On efficient estimators of the proportion of true null hypotheses in a multiple testing setup

Van Hanh Nguyen and Catherine Matias

CNRS - Laboratoire de Probabilités et Modèles Aléatoires, Paris
catherine.matias@math.cnrs.fr
http://cmatias.perso.math.cnrs.fr/

Journées MAS - Août 2014.

Outline

Multiple testing procedures (MTP)

- MT appears in many applications: microarray analysis, signal detection, astrophysics, ...
- Controlling the type I error of each test (e.g. nominal level α) may result in a large number of false positives.
- MTP aim at controlling global quantities, such as
  - Family-wise error rate: FWER$=\mathbb{P}(\mathsf{FP} \geq 1)$ (too stringent)
  - False discovery rate:
    $\mathsf{FDR} = \mathbb{E}\left(\frac{\mathsf{FP}}{\max(\mathsf{R},1)}\right) = \mathbb{E}\left(\frac{\mathsf{FP}}{\mathsf{R}} \middle| \mathsf{R} > 0\right) \mathbb{P}(\mathsf{R} > 0)$
  - Positive FDR: pFDR$=\mathbb{E}\left(\frac{\mathsf{FP}}{\mathsf{R}} | \mathsf{R} > 0\right)$
  - ...

	Accept H^i	Reject H^i	Total
H^i true	TN	FP	n_0
H^i false	FN	TP	n_1
Total	W	R	n

Table : Possible outcomes from testing n hypotheses $H^1, \ldots, H^n$

Error control vs error estimation

Two points of view on MTP

- Either estimate the error (FDR or pFDR or ...) for some fixed rejection region;
- Or, fix an a priori upper bound on the error and find a rejection region with controlled error.

Equivalent issues

- In fact these two points of view merge, as most of the MTP may be viewed as threshold procedures applied to estimates of FDR, pFDR ....
- Thus estimating FDR or pFDR is of major interest in MT.
- These quantities are closely related to the proportion of true null hypotheses and the density under the alternative hypothesis.

Outline

Semi-parametric mixture model

Notation

- Consider n identical hypotheses with test statistics $T_1, \ldots, T_n$ and p-values $P_1, \ldots, P_n$
- Let $H^i = 0$ if the i-th null hypothesis is true, and 1 otherwise. Assume H^i are i.i.d. variables.
- If $T_i | H^i = 0$ has a continuous distribution, then $P_i | H^i = 0 \sim \mathcal{U}([0,1])$
- Then the P_i are i.i.d. and follow a mixture distribution $g(x) = \theta 1_{[0,1]}(x) + (1-\theta) f(x), \ x \in [0,1]$
- $\theta \in [0,1]$ is the unknown proportion of true null hypotheses
- f is the unknown density of P_i under the alternative $H^i = 1$.

The model is parametrized by (θ, f).

Identifiability

Proposition

The parameter (θ, f) *is identifiable on a set* $(0,1) \times \mathcal{F}$ *if and only if for all* $f \in \mathcal{F}$ *and for all* $c \in (0,1)$*, we have* $c + (1-c)f \notin \mathcal{F}$*.*

Examples of sets $\mathcal{F}$

- Purity condition [Genovese & Wasserman, 2004]: $\inf_{x \in [0,1]} f(x) = 0$
- [Langaas *et al.*, 2005]: f is non-increasing with $f(1) = 0$
- [Pounds & Cheng, 2006, Celisse & Robin, 2010]: f vanishes in a neighborhood of 1 or an open interval in $(0,1)$.

In the following, we work on the set

$$\mathcal{F}_\lambda = \{f : [0,1] \mapsto \mathbb{R}^+, \text{ continuously non increasing density, positive on } [0,\lambda) \text{ and such that } f_{|[\lambda,1]} = 0\}.$$

Estimation of the proportion θ

Many proposals in the literature.

