

Rééchantillonnage et Sélection de modèles

Sylvain Arlot

Université Paris-Sud XI, Orsay
thèse dirigée par Pascal Massart

13 décembre 2007

Présentation générale

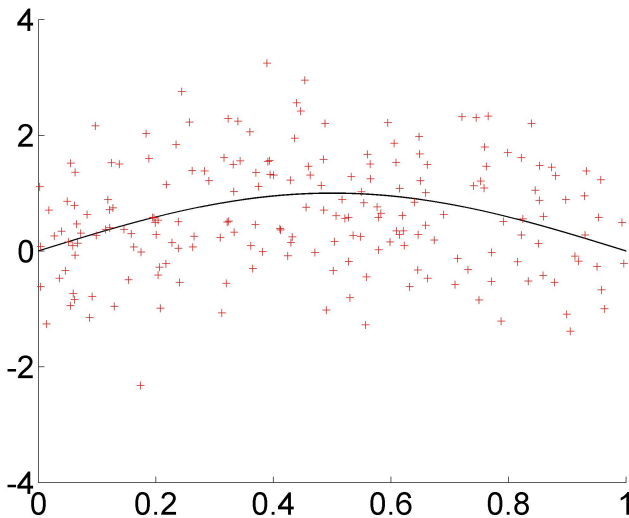
1 Sélection de modèles

- Validation croisée
- Pénalisation
 - Estimation de la forme de la pénalité
 - Calibration optimale

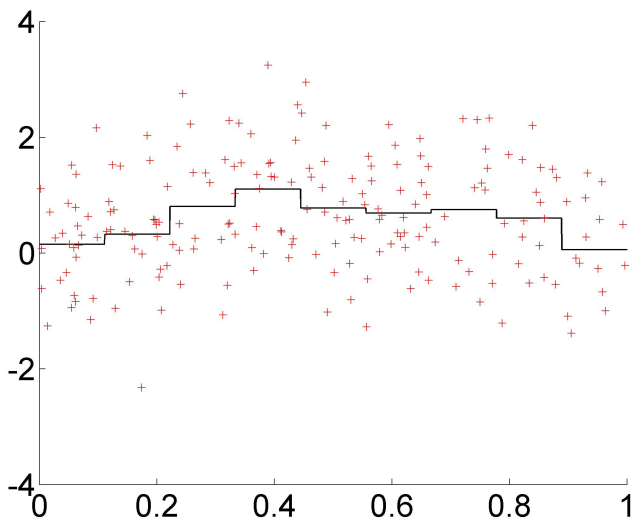
2 Régions de confiance

- Problèmes complexes
- Situation non-asymptotique
- Buts : résultats théoriques, calibration “optimale” en pratique

But : reconstruire le signal



L'oracle



Régression

$(X_1, Y_1), \dots, (X_n, Y_n) \in \mathcal{X} \times \mathcal{Y}$ i.i.d. $(X_i, Y_i) \sim P$ inconnue

$Y = s(X) + \sigma(X)\epsilon$ $X \in \mathcal{X} \subset \mathbb{R}^d$, $Y \in \mathcal{Y} = [0; 1]$ ou \mathbb{R}

bruit : $\mathbb{E}[\epsilon|X] = 0$ $\mathbb{E}[\epsilon^2|X] = 1$

intensité du bruit : $\sigma(X)$

prédicteur $t : \mathcal{X} \mapsto \mathcal{Y}$?

Régression

$(X_1, Y_1), \dots, (X_n, Y_n) \in \mathcal{X} \times \mathcal{Y}$ i.i.d. $(X_i, Y_i) \sim P$ inconnue

$$Y = s(X) + \sigma(X)\epsilon \quad X \in \mathcal{X} \subset \mathbb{R}^d, \quad Y \in \mathcal{Y} = [0; 1] \text{ ou } \mathbb{R}$$

$$\text{bruit :} \quad \mathbb{E}[\epsilon | X] = 0 \quad \mathbb{E}[\epsilon^2 | X] = 1$$

intensité du bruit : $\sigma(X)$

prédicteur $t : \mathcal{X} \mapsto \mathcal{Y}$?

Fonction de perte, estimateur des moindres carrés

- Fonction de perte des moindres carrés :

$$l(s, t) = P\gamma(t, \cdot) - P\gamma(s, \cdot) = \mathbb{E} [(t(X) - s(X))^2]$$

avec $\gamma(t, (x, y)) = (t(x) - y)^2$

- Minimiseur du risque empirique sur S_m (= modèle) :

$$\hat{s}_m \in \arg \min_{t \in S_m} P_n \gamma(t, \cdot) = \arg \min_{t \in S_m} \frac{1}{n} \sum_{i=1}^n (t(X_i) - Y_i)^2 .$$

- Exemple jouet : histogrammes sur une partition $(I_\lambda)_{\lambda \in \Lambda_m}$ de \mathcal{X} .

$$\hat{s}_m = \sum_{\lambda \in \Lambda_m} \hat{\beta}_\lambda \mathbf{1}_{I_\lambda} \quad \hat{\beta}_\lambda = \frac{1}{\text{Card}\{X_i \in I_\lambda\}} \sum_{X_i \in I_\lambda} Y_i .$$

Fonction de perte, estimateur des moindres carrés

- Fonction de perte des moindres carrés :

$$l(s, t) = P\gamma(t, \cdot) - P\gamma(s, \cdot) = \mathbb{E} [(t(X) - s(X))^2]$$

avec $\gamma(t, (x, y)) = (t(x) - y)^2$

- Minimiseur du risque empirique sur S_m (= modèle) :

$$\hat{s}_m \in \arg \min_{t \in S_m} P_n \gamma(t, \cdot) = \arg \min_{t \in S_m} \frac{1}{n} \sum_{i=1}^n (t(X_i) - Y_i)^2 .$$

- Exemple jouet : histogrammes sur une partition $(I_\lambda)_{\lambda \in \Lambda_m}$ de \mathcal{X} .

$$\hat{s}_m = \sum_{\lambda \in \Lambda_m} \hat{\beta}_\lambda \mathbb{1}_{I_\lambda} \quad \hat{\beta}_\lambda = \frac{1}{\text{Card}\{X_i \in I_\lambda\}} \sum_{X_i \in I_\lambda} Y_i .$$

Fonction de perte, estimateur des moindres carrés

- Fonction de perte des moindres carrés :

$$l(s, t) = P\gamma(t, \cdot) - P\gamma(s, \cdot) = \mathbb{E} [(t(X) - s(X))^2]$$

avec $\gamma(t, (x, y)) = (t(x) - y)^2$

- Minimiseur du risque empirique sur S_m (= modèle) :

$$\hat{s}_m \in \arg \min_{t \in S_m} P_n \gamma(t, \cdot) = \arg \min_{t \in S_m} \frac{1}{n} \sum_{i=1}^n (t(X_i) - Y_i)^2 .$$

- Exemple jouet : histogrammes sur une partition $(I_\lambda)_{\lambda \in \Lambda_m}$ de \mathcal{X} .

$$\hat{s}_m = \sum_{\lambda \in \Lambda_m} \hat{\beta}_\lambda \mathbf{1}_{I_\lambda} \quad \hat{\beta}_\lambda = \frac{1}{\text{Card}\{X_i \in I_\lambda\}} \sum_{X_i \in I_\lambda} Y_i .$$

Sélection de modèles

$$(S_m)_{m \in \mathcal{M}} \longrightarrow (\hat{S}_m)_{m \in \mathcal{M}} \longrightarrow \hat{S}_{\hat{m}} \quad ???$$

Objectifs :

- Inégalité-oracle :

$$\mathbb{E} [l(s, \hat{S}_{\hat{m}})] \leq C \mathbb{E} \left[\inf_{m \in \mathcal{M}} \{l(s, \hat{S}_m) + R(m, n)\} \right]$$

$$\mathbb{P} \left(l(s, \hat{S}_{\hat{m}}) \leq C \inf_{m \in \mathcal{M}} \{l(s, \hat{S}_m) + R(m, n)\} \right) \geq 1 - C' n^{-2}$$

- Adaptativité (e.g., (α, R) si $s \in \mathcal{H}(\alpha, R)$; $\sigma(X)$ dans le cas hétéroscédastique)

Sélection de modèles

$$(S_m)_{m \in \mathcal{M}} \longrightarrow (\hat{S}_m)_{m \in \mathcal{M}} \longrightarrow \hat{s}_{\hat{m}} \quad ???$$

Objectifs :

- Inégalité-oracle :

$$\mathbb{E} [I(s, \hat{s}_{\hat{m}})] \leq C \mathbb{E} \left[\inf_{m \in \mathcal{M}} \{I(s, \hat{S}_m) + R(m, n)\} \right]$$

$$\mathbb{P} \left(I(s, \hat{s}_{\hat{m}}) \leq C \inf_{m \in \mathcal{M}} \{I(s, \hat{S}_m) + R(m, n)\} \right) \geq 1 - C' n^{-2}$$

- Adaptativité (e.g., (α, R) si $s \in \mathcal{H}(\alpha, R)$; $\sigma(X)$ dans le cas hétéroscédastique)

Sélection de modèles

$$(S_m)_{m \in \mathcal{M}} \longrightarrow (\hat{S}_m)_{m \in \mathcal{M}} \longrightarrow \hat{s}_{\hat{m}} \quad ???$$

Objectifs :

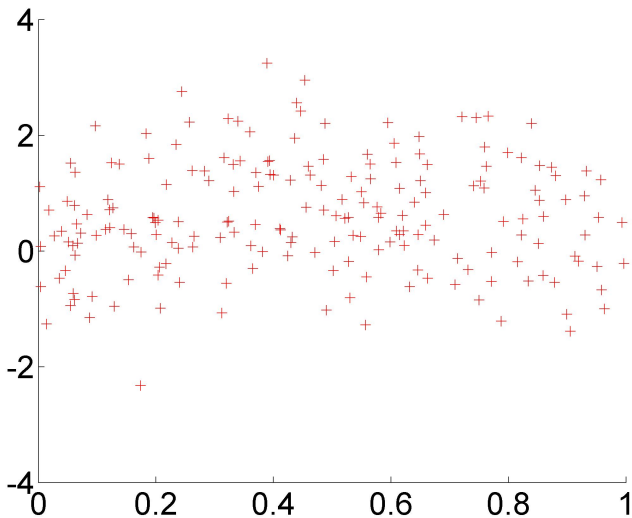
- Inégalité-oracle :

$$\mathbb{E} [l(s, \hat{s}_{\hat{m}})] \leq C \mathbb{E} \left[\inf_{m \in \mathcal{M}} \{l(s, \hat{s}_m) + R(m, n)\} \right]$$

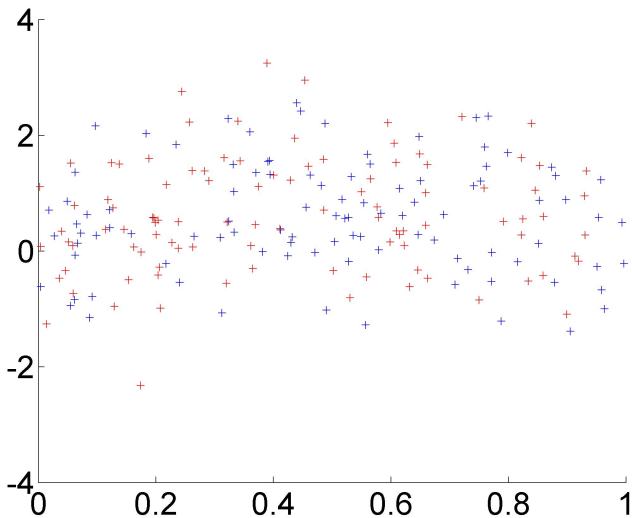
$$\mathbb{P} \left(l(s, \hat{s}_{\hat{m}}) \leq C \inf_{m \in \mathcal{M}} \{l(s, \hat{s}_m) + R(m, n)\} \right) \geq 1 - C' n^{-2}$$

- Adaptativité (e.g., (α, R) si $s \in \mathcal{H}(\alpha, R)$; $\sigma(X)$ dans le cas hétéroscédastique)

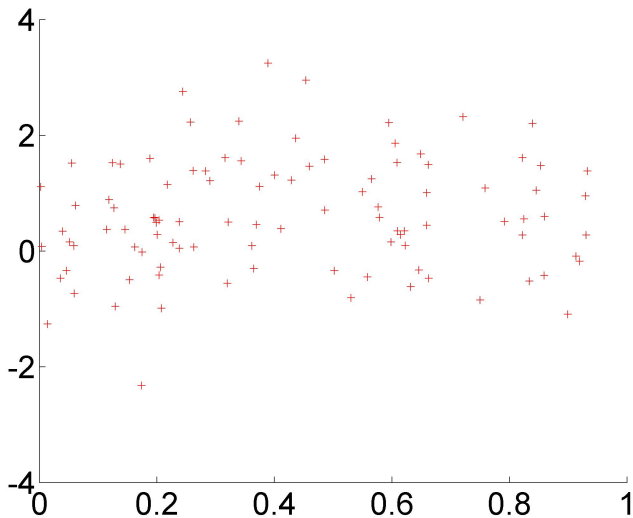
Validation



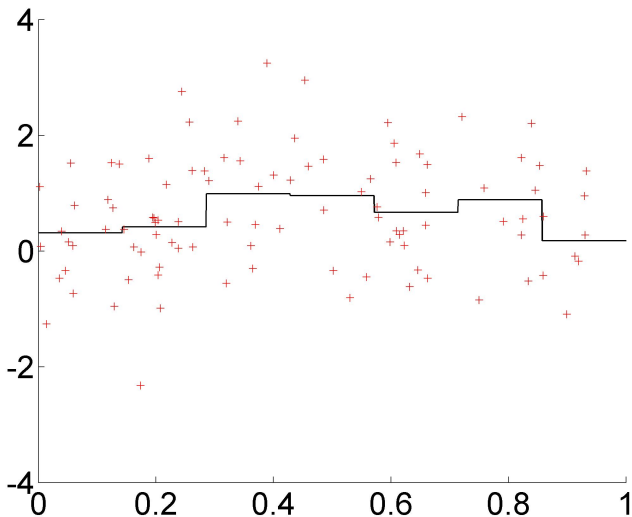
Validation



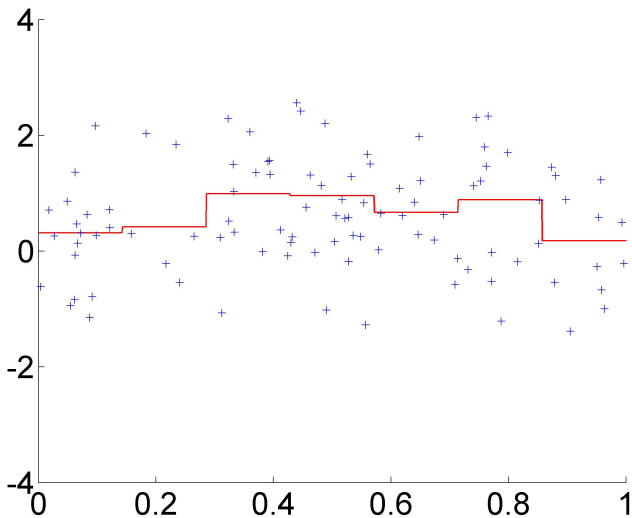
Validation : l'échantillon d'entraînement



