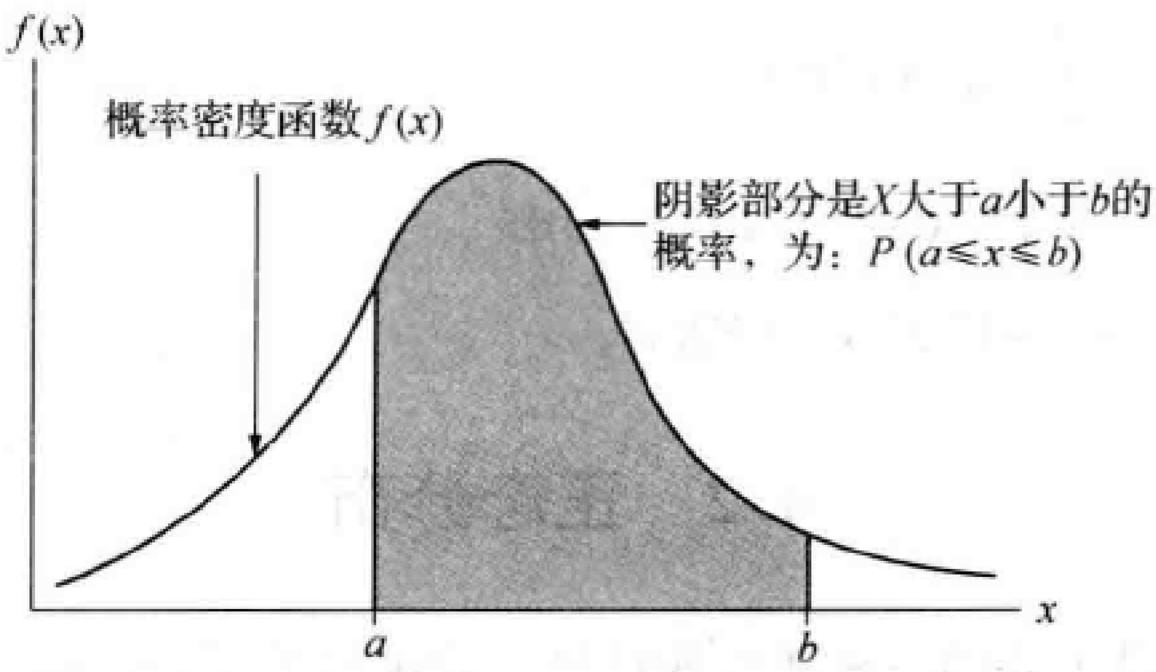

图 6－2　概率密度函数 $f(x)$

[**注意**] a. 连续随机变量取某个特定值的概率等于0，因为取某个特定值后，直方图中对应的矩形条退化成一条直线。所以我们一般考虑连续随机变量在一个区间内的概率。

概率密度函数 $f(x)$ 必须满足下列两个条件：

(1) 对于任何 x 的取值，$f(x)\geqslant 0$；

(2) 曲线 $f(x)$ 下面的总面积等于，$\int_{-\infty}^{+\infty} f(x)\mathrm{d}x = 1$。

b. $f(x)$ 不是概率函数。

3. 分布函数

连续型随机变量的概率也可以用分布函数 $F(x)$ 来表示。

分布函数(distribution function)定义为

$$F(x) = P(X \leqslant x) = \int_{-\infty}^{x} f(t)\mathrm{d}t$$

所以，根据分布函数，X 在区间 $[a,\ b]$ 上的概率为

$$P(a \leqslant x \leqslant b) = \int_{a}^{b} f(x)\mathrm{d}x = \int_{-\infty}^{b} f(x)\mathrm{d}x - \int_{-\infty}^{a} f(x)\mathrm{d}x = F(b) - F(a)$$

