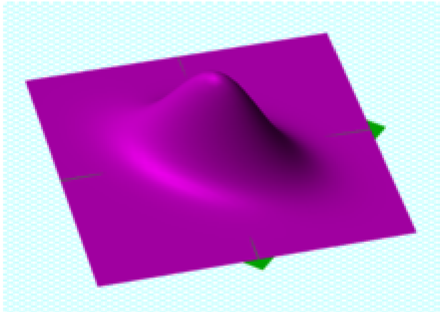
WIKIPEDIA

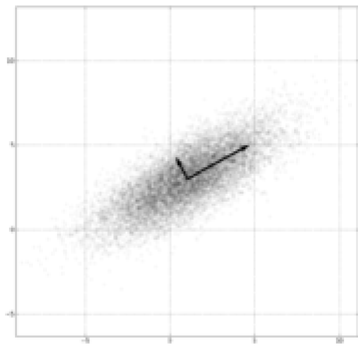
协方差矩阵

维基百科，自由的百科全书

在统计学与概率论中，**协方差矩阵**（也称**离差矩阵**、**方差-协方差矩阵**）是一个矩阵，其 i, j 位置的元素是第 i 个与第 j 个随机向量（即随机变量构成的向量）之间的协方差。这是从标量随机变量到高维度随机向量的自然推广。



中心为 (0, 0) 的一个二元高斯概率密度函数，协方差矩阵为 [1.00, 0.50 ; 0.50, 1.00]。



一个左下右上方向标准差为 3，正交方向标准差为 1 的多元高斯分布的样本点。由于 x 和 y 分量共变，x 与 y 的方差不能完全描述该分布；箭头的方向对应的协方差矩阵的特征向量，其长度为特征值的平方根。

目录

定义

假设X是以n个随机变數组成的列向量，

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$$

并且μ_i是X_i的期望值，即, $\mu_i = \mathrm{E}(X_i)$。协方差矩阵的第(i, j)項（第(i, j)項是一个协方差）被定义为如下形式：

$$\Sigma_{ij} = \mathrm{cov}(X_i, X_j) = \mathrm{E}\left[\,(X_i - \mu_i)(X_j - \mu_j)^{\mathrm{T}}\,\right]$$

而协方差矩阵为：

$$\Sigma = \mathrm{E}\left[\left(\mathbf{X} - \mathrm{E}[\mathbf{X}]\right)\left(\mathbf{X} - \mathrm{E}[\mathbf{X}]\right)^{\mathrm{T}}\right]$$

$$= \begin{bmatrix} \mathrm{E}[(X_1-\mu_1)(X_1-\mu_1)] & \mathrm{E}[(X_1-\mu_1)(X_2-\mu_2)] & \cdots & \mathrm{E}[(X_1-\mu_1)(X_n-\mu_n)] \\ \mathrm{E}[(X_2-\mu_2)(X_1-\mu_1)] & \mathrm{E}[(X_2-\mu_2)(X_2-\mu_2)] & \cdots & \mathrm{E}[(X_2-\mu_2)(X_n-\mu_n)] \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{E}[(X_n-\mu_n)(X_1-\mu_1)] & \mathrm{E}[(X_n-\mu_n)(X_2-\mu_2)] & \cdots & \mathrm{E}[(X_n-\mu_n)(X_n-\mu_n)] \end{bmatrix}$$

矩阵中的第(i,j)个元素是X_i与X_j的共變異數。这个概念是对于标量随机变數方差的一般化推广。

术语与符号分歧

共變異數矩阵有不同的术语。有些统计学家，沿用了概率学家威廉·费勒的说法，把这个矩阵称之为随机向量X的變異數（Variance of random vector X），这是从一维随机变量方差到高维随机向量的自然推广。另外一些则把它称为**共變異數矩阵**（Covariance matrix），因为它是随机向量里头每个标量元素的协方差的矩阵。不幸的是，这两种术语带来了一定程度上的冲突：

随机向量X的方差（Variance of random vector X）定义有如下两种形式：

$$\operatorname{var}(\mathbf{X}) = \operatorname{cov}(\mathbf{X}) = \mathrm{E}\left[(\mathbf{X}-\mathrm{E}[\mathbf{X}])(\mathbf{X}-\mathrm{E}[\mathbf{X}])^{\mathbf{T}}\right].$$

