

概率论与数理统计



第

7

讲

连续型随机变量及其分布

例1. 一个靶子是半径为2米的圆盘，设击中靶上任一同心圆盘上点的概率与该盘的面积成正比，并设射击都能中靶。以 X 表示弹着点与圆心的距离，试求随机变量 X 的分布函数。

解： X 的分布函数为 $F(x)=P\{X\leq x\}$

依题意，当 $x<0$ 时， $F(x)=P\{X\leq x\}=0$

当 $0\leq x\leq 2$ 时， $F(x)=P\{X\leq x\}=P\{X<0\}+P\{0\leq X\leq x\}=0+kx^2=kx^2$

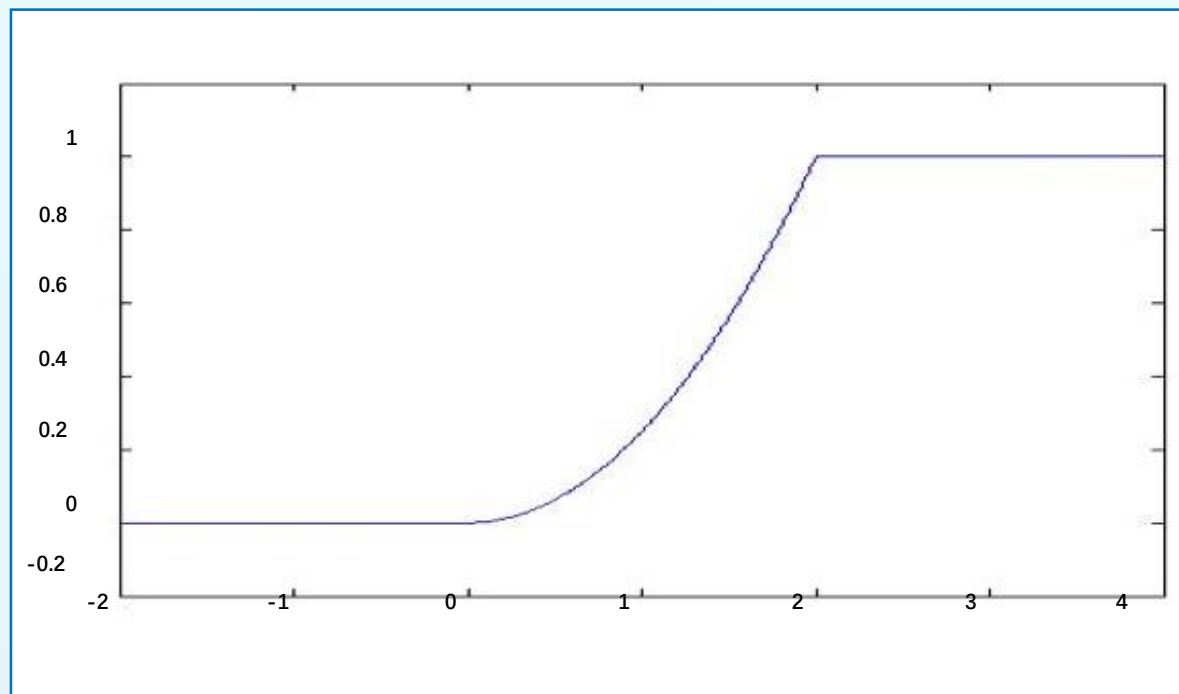
当 $x>2$ 时， $F(x)=P\{X\leq x\}=1$

另外依题意 $F(2)=P\{X\leq 2\}=k\cdot 2^2=1$ 解得 $k = \frac{1}{4}$

一. 连续型随机变量

所以 $F(x) = \begin{cases} 0, & x < 0 \\ \frac{x^2}{4}, & 0 \leq x \leq 2 \\ 1, & x > 2 \end{cases}$

易知 $f(x) = F'(x) = \begin{cases} \frac{x}{2}, & 0 \leq x \leq 2 \\ 0 & \text{其他} \end{cases}$



说明，存在一个非负可积函数 $f(x)$ ，使得下式成立 $F(x) = \int_{-\infty}^x f(t)dt$

定义： 设随机变量 X 的分布函数为 $F(x)$ ，如果存在一个非负可积函数 $f(x)$ ，使对任意的实数 x ，均有

$$F(x) = \int_{-\infty}^x f(t)dt$$

则称 X 是连续型随机变量(Continuous Random Variable)，称 $f(x)$ 是 X 的概率密度函数，简称概率密度(Probability Density Function)。

