

---

# Contrôle non-asymptotique adaptatif du *family-wise error rate* en tests multiples

---

G. Blanchard<sup>1</sup>

<sup>1</sup>Weierstrass Institut Berlin, Germany

Journées Statistiques du Sud, Porquerolles 17/06/09

- 1 Introduction
- 2 Contrôle du FWER et adaptativité à  $\pi_0$
- 3 Adaptativité à la dépendance et régions de confiance
  - Résultats théoriques
  - Quelques simulations

## 1 Introduction

## 2 Contrôle du FWER et adaptativité à $\pi_0$

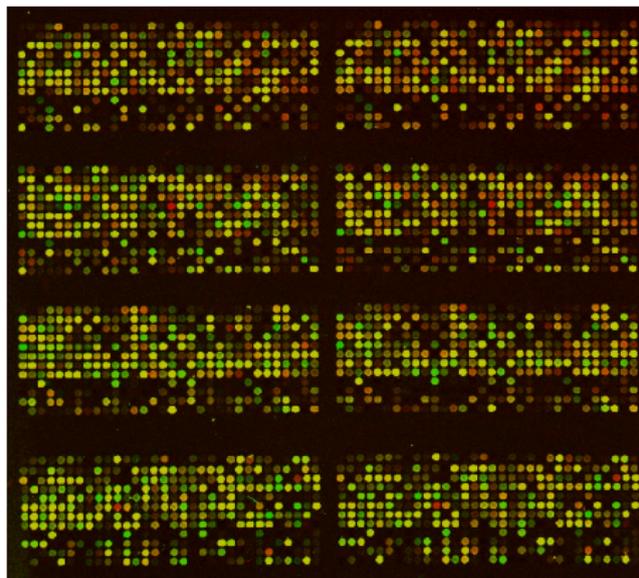
## 3 Adaptativité à la dépendance et régions de confiance

Résultats théoriques

Quelques simulations

# Tests d'hypothèses multiples : motivation

- ▶ problème de type “fouille de données” : déterminer des caractéristiques “marquantes” ou “significatives” d’une population, à partir d’un jeu de données, parmi une grande liste de candidats
- ▶ exemple typique : données en grande dimension, où chaque dimension correspond à un facteur
  - quels facteurs ont une moyenne nulle ?
  - comparaison de deux échantillons : quels facteurs ont une moyenne différente ?
  - détermination de relations d’indépendance entre facteurs (evt. conditionnelle) : modèles graphiques, sélection de variables pertinentes en régression
- ▶ problème central : complexité et contrôle de critères d’erreur de groupe
- ▶ on se concentrera en particulier sur des résultats non-asymptotiques



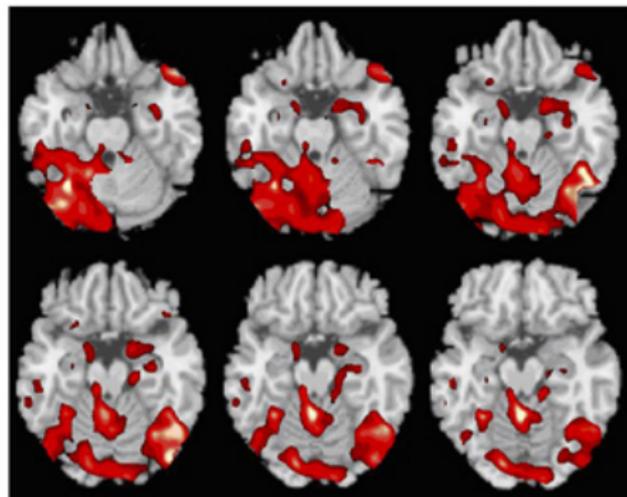
## Exemple – données microarray :

- ▶ petit nombre d'observations répétées en grande dimension
- ▶ quels gènes ont un niveau d'expression significativement élevée ou différencié ?
- ▶ chaque gène peut être testé séparément
- ▶ dizaines de milliers de gènes à tester

# Tests multiples – exemples

## Exemple – données d'imagerie fonctionnelle :

- ▶ petit nombre de séries temporelles répétées d'imagerie cérébrale pendant l'accomplissement d'une tâche
- ▶ quelles régions du cerveau sont-elles significativement actives ?
- ▶ tout point du cerveau (pixel) peut être testé séparément
- ▶ dépendances de courte et longue portée entre régions
- ▶ contraintes supplémentaires éventuelles sur la géométrie des régions actives



# Tests multiples – exemples

## Exemple – motifs significatifs d'une séquence ADN :

- ▶ modéliser l'ADN comme une chaîne de Markov d'ordre  $k$
- ▶ chercher les motifs de longueur  $\ell > k + 1$  qui ont une fréquence supérieure à la normale, suggérant un rôle biologique (par ex. marqueurs de "splicing")
- ▶ un motif spécifique peut être testé en utilisant des approximations du modèle Markovien (voir par ex. Roquain et Schbath, 2007)
- ▶ on veut tester tous les motifs de taille  $\ell$  ( $4^\ell$  possibilités)

```
AGGTTGCATATCGCAT-
TTCATCATAAAAGCTC-
CAGAACAACGACTAGC-
TGGACCAATCGGTCCG-
ATGAGGACAAGTTCCT-
ACAAGAAAGAGAGGGT-
CATCCACCCGCTTATT-
TATAGCTGGAATTCCT-
CTCCGGCATATACATA-
GTGTAAGGATGTAGCC
```

- ▶ un modèle de distributions  $\mathfrak{P}$  sur un espace  $\mathcal{X}$ , supposé contenir la vraie loi génératrice  $P$  (exemple :  $n$ -échantillon i.i.d. de Gaussiennes en dimension  $d$  de covariance  $I$  et de moyenne inconnue)
- ▶ une **hypothèse nulle**  $h$  est formellement un sous-ensemble de  $\mathfrak{P}$  (exemple : Gaussiennes dont la moyenne de la première coordonnée est nulle)
- ▶ test simple : fonction  $T_h : \mathcal{X} \rightarrow \{0, 1\}$ , indiquant une décision prise d'après l'observation, acceptant ou rejetant l'hypothèse  $P \in h$
- ▶ erreur de première espèce ou niveau :

$$\alpha = \sup_{P \in h} P(T(\mathbf{X}) = 1)$$

- ▶ le plus souvent, un test est basé sur une statistique  $Z(\mathbf{X}) \in \mathbb{R}$  et

$$T_h(\mathbf{X}) = \mathbf{1}\{Z(\mathbf{X}) \geq t\}$$

- ▶ plus généralement, une **famille croissante** de tests dépendant d'un paramètre  $\alpha$  tel que  $T_{h,\alpha}(\mathbf{X}) \leq T_{h,\alpha'}(\mathbf{X})$  si  $\alpha \leq \alpha'$ , et (après reparamétrisation monotone) telle que

$$\sup_{P \in h} P(T_\alpha(\mathbf{X}) = 1) = \alpha;$$

- ▶ on définit alors la fonction  **$p$ -value**  $p_h : \mathcal{X} \rightarrow [0, 1]$  associée,

$$p_h(\mathbf{X}) = \inf \{ \alpha \in [0, 1] \mid T_\alpha(\mathbf{X}) = 1 \},$$

caractérisée par :

$$\forall P \in h \quad \mathbb{P}_{\mathbf{X} \sim P} [p_h(\mathbf{X}) \leq \alpha] \leq \alpha. \quad (\star)$$

- ▶ Réciproquement, la donnée d'une fonction  $p$ -value  $p_h$  (càd satisfaisant la propriété  $(\star)$ ) assure le **contrôle de l'erreur de première espèce** au niveau  $\alpha$  pour le test

$$T_{h,\alpha}(\mathbf{X}) = \mathbf{1}\{p_h(\mathbf{X}) \leq \alpha\}$$

- ▶ le plus souvent, un test est basé sur une statistique  $Z(\mathbf{X}) \in \mathbb{R}$  et

$$T_h(\mathbf{X}) = \mathbf{1}\{Z(\mathbf{X}) \geq t\}$$

- ▶ plus généralement, une **famille croissante** de tests dépendant d'un paramètre  $\alpha$  tel que  $T_{h,\alpha}(\mathbf{X}) \leq T_{h,\alpha'}(\mathbf{X})$  si  $\alpha \leq \alpha'$ , et (après reparamétrisation monotone) telle que

$$\sup_{P \in h} P(T_\alpha(\mathbf{X}) = 1) = \alpha;$$

- ▶ on définit alors la fonction  **$p$ -value**  $p_h : \mathcal{X} \rightarrow [0, 1]$  associée,

$$p_h(\mathbf{X}) = \inf \{ \alpha \in [0, 1] \mid T_\alpha(\mathbf{X}) = 1 \},$$

caractérisée par :

$$\forall P \in h \quad \mathbb{P}_{X \sim P} [p_h(\mathbf{X}) \leq \alpha] \leq \alpha. \quad (\star)$$

- ▶ Réciproquement, la donnée d'une fonction  $p$ -value  $p_h$  (càd satisfaisant la propriété  $(\star)$ ) assure le **contrôle de l'erreur de première espèce** au niveau  $\alpha$  pour le test

$$T_{h,\alpha}(\mathbf{X}) = \mathbf{1}\{p_h(\mathbf{X}) \leq \alpha\}$$

- ▶ un ensemble (usuellement fini)  $\mathcal{H}$  d'hypothèses nulles à tester.
- ▶ procédure de test multiple  $R$  :

Observation  $\mathbf{X}$   $\rightarrow$  Hypothèses rejetées  $R(\mathbf{X}) \subset \mathcal{H}$

- ▶ on note  $\mathcal{H}_0(P) \subset \mathcal{H}$  l'ensemble des hypothèses nulles qui sont satisfaites (i.e. contiennent) la distribution génératrice  $P$ .
- ▶ on note  $\pi_0(P) = |\mathcal{H}_0|/|\mathcal{H}|$  la proportion d'hypothèses nulles.

