Statistiques et Sélection de modèles

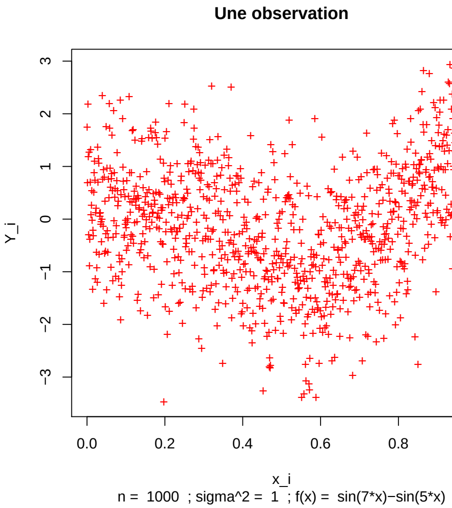
Sylvain Arlot

20 avril 2007

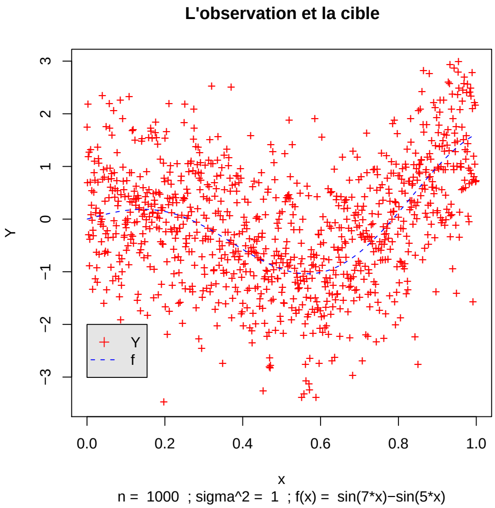
Part I

Introduction

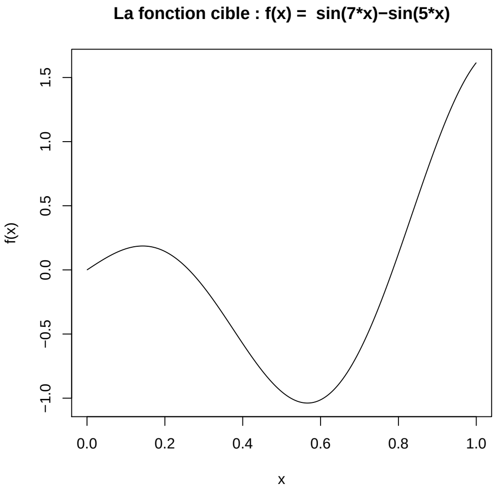
Données


Une observation
3
2
1
0
−1
−2
−3
Y_i
0.0
0.2
0.4
0.6
0.8
1.0
x_i
n = 1000 ; sigma^2 = 1 ; f(x) = sin(7*x)−sin(5*x)

But : retrouver le signal d'origine



Le signal d'origine



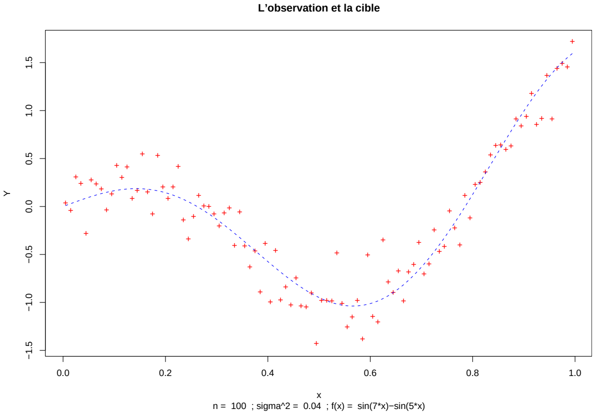
Exemples

Comment faire ?