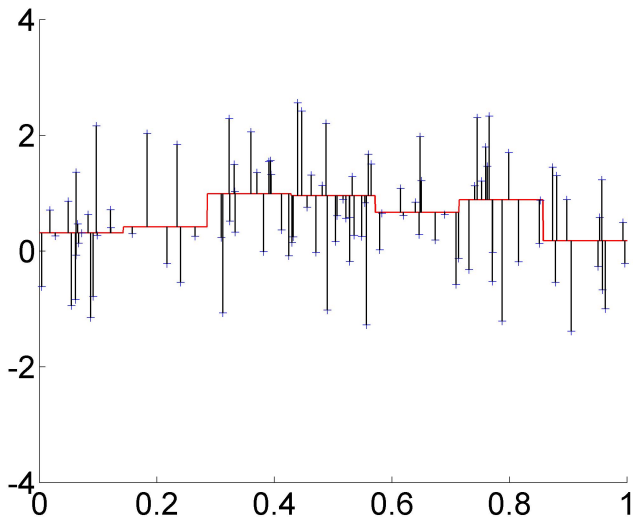
Validation : l'échantillon d'entraînement



Validation : l'échantillon de validation



Validation : l'échantillon de validation



Validation, validation croisée

$$\underbrace{(X_1, Y_1), \dots, (X_q, Y_q)}_{\text{Entraînement}}, \underbrace{(X_{q+1}, Y_{q+1}), \dots, (X_n, Y_n)}_{\text{Validation}}$$

$$\hat{s}_m^{(e)} \in \arg \min_{t \in S_m} \left\{ \sum_{i=1}^q \gamma(t, (X_i, Y_i)) \right\}$$

$$P_n^{(v)} = \frac{1}{n-q} \sum_{i=q+1}^n \delta_{(X_i, Y_i)} \quad \Rightarrow \quad P_n^{(v)} \gamma \left(\hat{s}_m^{(e)} \right)$$

Validation croisée "V-fold" : $(B_j)_{1 \leq j \leq V}$ partition de $\{1, \dots, n\}$

$$\Rightarrow \hat{m} \in \arg \min_{m \in \mathcal{M}} \left\{ \frac{1}{V} \sum_{j=1}^V P_n^j \gamma \left(\hat{s}_m^{(-j)} \right) \right\} \quad \tilde{s} = \hat{s}_{\hat{m}}$$

Validation, validation croisée

$$\underbrace{(X_1, Y_1), \dots, (X_q, Y_q)}_{\text{Entraînement}}, \underbrace{(X_{q+1}, Y_{q+1}), \dots, (X_n, Y_n)}_{\text{Validation}}$$

$$\hat{s}_m^{(e)} \in \arg \min_{t \in S_m} \left\{ \sum_{i=1}^q \gamma(t, (X_i, Y_i)) \right\}$$

$$P_n^{(v)} = \frac{1}{n-q} \sum_{i=q+1}^n \delta_{(X_i, Y_i)} \quad \Rightarrow \quad P_n^{(v)} \gamma \left(\hat{s}_m^{(e)} \right)$$

Validation croisée "V-fold" : $(B_j)_{1 \leq j \leq V}$ partition de $\{1, \dots, n\}$

$$\Rightarrow \hat{m} \in \arg \min_{m \in \mathcal{M}} \left\{ \frac{1}{V} \sum_{j=1}^V P_n^j \gamma \left(\hat{s}_m^{(-j)} \right) \right\} \quad \tilde{s} = \hat{s}_{\hat{m}}$$

Biais de la validation croisée V-fold

Critère idéal : $P\gamma(\hat{s}_m)$

Régression sur un modèle d'histogrammes :

$$\mathbb{E} [P\gamma(\hat{s}_m)] \approx P\gamma(s_m) + \frac{1}{n} \sum_{\lambda \in \Lambda_m} \sigma_\lambda^2$$

$$\mathbb{E} \left[P_n^{(j)} \gamma \left(\hat{s}_m^{(-j)} \right) \right] = \mathbb{E} \left[P\gamma \left(\hat{s}_m^{(-j)} \right) \right] \approx P\gamma(s_m) + \frac{V}{V-1} \frac{1}{n} \sum_{\lambda \in \Lambda_m} \sigma_\lambda^2$$

⇒ biais si V fixe ("surpénalisation")

⇒ sélection de modèles sous-optimale

Biais de la validation croisée V-fold

Critère idéal : $P\gamma(\hat{s}_m)$

Régression sur un modèle d'histogrammes :

$$\mathbb{E} [P\gamma(\hat{s}_m)] \approx P\gamma(s_m) + \frac{1}{n} \sum_{\lambda \in \Lambda_m} \sigma_\lambda^2$$

$$\mathbb{E} \left[P_n^{(j)} \gamma \left(\hat{s}_m^{(-j)} \right) \right] = \mathbb{E} \left[P\gamma \left(\hat{s}_m^{(-j)} \right) \right] \approx P\gamma(s_m) + \frac{V}{V-1} \frac{1}{n} \sum_{\lambda \in \Lambda_m} \sigma_\lambda^2$$

⇒ biais si V fixe ("surpénalisation")

⇒ sélection de modèles sous-optimale

Biais de la validation croisée V-fold

Critère idéal : $P\gamma(\hat{s}_m)$

Régression sur un modèle d'histogrammes :

$$\mathbb{E} [P\gamma(\hat{s}_m)] \approx P\gamma(s_m) + \frac{1}{n} \sum_{\lambda \in \Lambda_m} \sigma_\lambda^2$$

$$\mathbb{E} \left[P_n^{(j)} \gamma \left(\hat{s}_m^{(-j)} \right) \right] = \mathbb{E} \left[P\gamma \left(\hat{s}_m^{(-j)} \right) \right] \approx P\gamma(s_m) + \frac{V}{V-1} \frac{1}{n} \sum_{\lambda \in \Lambda_m} \sigma_\lambda^2$$

⇒ **biais** si V fixe ("surpénalisation")

⇒ sélection de modèles **sous-optimale**

Choix de V

- Biais : décroît avec V (peut être corrigé, Burman 1989)
- Variance : éviter V trop petit ($V = 2$), et parfois $V = n$ (algorithmes instables)
- Temps de calcul : complexité proportionnelle à V

⇒ **compromis** à trouver

“ $V = 5$ ou $V = 10$ pour la prédiction”

“ $V = 2$ pour l'identification ou les tests” (Zhang, 1993 ;
Dietterich, 1998)

Ce n'est plus vrai si le rapport signal sur bruit est faible...

Choix de V

- Biais : décroît avec V (peut être corrigé, Burman 1989)
- Variance : éviter V trop petit ($V = 2$), et parfois $V = n$ (algorithmes instables)
- Temps de calcul : complexité proportionnelle à V

⇒ **compromis** à trouver

“ $V = 5$ ou $V = 10$ pour la prédiction”

“ $V = 2$ pour l'identification ou les tests” (Zhang, 1993 ;
Dietterich, 1998)

Ce n'est plus vrai si le rapport signal sur bruit est faible...

Choix de V

- Biais : décroît avec V (peut être corrigé, Burman 1989)
- Variance : éviter V trop petit ($V = 2$), et parfois $V = n$ (algorithmes instables)
- Temps de calcul : complexité proportionnelle à V

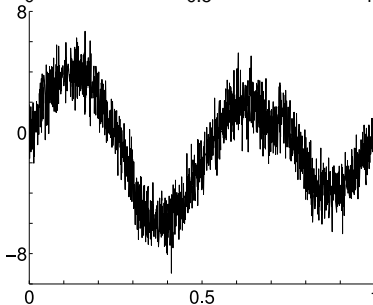
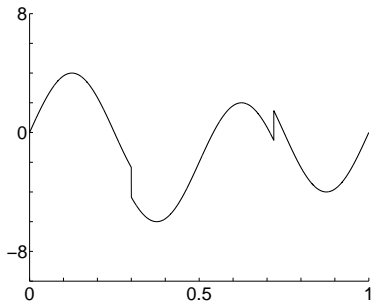
⇒ **compromis** à trouver

“ $V = 5$ ou $V = 10$ pour la prédiction”

“ $V = 2$ pour l'identification ou les tests” (Zhang, 1993 ;
Dietterich, 1998)

Ce n'est plus vrai si le rapport signal sur bruit est faible...

Simulations : HeaviSine, $n = 2048$, $\sigma \equiv 1$



$$C_{\text{classique}} = \frac{\mathbb{E} [l(s, \hat{s}_m)]}{\mathbb{E} [\inf_{m \in \mathcal{M}} l(s, \hat{s}_m)]}$$

2-fold	1.002 ± 0.003
5-fold	1.014 ± 0.003
10-fold	1.021 ± 0.003
20-fold	1.029 ± 0.004
leave-one-out	1.034 ± 0.004

Pénalisation

$$\hat{m} \in \arg \min_{m \in \mathcal{M}} \{P_n \gamma(\hat{S}_m) + \text{pen}(m)\}$$

$$\text{pen}(m) = \frac{2\sigma^2 D_m}{n} \quad (\text{Mallows 1973})$$

$$\text{pen}(m) = \frac{2\hat{\sigma}^2 D_m}{n} \quad \text{ou} \quad \hat{K} D_m$$

Mais des pénalités linéaires en D ne fonctionnent pas toujours : régression hétéroscédastique, classification, etc.

⇒ Pénalité idéale : $\text{pen}_{\text{id}}(m) = (P - P_n)(\gamma(\hat{S}_m, \cdot))$

Pénalisation

$$\hat{m} \in \arg \min_{m \in \mathcal{M}} \{P_n \gamma(\hat{S}_m) + \text{pen}(m)\}$$

$$\text{pen}(m) = \frac{2\sigma^2 D_m}{n} \quad (\text{Mallows 1973})$$

$$\text{pen}(m) = \frac{2\hat{\sigma}^2 D_m}{n} \quad \text{ou} \quad \hat{K} D_m$$

Mais des pénalités linéaires en D ne fonctionnent pas toujours : régression hétéroscédastique, classification, etc.

⇒ Pénalité idéale : $\text{pen}_{\text{id}}(m) = (P - P_n)(\gamma(\hat{S}_m, \cdot))$

Pénalisation

$$\hat{m} \in \arg \min_{m \in \mathcal{M}} \{P_n \gamma(\hat{S}_m) + \text{pen}(m)\}$$

$$\text{pen}(m) = \frac{2\sigma^2 D_m}{n} \quad (\text{Mallows 1973})$$

$$\text{pen}(m) = \frac{2\hat{\sigma}^2 D_m}{n} \quad \text{ou} \quad \hat{K} D_m$$

Mais des pénalités linéaires en D ne fonctionnent pas toujours :
régression hétéroscédastique, classification, etc.

⇒ Pénalité idéale : $\text{pen}_{\text{id}}(m) = (P - P_n)(\gamma(\hat{S}_m, \cdot))$

Heuristique de rééchantillonnage (bootstrap, Efron 1979)

Monde réel : $P \xrightarrow{\text{échantillonnage}} P_n \Longrightarrow \hat{S}_m$

$$\text{pen}_{\text{id}}(m) = (P - P_n)\gamma(\hat{S}_m) = F(P, P_n)$$

V-fold : $P_n^W = \frac{1}{n - \text{Card}(B_J)} \sum_{i \notin B_J} \delta_{(X_i, Y_i)}$ avec $J \sim \mathcal{U}(1, \dots, V)$

Heuristique de rééchantillonnage (bootstrap, Efron 1979)

Monde réel :

$$P \xrightarrow{\text{échantillonnage}} P_n \Longrightarrow \hat{S}_m$$



Monde bootstrap :

$$P_n$$

$$\text{pen}_{\text{id}}(m) = (P - P_n)\gamma(\hat{S}_m) = F(P, P_n)$$

$$V\text{-fold} : P_n^W = \frac{1}{n - \text{Card}(B_J)} \sum_{i \notin B_J} \delta_{(X_i, Y_i)} \quad \text{avec } J \sim \mathcal{U}(1, \dots, V)$$

Heuristique de rééchantillonnage (bootstrap, Efron 1979)

Monde réel :

$$P \xrightarrow{\text{échantillonnage}} P_n \Longrightarrow \hat{S}_m$$



Monde bootstrap :

$$P_n \xrightarrow{\text{rééchantillonnage}} P_n^W \Longrightarrow \hat{S}_m^W$$

$$(P - P_n)\gamma(\hat{S}_m) = F(P, P_n) \rightsquigarrow F(P_n, P_n^W) = (P_n - P_n^W)\gamma(\hat{S}_m^W)$$

$$V\text{-fold : } P_n^W = \frac{1}{n - \text{Card}(B_J)} \sum_{i \notin B_J} \delta_{(X_i, Y_i)} \quad \text{avec } J \sim \mathcal{U}(1, \dots, V)$$

Heuristique de rééchantillonnage (bootstrap, Efron 1979)

Monde réel :

$$P \xrightarrow{\text{échantillonnage}} P_n \Longrightarrow \hat{S}_m$$



Monde bootstrap :

$$P_n \xrightarrow{\text{sous-échantillonnage}} P_n^W \Longrightarrow \hat{S}_m^W$$

$$(P - P_n)\gamma(\hat{S}_m) = F(P, P_n) \rightsquigarrow F(P_n, P_n^W) = (P_n - P_n^W)\gamma(\hat{S}_m^W)$$

V-fold :
$$P_n^W = \frac{1}{n - \text{Card}(B_J)} \sum_{i \notin B_J} \delta_{(X_i, Y_i)} \quad \text{avec } J \sim \mathcal{U}(1, \dots, V)$$

Pénalisations par rééchantillonnage

- Estimer pen_{id} avec le **bootstrap** : Efron (1983), Shibata (1997)
- Estimer pen_{id} avec le **bootstrap “ m out of n ”** : Shao (1996)
⇒ identification
- ...
- Pénalités **globales en classification** : poids Rademacher (Koltchinskii, Panchenko 2001 ; Bartlett, Boucheron, Lugosi 2002), poids échangeables (Fromont 2004), ...
- Pénalité locales : **complexités de Rademacher locales** (Bartlett, Bousquet, Mendelson 2004 ; Koltchinskii 2006, etc.)