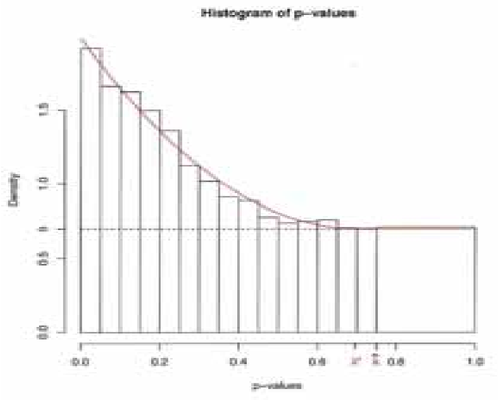
3 main types of estimators

- Histogram based estimators;
- Monotone density estimators;
- Regular density estimators.

Histogram based estimators I

Underlying assumption

f vanishes on some neighborhood of 1. *E.g.* [Schweder & Spjøtvoll, 1982]'s estimator



$$\hat{\theta}_n(\lambda) = \frac{\sharp\{P_i > \lambda, 1 \leq i \leq n\}}{n(1-\lambda)}$$

Choice of λ

- Fixed value: $\lambda = 1/2$ most popular choice;
- Adaptive choices. Many references, among which:
  - [Benjamini & Hochberg 2000]: detection of a change of slope;
  - [Storey 2002]: bootstrap procedure;
  - [Celisse & Robin, 2010]: cross-validation (LpO) procedure;

Histogram based estimators II

Convergence properties

- Very few convergence results have been established;
- [Celisse & Robin, 2010]'s estimator is proved to convergent in probability;
- Properties of [Schweder & Spjøtvoll, 1982]'s oracle version: if $f_{|[\lambda^\star,1]} = 0$ and $\lambda = \lambda^\star$ then $\sqrt{n}(\hat{\theta}_n(\lambda^\star) - \theta) \rightarrow^d_{n\rightarrow\infty} \mathcal{N}(0, \theta(\frac{1}{1-\lambda^\star} - \theta))$.

Monotone density and regular densities estimators

Other estimators

- Grenander's estimate is proposed by [Langaas *et al.*, 2005]; Converges at nonparametric rate $(\log n)^{1/3} n^{-1/3}$;
- Regular density estimators: [Neuvial, 2013] proposed a kernel based estimator; Converges at nonparametric rate $n^{-k/(2k+1)}\eta_n$, where $\eta_n \to \infty$ and k controls regularity of f near $x = 1$.

Issues

- When is it possible to construct an estimator converging at parametric rate?
- What is the optimal asymptotic variance of a parametric estimator and are there efficient estimators?

Results

Let us recall that we work on

$$f \in \mathcal{F}_\lambda = \{f : [0,1] \mapsto \mathbb{R}^+, \text{ continuously non increasing density, positive on } [0,\lambda) \text{ and such that } f_{|[\lambda,1]} = 0\}.$$

2 different cases occur

- When $\lambda = 1$: any estimator of θ cannot converge at parametric rate.
- When $\lambda < 1$: we can construct estimators converging at parametric rate but they are not asymptotically efficient (except for irregular models).

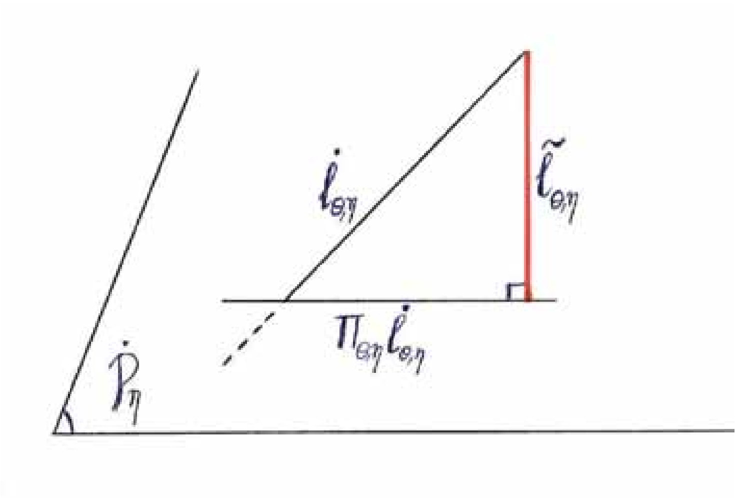
Outline

Asymptotic efficiency theory in semi-parametric models I

Let $\mathcal{P} = \{\mathbb{P}_{\theta,\eta} : \theta \in \Theta, \eta \in \mathcal{F}\}$, with $\Theta \subset \mathbb{R}$ an open set and $\mathcal{F}$ an infinite dimensional set.