分布函数 $F(x)$ 和概率密度 $f(x)$ 统称为 X 的概率分布，简称 X 的分布。

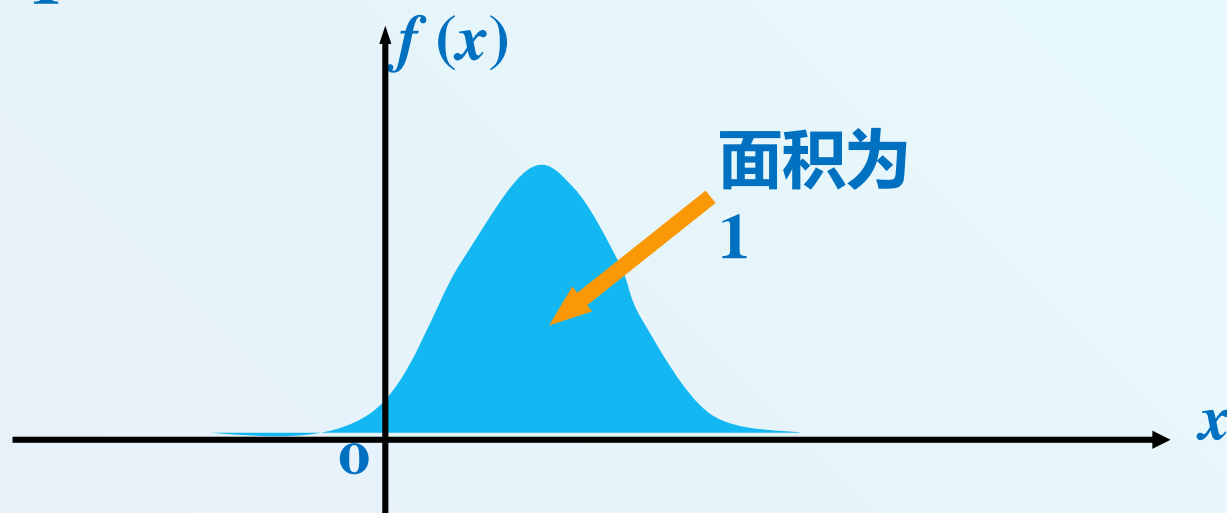
易知此时分布函数 $F(x)$ 是连续函数，即连续型随机变量的分布是连续函数。

概率密度函数的性质

(1) $f(x) \geq 0$

(2) $\int_{-\infty}^{+\infty} f(x)dx = 1$

这两条性质是判定一个函数 $f(x)$ 是否为某 $r.v.$ X 的概率密度函数的充要条件。



$$(3) P\{x_1 < X \leq x_2\} = F(x_2) - F(x_1) = \int_{-\infty}^{x_2} f(x)dx - \int_{-\infty}^{x_1} f(x)dx = \int_{x_1}^{x_2} f(x)dx$$

$$(4) P\{X \in G\} = \int_G f(x)dx \quad \text{如} \quad P\{X > x_1\} = \int_{x_1}^{+\infty} f(x)dx$$

(5) 在 $f(x)$ 的连续点 x 处, 有 $f(x) = F'(x)$

$$(6) \text{ 设 } x \text{ 为 } f(x) \text{ 的连续点, 当 } \Delta x \text{ 较小, 则有 } P\{x < X \leq x + \Delta x\} = \int_x^{x+\Delta x} f(t)dt \approx f(x)\Delta x$$

故 X 的密度 $f(x)$ 在 x 这一点的值, 恰好是 X 落在区间 $(x, x + \Delta x]$ 上的概率与区间长度 Δx 之比。它反映了 X 在 x 附近单位长区间上取值的概率。

连续型随机变量密度函数的性质与离散型随机变量分布律的性质非常相似, 但是, 密度函数值不是概率!