Part II

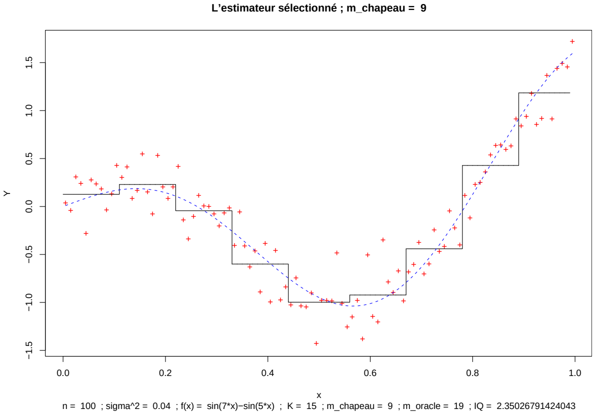
Estimation par histogrammes

Données



L'observation et la cible
Y
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
0.0
0.2
0.4
0.6
0.8
1.0
x
n = 100 ; sigma^2 = 0.04 ; f(x) = sin(7*x)−sin(5*x)

Histogramme à 9 pas



Quel est le “meilleur” histogramme estimant le signal ?

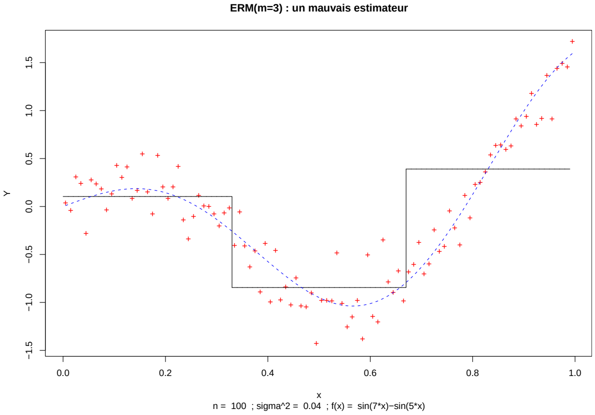
Part III

Sélection de modèles

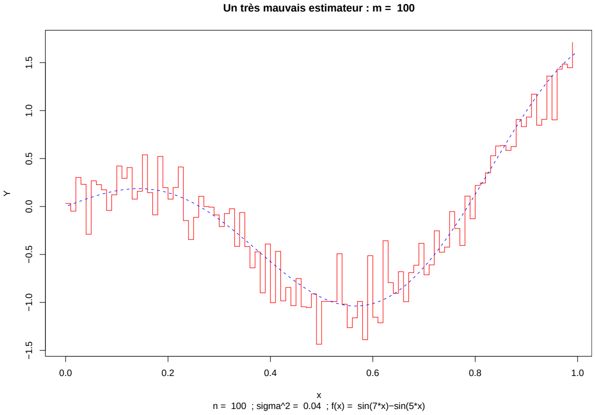
$1 \leq D \leq n \longrightarrow \widehat{s}_D$ histogramme à D pas

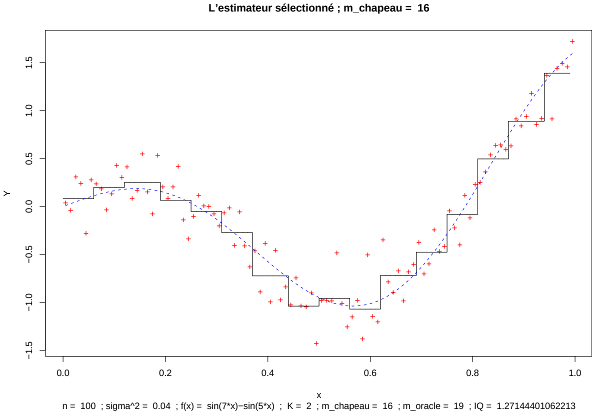
- Quel est le meilleur D ?
- Comment le trouver ?

