Resampling-based confidence regions and multiple tests for a correlated random vector

Sylvain Arlot[1,2]
joint work with Gilles Blanchard[3] Étienne Roquain[4]

[1]Université Paris-Sud XI, Orsay

[2]Projet Select, Inria-Futurs

[3]Fraunhofer FIRST.IDA, Berlin

[4]INRA, Jouy-en-Josas

Pascal Workshop and Pascal Challenge
Type I and Type II errors for Multiple Simultaneous Hypothesis Testing
May 16, 2007

Model

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

$Y^1, \ldots, Y^n \in \mathbb{R}^K$ i.i.d. symmetric, e.g. $\mathcal{N}(\mu, \Sigma)$

- Unknown mean $\mu = (\mu_k)_k$
- Unknown covariance matrix Σ
- Known upper bound $\sigma^2 \geq \max_k \mathrm{var}(Y_k^1)$
- $n << K$.

Aims : Find $\{k \text{ s.t. } \mu_k \neq 0\}$. + Confidence region for μ

Model

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. symmetric, e.g. $\mathcal{N}(\mu, \Sigma)$

- Unknown mean $\mu = (\mu_k)_k$
- Unknown covariance matrix Σ
- Known upper bound $\sigma^2 \geq \max_k \operatorname{var}(Y_k^1)$
- $n << K$.

Aims : Find $\{k \text{ s.t. } \mu_k \neq 0\}$. + Confidence region for μ

Model

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

$Y^1, \ldots, Y^n \in \mathbb{R}^K$ i.i.d. symmetric, e.g. $\mathcal{N}(\mu, \Sigma)$

- Unknown mean $\mu = (\mu_k)_k$
- Unknown covariance matrix Σ
- Known upper bound $\sigma^2 \geq \max_k \operatorname{var}(Y_k^1)$
- $n << K$.

Aims : Find $\{k \text{ s.t. } \mu_k \neq 0\}$. + Confidence region for μ

Model

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. symmetric, e.g. $\mathcal{N}(\mu, \Sigma)$

- Unknown mean $\mu = (\mu_k)_k$
- Unknown covariance matrix Σ
- Known upper bound $\sigma^2 \geq \max_k \operatorname{var}(Y_k^1)$
- $n << K$.

Aims : Find $\{k \text{ s.t. } \mu_k \neq 0\}$. + Confidence region for μ

Model

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. symmetric, e.g. $\mathcal{N}(\mu, \Sigma)$

- Unknown mean $\mu = (\mu_k)_k$
- Unknown covariance matrix Σ
- Known upper bound $\sigma^2 \geq \max_k \operatorname{var}(Y_k^1)$
- $n << K$.

Aims : Find $\{k \text{ s.t. } \mu_k \neq 0\}$. + Confidence region for μ

Model

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

$Y^1, \ldots, Y^n \in \mathbb{R}^K$ i.i.d. symmetric, e.g. $\mathcal{N}(\mu, \Sigma)$

- Unknown mean $\mu = (\mu_k)_k$
- Unknown covariance matrix Σ
- Known upper bound $\sigma^2 \geq \max_k \operatorname{var}(Y_k^1)$
- $n << K$.

Aims : Find $\{k \text{ s.t. } \mu_k \neq 0\}$. + Confidence region for μ

Model

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. symmetric, e.g. $\mathcal{N}(\mu, \Sigma)$

- Unknown mean $\mu = (\mu_k)_k$
- Unknown covariance matrix Σ
- Known upper bound $\sigma^2 \geq \max_k \mathrm{var}(Y_k^1)$
- $n << K$.

Aims : Find $\{k \text{ s.t. } \mu_k \neq 0\}$. + Confidence region for μ

Example : (neuro)images

- K pixels, each pixel k has an intensity Y_k
- n repetitions (independent copies)
- $n \ll K$

$\Rightarrow$ where is there some signal ?

- Spatial dependence $\Rightarrow$ correlations
- The true distance may be unknown
- Distant correlations are possible (non-markovian noise).