# Abstraction par $p$ -values

- ▶ Un problème souvent considéré est celui de test multiple construit à partir de tests simples connus.
- ▶ Supposons que pour tout  $h \in \mathcal{H}$ , un test simple  $T_h$  et une  $p$ -value correspondante  $p_h$  sont connus.
- ▶ On considère alors les procédures de tests multiples qui sont des fonctions de la famille des  $p$ -values :

$$\text{Données } \mathbf{X} \rightarrow p\text{-values } \mathbf{p} = (p_h(\mathbf{X}))_{h \in \mathcal{H}} \rightarrow R(\mathbf{p}) \subset \mathcal{H}$$

- ▶ Une sous-famille souvent considérée : rejet des  $p$ -values plus petites qu'un **seuil**,

$$R(\mathbf{p}) = \left\{ h \in \mathcal{H} : p_h(\mathbf{X}) \leq \hat{t} \right\},$$

$\hat{t}$  étant éventuellement un seuil aléatoire.

# Modèles de distribution des $p$ -values

- ▶ si  $h$  est satisfaite par  $P$ , alors  $p_h$  a une distribution stochastiquement bornée inférieurement par une variable  $U([0, 1])$  :

$$P \in h \Rightarrow P(p_h \leq t) \leq t$$

(c'est la définition d'une  $p$ -value)

- ▶ **(U<sub>0</sub>)** si  $h$  est satisfaite par  $P$ , alors  $p_h \sim U([0, 1])$
- ▶ **(A<sub>1</sub>)** si  $h$  n'est pas satisfaite par  $P$ , alors  $p_h \sim P_1$  pour une certaine loi  $P_1$
- ▶ **(I)** indépendance : les  $p$ -values sont indépendantes
- ▶ **(RE)** modèle "random effects", à tendance Bayésienne : soient  $h_1, \dots, h_N$  des variables de Bernoulli indépendantes de paramètre  $1 - \pi_0$ , les  $p$ -values sont indépendantes conditionnellement aux  $(h_i)$  avec

$$p_i \sim \begin{cases} U([0, 1]) & \text{si } h_i = 0, \\ P_1 & \text{si } h_i = 1. \end{cases}$$

Les hypothèses nulles vraies forment un ensemble aléatoire donné par  $\{i : h_i = 1\}$ .

# Critères d'erreur de première espèce

- ▶ l'erreur d'ensemble de  $R$  a famille (family-wise error rate ou **FWER**) est définie comme la probabilité que  $R(X)$  contienne au moins une hypothèse nulle :

$$\text{FWER}(R, P) = \mathbb{P}_{\mathbf{X} \sim P} [R \cap \mathcal{H}_0(P) \neq \emptyset]$$

- ▶ la proportion de fausses découvertes (false discovery proportion, FDP) est la variable aléatoire

$$\text{FDP}(R, P) = \frac{|\mathcal{H}_0(P) \cap R|}{|R|} \mathbf{1}_{\{|R| > 0\}}$$

(avec la convention  $= 0$  si  $|R| = 0$ ) ;

- ▶ le taux de fausses découvertes (false discovery rate, **FDR**) est

$$\text{FDR}(R, P) = \mathbb{E} [\text{FDP}(R, P)] = \mathbb{E} \left[ \frac{|\mathcal{H}_0(P) \cap R|}{|R|} \mathbf{1}_{\{|R| > 0\}} \right].$$

# Critères d'erreur de première espèce

- ▶ l'erreur d'ensemble de  $R$  a famille (family-wise error rate ou **FWER**) est définie comme la probabilité que  $R(X)$  contienne au moins une hypothèse nulle :

$$\text{FWER}(R, P) = \mathbb{P}_{\mathbf{X} \sim P} [R \cap \mathcal{H}_0(P) \neq \emptyset]$$

- ▶ la proportion de fausses découvertes (false discovery proportion, FDP) est la variable aléatoire

$$\text{FDP}(R, P) = \frac{|\mathcal{H}_0(P) \cap R|}{|R|} \mathbf{1}_{\{|R| > 0\}}$$

(avec la convention  $= 0$  si  $|R| = 0$ ) ;

- ▶ le taux de fausses découvertes (false discovery rate, **FDR**) est

$$\text{FDR}(R, P) = \mathbb{E} [\text{FDP}(R, P)] = \mathbb{E} \left[ \frac{|\mathcal{H}_0(P) \cap R|}{|R|} \mathbf{1}_{\{|R| > 0\}} \right].$$

- 1 Introduction
- 2 Contrôle du FWER et adaptativité à  $\pi_0$
- 3 Adaptativité à la dépendance et régions de confiance
  - Résultats théoriques
  - Quelques simulations

# Adaptativité ?

---

- ▶ le mot “adaptativité” utilisé en tests multiples est un peu galvaudé. Il s’agit ici plutôt d’un point de vue semi-paramétrique :
- ▶ “adaptativité” à des paramètres de nuisance ou annexes :
  - ▶ à la proportion  $\pi_0$  d’hypothèses nulles
  - ▶ à la structure de dépendance inconnue des  $p$ -values
  - ▶ à la loi des  $p$ -values des hypothèses non-nulles

# Adaptativité ?

---

- ▶ le mot “adaptativité” utilisé en tests multiples est un peu galvaudé. Il s’agit ici plutôt d’un point de vue semi-paramétrique :
- ▶ “adaptativité” à des paramètres de nuisance ou annexes :
- ▶ à la proportion  $\pi_0$  d’hypothèses nulles
- ▶ à la structure de dépendance inconnue des  $p$ -values
- ▶ à la loi des  $p$ -values des hypothèses non-nulles

# Adaptativité ?

---

- ▶ le mot “adaptativité” utilisé en tests multiples est un peu galvaudé. Il s’agit ici plutôt d’un point de vue semi-paramétrique :
- ▶ “adaptativité” à des paramètres de nuisance ou annexes :
- ▶ à la proportion  $\pi_0$  d’hypothèses nulles
- ▶ à la structure de dépendance inconnue des  $p$ -values
- ▶ à la loi des  $p$ -values des hypothèses non-nulles

# Adaptativité ?

---

- ▶ le mot “adaptativité” utilisé en tests multiples est un peu galvaudé. Il s’agit ici plutôt d’un point de vue semi-paramétrique :
- ▶ “adaptativité” à des paramètres de nuisance ou annexes :
- ▶ à la proportion  $\pi_0$  d’hypothèses nulles
- ▶ à la structure de dépendance inconnue des  $p$ -values
- ▶ à la loi des  $p$ -values des hypothèses non-nulles

# Contrôle non-adaptatif : correction de Bonferroni

- ▶ correction de Bonferroni : rejeter

$$R = \{h \in \mathcal{H} : p_h \leq \alpha/|\mathcal{H}|\} .$$

- ▶ alors le FWER est **contrôlé au niveau  $\alpha$**  car

$$\begin{aligned} \text{FWER}(R, P) &= \mathbb{P}_{\mathbf{X} \sim P} [R \cap \mathcal{H}_0 \neq \emptyset] \\ &= \mathbb{P}_{\mathbf{X} \sim P} \left[ \bigcup_{h \in \mathcal{H}_0} \{h \in R(\mathbf{X})\} \right] \\ &\leq \sum_{h \in \mathcal{H}_0(P)} \mathbb{P}_{\mathbf{X} \sim P} [p_h \leq \alpha/|\mathcal{H}|] \\ &\leq \pi_0 \alpha . \end{aligned}$$

# Contrôle non-adaptatif : correction de Bonferroni

- ▶ correction de Bonferroni : rejeter

$$R = \{h \in \mathcal{H} : p_h \leq \alpha/|\mathcal{H}|\} .$$

- ▶ alors le FWER est **contrôlé au niveau  $\alpha$**  car

$$\begin{aligned} \text{FWER}(R, P) &= \mathbb{P}_{\mathbf{X} \sim P} [R \cap \mathcal{H}_0 \neq \emptyset] \\ &= \mathbb{P}_{\mathbf{X} \sim P} \left[ \bigcup_{h \in \mathcal{H}_0} \{h \in R(\mathbf{X})\} \right] \\ &\leq \sum_{h \in \mathcal{H}_0(P)} \mathbb{P}_{\mathbf{X} \sim P} [p_h \leq \alpha/|\mathcal{H}|] \\ &\leq \pi_0 \alpha . \end{aligned}$$

# Contrôle non-adaptatif : correction de Bonferroni (pondérée)

- ▶ correction de Bonferroni pondérée : rejeter

$$R = \{h \in \mathcal{H} : p_h \leq \alpha \pi(h)\},$$

où  $\pi$  est une loi de probabilité (discrète) SUR  $\mathcal{H}$ .

