# 多维高斯分布

Dezeming Family

2021 年 9 月 18 日

DezemingFamily 系列书和小册子因为是电子书，所以可以很方便地进行修改和重新发布。如果您获得了 DezemingFamily 的系列书，可以从我们的网站 [https://dezeming.top/] 找到最新版。对书的内容建议和出现的错误欢迎在网站留言。

20210918：完成第一版。

## 目录



## 一 二维高斯分布的定义

我们经常可以看到二维高斯分布（正态分布）有多种表示形式，例如对于概率密度函数，我们有：

$$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x-\mu_1)(y-\mu_2)}{2\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right]} \tag{一.1}$$

其中，标准差 $\sigma_1>0$，$\sigma_2>0$，且相关系数 $|\rho|<1$。

或者是这种定义方式：

$$f(x,y)=\frac{1}{2\pi\sqrt{det(\Sigma)}}e^{-\frac{1}{2}\left[(x,y)-\overrightarrow{\mu}\right]^T\Sigma^{-1}\left[(x,y)-\overrightarrow{\mu}\right]} \tag{一.2}$$

其中，$\Sigma$ 是协方差矩阵：

$$\Sigma=\begin{bmatrix}cov(x,x) & cov(x,y)\\ cov(y,x) & cov(y,y)\end{bmatrix} \tag{一.3}$$

我们可以化简一下：

$$det(\Sigma)=cov(x,x)\times cov(y,y)-cov(x,y)\times cov(y,x) \tag{一.4}$$

$$=\sigma_x^2\times\sigma_y^2-\sigma_x^2\times\sigma_y^2\frac{cov(x,y)\times cov(y,x)}{\sigma_x^2\times\sigma_y^2} \tag{一.5}$$

$$=\sigma_x^2\times\sigma_y^2\Big(1-\frac{cov(x,y)}{\sigma_x\times\sigma_y}\frac{cov(y,x)}{\sigma_x\times\sigma_y}\Big) \tag{一.6}$$

根据相关系数的计算公式（见 DezemingFamily 的《协方差与相关系数》），可以前后两种表示关系是相等的。但第二种表示关系更简洁，我们由此引出多维高斯分布：

$$f(\overrightarrow{x}) = \frac{1}{\sqrt{(2\pi)^2 det(\Sigma)}} e^{-\frac{1}{2}\left[\overrightarrow{x}-\overrightarrow{\mu}\right]^T \Sigma^{-1} \left[\overrightarrow{x}-\overrightarrow{\mu}\right]} \tag{一.7}$$

$$\overrightarrow{x} = (x_1, x_2, ..., x_n)^T \tag{一.8}$$

$$\Sigma = \begin{bmatrix} cov(x_1,x_1) & cov(x_1,x_2) & \cdots & cov(x_1,x_n) \\ cov(x_2,x_1) & cov(x_2,x_2) & \cdots & cov(x_2,x_n) \\ \cdots & \cdots & \cdots & \cdots \\ cov(x_n,x_1) & cov(x_n,x_1) & \cdots & cov(x_n,x_n) \end{bmatrix} \tag{一.9}$$

# 二　二维高斯分布的原理

既然高斯分布是概率密度函数，也就是说，当 $f(\overrightarrow{x})$ 的值比较大的时候，说明生成向量 $f(\overrightarrow{x})$ 的概率会比较大（尽管严格意义上连续概率中生成某一点的概率都是 0）。

一维高斯分布的由来已经众说纷纭，有些数学大家都给出了自己的解释，包括高斯手稿的解读等，但没有非常统一的论述。尽管众说纷纭，但大家基本上都认为高斯分布的提出是数学家们尝试对误差分布的定量描述。

对于一个给定现象，我们进行多次观察，比如用尺子测量一段线段的长度，我们找 1000 个人去测量，得到测量数据。首先我们明确的是，线段的长度是一定的，不会因为人为测量而改变；但人的测量会存在误差。误差是对称分布的，而且大的误差出现频率会比较低，小的误差出现频率会比较高。最终我们会选择将这 1000 个人的测量结果取平均值，得到我们认为的比较准确的线段长度。

经过高斯的证明，得到了基于误差正态分布的最小二乘理论，由此可以作为高斯分布的开端。我们假设已经知道了一维高斯分布的情况，现在我们尝试拓展到二维以及多维。

考虑二维误差分布，比如人的身高体重的测量，定义为 $(height, weight)$，由于身高的测试比较容易，直接用量尺即可，所以测量误差一般比较小；而体重由于需要借助天平和砝码，考虑到各种摩擦力以及砝码误差，体重的测量误差一般会比较大，这个时候，二维正态分布就需要考虑两个维度分别的误差分布关系了。