Pénalisation V-fold

- Pénalité idéale :

$$(P - P_n)(\gamma(\hat{s}_m))$$

- Pénalité V-fold :

$$\text{pen}(m) = \frac{C}{V} \sum_{j=1}^V \left[(P_n - P_n^{(-j)}) \left(\gamma \left(\hat{s}_m^{(-j)} \right) \right) \right]$$

$$\hat{s}_m^{(-j)} \in \arg \min_{t \in S_m} P_n^{(-j)} \gamma(t)$$

avec $C \geq V - 1$ à choisir

($C = V - 1 \Rightarrow$ on retrouve Burman, 1989)

- Modèle sélectionné :

$$\hat{m} \in \arg \min_{m \in \mathcal{M}} \{P_n \gamma(\hat{s}_m) + \text{pen}(m)\}$$

Pénalisation V-fold

- Pénalité idéale :

$$(P - P_n)(\gamma(\hat{s}_m))$$

- Pénalité V-fold :

$$\text{pen}(m) = \frac{C}{V} \sum_{j=1}^V \left[(P_n - P_n^{(-j)}) \left(\gamma \left(\hat{s}_m^{(-j)} \right) \right) \right]$$

$$\hat{s}_m^{(-j)} \in \arg \min_{t \in S_m} P_n^{(-j)} \gamma(t)$$

avec $C \geq V - 1$ à choisir

($C = V - 1 \Rightarrow$ on retrouve Burman, 1989)

- Modèle sélectionné :

$$\hat{m} \in \arg \min_{m \in \mathcal{M}} \{P_n \gamma(\hat{s}_m) + \text{pen}(m)\}$$

Pénalisation V-fold

- Pénalité idéale :

$$(P - P_n)(\gamma(\hat{s}_m))$$

- Pénalité V-fold :

$$\text{pen}(m) = \frac{C}{V} \sum_{j=1}^V \left[(P_n - P_n^{(-j)}) \left(\gamma \left(\hat{s}_m^{(-j)} \right) \right) \right]$$

$$\hat{s}_m^{(-j)} \in \arg \min_{t \in S_m} P_n^{(-j)} \gamma(t)$$

avec $C \geq V - 1$ à choisir

($C = V - 1 \Rightarrow$ on retrouve Burman, 1989)

- **Modèle sélectionné :**

$$\hat{m} \in \arg \min_{m \in \mathcal{M}} \{ P_n \gamma(\hat{s}_m) + \text{pen}(m) \}$$

Inégalité-oracle “trajectorielle” non-asymptotique

- Régression sur des histogrammes, plan d'expérience aléatoire
 - Petit nombre de modèles (au plus $\diamond n^\alpha$)
 - V fixe ou $V = n$ (et plusieurs rééchantillonnages échangeables, dont le bootstrap)
 - $C \approx V - 1$ (ou $C \approx C_{W,\infty}$)
 - Des hypothèses “raisonnables” sur P , e.g. données bornées, bruit minoré et s régulière non-constante
- + quelques restrictions techniques

Théorème

Avec probabilité au moins $1 - \diamond n^{-2}$,

$$l(s, \widehat{s}_m) \leq \left(1 + \ln(n)^{-1/5}\right) \inf_{m \in \mathcal{M}} \{l(s, \widehat{s}_m)\}$$

Corollaires

- Inégalité-oracle classique :

$$\mathbb{E} [l(s, \widehat{s}_{\widehat{m}})] \leq \left(1 + \ln(n)^{-1/5}\right) \mathbb{E} \left[\inf_{m \in \mathcal{M}} \{l(s, \widehat{s}_m)\} \right] + \diamond n^{-2}$$

- Optimalité asymptotique si $C \sim_{n \rightarrow +\infty} V - 1$:

$$\frac{l(s, \widehat{s}_{\widehat{m}})}{\inf_{m \in \mathcal{M}} \{l(s, \widehat{s}_m)\}} \xrightarrow[n \rightarrow +\infty]{p.s.} 1$$

- Adaptativité à la régularité hölder (histogrammes réguliers, $\sup_{\mathcal{X}} s - \inf_{\mathcal{X}} s \geq \epsilon > 0$), même dans un cadre hétéroscédastique.

Corollaires

- Inégalité-oracle classique :

$$\mathbb{E} [l(s, \widehat{s}_{\widehat{m}})] \leq \left(1 + \ln(n)^{-1/5}\right) \mathbb{E} \left[\inf_{m \in \mathcal{M}} \{l(s, \widehat{s}_m)\} \right] + \diamond n^{-2}$$

- Optimalité asymptotique si $C \sim_{n \rightarrow +\infty} V - 1$:

$$\frac{l(s, \widehat{s}_{\widehat{m}})}{\inf_{m \in \mathcal{M}} \{l(s, \widehat{s}_m)\}} \xrightarrow[n \rightarrow +\infty]{p.s.} 1$$

- Adaptativité à la régularité hölder (histogrammes réguliers, $\sup_{\mathcal{X}} s - \inf_{\mathcal{X}} s \geq \epsilon > 0$), même dans un cadre hétéroscédastique.

Corollaires

- Inégalité-oracle classique :

$$\mathbb{E} [l(s, \widehat{s}_{\widehat{m}})] \leq \left(1 + \ln(n)^{-1/5}\right) \mathbb{E} \left[\inf_{m \in \mathcal{M}} \{l(s, \widehat{s}_m)\} \right] + \diamond n^{-2}$$

- Optimalité asymptotique si $C \sim_{n \rightarrow +\infty} V - 1$:

$$\frac{l(s, \widehat{s}_{\widehat{m}})}{\inf_{m \in \mathcal{M}} \{l(s, \widehat{s}_m)\}} \xrightarrow[n \rightarrow +\infty]{p.s.} 1$$

- Adaptativité à la régularité hölder (histogrammes réguliers, $\sup_{\mathcal{X}} s - \inf_{\mathcal{X}} s \geq \epsilon > 0$), même dans un cadre hétéroscédastique.

Schéma de la preuve

On montre, pour tout $m \in \mathcal{M}$,

$$\text{pen}_{\text{id}}(m) \approx \mathbb{E}[\text{pen}_{\text{id}}(m)] \propto \mathbb{E}[\text{pen}(m)] \approx \text{pen}(m)$$

avec des restes $\ll I(s, \hat{s}_m)$ quand $D_m \rightarrow +\infty$:

- Calculs explicites de pen_{id} et de pen
- Comparaison des espérances : $\mathbb{E}(\text{pen}_{\text{id}}) \propto \mathbb{E}(\text{pen})$ (si $\min_{\lambda \in \Lambda_m} \{n\mathbb{P}(X \in I_\lambda)\} \rightarrow +\infty$)
- Inégalités de moments (Boucheron, Bousquet, Lugosi, Massart 2005)
 - \Rightarrow inégalités de concentration (pour pen_{id} et pen)
- Hypothèses \Rightarrow contrôle des restes en termes de $I(s, \hat{s}_m)$

Schéma de la preuve

On montre, pour tout $m \in \mathcal{M}$,

$$\text{pen}_{\text{id}}(m) \approx \mathbb{E}[\text{pen}_{\text{id}}(m)] \propto \mathbb{E}[\text{pen}(m)] \approx \text{pen}(m)$$

avec des restes $\ll l(s, \hat{s}_m)$ quand $D_m \rightarrow +\infty$:

- **Calculs explicites de pen_{id} et de pen**
- Comparaison des espérances : $\mathbb{E}(\text{pen}_{\text{id}}) \propto \mathbb{E}(\text{pen})$ (si $\min_{\lambda \in \Lambda_m} \{n\mathbb{P}(X \in I_\lambda)\} \rightarrow +\infty$)
- Inégalités de moments (Boucheron, Bousquet, Lugosi, Massart 2005)
 \Rightarrow inégalités de concentration (pour pen_{id} et pen)
- Hypothèses \Rightarrow contrôle des restes en termes de $l(s, \hat{s}_m)$

Schéma de la preuve

On montre, pour tout $m \in \mathcal{M}$,

$$\text{pen}_{\text{id}}(m) \approx \mathbb{E}[\text{pen}_{\text{id}}(m)] \propto \mathbb{E}[\text{pen}(m)] \approx \text{pen}(m)$$

avec des restes $\ll l(s, \hat{s}_m)$ quand $D_m \rightarrow +\infty$:

- Calculs explicites de pen_{id} et de pen
- Comparaison des espérances : $\mathbb{E}(\text{pen}_{\text{id}}) \propto \mathbb{E}(\text{pen})$ (si $\min_{\lambda \in \Lambda_m} \{n\mathbb{P}(X \in I_\lambda)\} \rightarrow +\infty$)
- Inégalités de moments (Boucheron, Bousquet, Lugosi, Massart 2005)
 \Rightarrow inégalités de concentration (pour pen_{id} et pen)
- Hypothèses \Rightarrow contrôle des restes en termes de $l(s, \hat{s}_m)$

Schéma de la preuve

On montre, pour tout $m \in \mathcal{M}$,

$$\text{pen}_{\text{id}}(m) \approx \mathbb{E}[\text{pen}_{\text{id}}(m)] \propto \mathbb{E}[\text{pen}(m)] \approx \text{pen}(m)$$

avec des restes $\ll l(s, \hat{s}_m)$ quand $D_m \rightarrow +\infty$:

- Calculs explicites de pen_{id} et de pen
- Comparaison des espérances : $\mathbb{E}(\text{pen}_{\text{id}}) \propto \mathbb{E}(\text{pen})$ (si $\min_{\lambda \in \Lambda_m} \{n\mathbb{P}(X \in I_\lambda)\} \rightarrow +\infty$)
- Inégalités de moments (Boucheron, Bousquet, Lugosi, Massart 2005)
 \Rightarrow inégalités de concentration (pour pen_{id} et pen)
- Hypothèses \Rightarrow contrôle des restes en termes de $l(s, \hat{s}_m)$

Schéma de la preuve

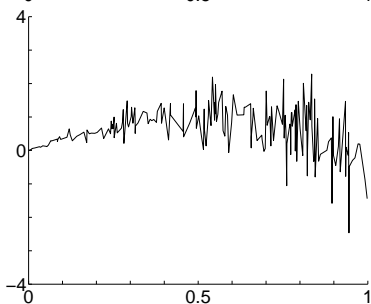
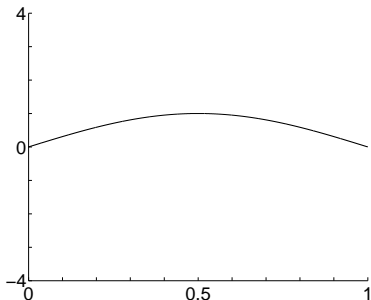
On montre, pour tout $m \in \mathcal{M}$,

$$\text{pen}_{\text{id}}(m) \approx \mathbb{E}[\text{pen}_{\text{id}}(m)] \propto \mathbb{E}[\text{pen}(m)] \approx \text{pen}(m)$$

avec des restes $\ll I(s, \hat{s}_m)$ quand $D_m \rightarrow +\infty$:

- Calculs explicites de pen_{id} et de pen
- Comparaison des espérances : $\mathbb{E}(\text{pen}_{\text{id}}) \propto \mathbb{E}(\text{pen})$ (si $\min_{\lambda \in \Lambda_m} \{n\mathbb{P}(X \in I_\lambda)\} \rightarrow +\infty$)
- Inégalités de moments (Boucheron, Bousquet, Lugosi, Massart 2005)
 - \Rightarrow inégalités de concentration (pour pen_{id} et pen)
- Hypothèses \Rightarrow contrôle des restes en termes de $I(s, \hat{s}_m)$

Simulations : \sin , $n = 200$, $\sigma(x) = x$, 2 pas



Mallows	3.86 ± 0.02
---------	-----------------

2-fold	2.54 ± 0.05
--------	-----------------

5-fold	2.58 ± 0.06
--------	-----------------

10-fold	2.60 ± 0.06
---------	-----------------

leave-one-out	2.59 ± 0.06
---------------	-----------------

pen 2-f	3.06 ± 0.07
---------	-----------------

pen 5-f	2.75 ± 0.06
---------	-----------------

pen 10-f	2.65 ± 0.06
----------	-----------------

pen 20-f	2.64 ± 0.06
----------	-----------------

pen Loo	2.59 ± 0.06
---------	-----------------

Mallows $\times 1.25$	4.05 ± 0.02
-----------------------	-----------------

pen 2-f $\times 1.25$	2.75 ± 0.06
-----------------------	-----------------

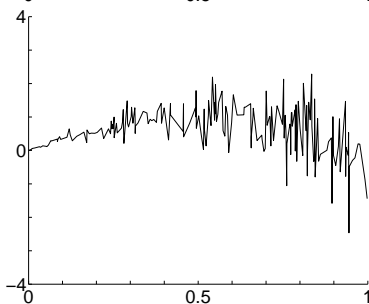
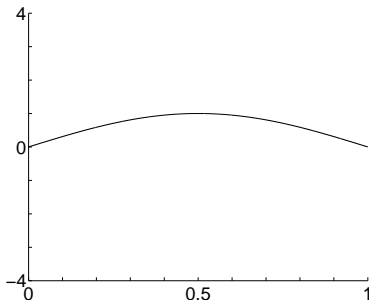
pen 5-f $\times 1.25$	2.38 ± 0.06
-----------------------	-----------------

pen 10-f $\times 1.25$	2.28 ± 0.05
------------------------	-----------------

pen 20-f $\times 1.25$	2.25 ± 0.06
------------------------	-----------------

pen Loo $\times 1.25$	2.21 ± 0.05
-----------------------	-----------------

Simulations : \sin , $n = 200$, $\sigma(x) = x$, 2 pas



Mallows	3.86 ± 0.02
2-fold	2.54 ± 0.05
5-fold	2.58 ± 0.06
10-fold	2.60 ± 0.06
leave-one-out	2.59 ± 0.06

pen 2-f	3.06 ± 0.07
pen 5-f	2.75 ± 0.06
pen 10-f	2.65 ± 0.06
pen 20-f	2.64 ± 0.06
pen Loo	2.59 ± 0.06

Mallows $\times 1.25$	4.05 ± 0.02
pen 2-f $\times 1.25$	2.75 ± 0.06
pen 5-f $\times 1.25$	2.38 ± 0.06
pen 10-f $\times 1.25$	2.28 ± 0.05
pen 20-f $\times 1.25$	2.25 ± 0.06
pen Loo $\times 1.25$	2.21 ± 0.05

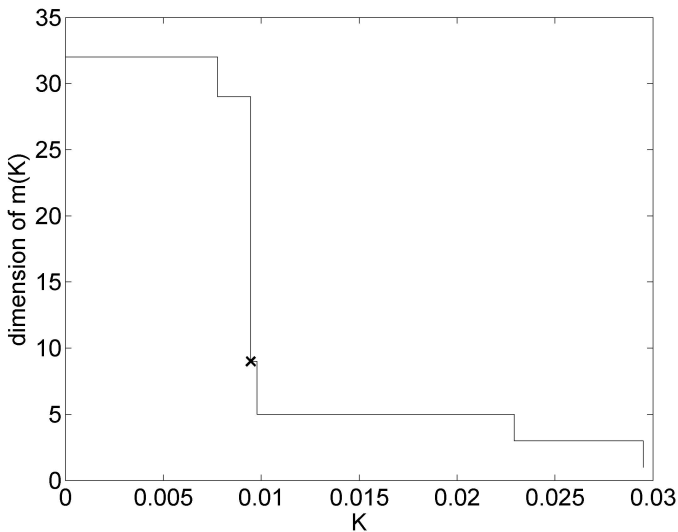
Calibration optimale de la pénalité

$$\hat{m}(K) \in \arg \min_{m \in \mathcal{M}_n} \{P_n \gamma(\hat{s}_m) + K \text{pen}_0(m)\}$$

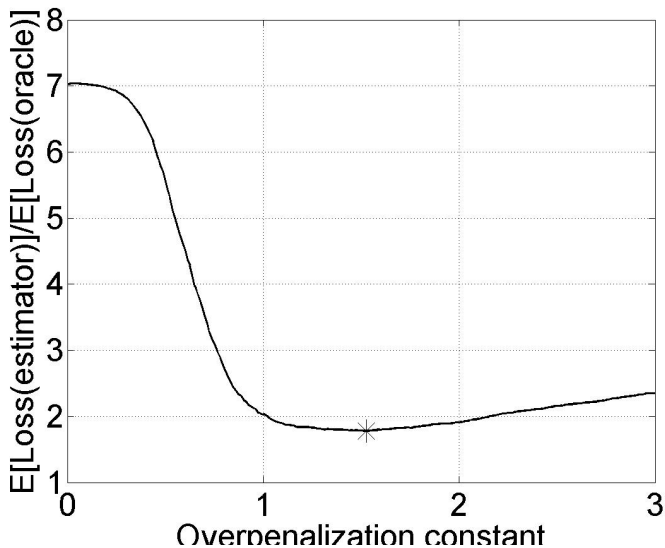
⇒ comment choisir K ?

