9 一元高斯分布

Univariate Gaussian Distribution

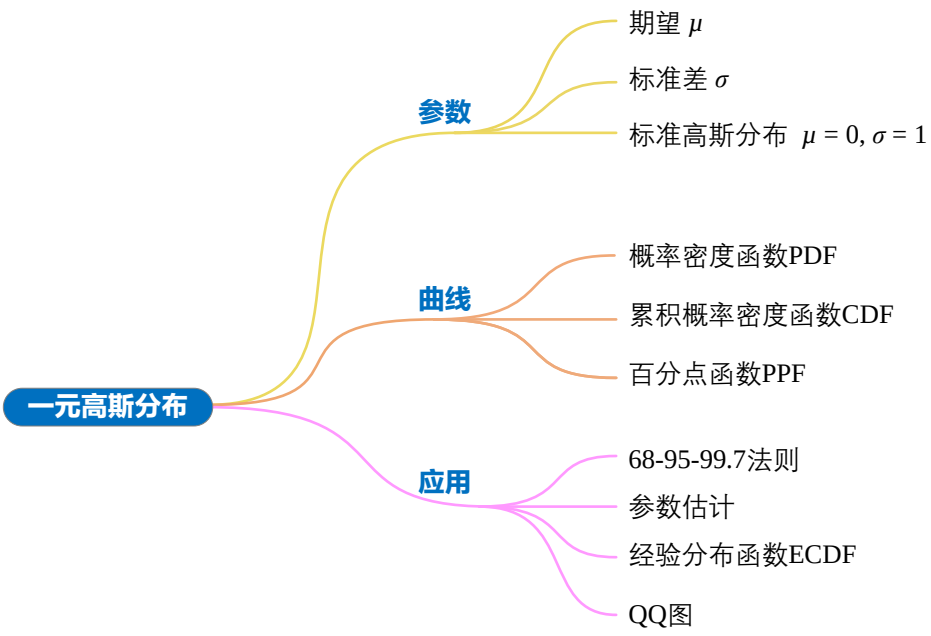
可能是应用最广泛的概率分布

数学家站在彼此的肩膀上。

Mathematicians stand on each other's shoulders.

—— 卡尔·弗里德里希·高斯 (Carl Friedrich Gauss) | 德国数学家、物理学家、天文学家 | 1777 ~ 1855

- `matplotlib.pyplot.axhline()` 绘制水平线
- `matplotlib.pyplot.axvline()` 绘制竖直线
- `matplotlib.pyplot.contour()` 绘制等高线图
- `matplotlib.pyplot.contourf()` 绘制填充等高线图
- `numpy.ceil()` 计算向上取整
- `numpy.copy()` 深拷贝数组，对新生成的对象修改删除操作不会影响到原对象
- `numpy.cumsum()` 计算累积和
- `numpy.floor()` 计算向下取整
- `numpy.meshgrid()` 生成网格数据
- `numpy.random.normal()` 生成满足高斯分布的随机数
- `scipy.stats.norm.cdf()` 高斯分布累积分布函数 CDF
- `scipy.stats.norm.pdf()` 高斯分布概率密度函数 PDF
- `scipy.stats.norm.ppf()` 高斯分布百分点函数 PPF



9.1 一元高斯分布：期望值决定位置，标准差决定形状

回顾上一章介绍**一元高斯分布** (univariate normal distribution)，其概率密度函数 PDF 如下：

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \tag{1}$$

其中，μ 为期望值，σ 为标准差。

期望值

一元高斯分布概率密度函数的形状为中间高两边低的钟形，其 PDF 最大值位于 $x = \mu$。

本书前文提过，一元高斯分布的概率密度函数以 $x = \mu$ 为轴左右对称，曲线向左右两侧远离 $x = \mu$ 呈逐渐均匀下降趋势，曲线两端与横轴 $y = 0$ 无限接近，但永不相交。

图 1 所示为 μ 对一元高斯分布 PDF 曲线位置的影响。

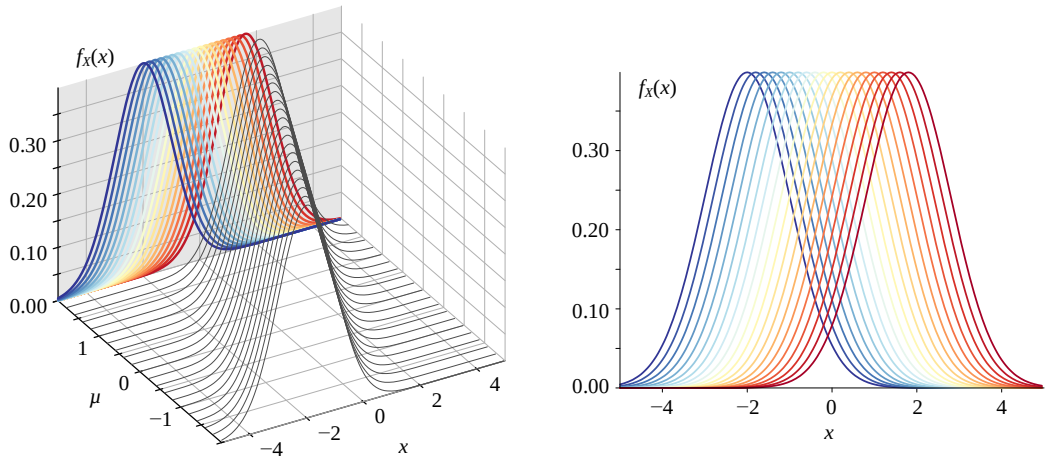


图 1. μ 对一元高斯分布 PDF 曲线位置的影响

标准差

σ 也称为高斯分布的形状参数，σ 越大，曲线越扁平；反之，σ 越小，曲线越瘦高。从数据角度，σ 描述数据分布的离散程度，σ 越大，数据分布越分散，σ 越小，数据分布越集中。图 2 所示为 σ 对一元高斯分布 PDF 曲线形状影响。

本书前文强调过，期望值、标准差的单位和随机变量的单位相同。因此，直方图、概率密度图上常常出现 $\mu \pm \sigma$、$\mu \pm 2\sigma$、$\mu \pm 3\sigma$ 等等。

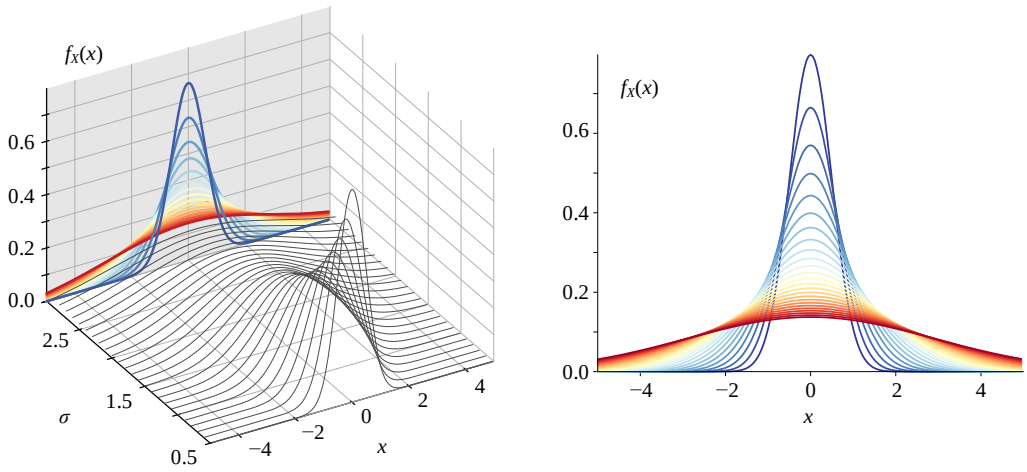


图 2. σ 对一元高斯分布 PDF 曲线形状影响

Bk5_Ch09_01.py 绘制图 1。请大家修改代码自行绘制图 2。代码自定义函数计算一元高斯分布概率密度，大家也可以使用 scipy.stats.norm.pdf() 函数获得一元高斯分布密度函数值。

在 Bk5_Ch09_01.py 基础上，我们用 Streamlit 制作了一个应用，大家可以改变 μ、σ 参数值，观察一元高斯 PDF 曲线变化。请大家参考 Streamlit_ Bk5_Ch09_01.py。

9.2 累积概率密度：对应概率值

一元高斯分布的累积概率密度函数 CDF：

$$F_X(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{-1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right)\mathrm{d}t \tag{2}$$

上式也可以用误差函数 erf() 来表达：

$$F_X(x)=\frac{1}{2}\left[1+\mathrm{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right] \tag{3}$$

《数学要素》第 18 章介绍过误差函数，请大家回顾。

期望值

图 3所示为 μ 对一元高斯分布 CDF 曲线位置的影响。随着 x 不断靠近−∞，CDF 取值不断接近于 0，但不等于 0；反之，随着 x 不断靠近+∞，CDF 取值不断接近于 1，但不等于 1。

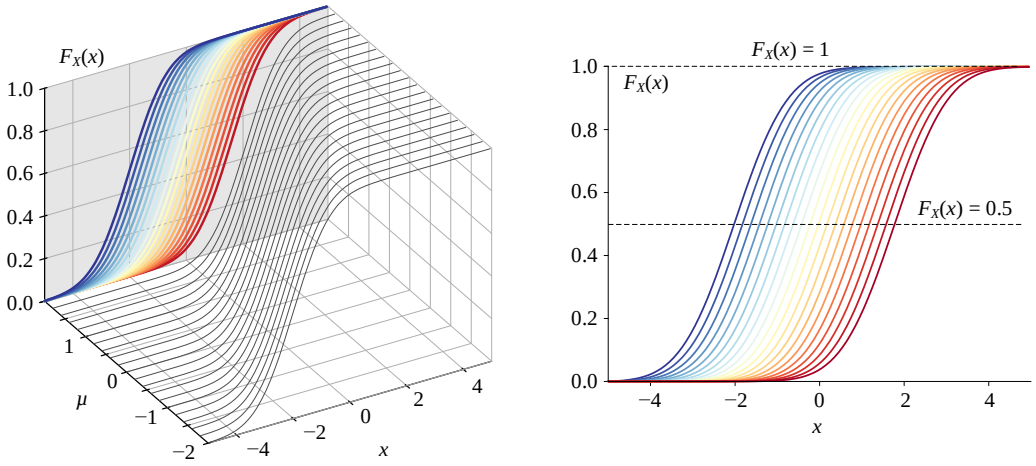


图 3. μ 对一元高斯分布 CDF 曲线位置的影响

标准差

图 4 所示为 σ 对一元高斯分布 CDF 曲线形状影响。σ 越小，CDF 曲线越陡峭；σ 越大，越平缓。从另外一个角度看一元高斯分布 CDF 曲线，它将位于实数轴 $(-\infty, +\infty)$ 之间的 x 转化为 $(0, 1)$ 之间的某个值，而这个值恰好对应一个概率。

