Influence function of the error rate of generalized k-means

Joint work with G. Haesbroeck

Ch. Ruwet

UNIVERSITY OF LIÈGE

30 March 2009

Introduction

The k-means clustering method

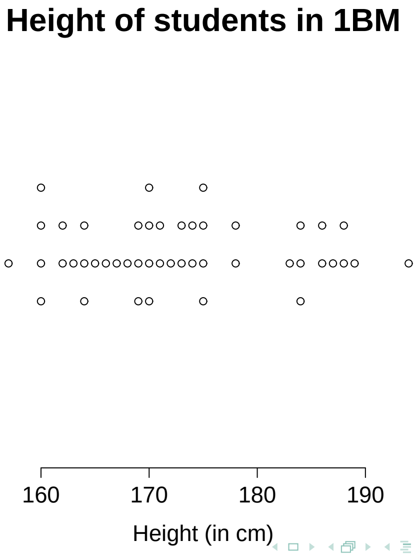
- Aim of clustering : Group similar observations in k clusters $C_1, \ldots, C_k$.
- The k-means algorithm constructs clusters in order to minimize the within cluster sum of squared distances.
- Let us focus on $k = 2$ groups.

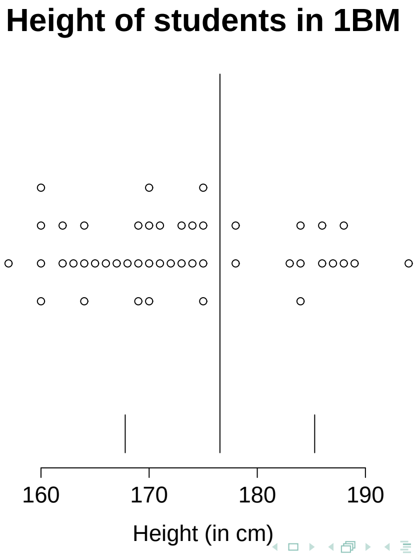
Classification based on clustering

In case of natural groups in the data, clustering may be used to find these groups via a classification rule :

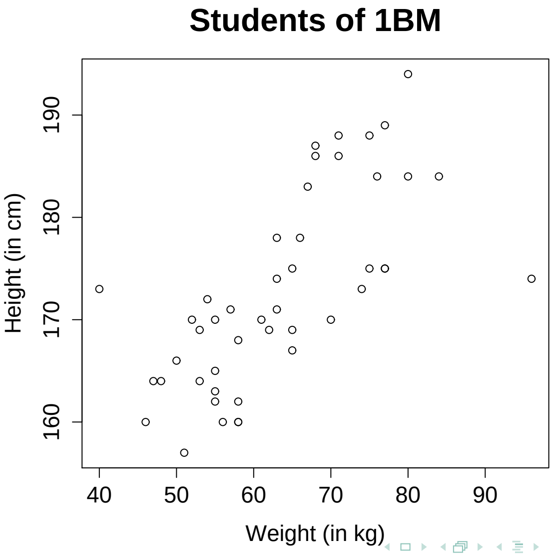
$$x \in C_j \Leftrightarrow \|x - \bar{x}_j\| = \min_{1 \leq i \leq 2} \|x - \bar{x}_i\|$$

Example in 1D

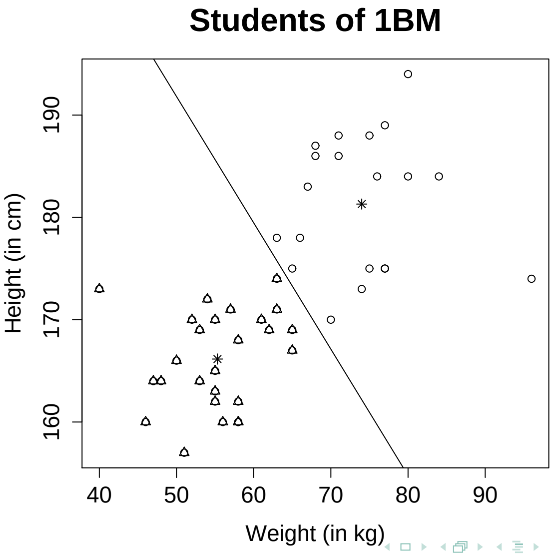



Height of students in 1BM
160
170
180
190
Height (in cm)