如图 6－3 所示，分布函数就是概率密度曲线下小于某个 x_0 的面积。

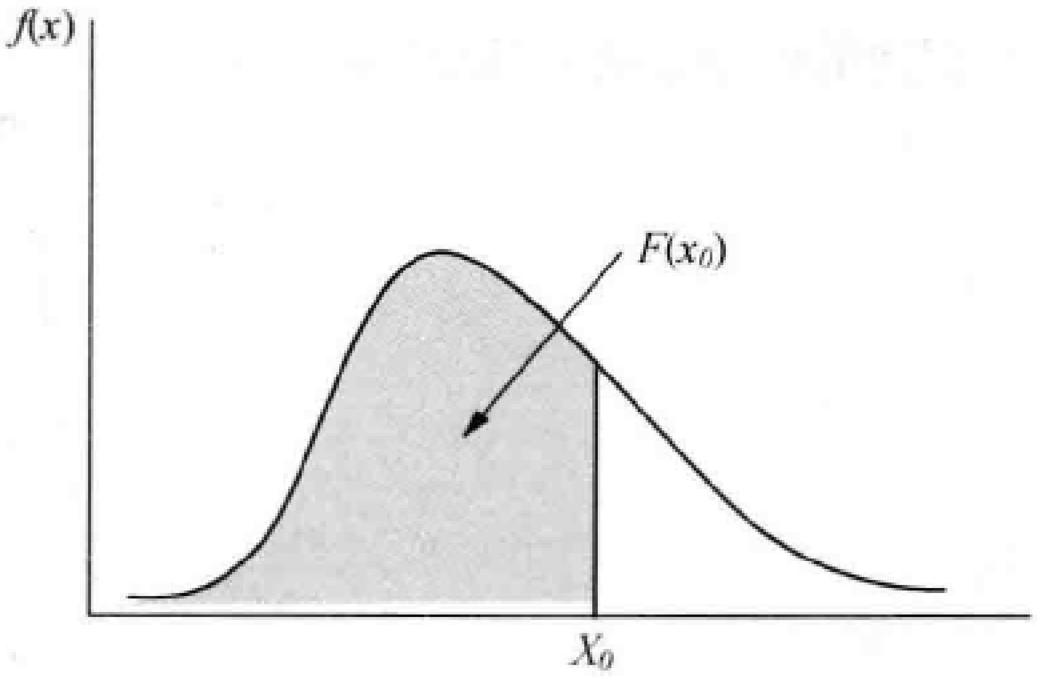

图 6－3　分布函数

4. 连续随机变量的期望和方差

在讨论离散随机变量的期望和方差时，我们使用连加符号的。连续随机变量必须使用积分符号。

(1) 期望：$E(X)=\int_{-\infty}^{+\infty}xf(x)\mathrm{d}x$；

(2) 方差：$\mathrm{var}(X)=\int_{-\infty}^{+\infty}(x-EX)^2f(x)\mathrm{d}x$

当然，同样有 $\mathrm{var}(X)=E(X^2)-(EX)^2$。

6.2 正态分布

连续随机变量中最重要的一种随机变量是正态随机变量，相应的概率分布称为正态分布(normal distribution)。它是一种对称钟形分布，如图 6－4 所示。