注意，图 1、图 2 的竖轴对应概率密度值，而图 3、图 4 的竖轴对应概率值。

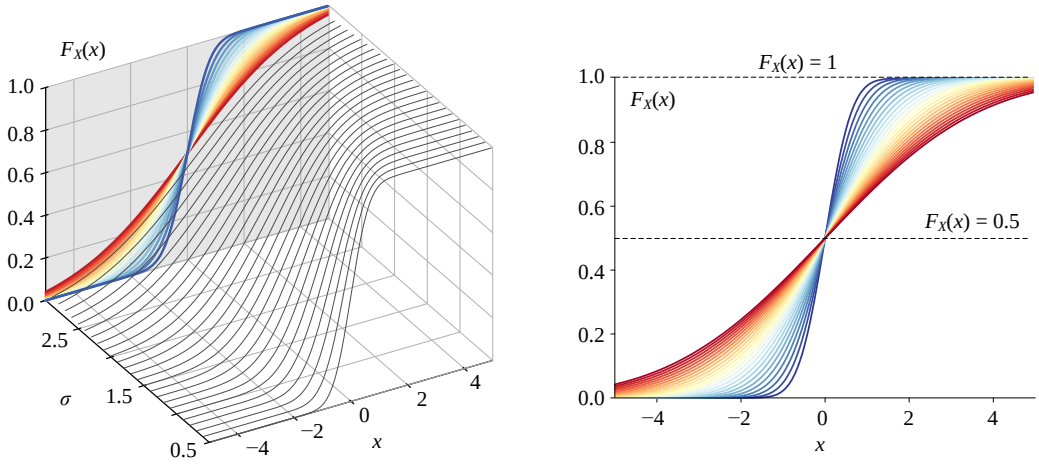


图 4. σ 对一元高斯分布 CDF 曲线形状影响

Bk5_Ch08_02.py 绘制图 3 和图 4。

PDF vs CDF

图 5 比较标准正态分布 $N(0, 1)$ 的 PDF 和 CDF 曲线。虽然两条曲线画在同一幅图上，它们的 y 轴数值的含义完全不同。对于 PDF 曲线，它的 y 轴数值代表概率密度，并不是概率值。而 CDF 曲线的 y 轴数值则代表概率值。

给定一点 x，图 5 中背景为蓝色区域面积对应 $F_X(x)=\int_{-\infty}^{x} f_X(t)\mathrm{d}t$，也就是 CDF 曲线的高度值。下一节还会继续讲解标准正态分布。

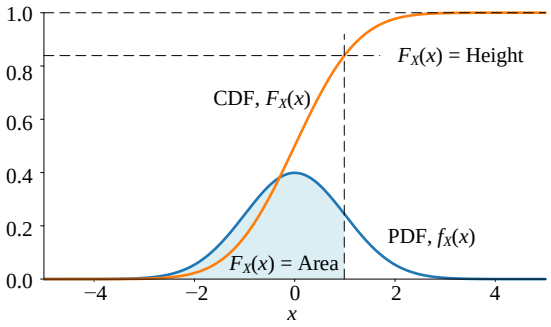


图 5. 比较标准正态分布的 PDF 和 CDF 曲线

百分点函数 PPF

我们把 Percent-Point Function (PPF) 直译为“**百分点函数**”。实际上，百分点函数 PPF 是 **CDF 的逆函数** (inverse CDF)。

如图 6 所示，给定 x，我们可以通过 CDF 曲线得到累积概率值 $F_X(x) = p$。而 PPF 曲线则正好相反，给定概率值 p，通过 PPF 曲线得到 x，即 $F_X^{-1}(p)=x$。

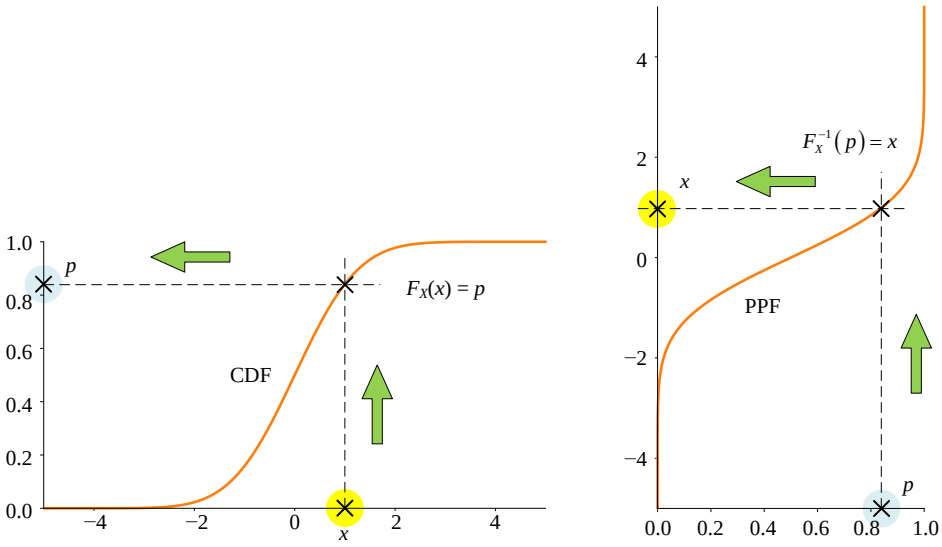


图 6. CDF 曲线和 PPF 曲线之间关系

9.3 标准高斯分布：期望为 0，标准差为 1

当 $\mu = 0$ 且 $\sigma = 1$ 时，高斯分布为**标准正态分布** (standard normal distribution)，记做 $N(0, 1)$。

本节用 Z 表示服从标准正态分布的连续随机变量，而 Z 的实数取值用 z 代表。因此，标准正态分布的 PDF 函数为：

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-z^2}{2}\right) \tag{4}$$

可以写成，$Z \sim N(0, 1)$。

图 7 (a) 所示为标准正态分布 PDF 曲线。特别地，$Z = 0$ 时，标准高斯分布的概率密度值为：

$$f_Z(0) = \frac{1}{\sqrt{2\pi}} \approx 0.39894 \tag{5}$$

这个值经常近似为 0.4。

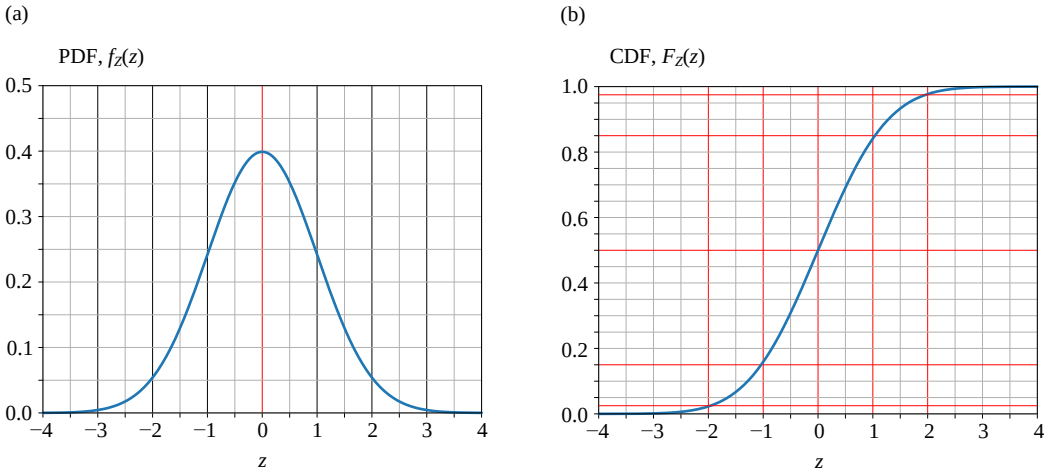


图 7. 标准高斯分布 PDF 和 CDF 曲线

容易发现，当 PDF 曲线随着 x 增大而增大时，PDF 的增幅先是逐渐变大，曲线逐渐变陡；然后，PDF 的增幅放缓，曲线坡度逐渐变得平缓，在 $x = \mu$ 曲线坡度为 0。

从一阶导数的角度来看，这一段，一阶导数值大于 0，直到 $x = \mu$ 处，一阶导数值为 0。然而，这段变化过程，二阶导数先为正，然后为负值，中间穿过 0。而 PDF 曲线二阶导数为 0 正好对应 $\mu \pm \sigma$ 这两点，这两点正是 PDF 曲线的拐点。

图 8 所示为标准正态分布 $N(0, 1)$ 的 CDF、PDF、PDF 一阶导数、PDF 二阶导数这四条曲线之。其中，黑色 × 对应 PDF 曲线的最大值处。红色 × 对应 PDF 曲线拐点。请大家自行分析这四幅图像中曲线的变化趋势。