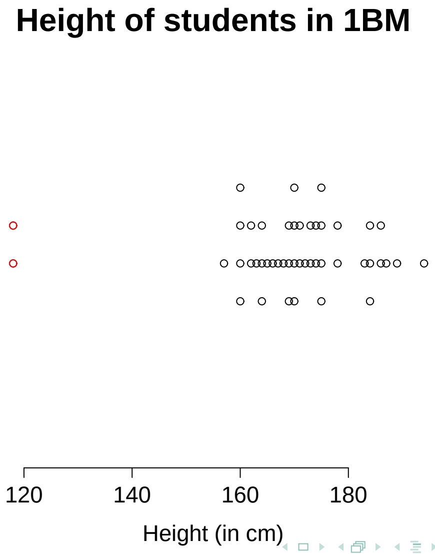
Example in 2D



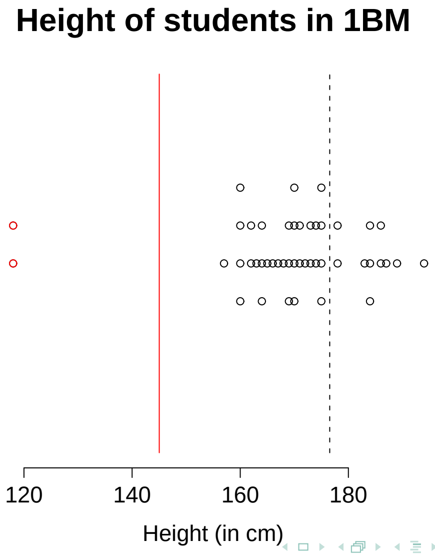
Example in 2D



Contaminated example in 1D



Contaminated example in 1D



The generalized 2-means clustering method

- The clusters' centers (T_1, T_2) are solution of

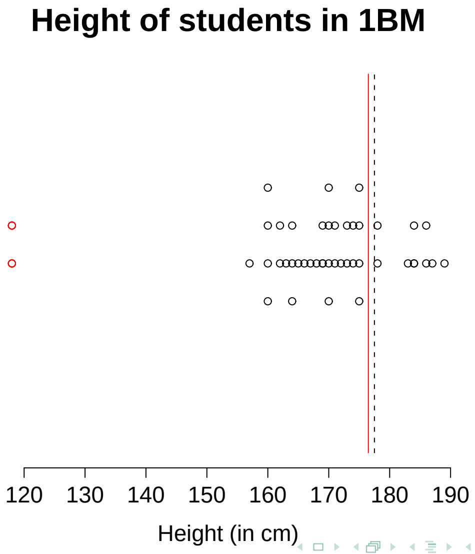
$$\min_{\{t_1,t_2\}\subset\mathbb{R}^2} \sum_{i=1}^{n} \Omega\left(\inf_{1\leq j\leq 2} \|x_i - t_j\|\right)$$

for a suitable nondecreasing penalty function Ω.

- Classical penalty functions :

$$\Omega(x) = x^2 \rightarrow \text{ 2-means method}$$
$$\Omega(x) = x \rightarrow \text{ 2-medoids method}$$

Contaminated example with the 2-medoids method



The generalized 2-means clustering method

- The classification rule :

$$x \in C_j \Leftrightarrow \Omega(\|x - T_j\|) = \min_{1 \leq i \leq 2} \Omega(\|x - T_i\|)$$

- In one dimension, the estimated clusters are simply:

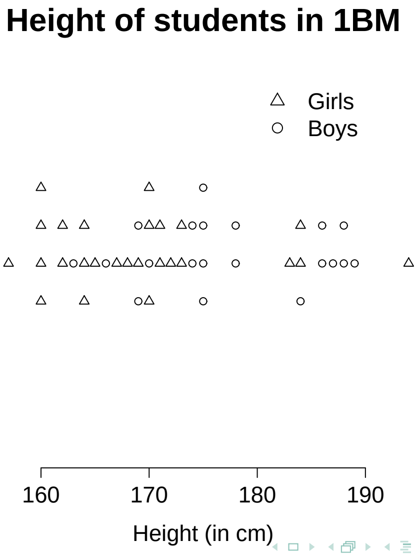
$$C_1 =]-\infty, C[$$

$$C_2 =]C, +\infty[$$

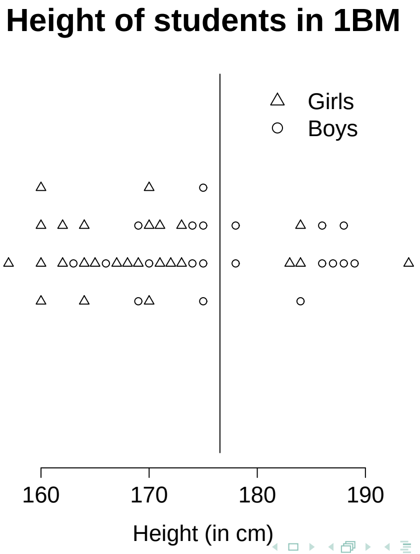
where $C = \dfrac{T_1 + T_2}{2}$ is the cut-off point.