- But : optimalité en termes de sélection de modèles
- Optimalité “asymptotique” : $\hat{K}_{opt} = 2\hat{K}_{min}$ (heuristique de pente, Birgé et Massart 2006).
⇒ encore valable dans un cadre hétéroscédastique (histogrammes).
- Optimalité “non-asymptotique” : il faut **surpénaliser** un peu

Calibration asymptotique : heuristique de pente



Calibration non-asymptotique : surpénalisation



Bilan

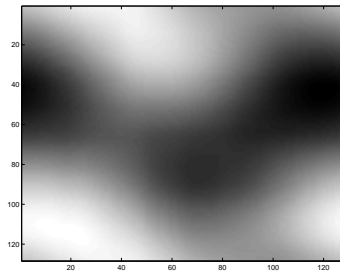
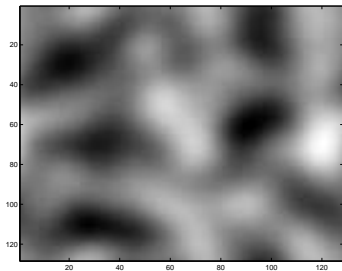
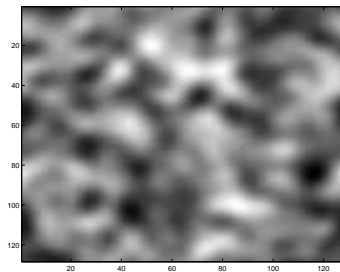
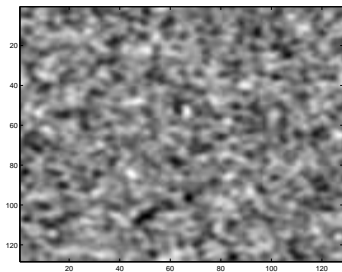
- Problème difficile (hétéroscédastique, *etc.*)
⇒ **estimation de la forme idéale**,
par rééchantillonnage V -fold ou échangeable
(preuve : histogrammes)
- Calibration de la constante multiplicative : **heuristique de pente**
(preuve : cadre non-gaussien hétéroscédastique,
histogrammes)
- Constante optimale en pratique : il faut **surpénaliser**. De combien ?

Régions de confiance et tests multiples

(collaboration avec G. Blanchard, E. Roquain ; version courte publiée aux actes de COLT 2007)

- Problématique : données de **grande dimension**, avec des **corrélations** générales inconnues, mais peu de répétitions
⇒ non-asymptotique
- Approches **par rééchantillonnage**
- Application pratique en imagerie cérébrale : collaboration en cours avec Sylvain Baillet (Pitié-Salpêtrière)

(Neuro-)images. $K = 16384 \gg n$, bruit corrélé.



Modèle

Observations : $\mathbf{Y} = (Y^1, \dots, Y^n) = \begin{pmatrix} Y_1^1 & \dots & Y_1^n \\ \vdots & & \vdots \\ \vdots & & \vdots \\ Y_K^1 & \dots & Y_K^n \end{pmatrix}$

$Y^1, \dots, Y^n \in \mathbb{R}^K$ i.i.d. **symétrique**

- Moyenne inconnue $\mu = (\mu_k)_k$
- Matrice de covariance inconnue Σ
- Majoration de la variance connue (ou à estimer séparément) : $\max_k \text{var}(Y_k^1) \leq \sigma^2$
- $n \ll K$.

Buts : Région de confiance pour μ ; Test multiple : $\{k/\mu_k \neq 0\} = ?$

Modèle

Observations : $\mathbf{Y} = (Y^1, \dots, Y^n) = \begin{pmatrix} Y_1^1 & \dots & Y_1^n \\ \vdots & & \vdots \\ \vdots & & \vdots \\ Y_K^1 & \dots & Y_K^n \end{pmatrix}$

$Y^1, \dots, Y^n \in \mathbb{R}^K$ i.i.d. **symétrique**

- Moyenne inconnue $\mu = (\mu_k)_k$
- Matrice de covariance inconnue Σ
- Majoration de la variance connue (ou à estimer séparément) : $\max_k \text{var}(Y_k^1) \leq \sigma^2$
- $n \ll K$.

Buts : Région de confiance pour μ ; Test multiple : $\{k/\mu_k \neq 0\} = ?$

Régions de confiance et tests multiples

Boule de confiance $B(\mathbf{Y}) = \left\{ \mu \in \mathbb{R}^K / \sup_k |\bar{\mathbf{Y}}_k - \mu_k| \leq t \right\}$,

où

- $\bar{\mathbf{Y}} = \frac{1}{n} \sum_{i=1}^n \mathbf{Y}^i$ moyenne empirique
- $t = t_\alpha(\mathbf{Y})$ seuil (indépendant de $k \in \{1, \dots, K\}$).

\Rightarrow test multiple : on rejette “ $\mu_k = 0$ ” pour tout

$$k \in R(\mathbf{Y}) = \{k / |\bar{\mathbf{Y}}_k| > t_\alpha(\mathbf{Y})\} .$$

$$\begin{aligned} \text{FWER}(R) &= \mathbb{P}(\exists k \ / \ \mu_k = 0 \text{ and } \sqrt{n}|\bar{\mathbf{Y}}_k| > t) \\ &\leq \mathbb{P}(\exists k \ / \ \sqrt{n}|\bar{\mathbf{Y}}_k - \mu_k| > t) \\ &= \mathbb{P}\left(\|\bar{\mathbf{Y}} - \mu\|_\infty > tn^{-1/2}\right) \end{aligned}$$

\Rightarrow région de confiance et contrôle du FWER

Régions de confiance et tests multiples

Boule de confiance $B(\mathbf{Y}) = \left\{ \mu \in \mathbb{R}^K / \sup_k |\bar{\mathbf{Y}}_k - \mu_k| \leq t \right\}$,

où

- $\bar{\mathbf{Y}} = \frac{1}{n} \sum_{i=1}^n Y^i$ moyenne empirique
- $t = t_\alpha(\mathbf{Y})$ seuil (indépendant de $k \in \{1, \dots, K\}$).

\Rightarrow test multiple : on rejette “ $\mu_k = 0$ ” pour tout

$$k \in R(\mathbf{Y}) = \{k / |\bar{\mathbf{Y}}_k| > t_\alpha(\mathbf{Y})\} .$$

$$\begin{aligned} \text{FWER}(R) &= \mathbb{P}(\exists k \ / \ \mu_k = 0 \text{ and } \sqrt{n}|\bar{\mathbf{Y}}_k| > t) \\ &\leq \mathbb{P}(\exists k \ / \ \sqrt{n}|\bar{\mathbf{Y}}_k - \mu_k| > t) \\ &= \mathbb{P}\left(\|\bar{\mathbf{Y}} - \mu\|_\infty > tn^{-1/2}\right) \end{aligned}$$

\Rightarrow région de confiance et contrôle du FWER

Seuil de Bonferroni, cas Gaussien

Borne d'union :

$$\begin{aligned} \text{FWER}(R) &\leq K \sup_k \mathbb{P}(\sqrt{n}|\bar{\mathbf{Y}}_k - \mu_k| > t) \\ &\leq 2K\bar{\Phi}(t/\sigma), \end{aligned}$$

où $\bar{\Phi}$ est la queue gaussienne standard.

Seuil de Bonferroni : $t_\alpha^{\text{Bonf}} = \sigma\bar{\Phi}^{-1}(\alpha/(2K))$.

- trop conservatif s'il y a de fortes corrélations entre les coordonnées

($Y_1 = \dots = Y_K$ revient à $K = 1$)

⇒ comment faire mieux ?

But

Trouver un seuil $t_\alpha(\mathbf{Y})$ tel que

$$\forall \mu \in \mathbb{R}^K \quad \mathbb{P}(\sqrt{n}\|\bar{\mathbf{Y}} - \mu\|_\infty > t_\alpha(\mathbf{Y})) \leq \alpha .$$

\Rightarrow FWER(R) $\leq \alpha$ + boule de confiance de niveau α pour μ .

- Non-asymptotique : $\forall K, n$
- Hypothèses : Y^i symétrique, **corrélations générales**

Seuil idéal : $t = q_\alpha^*$, quantile d'ordre $(1 - \alpha)$ de $\mathcal{D}(\sqrt{n}\|\bar{\mathbf{Y}} - \mu\|_\infty)$
 q_α^* dépend de $\mathcal{D}(\mathbf{Y})$ inconnue \Rightarrow estimé par **rééchantillonnage**.

But

Trouver un seuil $t_\alpha(\mathbf{Y})$ tel que

$$\forall \mu \in \mathbb{R}^K \quad \mathbb{P}(\sqrt{n} \|\bar{\mathbf{Y}} - \mu\|_\infty > t_\alpha(\mathbf{Y})) \leq \alpha .$$

$\Rightarrow \text{FWER}(R) \leq \alpha$ + boule de confiance de niveau α pour μ .

- Non-asymptotique : $\forall K, n$
- Hypothèses : Y^i symétrique, **corrélations générales**

Seuil idéal : $t = q_\alpha^*$, quantile d'ordre $(1 - \alpha)$ de $\mathcal{D}(\sqrt{n} \|\bar{\mathbf{Y}} - \mu\|_\infty)$

q_α^* dépend de $\mathcal{D}(\mathbf{Y})$ inconnue \Rightarrow estimé par **rééchantillonnage**.

Deux méthodes pour estimer q_α^*

- **Concentration** :

$\|\bar{\mathbf{Y}} - \mu\|_\infty$ se concentre autour de son espérance (écart-type $\leq \sigma n^{-1/2}$),

on estime $\mathbb{E} [\|\bar{\mathbf{Y}} - \mu\|_\infty]$ par rééchantillonnage

$$\Rightarrow q_\alpha^{\text{conc}}(\mathbf{Y}) = \text{cst} \times \sqrt{n} \mathbb{E} [\|\bar{\mathbf{Y}}_W - \bar{W} \bar{\mathbf{Y}}\|_\infty | \mathbf{Y}] + \text{reste}(\sigma, \alpha, n)$$

- **Quantiles** :

estimation directe de q_α^* par rééchantillonnage :

$$q_\alpha^{\text{quant}}(\mathbf{Y}) = \text{quantile d'ordre } (1 - \alpha) \text{ de}$$

$$\mathcal{D} \left(\sqrt{n} \left\| n^{-1} \sum_{i=1}^n W_i (\mathbf{Y}^i - \bar{\mathbf{Y}}) \right\|_\infty | \mathbf{Y} \right)$$

Deux méthodes pour estimer q_α^*

- Concentration :

$\|\bar{\mathbf{Y}} - \mu\|_\infty$ se concentre autour de son espérance (écart-type $\leq \sigma n^{-1/2}$),

on estime $\mathbb{E} [\|\bar{\mathbf{Y}} - \mu\|_\infty]$ par rééchantillonnage

$$\Rightarrow q_\alpha^{\text{conc}}(\mathbf{Y}) = \text{cst} \times \sqrt{n} \mathbb{E} [\|\bar{\mathbf{Y}}_W - \bar{W} \bar{\mathbf{Y}}\|_\infty | \mathbf{Y}] + \text{reste}(\sigma, \alpha, n)$$

- Quantiles :

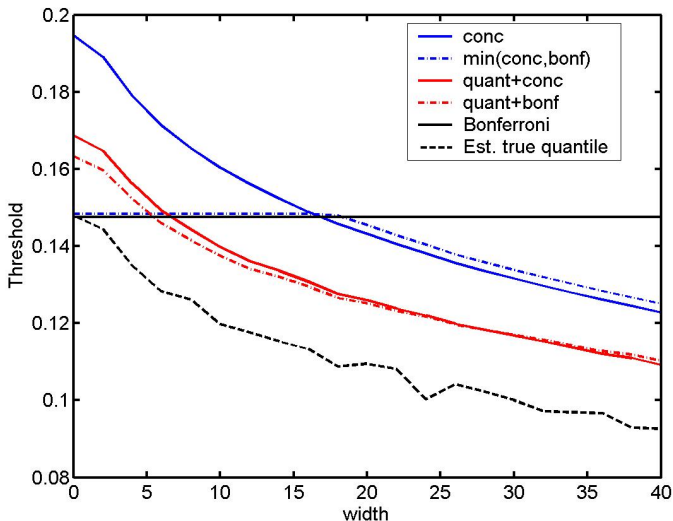
estimation directe de q_α^* par rééchantillonnage :

$$q_\alpha^{\text{quant}}(\mathbf{Y}) = \text{quantile d'ordre } (1 - \alpha) \text{ de } \mathcal{D}(\sqrt{n} \|n^{-1} \sum_{i=1}^n W_i (\mathbf{Y}^i - \bar{\mathbf{Y}})\|_\infty | \mathbf{Y})$$

Résultats théoriques

- **Concentration** : poids W échangeables généraux (ou V -fold), \mathbf{Y}^i gaussiens (ou symétriques bornés)
 ⇒ contrôle **non-asymptotique** du niveau.
- **Quantiles** : poids Rademacher i.i.d., \mathbf{Y}^i symétriques
 ⇒ besoin d'un terme supplémentaire pour contrôler le niveau
- Dans les deux cas : corrélations **générales et inconnues**, et $K \gg n$ est possible.

Simulations : $n = 1000$, $K = 16384$, $\sigma = 1$



Pourquoi rééchantillonner ?