We aim at estimating $\psi(\theta)$.

- The ordinary score function: $\dot{l}_{\theta,\eta} = \frac{\partial}{\partial\theta} \log d\mathbb{P}_{\theta,\eta}$.
- A tangent set for η:
  $$\dot{\mathcal{P}}_\eta = \left\{ \frac{\partial}{\partial t}\Big|_{t=0} \log d\mathbb{P}_{\theta,\eta_t} : \text{ for suitable paths } t \mapsto \eta_t \text{ in } \mathcal{F} \right\}$$
- The efficient score function: $\tilde{l}_{\theta,\eta} = \dot{l}_{\theta,\eta} - \Pi_{\theta,\eta}\dot{l}_{\theta,\eta}$, where $\Pi_{\theta,\eta}$ is the orthogonal projection onto $\overline{\text{lin}}\dot{\mathcal{P}}_\eta$ in $\mathbb{L}_2(\mathbb{P}_{\theta,\eta})$.
- The efficient information: $\tilde{I}_{\theta,\eta} = \mathbb{E}_{\theta,\eta}\tilde{l}^2_{\theta,\eta}$



Asymptotic efficiency theory in semi-parametric models II

Definition. An estimator $\hat{\theta}_n$ is asymptotically efficient if and only if it satisfies

$$\sqrt{n}(\hat{\theta}_n - \theta) = \tfrac{1}{\sqrt{n}} \textstyle\sum_{i=1}^n \tilde{I}_{\theta,\eta}^{-1} \tilde{l}_{\theta,\eta}(X_i) + o_{\mathbb{P}_{\theta,\eta}}(1).$$

As a consequence,

- By the central limit theorem and Slutsky's theorem,
$$\sqrt{n}(\hat{\theta}_n - \theta) \overset{\mathbb{P}_{\theta,\eta}}{\rightsquigarrow} N(0, \tilde{I}_{\theta,\eta}^{-1}).$$
- By the LAM theorem: the optimal variance is $\tilde{I}_{\theta,\eta}^{-1}$.

Efficient score and information in our case

$$\mathcal{P}_{\lambda^*} = \left\{\mathbb{P}_{\theta,f};\, \frac{d\mathbb{P}_{\theta,f}}{d\mu} = \theta + (1-\theta)f;\, (\theta,f) \in (0,1)\times\mathcal{F}_{\lambda^*}\right\}.$$

Proposition. The efficient score function $\tilde{l}_{\theta,f}$ and the efficient information $\tilde{I}_{\theta,f}$ for estimating θ in model $\mathcal{P}_{\lambda^*}$ are given by

$$\tilde{l}_{\theta,f}(x) = \tfrac{1}{\theta} - \tfrac{1}{\theta[1-\theta(1-\lambda^*)]}\mathbf{1}_{[0,\lambda^*)}(x) \text{ and } \tilde{I}_{\theta,f} = \tfrac{1-\lambda^*}{\theta[1-\theta(1-\lambda^*)]}.$$

Corollary.

- When $\lambda^* = 1$, we have $\tilde{I}_{\theta,f} = 0$, then there is no estimator of θ converging at parametric rate.
- When $\lambda^* < 1$, an estimator $\hat{\theta}_n$ of θ is asympt. eff. if and only if it satisfies

$$\hat{\theta}_n = \frac{\#\{X_i > \lambda^* : 1 \le i \le n\}}{n(1-\lambda^*)} + o_{\mathbb{P}_{\theta,f}}(n^{-1/2}),$$

with the optimal variance equal to $\theta\left(\frac{1}{1-\lambda^*} - \theta\right)$.

Case $\lambda^* < 1$

Let us further investigate what may be obtained in this case:

- Can we exhibit $\sqrt{n}$-consistent estimators?
- If yes, do they asymptotically achieve the optimal variance?