## 独立变量二维高斯分布

如果两个维度之间没有相关性关系，那么我们其实就可以用两个一维正态分布相乘来得到二维正态分布。

$$f(x_1) = \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}} \tag{二.1}$$

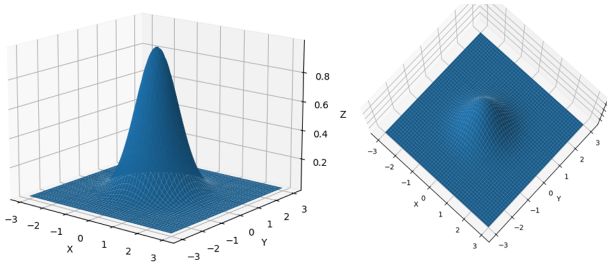
$$f(x_2) = \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{(x_2-\mu_2)^2}{2\sigma_2^2}} \tag{二.2}$$

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\left(\frac{(x_1-\mu_1)^2}{2\sigma_1^2} + \frac{(x_2-\mu_2)^2}{2\sigma_2^2}\right)} \tag{二.3}$$

这是二维高斯分布的特殊情况，注意这个时候的协方差矩阵是：

$$\Sigma = \begin{bmatrix} cov(x,x) = \sigma_x^2 & 0 \\ 0 & cov(y,y) = \sigma_y^2 \end{bmatrix} \tag{二.4}$$

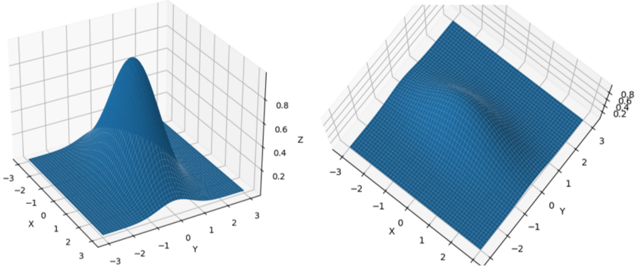
这个高斯分布比较规则，像一顶帽子：



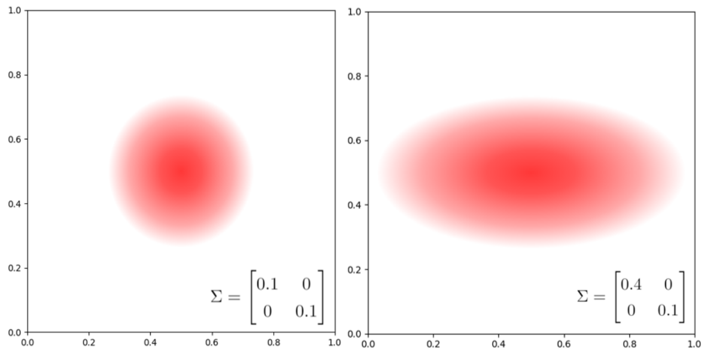
## 非独立变量二维高斯分布

但通常两个维度会存在一定的关系，比如身高和体重就具有一定的线性关系，一般长得高的人可能会重一些。

我们设 $\sigma_x^2=4$，$\sigma_y^2=1$，得到如下高斯分布：



可以看到，x 方向的分布比 y 方向更广了，但是它们之间还是没有什么相关性，椭圆的长短轴都是跟坐标轴对齐的。下图表示不同的协方差矩阵下的高斯分布情况：




现在考虑我们的协方差矩阵中非对角线元素也不为 0 的情况。

因为协方差矩阵是半正定的对称矩阵，可以分解为 $\Sigma=Q\Lambda Q^T$ 的形式（见 DezemingFamily 的《正定矩阵和负定矩阵》以及《矩阵的相似对角化》）由此我们可以得到 $\Sigma^{-1}=Q\Lambda^{-1}Q^T$。

$$\Lambda=\begin{bmatrix}\lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_m\end{bmatrix}\qquad \Lambda^{-1}=\begin{bmatrix}\frac{1}{\lambda_1} & & \\ & \frac{1}{\lambda_2} & \\ & & \frac{1}{\lambda_m}\end{bmatrix}\tag{二.5}$$