- Une méthode très générale : pénalisation, régions de confiance, *etc.*
et pour la surpénalisation ?
- Une méthode s'adaptant naturellement :
hétéroscédasticité (en régression, lorsque $s \in \mathcal{H}(\alpha, R)$,
 $\alpha \in (0, 1]$)
corrélations (régions de confiance ; non prouvé mais observé numériquement)
vitesses rapides en classification ? (résultats préliminaires uniquement). But : faire mieux que les complexités de Rademacher locales, la validation, l'agrégation.

Pourquoi rééchantillonner ?

- Une méthode très générale : pénalisation, régions de confiance, *etc.*
et pour la surpénalisation ?
- Une méthode s'adaptant naturellement :
hétéroscédasticité (en régression, lorsque $s \in \mathcal{H}(\alpha, R)$,
 $\alpha \in (0, 1]$)
corrélations (régions de confiance ; non prouvé mais observé numériquement)
vitesses rapides en classification ? (résultats préliminaires uniquement). But : faire mieux que les complexités de Rademacher locales, la validation, l'agrégation.

Pourquoi rééchantillonner ?

- Une méthode très générale : pénalisation, régions de confiance, *etc.*
et pour la surpénalisation ?
- Une méthode s'adaptant naturellement :
hétéroscédasticité (en régression, lorsque $s \in \mathcal{H}(\alpha, R)$,
 $\alpha \in (0, 1]$)
corrélations (régions de confiance ; non prouvé mais observé
numériquement)
vitesses rapides en classification ? (résultats préliminaires
uniquement). But : faire mieux que les complexités de
Rademacher locales, la validation, l'agrégation.

Pourquoi rééchantillonner ?

- Une méthode très générale : pénalisation, régions de confiance, *etc.*
et pour la surpénalisation ?
- Une méthode s'adaptant naturellement :
hétéroscédasticité (en régression, lorsque $s \in \mathcal{H}(\alpha, R)$,
 $\alpha \in (0, 1]$)
corrélations (régions de confiance ; non prouvé mais observé
numériquement)
vitesses rapides en classification ? (résultats préliminaires
uniquement). But : faire mieux que les complexités de
Rademacher locales, la validation, l'agrégation.

Comment rééchantillonner ?

- Rééchantillonnages échangeables : résultats précédents essentiellement asymptotiques (Barbe et Bertail 1995).

en sélection de modèles : bootstrap “ m out of n ” consistant

$\Leftrightarrow m \ll n$ (Shao 1996)

\Rightarrow Consistance pour l'identification avec des poids échangeables quelconques ?

résultats non-asymptotiques, histogrammes : poids échangeables un peu meilleurs que V -fold ; le bootstrap semble sous-pénaliser légèrement.

- Validation-croisée V -fold : choix de V , pas si simple (cf. Celisse et Robin 2007)

Question ouverte : tenir compte de la variabilité des estimateurs par rééchantillonnage ?

Comment rééchantillonner ?

- Rééchantillonnages échangeables : résultats précédents essentiellement asymptotiques (Barbe et Bertail 1995).

en **sélection de modèles** : bootstrap “ m out of n ” consistant

$\Leftrightarrow m \ll n$ (Shao 1996)

\Rightarrow **Consistance pour l'identification avec des poids échangeables quelconques ?**

résultats non-asymptotiques, histogrammes : poids échangeables un peu meilleurs que V -fold ; le bootstrap semble sous-pénaliser légèrement.

- Validation-croisée V -fold : choix de V , pas si simple (cf. Celisse et Robin 2007)

Question ouverte : tenir compte de la variabilité des estimateurs par rééchantillonnage ?

Comment rééchantillonner ?

- Rééchantillonnages échangeables : résultats précédents essentiellement asymptotiques (Barbe et Bertail 1995).

en sélection de modèles : bootstrap “ m out of n ” consistant

$\Leftrightarrow m \ll n$ (Shao 1996)

\Rightarrow Consistance pour l'identification avec des poids échangeables quelconques ?

résultats non-asymptotiques, histogrammes : poids échangeables un peu meilleurs que V -fold ; le bootstrap semble sous-pénaliser légèrement.

- Validation-croisée V -fold : choix de V , pas si simple (cf. Celisse et Robin 2007)

Question ouverte : tenir compte de la variabilité des estimateurs par rééchantillonnage ?

Comment rééchantillonner ?

- Rééchantillonnages échangeables : résultats précédents essentiellement asymptotiques (Barbe et Bertail 1995).

en sélection de modèles : bootstrap “ m out of n ” consistant

$\Leftrightarrow m \ll n$ (Shao 1996)

\Rightarrow Consistance pour l'identification avec des poids échangeables quelconques ?

résultats non-asymptotiques, histogrammes : poids échangeables un peu meilleurs que V -fold ; le bootstrap semble sous-pénaliser légèrement.

- **Validation-croisée V -fold** : choix de V , pas si simple (cf. Celisse et Robin 2007)

Question ouverte : tenir compte de la variabilité des estimateurs par rééchantillonnage ?

Comment rééchantillonner ?

- Rééchantillonnages échangeables : résultats précédents essentiellement asymptotiques (Barbe et Bertail 1995).

en sélection de modèles : bootstrap “ m out of n ” consistant

$\Leftrightarrow m \ll n$ (Shao 1996)

\Rightarrow Consistance pour l'identification avec des poids échangeables quelconques ?

résultats non-asymptotiques, histogrammes : poids échangeables un peu meilleurs que V -fold ; le bootstrap semble sous-pénaliser légèrement.

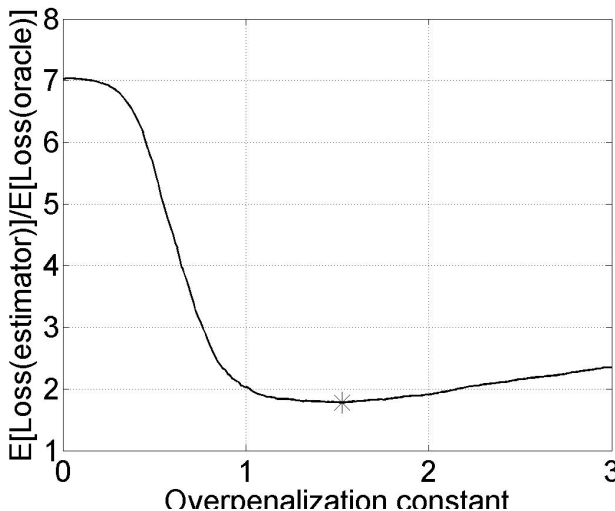
- Validation-croisée V -fold : choix de V , pas si simple (cf. Celisse et Robin 2007)

Question ouverte : tenir compte de la **variabilité** des estimateurs par rééchantillonnage ?

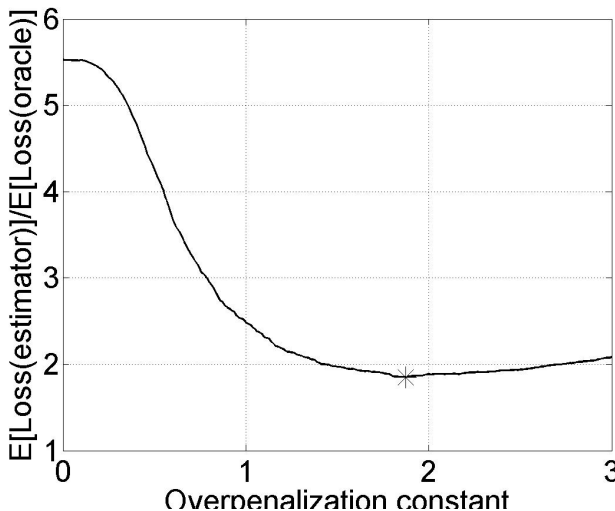
Annexes

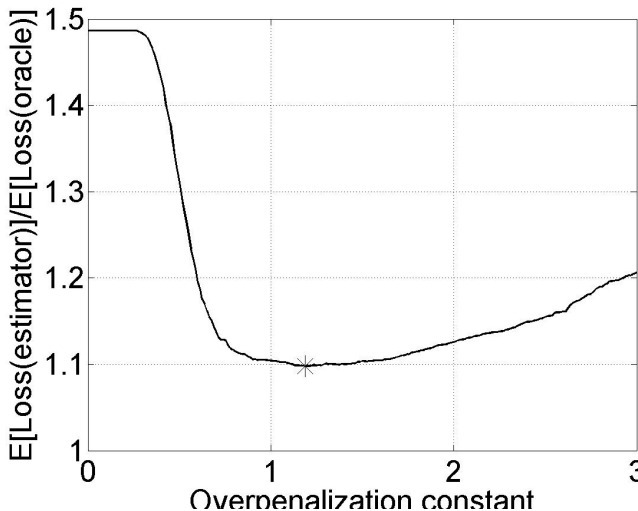
- 3 Surpénalisation
- 4 Heuristique de pente
- 5 Limites d'une pénalité linéaire
- 6 Pénalités par rééchantillonnage
- 7 Simulations
 - Validation-croisée V-fold
 - Pénalisation V-fold
 - Pénalisation par rééchantillonnage échangeable
- 8 Régions de confiance et tests multiples
 - Théorie
 - Régions de confiance
 - Simulations

Surpénalisation ($s = \sin$, $\sigma \equiv 1$, $n = 200$, Mallows)



Surpénalisation ($s = \sin$, $\sigma(x) = x$, $n = 200$, pen Rademacher)



Surpénalisation (HeaviSine, $n = 2048$, $\sigma(x) = x$, pen 10-f)

Premier pas vers l'oracle

Pénalité idéale : $\text{pen}_{\text{id}}(m) = (P - P_n)(\gamma(\hat{s}_m, \cdot))$

Pour tout $m \in \mathcal{M}$,

$$P_n \gamma(\hat{s}_{\hat{m}}, \cdot) + \text{pen}(\hat{m}) \leq P_n \gamma(\hat{s}_m, \cdot) + \text{pen}(m)$$

$$l(s, \hat{s}_{\hat{m}}) + [\text{pen}(\hat{m}) - \text{pen}_{\text{id}}(\hat{m})] \leq l(s, \hat{s}_m) + [\text{pen}(m) - \text{pen}_{\text{id}}(m)]$$

\Rightarrow il suffit d'avoir $\text{pen}(m') \geq \text{pen}_{\text{id}}(m') = (P - P_n)\gamma(\hat{s}_{m'}, \cdot)$ pour tout $m' \in \mathcal{M}$

Premier pas vers l'oracle

Pénalité idéale : $\text{pen}_{\text{id}}(m) = (P - P_n)(\gamma(\hat{s}_m, \cdot))$

Pour tout $m \in \mathcal{M}$,

$$P_n \gamma(\hat{s}_{\hat{m}}, \cdot) + \text{pen}(\hat{m}) \leq P_n \gamma(\hat{s}_m, \cdot) + \text{pen}(m)$$

$$l(s, \hat{s}_{\hat{m}}) + [\text{pen}(\hat{m}) - \text{pen}_{\text{id}}(\hat{m})] \leq l(s, \hat{s}_m) + [\text{pen}(m) - \text{pen}_{\text{id}}(m)]$$

\Rightarrow il suffit d'avoir $\text{pen}(m') \geq \text{pen}_{\text{id}}(m') = (P - P_n)\gamma(\hat{s}_{m'}, \cdot)$ pour tout $m' \in \mathcal{M}$

Premier pas vers l'oracle

$$\text{Pénalité idéale : } \text{pen}_{\text{id}}(m) = (P - P_n)(\gamma(\widehat{s}_m, \cdot))$$

Pour tout $m \in \mathcal{M}$,

$$P_n \gamma(\widehat{s}_{\widehat{m}}, \cdot) + \text{pen}(\widehat{m}) \leq P_n \gamma(\widehat{s}_m, \cdot) + \text{pen}(m)$$

$$l(s, \widehat{s}_{\widehat{m}}) + [\text{pen}(\widehat{m}) - \text{pen}_{\text{id}}(\widehat{m})] \leq l(s, \widehat{s}_m) + [\text{pen}(m) - \text{pen}_{\text{id}}(m)]$$

\Rightarrow il suffit d'avoir $\text{pen}(m') \geq \text{pen}_{\text{id}}(m') = (P - P_n)\gamma(\widehat{s}_{m'}, \cdot)$ pour tout $m' \in \mathcal{M}$

Bienfaits de la surpénalisation

- n petit, σ grand \Rightarrow surpénaliser peut être bénéfique (simulations)
- Optimalité asymptotique \neq Optimalité non-asymptotique
- $\text{pen}(m) \geq \text{pen}_{\text{id}}(m)$ doit être assuré avec grande probabilité pour tous les modèles
(plus important que $\text{pen}(m) \leq (1 + \epsilon) \text{pen}_{\text{id}}(m)$)

\Rightarrow Explique pourquoi $V = 2$ peut être "optimal"

\Rightarrow Facile de surpénaliser : $\text{pen}(m)$ estime $C \text{pen}_{\text{id}}(m)$ avec $C > 1$

Bienfaits de la surpénalisation

- n petit, σ grand \Rightarrow surpénaliser peut être bénéfique (simulations)
- **Optimalité asymptotique \neq Optimalité non-asymptotique**
- $\text{pen}(m) \geq \text{pen}_{\text{id}}(m)$ doit être assuré avec grande probabilité pour tous les modèles
(plus important que $\text{pen}(m) \leq (1 + \epsilon) \text{pen}_{\text{id}}(m)$)

\Rightarrow Explique pourquoi $V = 2$ peut être "optimal"

\Rightarrow Facile de surpénaliser : $\text{pen}(m)$ estime $C \text{pen}_{\text{id}}(m)$ avec $C > 1$

Bienfaits de la surpénalisation

- n petit, σ grand \Rightarrow surpénaliser peut être bénéfique (simulations)
- Optimalité asymptotique \neq Optimalité non-asymptotique
- $\text{pen}(m) \geq \text{pen}_{\text{id}}(m)$ doit être assuré avec grande probabilité pour tous les modèles (plus important que $\text{pen}(m) \leq (1 + \epsilon) \text{pen}_{\text{id}}(m)$)

\Rightarrow Explique pourquoi $V = 2$ peut être "optimal"

\Rightarrow Facile de surpénaliser : $\text{pen}(m)$ estime $C \text{pen}_{\text{id}}(m)$ avec $C > 1$