Histogrammes : $D = 3$

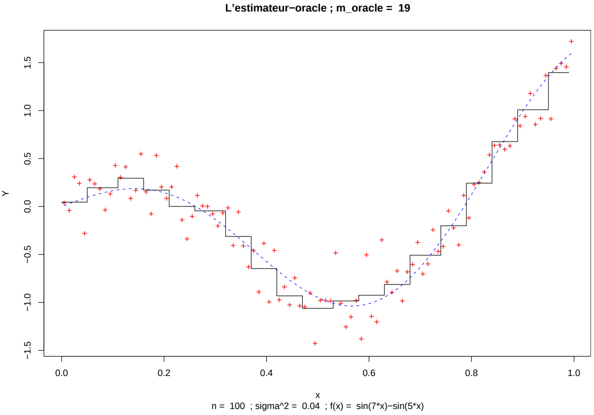


Histogrammes : $D = n = 100$

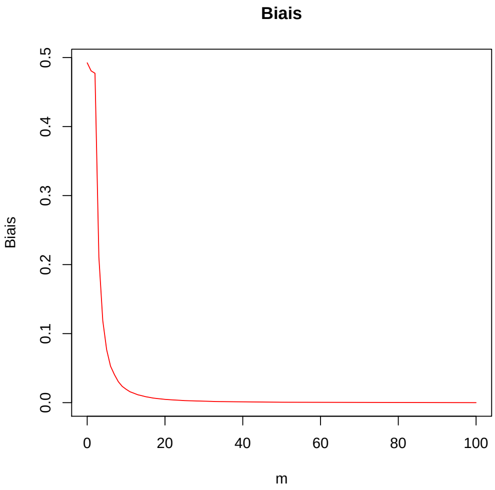



L'estimateur sélectionné ; m_chapeau = 16
Y
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
0.0
0.2
0.4
0.6
0.8
1.0
x
n = 100 ; sigma^2 = 0.04 ; f(x) = sin(7*x)−sin(5*x) ; K = 2 ; m_chapeau = 16 ; m_oracle = 19 ; IQ = 1.27144401062213

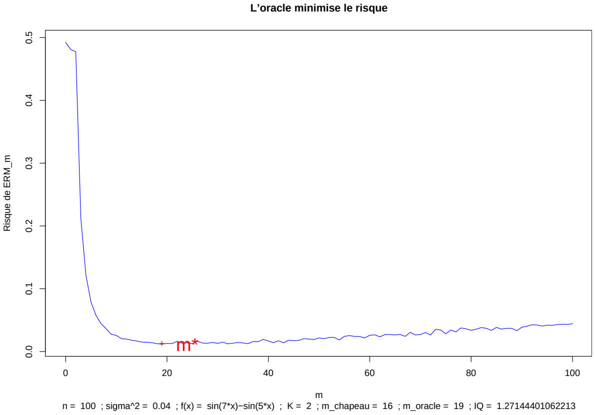
Histogrammes : $D = 19$



Le biais décroît : on ne cherche pas le “vrai” modèle



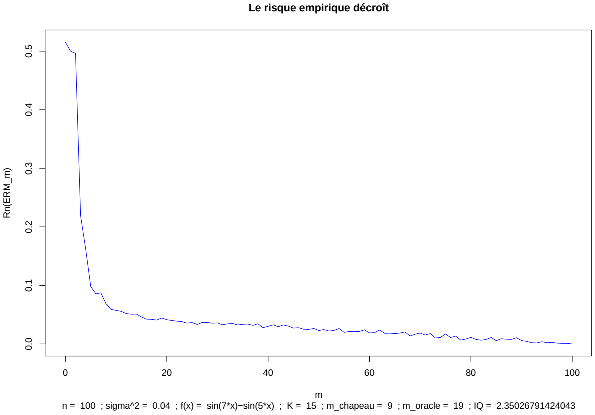
Risque en fonction du nombre D de paramètres



Part IV

Comment choisir le meilleur modèle ?