密度函数 $f(x)$ 在某点处 a 的函数值 $f(a)$ ，并不等于 X 取值为 a 的概率。但是，这个值 $f(a)$ 越大，则 X 在 a 附近取值的概率 $f(a)\Delta x$ 就越大。

也可以说，在某点密度曲线的函数值反映了概率集中在该点附近的程度，即 X 在该点附近取值的密集程度。

$$(7) P\{X=x_0\}=F(x_0)-F(x_0-0)=0$$

事实上，由于连续型随机变量的分布函数在整个数轴上都是连续函数，所以有

$$P\{X = x_1\} = F(x_1) - F(x_1 - 0) = 0$$

$$P\{x_1 < X \leq x_2\} = P\{x_1 \leq X \leq x_2\} = P\{x_1 < X < x_2\} = P\{x_1 \leq X < x_2\} = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x) dx$$

若 X 为连续型随机变量, 概率密度 $f(x)$ 唯一确定了分布函数 $F(x)$;

若随机变量 X 的分布函数 $F(x)$ 满足:

(1) $F(x)$ 连续;

(2)存在 $x_1 < x_2 < \dots < x_n (n \geq 0)$, 除这些点外, $F(x)$ 可导, 且导函数 $F'(x)$ 连续;

$$\text{令 } f(x) = \begin{cases} F'(x), & \text{当 } F'(x) \text{ 存在} \\ 0, & \text{当 } F'(x) \text{ 不存在} \end{cases}$$

则 $f(x)$ 必是 X 的概率密度。

例2. 设连续型随机变量 X 的概率密度为 $f(x) = \begin{cases} a - \frac{a}{2}x, & 0 < x < 2 \\ 0, & \text{其他} \end{cases}$

其中 a 为常数。求(1)常数 a 的值; (2) X 的分布函数; (3) $P\{1 < X < 3\}$ 。

解: (1) 由密度函数的归一性得

$$1 = \int_{-\infty}^{+\infty} f(x) dx = \int_0^2 a - \frac{a}{2}x dx = a$$

即 $a=1$

$$f(x) = \begin{cases} 1 - \frac{1}{2}x, & 0 < x < 2 \\ 0, & \text{其他} \end{cases} \quad (2) X \text{ 的分布函数};$$

解:(2) 当 $x \leq 0$ 时, $F(x) = \int_{-\infty}^x f(t)dt = 0$

当 $0 < x < 2$ 时, $F(x) = \int_{-\infty}^x f(t)dt = \int_{-\infty}^0 0dt + \int_0^x 1 - \frac{1}{2}tdt = x - \frac{1}{4}x^2$

当 $x \geq 2$ 时, $F(x) = \int_{-\infty}^x f(t)dt = \int_{-\infty}^0 0dx + \int_0^2 1 - \frac{1}{2}tdt + \int_2^x 0dt = 1$

所以
$$F(x) = \begin{cases} 0, & x \leq 0 \\ x - \frac{1}{4}x^2, & 0 < x < 2 \\ 1, & x \geq 2 \end{cases}$$

$$f(x) = \begin{cases} 1 - \frac{1}{2}x, & 0 < x < 2 \\ 0, & \text{其他} \end{cases}$$

解: (3) $P\{1 < X < 3\} = \int_1^3 f(x)dx = \int_1^2 (1 - \frac{1}{2}x)dx + \int_2^3 0dx = \frac{1}{4}$

或者 $P\{1 < X < 3\} = F(3) - F(1) = 1 - (1 - \frac{1}{4} \times 1^2) = \frac{1}{4}$

在上例中, 当 $x \notin (0, 2)$ 时, $f(x) = 0$, 即: $P\{X \notin (0, 2)\} = 0$, 为了方便, 我们说 X 只在 $(0, 2)$ 上取值。

一般地, 若随机变量 X 的概率密度 $f(x)$ 是如下分段函数: $f(x) = \begin{cases} g(x) > 0, & a \leq x \leq b \\ 0, & \text{其他} \end{cases}$
我们就说 X 只在 $[a, b]$ 上取值。

例3. 设连续型随机变量 X 的分布函数为

$$F(x) = \begin{cases} 0, & x < 0 \\ Cx^2, & 0 \leq x < 1 \\ 1, & x \geq 1 \end{cases}$$

求(1)常数 C 值; (2) X 取值于 $(0.3, 0.7)$ 内的概率; (3) X 的密度函数。

解: (1)应用连续型随机变量 X 的分布函数的连续性, 有

$$1 = F(1) = \lim_{x \rightarrow 1-} F(x) = C \quad \text{即 } C=1$$

$$(2) P\{0.3 < X < 0.7\} = F(0.7) - F(0.3) = 0.7^2 - 0.3^2 = 0.4$$

$$(3) f(x) = F'(x) = \begin{cases} 2x, & 0 < x < 1 \\ 0, & \text{其他} \end{cases}$$