Error rate

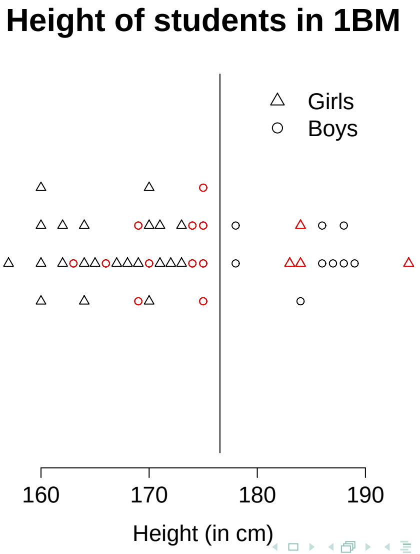
Example in 1D



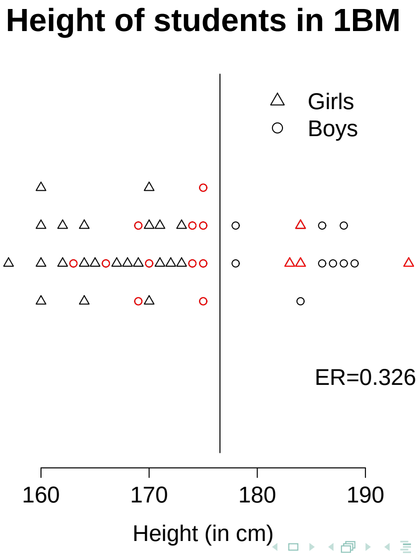
Example in 1D with the 2-means



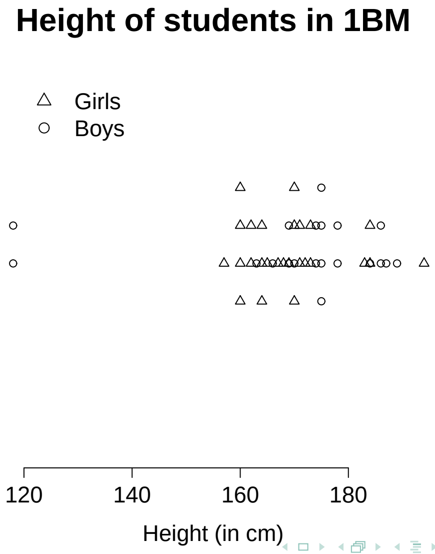
Example in 1D with the 2-means



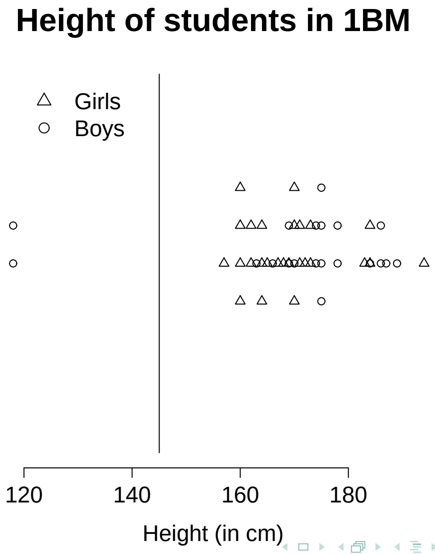
Example in 1D with the 2-means



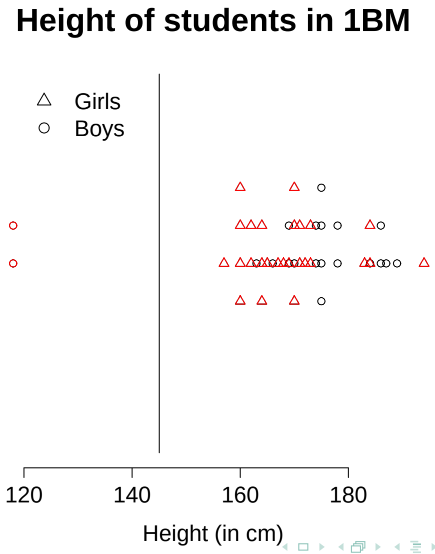
Contaminated example in 1D



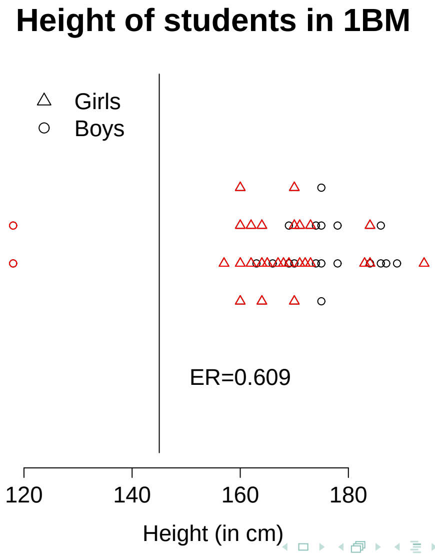
Contaminated example in 1D with the 2-means



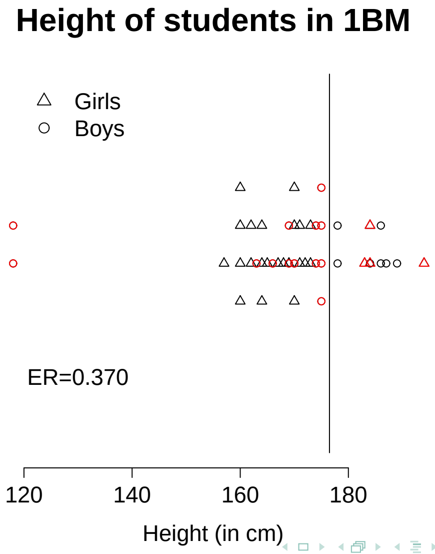
Contaminated example in 1D with the 2-means



Contaminated example in 1D with the 2-means



Contaminated example in 1D with the 2-medoids



Classification setting

Influence function of the error rate of generalized k-means

Ch. Ruwet

Introduction

Error rate

Influence function of the error rate

Bias of the error rate

Simulation study

Future research

Suppose

X arises from 2 groups G_1 and G_2 with $\pi_i(F) = \mathbb{P}_F[X \in G_i]$

then

F is a mixture of two distributions

$$F = \pi_1(F)F_1 + \pi_2(F)F_2$$

with densities f_1 and f_2.

Additional assumption : one dimension !

The generalized 2-means as statistical functionals

- The clusters' centers $(T_1(F), T_2(F))$ are solution of

$$\min_{\{t_1,t_2\}\subset\mathbb{R}^2} \int \Omega\left(\inf_{1\leq j\leq 2} \|x - t_j\|\right) dF(x)$$

for a suitable nondecreasing penalty function Ω.

- The classification rule is

$$R_F(x) = C_j(F) \Leftrightarrow \Omega(\|x - T_j(F)\|) = \min_{1\leq i\leq 2} \Omega(\|x - T_i(F)\|)$$

Optimality in classification

- A classification rule is optimal if the corresponding error rate is minimal
- The optimal classification rule is the Bayes rule :

$$x \in C_1 \Leftrightarrow \pi_1(F)f_1(x) > \pi_2(F)f_2(x)$$

(Anderson, 1958)

- The 2-means procedure is optimal under the model

$$F_N = 0.5\,N(\mu_1, \sigma^2) + 0.5\,N(\mu_2, \sigma^2) \text{ with } \mu_1 < \mu_2$$

(Qiu and Tamhane, 2007)

Theoretical vs empirical error rate

- Theoretical error rate :
  - Training sample according to F : estimation of the rule
  - Test sample according to F_m : evaluation of the rule
  - In ideal circumstances : $F = F_m$

$$\mathrm{ER}(F, F_m) = \sum_{j=1}^{2} \pi_j(F_m) \mathbb{P}_{F_m} \left[R_F(X) \neq C_j(F) \middle| G_j \right]$$

- Empirical error rate :
  - Training sample according to F : estimation and evaluation of the rule

$$\mathrm{ER}(F, F) = \sum_{j=1}^{2} \pi_j(F) \mathbb{P}_{F} \left[R_F(X) \neq C_j(F) \middle| G_j \right]$$

Theoretical vs empirical error rate

- Theoretical error rate :
  - Training sample according to F : estimation of the rule
  - Test sample according to F_m : evaluation of the rule
  - In ideal circumstances : $F = F_m$

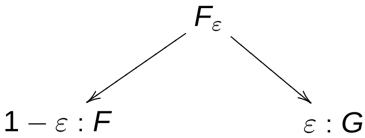
$$\mathrm{ER}(F, F_m) = \sum_{j=1}^{2} \pi_j(F_m) \mathbb{P}_{F_m} \left[R_F(X) \neq C_j(F) \middle| G_j \right]$$

- Empirical error rate :
  - Training sample according to F : estimation and evaluation of the rule

$$\mathrm{ER}(F, F) = \sum_{j=1}^{2} \pi_j(F) \mathbb{P}_{F} \left[R_F(X) \neq C_j(F) \middle| G_j \right]$$

Contaminated distribution

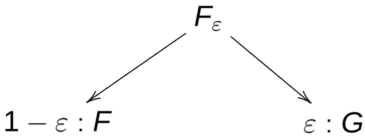
A contaminated distribution is defined by



where G is a arbitrary distribution function.