Décroissance du risque empirique

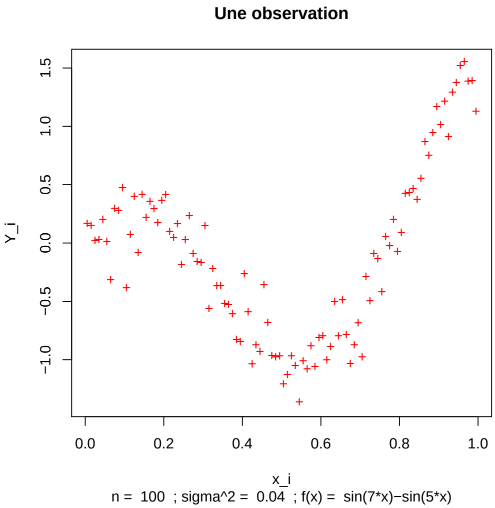


Plusieurs méthodes

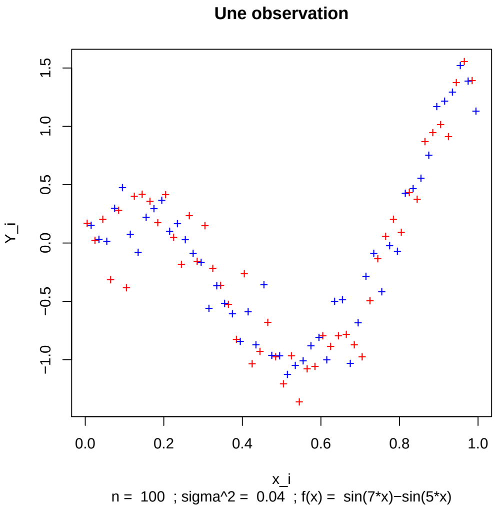
Part V

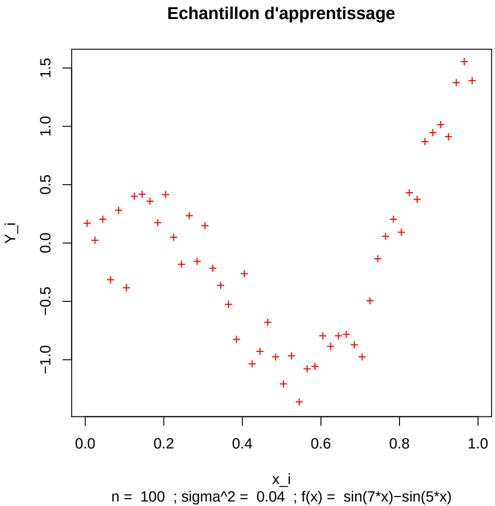
Première méthode : validation

Une observation

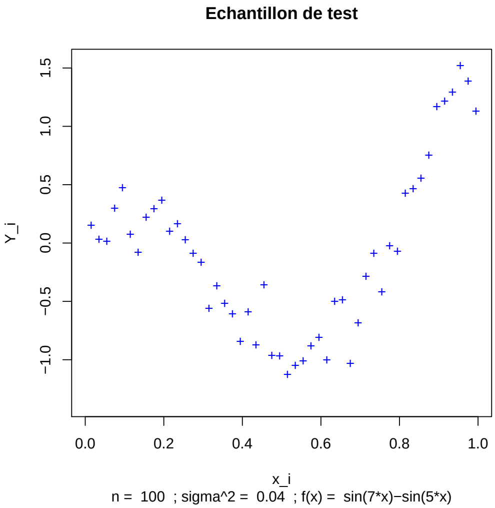


On la découpe en deux parties

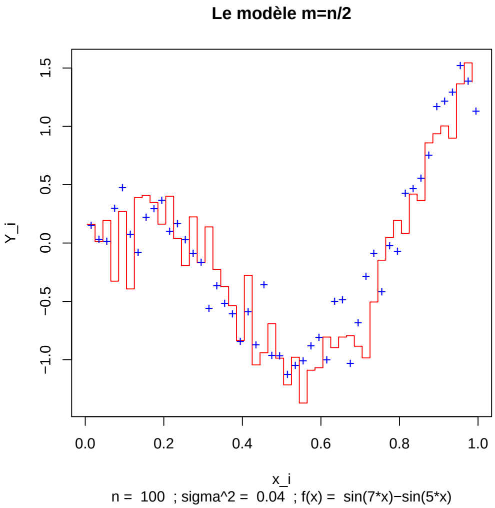


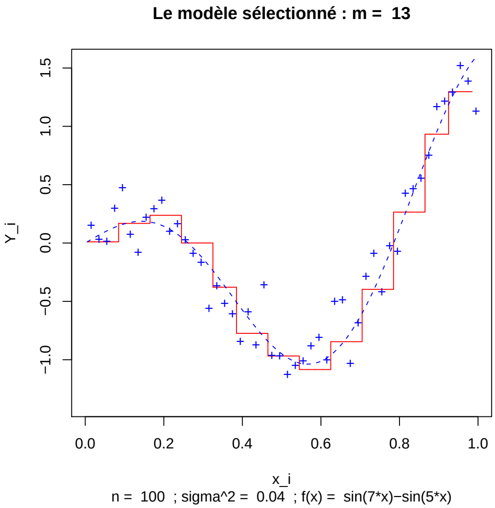

Echantillon d'apprentissage
