目录

文件名

第一部分　正态分布 and 标准正态分布

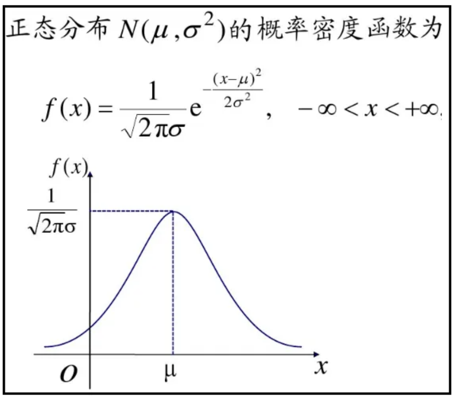
1　正态分布, normal distribution

正态分布, normal distribution, 直译过来就是 “最常态下的分布”, “一般最常见的分布”. 它又名高斯分布 Gaussian distribution.
正态分布, 是概率分布中最重要的分布. 在数学家眼里, 它是远远高于其他分布的. 有很多其他的分布，比如对数正态分布、T 分布、F 分布, 都是直接由 “正态分布” 推导出来的.

1.1　正态分布 - 概率密度函数: $\boxed{\varphi(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}}$

概率密度函数. 用小写的 φ 表示.
“正态分布” $N(\underbrace{\mu}_{\text{平均值.}}, \underbrace{\sigma^2}_{\sigma\text{是标准差. }\sigma^2\text{是方差}})$ 的概率函数是: $\boxed{\varphi(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \ (-\infty < x < +\infty)}$

记作: $\boxed{X \sim N(\mu, \sigma^2)}$ ← 称为: X 服从 “参数为 μ, σ 的正态分布 (或高斯分布)”.
- 这里的 N, 就是正态分布 (Normal distribution) 的英文首字母.
- μ 是 “平均值”
- σ 是 “标准差”
- 注意: 概率函数公式里, 这第二个参数写的是 σ^2, 而不是 σ! 所以, 比如对于 N(1, 100) 来说, 其 $\mu = 1$, $\sigma^2 = 100$, 即 $\sigma = 10$.
- 在正态分布中，“平均值 μ” 等于 “期望”，它决定了这条曲线的最高点; “方差 σ^2” 决定胖瘦，它决定曲线的弯曲度. 简单这两个数据，就确定了这条曲线的形状.



正态分布的 f(x) 曲线:
- 横坐标, 代表随机变量 X 的取值范围. 越往右, 随机变量的值就越大.
- 纵坐标, 则代表概率的大小，最底下的概率是 0，越往上概率越大.
这样，从曲线上随便找一点，确定它的横坐标、纵坐标，我们就知道了这个值出现的概率是多少.

因为这条曲线是左右对称的，所以:
- 中间的最高点，就代表 **“平均值” 出现的概率最大，数据最多.**
- 而两边陡峭下降，就意味着: **越靠近平均值，数据越多; 越远离平均值，数据就越少.**

1.2 正态分布 - 累加函数: $\boxed{F(x)=\frac{1}{\sqrt{2\pi}\cdot\sigma}\int_{-\infty}^{x}\left[e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right]dx}$

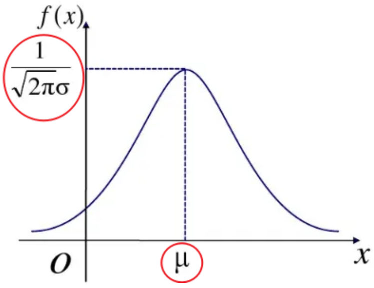
对 “概率函数 f(x)” 求积分, 其曲线下的阴影面积就是 “累加函数 F(x)”. 其面积 =1.

