

第11章 判别分析

李高荣

北京师范大学统计学院

E-mail: ligaorong@bnu.edu.cn



- 1 判别分析简介
- 2 判别分析的准则
- 3 概率分布已知的两个总体的判别方法
 - 先验分布已知的情形
 - 先验概率未知的情形
- 4 两个已知多元正态分布的判别
- 5 参数未知时两个正态总体的判别
 - 判别准则
 - 判别准则的分布
 - 判别准则的渐近分布
 - 似然比准则
- 6 错判概率
 - 基于 W 错判概率的渐近展开
 - 基于 Z 错判概率的渐近展开
- 7 多个总体的判别
- 8 多个多元正态分布的判别
- 9 案例及R语言计算

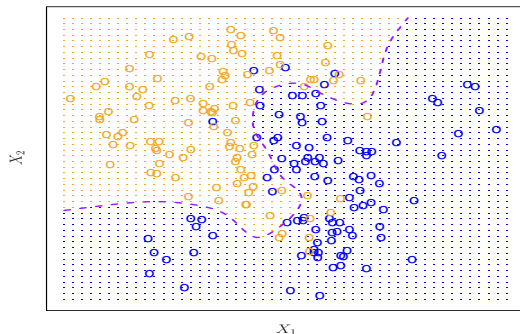


- 扫描二维码获取在线课件和相关教学电子资源
- 请遵守电子资源使用协议

- **判别分析(discriminant analysis)**: 使用具有类别信息的观测数据(training set) 建立一个**分类器(classifier)** 或者**分类法则(classification rule)**, 其可以最大可能的区分事先定义的类。
- **分类(Classification)**: 给定一组新的未知类别信息的观测数据集, 使用分类器将其分配到一些已知的类中。
- 实际应用中, 判别分析与分类常常混在一起:
 - 一个作为判别分析的 p 元函数也可以用于对新的观测数据进行分类
 - 一个分类准则常常作为判别法则使用

判别分析简介

- 假设有 K 个总体(类), $G = 1, 2, \dots, K$ 表示类别。 \mathbf{x} 为取值 Ω 上的多元观测, 且 $\mathbf{x}|G = g \sim f_g(\mathbf{x})$, $g = 1, 2, \dots, K$, 其中 f 为概率函数;
- 对任意给定的观测 \mathbf{x}_0 , 目的是把 \mathbf{x}_0 归到 K 个类中的某个。



例子

- 患者来到急诊室，医生通过患者的一系列症状将患者归到三类可能病症中的一类
- 网上银行服务需要基于用户的IP 地址、历史交易记录等信息辨别一个在线网上交易是否存在诈骗行为
- 一位生物学家对一定数量患先天疾病的患者和未患病的人进行DNA 排列信息分析，发现哪些DNA 突变是致病的，哪些是不致病的
- ...

判别分析的准则

- **目的:** 使错判概率最小, 或者使判别错误的平均值最小
- **问题:** 存在两个总体 π_1 和 π_2 , 给定一个个体的测量值向量 $\mathbf{x} = (x_1, \dots, x_p)'$, 判断该个体是来自总体 π_1 , 还是总体 π_2 ?
- **解决方案:** 寻找判别方法, 把 \mathbb{R}^p 空间分成两个区域 R_1 和 R_2 , 记为 $R = (R_1, R_2)$ 。如果这个个体的观测值向量 $\mathbf{x} \in R_1$, 则把它判断为来自总体 π_1 , 否则为来自总体 π_2 。
- **两种错误:** (1) 个体来自总体 π_1 , 却被误判为来自总体 π_2 ; (2) 个体来自总体 π_2 , 却被误判为来自总体 π_1 。

Table: 判别的两类错误

统计决策	π_1	π_2
π_1	0	$C(2 1)$
π_2	$C(1 2)$	0

- $C(2|1)$: 个体来自于总体 π_1 , 却被误判为来自总体 π_2 的损失
- $C(1|2)$: 个体来自于总体 π_2 , 却被误判为来自总体 π_1 的损失
- 好的判别方法的标准: 错判带来的损失达到最小

两个总体的情形

- **先验概率**：令 q_1 是来自 π_1 的观测值的先验概率， q_2 是来自 π_2 的观测值的先验概率，满足 $q_1 + q_2 = 1$
- **密度函数**：假设总体 π_1 的密度函数为 $f_1(\mathbf{x})$ ，总体 π_2 的密度函数为 $f_2(\mathbf{x})$
- **正确判别的概率**：若个体来自总体 π_1 ，则被正确判别的概率为：

$$\Pr(1|1, R) = \int_{R_1} f_1(\mathbf{x}) d\mathbf{x}, \quad d\mathbf{x} = dx_1 \cdots dx_p$$

- **错判概率**：若个体来自总体 π_1 ，则被错判为来自总体 π_2 的概率为：

$$\Pr(2|1, R) = \int_{R_2} f_1(\mathbf{x}) d\mathbf{x}$$

- **正确判别的概率**：同样，若个体来自总体 π_2 ，则被正确判别的概率为：

$$\Pr(2|2, R) = \int_{R_2} f_2(\mathbf{x}) d\mathbf{x}$$

- **错判概率**：若个体来自总体 π_2 ，则被错判为来自总体 π_1 的概率为：

$$\Pr(1|2, R) = \int_{R_1} f_2(\mathbf{x}) d\mathbf{x}$$

两个总体的情形

先验信息：观测值来自总体 π_1 的先验概率是 q_1 ，来自总体 π_2 的先验概率是 q_2 。

- **先验概率下正确判别的概率：**观测值来自总体 π_1 且被正确判别的概率为：

$$q_1 \Pr(1|1, R)$$

- **错判概率：**观测值来自总体 π_1 且被错判为来自总体 π_2 的错判概率为：

$$q_1 \Pr(2|1, R)$$

- 类似地，可以定义两种概率分别为：

$$q_2 \Pr(2|2, R), \quad q_2 \Pr(1|2, R)$$

- 错判的平均损失为：

$$\text{ECM}(R_1, R_2) = C(2|1) \Pr(2|1, R)q_1 + C(1|2) \Pr(1|2, R)q_2$$

- **目的**：找最优判别方法，把空间划分成 R_1 和 R_2 ，使得错判的平均损失达到最小。
- **Bayes判别方法**：给定先验概率 q_1 和 q_2 ，使错判的平均损失达到最小的方法称为**Bayes判别方法**。

两个总体的情形

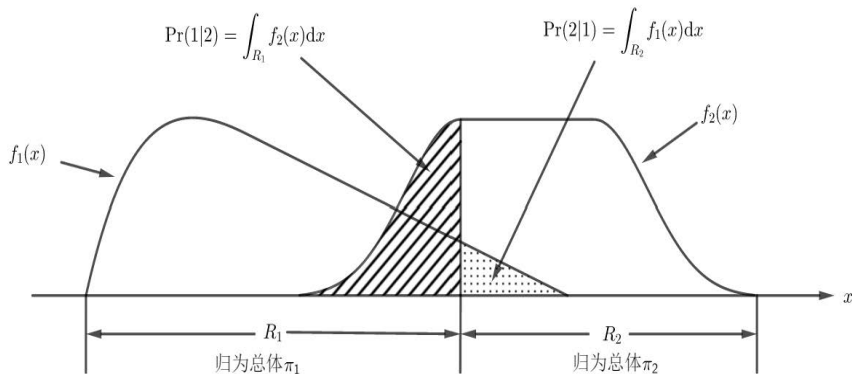


Figure: 当维数 $p = 1$ 时，错判概率的示意图。

两个总体的情形

先验分布未知情形：

- 观测值来自总体 π_1 的期望损失为：

$$C(2|1) \Pr(2|1, R) = r(1, R)$$

- 观测值来自总体 π_2 的期望损失为：

$$C(1|2) \Pr(1|2, R) = r(2, R)$$

- 在实际应用中，我们并不知道观测值是来自总体 π_1 ，还是来自总体 π_2 ，也不知道这两个类别的先验概率，这时如何解决？

两个总体的情形

设 R^* 是另一个判别方法，则

- 若 $r(1, R) \leq r(1, R^*)$ 且 $r(2, R) \leq r(2, R^*)$ ，则称 R 至少和 R^* 一样好；若其中一个不等号严格成立，则称 R 比 R^* 真好。
- **容许性**：如果存在一种方法，没有任何其他方法比它真好，或者至少它和其他方法一样好，那么说这个判别方法，或者这个样本空间的划分 R 是容许的。
- **Minimax原则**：如果一个方法的最大期望损失 $r(i, R)$ 达到最小，则称这个方法是minimax 的。
- **完备性**：对于方法类以外的任何一个方法 R^* ，如果能从这个方法类中找到一个方法 R 比 R^* 好，则称这个方法类是完备的。

先验分布已知的情形

问题：如何划分区域 R_1 和 R_2 ，使得错判的平均损失达到最小。

- 对先验概率 q_1 ，来自总体 π_1 的观测值的每一个随机变量小于 $\mathbf{y} = (y_1, \dots, y_p)'$ 对应的分量的概率为：

$$\int_{-\infty}^{y_1} \cdots \int_{-\infty}^{y_p} q_1 f_1(\mathbf{x}) d\mathbf{x}$$

- 计算在给定观测值的条件下，它来自某个总体的概率。如，给定观测值 \mathbf{x} ，它来自总体 π_1 的条件概率为：

$$\frac{q_1 f_1(\mathbf{x})}{q_1 f_1(\mathbf{x}) + q_2 f_2(\mathbf{x})}$$

先验分布已知的情形

- 期望损失，也称为错判概率，定义为：

$$q_1 \int_{R_2} f_1(\mathbf{x}) d\mathbf{x} + q_2 \int_{R_1} f_2(\mathbf{x}) d\mathbf{x} \quad (3.1)$$

- 对于一个给定的观测向量 \mathbf{x} ，我们将把这个个体指定到有最高条件概率的总体，这样就能使错判概率达到最小。
- 如果

$$\frac{q_1 f_1(\mathbf{x})}{q_1 f_1(\mathbf{x}) + q_2 f_2(\mathbf{x})} \geq \frac{q_2 f_2(\mathbf{x})}{q_1 f_1(\mathbf{x}) + q_2 f_2(\mathbf{x})}$$

这时把观测向量 \mathbf{x} 指定到总体 π_1 ，否则指定到总体 π_2 。

- 由于每个观测点上的错判概率都达到最小，因此可以说在整个空间上达到最小，故判别准则是：

$$\begin{aligned} R_1 &= \{\mathbf{x} : q_1 f_1(\mathbf{x}) > q_2 f_2(\mathbf{x})\} \\ R_2 &= \{\mathbf{x} : q_1 f_1(\mathbf{x}) < q_2 f_2(\mathbf{x})\} \end{aligned} \quad (3.2)$$

- 如果 $q_1 f_1(\mathbf{x}) = q_2 f_2(\mathbf{x})$ ，则观测值可以任意判别为来自总体 π_1 或 π_2 ；
- 如果 $q_1 f_1(\mathbf{x}) + q_2 f_2(\mathbf{x}) = 0$ ，则该观测值可以归类为任意区域，但是发生这种情形的概率为0。

- 现在证明判别准则(3.2)是最好的准则。对任意的划分 $R^* = (R_1^*, R_2^*)$, 错判概率为:

$$\begin{aligned} & q_1 \int_{R_2^*} f_1(\mathbf{x}) d\mathbf{x} + q_2 \int_{R_1^*} f_2(\mathbf{x}) d\mathbf{x} \\ &= \int_{R_2^*} [q_1 f_1(\mathbf{x}) - q_2 f_2(\mathbf{x})] d\mathbf{x} + q_2. \end{aligned}$$

- 既然 q_2 是一个给定的数, 故只需要选取划分使上式第一部分达到最小, 也就是说, 使 R_2^* 只包含 $q_1 f_1(\mathbf{x}) - q_2 f_2(\mathbf{x}) < 0$ 的点 \mathbf{x} , 不包含使 $q_1 f_1(\mathbf{x}) - q_2 f_2(\mathbf{x}) > 0$ 的点 \mathbf{x} 。

- 进一步假定，若

$$\Pr \left(\frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} = \frac{q_2}{q_1} \middle| \pi_i \right) = 0, \quad i = 1, 2,$$

则除了零概率集外，Bayes方法是唯一的。

- **总结：**给定非负先验概率 q_1 和 q_2 ，以及非负函数 $f_1(\mathbf{x})$ 和 $f_2(\mathbf{x})$ ，采用的判别准则就是划分(3.2)，使得期望损失或错判概率(3.1)达到最小。

- 考虑不同的损失，使

$$C(2|1) \Pr(2|1, R)q_1 + C(1|2) \Pr(1|2, R)q_2$$

达到最小，则有

$$\int_{R_2} \left([C(2|1)q_1]f_1(\mathbf{x}) - [C(1|2)q_2]f_2(\mathbf{x}) \right) d\mathbf{x} + C(1|2)q_2$$

- 这时判别准则为：

$$R_1 = \{\mathbf{x} : [C(2|1)q_1]f_1(\mathbf{x}) \geq [C(1|2)q_2]f_2(\mathbf{x})\}$$

$$R_2 = \{\mathbf{x} : [C(2|1)q_1]f_1(\mathbf{x}) < [C(1|2)q_2]f_2(\mathbf{x})\}$$

- 由于 $C(2|1)q_1$ 和 $C(1|2)q_2$ 是非负常数，则上面准则可以写成另一种形式：

$$\begin{aligned} R_1 &= \left\{ \mathbf{x} : \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq \frac{C(1|2)q_2}{C(2|1)q_1} \right\} \\ R_2 &= \left\{ \mathbf{x} : \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} < \frac{C(1|2)q_2}{C(2|1)q_1} \right\} \end{aligned} \quad (3.3)$$

定理11.2.1

若 q_1 和 q_2 分别是来自总体 π_1 和 π_2 的观测值的先验概率，其密度函数分别为 $f_1(\mathbf{x})$ 和 $f_2(\mathbf{x})$ ，且来自总体 π_1 的观测值错判到总体 π_2 的损失为 $C(2|1)$ ，而来自总体 π_2 的观测值错判到总体 π_1 的损失为 $C(1|2)$ ，则由式(3.3)定义的准则使判别区域 R_1 和 R_2 的期望损失达到最小。若进一步假定

$$\Pr \left(\frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} = \frac{C(1|2)q_2}{C(2|1)q_1} \middle| \pi_i \right) = 0, \quad i = 1, 2,$$

则除了零概率集外，此准则唯一。

- **问题：**在许多实际判别例子里，统计学家不能对两个总体假定有一个先验概率，这时如何进行判别分析？
- **解决办法：**寻找一类容许方法，也就是不能进一步改进的方法族。
具体证明如下：
- 如果两个总体的分布有相同的支撑集，则所有的Bayes方法是容许的。

- 给定先验概率 q_1, q_2 , 令 $R = (R_1, R_2)$ 是一个Bayes方法, 那么是否存在另一个划分 $R^* = (R_1^*, R_2^*)$, 使得

$$\Pr(1|2, R^*) \leq \Pr(1|2, R) \text{ 和 } \Pr(2|1, R^*) \leq \Pr(2|1, R)$$

同时成立且至少有一个不等号严格成立呢?