Y_i
1.5
1.0
0.5
0.0
−0.5
−1.0
0.0
0.2
0.4
0.6
0.8
1.0
x_i
n = 100 ; sigma^2 = 0.04 ; f(x) = sin(7*x)−sin(5*x)

Deuxième échantillon : test $\longrightarrow \widehat{D}$.



Un mauvais choix : $D = 50 = n/2$.

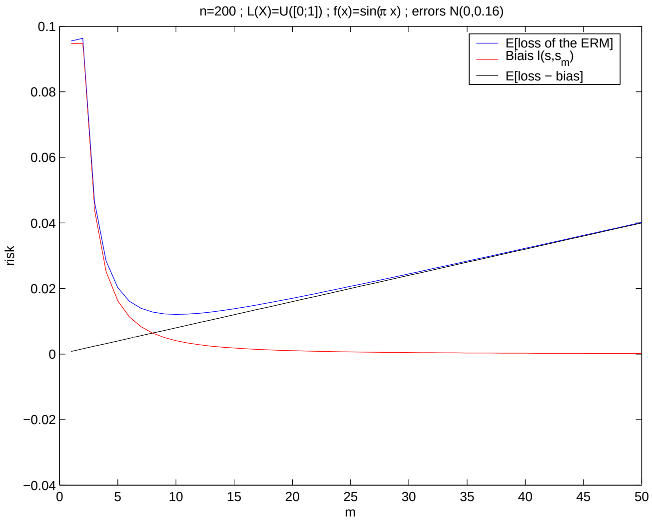



Le modèle sélectionné : m = 13
Y_i
x_i
n = 100 ; sigma^2 = 0.04 ; f(x) = sin(7*x)−sin(5*x)
1.5
1.0
0.5
0.0
−0.5
−1.0
0.0
0.2
0.4
0.6
0.8
1.0