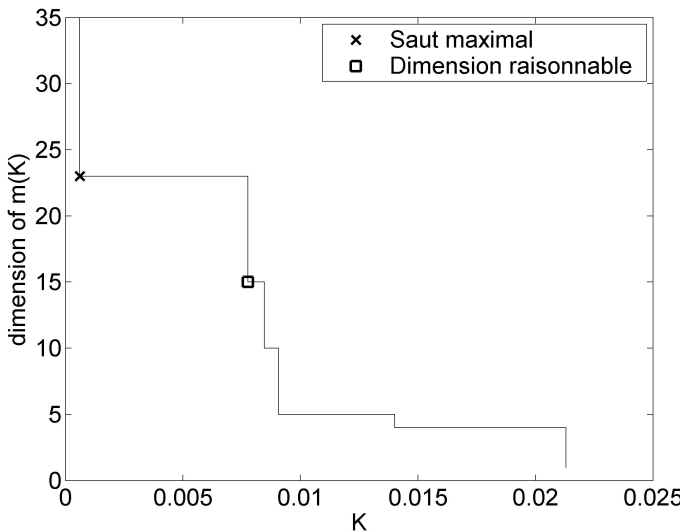
Bienfaits de la surpénalisation

- n petit, σ grand \Rightarrow surpénaliser peut être bénéfique (simulations)
- Optimalité asymptotique \neq Optimalité non-asymptotique
- $\text{pen}(m) \geq \text{pen}_{\text{id}}(m)$ doit être assuré avec grande probabilité pour tous les modèles
(plus important que $\text{pen}(m) \leq (1 + \epsilon) \text{pen}_{\text{id}}(m)$)

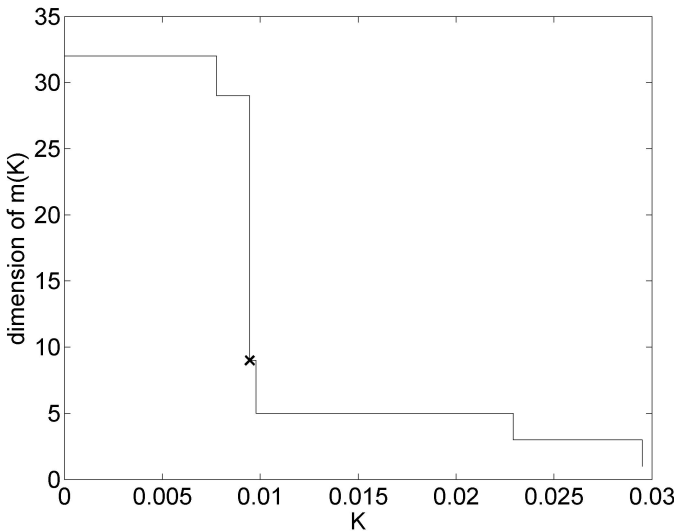
\Rightarrow Explique pourquoi $V = 2$ peut être "optimal"

\Rightarrow Facile de surpénaliser : $\text{pen}(m)$ estime $C \text{pen}_{\text{id}}(m)$ avec $C > 1$

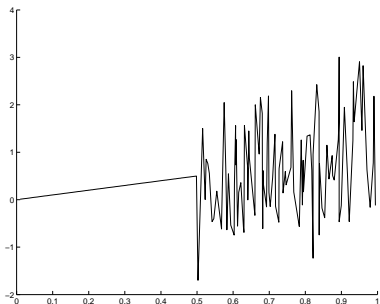
Calibration asymptotique : heuristique de pente



Calibration asymptotique : heuristique de pente



Limites d'une pénalité linéaire



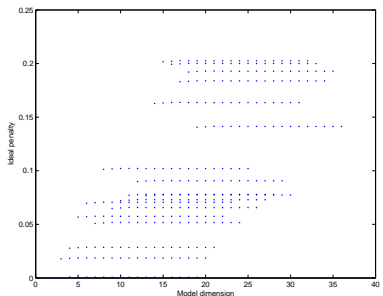
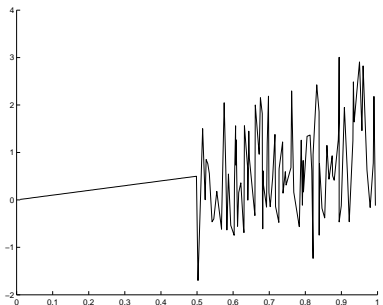
$$Y = X + \sigma(X)\epsilon$$

$$\sigma(X) = \mathbb{1}_{X \geq \frac{1}{2}} \quad \epsilon \sim \mathcal{N}(0, 1)$$

Histogrammes réguliers sur
 $[0; \frac{1}{2}]$ ($D_{m,1}$ morceaux) puis
 sur $[\frac{1}{2}; 1]$ ($D_{m,2}$ morceaux).

$\Rightarrow \text{pen}_{\text{id}}(m)$ n'est pas linéaire
 en D_m .

Limites d'une pénalité linéaire



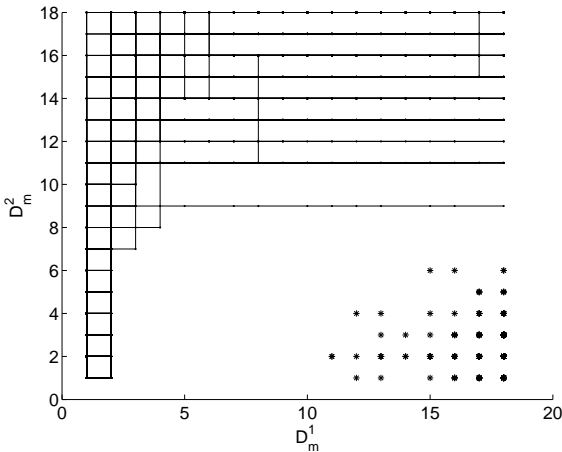
$$Y = X + \sigma(X)\epsilon$$

$$\sigma(X) = \mathbb{1}_{X \geq \frac{1}{2}} \quad \epsilon \sim \mathcal{N}(0, 1)$$

Histogrammes réguliers sur $[0; \frac{1}{2}]$ ($D_{m,1}$ morceaux) puis sur $[\frac{1}{2}; 1]$ ($D_{m,2}$ morceaux).

$\Rightarrow \text{pen}_{\text{id}}(m)$ n'est pas linéaire en D_m .

Limites d'une pénalité linéaire : $\hat{m}(K) \neq m^*$



Inégalité-oracle “trajectorielle” non-asymptotique

- Régression sur des histogrammes, plan d'expérience aléatoire
- Petit nombre de modèles (au plus $\diamond n^\alpha$)
- Présélection de modèles :

$$\forall m \in \mathcal{M}, \quad \min_{\lambda \in \Lambda_m} \{\text{Card} \{X_i \in I_\lambda\}\} \geq 2$$

- V fixe ou $V = n$ (et plusieurs rééchantillonnages échangeables, dont le bootstrap)
- $C \approx V - 1$ (ou $C \approx C_{W, \infty}$)

Théorème

Sous des hypothèses “raisonnables” sur P , avec probabilité au moins $1 - \diamond n^{-2}$,

$$I(s, \widehat{s}_{\widehat{m}}) \leq \left(1 + \ln(n)^{-1/5}\right) \inf_{m \in \mathcal{M}} \{I(s, \widehat{s}_m)\}$$

Hypothèses "raisonnables" sur P

- Données bornées : $\|Y\|_\infty \leq A < \infty$
- Niveau de bruit minoré :

$$0 < \sigma_{\min} \leq \sigma(X)$$

- Cible s régulière (α -hölderienne) non-constante
- Régularité de la partition : $\min_\lambda \mathbb{P}(X \in I_\lambda) \geq \diamond D_m^{-1}$

et ce n'est qu'un exemple de jeu d'hypothèses valide...

Hypothèses "raisonnables" sur P

- Données bornées : $\|Y\|_\infty \leq A < \infty$
- Niveau de bruit minoré :

$$0 < \sigma_{\min} \leq \sigma(X)$$

- Cible s régulière (α -hölderienne) non-constante
- Régularité de la partition : $\min_\lambda \mathbb{P}(X \in I_\lambda) \geq \diamond D_m^{-1}$

et ce n'est qu'un exemple de jeu d'hypothèses valide...

Simulations : cadre

$$Y_i = s(X_i) + \sigma(X_i)\epsilon_i$$

$$(X_i)_{1 \leq i \leq n} \sim^{\text{i.i.d.}} \mathcal{U}([0; 1]) \quad (\epsilon_i)_{1 \leq i \leq n} \sim^{\text{i.i.d.}} \mathcal{N}(0, 1)$$

$$\mathcal{M}_n = \left\{ \text{Histogrammes réguliers à } D \text{ morceaux, } 1 \leq D \leq \frac{n}{\log(n)} \right. \\ \left. \text{et tels que } \min_{\lambda \in \Lambda_m} \text{Card}\{X_i \in I_\lambda\} \geq 2 \right\}$$

⇒ Qualité des procédures :

$$C_{\text{classique}} = \frac{\mathbb{E} [l(s, \hat{s}_{\hat{m}})]}{\mathbb{E} [\inf_{m \in \mathcal{M}} l(s, \hat{s}_m)]}$$

calculé avec $N = 1000$ échantillons

Simulations : cadre

$$Y_i = s(X_i) + \sigma(X_i)\epsilon_i$$

$$(X_i)_{1 \leq i \leq n} \sim^{\text{i.i.d.}} \mathcal{U}([0; 1]) \quad (\epsilon_i)_{1 \leq i \leq n} \sim^{\text{i.i.d.}} \mathcal{N}(0, 1)$$

$$\mathcal{M}_n = \left\{ \text{Histogrammes réguliers à } D \text{ morceaux, } 1 \leq D \leq \frac{n}{\log(n)} \right. \\ \left. \text{et tels que } \min_{\lambda \in \Lambda_m} \text{Card}\{X_i \in I_\lambda\} \geq 2 \right\}$$

⇒ Qualité des procédures :

$$C_{\text{classique}} = \frac{\mathbb{E} [l(s, \hat{s}_{\hat{m}})]}{\mathbb{E} [\inf_{m \in \mathcal{M}} l(s, \hat{s}_m)]}$$

calculé avec $N = 1000$ échantillons

Simulations : cadre

$$Y_i = s(X_i) + \sigma(X_i)\epsilon_i$$

$$(X_i)_{1 \leq i \leq n} \sim^{\text{i.i.d.}} \mathcal{U}([0; 1]) \quad (\epsilon_i)_{1 \leq i \leq n} \sim^{\text{i.i.d.}} \mathcal{N}(0, 1)$$

$$\mathcal{M}_n = \left\{ \begin{array}{l} \text{Histogrammes réguliers à } D \text{ morceaux, } 1 \leq D \leq \frac{n}{\log(n)} \\ \text{et tels que } \min_{\lambda \in \Lambda_m} \text{Card}\{X_i \in I_\lambda\} \geq 2 \end{array} \right\}$$

⇒ Qualité des procédures :

$$C_{\text{classique}} = \frac{\mathbb{E} [l(s, \hat{s}_{\hat{m}})]}{\mathbb{E} [\inf_{m \in \mathcal{M}} l(s, \hat{s}_m)]}$$

calculé avec $N = 1000$ échantillons

Simulations : procédures considérées

- V -fold "classique" ($V \in \{2, 5, 10, 20\}$) :

$$\hat{m} \in \arg \min_{m \in \mathcal{M}} \left\{ \frac{1}{V} \sum_{j=1}^V P_n^j \gamma \left(\hat{s}_m^{(-j)} \right) \right\} \quad \tilde{s} = \hat{s}_{\hat{m}}$$

- Leave-one-out = n -fold "classique"

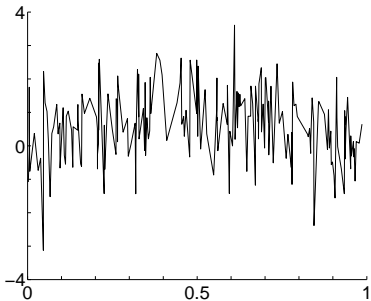
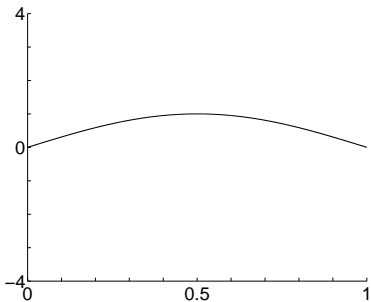
Simulations : procédures considérées

- V -fold “classique” ($V \in \{2, 5, 10, 20\}$) :

$$\hat{m} \in \arg \min_{m \in \mathcal{M}} \left\{ \frac{1}{V} \sum_{j=1}^V P_n^j \gamma \left(\hat{s}_m^{(-j)} \right) \right\} \quad \tilde{s} = \hat{s}_{\hat{m}}$$

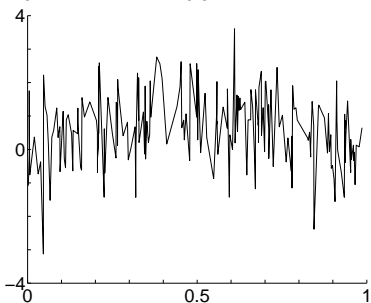
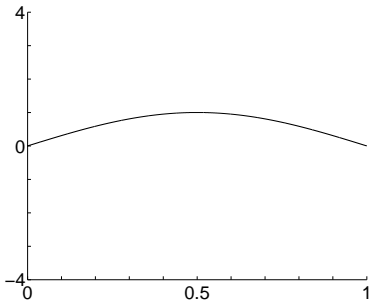
- Leave-one-out = n -fold “classique”

Choix de V pour VFCV : sin, $n = 200$, $\sigma \equiv 1$



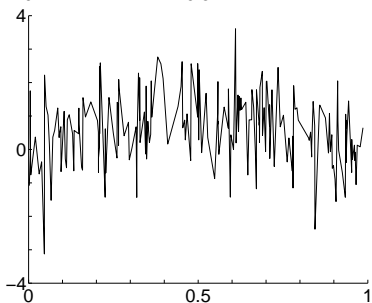
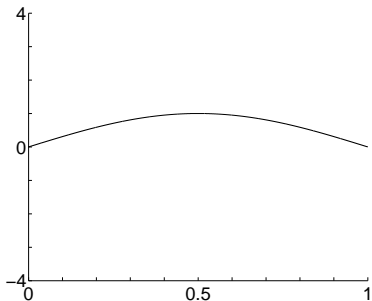
2-fold	2.08 ± 0.04
5-fold	2.14 ± 0.04
10-fold	2.10 ± 0.05
20-fold	2.09 ± 0.04
leave-one-out	2.08 ± 0.04

Simulations : \sin , $n = 200$, $\sigma \equiv 1$



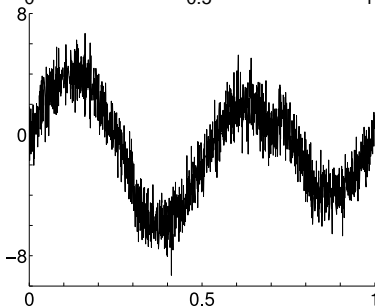
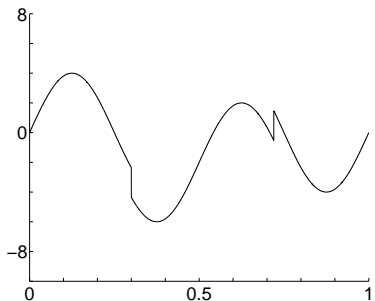
Mallows	1.93 ± 0.04
2-fold	2.08 ± 0.04
5-fold	2.14 ± 0.04
10-fold	2.10 ± 0.05
leave-one-out	2.08 ± 0.04
<hr/>	
pen 2-f	2.58 ± 0.06
pen 5-f	2.22 ± 0.05
pen 10-f	2.12 ± 0.05
pen 20-f	2.09 ± 0.04
pen Loo	2.08 ± 0.05
<hr/>	
Mallows $\times 1.25$	1.80 ± 0.03
pen 2-f $\times 1.25$	2.17 ± 0.05
pen 5-f $\times 1.25$	1.91 ± 0.05
pen 10-f $\times 1.25$	1.87 ± 0.03
pen 20-f $\times 1.25$	1.90 ± 0.05
pen Loo $\times 1.25$	1.84 ± 0.03