协方差矩阵**（Covariance matrix）定义如下：**

$$\operatorname{cov}(\mathbf{X},\mathbf{Y}) = \mathrm{E}\left[(\mathbf{X}-\mathrm{E}[\mathbf{X}])(\mathbf{Y}-\mathrm{E}[\mathbf{Y}])^{\top}\right]$$

第一个记号可以在威廉·费勒的广受推崇的两册概率论及其应用的书中找到。两个术语除了记法之外并没有不同。

性质

$\Sigma = \mathrm{E}\left[(\mathbf{X}-\mathrm{E}[\mathbf{X}])(\mathbf{X}-\mathrm{E}[\mathbf{X}])^{\top}\right]$ 与$\mu = \mathrm{E}(\mathbf{X})$ 满足下边的基本性质：

1. $\Sigma = \mathrm{E}(\mathbf{X}\mathbf{X}^{\top}) - \mu\mu^{\top}$
2. Σ是半正定的和对称的矩阵。
3. $\operatorname{var}(\mathbf{a}^{\top}\mathbf{X}) = \mathbf{a}^{\top}\operatorname{var}(\mathbf{X})\mathbf{a}$
4. $\Sigma \geq 0$
5. $\operatorname{var}(\mathbf{AX}+\mathbf{a}) = \mathbf{A}\operatorname{var}(\mathbf{X})\mathbf{A}^{\top}$
6. $\operatorname{cov}(\mathbf{X},\mathbf{Y}) = \operatorname{cov}(\mathbf{Y},\mathbf{X})^{\top}$
7. $\operatorname{cov}(\mathbf{X_1}+\mathbf{X_2},\mathbf{Y}) = \operatorname{cov}(\mathbf{X_1},\mathbf{Y}) + \operatorname{cov}(\mathbf{X_2},\mathbf{Y})$
8. 若 $p=q$，则有 $\operatorname{cov}(\mathbf{X}+\mathbf{Y}) = \operatorname{var}(\mathbf{X}) + \operatorname{cov}(\mathbf{X},\mathbf{Y}) + \operatorname{cov}(\mathbf{Y},\mathbf{X}) + \operatorname{var}(\mathbf{Y})$
9. $\operatorname{cov}(\mathbf{AX},\mathbf{BX}) = \mathbf{A}\operatorname{cov}(\mathbf{X},\mathbf{X})\mathbf{B}^{\top}$
10. 若$\mathbf{X}$ 与$\mathbf{Y}$ 是独立的，则有$\operatorname{cov}(\mathbf{X},\mathbf{Y}) = 0$
11. $\Sigma = \Sigma^{\top}$

其中 $\mathbf{X}, \mathbf{X_1}$ 与$\mathbf{X_2}$ 是随机$(\mathbf{p}\times\mathbf{1})$向量, $\mathbf{Y}$ 是随机$(\mathbf{q}\times\mathbf{1})$向量, $\mathbf{a}$ 是$(\mathbf{p}\times\mathbf{1})$ 向量, $\mathbf{A}$ 与$\mathbf{B}$ 是$(\mathbf{q}\times\mathbf{p})$ 矩阵。

尽管共變異數矩阵很简单，可它却是很多领域里的非常有力的工具。它能导出一个变换矩阵，这个矩阵能使数据完全去相关(decorrelation)。从不同的角度看，也就是说能够找出一组最佳的基以紧凑的方式来表达数据。(完整的证明请参考瑞利商)。 这个方法在统计学中被称为主成分分析(principal components analysis)，在图像处理中称为Karhunen-Loève 变

换(KL-变换)。

複随机向量

均值为μ的複随机标量变量的方差定义如下（使用共轭複数）：

$$\operatorname{var}(z) = \operatorname{E}[(z-\mu)(z-\mu)^*]$$

其中复数z的共轭记为z^*。

如果Z 是一个复列向量,则取其共轭转置，得到一个方阵:

$$\operatorname{E}[(Z-\mu)(Z-\mu)^*]$$

其中Z^*为共轭转置, 它对于标量也成立，因为标量的转置还是标量。

估计

多元正态分布的共變異數矩陣的估计的推导非常精致. 它需要用到谱定义以及为什么把标量看做1×1矩阵的迹更好的原因。参见共變異數矩陣的估计。

外部链接

- Covariance Matrix (http://mathworld.wolfram.com/CovarianceMatrix.html) at Mathworld