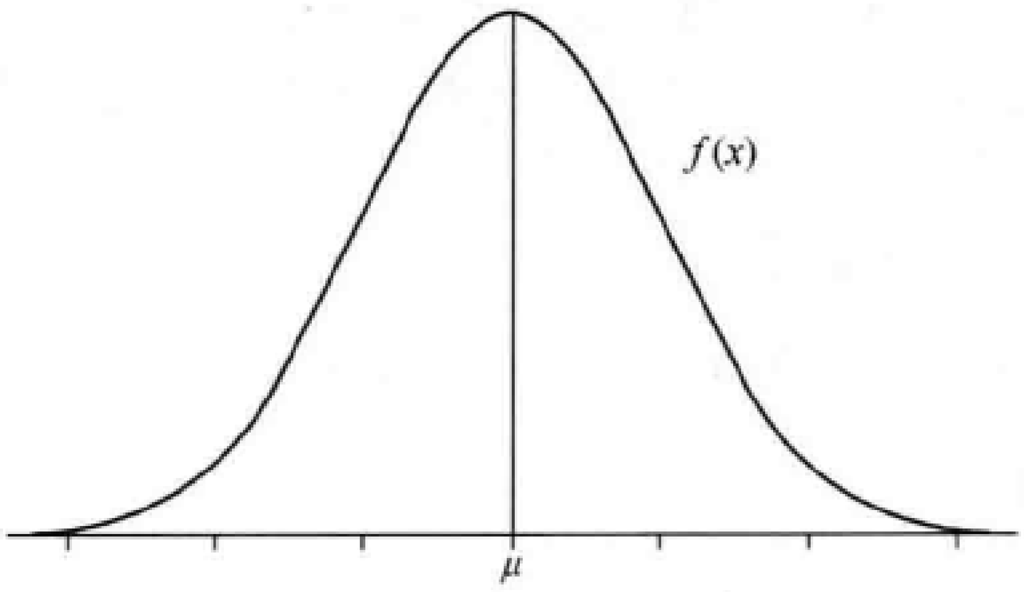

图 6－4　正态分布

正态分布又叫高斯分布(Gaussian distribution)，现实生活中近似服从正态分布的变量有很多，如测量误差、商品的重量或尺寸、某年龄人群的身高和体重等。

正态分布的概率密度函数为

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

随机变量 x 可以在实数轴上的任意地方取值。其中，μ 和 σ 是随机变量 x 所有可能观测值的总体的均值和标准差，是标识正态分布的两个参数。

(1) 均值不同、标准差相同的两条正态曲线(图 6－5)

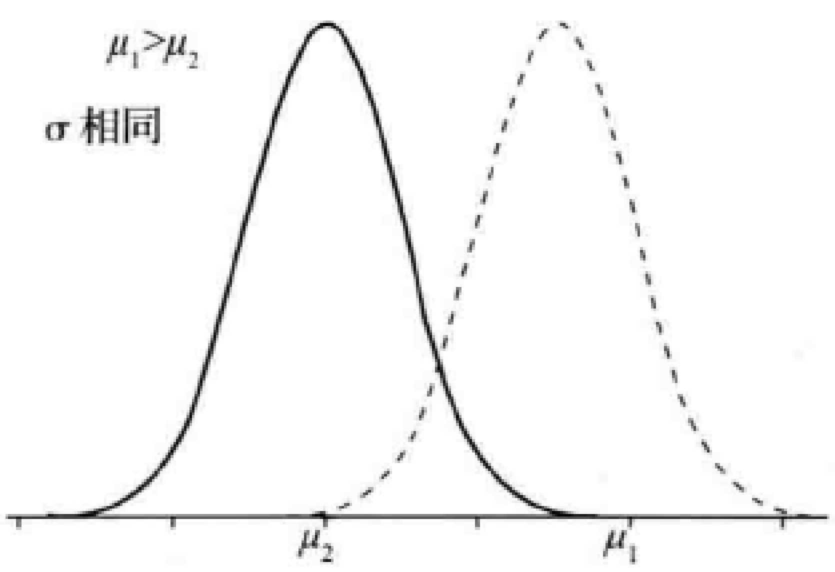

图 6－5　均值不同、标准差相同的两条正态曲线

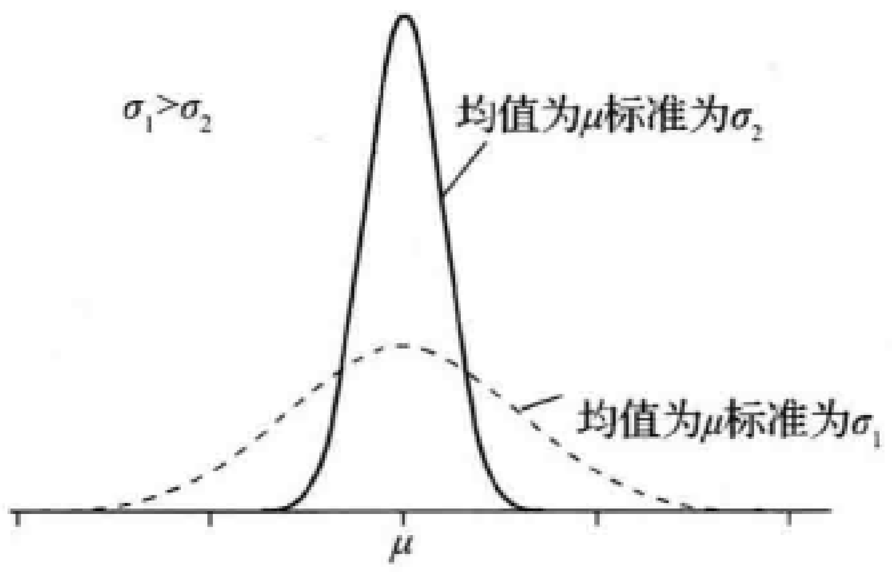

图 6－6　均值相同、标准差不同的两条正态曲线

(2) 均值相同、标准差不同的两条正态曲线(图 6－6)