Simulations : \sin , $n = 200$, $\sigma \equiv 1$



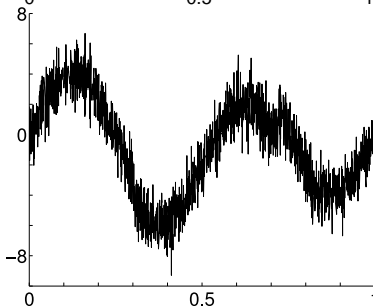
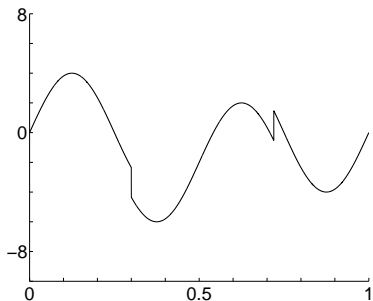
Mallows	1.93 ± 0.04
2-fold	2.08 ± 0.04
5-fold	2.14 ± 0.04
10-fold	2.10 ± 0.05
leave-one-out	2.08 ± 0.04
<hr/>	
pen 2-f	2.58 ± 0.06
pen 5-f	2.22 ± 0.05
pen 10-f	2.12 ± 0.05
pen 20-f	2.09 ± 0.04
pen Loo	2.08 ± 0.05
<hr/>	
Mallows $\times 1.25$	1.80 ± 0.03
pen 2-f $\times 1.25$	2.17 ± 0.05
pen 5-f $\times 1.25$	1.91 ± 0.05
pen 10-f $\times 1.25$	1.87 ± 0.03
pen 20-f $\times 1.25$	1.90 ± 0.05
pen Loo $\times 1.25$	1.84 ± 0.03

Simulations : HeaviSine, $n = 2048$, $\sigma \equiv 1$



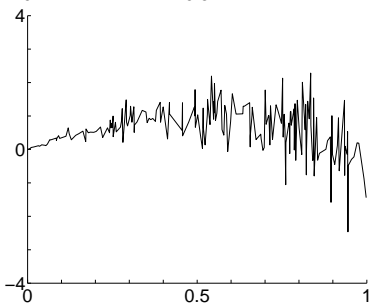
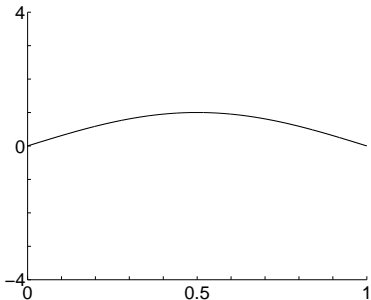
Mallows	1.606 ± 0.015
2-fold	1.002 ± 0.003
5-fold	1.014 ± 0.003
10-fold	1.021 ± 0.003
leave-one-out	1.034 ± 0.004
pen 2-f	1.038 ± 0.004
pen 5-f	1.037 ± 0.004
pen 10-f	1.034 ± 0.004
pen 20-f	1.034 ± 0.004
pen Loo	1.034 ± 0.004
Mallows $\times 1.25$	1.606 ± 0.015
pen 2-f $\times 1.25$	1.011 ± 0.003
pen 5-f $\times 1.25$	1.006 ± 0.003
pen 10-f $\times 1.25$	1.005 ± 0.003
pen 20-f $\times 1.25$	1.004 ± 0.003
pen Loo $\times 1.25$	1.004 ± 0.003

Simulations : HeaviSine, $n = 2048$, $\sigma \equiv 1$



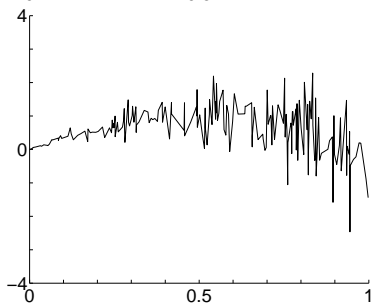
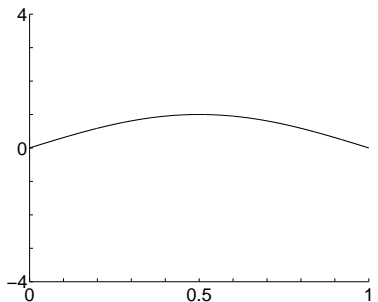
Mallows	1.606 ± 0.015
2-fold	1.002 ± 0.003
5-fold	1.014 ± 0.003
10-fold	1.021 ± 0.003
leave-one-out	1.034 ± 0.004
pen 2-f	1.038 ± 0.004
pen 5-f	1.037 ± 0.004
pen 10-f	1.034 ± 0.004
pen 20-f	1.034 ± 0.004
pen Loo	1.034 ± 0.004
Mallows $\times 1.25$	1.606 ± 0.015
pen 2-f $\times 1.25$	1.011 ± 0.003
pen 5-f $\times 1.25$	1.006 ± 0.003
pen 10-f $\times 1.25$	1.005 ± 0.003
pen 20-f $\times 1.25$	1.004 ± 0.003
pen Loo $\times 1.25$	1.004 ± 0.003

Simulations : \sin , $n = 200$, $\sigma(x) = x$, 2 pas



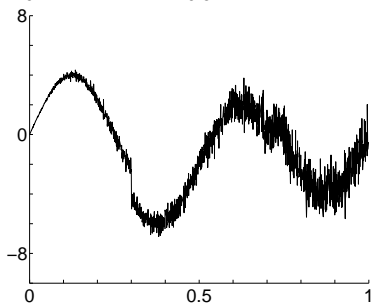
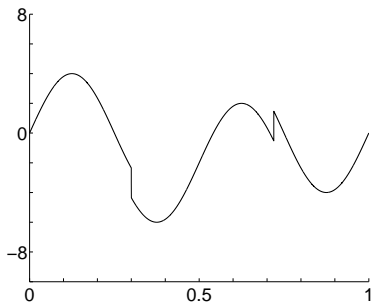
Mallows	3.86 ± 0.02
2-fold	2.54 ± 0.05
5-fold	2.58 ± 0.06
10-fold	2.60 ± 0.06
leave-one-out	2.59 ± 0.06
<hr/>	
pen 2-f	3.06 ± 0.07
pen 5-f	2.75 ± 0.06
pen 10-f	2.65 ± 0.06
pen 20-f	2.64 ± 0.06
pen Loo	2.59 ± 0.06
<hr/>	
Mallows $\times 1.25$	4.05 ± 0.02
pen 2-f $\times 1.25$	2.75 ± 0.06
pen 5-f $\times 1.25$	2.38 ± 0.06
pen 10-f $\times 1.25$	2.28 ± 0.05
pen 20-f $\times 1.25$	2.25 ± 0.06
pen Loo $\times 1.25$	2.21 ± 0.05

Simulations : \sin , $n = 200$, $\sigma(x) = x$, 2 pas



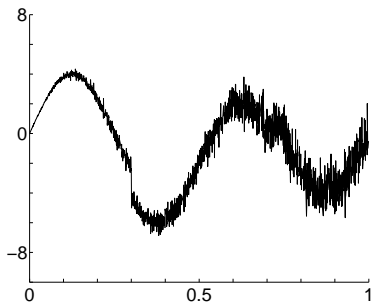
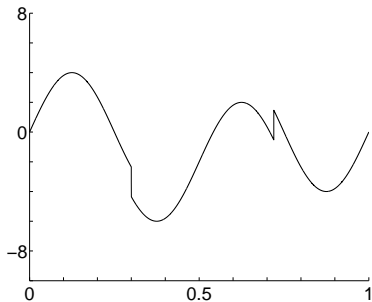
Mallows	3.86 ± 0.02
2-fold	2.54 ± 0.05
5-fold	2.58 ± 0.06
10-fold	2.60 ± 0.06
leave-one-out	2.59 ± 0.06
<hr/>	
pen 2-f	3.06 ± 0.07
pen 5-f	2.75 ± 0.06
pen 10-f	2.65 ± 0.06
pen 20-f	2.64 ± 0.06
pen Loo	2.59 ± 0.06
<hr/>	
Mallows $\times 1.25$	4.05 ± 0.02
pen 2-f $\times 1.25$	2.75 ± 0.06
pen 5-f $\times 1.25$	2.38 ± 0.06
pen 10-f $\times 1.25$	2.28 ± 0.05
pen 20-f $\times 1.25$	2.25 ± 0.06
pen Loo $\times 1.25$	2.21 ± 0.05

Simulations : HeaviSine, $n = 2048$, $\sigma(x) = x$, 2 pas



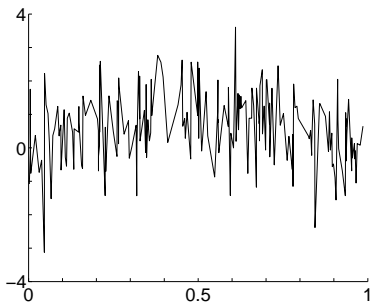
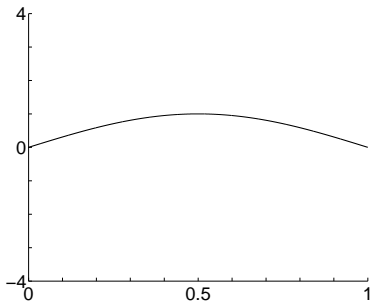
Mallows	1.487 ± 0.011
2-fold	1.184 ± 0.004
5-fold	1.115 ± 0.005
10-fold	1.109 ± 0.004
leave-one-out	1.105 ± 0.004
pen 2-f	1.103 ± 0.005
pen 5-f	1.104 ± 0.004
pen 10-f	1.104 ± 0.004
pen 20-f	1.105 ± 0.004
pen Loo	1.105 ± 0.004
Mallows $\times 1.25$	1.487 ± 0.011
pen 2-f $\times 1.25$	1.106 ± 0.004
pen 5-f $\times 1.25$	1.102 ± 0.004
pen 10-f $\times 1.25$	1.098 ± 0.004
pen 20-f $\times 1.25$	1.098 ± 0.004
pen Loo $\times 1.25$	1.096 ± 0.004

Simulations : HeaviSine, $n = 2048$, $\sigma(x) = x$, 2 pas



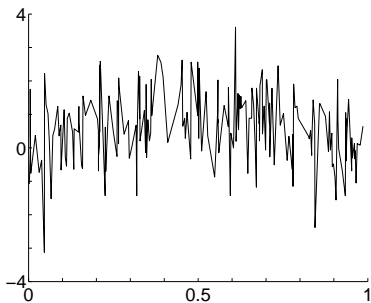
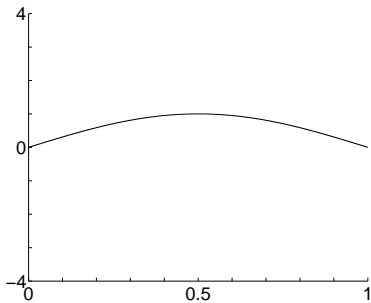
Mallows	1.487 ± 0.011
2-fold	1.184 ± 0.004
5-fold	1.115 ± 0.005
10-fold	1.109 ± 0.004
leave-one-out	1.105 ± 0.004
pen 2-f	1.103 ± 0.005
pen 5-f	1.104 ± 0.004
pen 10-f	1.104 ± 0.004
pen 20-f	1.105 ± 0.004
pen Loo	1.105 ± 0.004
Mallows $\times 1.25$	1.487 ± 0.011
pen 2-f $\times 1.25$	1.106 ± 0.004
pen 5-f $\times 1.25$	1.102 ± 0.004
pen 10-f $\times 1.25$	1.098 ± 0.004
pen 20-f $\times 1.25$	1.098 ± 0.004
pen Loo $\times 1.25$	1.096 ± 0.004

Simulations : \sin , $n = 200$, $\sigma \equiv 1$



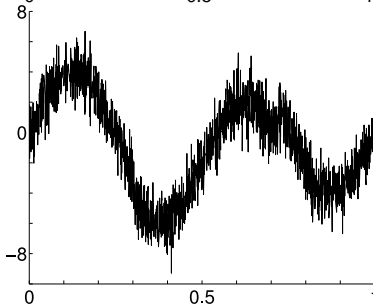
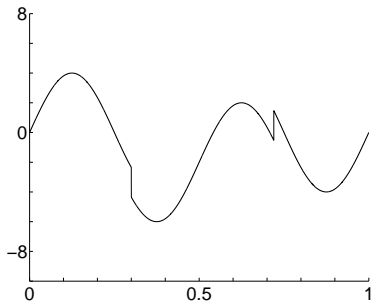
Mallows	1.93 ± 0.04
2-fold	2.08 ± 0.04
10-fold	2.10 ± 0.05
pen Efr	2.60 ± 0.07
pen Val al	1.98 ± 0.04
pen Loo	2.08 ± 0.05
pen 2-f	2.58 ± 0.06
pen 10-f	2.12 ± 0.05
Mallows × 1.25	1.80 ± 0.03
pen Efr × 1.25	2.02 ± 0.05
pen Val al × 1.25	1.80 ± 0.03
pen Loo × 1.25	1.84 ± 0.03
pen 2-f × 1.25	2.17 ± 0.05
pen 10-f × 1.25	1.87 ± 0.03

Simulations : \sin , $n = 200$, $\sigma \equiv 1$



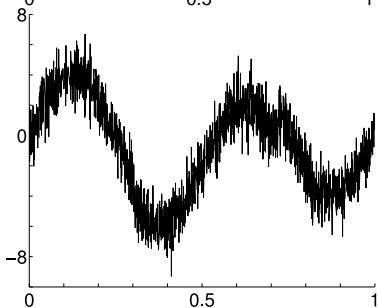
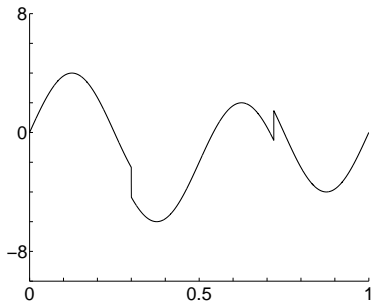
Mallows	1.93 ± 0.04
2-fold	2.08 ± 0.04
10-fold	2.10 ± 0.05
pen Efr	2.60 ± 0.07
pen Val al	1.98 ± 0.04
pen Loo	2.08 ± 0.05
pen 2-f	2.58 ± 0.06
pen 10-f	2.12 ± 0.05
Mallows $\times 1.25$	1.80 ± 0.03
pen Efr $\times 1.25$	2.02 ± 0.05
pen Val al $\times 1.25$	1.80 ± 0.03
pen Loo $\times 1.25$	1.84 ± 0.03
pen 2-f $\times 1.25$	2.17 ± 0.05
pen 10-f $\times 1.25$	1.87 ± 0.03

