

# Estimation par tests robustes.

Mathieu Sart

Université de Nice Sophia Antipolis

April 17, 2012

# Problème statistique

Problème statistique.

EMV

Procédure

Résultat

Sélection de modèles

Exemple de résultat.

Bornes de risques.

Frameworks.

- ▶ On observe  $(X_1, \dots, X_n)$  un  $n$ -échantillon admettant la densité  $f$  par rapport à Lebesgue.
- ▶ On utilise comme modèle statistique le modèle paramétrique de toutes les lois uniformes sur  $[0, \theta]$ .
- ▶ C'est à dire, que l'on définit

$$\mathcal{F} = \{\theta^{-1} \mathbb{1}_{[0, \theta]}, \theta \in ]0, 1]\}$$

et que l'on cherche un estimateur  $\hat{f}$  à valeurs dans  $\mathcal{F}$  "proche" de  $f$ .

# Problème statistique

Problème  
statistique.

EMV

Procédure

Résultat

Sélection de  
modèlesExemple de  
résultat.Bornes de  
risques.

Frameworks.

- ▶ On observe  $(X_1, \dots, X_n)$  un  $n$ -échantillon admettant la densité  $f$  par rapport à Lebesgue.
- ▶ On utilise comme modèle statistique le modèle paramétrique de toutes les lois uniformes sur  $[0, \theta]$ .
- ▶ C'est à dire, que l'on définit

$$\mathcal{F} = \{\theta^{-1} \mathbb{1}_{[0, \theta]}, \theta \in ]0, 1]\}$$

et que l'on cherche un estimateur  $\hat{f}$  à valeurs dans  $\mathcal{F}$  "proche" de  $f$ .

- ▶ On définit la distance de Hellinger  $H$  par

$$H^2(f, g) = \frac{1}{2} \int_{\mathbb{R}} \left( \sqrt{f(x)} - \sqrt{g(x)} \right)^2 dx$$

et on définit le risque de  $\hat{f}$  par  $\mathbb{E}_f \left[ H^2(\hat{f}, f) \right]$ .

- ▶ Si  $f = \theta^{-1} \mathbb{1}_{[0, \theta]}$ , l'EMV de  $\theta$  est  $X_{(n)} = \max_{1 \leq i \leq n} X_i$ . On estime donc  $f$  par  $\hat{f} = X_{(n)}^{-1} \mathbb{1}_{[0, X_{(n)}]}$ .

- ▶ Si  $f = \theta^{-1} \mathbb{1}_{[0, \theta]}$ , l'EMV de  $\theta$  est  $X_{(n)} = \max_{1 \leq i \leq n} X_i$ . On estime donc  $f$  par  $\hat{f} = X_{(n)}^{-1} \mathbb{1}_{[0, X_{(n)}]}$ .
- ▶ On peut alors montrer que

$$\forall f \in \mathcal{F}, \quad \mathbb{E}_f \left[ H^2(\hat{f}, f) \right] = \frac{1}{2n+1}.$$

- ▶ Que se passe-t-il si  $f \notin \mathcal{F}$ , mais  $f$  "proche" de  $\mathcal{F}$  ?

- ▶ Si  $f = \theta^{-1} \mathbb{1}_{[0, \theta]}$ , l'EMV de  $\theta$  est  $X_{(n)} = \max_{1 \leq i \leq n} X_i$ . On estime donc  $f$  par  $\hat{f} = X_{(n)}^{-1} \mathbb{1}_{[0, X_{(n)}]}$ .
- ▶ On peut alors montrer que

$$\forall f \in \mathcal{F}, \quad \mathbb{E}_f \left[ H^2(\hat{f}, f) \right] = \frac{1}{2n+1}.$$

- ▶ Que se passe-t-il si  $f \notin \mathcal{F}$ , mais  $f$  "proche" de  $\mathcal{F}$  ?
- ▶ On suppose que la densité  $f$  est

$$f = 10 \left[ (1 - 2n^{-1}) \mathbb{1}_{[0, 1/10]} + 2n^{-1} \mathbb{1}_{[9/10, 1]} \right]$$

- ▶  $f$  est proche de  $\mathcal{F}$ :  $H^2(f, 10 \mathbb{1}_{[0, 1/10]}) \leq 5/(4n)$ .