一个正态分布用 $N(\mu,\sigma^2)$ 表示，其中 μ 为均值，σ^2 为方差；也常用 $N(\mu,\sigma)$ 来表示，这里 σ 为标准差。

1. 正态分布的重要特性

(1) 正态分布的密度曲线是一个关于均值对称的钟型曲线。正态分布也是一族分布，各种正态分布根据它们的均值和标准差不同而有区别；

(2) 正态分布最高点在均值处，均值=中位数=众数；

(3) 正态曲线的尾部向左右两边无限延伸，永远不会与水平轴相交，而是无限地接近水平轴。因为 $f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}>0$；

(4) 曲线的位置由 μ 决定，曲线的陡缓程度由 σ 决定。σ 越大，曲线越平缓；σ 越小，曲线越陡峭。

2. 经验法则在正态分布里的细化

同经验法则一样，这里讨论 $\mu\pm\sigma$、$\mu\pm2\sigma$ 和 $\mu\pm3\sigma$ 三种情况，如图 6-7 所示。

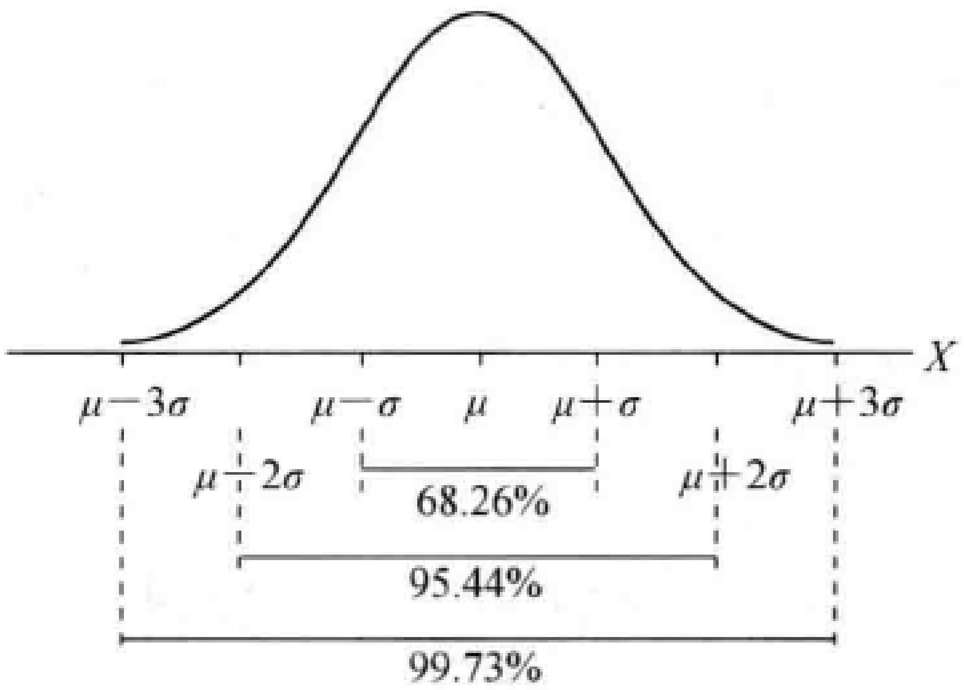

图 6-7 经验法则在正态分布里的细化

也就是说，正态分布在均值加减一个标准差的范围内包含 68.26%的数据，均值加减两个标准差的范围内包含 95.44%的数据，均值加减三个标准差的范围内包含 99.73%的数据。