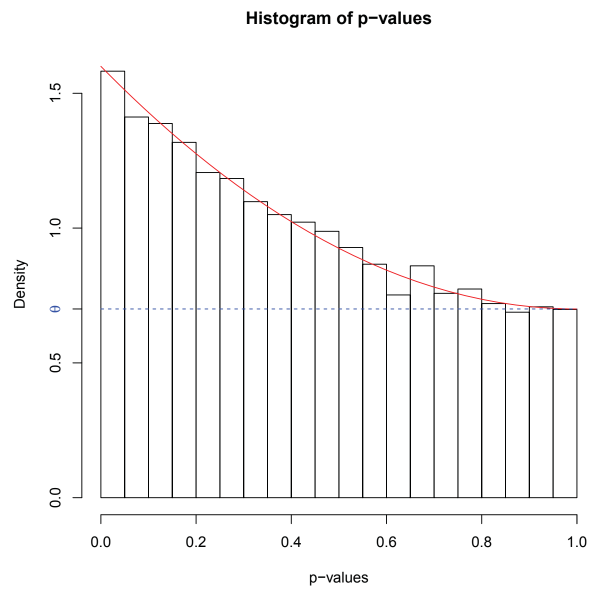
Outline

Case $\lambda^* < 1$: estimators with parametric rate I

A histogram based estimator

$\hat{f}_I$: a histogram estimator of f.
Define an estimator of θ as

$$\hat{\theta}_{I,n} = \min_{x \in [0,1]} \hat{f}_I(x)$$



Theorem

Suppose that $f \in \mathcal{F}^*_\lambda$ with $\lambda^* < 1$ and I is fine enough, then the estimator $\hat{\theta}_{I,n}$ has the following properties

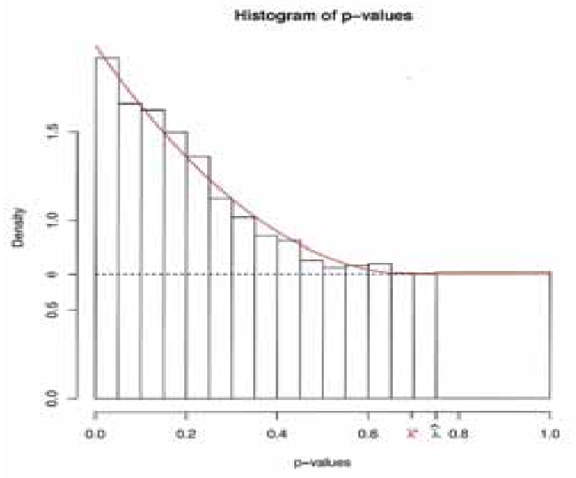
i) *$\hat{\theta}_{I,n}$ converges almost surely to θ,*

ii) $\limsup_{n \to \infty} n\mathbb{E}\left[(\hat{\theta}_{I,n} - \theta)^2\right] < +\infty.$

Case $\lambda^* < 1$: estimators with parametric rate II

Celisse & Robin [2010]'s procedure

$\hat{\theta}_n^{CR}$: estimator proposed by [Celisse & Robin, 2010]
$\hat{\lambda}$: chosen adaptively based on cross-validation method.



Theorem

Under some assumptions, the estimator $\hat{\theta}_n^{CR}$ has the following properties

i) $\hat{\theta}_n^{CR}$ *converges almost surely to* θ*,*
ii) $\hat{\theta}_n^{CR}$ *is* $\sqrt{n}$*-consistent, i.e.* $\sqrt{n}(\hat{\theta}_n^{CR} - \theta) = O_{\mathbb{P}}(1)$*,*
iii) *If the parameter* p *in leave-p-out estimator is fixed then* $\limsup_{n\to\infty} n\mathbb{E}\left[(\hat{\theta}_n^{CR} - \theta)^2\right] < +\infty$.

Case $\lambda^* < 1$: estimators with parametric rate III

Additional remarks

- We did not succeed in computing the asymptotic variance of these estimators;
- In the simulations, we further study this point.

