参考文献：

1. 盛骤，试式千，潘承毅．概率论与数理统计（第四版）：高等教育出版社，2008:p100-p112
2. <https://blog.csdn.net/zhongkelee/article/details/44064401>

# 二维的情况







令





# 高维的情况



 维正态随机变量 的概率密度定义为



# 协方差矩阵的计算

假设有 个样本



根据定义有





# PCA及投影方差

## 4.1 将协方差矩阵相似对角化



## 投影

注意：这里是指中心化后的数据的投影





得到结论



一般认为



此时有



一般将投影参数限制在平均值（这里认为是0）的3个标准差内（）内，即：



在信号处理中认为信号具有较大的方差，噪声有较小的方差，信噪比就是信号与噪声的方差比，越大越好。

因此取前个特征值对应的特征向量作为主成分。主成分所占整个信息的百分比可用下式计算：



## 简化计算

当样本数小于数据维度时可以进行简化计算：



如上所示，只需要计算矩阵的特征值 和特征向量 就可以得到协方差的特征值 对应的特征向量 。