$$
\begin{aligned}
\underbrace{F(x)}_{\text{累加函数}} &= \int \underbrace{f(x)}_{\text{概率函数}}\, dx \\
&= \int_{-\infty}^{x}\left[\frac{1}{\sqrt{2\pi}\cdot\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right]dx \\
&= \frac{1}{\sqrt{2\pi}\cdot\sigma}\int_{-\infty}^{x}\left[e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right]dx
\end{aligned}
$$

1.3 正态分布 (钟形曲线) 的 “概率函数” 的性质

1.3.1 以 “x= 均值 μ” 为对称轴

正态分布的 “概率函数” 曲线, 以 “x= 均值 μ” 为对称轴. 在此处, 函数的 y 值达到最大. 即此时 $y=\frac{1}{\sqrt{2\pi}\cdot\sigma}$



所以, 对称轴 μ, 能控制图像的 “左右平移”.

均值就是期望. 它们重合在同一点处 — 即 “曲线中间的最高点” 的 x 坐标处.
数学期望, 代表 “长期中的价值为几许?”. 在正态分布中，“平均值” 就能代表随机事件的长期价值.

为什么我们会用高考的平均成绩，衡量一所高中的教学质量? 为什么我们会用平均收益率，衡量一家基金公司的好坏? 原因很简单，因为高考成绩和基金公司的收益，都是服从 “正态分布” 的.

1.3.2 极端值很少. 且 “极端值” 对 “均值” 的影响很小

正态分布的图形, 越靠近平均值,这条曲线越高,出现的概率越大; 越远离平均值,这条曲线就越低,出现的概率就越小. 这就说明, **正态分布的大多数数据, 都集中在平均值附近, 极端值很少.**

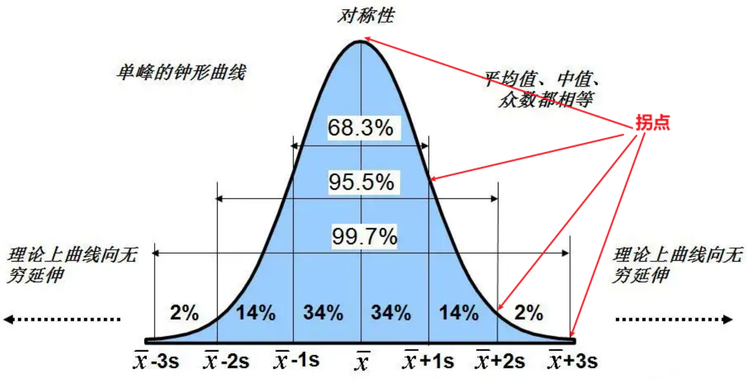
“极端值很少” 这句话，有两层含义:
- **一是极端值出现的概率很低**
- **二是极端值对 “均值” 的影响很小**

也因此，正态分布中的数据是非常稳定的. 比如人的身高，它大体服从正态分布，**所以即使姚明加入我们课程, 我们的平均身高也不会有太大变化 (即:“极端值”对“均值”的影响很小).**

1.3.3 以 x 轴为渐近线

就是说, 曲线的两端, 无限接近于 y=0, 而不会掉落到 -y 领域上去.

1.3.4 在 $x = \mu \pm \sigma$ 处有“拐点”