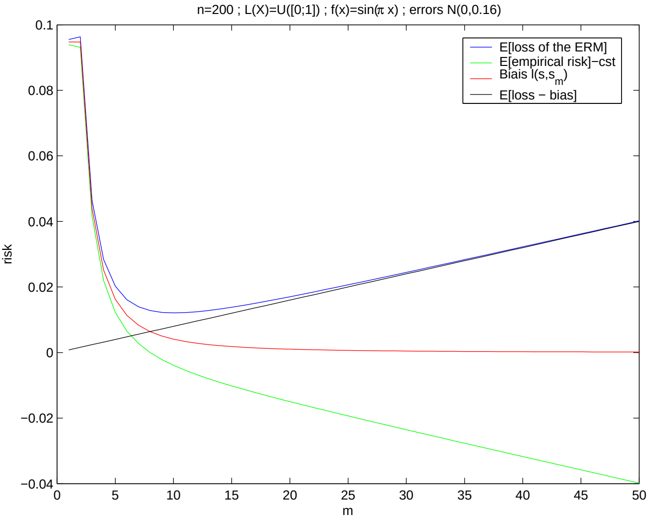
Part VI

Deuxième méthode : pénalisation

Décomposition biais-variance du risque



Risque, biais, risque empirique : pénalité idéale



Part VII

Estimation de la variance

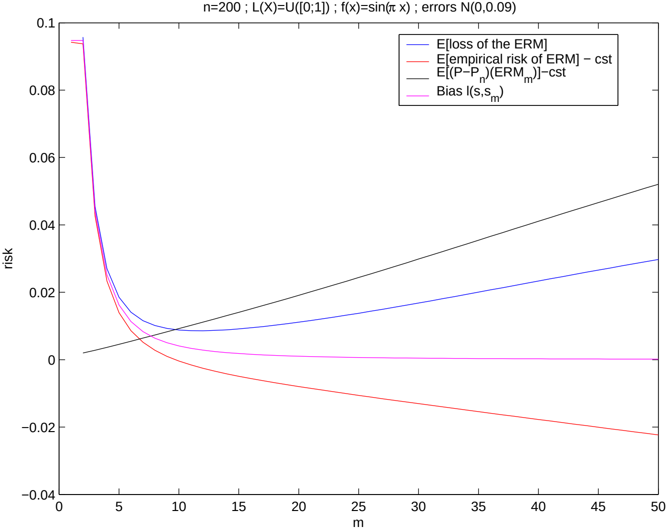
Si l'on connaît les ϵ_i

Si l'on connaît les $\epsilon_i + s_1$

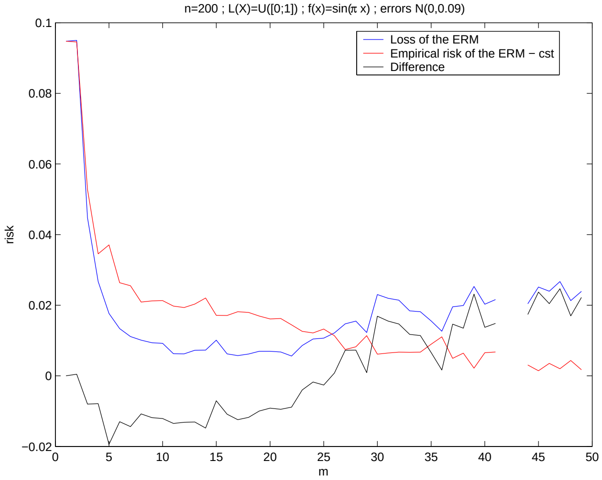
Si l'on ne connaît que $Y_i = \epsilon_i + s(X_i)$