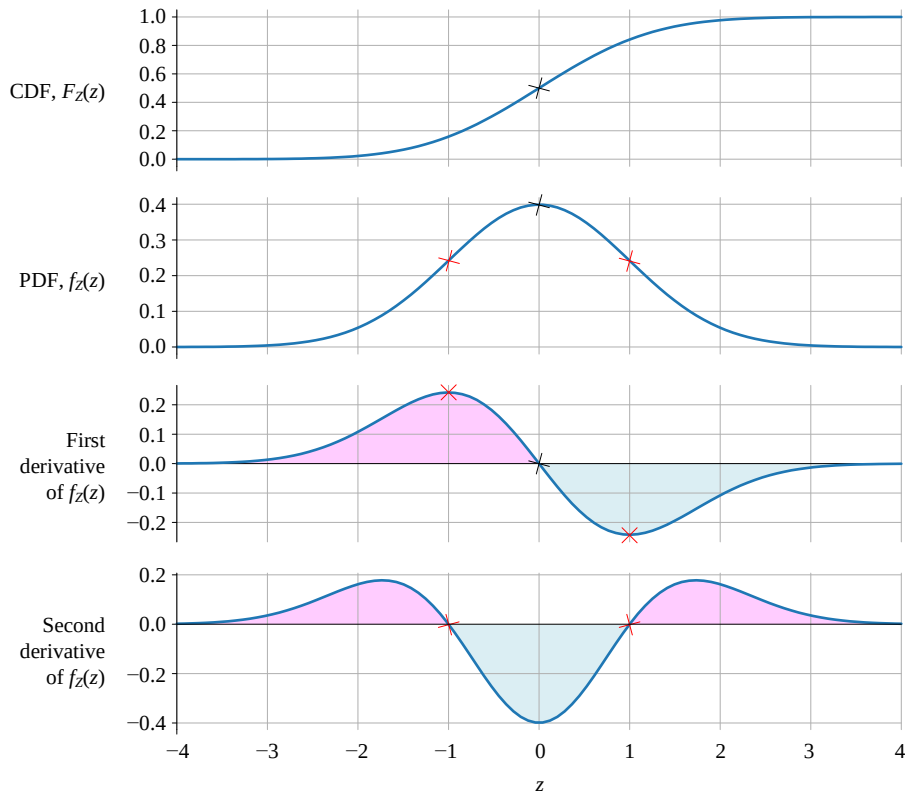


图 8. 四条曲线：标准正态分布 CDF、PDF、PDF 一阶导数、PDF 二阶导数

Z 分数：一种以标准差为单位的度量尺度

Z 分数 (Z-score)，也叫**标准分数** (standard score) 是样本值 x 与平均数 μ 的差再除以标准差 σ 的结果，对应的运算为：

$$z = \frac{x - \mu}{\sigma} \tag{6}$$

上述过程也叫做数据的**标准化** (standardize)。(6) 代表数据点 x 和均值 μ 之间的距离为 z 倍标准差。

本书前文强调，标准差和 x 具有相同的单位，而 (6) 分式消去了单位，这说明 Z 分数**无单位** (unitless)。样本数据的 Z 分数构成的分布有两个特点：a) 平均等于 0；b) 标准差等于 1。

⚠ 注意，本书把“normalize”翻译为“归一化”，它表示将一组数据转化为 [0, 1] 区间的数值。线性代数中，**向量单位化** (vector normalization) 指的是将非零向量转化成 L^2 模为 1 的单位向量。很多参考资料混用“standardize”和“normalize”，请大家注意区分。

图 9 所示为标准正态分布随机变量 z 值和 PDF 的对应关系。图 10 所示为标准正态分布 z 值到 CDF 值的映射关系。图 11 所示为 PPF 值到标准正态分布 z 值的映射关系。本章前文介绍过，CDF 和 PPF 互为反函数。

图 12 所示为标准正态分布中，不同 z 值对应的三类面积。我们一般会在 **Z 检验** (Z test) 中用到这个表。

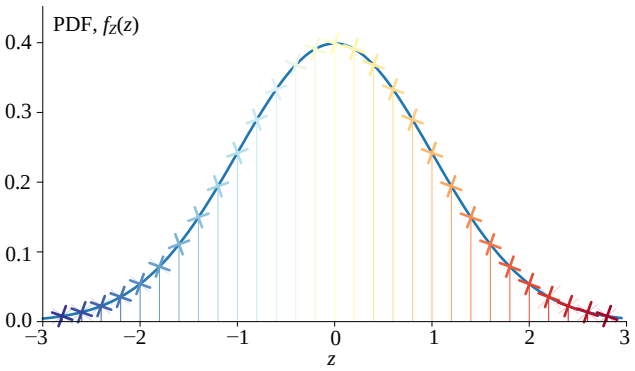


图 9. 标准正态分布随机变量取值 z 和 PDF 的对应关系

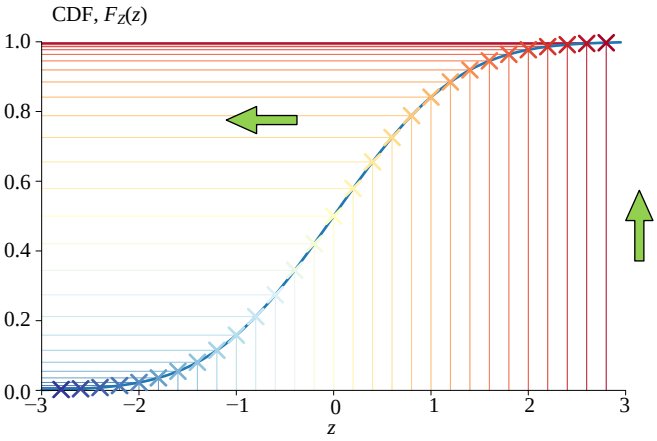


图 10. 标准正态分布随机变量取值 z 和 CDF 值的映射关系

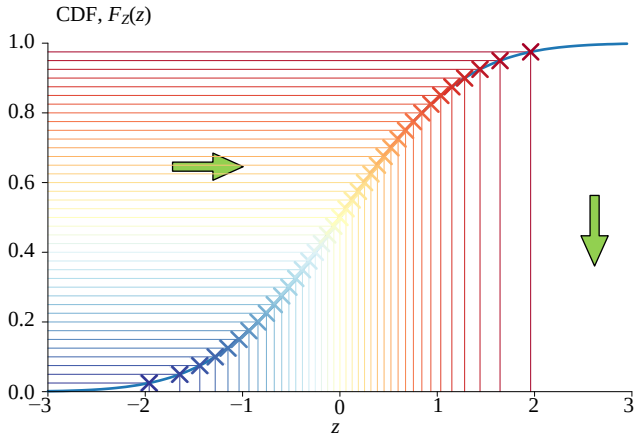
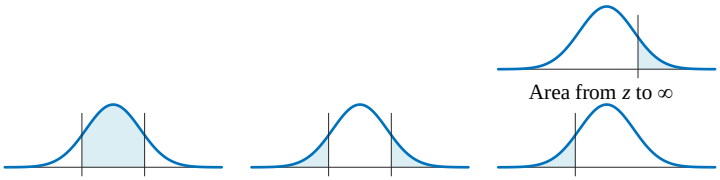
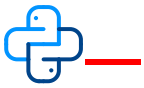


图 11. 标准正态分布随机变量取值 z 和 PPF 值的映射关系



z	Area from −z to z	Area from −∞ to −z and z to ∞	Area from −∞ to −z
1.64485	0.9	0.1	0.05
1.66959	0.905	0.095	0.0475
1.69540	0.91	0.09	0.045
1.72238	0.915	0.085	0.0425
1.75069	0.92	0.08	0.04
1.78046	0.925	0.075	0.0375
1.81191	0.93	0.07	0.035
1.84526	0.935	0.065	0.0325
1.88079	0.94	0.06	0.03
1.91888	0.945	0.055	0.0275
1.95996	0.95	0.05	0.025
2.00465	0.955	0.045	0.0225
2.05375	0.96	0.04	0.02
2.10836	0.965	0.035	0.0175
2.17009	0.97	0.03	0.015
2.24140	0.975	0.025	0.0125
2.32635	0.98	0.02	0.01
2.43238	0.985	0.015	0.0075
2.57583	0.99	0.01	0.005
2.80703	0.995	0.005	0.0025
3.29053	0.999	0.001	0.0005

图 12. 标准正态分布中，不同 z 值对应的四类面积



Bk5_Ch08_03.py 绘制本节之前大部分图像。

以鸢尾花数据为例

前文提过，Z 分数本质上代表一种距离，是标准化的“距离度量”。原始数据的 Z 分数代表距离均值若干倍的标准差偏移。比如，某个数据点的 Z 分数为 3，说明这个数据距离均值 3 倍标准差偏移。Z 分数的正负表达偏移的方向。