"One-step" estimator

- The one-step procedure is a general method for constructing an asymptotically efficient estimator starting from a $\sqrt{n}$-convergent one.

One step procedure

Construction

Let $\hat{\theta}_n$ a $\sqrt{n}$-consistent estimator of θ and $\hat{l}_{n,\theta}(\cdot) = \hat{l}_{n,\theta}(\cdot; X_1, \dots, X_n)$ an estimator of $\tilde{l}_{\theta,f}$. Denoting $m = \lfloor n/2 \rfloor$, we let

$$\hat{l}_{n,\theta,i}(\cdot) = \begin{cases} \hat{l}_{m,\theta}(\cdot; X_1, \dots, X_m) & \text{if} \quad i > m, \\ \hat{l}_{n-m,\theta}(\cdot; X_{m+1}, \dots, X_n) & \text{if} \quad i \leq m. \end{cases}$$

Then, a one-step estimator is constructed as

$$\tilde{\theta}_n = \hat{\theta}_n - \left(\sum_{i=1}^{n} \hat{l}^2_{n,\hat{\theta}_n,i}(X_i) \right)^{-1} \sum_{i=1}^{n} \hat{l}_{n,\hat{\theta}_n,i}(X_i).$$

Existence of asympt. eff. estimators

For an estimator $\hat{l}_{n,\theta}(\cdot) = \hat{l}_{n,\theta}(\cdot; X_1, \ldots, X_n)$ of $\tilde{l}_{\theta,f}$ and every sequence $\theta_n = \theta + O(n^{-1/2})$, introduce the following conditions

$$\sqrt{n}\mathbb{P}_{\theta_n,f}\hat{l}_{n,\theta_n} \xrightarrow[n\to\infty]{\mathbb{P}_{\theta,f}} 0, \tag{1}$$

$$\mathbb{P}_{\theta_n,f}\|\hat{l}_{n,\theta_n} - \tilde{l}_{\theta_n,f}\|^2 \xrightarrow[n\to\infty]{\mathbb{P}_{\theta,f}} 0 \tag{2}$$

Proposition (▸ Recall)

- *The existence of asympt. eff. estimator of $\theta \iff$ the existence of estimator $\hat{l}_{n,\theta}$ of $\tilde{l}_{\theta,f}$ satisfying (1) and (2).*
- *If $\tilde{l}_{\theta,f}$ is estimated through a plug-in estimate $\hat{\lambda}_n$ of λ^*, then this condition is equivalent to $\sqrt{n}(\hat{\lambda}_n - \lambda^*) = o_{\mathbb{P}}(1)$.*