- 假定这样的划分存在, 且不妨假定:

$$\Pr(1|2, R^*) < \Pr(1|2, R).$$

由于 R 是一个Bayes方法, 由定理11.2.1, 则有

$$q_1 \Pr(2|1, R) + q_2 \Pr(1|2, R) \leq q_1 \Pr(2|1, R^*) + q_2 \Pr(1|2, R^*).$$

- 这个不等式可以改写成：

$$q_1[\Pr(2|1, R) - \Pr(2|1, R^*)] \leq q_2[\Pr(1|2, R^*) - \Pr(1|2, R)]. \quad (3.4)$$

- 由 R^* 的选择可知，式(3.4)的左端非负，右端非正，故只能两端都等于0。
- 如果 $0 < q_1, q_2 < 1$ ，则有 $\Pr(2|1, R) = \Pr(2|1, R^*)$ 且 $\Pr(1|2, R^*) = \Pr(1|2, R)$ 。导致 R^* 比 R 真好发生矛盾。

先验概率未知的情形

- 如果 $q_1 = 0$, 则有 $q_2 = 1$ 。式(3.4)说明: $\Pr(1|2, R^*) = \Pr(1|2, R)$ 。
对Bayes方法来说, $q_1 = 0$ 说明 R_1 只能包含使 $f_2(\mathbf{x}) = 0$ 的点, 因此, $\Pr(1|2, R) = 0$ 。
- 由于 $\Pr(f_2(\mathbf{x}) = 0|\pi_1) = 0$, 故不妨假设 $R_1 = \{\mathbf{x} : f_2(\mathbf{x}) = 0\}$ 。进而由 $\Pr(1|2, R^*) = 0$ 知, $R_1^* \subset R_1$ 。
- 这时, R_2^* 的区域比 R_2 区域要大, 则

$$\Pr(2|1, R^*) \geq \Pr(2|1, R),$$

这又导致了与假设 R^* 比 R 真好的矛盾。

- 同理, $q_1 = 1$ 时也可以导出矛盾。因此, R 是容许的。

定理11.2.2

若 $\Pr(f_2(\mathbf{x}) = 0 | \pi_1) = 0$ 且 $\Pr(f_1(\mathbf{x}) = 0 | \pi_2) = 0$ ，则每个Bayes方法是容许的。

现在来证明这个定理的逆定理，即每个容许判别方法是Bayes判别方法。

● 假设

$$\Pr\left(\frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} = k \mid \pi_i\right) = 0, \quad i = 1, 2, \quad 0 \leq k \leq \infty, \quad (3.5)$$

则对任意 q_1 ，Bayes方法是唯一的，并且 $f_1(\mathbf{x})/f_2(\mathbf{x})$ 在总体 π_1 和 π_2 下的分布函数都是连续的。

- 令 R 是一个容许的判别方法，则存在 k 使得

$$\Pr(2|1, R) = \Pr\left(\frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \leq k \middle| \pi_1\right) = \Pr(2|1, R^*).$$

- 在这里 R^* 是Bayes方法，符合 $q_2/q_1 = k$ 。
- 由于 R 是容许的， $\Pr(2|1, R) \leq \Pr(2|1, R^*)$ 。由定理11.2.2，可知 R^* 是容许的，则有 $\Pr(2|1, R) \geq \Pr(2|1, R^*)$ ，这时有：
 $\Pr(2|1, R) = \Pr(2|1, R^*)$ 。
- 因此， R 也是一个Bayes方法，由Bayes方法的唯一性，可知 $R = R^*$ 。

定理11.2.3

若式(3.5)成立，则每个容许方法是一个Bayes方法。

定理11.2.3说明：

- Bayes方法是完备的：若 R 是这个集族外的任意方法，可以构造一个Bayes方法 R^* ，使得 $\Pr(2|1, R) = \Pr(2|1, R^*)$ 。又由于 R^* 是容许的，则 $\Pr(2|1, R) \geq \Pr(2|1, R^*)$ 。
- 若Bayes方法族与容许方法族完全相同，则Bayes方法族是最小完备类。

定理11.2.4

若式(3.5)成立，则Bayes方法族是最小完备类。

两个已知多元正态分布的判别

- 考虑两个正态总体 $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ 和 $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$, 其中

- $\boldsymbol{\mu}_i = (\mu_{i1}, \dots, \mu_{ip})'$ 为第 i 个总体的均值向量, $i = 1, 2$

- $\boldsymbol{\Sigma}_i$ 是两个总体的协方差矩阵, $i = 1, 2$

- 对 $i = 1, 2$, 第 i 个总体的密度函数为:

$$f_i(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} |\boldsymbol{\Sigma}_i|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_i)' \boldsymbol{\Sigma}_i^{-1} (\mathbf{x} - \boldsymbol{\mu}_i) \right]$$

- 则似然比可以定义为:

$$\frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} = \frac{|\boldsymbol{\Sigma}_2|^{1/2}}{|\boldsymbol{\Sigma}_1|^{1/2}} \times \frac{\exp \left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_1)' \boldsymbol{\Sigma}_1^{-1} (\mathbf{x} - \boldsymbol{\mu}_1) \right]}{\exp \left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}_2^{-1} (\mathbf{x} - \boldsymbol{\mu}_2) \right]}.$$

当 $\Sigma_1 \neq \Sigma_2$ 时，正态总体的判别

- 当 $\frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq k$ 时，判别区域 R_1 属于总体 π_1 。由于对数函数是单调递增的，该不等式可以写成：

$$\ln f_1(\mathbf{x}) - \ln f_2(\mathbf{x}) \geq \ln k.$$

- 记

$$\delta(\mathbf{x}) = -\frac{1}{2}\mathbf{x}'(\Sigma_1^{-1} - \Sigma_2^{-1})\mathbf{x} + (\mu_1'\Sigma_1^{-1} - \mu_2'\Sigma_2^{-1})\mathbf{x} - \xi.$$

- $\delta(\mathbf{x})$ 包含了二次项： $-\frac{1}{2}\mathbf{x}'(\Sigma_1^{-1} - \Sigma_2^{-1})\mathbf{x}$ ，它是关于 \mathbf{x} 的二次函数，故把 $\delta(\mathbf{x})$ 称为二次判别函数。

当 $\Sigma_1 \neq \Sigma_2$ 时，正态总体的判别

- 这里， $\xi = \frac{1}{2} \ln \left(\frac{|\Sigma_1|}{|\Sigma_2|} \right) + \frac{1}{2} (\mu_1' \Sigma_1^{-1} \mu_1 - \mu_2' \Sigma_2^{-1} \mu_2)$ 。

定理11.3.1：二次判别准则(QDA)

若总体 π_i 的密度函数为 $f_i(\mathbf{x})$, $i = 1, 2$, 对 $\Sigma_1 \neq \Sigma_2$, 则最好的判别区域为：

$$R_1 = \{\mathbf{x} : \delta(\mathbf{x}) \geq \ln k\}, \quad R_2 = \{\mathbf{x} : \delta(\mathbf{x}) < \ln k\}.$$

若先验概率 q_1 和 q_2 已知，则 $k = \frac{q_2 C(1|2)}{q_1 C(2|1)}$ 。

先验概率不存在的情形

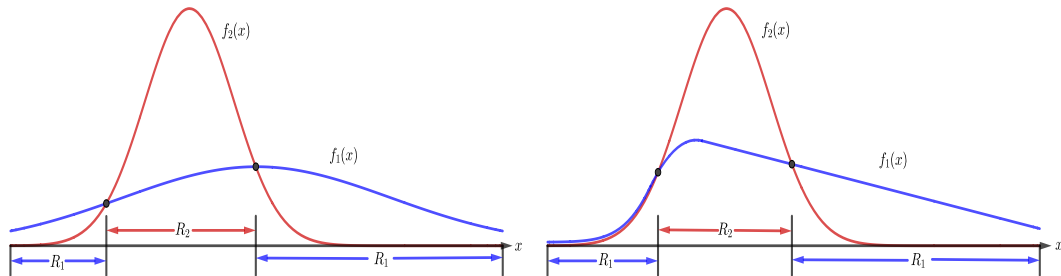


Figure: 当维数 $p = 1$ 时，两个总体的二次判别区域。左图：两个总体 π_1 和 π_2 都为正态分布情形；右图： π_1 总体为非正态分布， π_2 总体为正态分布。

当 $\Sigma_1 = \Sigma_2 = \Sigma$ 时，正态总体的判别

- 当协方差矩阵相等时，即 $\Sigma_1 = \Sigma_2 = \Sigma$ ，则似然比可以定义为：

$$\begin{aligned}\frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} &= \frac{\exp \left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_1)' \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}_1) \right]}{\exp \left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_2)' \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}_2) \right]} \\ &= \exp \left\{ -\frac{1}{2} \left[(\mathbf{x} - \boldsymbol{\mu}_1)' \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}_1) - (\mathbf{x} - \boldsymbol{\mu}_2)' \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}_2) \right] \right\}.\end{aligned}$$

当 $\Sigma_1 = \Sigma_2 = \Sigma$ 时，正态总体的判别

- 当 $\frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq k$ 时，判别区域 R_1 属于总体 π_1 。由于对数函数是单调递增的，该不等式可以写成：

$$-\frac{1}{2} \left[(\mathbf{x} - \boldsymbol{\mu}_1)' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}_1) - (\mathbf{x} - \boldsymbol{\mu}_2)' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}_2) \right] \geq \ln k.$$

- 对上式左边简单计算，有

$$\mathbf{x}' \Sigma^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) - \frac{1}{2} (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)' \Sigma^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2),$$

其中上式左边第一项为判别函数，或称为线性判别函数。

当 $\Sigma_1 = \Sigma_2 = \Sigma$ 时，正态总体的判别

定理11.3.2：线性判别准则(LDA)

若总体 π_i 的密度函数为多元正态密度函数 $f_i(\mathbf{x})$, $i = 1, 2$ ，则最好的判别区域为：

$$R_1 : \mathbf{x}'\Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) - \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)'\Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \geq \ln k,$$

$$R_2 : \mathbf{x}'\Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) - \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)'\Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) < \ln k.$$

若先验概率 q_1 和 q_2 已知，则 $k = \frac{q_2 C(1|2)}{q_1 C(2|1)}$ 。

- 考虑下面的模拟例子，从两个二元正态分布产生随机数据，比较Bayes判别方法、LDA方法和QDA方法的决策边界和错误率。
- 假设有两个正态总体，分别为 $\pi_1 \sim N_2(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ 和 $\pi_2 \sim N_2(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$ ，其中 $\boldsymbol{\mu}_1 = (-1, 1)'$ ， $\boldsymbol{\mu}_2 = (1.5, 2)'$ 。
- 分别考虑下面两种情形：

① $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix}$

② $\begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix} = \boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2 = \begin{pmatrix} 1 & 0.7 \\ 0.7 & 1 \end{pmatrix}$

模拟比较

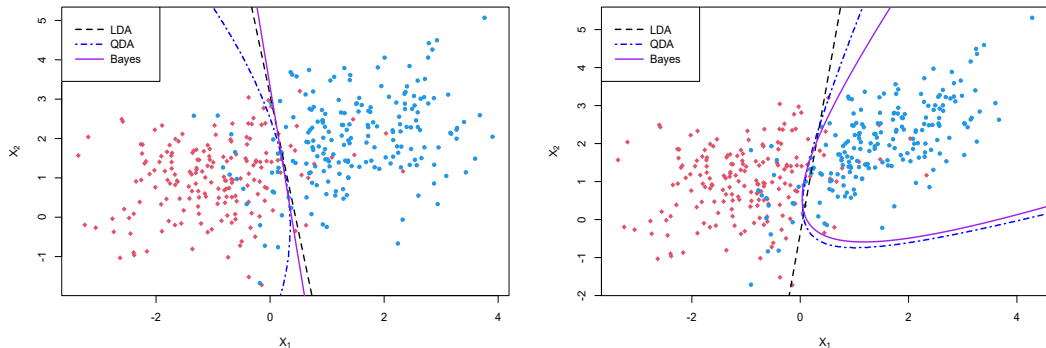


Figure: 左：当 $\Sigma_1 = \Sigma_2 = \Sigma$ 时，Bayes判别方法(紫色实线)、LDA方法(黑色虚线)和QDA 方法(蓝色点断线)的决策边界，三种方法的错误率分别为：0.1、0.108和0.103；右：当 $\Sigma_1 \neq \Sigma_2$ 时，Bayes判别方法(紫色实线)、LDA方法(黑色虚线)和QDA 方法(蓝色点断线)的决策边界，三种方法的错误率分别为：0.087、0.1和0.087。