任何尺度，甚至任何单位的样本数据经过标准化后可以比较大小。这样不同分布、不同单位的数据有了可比性。

图 13 所示为鸢尾花样本数据四个特征的 z 分数。

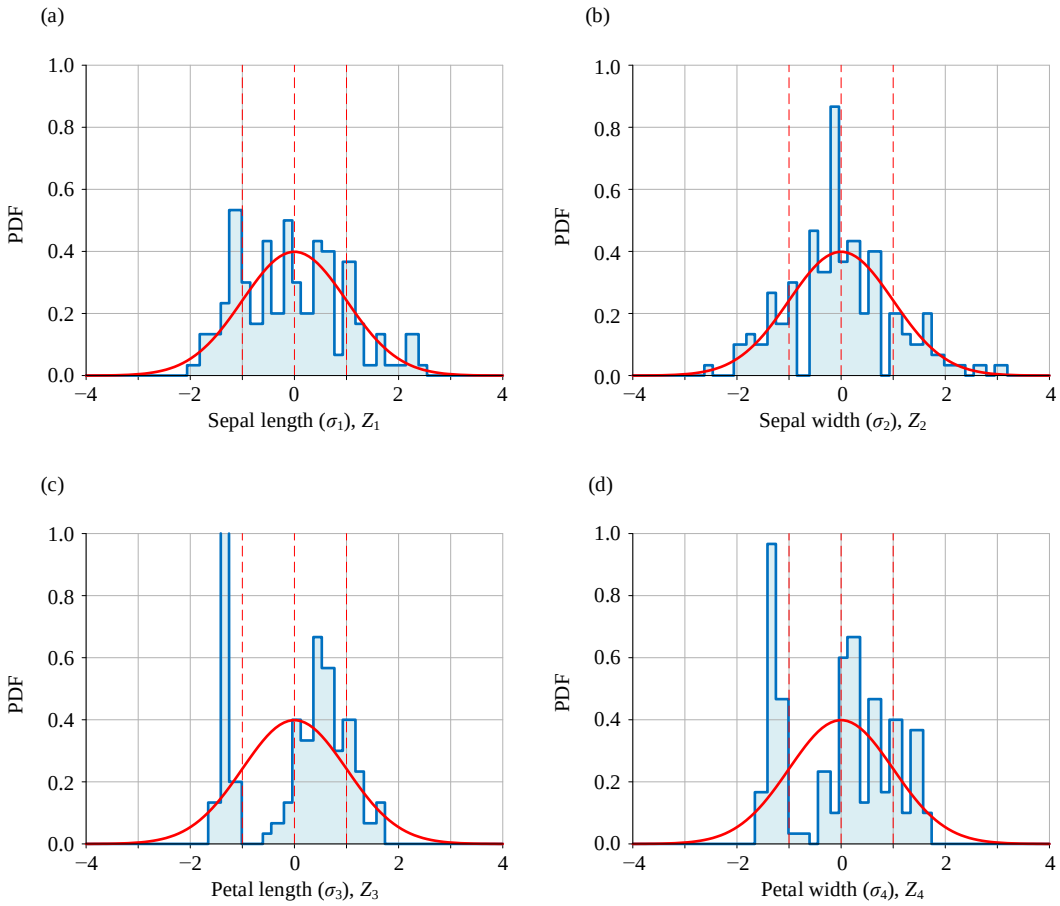


图 13. 鸢尾花四个特征的 z 分数，标准差距离

9.4 68-95-99.7 法则

一元高斯分布有所谓的 **68–95–99.7 法则** (68-95-99.7 Rule)，具体是指一组近乎满足正态分布的样本数据，约 68.3%、95.4%和 99.7%样本位于距平均值正负 1 个、2 个和 3 个标准差范围之内。

标准正态分布 $N(0, 1)$

以标准正态分布 $N(0, 1)$ 为例，整条标准正态分布曲线和横轴包裹的面积为 1。

如图 14 (a) 所示，$[-1, 1]$ 区间内，标准正态分布和横轴包裹的区域面积约为 0.68，即 68%。

如图 14 (b) 所示，$[-2, 2]$ 区间对应的阴影区域面积约为 0.95，即 95%。

如图 14 (c) 所示，$[-3, 3]$ 区间对应的阴影区域面积约为 0.997，即 99.7%。

写成具体的概率运算：

$$\begin{aligned} \Pr(-1 \le Z \le 1) &\approx 0.68 \\ \Pr(-2 \le Z \le 2) &\approx 0.95 \\ \Pr(-3 \le Z \le 3) &\approx 0.997 \end{aligned} \tag{7}$$

图 15 所示为标准正态分布 CDF 曲线上 68-95-99.7 法则对应的高度。

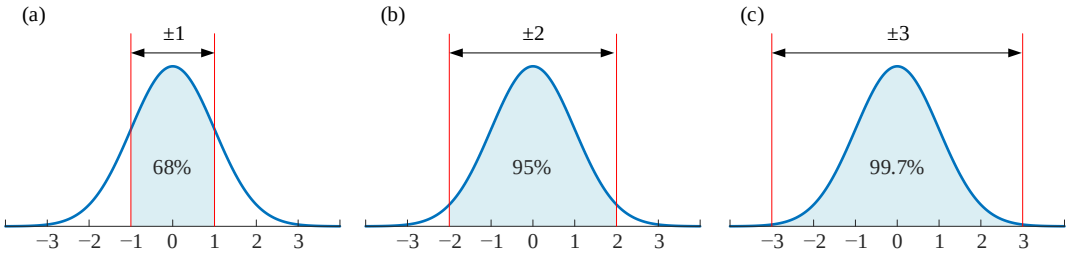


图 14. 68-95-99.7 法则，标准正态分布 PDF

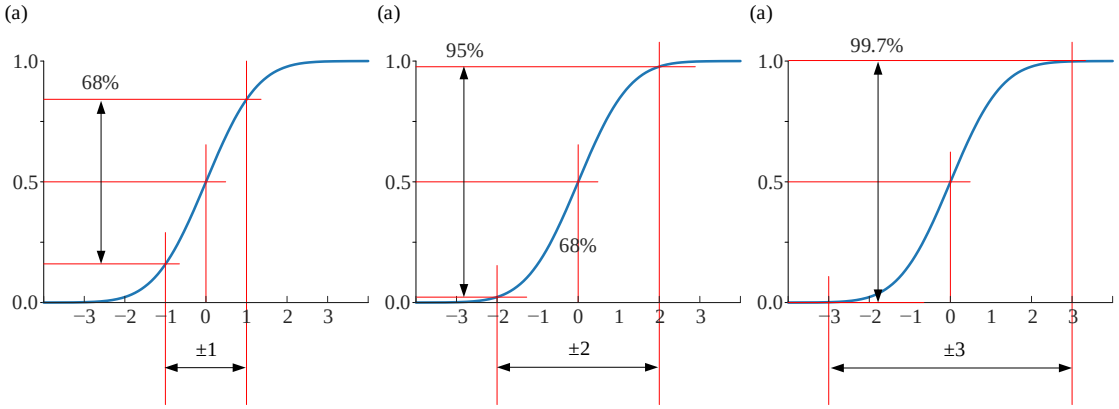


图 15. 68-95-99.7 法则，标准正态分布 CDF

正态分布 $N(\mu, \sigma^2)$

图 16 所示为一般正态分布 $N(\mu, \sigma^2)$ 中 68-95-99.7 法则对应的位置：

$$\begin{aligned} &\Pr\left(\mu-\sigma \le X \le \mu+\sigma\right) \approx 0.68 \\ &\Pr\left(\mu-2\sigma \le X \le \mu+2\sigma\right) \approx 0.95 \\ &\Pr\left(\mu-3\sigma \le X \le \mu+3\sigma\right) \approx 0.997 \end{aligned} \tag{8}$$

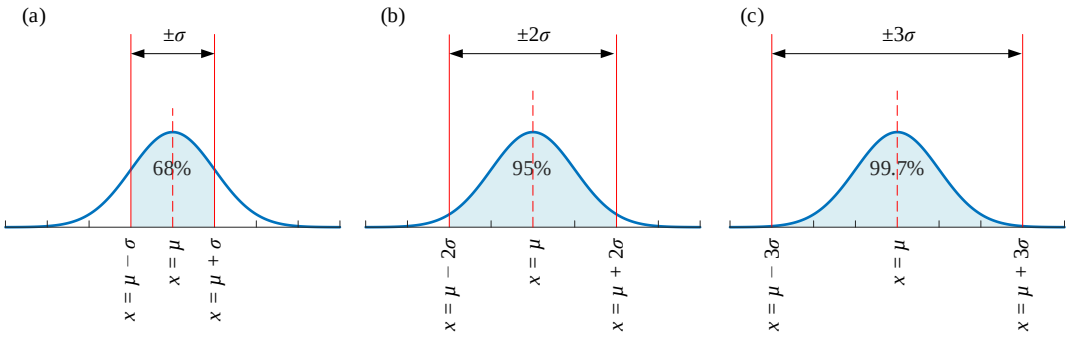


图 16. 68-95-99.7 法则，一般正态分布

和分位的关系

图 17 所示为 68-95-99.7 法则和四分位、十分位、二十分位、百分位关系。

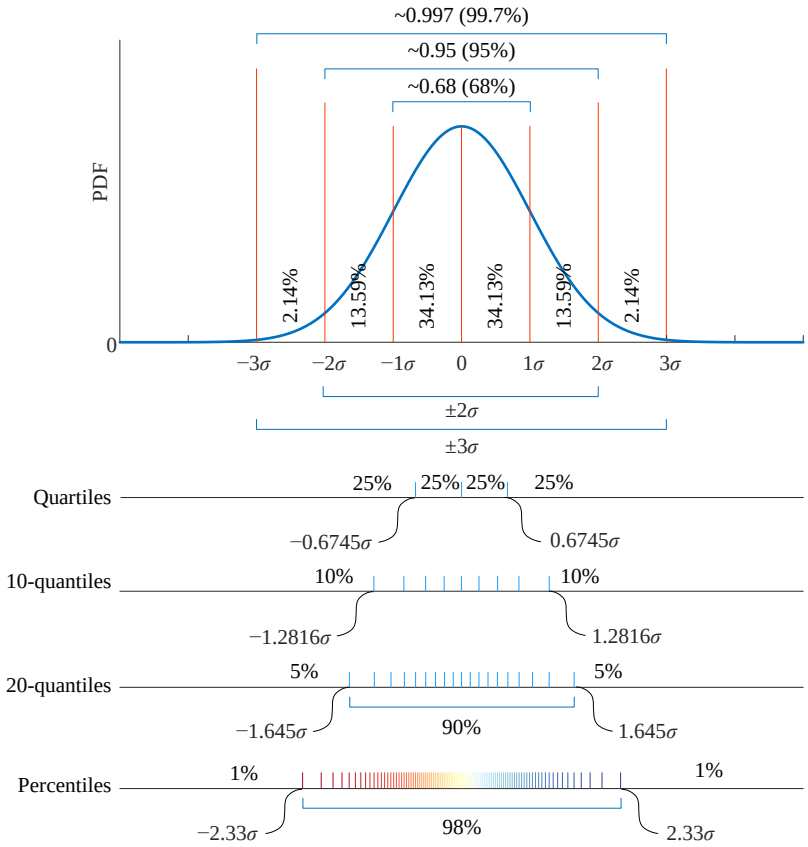


图 17. 68-95-99.7 法则和四分位、十分位、二十分位、百分位关系，注意图中并不区分总体标准差 σ 和样本标准差 s

随机数

如果随机数服从一元高斯分布 $N(\mu, \sigma^2)$，在 $[\mu - \sigma, \mu + \sigma]$ 这个 $\mu \pm \sigma$ 区间内，应该约有 68%的随机数。如图 18 所示，样本一共有 500 个随机数，约 340 个 (= 500 × 68%) 在 $\mu \pm \sigma$ 之内，约 160 个在 $\mu \pm \sigma$ 之外。