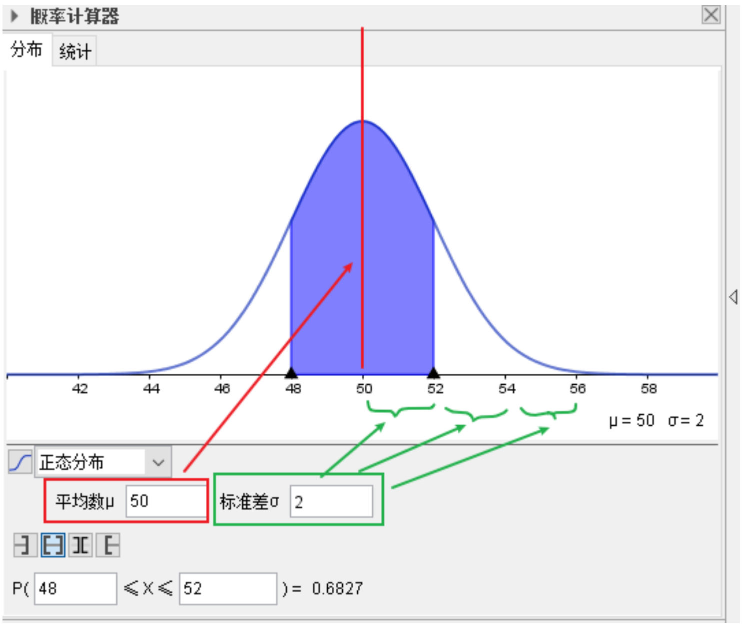
正态分布

1.3.5 “σ 标准差”参数, 控制图像的“矮胖”或“高瘦”

→ 若 σ 变小: 因为在 x= 处, y 有最大值是 $\frac{1}{\sqrt{2} \cdot \sigma}$. 所以当 σ 变小时, 分母变小, 则分数值就变大, 即 y 值变大, 所以图像会拉高, 变瘦高.
→ 若 σ 变大: 则最高点的 y 值变小, 图像会压低, 变矮胖.

- 标准差 σ 越大，数据的波动越剧烈，钟形曲线就越矮胖 (即 x 轴的横跨幅度越大).
- 标准差 σ 越小，波动就越小, 数据就越集中，钟形曲线就越高瘦 (即 x 轴的横跨幅度越窄).

但注意, 无论是变瘦高, 还是变矮胖, 曲线下的阴影面积, 始终是 =1, 不变的!



1.3.6 对不同的“正态分布”曲线, 进行比较

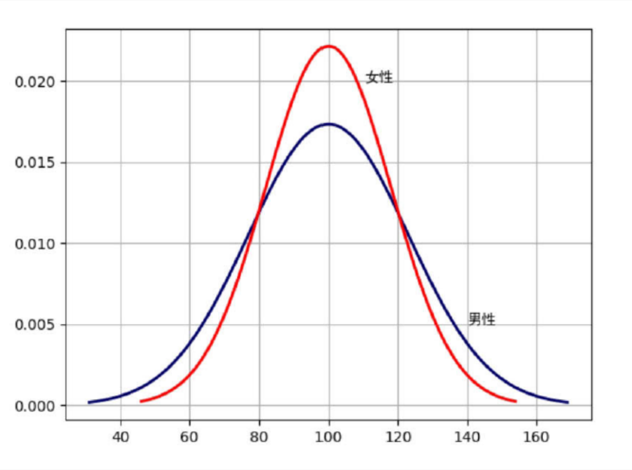
不同的正态分布曲线, 也能进行比较:

(1) **标准差 σ 相同, 均值 μ 不同, 能比较“好坏 (即价值高低)”.**
因为均值即”期望”, 期望就代表”长期价值”. 两个事物的“期望”不同, 自然它们的价值高低也不同.

(2) **均值 μ 相同, 标准差 σ 不同, 能比较“波动”(即风险性).**

例

男女智商的正态分布曲线如下,



能看出:
- 两者的均值相同. 说明男女智商, 没有高低之分.
- 但高矮胖瘦不一样 (即”标准差 σ”不一样, 波动程度不一样) :
→ 男性智商的波动性更大 (即 x 跨幅更宽, 更矮胖), 说明在智商高的人中间，男性的数量要多于女性 (比如, **我们看 X=140 智商处, 男性的蓝色曲线的高度, 要高于女性的红色曲线, 说明在 X=140 智商处, 男的概率, 要多于女的**);
→ 当然，**智商低下的人中间 (比如 X=60 智商处)，男性也同样比女性多.**