- ▶ alors le FWER est contrôlé au niveau  $\alpha$  car

$$\begin{aligned} \text{FWER}(R, P) &= \mathbb{P}_{\mathbf{X} \sim P} [R \cap \mathcal{H}_0 \neq \emptyset] \\ &= \mathbb{P}_{\mathbf{X} \sim P} \left[ \bigcup_{h \in \mathcal{H}_0} \{h \in R(\mathbf{X})\} \right] \\ &\leq \sum_{h \in \mathcal{H}_0(P)} \mathbb{P}_{\mathbf{X} \sim P} [p_h \leq \pi(h)\alpha] \\ &\leq \pi(\mathcal{H}_0)\alpha. \end{aligned}$$

# Bonferroni est “sharp”

- ▶ Dans le cas **(I)** et **(U)** ( $p$ -values indépendantes et  $\sim U([0, 1])$  pour les hypothèses nulles), on a pour un test multiple rejetant au seuil  $t$  :

$$\begin{aligned}\text{FWER}(R, P) &= \mathbb{P}_{\mathbf{x} \sim P} [\exists h \in \mathcal{H}_0 : p_h \leq t] \\ &= 1 - (1 - t)^{|\mathcal{H}_0|};\end{aligned}$$

- ▶ ainsi  $t = 1 - (1 - \alpha)^{1/|\mathcal{H}|}$  (**correction de Šidàk**) donne un contrôle exact au niveau  $1 - (1 - \alpha)^{|\mathcal{H}_0|/|\mathcal{H}|}$
- ▶ **équivalent** au seuil et niveau de Bonferroni quand  $\alpha \rightarrow 0$ .
- ▶ Remarque asymptotique : sous l'hypothèse **(RE)** Bonferroni comme Šidàk ont une puissance tendant vers 0 lorsque  $|\mathcal{H}| \rightarrow \infty$ , et le nombre d'hypothèses rejetées à tort converge vers une variable de Poisson de paramètre  $\pi_0 \alpha$  (Bonferroni) resp.  $-\log(1 - \pi_0 \alpha)$  (Šidàk).

# Bonferroni est “sharp”

- ▶ Dans le cas **(I)** et **(U)** ( $p$ -values indépendantes et  $\sim U([0, 1])$  pour les hypothèses nulles), on a pour un test multiple rejetant au seuil  $t$  :

$$\begin{aligned}\text{FWER}(R, P) &= \mathbb{P}_{\mathbf{x} \sim P} [\exists h \in \mathcal{H}_0 : p_h \leq t] \\ &= 1 - (1 - t)^{|\mathcal{H}_0|};\end{aligned}$$

- ▶ ainsi  $t = 1 - (1 - \alpha)^{1/|\mathcal{H}|}$  (**correction de Šidàk**) donne un contrôle exact au niveau  $1 - (1 - \alpha)^{|\mathcal{H}_0|/|\mathcal{H}|}$
- ▶ **équivalent** au seuil et niveau de Bonferroni quand  $\alpha \rightarrow 0$ .
- ▶ Remarque asymptotique : sous l’hypothèse **(RE)** Bonferroni comme Šidàk ont une puissance tendant vers 0 lorsque  $|\mathcal{H}| \rightarrow \infty$ , et le nombre d’hypothèses rejetées à tort converge vers une variable de Poisson de paramètre  $\pi_0 \alpha$  (Bonferroni) resp.  $-\log(1 - \pi_0 \alpha)$  (Šidàk).

# Procédure “oracle” et plug-in de Bonferroni

- ▶ pour “optimiser” la correction de Bonferroni on voudrait rejeter les hypothèses  $h$  telles que  $p_h \leq \alpha\pi(h)/\pi(\mathcal{H}_0(P))$ .
- ▶ cette procédure a son FWER contrôlé au niveau  $\alpha$ , et est plus puissante car  $\pi(\mathcal{H}_0(P)) \leq 1$ .
- ▶ Problème :  $\pi(\mathcal{H}_0(P)) \dots$  inconnu.
- ▶ supposons que  $\hat{\pi}_0(\mathbf{p})$  est un estimateur de  $\pi_0$  formé à partir de la collection de  $p$ -values
- ▶ on peut considérer le seuil “Bonferroni plug-in” (BPI) :  $\hat{\pi}_0^{-1}\alpha$

# Procédure “oracle” et plug-in de Bonferroni

- ▶ pour “optimiser” la correction de Bonferroni on voudrait rejeter les hypothèses  $h$  telles que  $p_h \leq \alpha\pi(h)/\pi(\mathcal{H}_0(P))$ .
- ▶ cette procédure a son FWER contrôlé au niveau  $\alpha$ , et est plus puissante car  $\pi(\mathcal{H}_0(P)) \leq 1$ .
- ▶ Problème :  $\pi(\mathcal{H}_0(P)) \dots$  inconnu.
- ▶ supposons que  $\hat{\pi}_0(\mathbf{p})$  est un estimateur de  $\pi_0$  formé à partir de la collection de  $p$ -values
- ▶ on peut considérer le seuil “Bonferroni plug-in” (BPI) :  $\hat{\pi}_0^{-1}\alpha$

# Calibration de Bonferroni plug-in (BPI)

- ▶ si  $\hat{\pi}_0(\mathbf{p})$  est croissant en les  $p$ -values, alors sous l'hypothèse **(I)** on a

$$FWER(R_{BPI}, P) \leq FWER(R_{BPI}, DU(n, n - n_0)),$$

où  $DU(n, k)$  est la distribution de  $p$ -values suivante :

- $k$  des  $p$ -values sont indistinctement nulles
  - $(n - k)$  des  $p$ -values sont  $U([0, 1])$
- ▶ on peut donc en principe calibrer cette méthode par simulation en calculant

$$\sup_{n_0 \leq n} FWER(R_{BPI}, DU(n, n - n_0)),$$

(pour développements voir Finner et Gontscharuk 2009)

# Le step-down de Holm

- ▶ Principe du **step-down** : utiliser le contrôle du FWER pour en déduire une borne de confiance supérieure sur  $\pi_0$  et itérer.
- ▶ utiliser la procédure  $R^{(1)}$  avec la correction de Bonferroni : rejeter

$$R^{(1)} = \{h \in \mathcal{H} : p_h \leq \alpha\pi(h)\} .$$

- ▶ itérer en utilisant l'étape précédente comme une borne supérieure sur  $\mathcal{H}_0$  : rejeter

$$R^{(k+1)} = \left\{ h \in \mathcal{H} : p_h \leq \alpha\pi(h)/\pi(R^{(k)}) \right\} .$$

- ▶ Arrêter si  $R^{(k)} = R^{(k+1)}$  .

# Le step-down de Holm

- ▶ Principe du **step-down** : utiliser le contrôle du FWER pour en déduire une borne de confiance supérieure sur  $\pi_0$  et itérer.
- ▶ utiliser la procédure  $R^{(1)}$  avec la correction de Bonferroni : rejeter

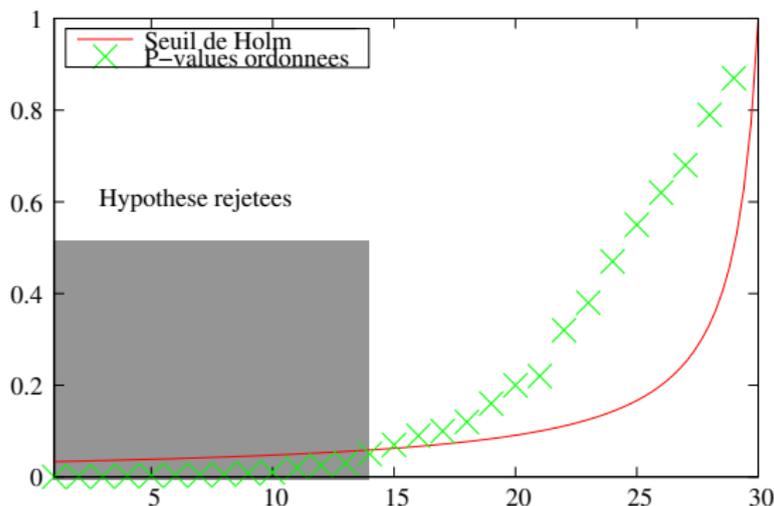
$$R^{(1)} = \{h \in \mathcal{H} : p_h \leq \alpha\pi(h)\} .$$

- ▶ itérer en utilisant l'étape précédente comme une borne supérieure sur  $\mathcal{H}_0$  : rejeter

$$R^{(k+1)} = \left\{ h \in \mathcal{H} : p_h \leq \alpha\pi(h)/\pi(R^{(k)}) \right\} .$$

- ▶ Arrêter si  $R^{(k)} = R^{(k+1)}$  .

# Interprétation graphique ( $\pi$ uniforme)



- ▶  $p^{(1)}, \dots, p^{(d)}$  les  $p$ -values ordonnées
- ▶ seuil de Holm :

$$t_{Holm} = \frac{\alpha}{n+1-k^*}, \quad k^* = \max \left\{ k \leq d : \forall k' \leq k, p^{(k')} \leq \frac{\alpha}{d+1-k'} \right\}$$

# Procédures de step-down

- ▶ considérons un cadre plus général (Romano et Wolf) : soit  $R(\mathcal{C})$  une famille de procédures de test multiple indexée par les sous-ensembles  $\mathcal{C} \subset \mathcal{H}$  et telle que

$$\mathcal{C} \subset \mathcal{C}' \Rightarrow R(\mathcal{C}) \supset R(\mathcal{C}') \quad (\text{p.s.})$$

et pour tout  $P \in \mathfrak{P}$

$$\text{FWER}(R(\mathcal{H}_0(P))) \leq \alpha.$$

- ▶ Le step-down généralisé est alors défini par la suite

$$\mathcal{C}_0 = \mathcal{H}; \quad \mathcal{C}_k = \mathcal{C}_{k-1} \setminus R(\mathcal{C}_{k-1}),$$

s'arrêtant à l'étape  $k^*$  lorsque  $\mathcal{C}_{k^*} = \mathcal{C}_{k^*-1}$  et rejetant  $R(\mathcal{C}_{k^*})$ .