在 $[\mu - 2\sigma, \mu + 2\sigma]$ 这个 $\mu \pm 2\sigma$ 区间内，应该约有 95%的随机数。如图 19 所示，样本数还是 500 个，约 475 个 (= 500 × 95%) 在 $\mu \pm 2\sigma$ 之内，约 25 个在 $\mu \pm 2\sigma$ 之外。

68-95-99.7 法则可以帮助大家直观地理解一元高斯分布的形态和特征，即大部分数据集中在均值周围，而远离均值的数据较为稀少。如果一组数据中存在明显偏离均值多个标准差的数据点，就有可能是异常值或者离群值，需要进一步检查和分析。

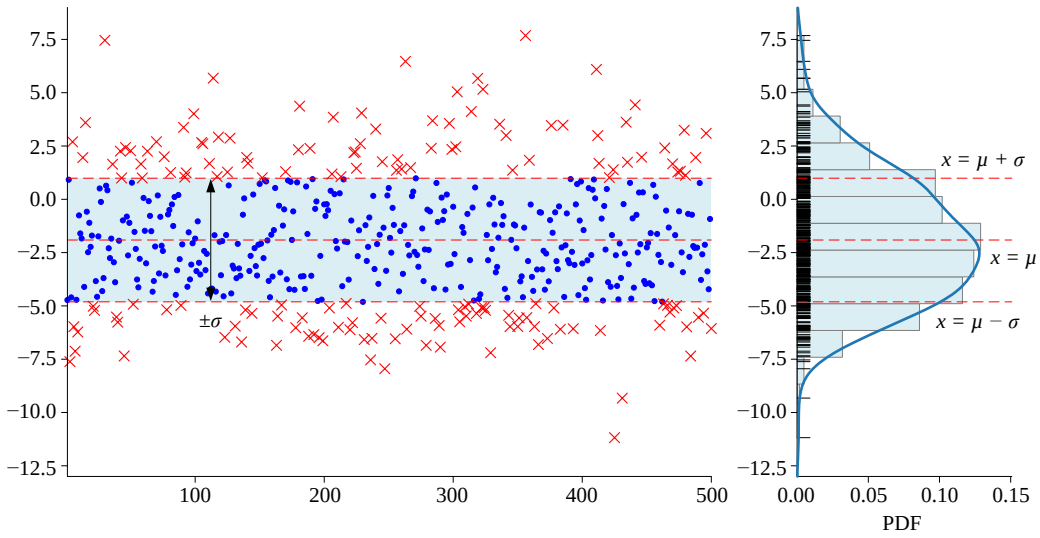


图 18. 500 个随机数和 $\mu \pm \sigma$

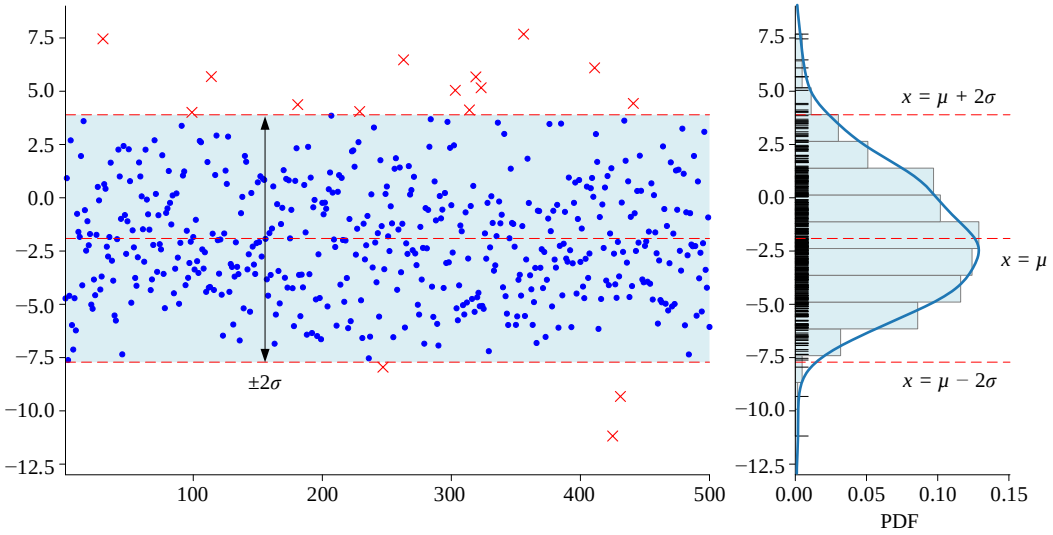


图 19. 500 个随机数和 $\mu \pm 2\sigma$

Bk5_Ch08_04.py 绘制图 18 和图 19。

9.5 用一元高斯分布估计概率密度

概率密度估计：参数估计

在数据科学和机器学习中，**概率密度估计** (probability density estimation) 是经常遇到的一个问题。简单来说，概率密度估计就是从离散的样本数据中估计得到连续的概率密度函数曲线。白话讲，找到一条 PDF 曲线尽可能贴合样本数据分布。

一元高斯分布 PDF 只需要两个参数——均值 (μ)、标准差 (σ)。很多时候，一元高斯分布是估计某个特定特征样本数据分布的一个不错且很便捷的选择。

以鸢尾花数据为例

举个例子，样本数据中花萼长度的均值为 $\mu_1 = 5.843$，标准差为 $\sigma_1 = 0.825$。注意，μ_1 和 σ_1 的单位均为厘米。有了这两个参数，我们可以估计鸢尾花花萼长度随机变量 X_1 概率密度函数可以用如下一元高斯分布估计：

$$f_{X1}(x) = \frac{1}{\sqrt{2\pi}\times 0.825}\exp\left(\frac{-1}{2}\left(\frac{x-5.843}{0.825}\right)^2\right) \tag{9}$$

用这种方法，我们可以得到图 20 所示四条 PDF 曲线。这四条曲线代表用一元高斯分布描述鸢尾花四个特征的概率密度函数。

有了概率密度函数，我们可以回答这样的问题，比如鸢尾花的花萼长度在 [4, 6] cm 区间的概率，利用积分运算就可以得到量化结果。

采用一元高斯分布 PDF 估计给定样本数据单一特征概率密度函数虽然很简单；但是，这种估算方法对应的问题也很明显。图 20 (a) 和 (b) 告诉我们，用高斯分布描述鸢尾花花萼长度和花萼宽度样本数据分布似乎还可以接受。

但是，比较图 20 (c) 和 (d) 中直方图和高斯分布，显然高斯分布不适合描述鸢尾花花瓣长度和宽度样本数据分布。

本书第 18 章将利用核密度估计解决这一问题。

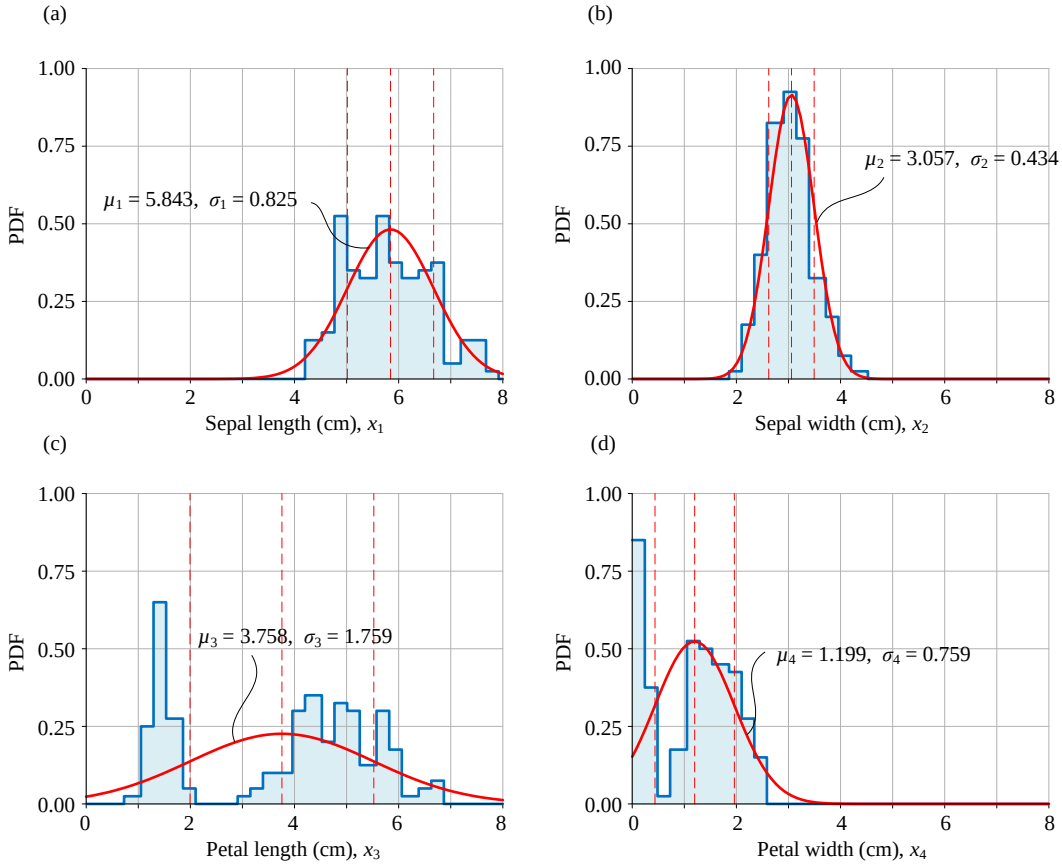


图 20. 比较概率密度直方图和高斯一元分布 PDF

9.6 经验分布函数

经验分布函数 (empirical cumulative distribution function, ECDF) 是用来描述一组样本数据分布情况的统计工具。ECDF 将样本数据按照大小排序，并计算每个数据点对应的累计比例，形成一个类似阶梯函数的曲线，它的纵坐标表示小于等于横坐标的数据比例，横坐标则表示数据的取值。

