# 多元统计分析导论

Ny *

nymath@163.com

2022 年 11 月 11 日

**摘要**

测试

# 目录



# 插图

*我会不定期更新笔记，如果感兴趣的话，可以前往https://github.com/nymath/notes4master获取最新版本。

# 1 样本几何

弹到坐标，必须得提到基，我们把坐标定义为一个矩阵。

**1.1 Definition 坐标**

Suppose $x \in \mathbb{R}^n$, then the matrix of $x$ is defined by

$$\mathcal{M}(x,e) = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

where $e$ is the standard orthonormal basis of $\mathbb{R}^n$.

**1.2 Theorem 坐标变换**

Suppose $x \in \mathbb{R}^n$, $e, \nu$ are basis of $\mathbb{R}^n$, then

$$\mathcal{M}(x,e) = \mathcal{M}(I,\nu,e)\mathcal{M}(x,\nu)$$

or equivalently,

$$\mathcal{M}(x,\nu) = \mathcal{M}(I,e,\nu)\mathcal{M}(x,e)$$

**1.3 Definition**

$$\mathcal{M}(I,e,\nu) = \frac{\partial \nu}{\partial x}$$

Example 1.4:

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = A \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

假定样本点为 P,p 在 e 下的坐标为 x，y，z，那么 p 在 v 在的坐标为 (u,v,w), 其中 v 可以通过 $A^{-1}e$ 得到

Remark: 可能你发现了，我在说基的时候，只写了 $e$，$\nu$. 思考一下我为什么写? (family 是映射，不是集合)

## 2 多元假设检验

# 3 判别分析

## 3.1 贝叶斯判别法

这里顺便复习一下，贝叶斯统计。假设总体有 k 个类别，我们用 $\Theta$ 表示，$\Theta \sim (p_1, \cdots, p_k)$. 用 $X$ 代表数据展现出的特征，同时我们有条件密度函数

$$f(x|\Theta = i) = f_i(x)$$

这个条件密度函数可以通过样本进行估计 (或者直接给定，但前者更好)。基于此，我们得到参数的后验分布，$\Theta_{|X=x}$，他服从

$$\Theta_{|X=x} \sim (\frac{f_i(x)P_i}{\sum_{i=1}^{k} P_i f_i(x)})$$

接下来，我们想选择一个随机变量 (估计量)，他与 $\Theta_{|X=x}$ 的" 距离" 充分接近。常用的备选有

- $L^2$ 距离: $L(\theta, \Theta_{|X=x}) = E(\theta - \Theta_{|X=x})^2$.
- $L^1$ 距离: $L(\theta, \Theta_{|X=x}) = E(|\theta - \Theta_{|X=x}|)$.
- 

Remark: **特别的，在 $L^2$ 距离意义下，我们的估计是 $E(\Theta|X = x)$.**

但是呢，在判别分析中，我们有更一般的损失函数，即

$$L(\theta, \Theta_{|X=x}) = E(c(\theta|\Theta_{|X=x})).$$

Example 3.1: 当 $c(\theta|\Theta_{|X=x}) = 1 - \chi_\theta(\Theta_{|X=x})$ 时，期望损失为

$$1 - \sum_{i=1}^{n} \chi_\theta(i) \frac{f_i(x)P_i}{\sum_{i=1}^{k} P_i f_i(x)}$$

因此，当 $\theta$ 等于 $\Theta_{|X=x}$ 的众数时，期望损失最小。

更一般的，我们的损失函数可以定义为

$$\boxed{L(\theta, \Theta_{|X=x}) = \sum_{j=1}^{k} c(\theta|j) \frac{f_j(x)P_j}{\sum_{i=1}^{k} P_i f_i(x)}}$$

## 3.2 Fisher 判别法

Fisher 判别法的思路在于，$\mathbb{R}^p$ 的一个向量，将我们的各组数据投影在这个向量上，使得组内距离充分小，组间距离充分大。(实际上是一维的方差分析)。由于我们的数据都在同一个方向，所以我们只需要计算各组数据与这个向量的内积即可。具体而言

$$
\begin{array}{lllll}
G_1: & x_1^{(1)}, & x_2^{(1)}, & \cdots, & x_{n_1}^{(1)} \\
G_2: & x_1^{(2)}, & x_2^{(2)}, & \cdots, & x_{n_2}^{(2)} \\
 & & \vdots & & \\
G_k: & x_1^{(k)}, & x_2^{(k)}, & \cdots, & x_{n_k}^{(k)}
\end{array}
$$

# 4 主成分分析

主成分分析可以用于降维，同时保持数据的大部分变差。一般来说，当数据相关性较强时，使用主成分分析的效果较好。

**4.1 Definition**

主成分应当具有如下性质:

- 每一个主成分是原始变量的线性组合
- 主成分数目远远小于变量的数量
- 主成分保留了原始变量的大部分信息
- 第一主成分应当保留最多的信息
- 各主成分之间互不相关

Remark: 我们这里提到的信息，使用方差进行刻画。

**4.2 Theorem** ***Real Spectrum***

实数域上的自伴随正算子 (在标准正交基下为正定的实对称矩阵)，有如下分解

$$\boxed{\mathcal{M}(T,\nu,\nu)=\mathcal{M}(I,e,\nu)\mathcal{M}(T,e,e)\mathcal{M}(I,\nu,e)}$$

其中 $\nu$ 为特征向量构成的基，而且

$$\boxed{\mathcal{M}(T,\nu,\nu)=diag(\lambda_1,\cdots,\lambda_n)}$$

$$\boxed{\mathcal{M}(I,\nu,e)=[\nu_1,\cdots,\nu_n]}$$

即 $T$ 在特征基下的矩阵为一个对角矩阵。

## 4.1 协方差矩阵法

**4.3 Theorem**

假定 $\Sigma$ 正定，$\lambda_i$ 为其特征值，$\nu_i$ 为对应的特征向量，则

$$\max_{x\perp\text{span}\{\nu_1,\cdots,\nu_k\}}\frac{\langle\Sigma x,x\rangle}{\langle x,x\rangle}=\lambda_{k+1}$$

**4.4 Theorem**

假定 p 维随机向量的协方差矩阵为 $\mathcal{M}(T,e,e)=\Sigma$，则正交变换

$$\boxed{Y=\mathcal{M}(T,e,\nu)X=P^{-1}X}.$$

生成了 p 个主成分，且 $Y[k]=\nu_k' X$ 为第 k 主成分。

值得注意的是，

$$Var(Y_i)=Var(\nu_k\prime X)=\nu_k\prime\Sigma\nu_k=\lambda_k.$$

**4.5 Theorem *PC* 的性质**

- $Y$ 的协方差矩阵为 $\mathcal{M}(T,\nu,\nu)$
- $Y$ 的总方差为 $\mathcal{M}(T,\nu,\nu)$ 的 trace.

**Proof**

$$\begin{aligned} Var(\iota' Y) &= tr(\mathcal{M}(T,\nu,\nu)) \\ &= tr(\mathcal{M}(T,e,e)) \end{aligned}$$

□

**4.6 Definition 贡献率**

主成分 $Y_i$ 的贡献率定义为:

$$\frac{\lambda_i}{\sum_{i=1}^{p}\lambda_i}$$

累积贡献率定义为

$$\frac{\sum_{i=1} m\lambda_i}{\sum_{i=1}^{p}\lambda_i}$$

**4.7 Definition 因子负荷量 *(Factor Loadings)***

$X_i$ 与 $Y_k$ 的 Pearson rho 定义为因子负荷量

$$\boxed{\rho(Y_k,X_i)=\frac{\nu_{ik}\sqrt{\lambda_k}}{\sqrt{\sigma_{ii}}}}$$

**Proof**

$$\begin{aligned} Cov(Y_k, X_i) &= Cov(\nu_k' X, e_i' X) \\ &= \nu_k' \Sigma e_i \\ &= \lambda_k \nu_{ik}. \end{aligned}$$

□

Remark: $\nu_{ik}$ 可以解释为第 k 个主成分对第 i 个变量的敏感性。
而因子负荷量的绝对值大小刻画了该主成分的主要意义及其成因，在解释第 i 个变量是第 k 个主成分的重要性时，应该根据因子负荷量进行分析，而不是仅仅 $\nu_{ik}$.

**4.8 Theorem**

$$\boxed{\sum_{i=1}^{p} \rho^2(Y_k, X_i)\sigma_{ii} = \lambda_k}$$

## 4.2 相关系数法

## 4.3 主成分的几何含义

随机向量 $X$ 在 $e$ 下的坐标为 $(X_1, \cdots, X_p)$，这些指标的 coordinate 不是独立的。然后呢，我们找了一个新的坐标系 $\nu = P \circ e$，在这个坐标系下，$X$ 的坐标为 $(Y_1, \cdots, Y_p)$，而且这些坐标之间是独立的。同时呢，我们也稍微的对 $P$ 的行 (列) 进行一些交换。这么做之后达到了一个怎么样的效果呢？$X$ 在 $\nu_1$ 方向上的变差，也就是 $Var(Y_1)$，最大。(这体现为，对 X 不断抽样，然后把它投影到 $\nu_1$ 上，轨迹应该最长才对)