先验概率不存在的情形

- 选择 $\ln k = c$ 。令 \mathbf{X} 是一个随机观测向量，希望找到下式定义的随机变量 U 的分布：

$$U = \mathbf{X}'\Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) - \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)'\Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2).$$

- 如果 $\mathbf{X} \sim N_p(\boldsymbol{\mu}_1, \Sigma)$ ，则 U 是正态分布，其均值为：

$$\begin{aligned} E_1(U) &= \boldsymbol{\mu}_1'\Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) - \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)'\Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \\ &= \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \end{aligned}$$

- 方差为：

$$\begin{aligned} \text{Var}_1(U) &= (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\Sigma^{-1}E_1[(\mathbf{X} - \boldsymbol{\mu}_1)(\mathbf{X} - \boldsymbol{\mu}_1)']\Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \\ &= (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \end{aligned}$$

- 记 $N_p(\mu_1, \Sigma)$ 和 $N_p(\mu_2, \Sigma)$ 之间的马氏距离为：

$$(\mu_1 - \mu_2)' \Sigma^{-1} (\mu_1 - \mu_2) = \Delta^2$$

- 则 $U \sim N(\Delta^2/2, \Delta^2)$ 。当 $X \sim N_p(\mu_2, \Sigma)$ 时，同样可计算均值为：

$$\begin{aligned} E_2(U) &= \mu_2' \Sigma^{-1} (\mu_1 - \mu_2) - \frac{1}{2} (\mu_1 + \mu_2)' \Sigma^{-1} (\mu_1 - \mu_2) \\ &= \frac{1}{2} (\mu_2 - \mu_1)' \Sigma^{-1} (\mu_1 - \mu_2) = -\frac{1}{2} \Delta^2 \end{aligned}$$

- 方差为： $\text{Var}_2(U) = \text{Var}_1(U)$ ，则 $U \sim N(-\Delta^2/2, \Delta^2)$ 。

先验概率不存在的情形

- 来自总体 π_1 的观测值被错判的概率为：

$$\begin{aligned}\Pr(2|1) &= \int_{-\infty}^c \frac{1}{\sqrt{2\pi}\Delta} \exp\left[-\frac{1}{2\Delta^2}\left(z - \frac{1}{2}\Delta^2\right)^2\right] dz \\ &= \int_{-\infty}^{(c-\Delta^2/2)/\Delta} \frac{1}{\sqrt{2\pi}} \exp(-y^2/2) dy\end{aligned}$$

- 来自总体 π_2 的观测值被错判的概率为：

$$\begin{aligned}\Pr(1|2) &= \int_c^{\infty} \frac{1}{\sqrt{2\pi}\Delta} \exp\left[-\frac{1}{2\Delta^2}\left(z + \frac{1}{2}\Delta^2\right)^2\right] dz \\ &= \int_{(c+\Delta^2/2)/\Delta}^{\infty} \frac{1}{\sqrt{2\pi}} \exp(-y^2/2) dy\end{aligned}$$

先验概率不存在的情形

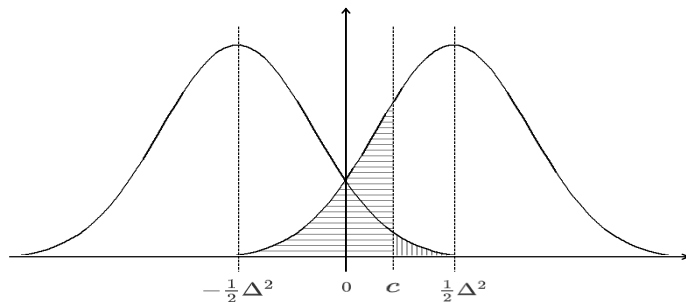


Figure: 两个错判概率的意义。

注：两个错判概率见图中阴影部分。

- 由minimax方法, 选择 c 使得

$$\begin{aligned} & C(1|2) \int_{(c+\Delta^2/2)/\Delta}^{\infty} \frac{1}{\sqrt{2\pi}} \exp(-y^2/2) dy \\ &= C(2|1) \int_{-\infty}^{(c-\Delta^2/2)/\Delta} \frac{1}{\sqrt{2\pi}} \exp(-y^2/2) dy \end{aligned} \quad (4.1)$$

定理11.3.3

若总体 π_i 的密度函数为多元正态密度函数 $f_i(\mathbf{x})$, $i = 1, 2$, 则minimax的判别区域在定理11.3.2中给出, 其中 $c = \ln k$ 由式(4.1)给出, $C(i|j)$ 为错判损失。

先验概率不存在的情形

- 当错判损失相等时，有 $c = 0$ ，且错判概率为：

$$\int_{\Delta/2}^{\infty} \frac{1}{\sqrt{2\pi}} \exp(-y^2/2) dy.$$

- 当错判损失不等的情况下， c 可以尝试由正态分布的性质来决定。
- 注意式(4.1)，两部分都含有共同的向量：

$$\theta = \Sigma^{-1}(\mu_1 - \mu_2).$$

- 问题：

- θ 的几何意义是什么？
- 它为什么会出现在判别函数中？

先验概率不存在的情形

- **夏普指数(Sharpe Ratio):** 在金融领域, 两个分布的中心相对于标准误差的距离, 称为夏普指数。
- 对于两个一元正态总体的判别分析中, 夏普指数越大就越容易判别, 即错判概率越小。
- **解决思路和办法:** 希望把一个多元向量投影到某一个方向上去, 使其最容易判别, 即考虑 $\xi'X$, 使得

$$\frac{[E_1(\xi'X) - E_2(\xi'X)]^2}{\text{Var}(\xi'X)} = \frac{\xi'(\mu_1 - \mu_2)(\mu_1 - \mu_2)'\xi}{\xi'\Sigma\xi}$$

达到最大。

先验概率不存在的情形

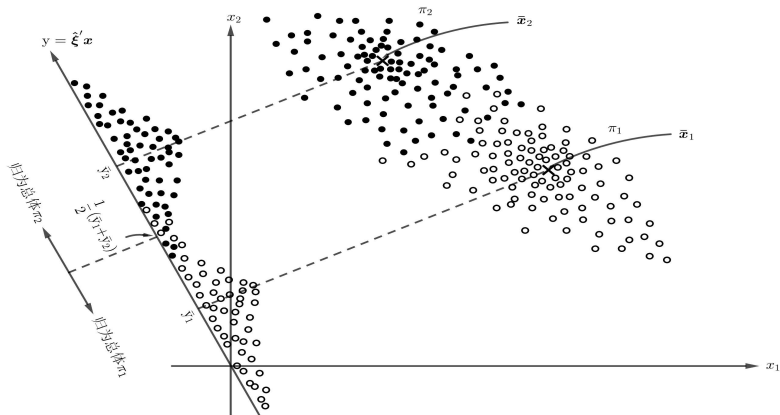
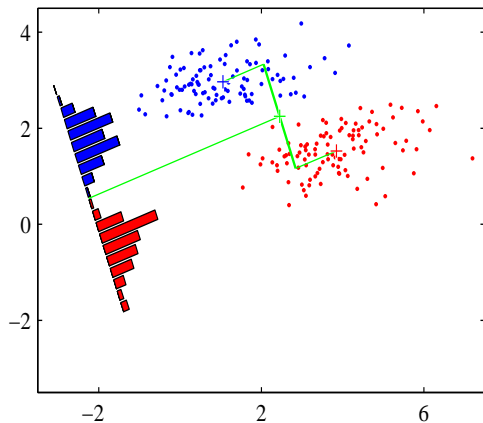
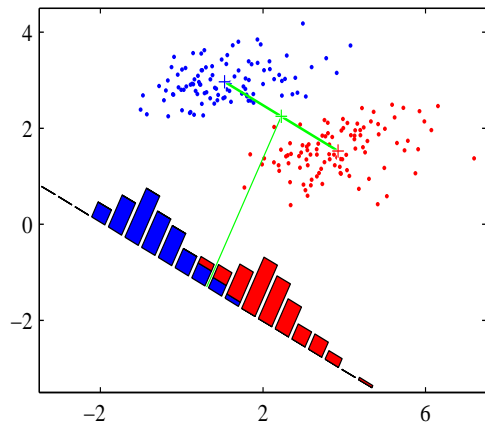


Figure: Fisher判别的示意图。

先验概率不存在的情形



先验概率不存在的情形

- 令 $\beta = \Sigma^{1/2}\xi$ ，代入上式，则有

$$\frac{\beta' \Sigma^{-1/2}(\mu_1 - \mu_2)(\mu_1 - \mu_2)' \Sigma^{-1/2} \beta}{\|\beta\|^2} \quad (4.2)$$

- 使式(4.2)达到最大值的 β 必须与 $\Sigma^{-1/2}(\mu_1 - \mu_2)$ 落在同一个方向上，即 $\beta = a\Sigma^{-1/2}(\mu_1 - \mu_2)$ 。
- 把 $\beta = a\Sigma^{-1/2}(\mu_1 - \mu_2)$ 代入式(4.2)，则式(4.2)得最大值 Δ^2 。
- 把 $\beta = a\Sigma^{-1/2}(\mu_1 - \mu_2)$ 代入 ξ ，则有

$$\xi = \Sigma^{-1/2}\beta = a\Sigma^{-1}(\mu_1 - \mu_2) = a\theta.$$

- 可见： ξ 与 θ 成比例，即 θ 是使得 X 投影最容易的判别方向，所以它会出现在最佳判别函数中。

参数未知时两个正态总体的判别

问题：在实际应用中，如果对两个总体一无所知，如何对观测值进行判别分析？

解决方案：

- 假设对两个总体存在两组历史样本，也称为训练样本，见如下
 - ① 设样本 $\mathbf{x}_1^{(1)}, \dots, \mathbf{x}_{n_1}^{(1)} \sim N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$
 - ② 设样本 $\mathbf{x}_1^{(2)}, \dots, \mathbf{x}_{n_2}^{(2)} \sim N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$
- 根据上面的训练样本，如何把一个新的观测值 \mathbf{x} 判别为来自总体 π_1 或 π_2

● 估计量:

$$① \quad \hat{\mu}_1 = \bar{\mathbf{x}}^{(1)} = \frac{1}{n_1} \sum_{i=1}^{n_1} \mathbf{x}_i^{(1)}$$

$$② \quad \hat{\mu}_2 = \bar{\mathbf{x}}^{(2)} = \frac{1}{n_2} \sum_{i=1}^{n_2} \mathbf{x}_i^{(2)}$$

$$③ \quad \hat{\Sigma} = \mathbf{S} = \frac{\mathbf{V}^{(1)} + \mathbf{V}^{(2)}}{n_1 + n_2 - 2}, \text{ 其中}$$

$$\mathbf{V}^{(1)} = \sum_{i=1}^{n_1} (\mathbf{x}_i^{(1)} - \bar{\mathbf{x}}^{(1)})(\mathbf{x}_i^{(1)} - \bar{\mathbf{x}}^{(1)})',$$

$$\mathbf{V}^{(2)} = \sum_{i=1}^{n_2} (\mathbf{x}_i^{(2)} - \bar{\mathbf{x}}^{(2)})(\mathbf{x}_i^{(2)} - \bar{\mathbf{x}}^{(2)})'$$

分别为两组样本的样本离差矩阵。

- Fisher (1936)建议把参数估计代入式(4.1), 可得下面的判别函数:

$$W(\mathbf{x}) = \mathbf{x}'\hat{\Sigma}^{-1}(\hat{\mu}_1 - \hat{\mu}_2) - \frac{1}{2}(\hat{\mu}_1 + \hat{\mu}_2)'\hat{\Sigma}^{-1}(\hat{\mu}_1 - \hat{\mu}_2). \quad (5.1)$$

讨论:

- 当总体分布的参数已知时, 可以说由式(4.1)定义的判别准则在某种意义下是最好的, 因为它在先验概率已知的情况下, 它能使得期望损失达到最小。
- 当先验概率未知时, 我们可以用容许性来评价判别方法的好坏。但是不能用同样的方法对式(5.1)定义的判别准则进行评价。

判别准则

- 假设有一个来自总体 π_1 或 π_2 的样本 $\mathbf{x}_1, \dots, \mathbf{x}_n$, 下面我们将样本视为一个整体来进行判别。

- 估计量:

① $\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i$

② $\hat{\Sigma} = \mathbf{S} = \frac{\mathbf{V}^{(1)} + \mathbf{V}^{(2)} + \mathbf{V}}{n_1 + n_2 + n - 3}$, 其中 $\mathbf{V} = \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'$ 为新样本的样本离差矩阵。

- 则定义Fisher判别准则为:

$$\left[\bar{\mathbf{x}} - \frac{1}{2}(\hat{\mu}_1 + \hat{\mu}_2) \right]' \hat{\Sigma}^{-1} (\hat{\mu}_1 - \hat{\mu}_2). \quad (5.2)$$