Contaminated distribution

A contaminated distribution is defined by

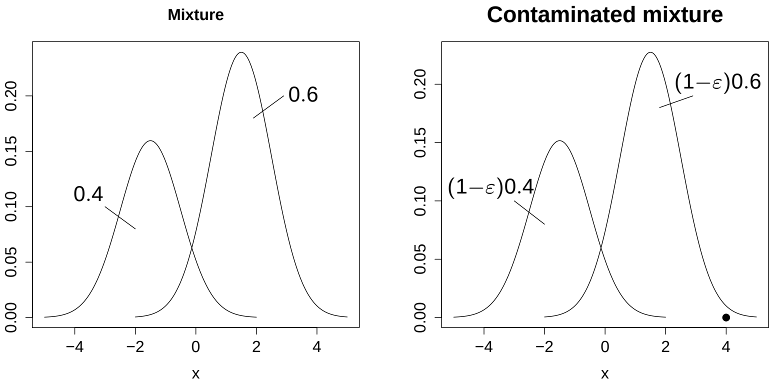


where G is a arbitrary distribution function.

To see the influence of one singular point x, $G = \Delta_x$ leading to

$$F_\varepsilon = (1-\varepsilon)F + \varepsilon\Delta_x$$

Contaminated mixture



Theoretical vs empirical error rate under contamination

Now, the training sample is distributed as F_ε which is a contaminated mixture.

- Theoretical error rate :

$$\mathrm{ER}(F_\varepsilon, F_m) = \sum_{j=1}^{2} \pi_j(F_m)\mathbb{P}_{F_m}\left[R_{F_\varepsilon}(X) \neq C_j(F_\varepsilon) \middle| G_j\right]$$

- Empirical error rate :

$$\mathrm{ER}(F_\varepsilon, F_\varepsilon) = \sum_{j=1}^{2} \pi_j(F_\varepsilon)\mathbb{P}_{F_\varepsilon}\left[R_{F_\varepsilon}(X) \neq C_j(F_\varepsilon) \middle| G_j\right]$$

Theoretical vs empirical error rate under contamination

Now, the training sample is distributed as F_ε which is a contaminated mixture.

- Theoretical error rate :

$$\mathrm{ER}(F_\varepsilon, F_m) = \sum_{j=1}^{2} \pi_j(F_m) \mathbb{P}_{F_m} \left[R_{F_\varepsilon}(X) \neq C_j(F_\varepsilon) \middle| G_j \right]$$

- Empirical error rate :

$$\mathrm{ER}(F_\varepsilon, F_\varepsilon) = \sum_{j=1}^{2} \pi_j(F_\varepsilon) \mathbb{P}_{F_\varepsilon} \left[R_{F_\varepsilon}(X) \neq C_j(F_\varepsilon) \middle| G_j \right]$$

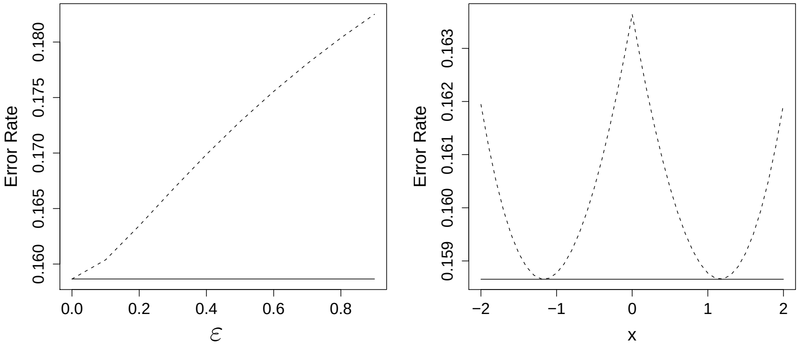
Theoretical vs empirical error rate under contamination

Graphically :

- $F_m = F_N \equiv 0.5\, N(-1,1) + 0.5\, N(1,1)$ an optimal model
- $C(F_N) = \frac{-1+1}{2} = 0$ (Qiu and Tamhane, 2007)
- $F_\varepsilon = (1-\varepsilon)F_m + \varepsilon\Delta_x$
- ε varying and $x = -0.5$
- $x \in G_1$ varying and $\varepsilon = 0.1$

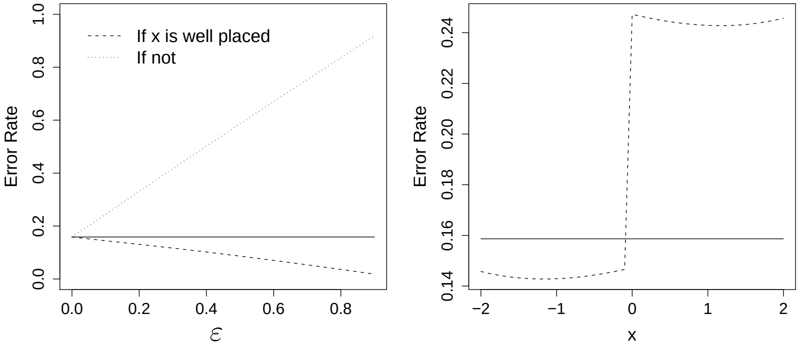
Theoretical vs empirical error rate under contamination

- Theoretical error rate (with the 2-means) :



Theoretical vs empirical error rate under contamination

- Empirical error rate (with the 2-means) :



Influence function of the error rate

Influence functions

Hampel et al (1986) : For any statistical functional T and any distribution F,

- $\mathrm{IF}(x; \mathrm{T}, F) = \lim_{\varepsilon \to 0} \frac{\mathrm{T}(F_\varepsilon) - \mathrm{T}(F)}{\varepsilon} = \frac{\partial}{\partial \varepsilon} \mathrm{T}(F_\varepsilon)\Big|_{\varepsilon=0}$ where $F_\varepsilon = (1-\varepsilon)F + \varepsilon\Delta_x$;
- $E_F[\mathrm{IF}(X; \mathrm{T}, F)] = 0$;
- $\mathrm{T}(F_\varepsilon) \approx \mathrm{T}(F) + \varepsilon \mathrm{IF}(x; \mathrm{T}, F)$ for ε small enough (First-order von Mises expansion of T at F).