- ▶ alors pour tout  $P \in \mathfrak{P}$

$$\text{FWER}(R(\mathcal{C}_{k^*})) \leq \alpha.$$

# Procédures de step-down (preuve)

- ▶ considérons l'événement

$$\Omega(\mathcal{H}_0) = \{R(\mathcal{H}_0) \cap \mathcal{H}_0 = \emptyset\} .$$

- ▶ Si  $\Omega(\mathcal{H}_0)$  est satisfait, alors par l'hypothèse de monotonie

$$\begin{aligned} \mathcal{C}_k \supset \mathcal{H}_0 &\Rightarrow R(\mathcal{C}_k) \cap \mathcal{H}_0 \subset R(\mathcal{H}_0) \cap \mathcal{H}_0 = \emptyset \\ &\Rightarrow \mathcal{C}_{k+1} = \mathcal{C}_k \setminus R(\mathcal{C}_k) \supset \mathcal{H}_0 . \end{aligned}$$

- ▶ ainsi par récurrence, sur  $\Omega(\mathcal{H}_0)$  on a  $R(\mathcal{C}_{k^*}) \cap \mathcal{H}_0 = \emptyset$  et par l'hypothèse sur la procédure oracle

$$\begin{aligned} \text{FWER}(R(\mathcal{C}_{k^*})) &= \mathbb{P}[\Omega(\mathcal{H}_0) \cap \{R(\mathcal{C}_{k^*}) \cap \mathcal{H}_0 = \emptyset\}] + \mathbb{P}[\Omega(\mathcal{H}_0)^c] \\ &\leq \text{FWER}(R(\mathcal{H}_0)) \leq \alpha . \end{aligned}$$

- 1 Introduction
- 2 Contrôle du FWER et adaptativité à  $\pi_0$
- 3 Adaptativité à la dépendance et régions de confiance**
  - Résultats théoriques
  - Quelques simulations

- ▶ On a vu que le seuil de Bonferroni est approximativement “sharp” dans le cas de  $p$ -values indépendantes, ce qui correspond donc au “pire” cas
- ▶ Peut-on tenir compte d'information supplémentaire sur la dépendance des  $p$ -values ?
- ▶ Les  $p$ -values sont souvent calculées en utilisant un échantillon  $\mathbf{X} = X_1, \dots, X_n$  i.i.d.
- ▶ On considère un cadre particulier :  $X_i \in \mathbb{R}^d$ , i.i.d., de moyenne  $\mu$ , et on veut tester la nullité de chacune des coordonnées de  $\mu$ .
- ▶ Hypothèses considérées :
  - **(Gaus.)** : les  $X_i$  sont Gaussiens,  $\sigma_k = \text{Var}(X^k)$  connue
  - **(SB)** : la distribution de  $X_i$  est symétrique par rapport à  $\mu$ , et  $|X_i| \leq M$  p.s.
- ▶ On s'intéresse à la statistique de test  $|\overline{X^k}|$  pour chaque coordonnée  $k$ .

# Test par seuillage : procédure oracle

- ▶ Supposons qu'on rejette  $R(t) = \{h \in \mathcal{H} | p_h \leq t\}$  ;
- ▶ On cherche un seuil adéquat  $t$  tel que

$$\mathbb{P} \left[ \sup_{k: \mu_k=0} |\overline{X^k}| > t \right] \leq \alpha .$$

- ▶ sous cette condition, rejeter au seuil  $t$  conduit à un contrôle du FWER au niveau  $\alpha$  .
- ▶ le seuil "oracle" est donc le  $(1 - \alpha)$ -quantile de la variable  $\sup_{k: \mu_k=0} |\overline{X^k}|$  .

# Test par seuillage : procédure oracle

- ▶ Supposons qu'on rejette  $R(t) = \{h \in \mathcal{H} | p_h \leq t\}$  ;
- ▶ On cherche un seuil adéquat  $t$  tel que

$$\mathbb{P} \left[ \sup_{k: \mu_k=0} |\overline{X^k}| > t \right] \leq \alpha .$$

- ▶ sous cette condition, rejeter au seuil  $t$  conduit à un contrôle du FWER au niveau  $\alpha$  .
- ▶ le seuil “oracle” est donc le  $(1 - \alpha)$ -quantile de la variable  $\sup_{k: \mu_k=0} |\overline{X^k}|$  .

# Tests “exacts” par randomisation

- ▶ idée de type “test exact” en utilisant l’hypothèse de symétrie :
- ▶ notons  $X^0$  la projection de  $X$  sur les coordonnées de  $\mathcal{H}_0$  ( ayant une moyenne nulle), alors

$$\mathcal{D}(X^0) = c\mathcal{D}(-X^0);$$

- ▶ considérons  $W = (W_i)_{1 \leq i \leq n} \in \{-1, 1\}$  une famille de signes i.i.d. aléatoires (var. de Rademacher)
- ▶ on note

$$\mathbf{x}^0 \bullet W = (X_1^0 W_1, \dots, X_n^0 W_n),$$

et

$$\overline{X^0}^{(W)} = \overline{\mathbf{x}^0 \bullet W} = \frac{1}{n} \sum_{i=1}^n W_i X_i^0.$$

- ▶ On considère le  $q_\alpha(\mathbf{X}^0)$  le  $(1 - \alpha)$ -quantile de

$$\mathcal{D}(|\overline{X^0}^{(W)}| \mid \mathbf{X}^0),$$

- ▶ alors

$$\begin{aligned} \mathbb{P} \left[ \sup_{k: \mu_k=0} |\overline{X^k}| > q_\alpha(\mathbf{X}^0) \right] &= \mathbb{E}_W \left[ \mathbb{P} \left[ \sup_{k: \mu_k=0} |\overline{X^k}^{(W)}| > q_\alpha(\mathbf{X}^0 \bullet W) \right] \right] \\ &= \mathbb{E}_{\mathbf{X}^0} \left[ \mathbb{P}_W \left[ \sup_{k: \mu_k=0} |\overline{X^k}^{(W)}| > q_\alpha(\mathbf{X}^0) \right] \right] \\ &\leq \alpha \end{aligned}$$

- ▶ ... comme  $\mathcal{H}_0$  est inconnu, on peut prendre a fortiori le seuil calculable

$$q_\alpha(\mathbf{X}) \geq q_\alpha(\mathbf{X}^0).$$

- ▶ Noter qu'on pourra appliquer ici le principe de step-down ! On y reviendra.

# Désavantage du seuil test exact

- ▶ le seuil randomisé  $q_\alpha(\mathbf{X})$  inclut des coordonnées ayant une moyenne non nulle.
- ▶ si la moyenne de ces coordonnées est grande par rapport au bruit, c'est elles qui vont avoir la contribution la plus importante dans ce seuil !
- ▶ un seuil plus adéquat serait  $q_\alpha((\mathbf{X} - \mu))$  ... mais  $\mu$  est inconnu.
- ▶ suggestion naturelle : remplacer  $\mu$  par la moyenne empirique et considérer

$$q_\alpha((\mathbf{X} - \bar{X})),$$

le quantile de randomisation des données empiriquement recentrées.

- ▶ peut-on comparer ce seuil à  $q_\alpha((\mathbf{X} - \mu))$  ?
- ▶ interprétation comme une méthode de rééchantillonnage : on veut "imiter" les variations de

$$(\bar{X} - \mu) \quad \text{par celles de} \quad \overline{(X - \bar{X})}^{(W)} \quad (\text{condt. à } \mathbf{X}).$$

# Désavantage du seuil test exact

- ▶ le seuil randomisé  $q_\alpha(\mathbf{X})$  inclut des coordonnées ayant une moyenne non nulle.
- ▶ si la moyenne de ces coordonnées est grande par rapport au bruit, c'est elles qui vont avoir la contribution la plus importante dans ce seuil !
- ▶ un seuil plus adéquat serait  $q_\alpha((\mathbf{X} - \mu))$  ... mais  $\mu$  est inconnu.
- ▶ suggestion naturelle : remplacer  $\mu$  par la moyenne empirique et considérer

$$q_\alpha((\mathbf{X} - \bar{X})),$$

le quantile de randomisation des données empiriquement recentrées.