Simulations : HeaviSine, $n = 2048$, $\sigma \equiv 1$



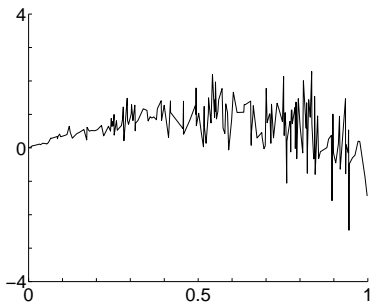
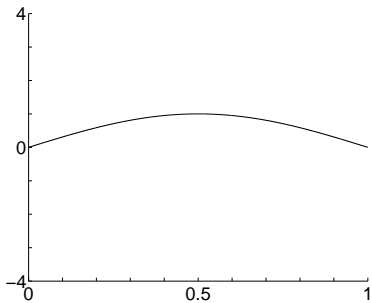
Mallows	1.606 ± 0.015
2-fold	1.002 ± 0.003
10-fold	1.021 ± 0.003
pen Efr	1.067 ± 0.005
pen Val al	1.018 ± 0.003
pen Loo	1.034 ± 0.004
pen 10-f	1.034 ± 0.004
Mallows $\times 1.25$	1.606 ± 0.015
pen Efr $\times 1.25$	1.011 ± 0.003
pen Val al $\times 1.25$	1.002 ± 0.003
pen Loo $\times 1.25$	1.004 ± 0.003
pen 10-f $\times 1.25$	1.005 ± 0.003

Simulations : HeaviSine, $n = 2048$, $\sigma \equiv 1$



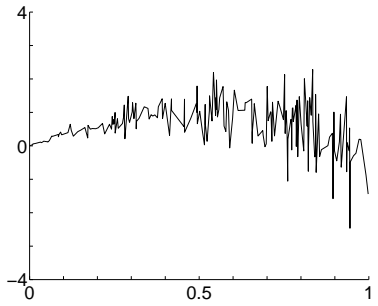
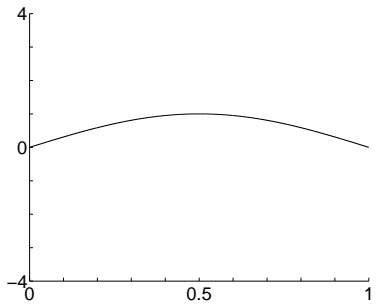
Mallows	1.606 ± 0.015
2-fold	1.002 ± 0.003
10-fold	1.021 ± 0.003
pen Efr	1.067 ± 0.005
pen Val al	1.018 ± 0.003
pen Loo	1.034 ± 0.004
pen 10-f	1.034 ± 0.004
Mallows $\times 1.25$	1.606 ± 0.015
pen Efr $\times 1.25$	1.011 ± 0.003
pen Val al $\times 1.25$	1.002 ± 0.003
pen Loo $\times 1.25$	1.004 ± 0.003
pen 10-f $\times 1.25$	1.005 ± 0.003

Simulations : \sin , $n = 200$, $\sigma(x) = x$, 2 pas



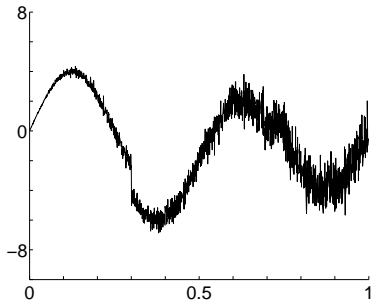
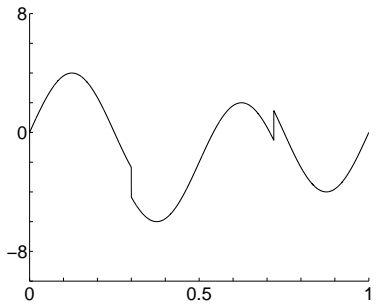
Mallows	3.86 ± 0.02
2-fold	2.54 ± 0.05
10-fold	2.60 ± 0.06
pen Efr	3.15 ± 0.07
pen Val al	2.50 ± 0.06
pen Loo	2.59 ± 0.06
pen 10-f	2.65 ± 0.06
Mallows $\times 1.25$	4.05 ± 0.02
pen Efr $\times 1.25$	2.61 ± 0.06
pen Val al $\times 1.25$	2.14 ± 0.05
pen Loo $\times 1.25$	2.21 ± 0.05
pen 10-f $\times 1.25$	2.25 ± 0.05

Simulations : \sin , $n = 200$, $\sigma(x) = x$, 2 pas



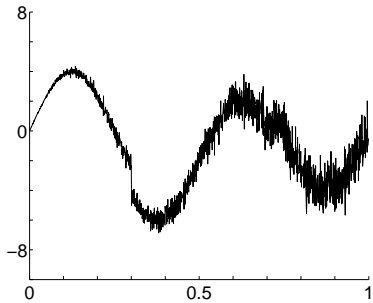
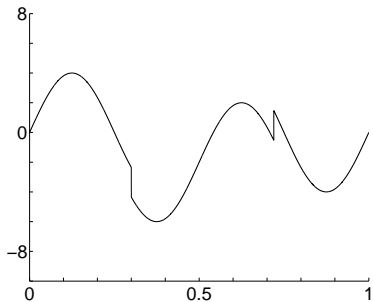
Mallows	3.86 ± 0.02
2-fold	2.54 ± 0.05
10-fold	2.60 ± 0.06
pen Efr	3.15 ± 0.07
pen Val al	2.50 ± 0.06
pen Loo	2.59 ± 0.06
pen 10-f	2.65 ± 0.06
Mallows $\times 1.25$	4.05 ± 0.02
pen Efr $\times 1.25$	2.61 ± 0.06
pen Val al $\times 1.25$	2.14 ± 0.05
pen Loo $\times 1.25$	2.21 ± 0.05
pen 10-f $\times 1.25$	2.25 ± 0.05

Simulations : HeaviSine, $n = 2048$, $\sigma(x) = x$, 2 pas



Mallows	1.487 ± 0.011
2-fold	1.184 ± 0.004
10-fold	1.109 ± 0.004
pen Efr	1.114 ± 0.005
pen Val al	1.103 ± 0.004
pen Loo	1.105 ± 0.004
pen 10-f	1.104 ± 0.004
Mallows $\times 1.25$	1.487 ± 0.011
pen Efr $\times 1.25$	1.097 ± 0.004
pen Val al $\times 1.25$	1.095 ± 0.004
pen Loo $\times 1.25$	1.096 ± 0.004
pen 10-f $\times 1.25$	1.098 ± 0.004

Simulations : HeaviSine, $n = 2048$, $\sigma(x) = x$, 2 pas



Mallows	1.487 ± 0.011
2-fold	1.184 ± 0.004
10-fold	1.109 ± 0.004
pen Efr	1.114 ± 0.005
pen Val al	1.103 ± 0.004
pen Loo	1.105 ± 0.004
pen 10-f	1.104 ± 0.004
Mallows $\times 1.25$	1.487 ± 0.011
pen Efr $\times 1.25$	1.097 ± 0.004
pen Val al $\times 1.25$	1.095 ± 0.004
pen Loo $\times 1.25$	1.096 ± 0.004
pen 10-f $\times 1.25$	1.098 ± 0.004

Théorème de concentration

Theorem

Y^i Gaussiennes. W Rademacher. Pour tout $\alpha \in (0; 1)$,

$$q_\alpha^{\text{conc},1}(\mathbf{Y}) := \frac{\sqrt{n} \mathbb{E} [\|\bar{\mathbf{Y}}_W - \bar{W} \bar{\mathbf{Y}}\|_\infty | \mathbf{Y}]}{B_W} + \sigma \bar{\Phi}^{-1}(\alpha/2) \left[\frac{C_W}{\sqrt{n} B_W} + 1 \right]$$

vérifie

$$\mathbb{P}(\sqrt{n} \|\bar{\mathbf{Y}} - \mu\|_\infty > q_\alpha^{\text{conc},1}(\mathbf{Y})) \leq \alpha$$

avec $\sigma^2 := \max_k \text{var}(Y_k^1)$,

$$B_W := \mathbb{E} \left(\frac{1}{n} \sum_{i=1}^n (W_i - \bar{W})^2 \right)^{1/2} = 1 - \mathcal{O}(n^{-1/2}) \quad \text{et} \quad C_W = 1$$

Outil principal : Théorème de concentration gaussienne [Cirel'son, Ibragimov and Sudakov, 1976]

Méthode par quantiles

- $q_\alpha^{\text{quant}}(\mathbf{Y}) = \text{quantile d'ordre } (1 - \alpha) \text{ de}$

$$\mathcal{D}(\sqrt{n}\|\bar{\mathbf{Y}}_W - \bar{W}\bar{\mathbf{Y}}\|_\infty | \mathbf{Y})$$

Heuristique \Rightarrow devrait vérifier

$$\mathbb{P}(\|\bar{\mathbf{Y}} - \mu\|_\infty > q_\alpha^{\text{quant}}(\mathbf{Y})) \leq \alpha.$$

- \neq “quantile non recentré”, le quantile d'ordre $(1 - \alpha)$ de

$$\mathcal{D}(\sqrt{n}\|\bar{\mathbf{Y}}_W\|_\infty | \mathbf{Y})$$

Méthode par quantiles

- $q_\alpha^{\text{quant}}(\mathbf{Y}) =$ quantile d'ordre $(1 - \alpha)$ de

$$\mathcal{D}(\sqrt{n}\|\bar{\mathbf{Y}}_W - \bar{W}\bar{\mathbf{Y}}\|_\infty | \mathbf{Y})$$

Heuristique \Rightarrow devrait vérifier

$$\mathbb{P}(\|\bar{\mathbf{Y}} - \mu\|_\infty > q_\alpha^{\text{quant}}(\mathbf{Y})) \leq \alpha.$$

- \neq "quantile non recentré", le quantile d'ordre $(1 - \alpha)$ de

$$\mathcal{D}(\sqrt{n}\|\bar{\mathbf{Y}}_W\|_\infty | \mathbf{Y})$$

Théorème quantile : besoin d'un terme supplémentaire

Theorem

Y symétrique. W Rademacher. $\alpha, \delta, \gamma \in (0, 1)$.

Si f est un seuil positif, de FWER contrôlé par $\alpha\gamma/2$:

$$\mathbb{P}(\sqrt{n}\|\bar{\mathbf{Y}} - \mu\|_{\infty} > f(\mathbf{Y})) \leq \frac{\alpha\gamma}{2}$$

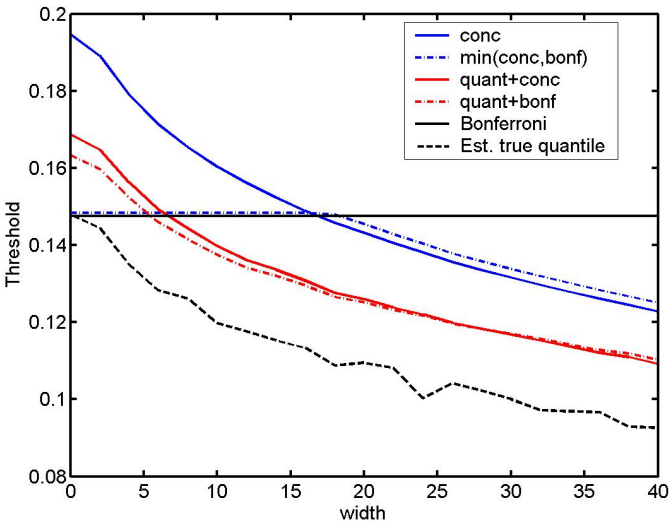
Alors,

$$q_{\alpha}^{\text{quant}+f}(\mathbf{Y}) = q_{\alpha(1-\delta)(1-\gamma)}^{\text{quant}}(\mathbf{Y}) + \sqrt{\frac{2 \log(2/(\delta\alpha))}{n}} f(\mathbf{Y})$$

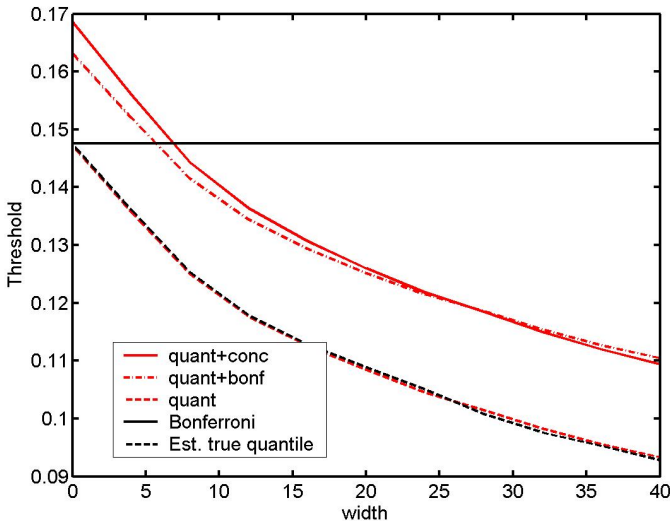
a un FWER contrôlé par α :

$$\mathbb{P}\left(\sqrt{n}\|\bar{\mathbf{Y}} - \mu\|_{\infty} > q_{\alpha}^{\text{quant}+f}(\mathbf{Y})\right) \leq \alpha$$

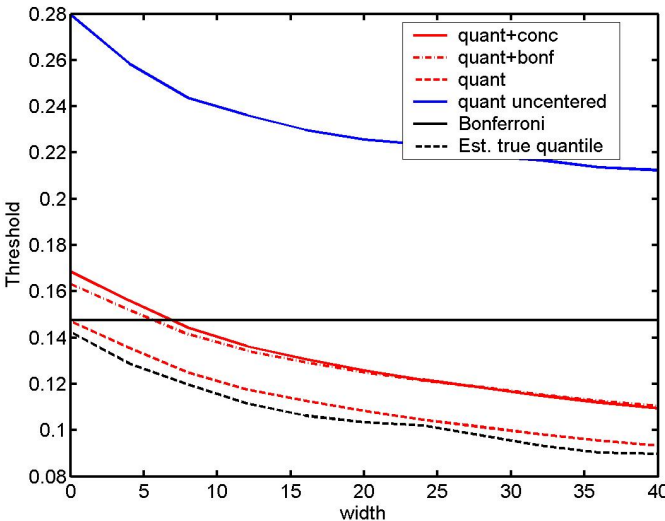
Simulations : $n = 1000$, $K = 16384$, $\sigma = 1$



Simulations : $n = 1000$, $K = 16384$, $\sigma = 1$, $\mu = 0$



Simulations : $n = 1000$, $K = 16384$, $\sigma = 1$, $0 \leq \mu_k \leq 2.9$



Simulations : $0 \leq \mu_k \leq 2.9$; power