(3) **均值和标准差, 都不同. 那也能比较“专业和业余”**
比如, 某体育项目, 你和世界冠军同台比赛, 他比你得分高 (期望大), 又成绩稳定 (方差小), 所以这两项都比你强, 就说明他比你“专业”.
所以, **专业就是“均值 μ 更高，标准差 σ 更小”. 而业余则恰恰相反.**

---

2 标准正态分布 ← 即当均值 $\mu=0$, 标准差 $\sigma=1$ 时的“正态分布”

2.1 标准正态分布 - 概率函数: $\boxed{\phi_0(x)=f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}}$

我们把 $\mu=0,\sigma=1$, 代入正态分布的 PDF 和 CDF 函数中, 就得到:

“标准正态分布”的“概率函数 PDF”(专门记作 $\phi_0(x)$) :

$$\phi_0(x)=f(x)=\frac{1}{\sqrt{2\pi}\cdot\underbrace{\sigma}_{=1}}e^{-\frac{(x-\overbrace{\mu}^{=0})^2}{2\sigma^2}}=\frac{1}{\sqrt{2\pi}\cdot 1}e^{-\frac{(x-0)^2}{2\cdot 1^2}}=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$$

即: $\boxed{\phi_0(x)=f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}}$

2.2 标准正态分布 - 累加函数: $\boxed{\Phi_0(x)=F(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\left[e^{-\frac{x}{2}}\right]dx}$

"标准正态分布" 的 "累加函数 CDF" (专门记作 $\Phi_0(x)$) :

$$\begin{aligned}\Phi_0(x)&=F(x)\\&=\frac{1}{\sqrt{2\pi}\cdot\underbrace{\sigma}_{=1}}\int_{-\infty}^{x}\left[e^{-\frac{(x-\overbrace{\mu}^{=0})^2}{2\sigma^2}}\right]dx\\&=\frac{1}{\sqrt{2\pi}\cdot 1}\int_{-\infty}^{x}\left[e^{-\frac{(x-0)^2}{2\cdot 1^2}}\right]dx\\&=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\left[e^{-\frac{x}{2}}\right]dx\end{aligned}$$

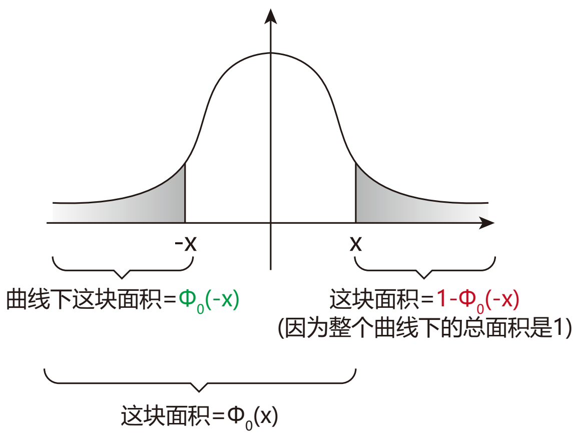
即: $\boxed{\Phi_0(x)=F(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\left[e^{-\frac{x}{2}}\right]dx}$

2.3 标准正态分布的性质

2.3.1 因为它的 $\mu=0$, 所以它的函数曲线, 关于 x=0 对称, 即 y 轴是对称轴.

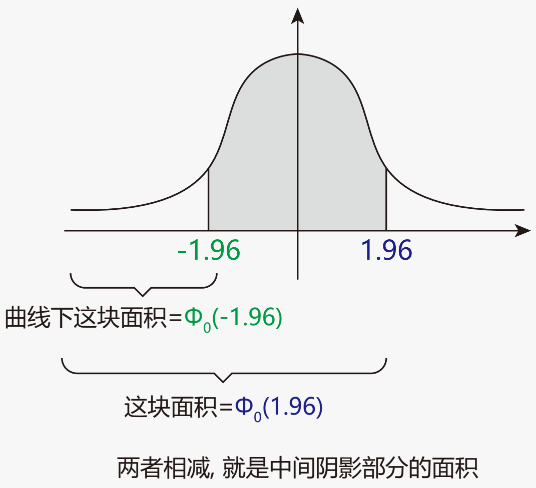
所以它就是个偶函数. 有:
- 概率函数 $\varphi_0(x)=\varphi_0(-x)$ ← 我们在下标处加个 0, 来表示它是 "标准" 的正态分布函数的 "概率函数" 或 "累加函数".