Influence functions

Hampel et al (1986) : For any statistical functional T and any distribution F,

- $\mathrm{IF}(x;\mathrm{T},F) = \lim_{\varepsilon\to 0} \frac{\mathrm{T}(F_\varepsilon)-\mathrm{T}(F)}{\varepsilon} = \frac{\partial}{\partial\varepsilon}\mathrm{T}(F_\varepsilon)\Big|_{\varepsilon=0}$ where $F_\varepsilon = (1-\varepsilon)F + \varepsilon\Delta_x$;
- $E_F[\mathrm{IF}(X;\mathrm{T},F)] = 0$;
- $\mathrm{T}(F_\varepsilon) \approx \mathrm{T}(F) + \varepsilon\mathrm{IF}(x;\mathrm{T},F)$ for ε small enough (First-order von Mises expansion of T at F).

Theoretical vs empirical error rate

$$\mathrm{ER}(F_\varepsilon, F) \approx \mathrm{ER}(F, F) + \varepsilon \mathrm{IF}(x; \mathrm{ER}, F)$$

$$\mathrm{ER}(F_\varepsilon, F_\varepsilon) \approx \mathrm{ER}(F, F) + \varepsilon \mathrm{IF}(x; \mathrm{ER}, F)$$

- Theoretical error rate :

$$\mathrm{ER}(F_\varepsilon, F_N) \geq \mathrm{ER}(F_N, F_N) \Rightarrow \mathrm{IF}(x; \mathrm{ER}, F_N) \equiv 0$$

- Empirical error rate : The IF does not vanish!
  From now on, $\mathrm{ER}(F) = \mathrm{ER}(F, F)$.

Theoretical vs empirical error rate

$$\mathrm{ER}(F_\varepsilon, F) \approx \mathrm{ER}(F, F) + \varepsilon \mathrm{IF}(x; \mathrm{ER}, F)$$

$$\mathrm{ER}(F_\varepsilon, F_\varepsilon) \approx \mathrm{ER}(F, F) + \varepsilon \mathrm{IF}(x; \mathrm{ER}, F)$$

- Theoretical error rate :

$$\mathrm{ER}(F_\varepsilon, F_N) \geq \mathrm{ER}(F_N, F_N) \Rightarrow \mathrm{IF}(x; \mathrm{ER}, F_N) \equiv 0$$

- Empirical error rate : The IF does not vanish! From now on, $\mathrm{ER}(F) = \mathrm{ER}(F, F)$.

Theoretical vs empirical error rate

$$\mathrm{ER}(F_\varepsilon, F) \approx \mathrm{ER}(F, F) + \varepsilon \mathrm{IF}(x; \mathrm{ER}, F)$$

$$\mathrm{ER}(F_\varepsilon, F_\varepsilon) \approx \mathrm{ER}(F, F) + \varepsilon \mathrm{IF}(x; \mathrm{ER}, F)$$

- Theoretical error rate :

$$\mathrm{ER}(F_\varepsilon, F_N) \geq \mathrm{ER}(F_N, F_N) \Rightarrow \mathrm{IF}(x; \mathrm{ER}, F_N) \equiv 0$$

- Empirical error rate : The IF does not vanish!
  From now on, $\mathrm{ER}(F) = \mathrm{ER}(F, F)$.

$\mathrm{ER}(F_\varepsilon) = ?$

$$
\begin{aligned}
\mathrm{ER}(F) &= \sum_{j=1}^{2} \pi_j(F) \mathbb{P}_F \left[R_F(X) \neq C_j(F) \middle| G_j \right] \\
&= \pi_1(F)\{1 - F_1(C(F))\} + \pi_2(F) F_2(C(F))
\end{aligned}
$$

$\mathrm{ER}(F_\varepsilon) = ?$

$$\mathrm{ER}(F) = \sum_{j=1}^{2} \pi_j(F) \mathbb{P}_F \left[R_F(X) \neq C_j(F) \middle| G_j \right]$$
$$= \pi_1(F)\{1 - F_1(C(F))\} + \pi_2(F) F_2(C(F))$$

Under $F_\varepsilon = (1-\varepsilon)F + \varepsilon\Delta_x$, one has

$$\mathrm{ER}(F_\varepsilon) = \pi_1(F_\varepsilon)\left\{1 - F_{1,\varepsilon}(C(F_\varepsilon))\right\} + \pi_2(F_\varepsilon) F_{2,\varepsilon}(C(F_\varepsilon))$$

$\mathrm{ER}(F_\varepsilon) = ?$

$$\mathrm{ER}(F) = \sum_{j=1}^{2} \pi_j(F) \mathbb{P}_F \left[R_F(X) \neq C_j(F) \middle| G_j \right]$$
$$= \pi_1(F)\{1 - F_1(C(F))\} + \pi_2(F) F_2(C(F))$$

Under $F_\varepsilon = (1-\varepsilon)F + \varepsilon\Delta_x$, one has

$$\mathrm{ER}(F_\varepsilon) = \pi_1(F_\varepsilon)\left\{1 - F_{1,\varepsilon}(C(F_\varepsilon))\right\} + \pi_2(F_\varepsilon) F_{2,\varepsilon}(C(F_\varepsilon))$$