- ▶ peut-on comparer ce seuil à  $q_\alpha((\mathbf{X} - \mu))$  ?
- ▶ interprétation comme une méthode de rééchantillonnage : on veut "imiter" les variations de

$$(\bar{X} - \mu) \quad \text{par celles de} \quad \overline{(X - \bar{X})}^{(W)} \quad (\text{condt. à } \mathbf{X}).$$

- ▶ on va considérer plus généralement des seuils de type

$$q_\alpha(\mathbf{X}, \rho) = (1 - \alpha)\text{-quantile de } \mathcal{D} \left( \left\| \bar{X}^{(W)} \right\|_\rho \mid \mathbf{X} \right)$$

où  $\rho \in [1, \infty]$ ;

- ▶ avec pour but d'estimer  $q_\alpha(\mathbf{X} - \mu, \rho)$ .
- ▶ cela permettra notamment d'obtenir des  $\psi$ -régions de confiance pour  $\mu$  du type

$$\mathcal{G}(\mathbf{X}, t) = \left\{ z : \left\| \bar{X} - z \right\|_\rho \leq t \right\}.$$

- ▶ on va considérer plus généralement des seuils de type

$$q_\alpha(\mathbf{X}, \rho) = (1 - \alpha)\text{-quantile de } \mathcal{D} \left( \left\| \bar{X}^{(W)} \right\|_\rho \mid \mathbf{X} \right)$$

où  $\rho \in [1, \infty]$ ;

- ▶ avec pour but d'estimer  $q_\alpha(\mathbf{X} - \mu, \rho)$ .
- ▶ cela permettra notamment d'obtenir des  $\psi$ -regions de confiance pour  $\mu$  du type

$$\mathcal{G}(\mathbf{X}, t) = \left\{ z : \left\| \bar{X} - z \right\|_\rho \leq t \right\}.$$

## Théorème

Soient  $\alpha, \delta, \gamma \in ]0, 1[$  et  $f : \mathbb{R}^{d \times n} \rightarrow \mathbb{R}$  une fonction positive telle que

$$\mathbb{P} \left[ \|\bar{\mathbf{X}} - \mu\|_p > f(\mathbf{X}) \right] \leq \frac{\alpha\gamma}{2};$$

alors le seuil

$$t_{\alpha}^{q+f}(\mathbf{Y}) := q_{\alpha(1-\delta)(1-\gamma)}(\mathbf{X} - \bar{\mathbf{X}}, p) + \sqrt{\frac{2 \log(2/(\delta\alpha))}{n}} f(\mathbf{X})$$

vérifie

$$\mathbb{P} \left[ \|\bar{\mathbf{X}} - \mu\|_p > t_{\alpha}^{q+f}(\mathbf{X}) \right] \leq \alpha.$$

$$t_{\alpha}^{q+f}(\mathbf{X}) = q_{\alpha(1-\delta)(1-\gamma)}(\mathbf{X} - \bar{X}, p) + \sqrt{\frac{2 \log(2/(\delta\alpha))}{n}} f(\mathbf{X})$$

- ▶ la seule hypothèse sur la loi de  $X$  est la symétrie par rapport à la moyenne.
- ▶ pour obtenir un seuil observable, il faut une borne – la fonction  $f$  – sur un quantile “extrême” de cette distribution.
- ▶ la fonction  $f$  apparaît dans un terme de 2e ordre : la borne  $f$  elle-même n’a pas besoin d’être extrêmement précise.
- ▶ en utilisant un principe de Monte-Carlo avec  $B$  tirages (des signes  $W$ ), on perd au plus  $(B + 1)^{-1}$  dans le niveau, et on ne perd rien si  $\alpha(1 - \gamma)(1 - \delta)$  est un multiple de  $(B + 1)^{-1}$ .

# Exemple

$$t_{\alpha}^{q+f}(\mathbf{X}) = q_{\alpha(1-\delta)(1-\gamma)}(\mathbf{X} - \bar{X}, p) + \sqrt{\frac{2 \log(2/(\delta\alpha))}{n}} f(\mathbf{X})$$

- ▶ Niveau cible  $\alpha$  ;
- ▶  $\gamma = \delta = \Theta(n^{-2})$
- ▶ sous **(Gaus.)**, prendre

$$f(\mathbf{X}) = \frac{\|\sigma\|_p}{\sqrt{n}} \bar{\Phi}\left(\frac{\alpha\gamma}{2d}\right) = \mathcal{O}\left(\frac{\log(d) + \log n}{\sqrt{n}}\right),$$

- ▶ le terme quantile randomisé est calculé à un niveau  $\alpha(1 - \Theta(n^{-2}))$  ;
- ▶ le terme résiduel est en

$$\mathcal{O}\left(\frac{\log(d) + \log n}{n}\right)$$

# Rééchantillonnage plus général : concentration

## Théorème

Supposons (**Gaus.**),  $p \in [1, \infty]$ ,  $W$  vecteur de poids aléatoires échangeables, de carré intégrable.

Alors pour tout  $\alpha \in (0, 1)$ , le seuil

$$t_{\alpha}^{\text{conc}}(\mathbf{X}) := \frac{\mathbb{E}_W \left[ \left\| \overline{\mathbf{X}} - \overline{\mathbf{X}}^{(W)} \right\|_p \right]}{B_W} + \frac{\|\sigma\|_p}{\sqrt{n}} \Phi^{-1}(\alpha/2) \left[ \frac{C_W}{\sqrt{n} B_W} + 1 \right]$$

vérifie

$$\mathbb{P} \left[ \left\| \overline{\mathbf{X}} - \mu \right\|_p > t_{\alpha}^{\text{conc}} \right] \leq \alpha .$$

Avec  $\sigma_k^2 = \text{Var}[X_k]$ ,

$$B_W = \mathbb{E} \left[ \left( \frac{1}{n} \sum_{i=1}^n (W_i - \bar{W})^2 \right)^{\frac{1}{2}} \right] ; \quad C_W = \left( \frac{n}{n-1} \mathbb{E} [(W_1 - \bar{W})^2] \right)^{\frac{1}{2}}$$

► Comparaison des moyennes :

$$B_W \mathbb{E} \left[ \|\bar{\mathbf{X}} - \mu\|_p \right] = \mathbb{E} \left[ \left\| \overline{\mathbf{X}}^{(W)} - \mu \right\|_p \right]$$

- Concentration Gaussienne (Cirels'on, Ibragimov and Sudakov 1976)
- Sous **(SB)** (variables symétriques et bornées), on a des résultats comparables en utilisant notamment l'approche de Fromont (2006).

- ▶ Comparaison des moyennes :

$$B_W \mathbb{E} \left[ \|\bar{\mathbf{X}} - \mu\|_p \right] = \mathbb{E} \left[ \left\| \overline{\mathbf{X}}^{(W)} - \mu \right\|_p \right]$$

- ▶ Concentration Gaussienne (Cirels'on, Ibragimov and Sudakov 1976)
- ▶ Sous **(SB)** (variables symétriques et bornées), on a des résultats comparables en utilisant notamment l'approche de Fromont (2006).

# Et pour $\sigma$ ?

- ▶ Peu satisfaisant de postuler une borne a priori sur  $\|\sigma\|_p$
- ▶ On peut considérer la variance empirique

$$\hat{\sigma} = \left( \sqrt{\frac{1}{n} \sum_{i=1}^n (X_k^i - \bar{\mathbf{X}}_k)^2} \right)_{1 \leq k \leq d}$$

## Proposition

Sous **(Gaus.)**, avec probabilité au moins  $1 - \delta$  :

$$\|\sigma\|_p \leq \left( C_n - n^{-\frac{1}{2}} \bar{\Phi}(\delta/2) \right)^{-1} \|\hat{\sigma}\|_p,$$

pour une constante explicite  $C_n = 1 - O(n^{-1})$ .

# Et pour $\sigma$ ?

- ▶ Peu satisfaisant de postuler une borne a priori sur  $\|\sigma\|_p$
- ▶ On peut considérer la variance empirique

$$\hat{\sigma} = \left( \sqrt{\frac{1}{n} \sum_{i=1}^n (X_k^i - \bar{\mathbf{X}}_k)^2} \right)_{1 \leq k \leq d}$$

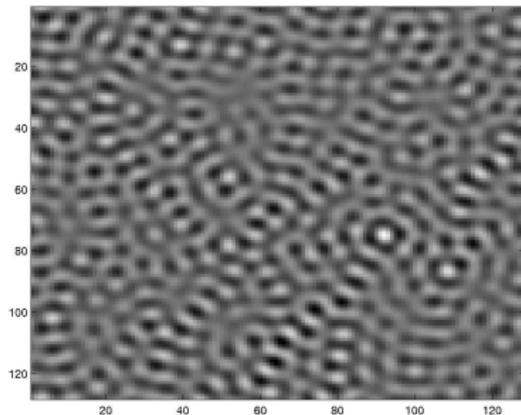
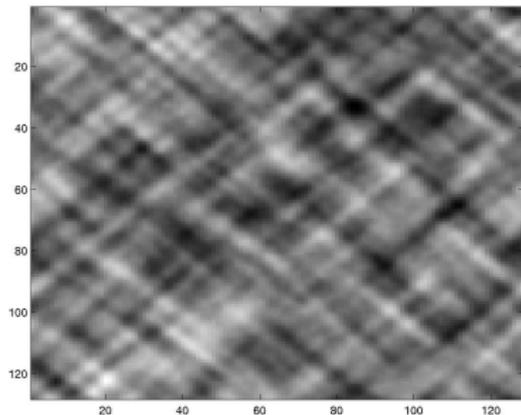
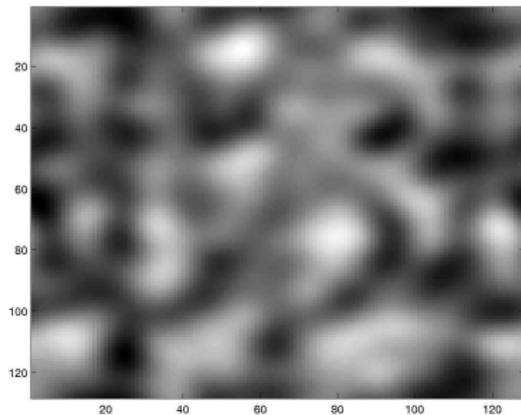
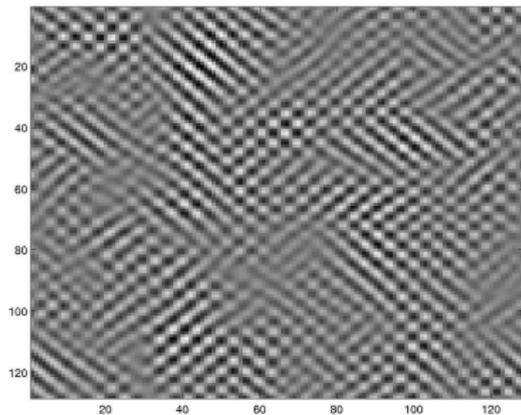
## Proposition