- 其 "累加函数" 有: $\boxed{\Phi_0(-x)=1-\Phi_0(x)}$ ← 这个公式很重要!
比如: $\Phi_0(-4)=1-\Phi_0(4)$



例

$$\begin{aligned}\text{求 } P\{|x|\le 1.96\}&=P\{-1.96\le X\le 1.96\}\\&=\Phi_0(1.96)-\underbrace{\Phi_0(-1.96)}_{=1-\Phi_0(1.96)}\\&=\Phi_0(1.96)-1+\Phi_0(1.96)\\&=2\Phi_0(1.96)-1\end{aligned}$$

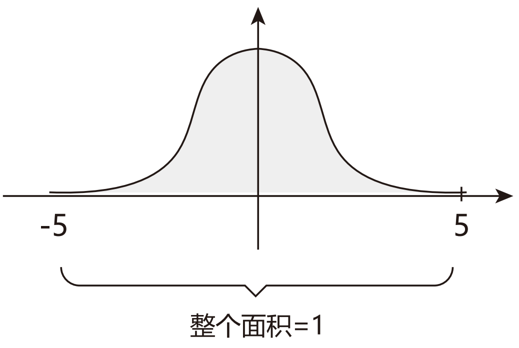


2.3.2 对于 $x \geq 5$ 的 y 值, 已经非常靠近 y=0 了

正态分布的值, 怎么算? — 查表.
一般, 书上给出的都是 "标准正态分布" 的表. 所以如果你是普通的 "正态分布", 必须先把它转成 "标准正态分布", 再来查表.

并且, 表的范围, 只给出了 $0 \leq x < 5$ 的值. **因为对于 $x \geq 5$ 的值, 此时的曲线高度, 即 y 值, 已经非常靠近 y=0 了. 所以我们就可以认为: 对于 $x \geq 5$ 的标准正态分布的 "概率函数 $\varphi_0(x)$" 的 y 值, 都 =0.**

同样, 来看累加函数 CDF:
对于 $x \geq 5$ 时, 其位置已经非常靠近整个曲线的右端末尾了, 而整个函数曲线下的面积也就 =1, 所以, 在 $x \geq 5$ 处的 "累加函数 $\Phi_0(x)$", 其值我们就可以认为是 1.



即:

标准正态分布	概率函数 $\varphi_0(x)$	累加函数 $\Phi_0(x)$
当 $x \leq -5$ 时	$y \approx 0$	$y \approx 0$
当 $x \geq 5$ 时	$y \approx 0$	$y \approx 1$

3 普通的“正态分布”，怎样转化成“标准正态分布”?

3.1 “概率函数”的转化公式是: $\boxed{\varphi(x)=\frac{1}{\sigma}\cdot\varphi_0(\frac{x-\mu}{\sigma})}$

普通“正态分布”的概率函数公式是:

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

我们来给它做一下变形（注意标出颜色的地方的变化）:

$$\begin{aligned}\varphi(x)&=\frac{1}{\sqrt{2\pi}\cdot\sigma}e^{-\frac{(\frac{x-\mu}{\sigma})^2}{2}}\\&=\frac{1}{\sigma}\frac{1}{\sqrt{2\pi}}e^{-\frac{(\frac{x-\mu}{\sigma})^2}{2}}\end{aligned}$$

即: $$\underbrace{\varphi(x)}_{\text{正态分布的概率函数}}=\frac{1}{\sigma}\frac{1}{\sqrt{2\pi}}e^{-\frac{(\frac{x-\mu}{\sigma})^2}{2}}$$

变到这里后，你来和“标准正态分布”的概率函数

$$\underbrace{\varphi_0(x)}_{\text{标准正态分布的概率函数}}=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$$

来做对比. 会发现: 两者的差别只在于红色标出的地方.