例4.某电子元件的寿命 X (单位: 小时)的概率密度为
$$f(x) = \begin{cases} 0, & x \leq 100 \\ 100/x^2, & x > 100 \end{cases}$$

求5个同类型的元件在使用的前150小时内恰有2个需要更换的概率。

解: 设 $A = \{\text{一元件在前150 小时内需要更换}\}$

$B = \{\text{5个同类型的元件在使用的前150小时内恰有2个需要更换}\}$

$$P(A) = P\{X \leq 150\} = \int_{-\infty}^{150} f(x)dx = \int_{-\infty}^{100} 0dx + \int_{100}^{150} \frac{100}{x^2} dx = \frac{1}{3}$$

例4.某电子元件的寿命 X (单位: 小时)的概率密度为 $f(x) = \begin{cases} 0, & x \leq 100 \\ 100/x^2, & x > 100 \end{cases}$

求5个同类型的元件在使用的前150小时内恰有2个需要更换(B)的概率。

解: 设 Y 表示5个元件中需要更换的个数。则

$$Y \sim b(5, 1/3)$$

所以 $P(B) = P\{Y = 2\} = C_5^2 \times \left(\frac{1}{3}\right)^2 \times \left(\frac{2}{3}\right)^3 = \frac{80}{243}$

随机变量的分类:



注 意



存在既不是连续型随机变量，又不是离散型随机变量的例子。

如：设 $X \sim U(0,1)$, 令 $Y = \min\{X, 0.75\}$, 则 Y 为非连续非离散型随机变量。

1. 均匀分布(Uniform Distribution)

(1) 若随机变量 X 的概率密度为

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

区间 (a,b) 上的示性函数

则称 X 在 (a, b) 上服从均匀分布, 记为 $X \sim U(a, b)$

$$I_{(a,b)} = I_{(a,b)}(x) = \begin{cases} 1, & x \in (a, b) \\ 0, & x \notin (a, b) \end{cases}$$

类似地, 我们可以定义区间 $[a, b)$ 、 $(a, b]$ 和 $[a, b]$ 上的均匀分布。

可以看到，均匀分布的密度函数 $f(x)$ 区间 (a,b) 上为常数，故在这个区间上概率在各点处的密集程度是一样的，也可以说，概率均匀地分布在这个区间上，这可以理解为均匀分布这个名称的由来。

一般地，设 D 是数轴上一些不相交的区间之和，若 X 的概率密度为

$$f(x) = \begin{cases} \frac{1}{D\text{的长度}}, & x \in D \\ 0, & x \notin D \end{cases}$$

则称 X 在 D 上服从均匀分布。

若 $X \sim U(a, b)$, X 的分布函数为

$$F(x) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a < x < b \\ 1, & \text{其他} \end{cases}$$

对于满足 $a \leq c < d \leq b$ 的任意的 c, d , 有

$$P\{c < X \leq d\} = \frac{d-c}{b-a}$$

表明, 随机变量 X 落在区间 (a, b) 内的任一子区间 (c, d) 的概率, 只与该子区间的长度 $d-c$ 成正比, 而与该子区间在区间 (a, b) 中的具体位置无关。

例5. 设公共汽车站从上午7时起每隔15分钟来一班车，如果某乘客到达此站的时间是7:00到7:30之间的均匀随机变量。试求该乘客候车时间不超过5分钟的概率。

解：设该乘客于7时 X 到达此站，则 X 服从 $[0, 30]$ 上的均匀分布

$$f(x) = \begin{cases} \frac{1}{30} & 0 \leq x \leq 30 \\ 0 & \text{其它} \end{cases}$$

令 $B = \{\text{候车时间不超过5分钟}\}$

$$P(B) = P\{10 \leq X \leq 15\} + P\{25 \leq X \leq 30\} = \int_{10}^{15} \frac{1}{30} dx + \int_{25}^{30} \frac{1}{30} dx = \frac{1}{3}$$

例6. 设随机变量 $X \sim U(a, b)$, $a > 0$, 且 $P\{0 < X < 3\} = 1/4$, $P\{X > 4\} = 1/2$, 求 $P\{1 < X < 5\}$ 。