- 如果样本量 n 越大, 错判概率就越小。

● 令

$$\begin{aligned} W &= \mathbf{X}'\mathbf{S}^{-1}(\bar{\mathbf{X}}^{(1)} - \bar{\mathbf{X}}^{(2)}) - \frac{1}{2}(\bar{\mathbf{X}}^{(1)} + \bar{\mathbf{X}}^{(2)})'\mathbf{S}^{-1}(\bar{\mathbf{X}}^{(1)} - \bar{\mathbf{X}}^{(2)}) \\ &= \left[\mathbf{X} - \frac{1}{2}(\bar{\mathbf{X}}^{(1)} + \bar{\mathbf{X}}^{(2)}) \right]' \mathbf{S}^{-1}(\bar{\mathbf{X}}^{(1)} - \bar{\mathbf{X}}^{(2)}), \end{aligned}$$

其中 \mathbf{X} , $\bar{\mathbf{X}}^{(1)}$, $\bar{\mathbf{X}}^{(2)}$ 和 \mathbf{S} 是随机的。

- W 的分布是相当复杂的，因为它不仅依赖于样本量大小，而且还和未知量 Δ^2 有关。

- 令

$$Y_1 = c_1[\mathbf{X} - (n_1 + n_2)^{-1}(n_1\bar{\mathbf{X}}^{(1)} + n_2\bar{\mathbf{X}}^{(2)})], \quad Y_2 = c_2(\bar{\mathbf{X}}^{(1)} - \bar{\mathbf{X}}^{(2)}),$$

其中 $c_1 = \sqrt{(n_1 + n_2)/(n_1 + n_2 + 1)}$ 和 $c_2 = \sqrt{n_1 n_2 / (n_1 + n_2)}$ 。

- 可以算出 Y_1 和 Y_2 分别服从具有协方差矩阵 Σ 的两个独立正态分布，且 Y_2 的数学期望总是 $c_2(\mu_1 - \mu_2)$ ；
- Y_1 的数学期望为：

$$E(Y_1) = \begin{cases} c_1[n_2/(n_1 + n_2)](\mu_1 - \mu_2), & \text{若 } \mathbf{X} \sim \pi_1; \\ -c_1[n_1/(n_1 + n_2)](\mu_1 - \mu_2), & \text{若 } \mathbf{X} \sim \pi_2. \end{cases}$$

- 若记 $\mathbf{Y} = (Y_1, Y_2)$, 且

$$\mathbf{M} = \mathbf{Y}'\mathbf{S}^{-1}\mathbf{Y} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix}$$

- 则有: $W = \sqrt{\frac{n_1 + n_2 + 1}{n_1 n_2}} m_{12} + \frac{n_1 - n_2}{2n_1 n_2} m_{22}$.
- Sitgreaves (1952)给出了 \mathbf{M} 的密度函数, Anderson (1951)和Wald (1944)也对 W 的分布做了研究。

判别准则的渐近分布

如果训练样本足够大，可以通过极限分布理论导出判别准则的渐近分布。

- 由Kolmogorov强大数定律知，当 $n_i \rightarrow \infty$ 时， $\bar{X}^{(i)}$ 是 μ_i 的强相合估计， $i = 1, 2$ 。
- 当 $n_1 + n_2 \rightarrow \infty$ 时， \mathbf{S} 是 Σ 的强相合估计。
- 可以证明， \mathbf{S}^{-1} 存在，并且有

$$\lim \mathbf{S}^{-1} = \Sigma^{-1}, \quad \text{a.s.}$$

- 因此，有

$$\lim_{n_1, n_2 \rightarrow \infty} \mathbf{S}^{-1}(\bar{X}^{(1)} - \bar{X}^{(2)}) = \Sigma^{-1}(\mu_1 - \mu_2),$$

$$\lim_{n_1, n_2 \rightarrow \infty} (\bar{X}^{(1)} + \bar{X}^{(2)})' \mathbf{S}^{-1}(\bar{X}^{(1)} - \bar{X}^{(2)}) = (\mu_1 + \mu_2)' \Sigma^{-1}(\mu_1 - \mu_2).$$

判别准则的渐近分布

- 由上面的讨论，可以看出，当 $n_1, n_2 \rightarrow \infty$ 时， W 的极限分布就是 U 的分布。这时，给出下面的定理。

定理11.4.1

在 W 的定义中， $\bar{X}^{(1)}$ 是由总体 $N_p(\mu_1, \Sigma)$ 中抽取的容量为 n_1 的样本均值向量， $\bar{X}^{(2)}$ 是由总体 $N_p(\mu_2, \Sigma)$ 中抽取的容量为 n_2 的样本均值向量，且 S 是 Σ 的估计量。当 $n_1, n_2 \rightarrow \infty$ 时，且 $X \sim N_p(\mu_1, \Sigma)$ 时，则 W 的极限分布为 $N(\Delta^2/2, \Delta^2)$ ；当 $X \sim N_p(\mu_2, \Sigma)$ 时，则 W 的极限分布为 $N(-\Delta^2/2, \Delta^2)$ 。

似然比准则：从假设检验的角度介绍另一种判别准则，即似然比准则。

● 考虑下面的假设检验问题：

① 原假设 H_0 ：样本 \mathbf{x} , $\mathbf{x}_1^{(1)}, \dots, \mathbf{x}_{n_1}^{(1)}$ 来自总体 $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$,
且样本 $\mathbf{x}_1^{(2)}, \dots, \mathbf{x}_{n_2}^{(2)}$ 来自总体 $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$;

② 备择假设 H_1 ：样本 $\mathbf{x}_1^{(1)}, \dots, \mathbf{x}_{n_1}^{(1)}$ 来自总体 $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$,
且样本 $\mathbf{x}, \mathbf{x}_1^{(2)}, \dots, \mathbf{x}_{n_2}^{(2)}$ 来自总体 $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$ 。

● $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$ 和 $\boldsymbol{\Sigma}$ 是未知的。

● 令

$$\mathbf{A} = \sum_{k=1}^2 \sum_{i=1}^{n_k} (\mathbf{x}_i^{(k)} - \bar{\mathbf{x}}^{(k)})(\mathbf{x}_i^{(k)} - \bar{\mathbf{x}}^{(k)})'.$$

在原假设 H_0 下, μ_1, μ_2 和 Σ 的极大似然估计:

- $\hat{\mu}_{11} = \frac{n_1 \bar{\mathbf{x}}^{(1)} + \mathbf{x}}{n_1 + 1}, \quad \hat{\mu}_{21} = \bar{\mathbf{x}}^{(2)}$

- Σ 的极大似然估计为:

$$\begin{aligned} \hat{\Sigma}_1 = & \frac{1}{n_1 + n_2 + 1} \left[\sum_{i=1}^{n_1} (\mathbf{x}_i^{(1)} - \hat{\mu}_{11})(\mathbf{x}_i^{(1)} - \hat{\mu}_{11})' \right. \\ & \left. + (\mathbf{x} - \hat{\mu}_{11})(\mathbf{x} - \hat{\mu}_{11})' + \sum_{i=1}^{n_2} (\mathbf{x}_i^{(2)} - \hat{\mu}_{21})(\mathbf{x}_i^{(2)} - \hat{\mu}_{21})' \right] \end{aligned}$$

• 由于

$$\begin{aligned} & \sum_{i=1}^{n_1} (\mathbf{x}_i^{(1)} - \hat{\boldsymbol{\mu}}_{11})(\mathbf{x}_i^{(1)} - \hat{\boldsymbol{\mu}}_{11})' + (\mathbf{x} - \hat{\boldsymbol{\mu}}_{11})(\mathbf{x} - \hat{\boldsymbol{\mu}}_{11})' \\ &= \sum_{i=1}^{n_1} (\mathbf{x}_i^{(1)} - \bar{\mathbf{x}}^{(1)})(\mathbf{x}_i^{(1)} - \bar{\mathbf{x}}^{(1)})' + n_1(\bar{\mathbf{x}}^{(1)} - \hat{\boldsymbol{\mu}}_{11})(\bar{\mathbf{x}}^{(1)} - \hat{\boldsymbol{\mu}}_{11})' + (\mathbf{x} - \hat{\boldsymbol{\mu}}_{11})(\mathbf{x} - \hat{\boldsymbol{\mu}}_{11})' \\ &= \sum_{i=1}^{n_1} (\mathbf{x}_i^{(1)} - \bar{\mathbf{x}}^{(1)})(\mathbf{x}_i^{(1)} - \bar{\mathbf{x}}^{(1)})' + \frac{n_1}{n_1 + 1}(\mathbf{x} - \bar{\mathbf{x}}^{(1)})(\mathbf{x} - \bar{\mathbf{x}}^{(1)})' \end{aligned}$$

• 则有

$$\hat{\boldsymbol{\Sigma}}_1 = \frac{1}{n_1 + n_2 + 1} \left[\mathbf{A} + \frac{n_1}{n_1 + 1}(\mathbf{x} - \bar{\mathbf{x}}^{(1)})(\mathbf{x} - \bar{\mathbf{x}}^{(1)})' \right]$$

在备择假设 H_1 下, μ_1, μ_2 和 Σ 的极大似然估计:

- $\hat{\mu}_{12} = \bar{\mathbf{x}}^{(1)}$

- $\hat{\mu}_{22} = \frac{n_2 \bar{\mathbf{x}}^{(2)} + \mathbf{x}}{n_2 + 1}$

- Σ 的极大似然估计为:

$$\hat{\Sigma}_2 = \frac{1}{n_1 + n_2 + 1} \left[\mathbf{A} + \frac{n_2}{n_2 + 1} (\mathbf{x} - \bar{\mathbf{x}}^{(2)}) (\mathbf{x} - \bar{\mathbf{x}}^{(2)})' \right]$$

- 因此，似然比准则为：

$$\frac{|\hat{\Sigma}_2|}{|\hat{\Sigma}_1|} = \frac{\left| \mathbf{A} + \frac{n_2}{n_2 + 1} (\mathbf{x} - \bar{\mathbf{x}}^{(2)}) (\mathbf{x} - \bar{\mathbf{x}}^{(2)})' \right|}{\left| \mathbf{A} + \frac{n_1}{n_1 + 1} (\mathbf{x} - \bar{\mathbf{x}}^{(1)}) (\mathbf{x} - \bar{\mathbf{x}}^{(1)})' \right|}$$

- 令 $n = n_1 + n_2 - 2$ ，进一步可以改写成：

$$\frac{|\hat{\Sigma}_2|}{|\hat{\Sigma}_1|} = \frac{1 + \frac{n_2}{n_2 + 1} (\mathbf{x} - \bar{\mathbf{x}}^{(2)})' \mathbf{A}^{-1} (\mathbf{x} - \bar{\mathbf{x}}^{(2)})}{1 + \frac{n_1}{n_1 + 1} (\mathbf{x} - \bar{\mathbf{x}}^{(1)})' \mathbf{A}^{-1} (\mathbf{x} - \bar{\mathbf{x}}^{(1)})} = \frac{n + \frac{n_2}{n_2 + 1} (\mathbf{x} - \bar{\mathbf{x}}^{(2)})' \mathbf{S}^{-1} (\mathbf{x} - \bar{\mathbf{x}}^{(2)})}{n + \frac{n_1}{n_1 + 1} (\mathbf{x} - \bar{\mathbf{x}}^{(1)})' \mathbf{S}^{-1} (\mathbf{x} - \bar{\mathbf{x}}^{(1)})}.$$

- 对给定的 $K_n > 0$, 当 $\frac{|\hat{\Sigma}_2|}{|\hat{\Sigma}_1|} \geq K_n$ 时, 判别区域落在总体 π_1 中, 即

$$\begin{aligned} R_1 &= \left\{ \mathbf{x} : n + \frac{n_2}{n_2 + 1} (\mathbf{x} - \bar{\mathbf{x}}^{(2)})' \mathbf{S}^{-1} (\mathbf{x} - \bar{\mathbf{x}}^{(2)}) \right. \\ &\quad \left. \geq K_n \left[n + \frac{n_1}{n_1 + 1} (\mathbf{x} - \bar{\mathbf{x}}^{(1)})' \mathbf{S}^{-1} (\mathbf{x} - \bar{\mathbf{x}}^{(1)}) \right] \right\} \end{aligned}$$

- 若令 $K_n = 1 + 2c/n$, 且 n_1 和 n_2 都很大, 则 R_1 近似为 $\{\mathbf{x} : W(\mathbf{x}) \geq c\}$ 。
- 若令 $K_n = 1$, 则当 $\frac{|\hat{\Sigma}_2|}{|\hat{\Sigma}_1|} \geq 1$ 时, 判定为 π_1 ;

- 当 $\frac{|\hat{\Sigma}_2|}{|\hat{\Sigma}_1|} < 1$ 时, 判定为 π_2 。这就是极大似然比准则。

- 令

$$Z = \frac{1}{2} \left[\frac{n_2}{n_2 + 1} (\mathbf{x} - \bar{\mathbf{x}}^{(2)})' \mathbf{S}^{-1} (\mathbf{x} - \bar{\mathbf{x}}^{(2)}) - \frac{n_1}{n_1 + 1} (\mathbf{x} - \bar{\mathbf{x}}^{(1)})' \mathbf{S}^{-1} (\mathbf{x} - \bar{\mathbf{x}}^{(1)}) \right]$$

- 极大似然准则是: 当 $Z > 0$, 归类为 π_1 ; 当 $Z < 0$, 归类为 π_2 。
- 容易看到: 把观测值 \mathbf{x} 判别到哪个总体, 就看它到哪个总体的样本均值的马氏距离比较近。如, 到 $\bar{\mathbf{x}}^{(1)}$ 的距离近, 就判别到 π_1 , 否则判别到 π_2 。