6.3 标准正态分布

前面说过，正态分布由均值 μ 和标准差 σ 这两个参数决定。由于存在许多均值 μ 和标准差 σ 的组合(这种组合有无限多个)，我们希望能做出一张特殊的正态曲线面积的表格，它适用于所有正态曲线面积的计算。这样的正态曲线面积的表格是存在的(限于篇幅，我们把表格放在附录里)。

考虑一个均值为 μ、标准差为 σ 的正态分布的随机变量 x，联系前面第三章讲过的标准化分数，将变量 x 标准化得到：

$$z=\frac{x-\mu}{\sigma}$$

由于这个变换是线性变换，不改变分布的形状。所以变换后的变量 z 也服从正态分布，其均值为 0、标准差为 1。我们把 z 的分布称为标准正态分布(standard normal distribution)，如图 6-8 所示。

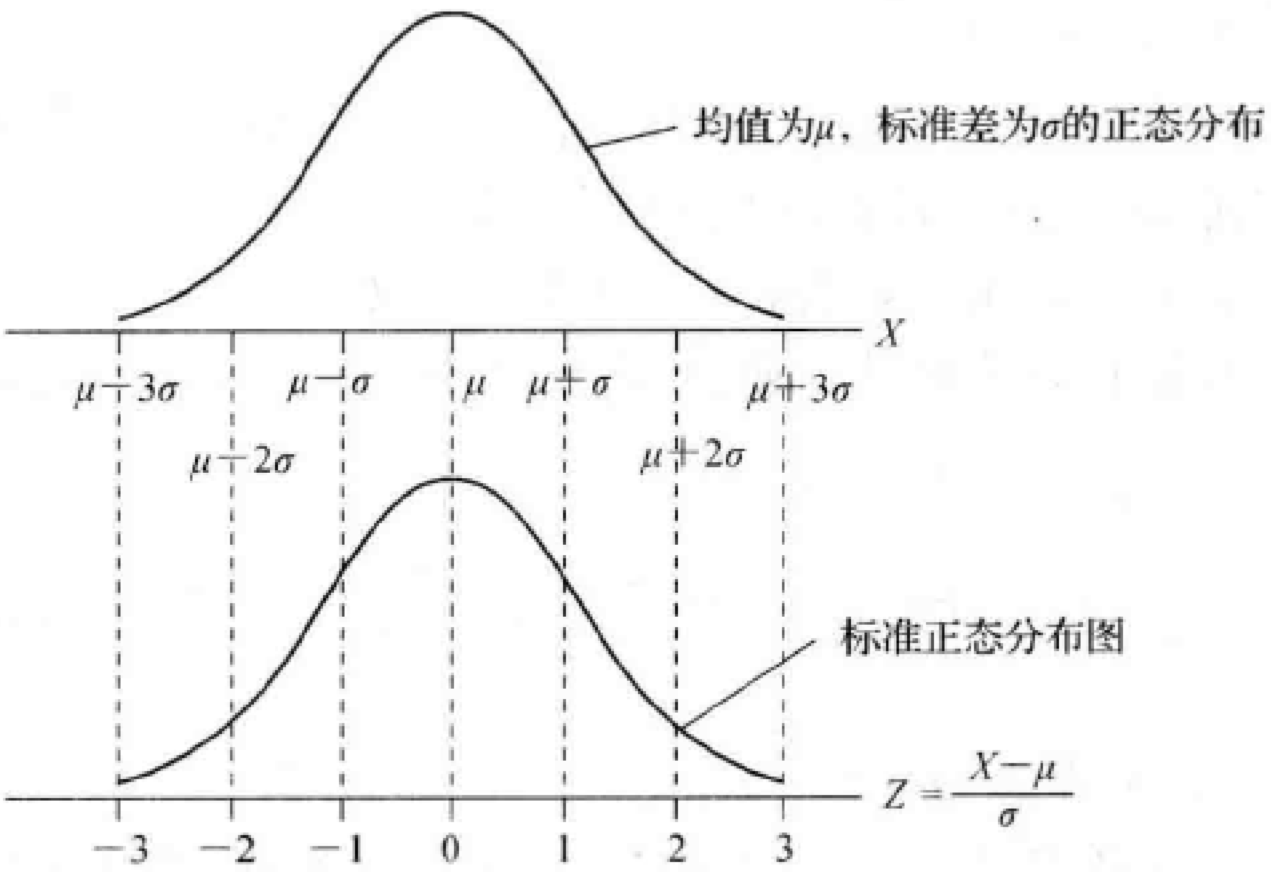

图 6-8　标准正态分布

由图 6-8 可以看出，$P(\mu-3\sigma\leqslant x\leqslant\mu)=P(-3\leqslant z\leqslant 0)$。一般有：

