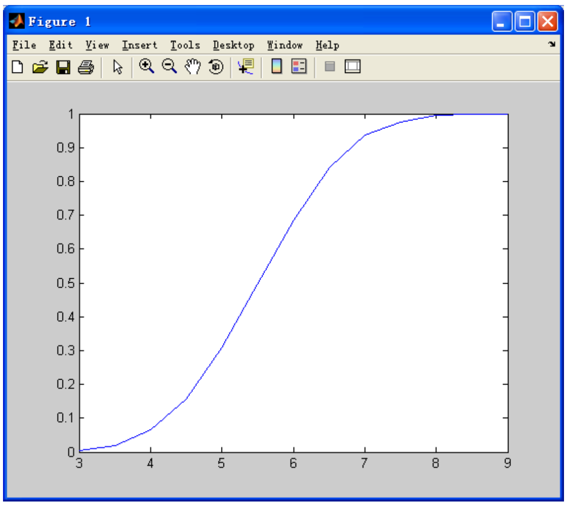

图 4-11 正态分布 N(6,1)的随机数累积百分比图

12. **标准正态分布的上α分位点**

基本数学原理: 对标准正态分布 X~N(0,1), 若 z_α满足 $P\{X>z_\alpha\}=\alpha$, 则称点 z_α为标准正态分布的上α分位点.

在 MATLAB 中, 我们可以通过用逆累积分布函数来求上α分位点. 正态分布的逆累积

分布函数是 norminv，其基本调用格式是：

- x=norminv(p, mu, sigma)　　% 给定 alpha，则分位点是 x=norminv(1-alpha, 0, 1).

例 4-15 对于 α =0.1，画出标准正态分布概率密度的图像(即正态曲线)，并标出 α =0.1 时的概率几何表示.

解 在命令窗口中输入：

```
alpha=0.1;
xmax=3.5;
ymax=0.5;
x=linspace(-xmax,xmax,10000);  %等分区间(-xmax,xmax)，产生 10000 个划分点.
y=normpdf(x,0,1);
plot(x,y,'k-',[0,0],[0,ymax],':k')   % plot([0,0],[0,ymax],':k') 在点(0,0)和 (0,ymax)之间以“:”画线，即画对称轴.
xa=norminv(1-alpha,0,1);   %求上 alpha 分位点,注意这里用 1-alpha 表示已知概率.
xx=linspace(xa,xmax,1000); %等分区间(上分位点，xmax)，产生 1000 个划分点,为填色用.
yy=normpdf(xx,0,1);
hold on
fill([xx(1),xx,xx(end)],[0,yy,0],'c');   % 在 点 (xx(1),0),( xx, yy),( xx(end),0)之间填青色.
axis([-xmax,xmax,0,ymax])  % 确定坐标轴的变化范围.
hold off
```

运行后生成下面的图像(见图 4-12)：

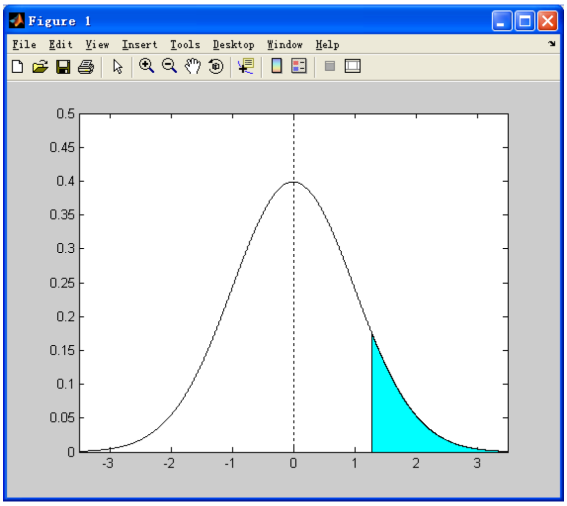

图 4-12 上 α 分位点图

由标准正态分布的概率密度的图像可知，标准正态分布的上 α 分位点 z_α 满足：$z_{1-\alpha}=-z_\alpha$.

三、实践结论与总结

在 MATLAB 中，用 normpdf 计算正态分布的概率密度的值. 用 normcdf 计算正态分布的分布函数的值. 对于一般正态分布 $N(\mu,\sigma^2)$，都可以通过作标准化变换 $Z=\frac{X-\mu}{\sigma}$ 化为标准正态分布. 通常是先求得服从标准正态分布的 X 值，然后通过变换 $Y=\mu+\sigma X$ 得到服从正态分布 $N(\mu,\sigma^2)$ 的 Y 值. 正态分布的值落在(μ -3 σ， μ +3 σ)内的概率大约是 0.9973 左右，几乎是肯定的事件，“3 σ 准则”.