- 计算 W 与 Z 之间的差为：

$$W - Z = \frac{1}{2} \left[\frac{1}{n_2 + 1} (\mathbf{x} - \bar{\mathbf{x}}^{(2)})' \mathbf{S}^{-1} (\mathbf{x} - \bar{\mathbf{x}}^{(2)}) - \frac{1}{n_1 + 1} (\mathbf{x} - \bar{\mathbf{x}}^{(1)})' \mathbf{S}^{-1} (\mathbf{x} - \bar{\mathbf{x}}^{(1)}) \right]$$

- 当 $n_1, n_2 \rightarrow \infty$ 时, $W - Z \xrightarrow{P} 0$ 。因此, 从大样本的角度可以看出, 判别准则 W 和 Z 导致的错判概率是渐近相等的。
- 当 $n_1 = n_2$, $Z = [n_1 / (n_1 + 1)]W$, 对于临界点 $c = 0$ 而言, 用 W 和 Z 进行判别是等价的。

- 在实际应用中，当决定抽取这两个训练样本之前，需要预先知道错判概率；
- 或在抽取这两个训练样本之后，需要计算错判的条件概率；
- 但是如前面所述， W 和 Z 的精确分布很难得到，这时很难计算错判概率。
- 当 $n_1, n_2 \rightarrow \infty$ 时，前面已经证明：
 - ① 若 \mathbf{x} 来自总体 π_1 时， W 和 Z 的极限分布是 $N(\Delta^2/2, \Delta^2)$ ；
 - ② 若 \mathbf{x} 来自总体 π_2 时， W 和 Z 的极限分布是 $N(-\Delta^2/2, \Delta^2)$ 。
- 因此，当 $n_1, n_2 \rightarrow \infty$ 时，可以利用它们的极限分布计算错判概率的渐近展开。

基于 W 错判概率的渐近展开

- Okamoto (1963)将 W 的分布展开到 n^{-2} 阶;
- Siotani 和Wang (1975,1977)将 W 的分布展开到 n^{-3} 阶。
- 令 $\Phi(\cdot)$ 和 $\phi(\cdot)$ 分别表示标准正态分布 $N(0, 1)$ 的分布函数和密度函数。

定理11.5.1

当 $n_1 \rightarrow \infty, n_1/n_2 \rightarrow \kappa > 0, n = n_1 + n_2 - 2$ 时, 对任何的 u , 有

$$\begin{aligned} \Pr\left(\frac{W - \frac{1}{2}\Delta^2}{\Delta} \leq u \middle| \pi_1\right) &= \Phi(u) - \phi(u) \left\{ \frac{1}{2n_1\Delta^2} [u^3 + (p-3)u - p\Delta] \right. \\ &\quad + \frac{1}{2n_2\Delta^2} [u^3 + 2\Delta u^2 + (p-3+\Delta^2)u + (p-2)\Delta] \\ &\quad \left. + \frac{1}{4n} [4u^3 + 4\Delta u^2 + (6p-6+\Delta^2)u + 2(p-1)\Delta] \right\} + O(n^{-2}). \end{aligned}$$

当 n_1 和 n_2 互换时, $\Pr(-(W + \Delta^2/2)/\Delta \leq u | \pi_2)$ 也是上式。

统计量 W 的判别准则：

- 若 $W(\mathbf{x}) > c$ ，则认为观测值 \mathbf{x} 来自于总体 π_1 ；错判概率由定理11.5.1给出，均值为 $u = (c - \Delta^2/2)/\Delta$ ；
- 若 $W(\mathbf{x}) \leq c$ ，则认为观测值 \mathbf{x} 来自于总体 π_2 ；错判概率由定理11.5.1给出，均值为 $u = -(c + \Delta^2/2)/\Delta$ ；
- 当 $c = 0, u = -\Delta/2$ 时，若 $n_1 = n_2$ ，则所定义的判别是minimax 的。

推论11.5.1

当 n_1 和 n_2 充分接近时, 则

$$\begin{aligned}\Pr\left(W \leq 0 \middle| \pi_1\right) &= \Phi\left(-\frac{\Delta}{2}\right) + \frac{1}{n}\phi\left(\frac{\Delta}{2}\right)\left(\frac{p-1}{\Delta} + \frac{p}{4}\Delta\right) + o(n^{-1}) \\ &= \Pr\left(W \geq 0 \middle| \pi_2\right).\end{aligned}$$

- 渐近展开式中的第一项是均值向量和协方差矩阵已知时的错判概率;
- 第二项是修正项, 它永远是正的;
- 说明由训练样本得到的错判概率比参数已知时的错判概率要大;
- 修正项随着维数 p 的增大而增大, 随着 Δ 和 n 的增大而减小, 且修正项的阶数是 n^{-1} 。

- 由于 Δ 未知，但它与学生化的 W 有关，则样本的马氏平方距离为：

$$D^2 = (\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)})' \mathbf{S}^{-1} (\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)}).$$

- 可见，它是两个总体间马氏平方距离 Δ^2 的一个估计。
- D^2 的期望为：

$$E(D^2) = \frac{n}{n-p-1} \left[\Delta^2 + p \left(\frac{1}{n_1} + \frac{1}{n_2} \right) \right].$$

- 显然，当 $n_1 \rightarrow \infty$ 和 $n_2 \rightarrow \infty$ 时， $E(D^2) \rightarrow \Delta^2$ 。

Anderson (1973)给出了下面定理:

定理11.5.2

若 $\lim_{n \rightarrow \infty} \frac{n_1}{n_2} \rightarrow \kappa > 0$ 时, 其中 $n = n_1 + n_2 - 2$, 对任何的 u , 则

$$\Pr \left(\frac{W - \frac{1}{2}D^2}{D} \leq u \middle| \pi_1 \right) = \Phi(u) - \phi(u) \left\{ \frac{1}{n_1} \left(\frac{u}{2} - \frac{p-1}{\Delta} \right) + \frac{1}{n} \left[\frac{u^3}{4} + \left(p - \frac{3}{4} \right) u \right] \right\} + O(n^{-2}),$$
$$\Pr \left(-\frac{W + \frac{1}{2}D^2}{D} \leq u \middle| \pi_2 \right) = \Phi(u) - \phi(u) \left\{ \frac{1}{n_2} \left(\frac{u}{2} - \frac{p-1}{\Delta} \right) + \frac{1}{n} \left[\frac{u^3}{4} + \left(p - \frac{3}{4} \right) u \right] \right\} + O(n^{-2}).$$

基于 W 错判概率的渐近展开

- 可以通过选择临界点 c 来控制某一个错判概率, 令 $\Pr(W < c | \pi_1) = \alpha$, Anderson (1973)给出了下面的定理。

定理11.5.3

令 u_0 使得 $\Phi(u_0) = \alpha$, 以及

$$u = u_0 - \frac{1}{n_1} \left(\frac{p-1}{D} - \frac{u_0}{2} \right) + \frac{1}{n} \left[\left(p - \frac{3}{4} \right) u_0 + \frac{u_0^3}{4} \right],$$

当 $n_1 \rightarrow \infty, n_2 \rightarrow \infty, n_1/n_2 \rightarrow \kappa > 0$ 时, 则

$$\Pr \left(\frac{W - \frac{1}{2}D^2}{D} \leq u \middle| \pi_1 \right) = \alpha + O(n^{-2}),$$

且 $c = Du + D^2/2$ 能达到期望的概率 α 。

基于 W 错判概率的渐近展开

下面计算给定两个训练样本后的条件错判概率：

- 给定 $\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}$ 和 \mathbf{S} ，当 \mathbf{x} 来自总体 $\pi_k, k = 1, 2$ ，因为随机变量 W 是正态分布

① 则其条件均值为：

$$\begin{aligned} E(W|\pi_k, \bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}, \mathbf{S}) &= \left[\boldsymbol{\mu}_k - \frac{1}{2}(\bar{\mathbf{x}}^{(1)} + \bar{\mathbf{x}}^{(2)}) \right]' \mathbf{S}^{-1}(\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)}) \\ &=: u_k(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}, \mathbf{S}) \end{aligned}$$

② 则其条件方差为：

$$\begin{aligned} \text{Var}(W|\pi_k, \bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}, \mathbf{S}) &= (\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)})' \mathbf{S}^{-1} \boldsymbol{\Sigma} \mathbf{S}^{-1} (\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)}) \\ &=: \sigma^2(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}, \mathbf{S}) \end{aligned}$$

- 注意到：条件均值和条件方差是样本的函数，且极限为：

$$\lim_{n_1, n_2 \rightarrow \infty} u_k(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}, \mathbf{S}) = (-1)^{k-1} \frac{1}{2} \Delta^2,$$

$$\lim_{n_1, n_2 \rightarrow \infty} \sigma^2(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}, \mathbf{S}) = \Delta^2.$$

- 说明：对于大的样本量 n_1, n_2 来说，条件错判概率近似于参数已知时的错判概率。

- 如果 c 是临界点, 则给定 $\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}$ 和 \mathbf{S} 的条件错判概率分别为:

$$\Pr(2|1, c, \bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}, \mathbf{S}) = \Phi \left[\frac{c - u_1(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}, \mathbf{S})}{\sigma(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}, \mathbf{S})} \right], \quad (6.1)$$

$$\Pr(1|2, c, \bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}, \mathbf{S}) = 1 - \Phi \left[\frac{c - u_2(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}, \mathbf{S})}{\sigma(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}, \mathbf{S})} \right]. \quad (6.2)$$

基于 W 错判概率的渐近展开

- 在式(6.1)中, 令 $c = Du_1 + D^2/2$, 则式(6.1)中 $\Phi(\cdot)$ 的变量变为:

$$u_1 D / \sigma + (\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)})' \mathbf{S}^{-1} (\bar{\mathbf{x}}^{(1)} - \boldsymbol{\mu}_1) / \sigma.$$

- 当 $n_1 \rightarrow \infty, n_2 \rightarrow \infty$ 时, 则

$$u_1 D / \sigma \xrightarrow{P} u_1$$

$$(\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)})' \mathbf{S}^{-1} (\bar{\mathbf{x}}^{(1)} - \boldsymbol{\mu}_1) / \sigma \xrightarrow{P} 0$$

- 这时式(6.1)收敛到 $\Phi(u_1)$ 。

基于 W 错判概率的渐近展开

- 在式(6.2)中, 令 $c = Du_2 - D^2/2$, 则式(6.2)中 $\Phi(\cdot)$ 的变量变为:

$$u_2 D / \sigma + (\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)})' \mathbf{S}^{-1} (\bar{\mathbf{x}}^{(2)} - \boldsymbol{\mu}_2) / \sigma.$$

- 当 $n_1 \rightarrow \infty, n_2 \rightarrow \infty$ 时, 则

$$u_2 D / \sigma \xrightarrow{P} u_2$$

$$(\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)})' \mathbf{S}^{-1} (\bar{\mathbf{x}}^{(2)} - \boldsymbol{\mu}_2) / \sigma \xrightarrow{P} 0$$

- 这时式(6.2)收敛到 $1 - \Phi(u_2)$ 。

基于 W 错判概率的渐近展开

- 对于式(6.1), 考虑当 $c = Du_1 + D^2/2$, 通过选择适当的 u_1 来控制条件错判概率 $\Pr(2|1, c, \bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}, \mathbf{S})$ 。
- 作为 $\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}$ 和 \mathbf{S} 的函数, 这个条件错判概率是一个随机变量, 它的分布是可以逼近的。

定理11.5.4

当 $n_1 \rightarrow \infty, n_1/n_2 \rightarrow \kappa > 0$ 时, 则

$$\begin{aligned} & \Pr \left(\frac{\Pr \left(2|1, Du_1 + \frac{1}{2}D^2, \bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}, \mathbf{S} \right) - \Phi(u_1)}{\phi(u_2) \left(\frac{1}{2}u_1^2 + \frac{n}{n_1} \right)^{1/2}} \leq x \right) \\ &= \Phi \left(x - \frac{(p-1)\frac{n}{n_1} - \left(p - \frac{3}{4} + \frac{n}{n_1} \right) u_1 - u_1^3/4}{\sqrt{n} \left(\frac{1}{2}u_1^2 + \frac{n}{n_1} \right)^{1/2}} \right) + O(n^{-2}). \end{aligned}$$

基于Z错判概率的渐近展开

- Memon 和Okamoto (1971)将Z的分布展开到 n^{-2} 阶;
- Siotani 和Wang (1975,1977)将Z的分布展开到 n^{-3} 阶。

定理11.5.5

当 $n_1 \rightarrow \infty, n_1/n_2 \rightarrow \kappa > 0, n = n_1 + n_2 - 2$ 时, 对任何的 u , 则

$$\begin{aligned} \Pr \left(\frac{Z - \frac{1}{2}\Delta^2}{\Delta} \leq u \middle| \pi_1 \right) &= \Phi(u) - \phi(u) \left\{ \frac{1}{2n_1\Delta^2} [u^3 + \Delta u^2 \right. \\ &\quad \left. + (p-3)u - \Delta] + \frac{1}{2n_2\Delta^2} [u^3 + \Delta u^2 + (p-3-\Delta^2)u - \Delta^3 - \Delta] \right. \\ &\quad \left. + \frac{1}{4n} [4u^3 + 4\Delta u^2 + (6p-6+\Delta^2)u + 2(p-1)\Delta] \right\} + O(n^{-2}). \end{aligned}$$