$$P(a\leqslant x\leqslant b)=P\left(\frac{a-\mu}{\sigma}\leqslant z\leqslant\frac{b-\mu}{\sigma}\right)$$

所以，所有均值为 μ、标准差为 σ 的正态分布都可以转化为均值为 0、标准差为 1 的标准正态分布。

6.4　正态分布的相关计算

6.4.1　常用的概率和标准化分数

1. 中间的概率为 0.9、0.95 和 0.99(图 6-9)

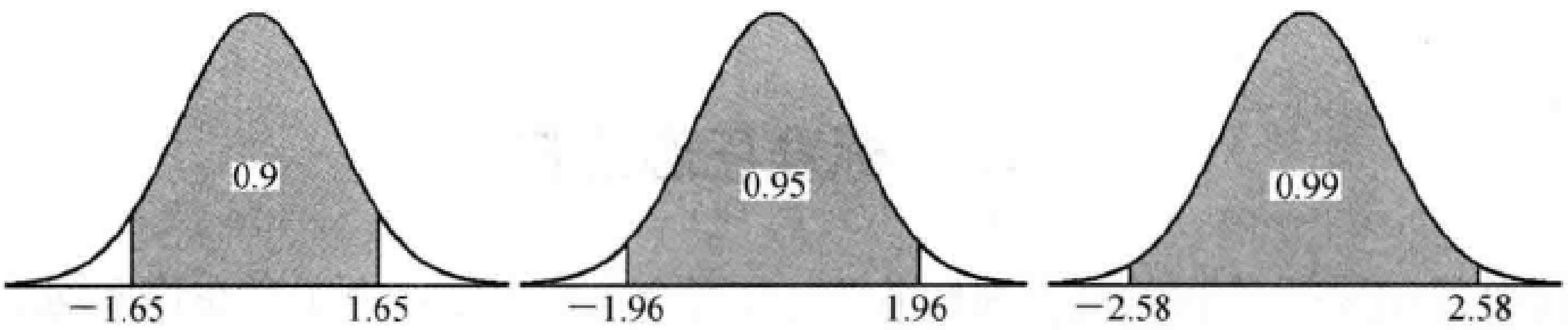

图 6-9

2. 左尾概率为 0.9、0.95 和 0.99(图 6-10)

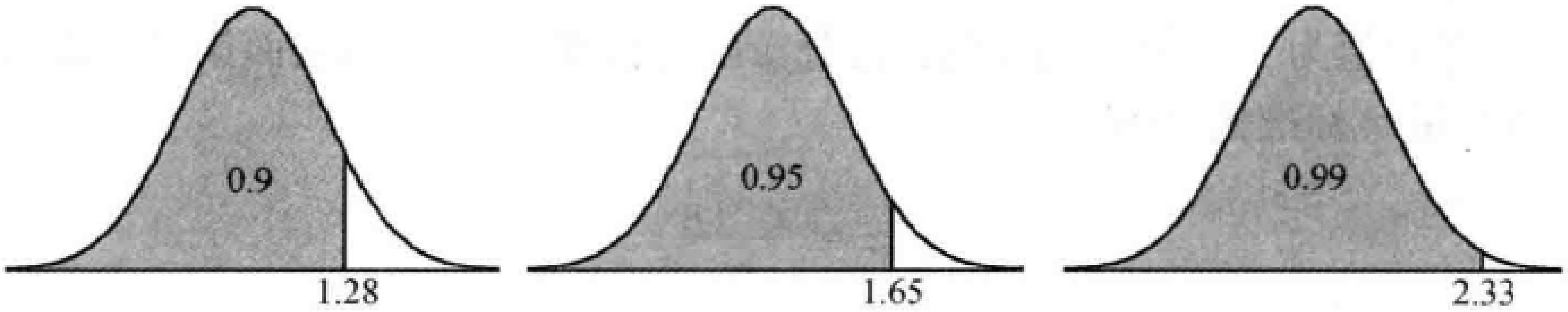

图 6-10

3. 右尾概率为 0.9、0.95 和 0.99(图 6 - 11)