À partir de la courbe $R_n(D)$ du risque empirique

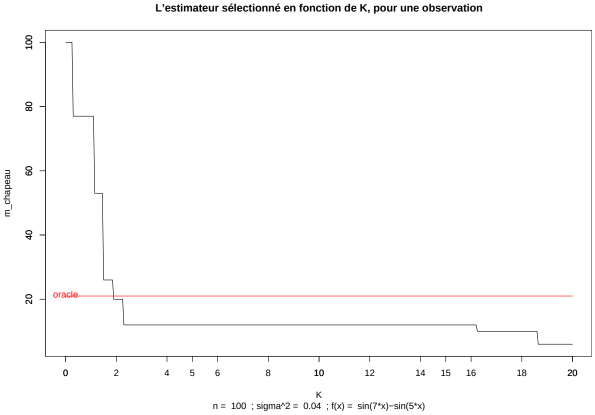
Heuristique de pente : espérances



Heuristique de pente : une réalisation



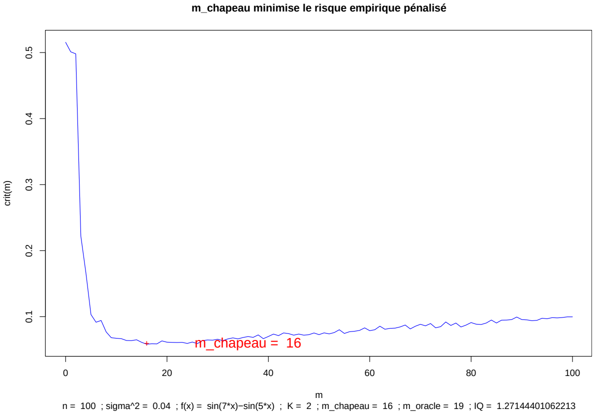
Heuristique de pente : $\widehat{D}(K)$

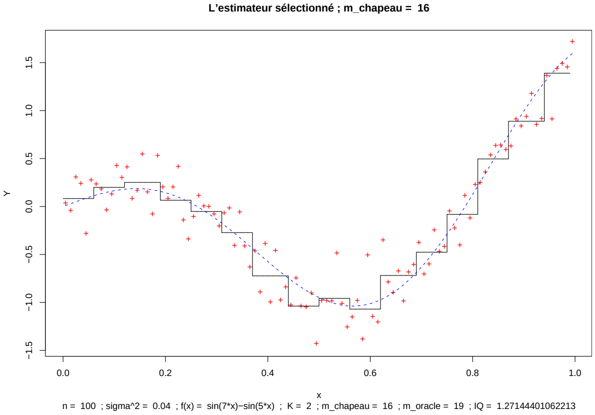


Part VIII

Pénalisation : simulations numériques

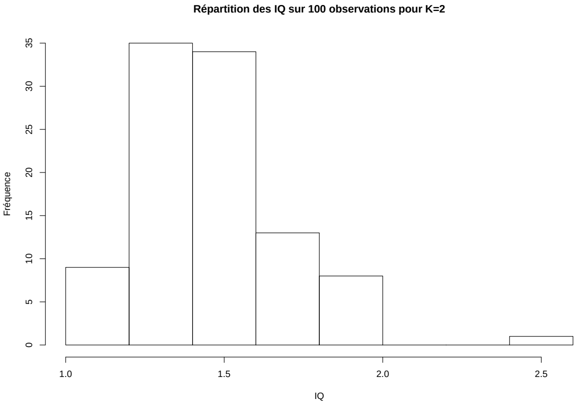
Pénalité de Mallows : crit(m).





L'estimateur sélectionné ; m_chapeau = 16
Y
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
0.0
0.2
0.4
0.6
0.8
1.0
x
n = 100 ; sigma^2 = 0.04 ; f(x) = sin(7*x)−sin(5*x) ; K = 2 ; m_chapeau = 16 ; m_oracle = 19 ; IQ = 1.27144401062213

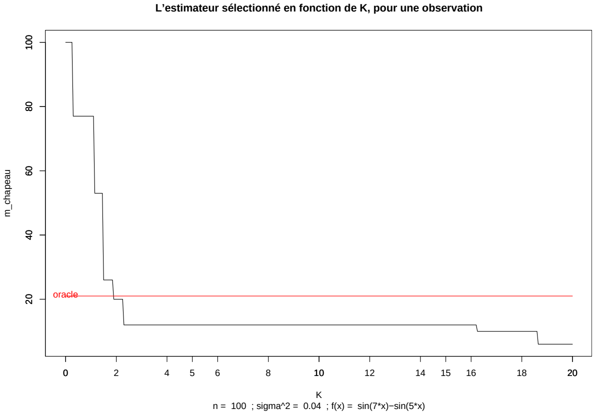
$K = 2$: répartition de $\mathcal{I}_Q = \frac{d_n(s,\widehat{s}_{\widehat{D}})^2}{d_n(s,\widehat{s}_{D^\star})^2}$ (moy. ≈ 1.45)



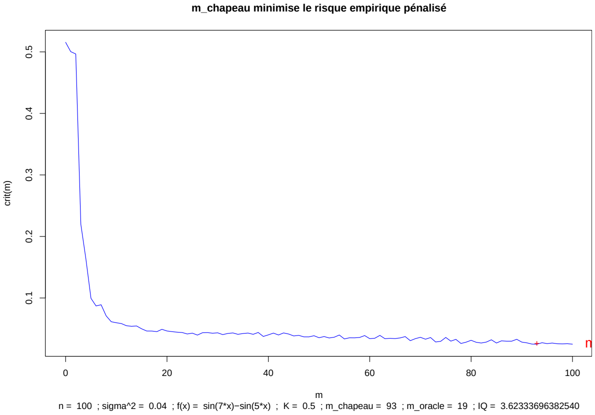
Part IX

Doit-on sur-pénaliser ou sous-pénaliser ?

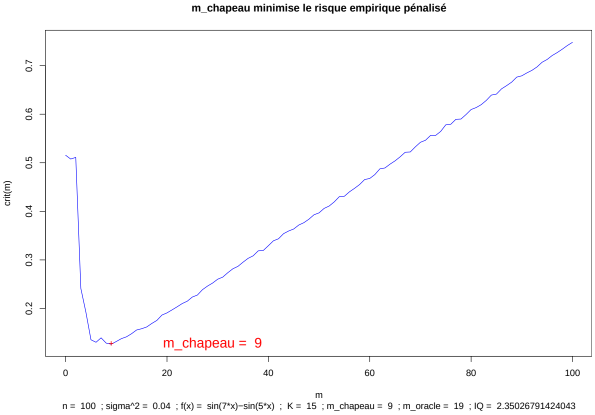
Choix de K dans la pénalité $K \times D/n$?