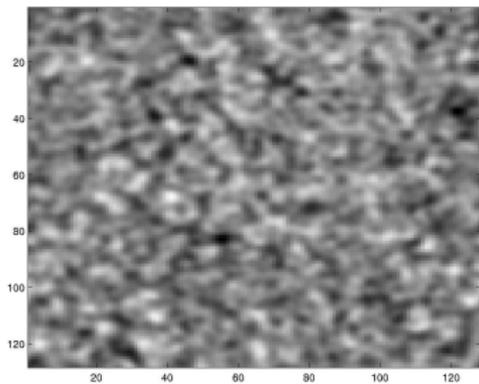
Sous **(Gaus.)**, avec probabilité au moins  $1 - \delta$  :

$$\|\sigma\|_p \leq \left( C_n - n^{-\frac{1}{2}} \bar{\Phi}(\delta/2) \right)^{-1} \|\hat{\sigma}\|_p ,$$

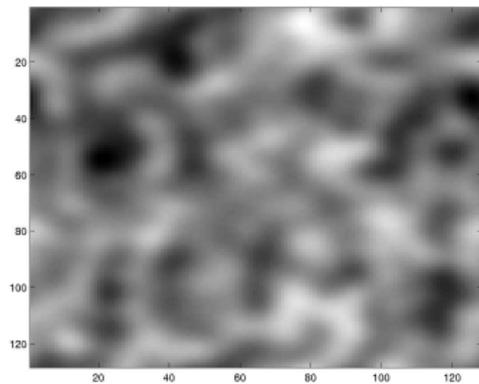
pour une constante explicite  $C_n = 1 - O(n^{-1})$ .

- ▶ différents résultats non-asymptotiques pour les régions de confiance issues de rééchantillonnage en grande dimension
- ▶ termes résiduels
- ▶ peut donner lieu à des études asymptotiques “non classiques” ( $d(n) \gg n$ )

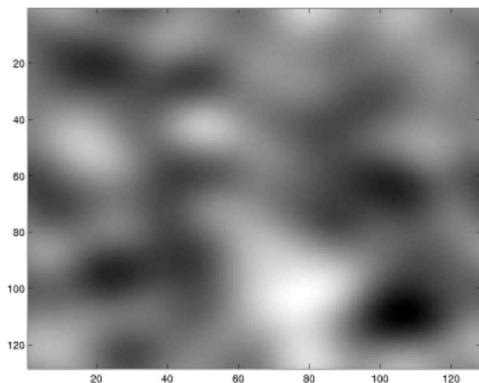




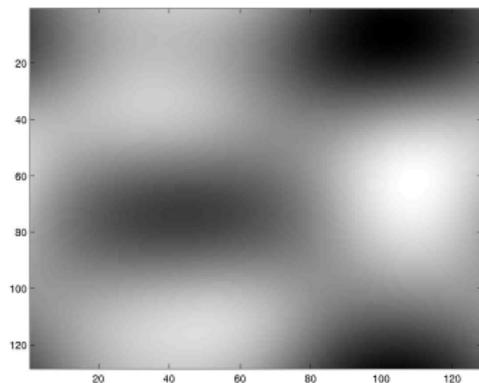
**b=2**



**b=6**

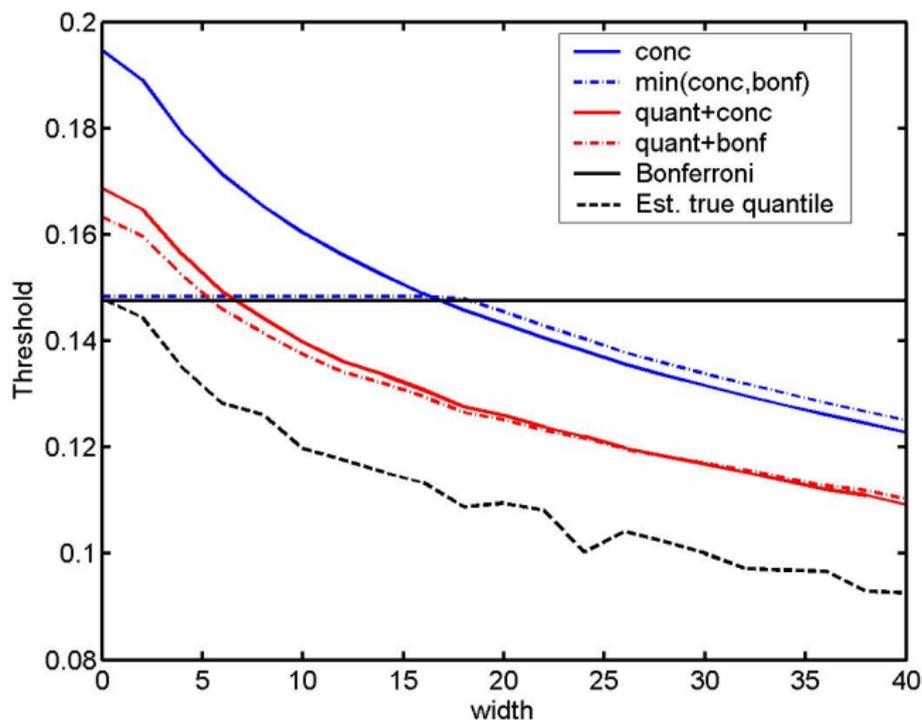


**b=12**

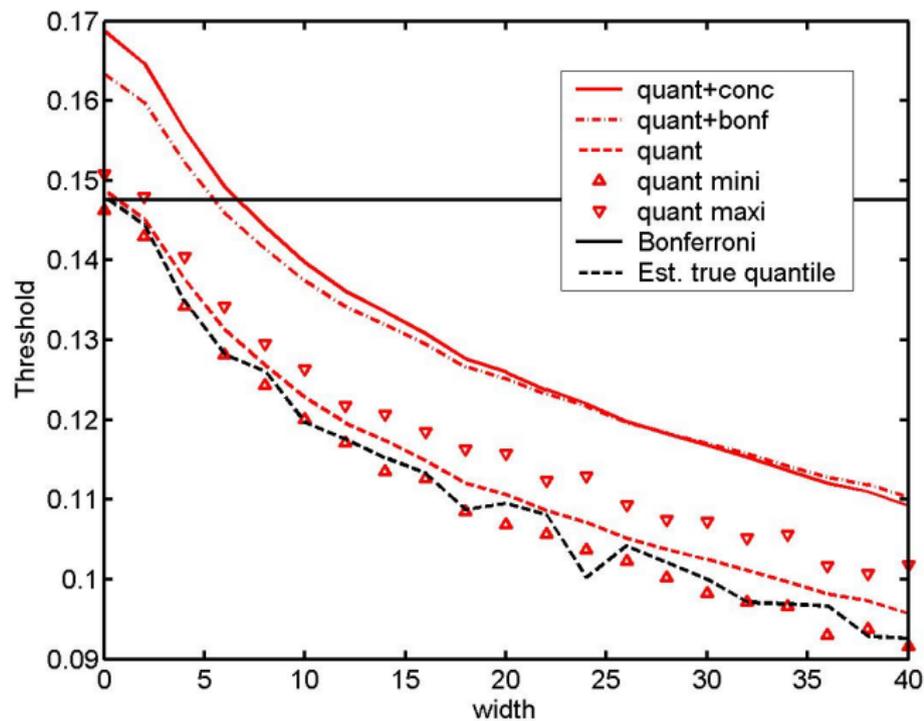


**b=30**

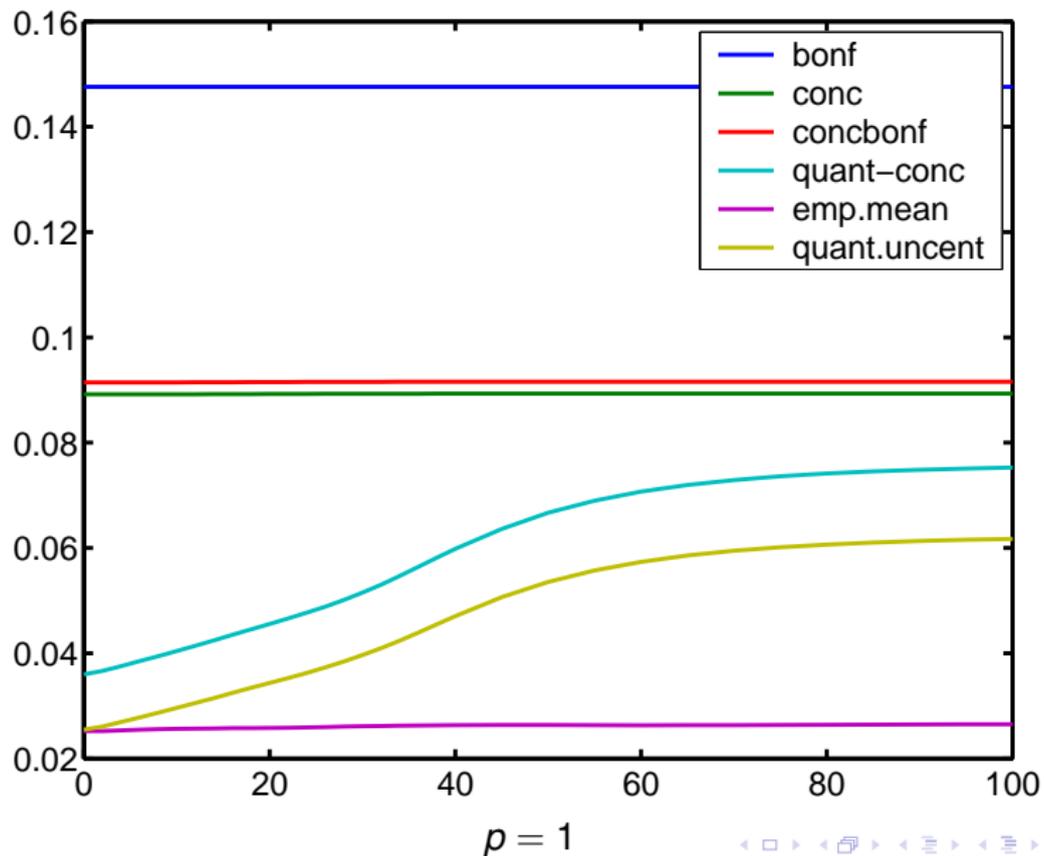
Simulations :  $n=1000$ ,  $d = 128^2$ ,  $\sigma = 1$ ,  $\|\cdot\|_\infty$