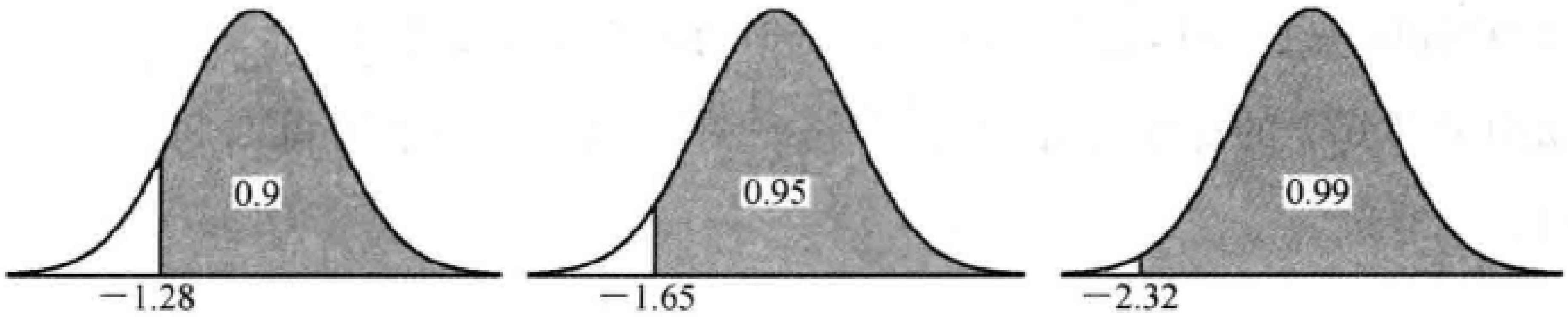

图 6 - 11

6.4.2 使用分布表计算

使用标准正态概率分布表可以计算相关数据,如概率值、变量值、均值和标准差。主要围绕标准化形式 $z=\frac{x-\mu}{\sigma}$ 计算。

1. 计算概率

例 6 - 1 某电器元件的使用寿命服从均值为 1 500 小时、标准差为 75 小时的正态分布。

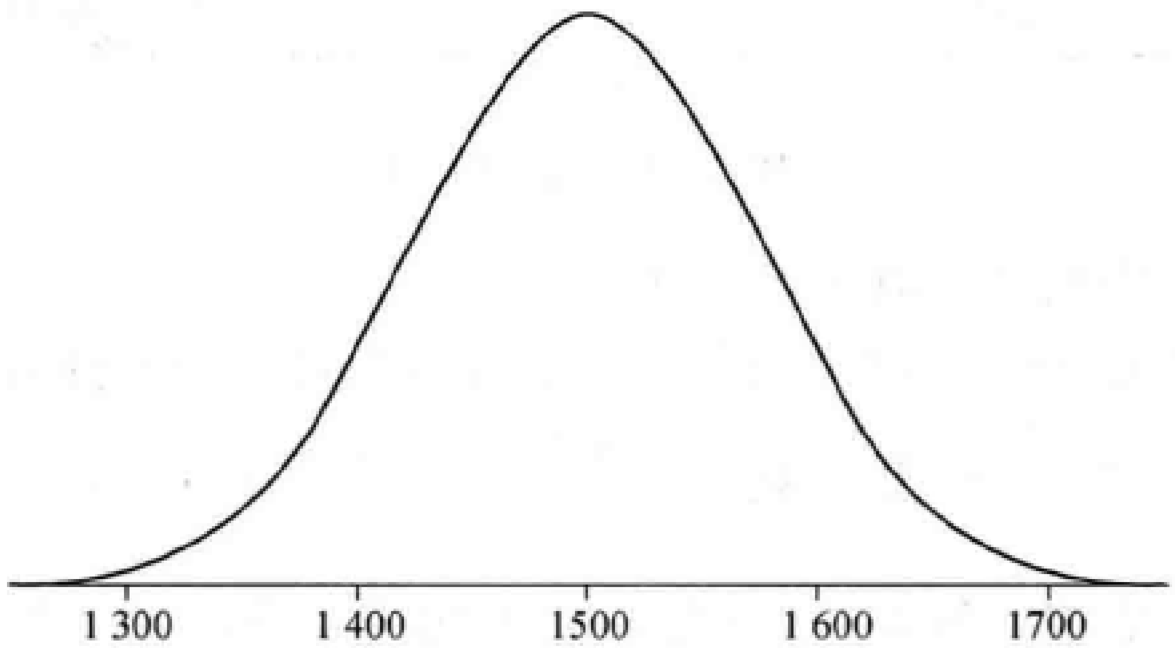

图 6 - 12 正态分布图