当 n_1 和 n_2 互换时, $\Pr(-(Z + \Delta^2/2)/\Delta \leq u | \pi_2)$ 也是上式。

统计量 Z 的判别准则：

- 若 $Z > c$ ，则认为观测值 \mathbf{x} 来自于总体 π_1 ；错判概率由定理11.5.5给出，均值为 $u = (c - \Delta^2/2)/\Delta$ ；
- 若 $Z \leq c$ ，则认为观测值 \mathbf{x} 来自于总体 π_2 ；错判概率由定理11.5.5给出，均值为 $u = -(c + \Delta^2/2)/\Delta$ ；
- 当 $c = 0, u = -\Delta/2$ 时，若 $n_1 = n_2$ ，则用 Z 的判别准则和用 W 的判别准则是等价的。

基于Z错判概率的渐近展开

Fujikoshi 和Kanazawa (1976)给出了下面定理:

定理11.5.6

若 $\lim_{n \rightarrow \infty} \frac{n_1}{n_2} \rightarrow \kappa > 0$ 时, 其中 $n = n_1 + n_2 - 2$, 对任何的 u , 则

$$\begin{aligned} \Pr \left(\frac{Z - \frac{1}{2}D^2}{D} \leq u \middle| \pi_1 \right) &= \Phi(u) - \phi(u) \left\{ \frac{1}{2n_1\Delta} [u^2 + \Delta u - (p-1)] \right. \\ &\quad \left. - \frac{1}{2n_2\Delta} [u^2 + 2\Delta u + p - 1 + \Delta^2] + \frac{1}{4n} [u^3 + (4p-3)u] \right\} + o(n^{-2}), \\ \Pr \left(-\frac{Z + \frac{1}{2}D^2}{D} \leq u \middle| \pi_2 \right) &= \Phi(u) - \phi(u) \left\{ -\frac{1}{2n_1\Delta} [u^2 + 2\Delta u \right. \\ &\quad \left. + (p-1) + \Delta^2] + \frac{1}{2n_2\Delta} [u^2 + \Delta u - (p-1)] + \frac{1}{4n} [u^3 + (4p-3)u] \right\} + o(n^{-2}). \end{aligned}$$

基于Z错判概率的渐近展开

- Kanazawa (1979)给出了下面定理:

定理11.5.7

令 u_0 使得 $\Phi(u_0) = \alpha$, 以及

$$u = u_0 + \frac{1}{2n_1 D} [u_0^2 + Du_0 - (p-1)] \\ - \frac{1}{2n_2 D} [u_0^2 + Du_0 + (p-1) - D^2] + \frac{1}{4n} [u_0^3 + (4p-5)u_0],$$

当 $n_1 \rightarrow \infty, n_1/n_2 \rightarrow \kappa > 0$ 时, 则

$$\Pr \left(\frac{Z - \frac{1}{2}D^2}{D} \leq u \middle| \pi_1 \right) = \alpha + O(n^{-2}).$$

基于Z错判概率的渐近展开

- 考虑训练样本给定后的错判概率，除非当 $n_1 = n_2$ 时， Z 是 \mathbf{x} 的二次型，所以 Z 的条件分布不是正态的，这时没有类似式(6.1)和(6.2)的表达式。
- Siotani (1980)证明了下面的定理。

定理11.5.8

当 $n_1 \rightarrow \infty, n_1/n_2 \rightarrow \kappa > 0$ 时，则

$$\begin{aligned} & \Pr \left(2\sqrt{\frac{n_1 n_2}{n_1 + n_2}} \frac{\Pr(2|1, 0, \bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}, \mathbf{S}) - \Phi(-\frac{1}{2}\Delta)}{\phi(\frac{1}{2}\Delta)} \leq x \right) \\ &= \Phi \left(x - 2\sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left\{ \frac{1}{16n_1 \Delta} [4(p-1) - \Delta^2] \right. \right. \\ & \quad \left. \left. + \frac{1}{16n_2} [4(p-1) + 3\Delta^2] - \frac{(p-1)\Delta}{4n} \right\} \right) + o(n^{-2}). \end{aligned}$$

多个总体的判别

现在考虑多个总体的判别问题：

- 令 π_1, \dots, π_K 是 K 个总体，其密度函数分别为

$$f_1(\mathbf{x}), \dots, f_K(\mathbf{x})$$

- 将整个空间划分成 K 个互不相交的区域 R_1, \dots, R_K
- 若观测值 \mathbf{x} 落在区域 R_i 中，则把它判定为来自总体 π_i
- 令来自总体 π_i 的观测值错判到 $\pi_j (j \neq i)$ 的损失效应为 $C(j|i)$ ，则错判概率为：

$$\Pr(j|i, R) = \int_{R_j} f_i(\mathbf{x}) d\mathbf{x}.$$

多个总体的判别

- 假设对每个总体，有先验概率 q_1, \dots, q_K ，则期望损失为：

$$\sum_{i=1}^K q_i \left\{ \sum_{j=1, j \neq i}^K C(j|i) \Pr(j|i, R) \right\}.$$

- 目的：**选择区域 R_1, \dots, R_K 使得上面的期望损失达到最小。
- 如何完成上面的目标？

多个总体的判别

- 由于已知总体的先验概率，可以定义观测值的条件概率，观测来自总体 π_i 的条件概率为：

$$\frac{q_i f_i(\mathbf{x})}{\sum_{k=1}^K q_k f_k(\mathbf{x})}.$$

- 若将观测值 \mathbf{x} 判别到总体 π_j 中，则给定 \mathbf{x} 时其条件期望损失为：

$$\sum_{i=1, i \neq j}^K \frac{q_i f_i(\mathbf{x}) C(j|i)}{\sum_{k=1}^K q_k f_k(\mathbf{x})}. \quad (7.1)$$

- 选择 j 使得式(7.1)达到最小，也就是使期望损失达到最小。

多个总体的判别

- 对所有的 j , 考虑

$$\sum_{i=1, i \neq j}^K q_i f_i(\mathbf{x}) C(j|i). \quad (7.2)$$

- 选择使式(7.2)最小的 j 。通过该方法, 可以定义 $\mathbf{x} \in R_j$ 。
- 对每一个 \mathbf{x} , 可以使用上述方法划分区域 R_1, \dots, R_K 。
- 判别方法:** 当观测值 \mathbf{x} 落在区域 R_j 时, 就将它归类为来自总体 π_j 。

多个总体的判别

定理11.6.1

假设观测值来自总体 π_i 的先验概率为 q_i , π_i 的密度函数为 $f_i(\mathbf{x})$, $i = 1, \dots, K$, 来自总体 π_i 的观测值被错误地归类到总体 π_j 的损失为 $C(j|i)$, 则使得期望损失最小的判别方法可由下式定义。如果

$$\sum_{i=1, i \neq k}^K q_i f_i(\mathbf{x}) C(\textcolor{red}{k}|i) < \sum_{i=1, i \neq j}^K q_i f_i(\mathbf{x}) C(\textcolor{red}{j}|i), \quad j = 1, \dots, K, \quad j \neq k, \quad (7.3)$$

则将观测值 \mathbf{x} 归类为 R_k 。对于任意的 k 和 j , 若式(7.3)左右两边相等的概率为0, 则除了零概率集之外最优判别方法是唯一的。

多个总体的判别

证明： 令 $h_j(\mathbf{x}) = \sum_{i=1, i \neq j}^K q_i f_i(\mathbf{x}) C(j|i)$, 则判别方法 R 的期望损失, 即风险为

$$\sum_{j=1}^K \int_{R_j} h_j(\mathbf{x}) d\mathbf{x} = \int h(\mathbf{x}|R) d\mathbf{x},$$

其中如果 $\mathbf{x} \in R_j$, $h(\mathbf{x}|R) = h_j(\mathbf{x})$ 。在定理11.6.1 中描述的Bayes方法 R^* , 满足 $h(\mathbf{x}|R^*) = \min_i h_i(\mathbf{x})$ 。因此, 任意判别方法 R 和Bayes方法 R^* 的风险之差为:

$$\int [h(\mathbf{x}|R) - h(\mathbf{x}|R^*)] d\mathbf{x} = \sum_j \int_{R_j} [h_j(\mathbf{x}) - \min_i h_i(\mathbf{x})] d\mathbf{x} \geq 0.$$

上式等号成立当且仅当除了零概率集外 $h_j(\mathbf{x}) = \min_i h_i(\mathbf{x})$ 对任意的 $\mathbf{x} \in R_j$ 成立。

多个总体的判别

现在考虑损失 $C(j|i) = 1, \forall i, j, i \neq j$ 的情形：

- 当 $\mathbf{x} \in R_k$ 时，

$$\sum_{i=1, i \neq k}^K q_i f_i(\mathbf{x}) < \sum_{i=1, i \neq j}^K q_i f_i(\mathbf{x}), \quad j \neq k \quad (7.4)$$

- 从式(7.4)两边减去 $\sum_{i=1, i \neq k, j}^K q_i f_i(\mathbf{x})$ ，可得

$$q_j f_j(\mathbf{x}) < q_k f_k(\mathbf{x}), \quad j \neq k \quad (7.5)$$

- 可知，当点 $\mathbf{x} \in R_k$ 时， k 就是那个使得 $q_i f_i(\mathbf{x})$ 达到最大值的下标。即断定 π_k 是最可能的总体。

多个总体的判别

- 假定没有先验概率，不能定义判别准则的无条件期望损失。
- 但在观测值来自给定总体的条件下，可以定义一个期望损失，若观测值来自总体 π_i ，则

$$\sum_{j=1, j \neq i}^K C(j|i) \Pr(j|i, R) = r(i, R).$$

- ① 若 $r(i, R) \leq r(i, R^*)$, $i = 1, \dots, m$ ，则说明方法 R 至少和 R^* 一样好；
- ② 若有一个不等号严格成立，则说方法 R 比 R^* 真好；
- ③ **容许性**：若没有方法 R^* 比 R 真好，则说 R 是容许的；
- ④ **完备性**：若对给定判别方法类外的每一个判别方法 R ，判别方法类中都有一个方法 R^* 比 R 好，则称该判别方法类是完备的。

多个总体的判别

现在来证明Bayes方法是容许的：

- 令 R 是一个Bayes判别方法， R^* 是另一个判别方法。由于 R 是Bayes方法，则

$$\sum_{i=1}^K q_i r(i, R) \leq \sum_{i=1}^K q_i r(i, R^*).$$

- 假设 $q_1 > 0, q_2 > 0, r(2, R^*) \leq r(2, R)$ 且 $r(i, R^*) \leq r(i, R), i = 3, \dots, K$, 则

$$q_1 [r(1, R) - r(1, R^*)] \leq \sum_{i=2}^K q_i [r(i, R^*) - r(i, R)] < 0$$

- 显然， $r(1, R) < r(1, R^*)$ 。因此， R^* 不会比 R 真好。

多个总体的判别

定理11.6.2

若先验概率 $q_i > 0, i = 1, \dots, K$, 则Bayes方法是容许的。

定理11.6.3

若 $C(i|j) = 1, i \neq j$ 且 $\Pr\{f_i(\mathbf{x}) = 0 | \pi_j\} = 0$, 则任何Bayes方法都是容许的。

定理11.6.3的证明作为练习, 课下参考Anderson (2003)。

定理11.6.4

每一个容许的判别方法是Bayes判别方法。

多个多元正态分布的判别

- 令 $N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$ 是 π_i 的分布，密度函数用 $f_i(\mathbf{x})$ 表示， $i = 1, \dots, K$ 。
- 假定错判损失是相等的，则使用函数

$$\begin{aligned} u_{jk}(\mathbf{x}) &= \ln \frac{f_j(\mathbf{x})}{f_k(\mathbf{x})} \\ &= \left[\mathbf{x} - \frac{1}{2}(\boldsymbol{\mu}_j + \boldsymbol{\mu}_k) \right]' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_j - \boldsymbol{\mu}_k). \end{aligned} \quad (8.1)$$

- 如果先验概率已知，则区域 R_j 有如下的定义：

$$R_j = \left\{ \mathbf{x} : u_{jk}(\mathbf{x}) > \ln \frac{q_k}{q_j}, k = 1, \dots, K, k \neq j \right\}. \quad (8.2)$$

多个多元正态分布的判别

定理11.7.1

若 q_i 是观测值来自 $\pi_i = N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}), i = 1, \dots, K$ 的先验概率, 且错判损失是相等的, 则由式(8.2)的判别区域 R_1, \dots, R_K 使得期望损失达到最小, 其中 $u_{jk}(\mathbf{x})$ 由式(8.1)所定义, 其中 $j \neq k$ 。

- ① $u_{jk}(\mathbf{x})$ 是与第 j 个总体和第 k 个总体相关的判别函数, 且 $u_{jk}(\mathbf{x}) = -u_{kj}(\mathbf{x})$ 。
- ② 既然这些判别函数是线性的, 所以区域 R_i 是被某些超平面界定的。
- ③ 对于每个 $j \neq k$, 不等式(8.2)是用超平面 $u_{jk}(\mathbf{x}) = \ln(q_k/q_j)$ 把欧氏空间 \mathbb{R}^p 分成了两部分, R_j 落在 $u_{jk} > \ln(q_k/q_j)$ 的那半个空间里。

多个多元正态分布的判别

- 在先验概率未知的情况下，区域 R_j 由下列不等式定义：

$$u_{jk}(\mathbf{x}) \geq c_j - c_k, \quad k = 1, \dots, K, \quad k \neq j, \quad (8.3)$$

其中 c_k 可取为负。这些区域形成一个容许方法族。

- 对于minimax方法，常数的选取方法是使所有的 $\Pr(i|i, R)$ 都相等。
- 现在来看如何计算正确判别的概率。若 \mathbf{X} 是一个随机观测向量，考虑下面的随机变量：

$$U_{ji} = \left[\mathbf{X} - \frac{1}{2}(\boldsymbol{\mu}_j + \boldsymbol{\mu}_i) \right]' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_j - \boldsymbol{\mu}_i) = -U_{ij}.$$