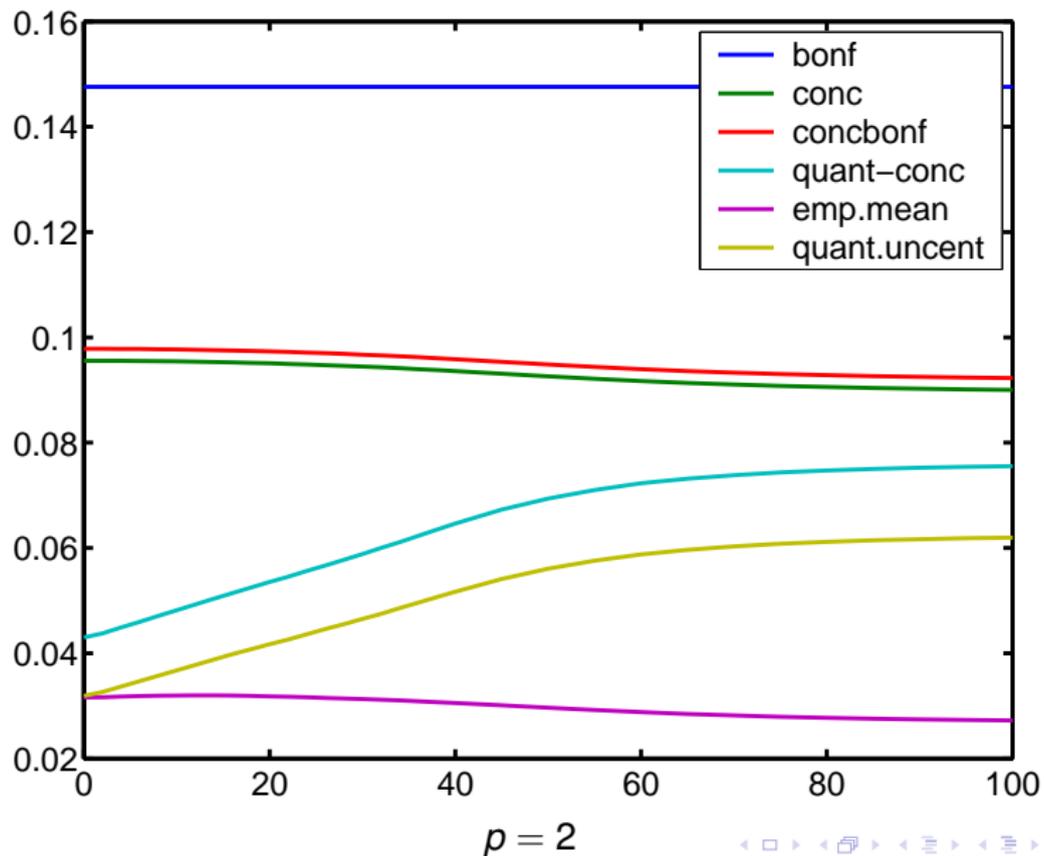
# Simulations : en oubliant les termes résiduels



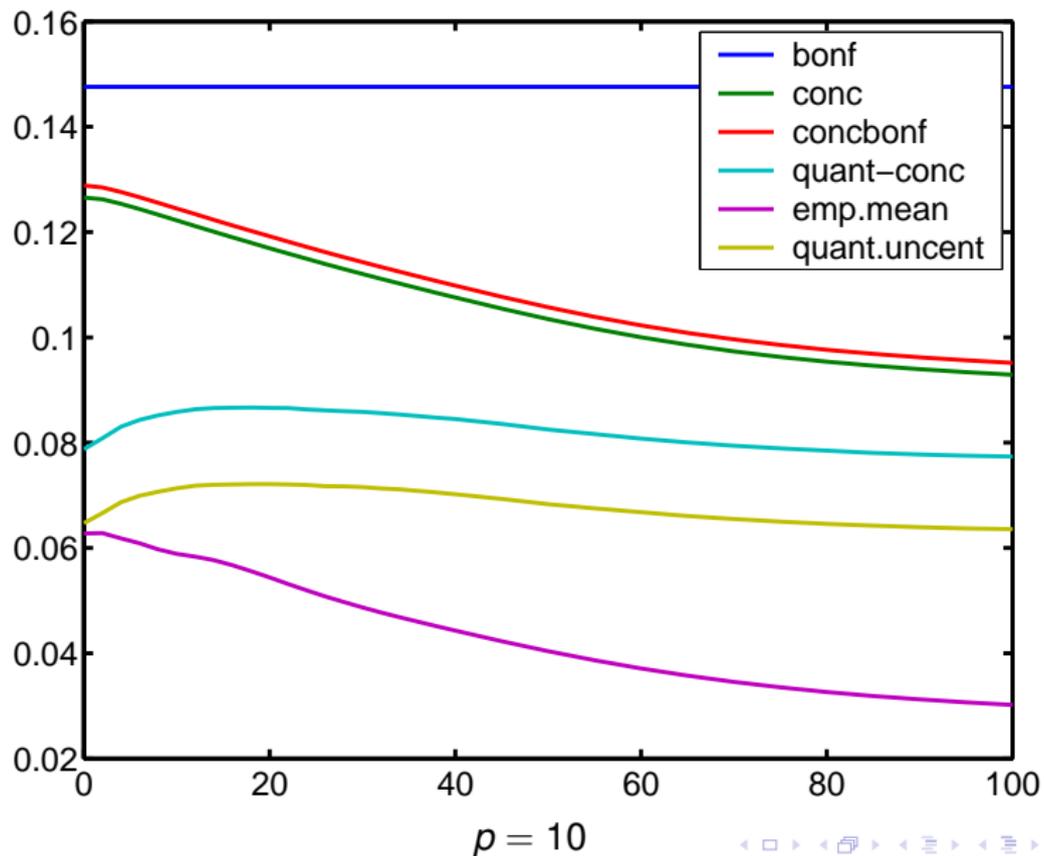
# Simulations : $n=1000$ , $\|\cdot\|_p$



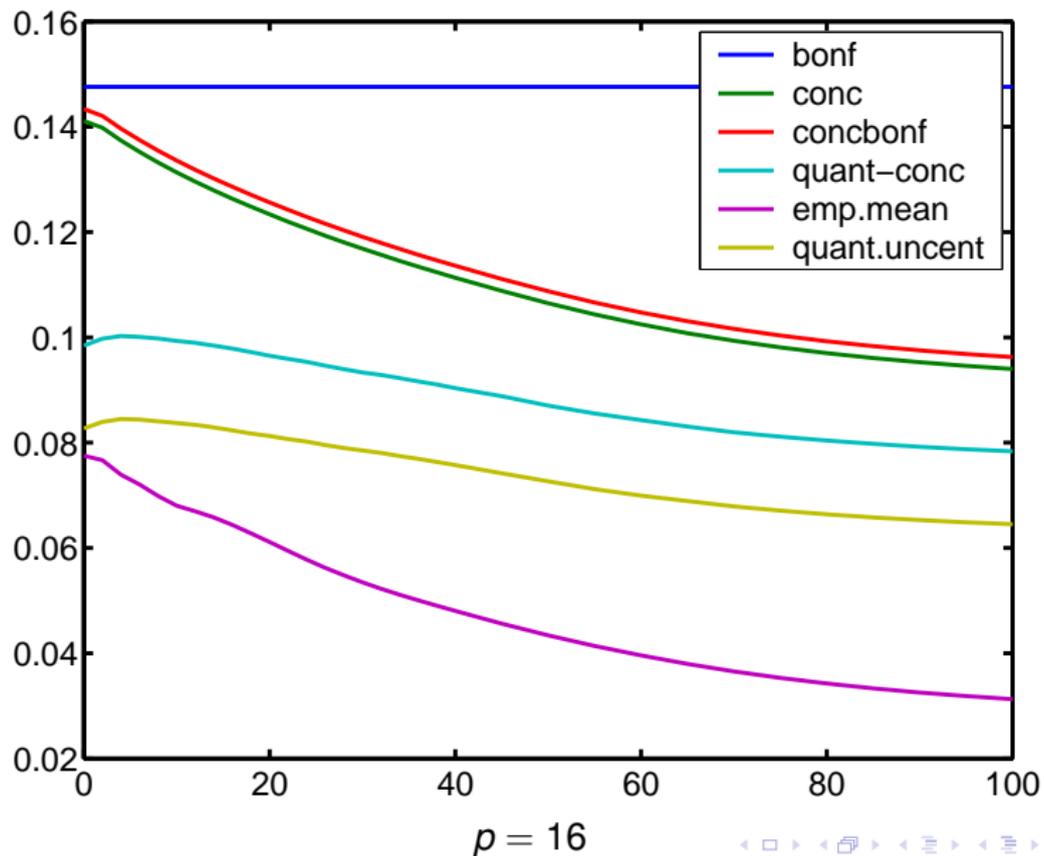
# Simulations : $n=1000$ , $\|\cdot\|_p$