$\pi_i(F_\varepsilon) =?$ and $F_{i,\varepsilon} =?$

$$F_\varepsilon = (1-\varepsilon)F + \varepsilon\Delta_x$$

- $\pi_i(F_\varepsilon) = (1-\varepsilon)\pi_i(F) + \varepsilon \mathrm{I}\{x \in G_i\}$
- $F_{i,\varepsilon} = \left(1 - \frac{\varepsilon \mathrm{I}\{x \in G_i\}}{\pi_i(F_\varepsilon)}\right) F_i + \frac{\varepsilon \mathrm{I}\{x \in G_i\}}{\pi_i(F_\varepsilon)} \Delta_x$

$$\Rightarrow F_\varepsilon = \pi_1(F_\varepsilon)F_{1,\varepsilon} + \pi_2(F_\varepsilon)F_{2,\varepsilon}$$

$\pi_i(F_\varepsilon) = ?$ and $F_{i,\varepsilon} = ?$

$$F_\varepsilon = (1-\varepsilon)F + \varepsilon\Delta_x$$

- $\pi_i(F_\varepsilon) = (1-\varepsilon)\pi_i(F) + \varepsilon \mathrm{I}\{x \in G_i\}$
- $F_{i,\varepsilon} = \left(1 - \frac{\varepsilon \mathrm{I}\{x \in G_i\}}{\pi_i(F_\varepsilon)}\right) F_i + \frac{\varepsilon \mathrm{I}\{x \in G_i\}}{\pi_i(F_\varepsilon)} \Delta_x$

$$\Rightarrow F_\varepsilon = \pi_1(F_\varepsilon)F_{1,\varepsilon} + \pi_2(F_\varepsilon)F_{2,\varepsilon}$$

$\pi_i(F_\varepsilon) = ?$ and $F_{i,\varepsilon} = ?$

$$F_\varepsilon = (1-\varepsilon)F + \varepsilon\Delta_x$$

- $\pi_i(F_\varepsilon) = (1-\varepsilon)\pi_i(F) + \varepsilon \mathrm{I}\{x \in G_i\}$
- $F_{i,\varepsilon} = \left(1 - \frac{\varepsilon \mathrm{I}\{x \in G_i\}}{\pi_i(F_\varepsilon)}\right) F_i + \frac{\varepsilon \mathrm{I}\{x \in G_i\}}{\pi_i(F_\varepsilon)} \Delta_x$

$$\Rightarrow F_\varepsilon = \pi_1(F_\varepsilon)F_{1,\varepsilon} + \pi_2(F_\varepsilon)F_{2,\varepsilon}$$

Influence function of the empirical error rate

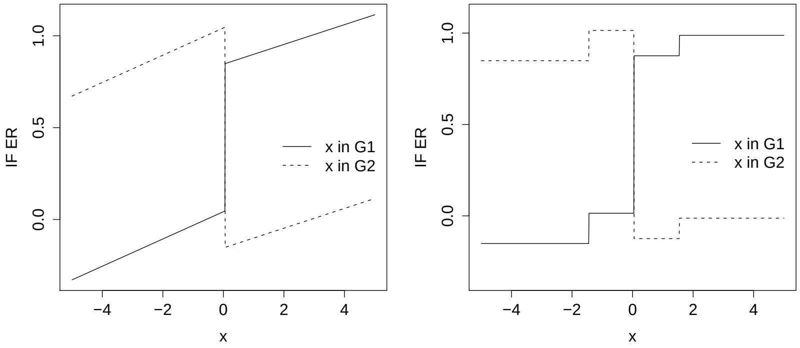
Proposition

For all $x \neq C(F)$,

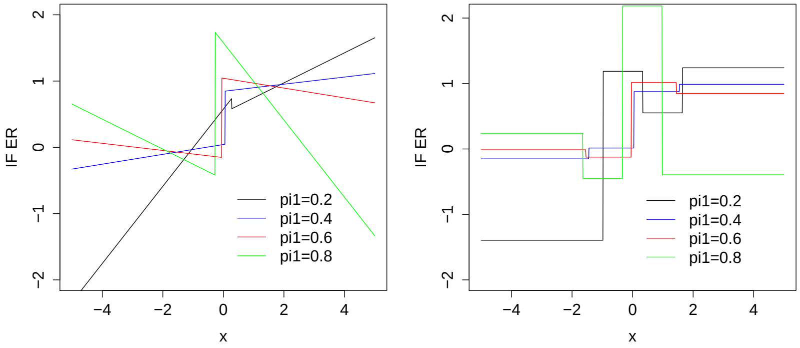
$$\begin{aligned}\mathrm{IF}(x;\mathrm{ER},F) = &-\mathrm{ER}(F)\\ &+ \mathrm{I}\{x \leq C(F)\}(1-2\,\mathrm{I}\{x \in G_1\}) + \mathrm{I}\{x \in G_1\}\\ &+ \frac{1}{2}(\mathrm{IF}(x;T_1,F)+\mathrm{IF}(x;T_2,F))\\ &\{\pi_2(F)f_2(C(F)) - \pi_1(F)f_1(C(F))\}.\end{aligned}$$

Expressions of $\mathrm{IF}(x;T_1,F)$ and $\mathrm{IF}(x;T_2,F)$ have been computed by García-Escudero and Gordaliza (1999).

Graphics of IF(x; ER, F)

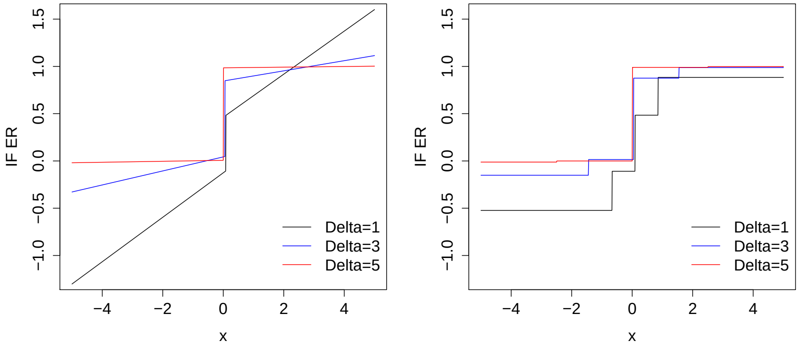


Graphics of IF(x; ER, F)