多个多元正态分布的判别

如果这 K 个总体的均值向量张成一个 $K - 1$ 维空间(即指 K 个均值向量线性无关), 则可以利用上述 $K(K - 1)/2$ 个判别函数来进行判别。方法如下:

- 若 X 来自总体 π_j , 则 U_{ji} 的分布为 $N(\Delta_{ji}^2/2, \Delta_{ji}^2)$, 其中

$$\Delta_{ji}^2 = (\boldsymbol{\mu}_j - \boldsymbol{\mu}_i)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_j - \boldsymbol{\mu}_i).$$

- U_{ji} 和 U_{jk} 的协方差为: $\Delta_{jk,ji} = (\boldsymbol{\mu}_j - \boldsymbol{\mu}_i)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_j - \boldsymbol{\mu}_k)$ 。
- 为了确定常数 c_j , 考虑积分

$$\Pr(j|j, R) = \int_{c_j - c_m}^{\infty} \cdots \int_{c_j - c_1}^{\infty} f_j du_{j1} \cdots du_{j,j-1} du_{j,j+1} \cdots du_{jm},$$

其中 f_j 是 $\{U_{ji}, i = 1, \cdots, K, i \neq j\}$ 的联合密度函数。

定理11.7.2

若 $\pi_i = N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$, $i = 1, \dots, K$, 且错判损失是相等的, 则判别区域 R_1, \dots, R_K 使得由式(8.3)定义的最大条件期望损失达到最小, 其中 $u_{jk}(\mathbf{x})$ 由式(8.1)所定义, 其中 $j \neq k$, 常数 c_j 的选取是由上述积分相等来确定。

- ① 上面的定理的结果是在参数已知的情况下得到的, 在实际问题中这些参数是未知。
- ② 如果这些参数是未知的, 且每个总体有一个训练样本可用, 则可以把这些未知参数通过训练样本估计出来, 代入到 $u_{ij}(\mathbf{x})$ 的定义中, 进而进行判别分析。

多个多元正态分布的判别

- 设有样本 $\mathbf{x}_1^{(i)}, \dots, \mathbf{x}_{n_i}^{(i)} \sim N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}), i = 1, \dots, K$
- 均值向量 $\boldsymbol{\mu}_i$ 的估计: $\hat{\boldsymbol{\mu}}_i = \bar{\mathbf{x}}^{(i)} = \frac{1}{n_i} \sum_{k=1}^{n_i} \mathbf{x}_k^{(i)}$
- $\boldsymbol{\Sigma}$ 的估计为 \mathbf{S} , 定义为:

$$\left(\sum_{i=1}^K n_i - K \right) \mathbf{S} = \sum_{i=1}^K \sum_{k=1}^{n_i} (\mathbf{x}_k^{(i)} - \bar{\mathbf{x}}^{(i)}) (\mathbf{x}_k^{(i)} - \bar{\mathbf{x}}^{(i)})'$$

- 则可定义下面的判别函数:

$$w_{ij}(\mathbf{x}) = \left[\mathbf{x} - \frac{1}{2}(\bar{\mathbf{x}}^{(i)} + \bar{\mathbf{x}}^{(j)}) \right]' \mathbf{S}^{-1} (\bar{\mathbf{x}}^{(i)} - \bar{\mathbf{x}}^{(j)}), \quad i \neq j.$$

- Rao (1948)考虑了贵族(π_1), 工匠(π_2)和贫民(π_3)三个阶层。每个阶层的测量值为 x_1 (身高), x_2 (坐高), x_3 (鼻深)和 x_4 (鼻长)。这些变量的均值由下表给出。

Table: 三个印度阶层关于身高、坐高、鼻深和鼻长的均值。

测量	贵族(π_1)	工匠(π_2)	贫民(π_3)
x_1 (身高)	164.51	160.53	158.17
x_2 (坐高)	86.43	81.47	81.16
x_3 (鼻深)	25.49	23.84	21.44
x_4 (鼻长)	51.24	48.62	46.72

多个多元正态分布的判别

- 所有总体的相关系数矩阵为：

$$\begin{pmatrix} 1.0000 & 0.5849 & 0.1774 & 0.1974 \\ 0.5849 & 1.0000 & 0.2094 & 0.2170 \\ 0.1774 & 0.2094 & 1.0000 & 0.2910 \\ 0.1974 & 0.2170 & 0.2910 & 1.0000 \end{pmatrix}$$

- 标准差分别为： $\sigma_1 = 5.74$, $\sigma_2 = 3.20$, $\sigma_3 = 1.75$, $\sigma_4 = 3.50$ 。
- **问题：**假定每个总体是正态的，将由四个变量 x_1 , x_2 , x_3 , x_4 形成的空间划分成三个判别区域。

多个多元正态分布的判别

假定错判损失是相等的，要找到：

- ① 一个区域族，使得一个新的观测值来自每个总体是等可能的，即 $q_1 = q_2 = q_3 = \frac{1}{3}$
- ② 该区域族使得错判的最大概率达到最小，即minimax判别。

求解：

- 首先计算： $\Sigma^{-1}(\mu_1 - \mu_2)$ 和 $\Sigma^{-1}(\mu_1 - \mu_3)$ 的系数，即

$$\Sigma^{-1}(\mu_2 - \mu_3) = \Sigma^{-1}(\mu_1 - \mu_3) - \Sigma^{-1}(\mu_1 - \mu_2).$$

- 然后计算：

$$\frac{1}{2}(\mu_i + \mu_j)' \Sigma^{-1}(\mu_i - \mu_j).$$

多个多元正态分布的判别

- 可获得判别函数：

$$u_{12}(\mathbf{x}) = -0.0708x_1 + 0.4990x_2 + 0.3373x_3 + 0.0887x_4 - 43.13,$$

$$u_{13}(\mathbf{x}) = 0.0003x_1 + 0.3550x_2 + 1.1063x_3 + 0.1375x_4 - 62.49,$$

$$u_{23}(\mathbf{x}) = 0.0711x_1 - 0.1440x_2 + 0.7690x_3 + 0.0488x_4 - 19.36.$$

- 其他三个函数为： $u_{21}(\mathbf{x}) = -u_{12}(\mathbf{x})$, $u_{31}(\mathbf{x}) = -u_{13}(\mathbf{x})$ 和 $u_{32}(\mathbf{x}) = -u_{23}(\mathbf{x})$ 。
- 若它们有一个先验概率，并且相等，则最好的判别区域为：

$$R_1 : u_{12}(\mathbf{x}) \geq 0, \quad u_{13}(\mathbf{x}) \geq 0;$$

$$R_2 : u_{21}(\mathbf{x}) \geq 0, \quad u_{23}(\mathbf{x}) \geq 0;$$

$$R_3 : u_{31}(\mathbf{x}) \geq 0, \quad u_{32}(\mathbf{x}) \geq 0.$$

多个多元正态分布的判别

- 例如，若测量值 \mathbf{x} 的个体满足： $u_{12}(\mathbf{x}) \geq 0$ 且 $u_{13}(\mathbf{x}) \geq 0$ ，则把它归类为贵族。
- 当一个个体来自总体 π_k 时，为了找到错判概率，需要知道 u 中任意两个均值，方差和协方差，见下表。

Table: 三个印度阶层的判别函数

总体	u	均值	标准差	相关系数
π_1	u_{12}	1.491	1.727	0.8658
	u_{13}	3.487	2.641	
π_2	u_{21}	1.491	1.727	-0.3894
	u_{23}	1.031	1.436	
π_3	u_{31}	3.487	2.641	0.7983
	u_{32}	1.031	1.436	

多个多元正态分布的判别

- 错判概率可以通过二元正态分布表得到，这些错判概率分别为：
 - ① 样本来于 π_1 时0.21;
 - ② 样本来于 π_2 时0.42;
 - ③ 样本来于 π_3 时0.25。
- 例如，若个体来自贵族，却将其错误归类到工匠或贫民的概率为0.21。
- Minimax方法是通过式(8.3)选择 c_1, c_2, c_3 使得错判概率相等，这些判别区域为：

$$R'_1 : u_{12}(\mathbf{x}) \geq 0.54, \quad u_{13}(\mathbf{x}) \geq 0.29;$$

$$R'_2 : u_{21}(\mathbf{x}) \geq -0.54, \quad u_{23}(\mathbf{x}) \geq -0.25;$$

$$R'_3 : u_{31}(\mathbf{x}) \geq -0.29, \quad u_{32}(\mathbf{x}) \geq 0.25.$$

- 错判的共同概率为0.30，因此，错判的最大概率从0.42减少到0.30。

- 在R语言中，QDA和LDA方法分别可以使用程序包MASS中的函数`qda()`和`lda()`来实现，它们有三种调用格式。

第一种调用格式：

```
qda(formula, data, ..., subset, na.action)
```

```
lda(formula, data, ..., subset, na.action)
```

其中`formula`表示判别公式，形式为：`groups~x1+x2+...`；

`data`表示数据集；`subset`表示子样本；`na.action`表示处理缺失值的方法，默认为如果样本中有缺失值，则函数无法运行；如果设置为`na.omit`，则表示自动删除样本中的缺失值，然后进行计算。

第二种调用格式:

```
qda(x, grouping, prior = proportions, method, CV = FALSE, nu, ...)  
lda(x, grouping, prior = proportions, tol = 1.0e-4,  
    method, CV = FALSE, nu, ...)
```

其中x表示数据框数据；grouping表示每个观测样本的所属类别；prior表示各类别的先验概率，默认取训练集中各样本的比例；tol表示筛选变量，默认取0.0001；如果CV=TRUE，表示返回结果是采用了去一交叉验证(LOOCV)方法得到的分类结果或后验概率；其他参数见在线帮助。

第三种调用格式:

```
qda(x, grouping, ..., subset, na.action)  
lda(x, grouping, ..., subset, na.action)
```

其中x表示数据矩阵，其他参数相同于上面两种调用格式。

■ 例：三文鱼数据

■ 目的：试用LDA和QDA 方法对三文鱼数据集进行判别分析，并对结果进行可视化分析。

■ 数据集来自程序包rrcov，数据的值放大了100 倍，单位是英寸

■ X_1 表示第一年淡水(Freshwater)生长年轮直径大小

■ X_2 表示第一年深海(Marine)生长年轮直径大小

■ “1” 表示雌性，“2” 表示雄性

表 11.4 三文鱼数据

Alaskan						Canadian					
Gender	X_1	X_2	Gender	X_1	X_2	Gender	X_1	X_2	Gender	X_1	X_2
2	108	368	2	85	444	1	129	420	2	134	383
1	131	355	1	109	397	1	148	371	1	117	355
1	105	469	2	106	442	1	179	407	2	126	345
2	86	506	1	82	431	2	152	381	1	118	379
1	99	402	2	118	381	2	166	377	2	120	369
2	87	423	1	105	388	2	124	389	1	153	403
1	94	440	1	121	403	1	156	419	2	150	354
2	117	489	1	85	451	2	131	345	1	154	390
2	79	432	1	83	453	1	140	362	1	155	349
1	99	403	1	53	427	2	144	345	2	109	325
1	114	428	1	95	411	2	149	393	2	117	344
2	123	372	1	76	442	1	108	330	1	128	400
1	123	372	1	95	426	1	135	355	1	144	403
2	109	420	2	87	402	2	170	386	2	163	370
2	112	394	1	70	397	1	152	301	2	145	355
1	104	407	2	84	511	1	153	397	1	133	375
2	111	422	2	91	469	1	152	301	1	128	383
2	126	423	1	74	451	2	136	438	2	123	349
2	105	434	2	101	474	2	122	306	1	144	373
1	119	474	1	80	398	1	148	383	2	140	388
1	114	396	1	95	433	2	90	385	2	150	339
2	100	470	2	92	404	1	145	337	2	124	341
2	84	399	1	99	481	1	123	364	1	125	346
2	102	429	2	94	491	2	145	376	1	153	352
2	101	469	1	87	480	2	115	354	1	108	339

案例及R语言计算

```
library(rrcov); library(MASS); data(salmon)
lda.fit = lda(Origin~Freshwater+Marine,data=salmon,prior=c(1,1)/2)
#### 函数lda()的输出结果:
Call:
lda(Origin ~ Freshwater+Marine, data=salmon, prior=c(1, 1)/2)
Prior probabilities of groups:
  Alaskan Canadian
      0.5      0.5
Group means:
      Freshwater Marine
Alaskan      98.38 429.66
Canadian     137.46 366.62
Coefficients of linear discriminants:
      LD1
Freshwater  0.04458572
Marine      -0.01803856
#### 利用函数predict()进行预测:
lda.predict = predict(lda.fit, newdata=salmon)
table(salmon[,4],lda.predict$class)
#### 输出混淆矩阵:
      Alaskan Canadian
Alaskan      44      6
Canadian      1     49
```

- 函数`plot()`绘制线性判别图像
- 使用程序包`klaR`中的函数`partimat()`对判别分析效果进行可视化
- 不仅给出了分类的错判率，还给出了分类情况和决策边界

函数`plot()`绘制线性判别图像：

```
plot(lda.fit)
```

```
library(klaR)
```

```
partimat(Origin ~ Marine+Freshwater, data=salmon,  
         method="lda",  
         main = "Linear Discriminant Analysis")
```