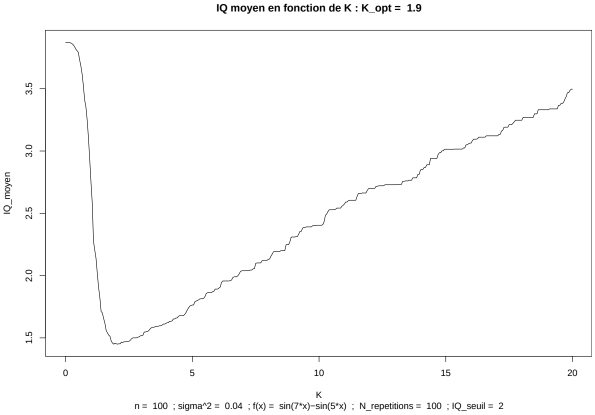
Pourquoi éviter K petit : crit(m)



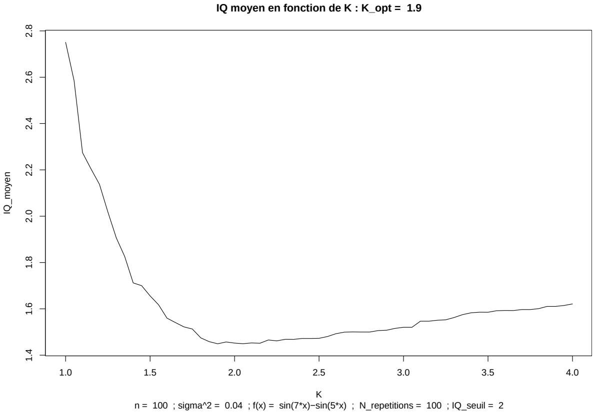
Pourquoi éviter K grand : crit(m)



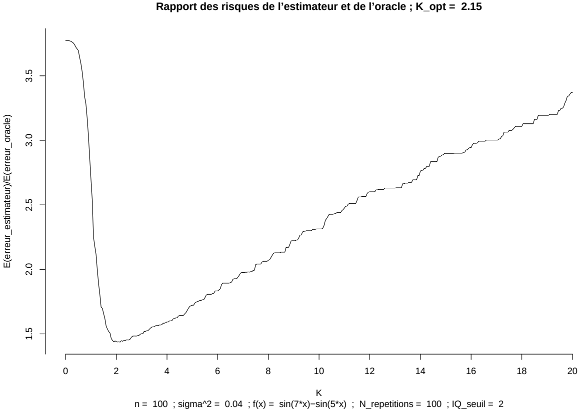
Il faut éviter $K \leq 1$ ou $K \gg 1$.



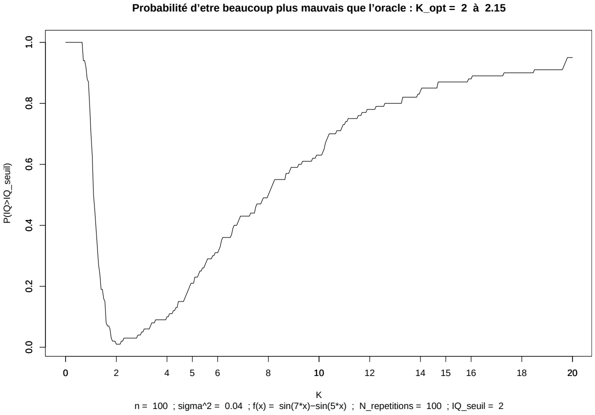
$1{,}5 \leq K \leq 4$ semble un bon choix.



$\dfrac{\text{Risque}(\widehat{s}_{\widehat{D}})}{\text{Risque}(\widehat{s}_{D^\star})}$ en fonction de K : $K_{\text{opt}} = 2{,}15$.



$\mathbb{P}(\mathcal{I}_Q > 2)$ en fonction de K : $2 \le K_{\text{opt}} \le 2{,}15$



Part X

Apprentissage