对于二维高斯分布：

$$f(x,y)=\frac{1}{2\pi\sqrt{det(\Sigma)}}e^{-\frac{1}{2}\left[(x,y)-\vec{\mu}\right]^T\Sigma^{-1}\left[(x,y)-\vec{\mu}\right]}\tag{二.6}$$

我们可以看到，变量仅在 $e$ 的指数上，所以我们可以只看 $e$ 的指数部分（其他部分的作用是为了保证概率密度积分为 1）：

$$-\frac{1}{2}\left[(x,y)-\vec{\mu}\right]^T\Sigma^{-1}\left[(x,y)-\vec{\mu}\right]=\tag{二.7}$$

$$-\frac{1}{2}\left[(x,y)-\vec{\mu}\right]^T Q\Lambda^{-1}Q^T\left[(x,y)-\vec{\mu}\right]=\tag{二.8}$$

$$-\frac{1}{2}\Big(\left[(x,y)-\vec{\mu}\right]^T Q\Big)\Lambda^{-1}\Big(Q^T\left[(x,y)-\vec{\mu}\right]\Big)\tag{二.9}$$

我们可以看出：

$$\left(\left[(x,y)-\vec{\mu}\right]^T Q\right)=\left(Q^T\left[(x,y)-\vec{\mu}\right]\right)^T \tag{二.10}$$

因此，我们设：

$$(a,b)=\left(\left[(x,y)-\vec{\mu}\right]^T Q\right) \tag{二.11}$$

因此得到指数部分为：

$$-\frac{1}{2}(a,b)\Lambda^{-1}(a,b)^T \tag{二.12}$$

由 DezemingFamily 的《矩阵二次型》可知，指数部分其实就是一个椭圆，而且该椭圆的长短轴的值是 $\Lambda^{-1}$ 的对角线元素值。二维向量构成的二次型矩阵中，向量做变量时其实就是构成一个椭圆（其中 $C$ 是我们设置的某个常数）：

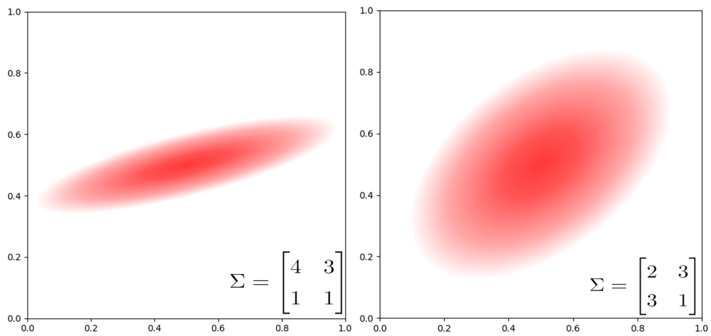
$$\left(\left[(x,y)-\vec{\mu}\right]^T Q\right)\Lambda^{-1}\left(Q^T\left[(x,y)-\vec{\mu}\right]\right)=C \tag{二.13}$$

因为 $e^x$ 指数函数单调递增，所以整个二维高斯分布函数等于某个定值时，其实也构成一个椭圆：

$$e^{-\frac{1}{2}\left(\left[(x,y)-\vec{\mu}\right]^T Q\right)\Lambda^{-1}\left(Q^T\left[(x,y)-\vec{\mu}\right]\right)}=e^{-\frac{1}{2}C} \tag{二.14}$$

在这个椭圆上的所有点具有相同的概率密度函数。

我们画一下不同 $\Sigma$ 值的高斯分布：




左图的 $Sigma$ 两个特征值约为 $4.8, 0.21$，右图的两个特征值约为 $4.5, -1.5$，很明显左图中横纵轴方向变化的差异性就会比较大。

注意：

我们在计算特征值的时候无需计算 $\Sigma^{-1}$ 的特征值，因为要知道，高斯分布的方差是作用在分母上的：

$$f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{二.15}$$

这里求出来的特征值就是代表了 $\sigma^2$ 的含义，特征值越大，表示在该方向分布越广，特征值越小表示在该方向分布越密集。大家不要把 $\Sigma^{-1}$ 和 $\Sigma$ 的形式弄混了。

# 参考文献


[1] https://www.zhihu.com/question/66005432

[2] https://blog.csdn.net/farmwang/article/details/78699926

[3] https://cosx.org/2013/01/story-of-normal-distribution-1

[4] https://blog.csdn.net/weixin_43716250/article/details/109706857

[5] https://www.zhihu.com/question/36339816