解: 易知 $0 < a \leq 3$, 否则 $P\{0 < X < 3\} = 0$, 同理, $b > 4$, 否则 $P\{X > 4\} = 0$ 。

$$\text{又有 } P\{0 < X < 3\} = \frac{3-a}{b-a} = \frac{1}{4} \quad P\{X > 4\} = \frac{b-4}{b-a} = \frac{1}{2}$$

$$\text{所以得方程组 } \begin{cases} 3a + b = 12 \\ a + b = 8 \end{cases} \quad \text{解得 } \begin{cases} a = 2 \\ b = 6 \end{cases}$$

所以

$$P\{1 < X < 5\} = \frac{5-2}{6-2} = \frac{3}{4}$$

2.指数分布(Exponential Distribution)

若随机变量 X 的概率密度为

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

其中常数 $\lambda > 0$, 则称 X 服从参数为 λ 的指数分布。记为 $E(\lambda)$

易求得 X 的分布函数为

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

指数分布的另一种等价定义

定义：设连续型随机变量 X 的概率密度为 $f(x) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}}, & x > 0 \\ 0, & x \leq 0 \end{cases}$

其中 $\theta > 0$ 为常数，则称 X 服从参数为 θ 的指数分布。

服从指数分布的随机变量 X 具有以下性质： **无记忆性或无后效性**

即对于任意 $s, t > 0$ ，有 $P\{X > s+t \mid X > s\} = P\{X > t\}$

事实上
$$P\{X > s+t \mid X > s\} = \frac{P\{X > s+t, X > s\}}{P\{X > s\}} = \frac{P\{X > s+t\}}{P\{X > s\}}$$
$$= \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = 1 - F(t) = P\{X > t\}$$

如果 X 表示某仪器的工作寿命，无后效性的解释是：当仪器工作了 s 小时后再能继续工作 t 小时的概率等于该仪器刚开始就能工作 t 小时的概率。

说明该仪器的使用寿命不随使用时间的增加发生变化，或说仪器是“永葆青春”的。

一般来说，电子元件等具备这种性质，它们本身的老化是可以忽略不计的，造成损坏的原因是意外的高电压等等。

例7. 设某大型设备在任何长度为 t (小时) 的时间内发生故障的次数 $N(t)$ 服从参数为 λt 的泊松分布。

- 求 (1) 相继出现两次故障之间的时间间隔 T 的概率分布;**
(2) 设备已经无故障工作 a 小时的情况下, 再无故障工作 b 小时的概率。

解: (1) 已知 $N(t)$ 服从泊松分布 $\pi(\lambda t)$ 。

所以
$$P\{N(t) = k\} = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, k = 0, 1, \dots$$

而两次故障之间的时间间隔 T 是非负随机变量, 且对任意的 $t > 0$, 事件 $\{T > t\}$ 表明该设备在 $[0, t]$ 内没有发生故障, 即 $\{T > t\}$ 等价于 $\{N(T) = 0\}$

当 $t \leq 0$ 时, $F(t) = P\{T \leq t\} = 0$

当 $t > 0$ 时, $F(t) = P\{T \leq t\} = 1 - P\{T > t\} = 1 - P\{N(t) = 0\} = 1 - \frac{(\lambda t)^0}{0!} e^{-\lambda t} = 1 - e^{-\lambda t}$

即
$$F(t) = \begin{cases} 1 - e^{-\lambda t}, & t \geq 0 \\ 0, & t < 0 \end{cases}$$

(2) 利用指数分布的无记忆性

$$P\{X > a + b \mid X > a\} = P\{X > b\} = 1 - F(b) = e^{-b\lambda}$$

3.正态分布(Normal Distribution)