# Simulations : $n=1000$ , $\|\cdot\|_p$



# Simulations : $n=1000$ , $\|\cdot\|_p$



# Où l'on revient enfin aux tests multiples

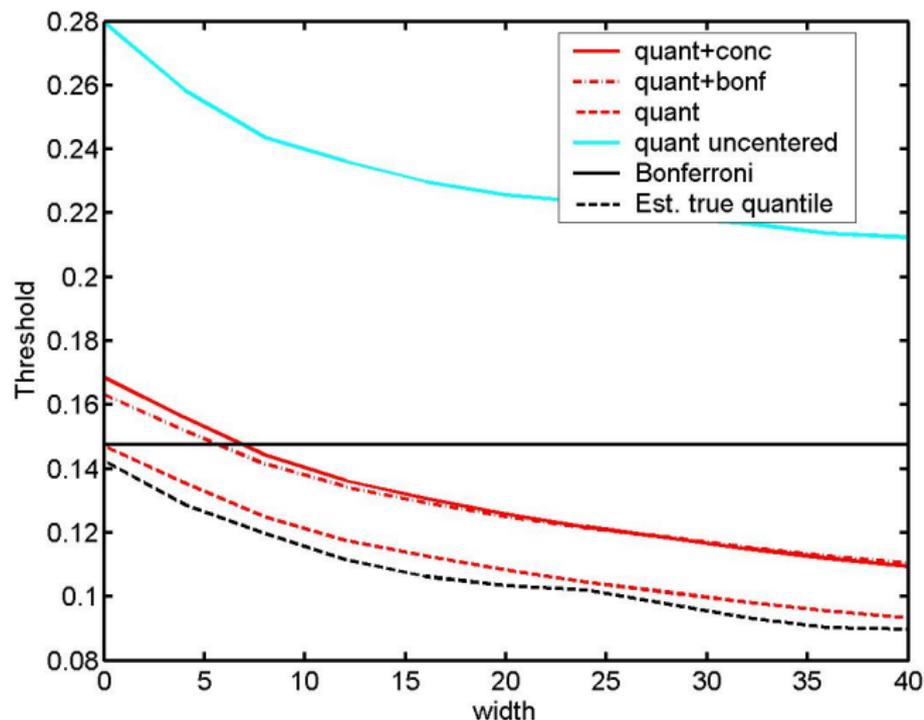
---

- ▶ rappel : en test les différents seuils obtenus ci-dessus ont un concurrent naturel, le quantile randomisé non-recentré :

$$t_{\alpha}^*(\mathbf{X}) = q_{\alpha}(\mathbf{X}) ;$$

- ▶ différences avec les seuils obtenus pour des régions de confiance :
  - pas de termes résiduels ni de diminution du niveau pour le calcul du quantile
  - pas de recentrage empirique  $\Rightarrow$  influence des moyennes non-nulles sur le seuil.
- ▶ de plus, tous les seuils considérés peuvent être utilisés dans un principe de step-down.

# Simulations : moyennes non nulles $\mu_k \in [0, 3]$

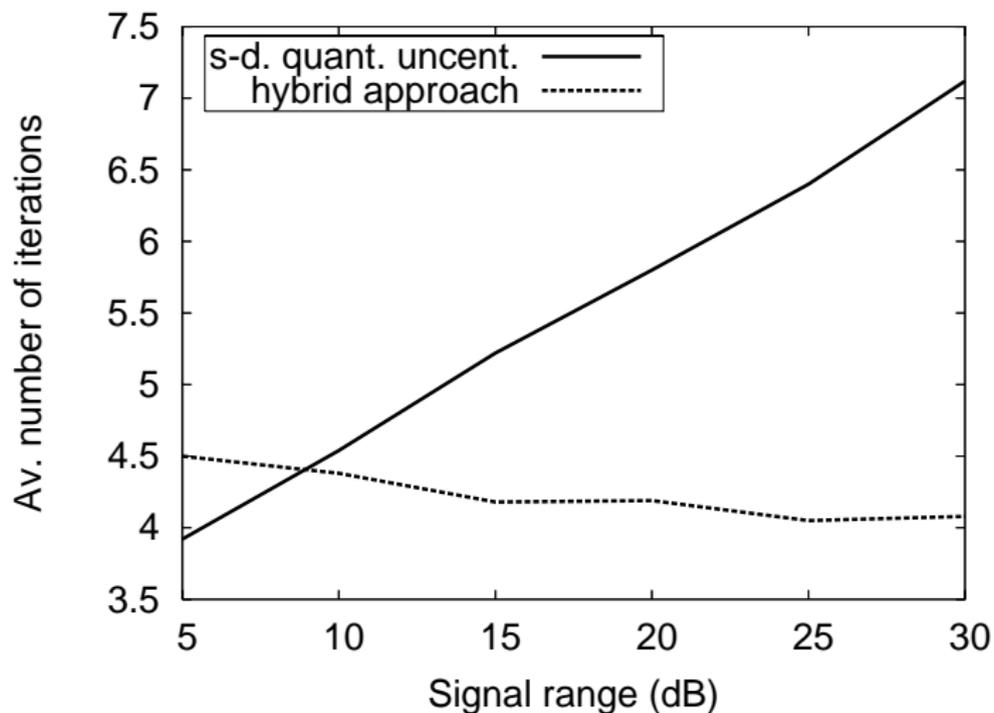


# Step-down hybride

---

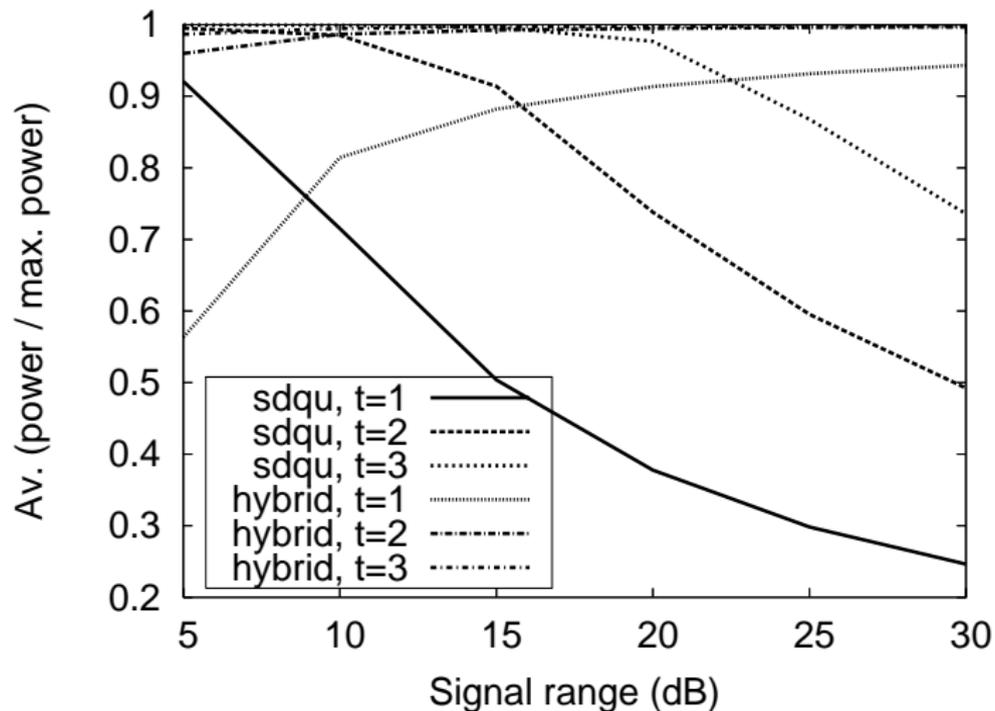
- ▶ Le seuil 'quantile randomisé non-recentré' devient plus performant au fur et à mesure des itérations de step-down car les moyennes les plus grandes sont éliminées
- ▶ Le seuil 'quantile randomisé recentré' reste utile pour éliminer une majorité des moyennes non-nulles en une étape
- ▶ **Step-down hybride** : appliquer à la première étape le seuil recentré, puis pour les itérations suivantes le seuil non-recentré calculé sur les hypothèses restantes
- ▶ Intérêt : accélération de la procédure – en particulier du fait qu'une seule étape de rééchantillonnage peut être coûteuse en temps de calcul

# Simulations : $n = 100$ , $b = 30$ , $d = 128^2$



Nombre moyen d'itérations, procédures complètes

# Simulations : $n = 100$ , $b = 30$ , $d = 128^2$



Puissance des procédures avec arrêt anticipé

# Références bibliographiques

---



J.P. Romano et M. Wolf

Exact and approximate stepdown methods for multiple hypothesis testing.  
JASA 100(469) (2005) 94- 108



S.Arlot, G. Blanchard, E.Roquain

Some non-asymptotic results on resampling in high dimension,  
I : confidence regions  
II : multiple tests  
Annals of Statistics, to appear