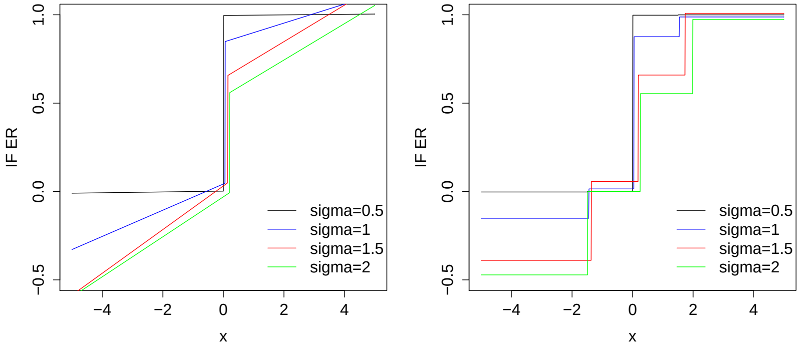


Graphics of IF(x; ER, F)

$$\Delta = \mu_2 - \mu_1$$



Graphics of IF(x; ER, F)



Bias of the error rate

Definition

- In practice, F is replace by F_n, the empirical cdf
- The bias is defined by $B_n(\text{ER}) = E_F[\text{ER}(F_n) - \text{ER}(F)]$
- Fernholz (2001) :

$$B_n(\text{ER}) = \frac{1}{2n} E_F[\text{IF2}(X; \text{ER}, F)] + o(n^{-1})$$

where

$$\text{IF2}(x; \text{ER}, F) = \frac{\partial^2}{\partial \varepsilon^2} \text{ER}((1-\varepsilon)F + \varepsilon \Delta_x)\Big|_{\varepsilon=0}.$$

This result is true if ER(.) is Hadamard or Fréchet differentiable.

Computation of the bias

Proposition

Under asymptotic normality and consistency of T_1 and T_2 (Pollard, 1981 and 1982) :

$$\begin{aligned} B_n(\mathrm{ER}) \approx \frac{1}{4n} & \{\pi_2(F) f_2(C(F)) - \pi_1(F) f_1(C(F))\} \\ & (E_F[\mathrm{IF2}(X; T_1, F)] + E_F[\mathrm{IF2}(X; T_2, F)]) \\ + \frac{1}{8n} & \{\pi_2(F) f_2'(C(F)) - \pi_1(F) f_1'(C(F))\} \\ & (\mathrm{ASV}(T_1) + \mathrm{ASV}(T_2) + 2\,\mathrm{ASC}(T_1, T_2)). \end{aligned}$$

Expressions of $\mathrm{IF2}(X; T_1, F)$ and $\mathrm{IF2}(X; T_2, F)$ have been computed.

Asymptotic difference

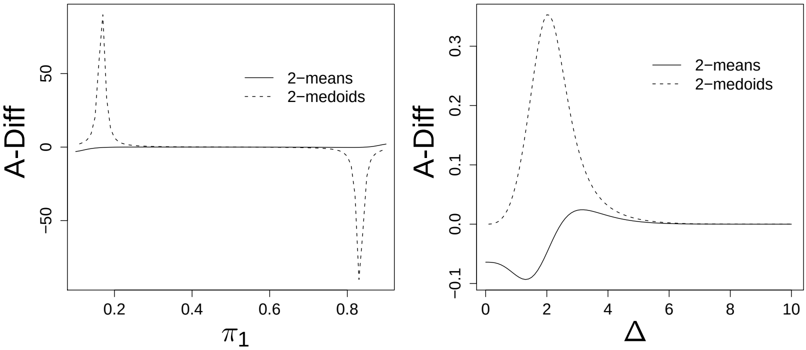
How much difference in error rate is to be expected by estimating a clustering rule from a finite sample ?

$$\text{A-Diff(ER)} = \lim_{n\to\infty} n\, B_n(\text{ER})$$

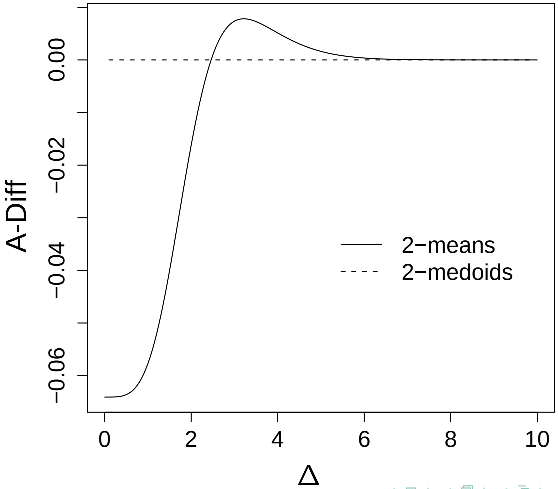
$\Rightarrow$ Graphical comparisons of the 2-means and 2-medoids methods :

- $F = \pi_1 N(-\Delta/2, 1) + (1 - \pi_1) N(\Delta/2, 1)$
  - π_1 varying and $\Delta = 3$
  - Δ varying and $\pi_1 = 0.4$
- $F_N = 0.5\, N(-\Delta/2, 1) + 0.5\, N(\Delta/2, 1)$

Graphics of A-Diff(ER) under *F*



Graphic of A-Diff(ER) under F_N



Simulation study

Simulation settings

- $F = 0.4\,N(-1.5, 1) + 0.6\,N(1.5, 1)$
- $n = 100$
- $F_\varepsilon = (1 - \varepsilon)F + \varepsilon\Delta_x$ with
  - $\varepsilon = 0.01$ and $|x| = 5$
  - $\varepsilon = 0.05$ and $|x| = 5$
  - $\varepsilon = 0.01$ and $|x| = 50$
  - $\varepsilon = 0.05$ and $|x| = 50$
- $N = 1000$

Simulation results

	2-means (0.071)		2-medoids (0.072)	
	in C_1	in C_2	in C_1	in C_2
from G_1				
from G_2				

Simulation results

	2-means (0.071)		2-medoids (0.072)	
	in C_1	in C_2	in C_1	in C_2
from G_1	0.068	0.083		
from G_2	0.078	0.072		