若随机变量 X 的概率密度为

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

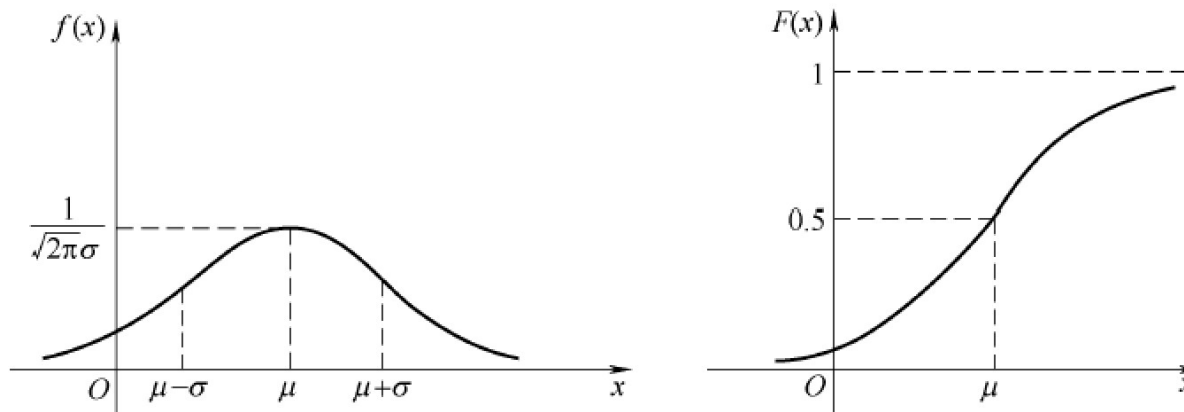
其中 μ, σ 均为常数, 且 $\sigma > 0$, 则称 X 服从参数为 μ 和 σ 的正态分布。

记作 $X \sim N(\mu, \sigma^2)$

分布函数为

$$F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{1}{2\sigma^2}(t-\mu)^2} dt$$

二. 几种重要的连续型随机



a. 正态分布的密度曲线是一条关于 μ 对称的钟形曲线。 $f(\mu+x)=f(\mu-x)$

特点是“两头小，中间大，左右对称”。

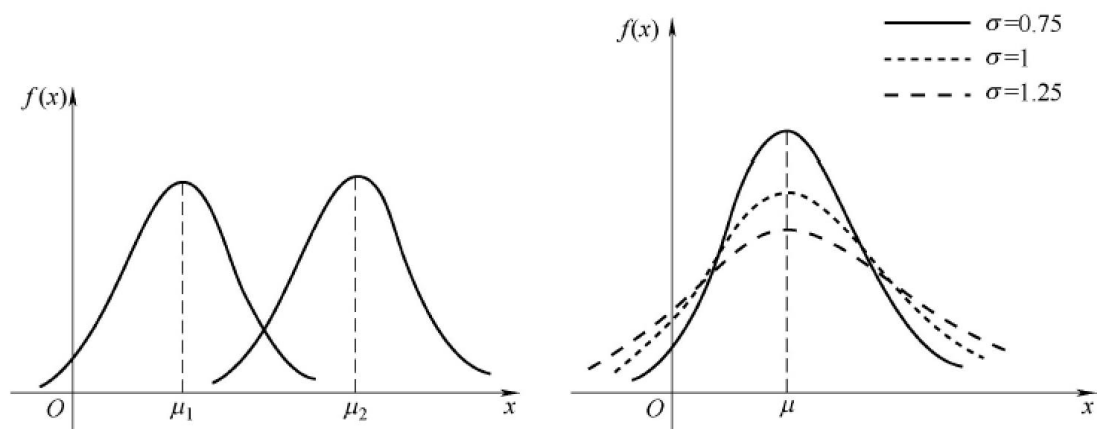
b. 只有一个峰, 在 $x=\mu$ 处达到最大值: $f(\mu) = \frac{1}{\sqrt{2\pi}\sigma}$

说明 X 落在 μ 附近的概率最大, 或者说 X 的取值在 μ 附近最密集。

d. 曲线 $f(x)$ 向左右伸展时, 越来越贴近 x 轴, 即 $f(x)$ 以 x 轴为渐近线, 当 $x \rightarrow \pm\infty$ 时, $f(x) \rightarrow 0$ 。

d. $x = \mu \pm \sigma$ 为 $f(x)$ 的两个拐点的横坐标。

e. μ 决定了图形的中心位置, 称为位置参数; σ 决定了图形中峰的陡峭程度, 称为形状参数或者刻度参数。



正态分布是概率论与数理统计中最重要的分布之一，德国数学家高斯（Gauss, 1777—1855）在研究误差理论时首先用正态分布来刻画误差的分布，所以正态分布也称为**高斯分布**。