所以，两者“概率函数”的转化公式，就是:

$$\boxed{\underbrace{\varphi(x)}_{\text{正态分布的概率函数}}=\frac{1}{\sigma}\cdot\underbrace{\varphi_0(\overbrace{\frac{x-\mu}{\sigma}}^{\text{这一块，就是标准正态分布的}x\text{值}})}_{\text{标准正态分布的概率函数}}}$$

例

你每天等公交的时间，是一个随机变量 X，这个变量服从正态分布. 过去 20 天，你等公交的时间 (分钟) 分别是: 26、33、65、28、34、55、25、44、50、36、26、37、43、62、35、38、45、32、28、34.

那么，你现在等公交会耗费 30-45 分钟的概率是多少? — 即求: $P(30<X<45)$.

(1) 先算“正态分布: 所需要的两个参数: 均数 μ, 标准差 σ.

3.2 “累加函数”的转化公式是: $\boxed{\Phi(x)=\Phi_0(\frac{x-\mu}{\sigma})}$

普通“正态分布”的“累加函数”公式是:

$$\Phi(x)=\frac{1}{\sqrt{2\pi}\cdot\sigma}\int_{-\infty}^{x}\left[e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right]dx$$

我们来给它做一下变形(注意标出颜色的地方的变化):

$$\begin{aligned}\Phi(x)&=\frac{1}{\sqrt{2\pi}\cdot\sigma}\int_{-\infty}^{x}\left[e^{-\frac{(\frac{x-\mu}{\sigma})^2}{2}}\right]dx\\&=\frac{1}{\sigma}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\left[e^{-\frac{(\frac{x-\mu}{\sigma})^2}{2}}\right]d(x-\underbrace{\mu}_{\text{常数}})\\&=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\left[e^{-\frac{(\frac{x-\mu}{\sigma})^2}{2}}\right]d\left(\frac{x-\mu}{\sigma}\right)\end{aligned}$$

即: $$\underbrace{\Phi(x)}_{\text{正态分布的累加函数}}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\left[e^{-\frac{(\frac{x-\mu}{\sigma})^2}{2}}\right]d\left(\frac{x-\mu}{\sigma}\right)$$

变到这里后, 你来和“标准正态分布”的“累加函数”

$$\underbrace{\Phi_0(x)}_{\text{标准正态分布的累加函数}}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\left[e^{-\frac{x}{2}}\right]dx$$

来做对比. 会发现: 两者的差别只在于红色标出的地方.

所以, 两者概率函数的转化公式, 就是:

$$\boxed{\underbrace{\Phi(x)}_{\text{正态分布的累加函数}}=\underbrace{\Phi_0(\frac{x-\mu}{\sigma})}_{\text{标准正态分布的累加函数}}}$$

例

有 $X\sim N(1,4)$, 即 $\mu=1,\ \sigma^2=4,\ \sigma=2$

到注意: 这只是一个普通的正态分布, 我们必须先把它转成“标准正态分布”再来做.

$\rightarrow$ 求 $P\{0<X<1.6\}=\Phi(1.6)-\Phi(0)$

先把这个累加函数 $\Phi(x)$ (正态分布的), 转成 $\Phi_0(x)$ (标准正态分布的). 套用转化公式, 就有:

$\Phi(1.6)=\Phi_0\left(\frac{x-\mu}{\sigma}\right)=\Phi_0\left(\frac{1.6-1}{2}\right)=\Phi_0(0.3)$

$\Phi(0)=\Phi_0\left(\frac{x-\mu}{\sigma}\right)=\Phi_0\left(\frac{0-1}{2}\right)=\Phi_0(-0.5)$