- ▶ Si  $f = \theta^{-1} \mathbb{1}_{[0, \theta]}$ , l'EMV de  $\theta$  est  $X_{(n)} = \max_{1 \leq i \leq n} X_i$ . On estime donc  $f$  par  $\hat{f} = X_{(n)}^{-1} \mathbb{1}_{[0, X_{(n)}]}$ .
- ▶ On peut alors montrer que

$$\forall f \in \mathcal{F}, \quad \mathbb{E}_f \left[ H^2(\hat{f}, f) \right] = \frac{1}{2n+1}.$$

- ▶ Que se passe-t-il si  $f \notin \mathcal{F}$ , mais  $f$  "proche" de  $\mathcal{F}$  ?
- ▶ On suppose que la densité  $f$  est

$$f = 10 \left[ (1 - 2n^{-1}) \mathbb{1}_{[0, 1/10]} + 2n^{-1} \mathbb{1}_{[9/10, 1]} \right]$$

- ▶  $f$  est proche de  $\mathcal{F}$ :  $H^2(f, 10 \mathbb{1}_{[0, 1/10]}) \leq 5/(4n)$ .
- ▶ Pourtant

$$\mathbb{E}_f \left[ H^2(\hat{f}, f) \right] > 0.38$$

# EMV vs Tests robustes

Problème  
statistique.

**EMV**

Procédure

Résultat

Sélection de  
modèles

Exemple de  
résultat.

Bornes de  
risques.

Frameworks.

- ▶ De manière générale, l'EMV peut être très mauvais dans certaines situations. Par exemple lorsque le modèle est légèrement faux.
- ▶ Cela justifie de chercher des méthodes alternatives.

# Définition d'un test.

Problème statistique.

EMV

Procédure

Résultat

Sélection de modèles

Exemple de résultat.

Bornes de risques.

Frameworks.

Soient  $f_1$  et  $f_2$  deux densités. On note  $T(f_1, f_2, \mathbf{X})$  un test à valeurs dans  $\{0, 1\}$ .

- ▶  $T(f_1, f_2, \mathbf{X}) = 1$  signifie que l'on préfère  $f_2$  à  $f_1$ .
- ▶  $T(f_1, f_2, \mathbf{X}) = 0$  signifie que l'on préfère  $f_1$  à  $f_2$ .

Il vérifie en outre  $T(f_1, f_2, \mathbf{X}) = 1 - T(f_2, f_1, \mathbf{X})$

# Choisir parmi un ensemble fini de points

Problème  
statistique.

EMV

**Procédure**

Résultat

Sélection de  
modèles

Exemple de  
résultat.

Bornes de  
risques.

Frameworks.

Soit  $S$  un ensemble fini ou dénombrable de densités.

But: choisir l'élément de  $S$  le plus proche possible de la vraie densité.

# La procédure.

Problème  
statistique.

EMV

Procédure

Résultat

Sélection de  
modèlesExemple de  
résultat.Bornes de  
risques.

Frameworks.

- ▶ Soit  $S$  un ensemble fini (ou dénombrable).
- ▶ Pour toute densité  $f \in S$  notons

$$\mathcal{P}_{\text{referable}}(f) = \{f' \in S \setminus \{f\}, T(f, f', \mathbf{X}) = 1\}$$

- ▶ Définissons

$$\mathcal{D}(f) = \sup \{H^2(f, f'), f' \in \mathcal{P}_{\text{referable}}(f)\}$$

avec la convention  $\mathcal{D}(f) = 0$  si  $\mathcal{P}_{\text{referable}}(f) = \emptyset$ .

- ▶ Un  $T$  estimateur est  $\hat{f} \in S$  tel que

$$\mathcal{D}(\hat{f}) = \inf \{\mathcal{D}(f), f \text{ densité}\}.$$

# Illustration pour notre problème.

Problème statistique.

EMV

Procédure

Résultat

Sélection de modèles

Exemple de résultat.

Bornes de risques.

Frameworks.

- Dans notre problème, on souhaite trouver la fonction  $\hat{f}$  de

$$\mathcal{F} = \{\theta^{-1} \mathbb{1}_{[0, \theta]}, 0 < \theta \leq 1\}$$

la plus proche possible de  $f$ .

# Illustration pour notre problème.

Problème statistique.

EMV

Procédure

Résultat

Sélection de modèles

Exemple de résultat.

Bornes de risques.

Frameworks.

- ▶ Dans notre problème, on souhaite trouver la fonction  $\hat{f}$  de

$$\mathcal{F} = \{\theta^{-1} \mathbb{1}_{[0,\theta]}, 0 < \theta \leq 1\}$$

la plus proche possible de  $f$ .

- ▶ Pour ce faire, on discrétise  $\mathcal{F}$ , ce qui conduit à un ensemble fini (ou dénombrable)  $S$ .
- ▶ On sélectionne alors le meilleur élément de  $S$  par la procédure précédente.