Simulation results

	2-means (0.071)		2-medoids (0.072)	
	in C_1	in C_2	in C_1	in C_2
from G_1	0.068	0.083	0.071	0.083
from G_2	0.078	0.072	0.081	0.073

Simulation results

	2-means (0.071)		2-medoids (0.072)	
	in C_1	in C_2	in C_1	in C_2
from G_1	0.068 0.064	0.083 0.139	0.071	0.083
from G_2	0.078 0.119	0.072 0.083	0.081	0.073

Simulation results

	2-means (0.071)		2-medoids (0.072)	
	in C_1	in C_2	in C_1	in C_2
from G_1	0.068 0.064	0.083 0.139	0.071 0.066	0.083 0.127
from G_2	0.078 0.119	0.072 0.083	0.081 0.116	0.073 0.072

Simulation results

2-means (0.071) 2-medoids (0.072)

	2-means in C_1	2-means in C_2	2-medoids in C_1	2-medoids in C_2
from G_1	0.068 0.064 0.39	0.083 0.139 0.61	0.071 0.066	0.083 0.127
from G_2	0.078 0.119 0.41	0.072 0.083 0.59	0.081 0.116	0.073 0.072

Simulation results

	2-means (0.071)		2-medoids (0.072)	
	in C_1	in C_2	in C_1	in C_2
from G_1	0.068 0.064 0.39	0.083 0.139 0.61	0.071 0.066 0.072	0.083 0.127 0.083
from G_2	0.078 0.119 0.41	0.072 0.083 0.59	0.081 0.116 0.081	0.073 0.072 0.073

Simulation results

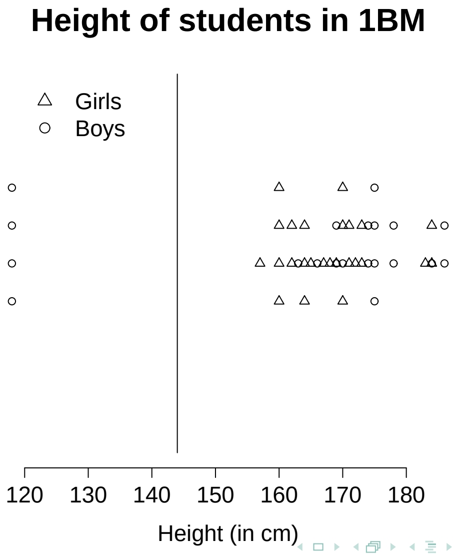
2-means (0.071) 2-medoids (0.072)

	2-means in C_1	2-means in C_2	2-medoids in C_1	2-medoids in C_2
from G_1	0.068	0.083	0.071	0.083
	0.064	0.139	0.066	0.127
	0.39	0.61	0.072	0.083
	0.35	0.65		
from G_2	0.078	0.072	0.081	0.073
	0.119	0.083	0.116	0.072
	0.41	0.59	0.081	0.073
	0.45	0.55		

Simulation results

	2-means (0.071)		2-medoids (0.072)	
	in C_1	in C_2	in C_1	in C_2
from G_1	0.068	0.083	0.071	0.083
	0.064	0.139	0.066	0.127
	0.39	0.61	0.072	0.083
	0.35	0.65	0.35	0.65
from G_2	0.078	0.072	0.081	0.073
	0.119	0.083	0.116	0.072
	0.41	0.59	0.081	0.073
	0.45	0.55	0.45	0.55

Contaminated example with the 2-medoids



Conclusion of these simulations

- A well-placed contamination in the smallest group makes the error rate decrease
- Too much contamination makes the error rate of the 2-means break down
- Unfortunately, the error rate of the 2-medoids also breaks down when there are too much and too far outliers !

Future research

Future research

- Generalized trimmed 2-means : for $\alpha \in [0,1]$, $(T_1(F), T_2(F))$ are solution of

$$\min_{\{A:F(A)=1-\alpha\}} \min_{\{t_1,t_2\}\subset\mathbb{R}} \int_A \Omega\left(\inf_{1\leq j\leq 2} \|x-t_j\|\right) dF(x).$$

- Theoretical error rate

$$\mathrm{ER}(F_\varepsilon, F_m) = \sum_{j=1}^{2} \pi_j(F_m) \mathbb{P}_{F_m}\left[R_{F_\varepsilon}(X) \neq C_j(F_\varepsilon) \middle| G_j\right]$$

- More than 1 dimension or more than 2 groups

Thank you for your attention!

Bibliography

- CROUX C., FILZMOSER P. and JOOSSENS K. (2008), Classification efficiences for robust linear discriminant analysis, *Statistica Sinica* 18, pp. 581-599.
- CROUX C., HAESBROECK G. and JOOSSENS K. (2008), Logistic discrimination using robust estimators : an influence function approach, *The Canadian Journal of Statistics*, 36, pp. 157-174.
- FERNHOLZ L. T., On multivariate higher order von Mises expansions in *Metrika* 2001, vol. 53, pp. 123-140.

Bibliography

- GARCÍA-ESCUDERO L. A. and GORDALIZA A., Robustness Properties of k Means and Trimmed k Means, *Journal of the American Statistical Association*, September 1999, Vol. 94, nř 447, pp. 956-969.
- HAMPEL F.R., RONCHETTI E.M., ROUSSEEUW P.J., STAHEL W.A., Robust Statistics : The Approach Based on Influence Functions, John Wiley and Sons, New-York, 1986.
- ANDERSON T.W., *An Introduction to Multivariate Statistical Analysis*, Wiley, New-York, 1958, pp. 126-133.

Bibliography

- POLLARD D., Strong Consistency of k-Means Clustering, *The Annals of Probability*, 1981, Vol.9, nř4, pp.919-926.
- POLLARD D., A Central Limit Theorem for k-Means Clustering, *The Annals of Probability*, 1982, Vol.10, nř1, pp.135-140.
- QIU D. and TAMHANE A. C. (2007), A comparative study of the *k*-means algorithm and the normal mixture model for clustering : Univariate case, *Journal of Statistical Planning and Inference*, 137, pp. 3722-3740.