具体来说，如果有 n 个样本，ECDF 是在所有 n 个数据点上都跳跃 $1/n$ 的阶跃函数。ECDF 也可以用来与理论分布函数进行比较，以检验样本数据是否符合某种假设的分布。

图 21 比较鸢尾花不同特征样本数据的 ECDF 和对应的高斯分布 CDF 曲线。

显然，累积概率函数是一个双射函数。从函数角度来讲，**双射** (bijection) 指的是每一个输入值都有正好一个输出值，并且每一个输出值都有正好一个输入值。

图 22 比较逆经验累积分布函数和高斯分布 PPF 曲线。

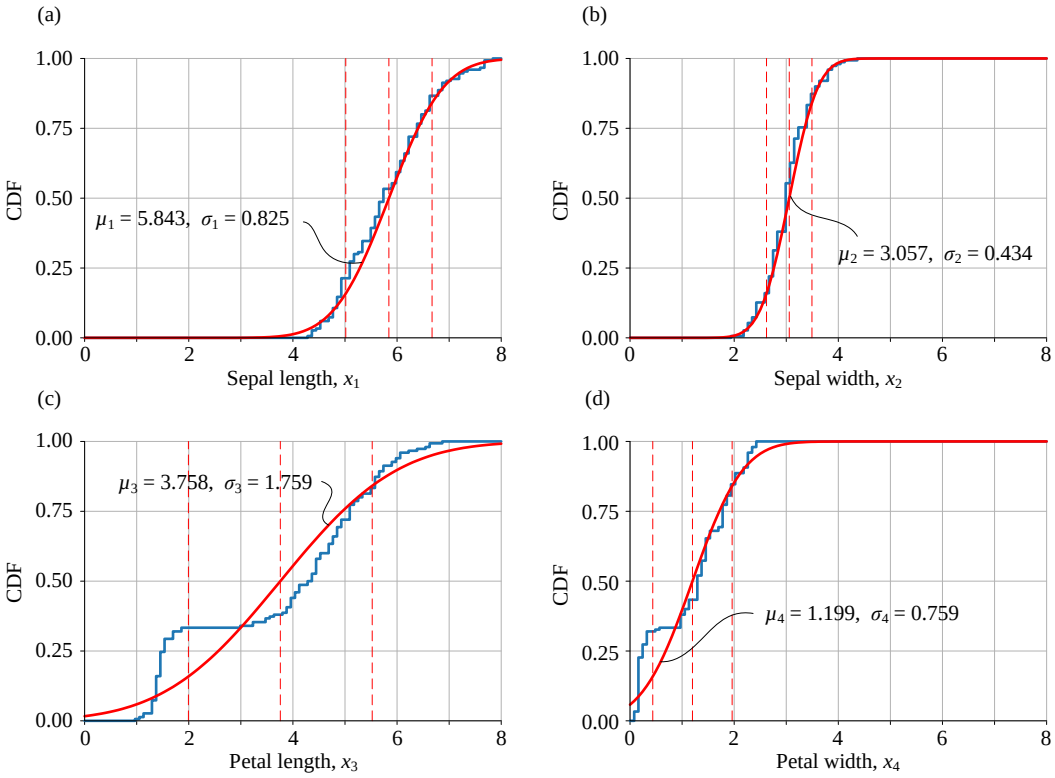


图 21. 比较 ECDF 和高斯 CDF

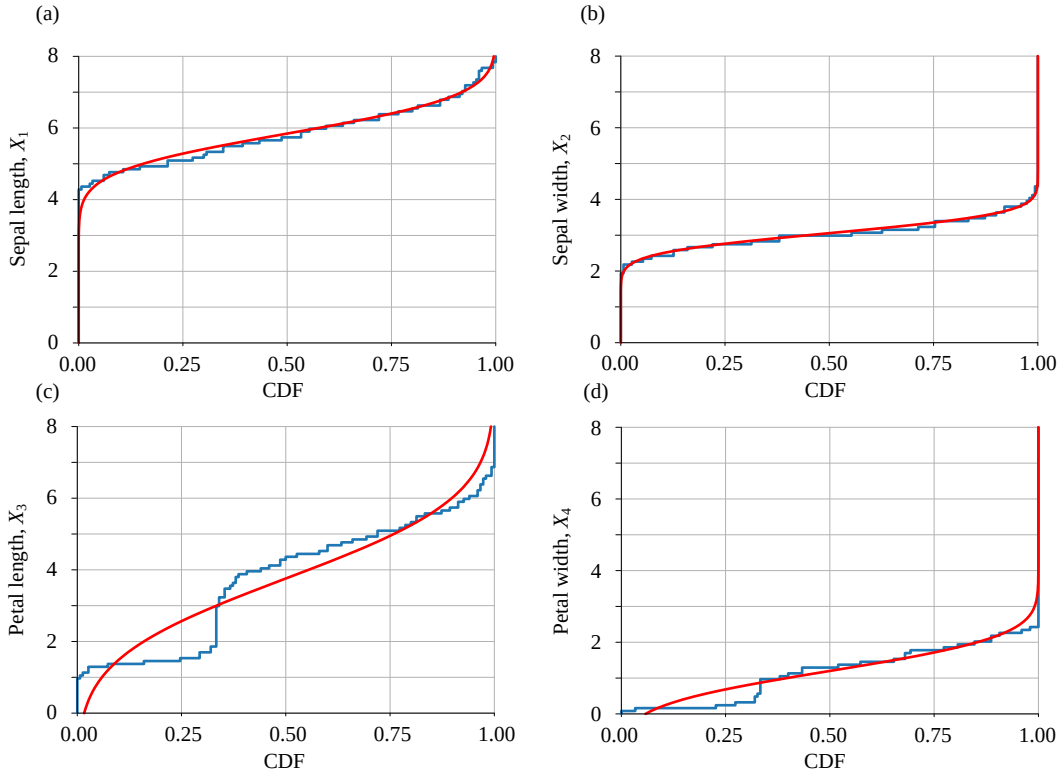


图 22. 逆经验累积分布函数和高斯 PPF

9.7 QQ 图：分位-分位图

QQ 图 (quantile-quantile plot, QQ plot) 中的 Q 代表分位数。QQ 图是散点图，横坐标为某一样本的分位数，纵坐标为另一样本的分位数，横坐标和纵坐标组成的散点图代表同一累计概率所对应的分位数。