(1) 左尾概率(X 小于等于某个数值的概率)

① 随机抽取一个元件,它的寿命不超过 1 410 小时的概率是多少?

我们知道,1 410 的标准化分数是 $\frac{1\,410-1\,500}{75}=-1.2$,查附录的表 A 得到$-1.2$ 的左尾概率为 0.115。

$$P(x<1\,410)=P\left(\frac{x-1\,500}{75}<\frac{1\,410-1\,500}{75}\right)=0.115$$

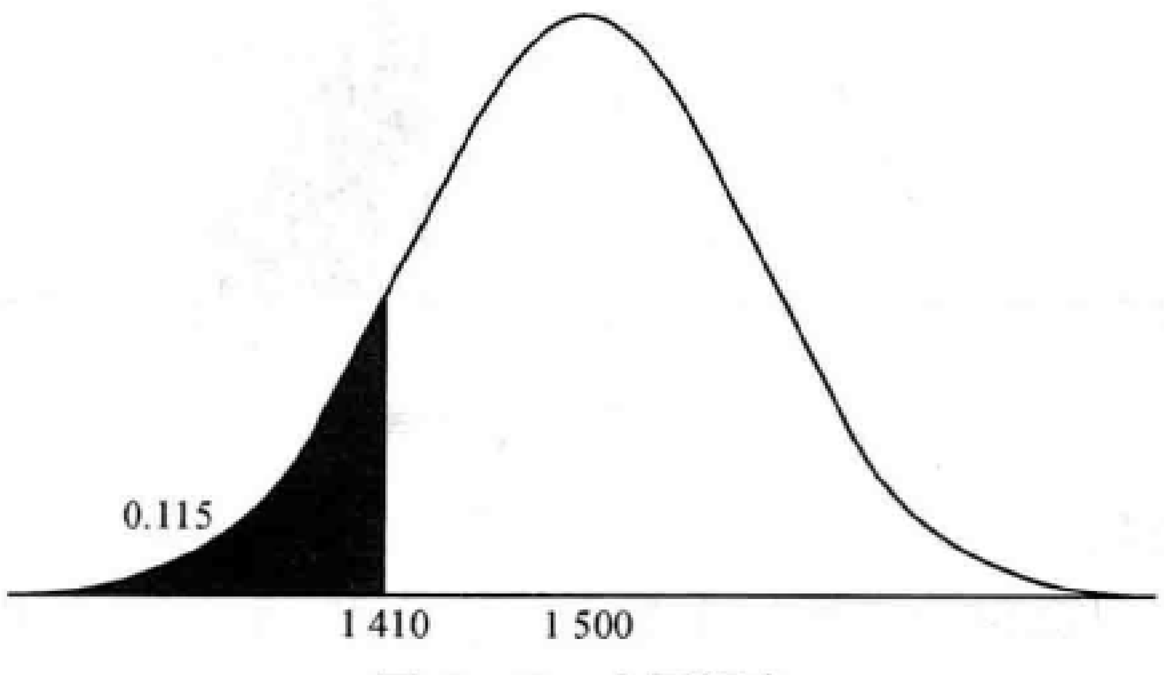

图 6 - 13 左尾概率