# Illustration pour notre problème.

## Proposition ([Birgé, 2006])

Soient  $X_1, \dots, X_n$ ,  $n$  variables aléatoires i.i.d de densité  $f$ . Il existe un estimateur  $\hat{f} \in \mathcal{F}$  tel que

$$C\mathbb{E}_f \left[ H^2(f, \hat{f}) \right] \leq H^2(f, \mathcal{F}) + \frac{1}{n}$$

où

$$\mathcal{F} = \{ \theta^{-1} \mathbb{1}_{[0, \theta]}, 0 < \theta \leq 1 \} \quad \text{et} \quad H^2(f, \mathcal{F}) = \inf_{g \in \mathcal{F}} H^2(f, g)$$

et où  $C > 0$  est une constante universelle.

# Illustration pour notre problème.

## Proposition ([Birgé, 2006])

Soient  $X_1, \dots, X_n$ ,  $n$  variables aléatoires i.i.d de densité  $f$ . Il existe un estimateur  $\hat{f} \in \mathcal{F}$  tel que

$$CE_f \left[ H^2(f, \hat{f}) \right] \leq H^2(f, \mathcal{F}) + \frac{1}{n}$$

où

$$\mathcal{F} = \{ \theta^{-1} \mathbb{1}_{[0, \theta]}, 0 < \theta \leq 1 \} \quad \text{et} \quad H^2(f, \mathcal{F}) = \inf_{g \in \mathcal{F}} H^2(f, g)$$

et où  $C > 0$  est une constante universelle.

- ▶ Si  $f \in \mathcal{F}$  on a bien une vitesse en  $n^{-1}$ .
- ▶ Si par exemple ,

$$f = 10 \left[ (1 - 2n^{-1}) \mathbb{1}_{[0, 1/10]} + 2n^{-1} \mathbb{1}_{[9/10, 1]} \right]$$

on garde la vitesse en  $n^{-1}$ .

# Vers un théorème de sélection de modèle.

## Proposition

Soient  $X_1, \dots, X_n$ ,  $n$  variables aléatoires i.i.d de densité  $f$ , et soit  $\mathbb{V}$  une collection d'espaces vectoriels de dimension finie de  $\mathbb{L}^2(\mathbb{R}, dx)$ . Soit  $\Delta$  une application positive sur  $\mathbb{V}$  telle que

$$\sum_{V \in \mathbb{V}} e^{-\Delta(V)} \leq 1$$

Il existe un estimateur  $\hat{f}$  tel que

$$C \mathbb{E}_f \left[ H^2(f, \hat{f}) \right] \leq \inf_{V \in \mathbb{V}} \left\{ d^2 \left( \sqrt{f}, V \right) + \frac{\dim V + \Delta(V)}{n} \right\}$$

où  $C > 0$  est une constante universelle et où  $d$  est la distance de  $\mathbb{L}^2(\mathbb{R}, dx)$ .

# Résultats

Problème  
statistique.

EMV

Procédure

Résultat

Sélection de  
modèles

Exemple de  
résultat.

Bornes de  
risques.

Frameworks.

Cette méthode conduit à un estimateur vérifiant des inégalités oracles très générales. On peut en déduire des bornes de risques non asymptotiques sous des hypothèses

- ▶ de régularité sur  $f$
- ▶ structurelles sur  $f$

En outre, l'estimateur est adaptatif et robuste par rapport à ces hypothèses.

# Une méthode souple.

Cette méthode peut être utilisée dans différents problèmes statistiques:

- ▶ Estimation d'une densité. [Birgé, 2006] [Birgé, 2008]
- ▶ Estimation de l'intensité d'un processus de Poisson. [Birgé, 2007] [Baraud, 2010]
- ▶ Estimation de l'intensité d'un processus de Poisson avec covariables. [Sart, 2012]
- ▶ Estimation de la moyenne de variables aléatoires positives. [Baraud, 2010]
- ▶ Régression non gaussienne. [Birgé, 2006] [Baraud, 2010]
- ▶ Estimation de la densité d'un processus déterminantal. [Baraud, 2012]



Baraud, Y. (2010).

Estimator selection with respect to hellinger-type risks.

*Probability Theory and Related Fields*, pages 1–49.

[10.1007/s00440-010-0302-y](https://doi.org/10.1007/s00440-010-0302-y).



Baraud, Y. (2012).

Estimation of the density of a determinantal process.



Birgé, L. (2006).

Model selection via testing: an alternative to (penalized) maximum likelihood estimators.

*Annales de l'Institut Henri Poincaré (B) Probability and Statistics*, 42(3):273 – 325.



Birgé, L. (2007).

Model selection for poisson processes.

*Lecture Notes-Monograph Series*, 55:pp. 32–64.



Birgé, L. (2008).

Model selection for density estimation with  $\mathbb{L}_2$ -loss.

*ArXiv e-prints*.



Sart, M. (2012).

Model selection for poisson processes with covariates.