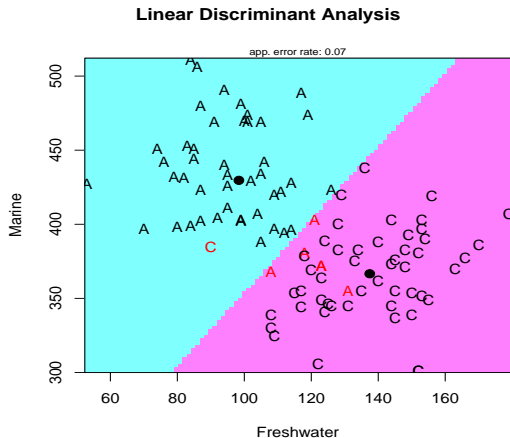
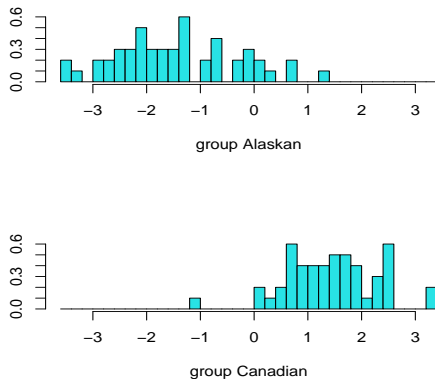



Figure: 左：线性判别图像；右：线性判别函数的决策边界。

- (1) 函数`lda()`输出了先验概率，类平均值，以及线性判别函数中的淡水和深海的组合系数，用来形成LDA方法的决策准则，线性判别函数为：

$$0.044 \times \text{Freshwater} - 0.018 \times \text{Marine};$$

- (2) 函数`predict()`返回一个三元列表，第一个元素`$class`给出了LDA方法关于三文鱼的预测结果；第二个元素`$posterior`是一个 100×2 的矩阵，其中第 k 列是观测值属于第 k 类的后验概率，且 $k = 1, 2$ ；第三个元素`$x`给出了每个样本的线性判别；
- (3) 由预测结果可知，当线性判别的值为正时，把该样本判别为加拿大三文鱼，否则判别为阿拉斯加三文鱼；

- (4) 从混淆矩阵可知，LDA 方法的错判率为 $(6 + 1)/100 = 0.07$ ；
- (5) 线性判别的决策边界是一条线性函数，LDA 方法只能用于协方差矩阵相等的情形，而经计算，阿拉斯加和加拿大的样本协方差矩阵分别为：

$$\mathbf{S}_A = \begin{pmatrix} 260.6078 & -188.0927 \\ -188.0927 & 1399.0861 \end{pmatrix},$$

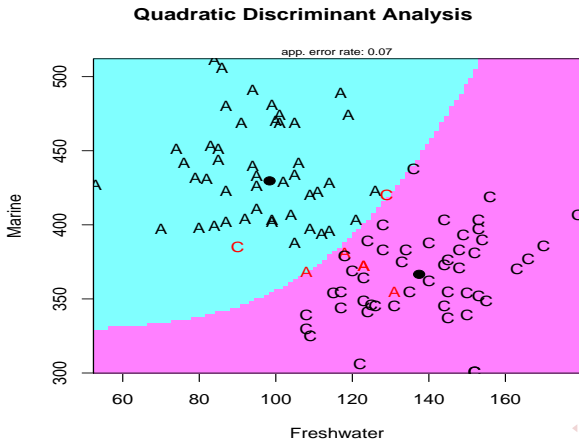
$$\mathbf{S}_C = \begin{pmatrix} 326.0902 & 133.5049 \\ 133.5049 & 893.2608 \end{pmatrix}.$$

案例及R语言计算

```
qda.fit = qda(Origin~Marine+Freshwater, data=salmon)
#### 函数qda()的输出结果:
Call:
qda(Origin ~ Marine + Freshwater, data = salmon)
Prior probabilities of groups:
  Alaskan Canadian
      0.5      0.5
Group means:
      Marine Freshwater
Alaskan  429.66      98.38
Canadian 366.62     137.46
#### 利用函数predict()进行预测:
qda.predict = predict(object=qda.fit, newdata=salmon)
table(salmon[,4], qda.predict$class)
#### 输出混淆矩阵:
      Alaskan Canadian
Alaskan    45       5
Canadian    2      48
```

案例及R语言计算

```
partimat(Origin ~ Marine+Freshwater, data=salmon,  
method="qda", main= "Quadratic Discriminant Analysis")
```



● 例：Fisher Iris数据集

- ① R语言中自带的Iris数据有四个属性，萼片长度、萼片宽度、花瓣长度和花瓣宽度
- ② 数据共有150个样本，分为三类：前50个样本是属于第一类—Setosa，中间的50个样本属于第二类—Versicolor，最后50个样本属于第三类—Virginica

● 现将该数据集随机分成两部分：

- ① 75个样本作为训练集训练判别函数
- ② 75个样本作为测试集来测试判别精度

● 目的：试用LDA和QDA方法建立判别分类器，并对结果进行可视化分析。

案例及R语言计算

```
library(MASS); data(iris); set.seed(2021)
train = sample(1:150, 75)
table(iris$Species[train])          ## 展示训练样本
      setosa versicolor  virginica
      24         21         30

lda.iris=lda(Species ~ ., iris, prior=c(24,21,30)/75, subset=train)
plot(lda.iris, abbrev=T, col=as.numeric(iris$Species[train]))

#### 输出结果:

Call:
lda(Species ~ ., data=iris, prior=c(24, 21, 30)/75, subset=train)

Prior probabilities of groups:
      setosa versicolor  virginica
      0.32         0.28         0.40
```

案例及R语言计算

Group means:

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
setosa	5.079167	3.541667	1.462500	0.2666667
versicolor	5.961905	2.704762	4.228571	1.2857143
virginica	6.570000	2.930000	5.553333	2.0100000

Coefficients of linear discriminants:

	LD1	LD2
Sepal.Length	1.276769	0.06739943
Sepal.Width	1.204560	-2.29478268
Petal.Length	-2.737373	0.51175177
Petal.Width	-2.104923	-2.18494902

Proportion of trace:

	LD1	LD2
	0.9901	0.0099

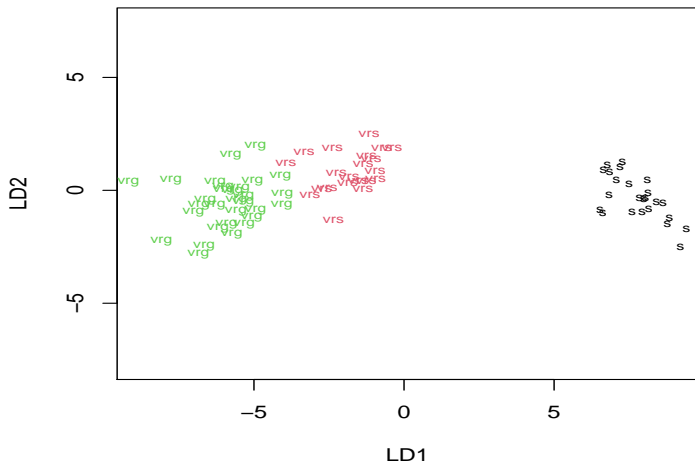


Figure: 基于Fisher Iris训练数据的线性判别图像。

- 函数lda()输出的结果包括每个类别的先验概率、每个类别数据的组平均值、第一和第二线性判别函数的系数(因为有三个类别,所以需要两个判别函数)、以及第一和第二线性判别函数解释方差的比例
- 第一线性判别函数能够解释总体数据99.01%的方差
- 两个线性判别函数为:

$$\begin{aligned}W_1 &= 1.28 \times \text{Sepal.Length} + 1.20 \times \text{Sepal.Width} \\&\quad - 2.74 \times \text{Petal.Length} - 2.10 \times \text{Petal.Width}, \\W_2 &= 0.07 \times \text{Sepal.Length} - 2.29 \times \text{Sepal.Width} \\&\quad + 0.51 \times \text{Petal.Length} - 2.18 \times \text{Petal.Width}.\end{aligned}$$

案例及R语言计算

```
piris.lda = predict(lda.iris, iris[-train, ])$class
cl.test = iris$Species[-train]
table(cl.test, piris.lda)
#### 输出结果:
```

```
      piris.lda
cl.test setosa versicolor virginica
setosa      26           0           0
versicolor   0          27           2
virginica     0           0          20
```

```
#### 利用函数classError() 计算分类错误率:
```

```
library(mclust); classError(piris.lda,cl.test)
$misclassified
[1] 41 47
$errorRate
[1] 0.02666667
```

- 采用LOOCV方法在整个数据集上对LDA方法的判别精度进行评价，只需要在函数lda() 中设置参数CV=TRUE

```
cv.lda=lda(Species~.,iris,prior=c(24,21,30)/75,CV=TRUE)
```

```
table(iris$Species, cv.lda$class)
```

输出混淆矩阵:

	setosa	versicolor	virginica
setosa	50	0	0
versicolor	0	48	2
virginica	0	1	49

```
> mean(iris$Species!=cv.lda$class)
```

```
[1] 0.02
```

```
set.seed(2021); tr.id = sample(1:50, 25)
train = rbind(iris3[tr.id,,1], iris3[tr.id,,2], iris3[tr.id,,3])
test = rbind(iris3[-tr.id,,1], iris3[-tr.id,,2], iris3[-tr.id,,3])
cl = factor(c(rep("s",25), rep("c",25), rep("v",25)))
qda.iris = qda(train, cl, method="mle")
piris.qda = predict(qda.iris,test)$class
classError(piris.qda,cl)

[1] 45 70

$errorRate

[1] 0.02666667
```

LOOCV方法的QDA评价:

```
cv.qda = qda(Species~., data=iris, CV=TRUE)
```

```
cv.qda$class
```

```
table(iris$Species, cv.qda$class)
```

输出混淆矩阵:

	setosa	versicolor	virginica
setosa	50	0	0
versicolor	0	47	3
virginica	0	1	49

```
> mean(iris$Species!=cv.qda$class)
```

```
[1] 0.02666667
```

- 使用程序包klaR中的函数partimat()对判别分析效果进行可视化

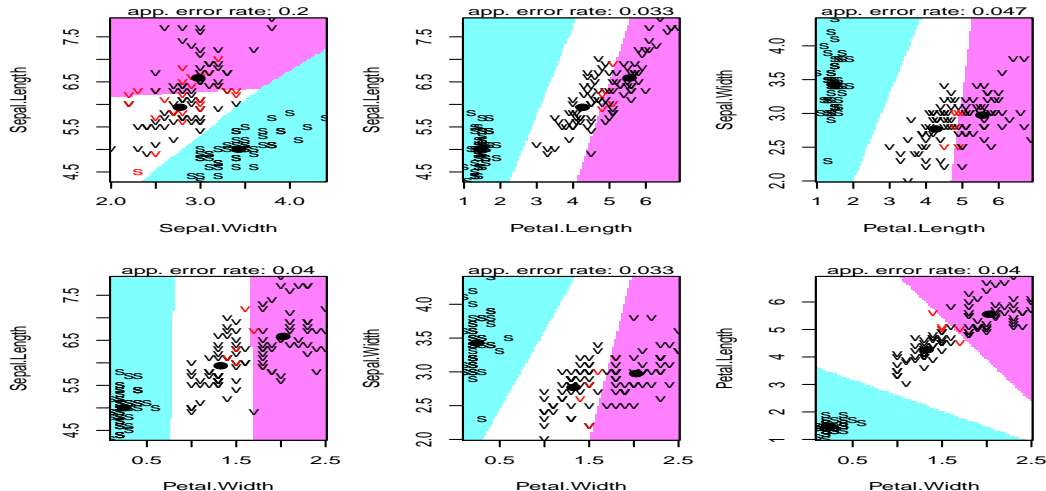
```
library(MASS)

library(klaR)

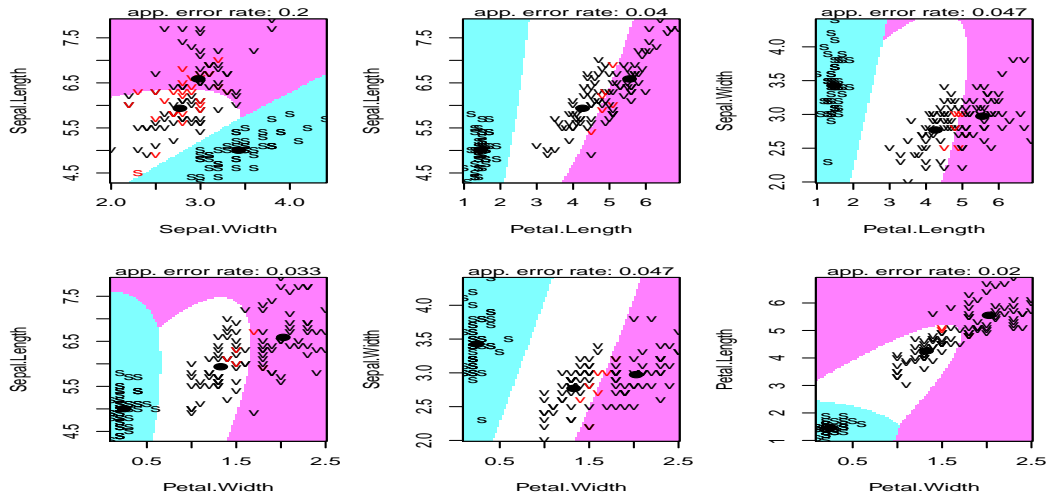
partimat(Species ~ ., data = iris, method = "lda",
         main = "Linear Discriminant Analysis")

partimat(Species ~ ., data = iris, method = "qda",
         main = "Quadratic Discriminant Analysis")
```

Linear Discriminant Analysis



Quadratic Discriminant Analysis





谢谢，请多提宝贵意见！