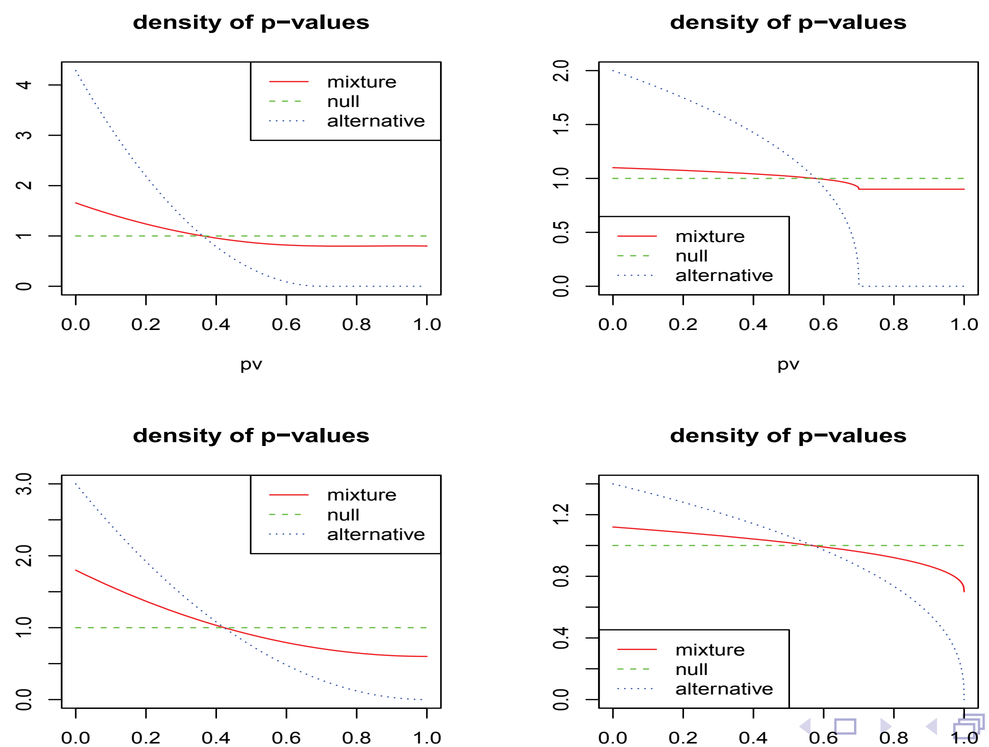
Existence

- Irregular models: f has a jump point at λ^*, YES
- Regular models: conjecture that NO

Outline

Simulations setup

- Consider the alternative density
$$f_1(x) = \frac{s}{\lambda^*}\left(1 - \frac{x}{\lambda^*}\right)^{s-1} \mathbf{1}_{[0,\lambda^*]}(x)$$
- Different parameter values:
$s \in \{1.4; 3\}, \lambda^* \in \{0.7; 1\}, \theta \in \{0.6; 0.7; 0.8; 0.9\}$
- Sample size
$n \in \{5000; 7000; 9000; 10000; 12000; 14000; 15000\}$ and
$S = 100$ repetitions.