■ 实际中，很多随机变量都服从或近似服从正态分布

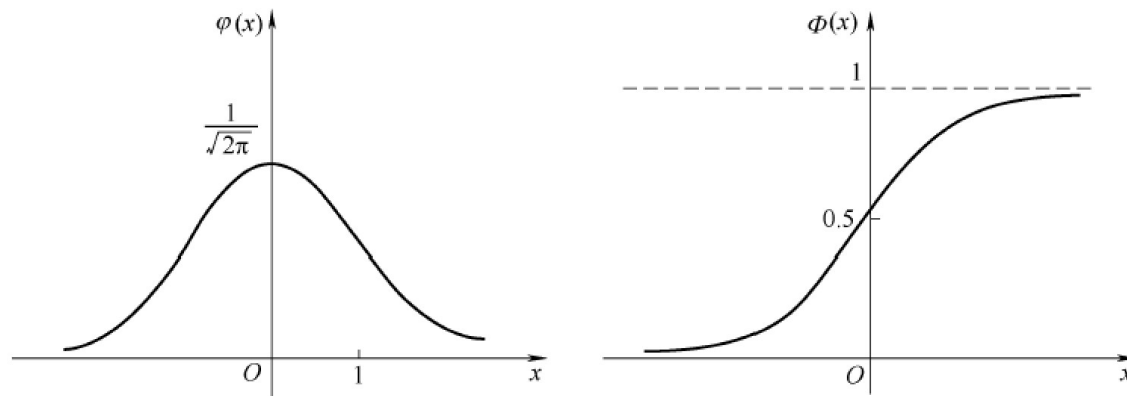


例如，零件长度的测量误差，某地区居民的收入，年降雨量、同龄人身高、在正常条件下各种产品的质量指标——如零件的尺寸、纤维的强度和张力、农作物的产量，小麦的穗长、株高、测量误差、射击目标的水平或垂直偏差、信号噪声等等，都服从或近似服从正态分布。

$\mu=0, \sigma=1$ 的正态分布称为标准正态分布

其密度函数和分布函数常用 $\varphi(x)$ 和 $\Phi(x)$ 表示

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



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

注 意

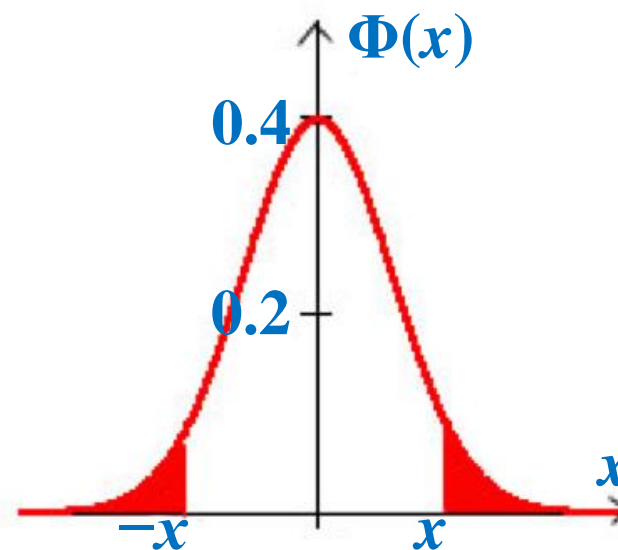
$\Phi(0)=0.5$, $\Phi(-x)=1-\Phi(x)$

若 $X \sim N(0, 1)$, 对任意的实数 x_1, x_2 ($x_1 < x_2$), 有

$P\{X \leq x_1\} = \Phi(x_1)$ $P\{X > x_1\} = 1 - \Phi(x_1)$

$P\{x_1 \leq X \leq x_2\} = \Phi(x_2) - \Phi(x_1)$

人们已编制了 $\Phi(x)$ 的函数表, 可供查用。



正态分布的计算

$$F(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt = \int_{-\infty}^{(x-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

对任意的实数 x_1, x_2 ($x_1 < x_2$), 有

$$P\{X \leq x_1\} = F(x_1) = \Phi\left(\frac{x_1 - \mu}{\sigma}\right)$$

$$P\{X > x_1\} = 1 - F(x_1) = 1 - \Phi\left(\frac{x_1 - \mu}{\sigma}\right)$$

$$P\{x_1 < X \leq x_2\} = F(x_2) - F(x_1) = \Phi\left(\frac{x_2 - \mu}{\sigma}\right) - \Phi\left(\frac{x_1 - \mu}{\sigma}\right)$$

例8. 某种电子元件在电源电压不超过200伏, 200~240伏, 超过240伏三种情况下损坏的概率分别为0.1、 0.001、 0.2, 设电源电压 $X \sim N(220, 25^2)$ 。