如果两分布相似，散点在 QQ 图上趋近于落在一条直线上。QQ 图的横坐标一般是正态分布，当然也可以是其他分布。也就是说，横轴一般是某个理论分布的分位数。

图 23 所示为 QQ 图原理。

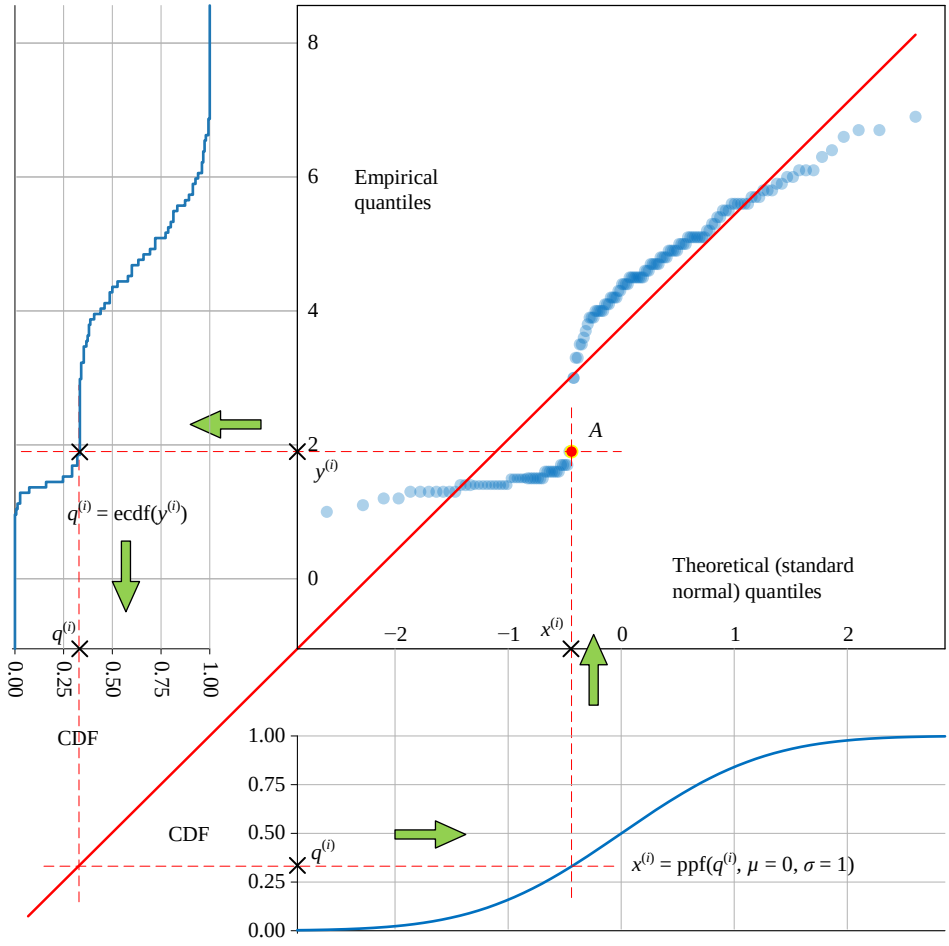


图 23. QQ 图原理，横轴为正态分布

以鸢尾花数据为例

图 24 所示为鸢尾花数据四个特征样本数据的 QQ 图。通过观察这四幅图像，大家应该能够看出那个特征的数据分布更类似 (贴合) 正态分布。

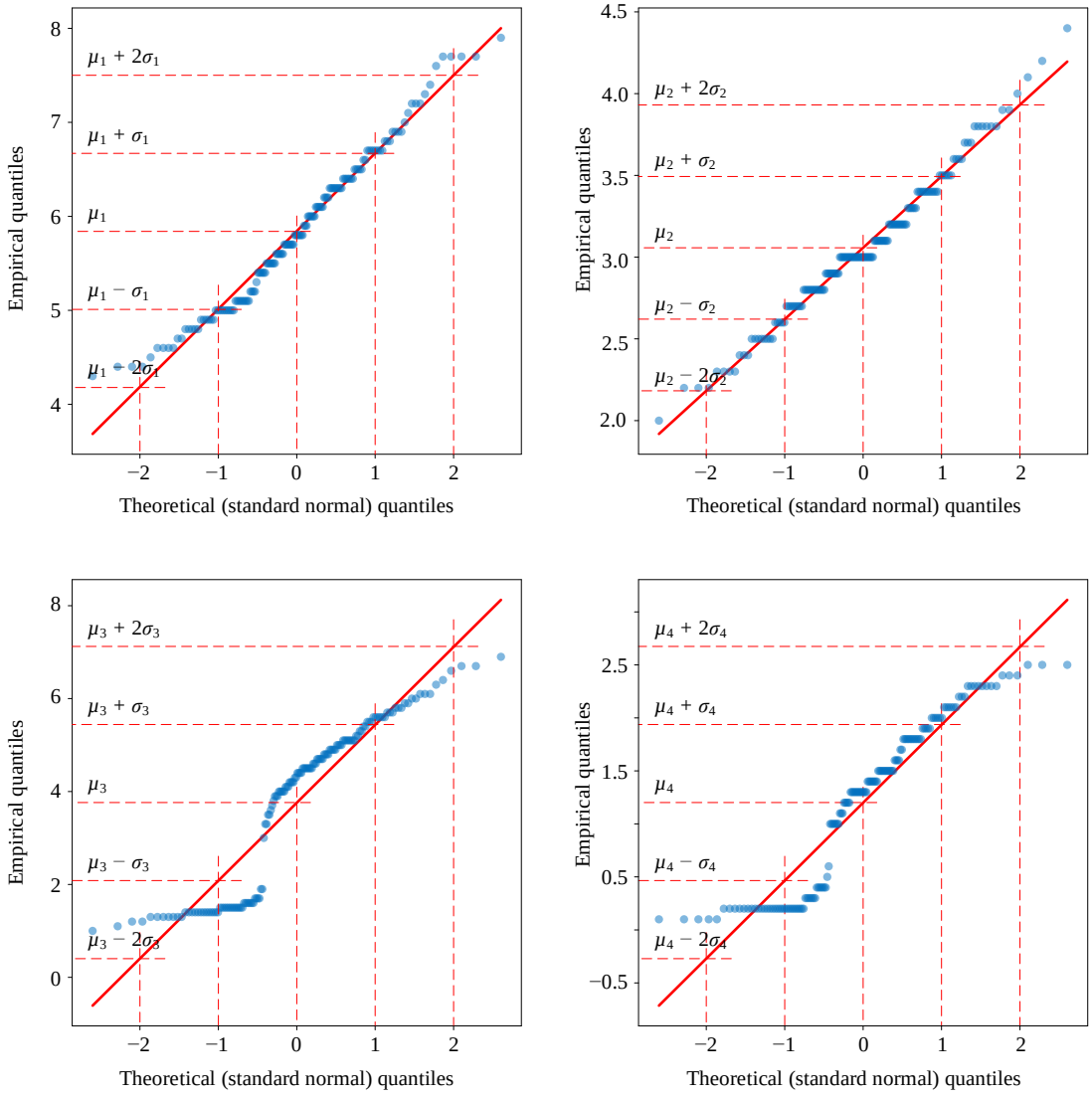


图 24. 鸢尾花数据四个特征样本数据的 QQ 图

Bk5_Ch08_05.py 绘制 8.5 ~ 8.7 节大部分图像。

特殊分布的 QQ 图特征

图 25 所示为几种常见特殊分布对比正态分布的 QQ 图。如图 25 (a) 所示，当样本数据分布近似服从正态分布时，QQ 图中散点几乎在一条直线上。通过散点图的形态，我们还可以判断分布是否有双峰 (图 25 (b))、瘦尾 (图 25 (c))、肥尾 (图 25 (d))、左偏 (图 25 (e))、右偏 (图 25 (f))。

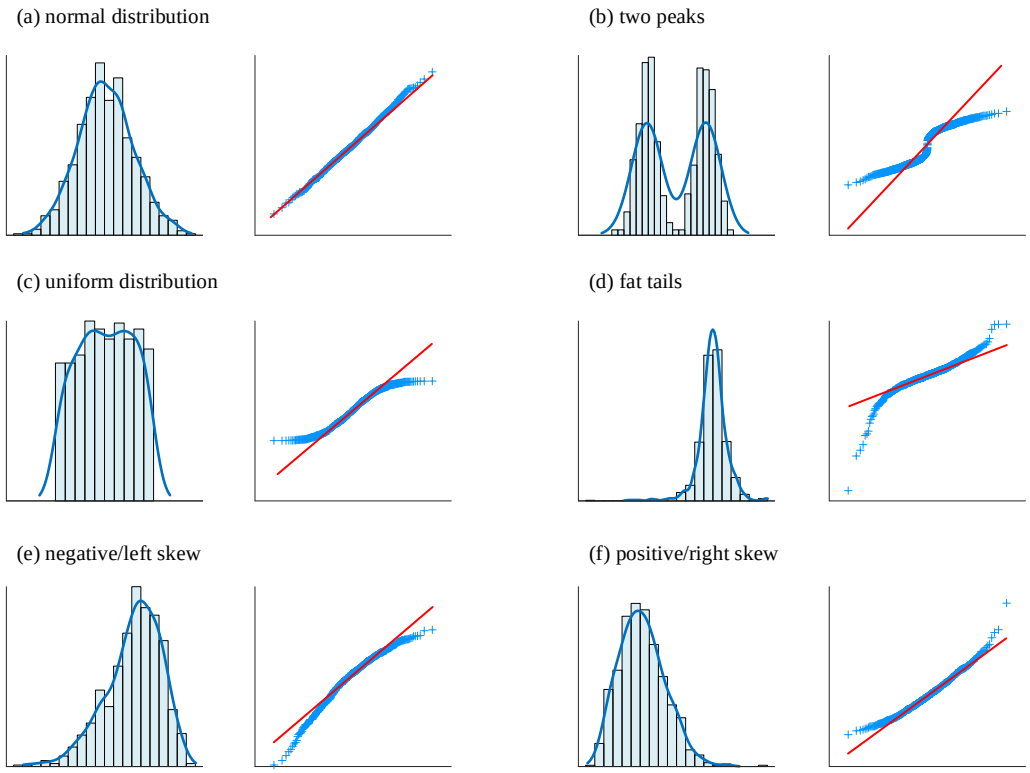


图 25. 几种特殊分布的 QQ 图特点，对比纵坐标正态分布

当然 QQ 图的横轴也可以是其他分布的 CDF。图 26 所示为横轴为均匀分布的 QQ 图，即横轴为理论均匀分布，纵轴为近似均匀分布的样本数据。

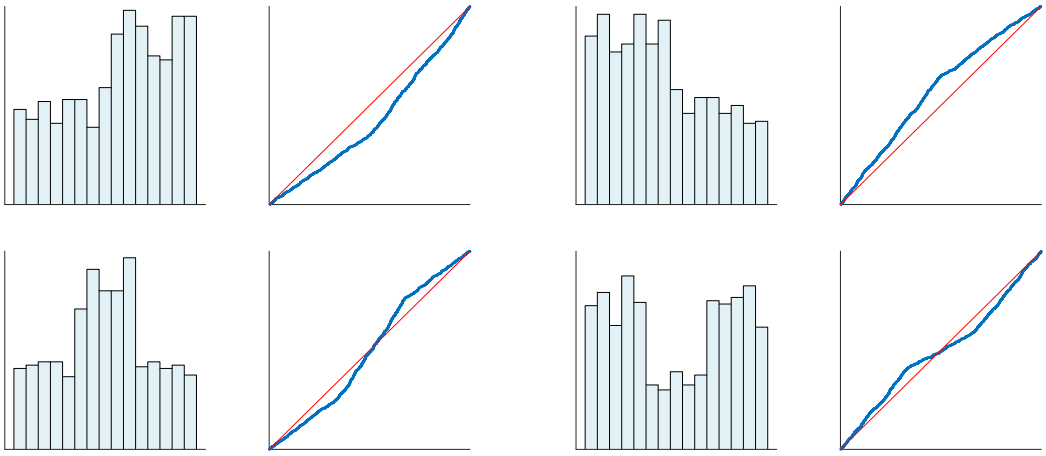


图 26. 几种特殊分布的 QQ 图特点，横轴为理论均匀分布，纵轴为近似均匀分布的样本数据

9.8 从距离到一元高斯分布

现在回过头来再看一元高斯分布的 PDF 解析式：

$$f_X(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \tag{10}$$

而标准正态分布的 PDF 解析式为：

$$f_Z(z)=\frac{1}{\sqrt{2\pi}}\exp\left(\frac{-z^2}{2}\right) \tag{11}$$

几何变换：平移 + 缩放

比较 (1) 和 (4)，我们容易发现满足 $N(\mu, \sigma^2)$ 的 X 可以通过“平移 + 缩放”变成满足 $N(0, 1)$ 的 Z。$X \rightarrow Z$ 对应的运算为：

$$Z=\underbrace{\frac{X-\mu}{\sigma}}_{\text{Scale}}\quad\text{Translate} \tag{12}$$