Simulations

$\lambda^* = 0.7$ $\qquad\qquad$ $\lambda^* = 1$

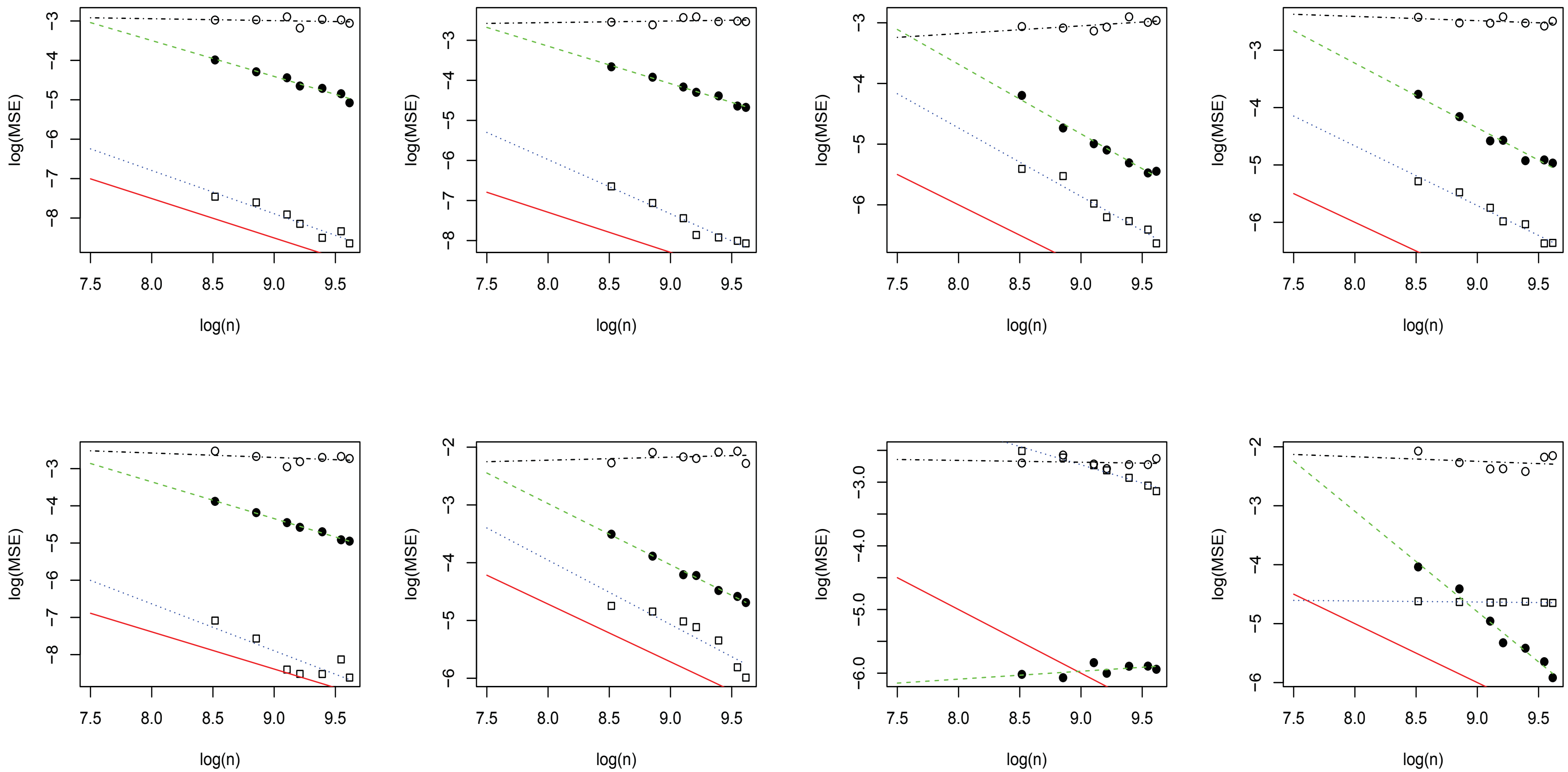


Figure : Logarithm of MSE as a function of $\log(n)$ and linear regression for $\hat{\theta}_n^L$ (black line), $\hat{\theta}_n^{CR}$ (blue line) and $\hat{\theta}_{I,n}$ (green line). Red line: $y = -\log(n) + \log[\theta((1-\lambda^*)^{-1} - \theta)]$ (oracle version).

Outline

Conclusions

Conclusions

Consider a mixture $g(x) = \theta \mathbf{1}_{[0,1]}(x) + (1-\theta)f(x)$, where the density f is non-increasing and its support stops at λ^*.

- $\lambda^* = 1$: there is no estimator of θ converging at parametric rate
- $\lambda^* < 1$:
  - Two estimators of θ converging at parametric rate
  - Irregular model: it is possible to construct an asymptotically efficient estimator of θ.
  - Regular models: we conjecture that asymptotically efficient estimators of θ do not exist.

References I

Y. Benjamini and Y. Hochberg.
On the adaptive control of the false discovery rate in multiple testing with independent statistics.
J. Educ. Behav. Stat. Ser., 25, 2000.

A. Celisse and S. Robin.
A cross-validation based estimation of the proportion of true null hypotheses.
J. Statist. Plann. Inference, 140(11):3132–3147, 2010.

C. Genovese and L. Wasserman.
A stochastic process approach to false discovery control.
Ann. Statist., 32(3):1035–1061, 2004.

M. Langaas, B. H. Lindqvist, and E. Ferkingstad.
Estimating the proportion of true null hypotheses, with application to DNA microarray data.
J. R. Stat. Soc. Ser. B Stat. Methodol., 67(4):555–572, 2005.

P. Neuvial.
Asymptotic results on adaptive false discovery rate controlling procedures based on kernel estimators.
Journal of Machine Learning Research, 14:1423–1459, 2013.

S. Pounds and C. Cheng.
Robust estimation of the false discovery rate.
Bioinformatics, 22(16):1979–1987, 2006.

T. Schweder and E. Spjøtvoll.
Plots of p-values to evaluate many tests simultaneously.
Biometrika, 69(3):493–502, 1982.

J. D. Storey.
A direct approach to false discovery rates.
J. R. Stat. Soc. Ser. B Stat. Methodol., 64(3):479–498, 2002.

References II

A. van der Vaart.
Asymptotic statistics.
Cambridge Series in Statistical and Probabilistic Mathematics, Vol 3, 1998.