所以回到原题, $\Phi(1.6)-\Phi(0)=\Phi_0(0.3)-\underbrace{\Phi_0(-0.5)}_{=1-\Phi_0(0.5)}$

→ 求 $P\{|X|\le 2\}$
方法 1:

$$\begin{aligned}P\{|X|\le 2\} &= P\{-2\le X\le 2\}\\ &= \Phi(2)-\Phi(-2) \ \leftarrow \text{转成“标准正态分布”的累加函数}\\ &= \Phi_0\left(\frac{2-\overbrace{1}^{\mu}}{\underbrace{1}_{\sigma}}\right)-\Phi_0\left(\frac{-2-1}{1}\right)\\ &= \Phi_0(0.5)-\Phi_0(-1.5)\\ &= \Phi_0(0.5)-[1-\Phi_0(1.5)]\end{aligned}$$

方法 2:

$$\begin{aligned}P\{|X|\le 2\} &= P\{-2\le X\le 2\}\\ &= P\{(-2-1)\le (X-1)\le (2-1)\} \ \leftarrow \text{对}X\text{左中右, 同时减去}\mu\text{值1}\\ &= P\left\{\frac{-2-1}{2}\le \frac{X-\overbrace{1}^{\mu}}{\underbrace{2}_{\sigma}}\le \frac{2-1}{2}\right\} \ \leftarrow \text{对}X\text{左中右, 同时除以}\sigma\text{值2}\end{aligned}$$

注意, 现在中间的$\frac{X-1}{2}$,就已经被我们转化成了“标准正态分布”的 x 形式了, 即从x转成了$\frac{x-\mu}{\sigma}$

$$\begin{aligned}&= P\left\{-1.5\le \frac{X-1}{2}\le 0.5\right\}\\ &= \Phi_0(0.5)-\Phi_0(-1.5)\end{aligned}$$

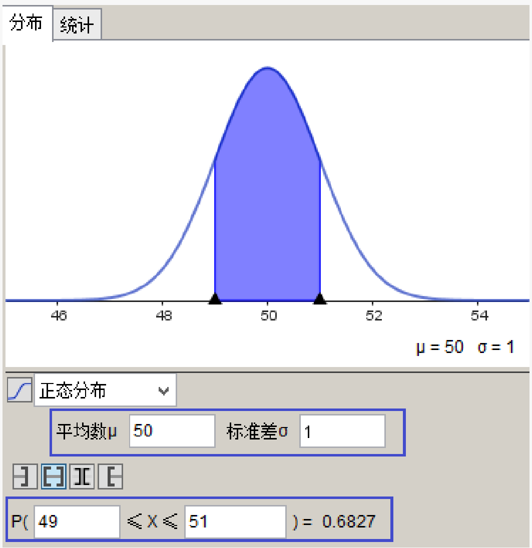
例

飞船零件, 尺寸符合“正态分布” $X\sim N\left(\underbrace{50}_{\mu},\underbrace{1}_{\sigma^2}\right)$, 即 $\sigma=1$

其尺寸只有在 50 ± 1 时, 才是合格的.

那么问:
→ 生产出的单个飞船零件, 合格概率是?

$$\begin{aligned}P\{49\le X\le 51\} &= \Phi(51)-\Phi(49)\\ &= \Phi_0\left(\frac{51-\overbrace{50}^{\mu}}{\underbrace{1}_{\sigma}}\right)-\Phi_0\left(\frac{49-50}{1}\right)\\ &= \Phi_0(1)-\Phi_0(-1)\\ &= \Phi_0(1)-(1-\Phi_0(1))=0.68269\end{aligned}$$



→ 重复抽检 3 次, 至少有 1 个零件是合格的概率?