求(1)此种电子元件的损坏率;

(2)此种电子元件损坏时, 电源电压在200~240伏的概率。

解: (1)设 $A=\{\text{电子元件损坏}\}$, $B_1=\{\text{电压不超过200伏}\}$, $B_2=\{\text{电压为200~240伏}\}$, $B_3=\{\text{电压超过240伏}\}$, 且 X 的分布函数为 $F(x)$ 。

由已知得 $P(A/B_1)=0.1$, $P(A/B_2)=0.001$, $P(A/B_3)=0.2$ 。

$$P(B_1) = P\{X \leq 200\} = F(200) = \Phi\left(\frac{200 - 220}{25}\right) = 1 - \Phi(0.8) = 0.2119$$

$$\begin{aligned} P(B_2) &= P\{200 < X < 240\} = F(240) - F(200) \\ &= \Phi\left(\frac{240-220}{25}\right) - \Phi\left(\frac{200-220}{25}\right) = 2\Phi(0.8) - 1 = 0.5762 \end{aligned}$$

$$P(B_3) = P\{X \geq 240\} = 1 - F(240) = 1 - \Phi\left(\frac{240-220}{25}\right) = 1 - \Phi(0.8) = 0.2119$$

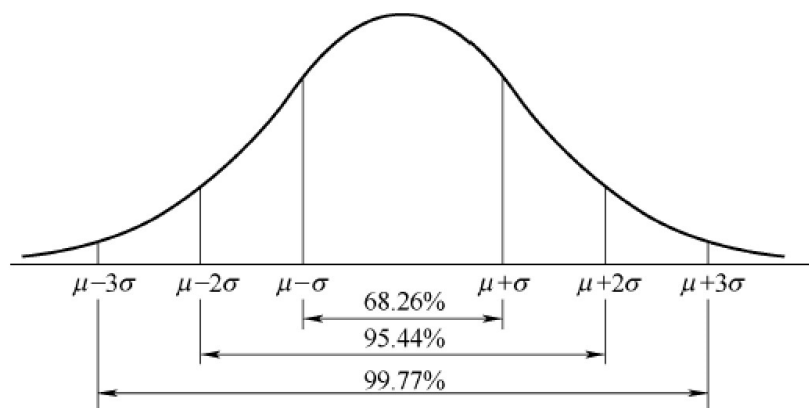
故，由全概率公式得

$$P(A) = \sum_{i=1}^3 P(B_i) P(A|B_i) = 0.1 \times 0.2119 + 0.001 \times 0.5762 + 0.2 \times 0.2119 = 0.0641$$

$$(2) \text{由贝叶斯公式得 } P(B_2|A) = \frac{P(B_2) P(A|B_2)}{P(A)} = \frac{0.001 \times 0.5762}{0.0641} = 0.009$$

例9. 设 $X \sim N(\mu, \sigma^2)$, 求 $P\{|X - \mu| < k\sigma\}$ 的值, $k=1, 2, 3$ 。

解:
$$P\{|X - \mu| < k\sigma\} = P\{\mu - k\sigma < X < \mu + k\sigma\} = F(\mu + k\sigma) - F(\mu - k\sigma)$$
$$= \Phi\left(\frac{\mu + k\sigma - \mu}{\sigma}\right) - \Phi\left(\frac{\mu - k\sigma - \mu}{\sigma}\right) = \Phi(k) - \Phi(-k) = \Phi(k) - [1 - \Phi(k)] = 2\Phi(k) - 1$$



由此可以看出, 尽管正态分布的随机变量的取值范围为 $(-\infty, +\infty)$, 但它的值落在 $(\mu - 3\sigma, \mu + 3\sigma)$ 内的概率达到 0.9974, 这是几乎肯定的事. 这就是统计中正态分布的 “3 σ 原则”。

质量控制中的 3σ 原则

设在正常生产的情况下，某零件的尺寸 X 服从正态分布 $N(\mu, \sigma^2)$ ，为了在生产过程中随时检查有无系统性误差出现，人们画了一个质量控制图。

每隔一定时间，对产品尺寸进行检查，测量的产品的尺寸应落在上、下控制线之内。如果超出控制线，则很有可能是生产出现了异常情况，应该暂停生产进行检查。当然也可能虚报，但虚报的可能性比较小。



作业： 25,27,31,32,35,36,,40,42,46

第

7

讲

谢谢观看