相反，$Z \rightarrow X$ 对应“缩放 + 平移”：

$$X=\overbrace{Z\sigma+\mu}^{\text{Translate}}\quad\text{Scale} \tag{13}$$

图 27 所示为满足 $N(10, 4)$ 的一元高斯分布通过“平移 + 缩放”变成标准高斯分布的过程。

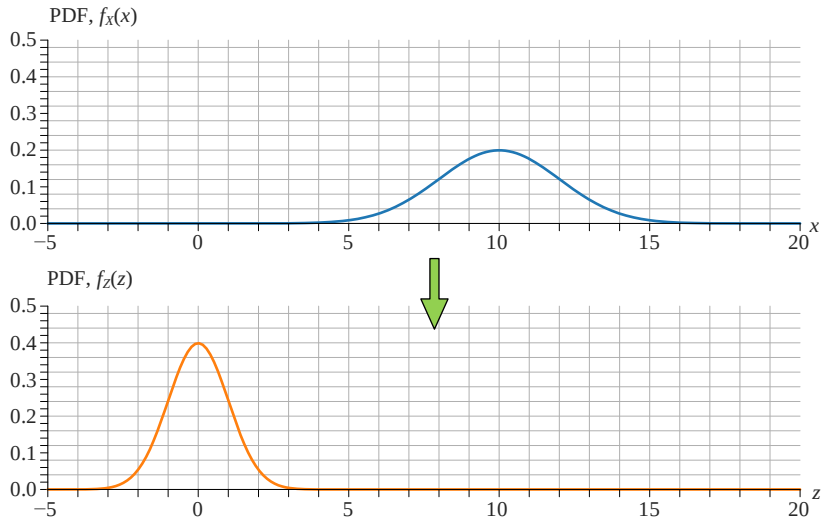


图 27. 随机变量 X 线性变换得到 Z 的过程

如图 28 所示，平移仅改变随机数的均值位置，不影响随机数的分布情况。如图 29 所示，缩放改变随机数的分布离散程度。

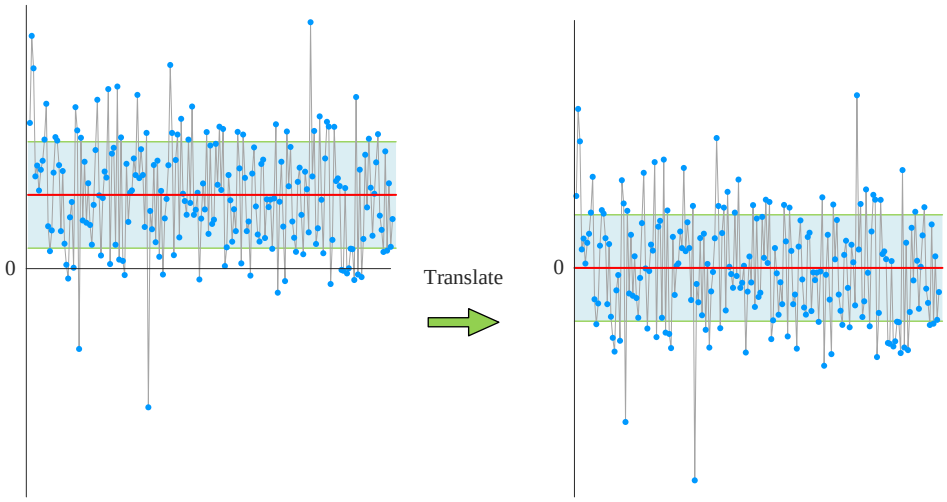


图 28. 平移

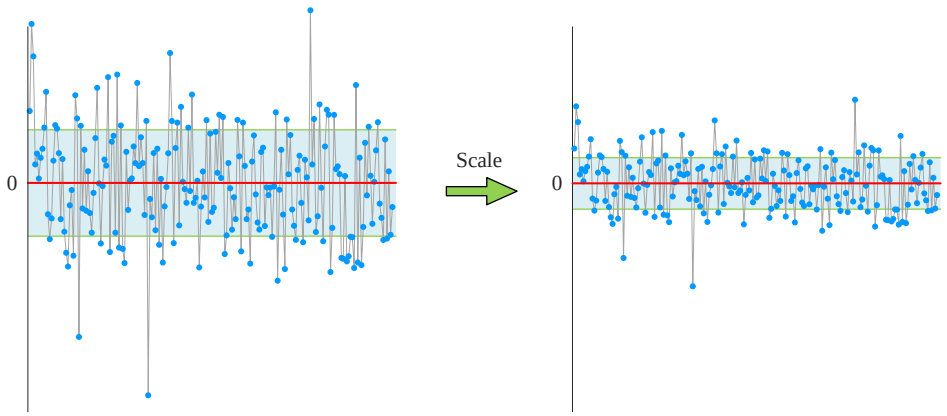


图 29. 缩放

假设 X 是连续随机变量，它的概率密度函数 PDF 为 $f_X(x)$，经过如下线性变换得到 Y:

$$Y = aX + b \tag{14}$$

Y 的 PDF 为:

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right) \tag{15}$$

这样就解释了 (10) 和 (11) 的关系。

此外，服从正态分布的随机变量，在进行线性变换后，正态性保持不变。比如，X 为服从 $N(\mu, \sigma^2)$ 的随机变量；$Y = aX + b$ 仍然服从正态分布。Y 的均值、方差分别为：

$$\mathrm{E}(Y) = a\mu + b, \quad \mathrm{var}(Y) = a^2\sigma^2 \tag{16}$$

图 30 所示为随机变量线性变换的示意图。

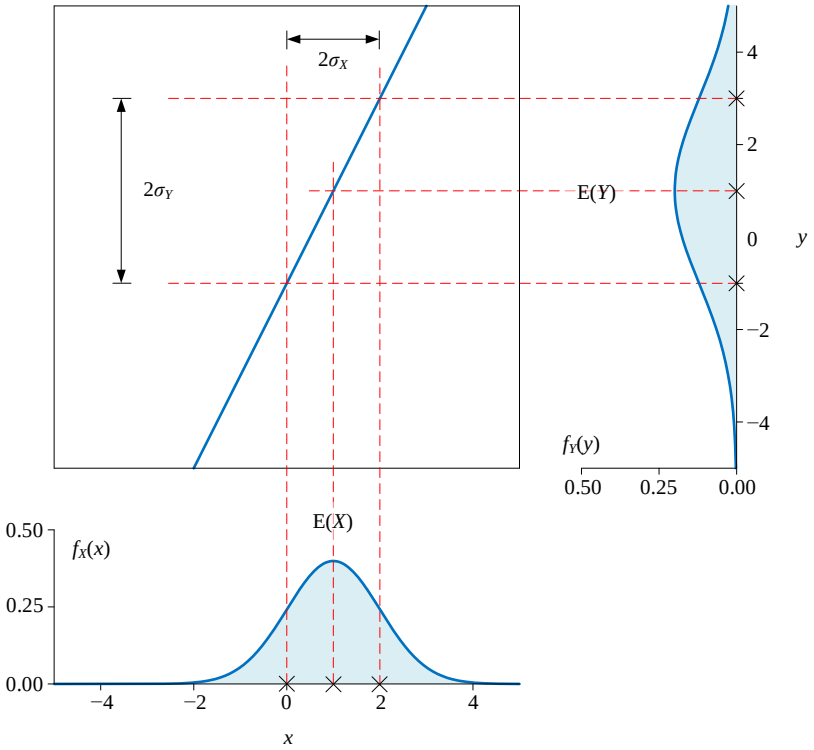


图 30. 线性变换对均值和方差的影响

面积归 1

$f_X(x)$ 作为一个概率密度函数的基本要求：1) 非负；2) 面积为 1：

$$\begin{aligned} f_X(x) &\geq 0 \\ \int_{-\infty}^{+\infty} f_X(x)\,\mathrm{d}x &= 1 \end{aligned} \tag{17}$$

这便解释了为什么 (1) 分母上要除以 $\sqrt{2\pi}$？因为如下高斯函数积分结果为 $\sqrt{2\pi}$：

$$\int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2}\right)\mathrm{d}x = \sqrt{2\pi} \tag{18}$$

也就是说：

$$\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)\mathrm{d}x=1 \tag{19}$$

下面，推导 (1) 和整个横轴围成的面积为 1：

$$\begin{aligned}\int_{-\infty}^{+\infty}f_X(x)\mathrm{d}x&=\int_{-\infty}^{+\infty}\frac{1}{\sigma\sqrt{2\pi}}\exp\left(\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)\mathrm{d}x\\&=\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}\exp\left(\frac{-1}{2}\left(\frac{x-\mu}{\sigma}_z\right)^2\right)\mathrm{d}\left(\frac{x-\mu}{\sigma}_z\right)\\&=\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}\exp\left(\frac{-1}{2}z^2\right)\mathrm{d}z=\frac{\sqrt{2\pi}}{\sqrt{2\pi}}=1\end{aligned} \tag{20}$$

历史上，以下两个函数都曾多作为正态函数 PDF 解析式：

$$\begin{aligned}f_1(x)&=\frac{1}{\sqrt{\pi}}\exp\left(-x^2\right)\\f_2(x)&=\exp\left(-\pi x^2\right)\end{aligned} \tag{21}$$

它们之所以被大家放弃，都是因为方差不方便。$f_1(x)$ 的方差为 1/2。$f_2(x)$ 的方差为 $1/(2\pi)$。显而易见，作为标准正态分布的 PDF，(4) 更方便，因为它的方差为 1，标准差也是 1。

距离 → 亲密度

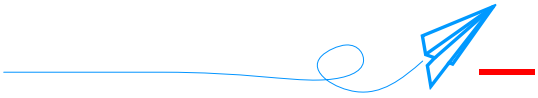
大家可能还有印象，我们在《数学要素》第 12 章讲过讲高斯函数：

$$f(x)=\exp\left(-x^2\right) \tag{22}$$

注意，(22) 的积分为：

$$\int_{-\infty}^{\infty}\exp\left(-x^2\right)\mathrm{d}x=\sqrt{\pi} \tag{23}$$

前文提过几次，Z 分数代表“距离”，而利用类似 (22) 这种高斯函数，我们将“距离”转换成“亲近度”。这样我们更容易理解 (10)，距离期望值 μ 越近，亲近度越大，代表可能性越大，概率密度越大；反之，离 μ 越远，越疏远，代表可能性越小，概率密度越小。本书后文还会用这个视角分析其他高斯分布。



在实际应用中，高斯分布经常用于建模和分析连续型数据，如测量值、物理量和经济指标等。在机器学习和数据分析中，高斯分布也被广泛应用于分类、聚类、离群点检测等问题中。但是，仅仅掌握一元高斯分布的知识是不够的。从下一章开始，我们将探讨二元、多元高斯分布，以及高斯分布背后的协方差矩阵。