$$\begin{aligned} P \underbrace{\{Y \geq 1\}}_{\text{我们令}Y\text{表示合格品的数量}} &= 1 - \underbrace{P\{Y = 0\}}_{\text{所抽的 3 件都不合格 的概率}} \\ &= 1 - \Bigg(1 - \underbrace{0.6827}_{\text{单个产品的合格率}}\Bigg)^3 = 0.968054 \end{aligned}$$

例

X 是服从正态分布的, 即 $X \sim N(\mu, \sigma^2)$, 求:

→

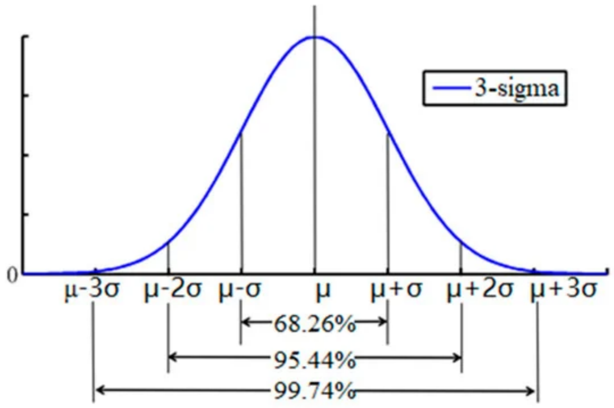
$$\begin{aligned} P\{|X-\mu| < \sigma\} &= P\{-\sigma < X - \mu < \sigma\} \\ &= P\{-\sigma + \mu < X < \sigma + \mu\} \\ &= \Phi\underbrace{(\sigma+\mu)} - \Phi\underbrace{(-\sigma+\mu)} \leftarrow \text{先转成“标准正态分布” 的}\Phi_0 \\ &= \Phi_0\left(\frac{\overbrace{(\sigma+\mu)} - \mu}{\sigma}\right) - \Phi_0\left(\frac{\overbrace{(-\sigma+\mu)} - \mu}{\sigma}\right) \\ &= \Phi_0(1) - \Phi_0(-1) \\ &= \Phi_0(1) - [1 - \Phi_0(1)] = 0.6827 \end{aligned}$$

→ $P\{|X-\mu| < 2\sigma\}$ =? 根据上面的方法, 同理 $= 0.9544$
→ $P\{|X-\mu| < 3\sigma\}$ =? 根据上面的方法, 同理 $= 0.9974$

这个例题, 就引出了“3 准则”.

4 3σ 准则 (pauta criterion)

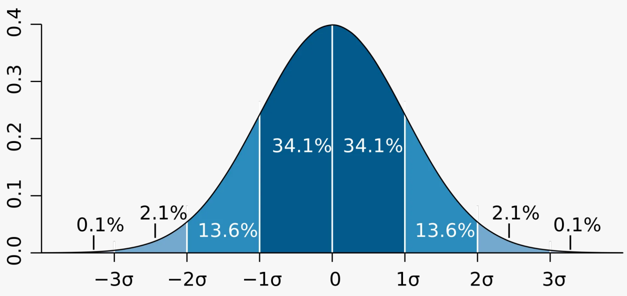
落在 $|3\sigma|$ 区间段内的概率, 是 0.99



例

一些软件, 告诉你开机时间多少秒, 打败了全国 97% 的用户, 这个是怎么算出来的? 就是利用了 "正态分布模型".

只要随机抽取一部分用户的开机数据，算出 "均值 μ" 和 "标准差 ", 就可以确定一条正态分布曲线. 在正态分布中，一个标准差, 覆盖 68.26% 的数据; 两个标准差, 覆盖 95.44% 的数据. 软件**只需要比较你的 "开机时间" 离 "均值 μ" 的差距，就能知道你距离均值多少个标准差 ， 也就知道你的排名了.**



所以, 正态分布，为我们提供了一个 "估算个体在整体中位置" 的便捷方法. 像智商、身高、考试成绩，只要服从 "正态分布", 都能使用这种方法, 快速得到答案.