$\Rightarrow$ unknown correlations

n series of images $\Rightarrow$ spatial and time correlations

Example : (neuro)images

- K pixels, each pixel k has an intensity Y_k
- n repetitions (independent copies)
- $n \ll K$

$\Rightarrow$ where is there some signal ?

- Spatial dependence $\Rightarrow$ correlations
- The true distance may be unknown
- Distant correlations are possible (non-markovian noise).

$\Rightarrow$ unknown correlations

n series of images $\Rightarrow$ spatial and time correlations

Example : (neuro)images

- K pixels, each pixel k has an intensity Y_k
- n repetitions (independent copies)
- $n \ll K$

$\Rightarrow$ where is there some signal ?

- Spatial dependence $\Rightarrow$ correlations
- The true distance may be unknown
- Distant correlations are possible (non-markovian noise).

$\Rightarrow$ unknown correlations

n series of images $\Rightarrow$ spatial and time correlations

Example : (neuro)images

- K pixels, each pixel k has an intensity Y_k
- n repetitions (independent copies)
- $n \ll K$

$\Rightarrow$ where is there some signal ?

- Spatial dependence $\Rightarrow$ correlations
- The true distance may be unknown
- Distant correlations are possible (non-markovian noise).

$\Rightarrow$ unknown correlations

n series of images $\Rightarrow$ spatial and time correlations

Example : (neuro)images

- K pixels, each pixel k has an intensity Y_k
- n repetitions (independent copies)
- $n \ll K$

$\Rightarrow$ where is there some signal ?

- Spatial dependence $\Rightarrow$ correlations
- The true distance may be unknown
- Distant correlations are possible (non-markovian noise).

$\Rightarrow$ unknown correlations

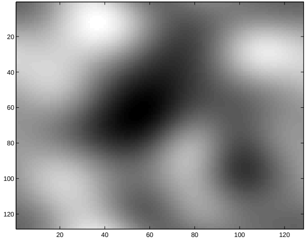
n series of images $\Rightarrow$ spatial and time correlations

Example : simulation experiment

$$Y_t = \mu_t + G_t \qquad t \in \mathbb{T}_m^2 = (\mathbb{Z}/m\mathbb{Z})^2 \qquad K = m^2$$

- $G = N * F$
- N = white noise on $\mathbb{T}_m^2$
- $F : \mathbb{T}_m^2 \to \mathbb{R}$ such that $\sum_t F(t)^2 = 1$.

$\Rightarrow$ G stationary Gaussian process on $\mathbb{T}_m^2$, centered, with variance 1.



$$F_b(t) = C_b \exp\left(-d_{\mathbb{T}_m^2}(0,t)^2/b^2\right)$$

b = bandwidth

correlations increase with b.

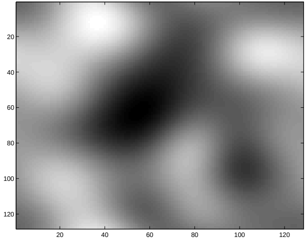
$(b = 20, \ m = 128)$

Example : simulation experiment

$$Y_t = \mu_t + G_t \qquad t \in \mathbb{T}_m^2 = (\mathbb{Z}/m\mathbb{Z})^2 \qquad K = m^2$$

- $G = N * F$
- $N =$ white noise on $\mathbb{T}_m^2$
- $F : \mathbb{T}_m^2 \to \mathbb{R}$ such that $\sum_t F(t)^2 = 1$.

$\Rightarrow$ G stationary Gaussian process on $\mathbb{T}_m^2$, centered, with variance 1.



$$F_b(t) = C_b \exp\left(-d_{\mathbb{T}_m^2}(0,t)^2/b^2\right)$$

$b =$ bandwidth

correlations increase with b.

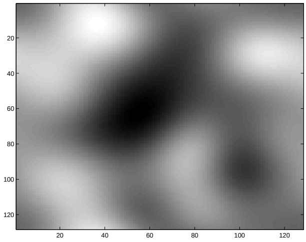
$(b = 20, \ m = 128)$

Example : simulation experiment

$$Y_t = \mu_t + G_t \qquad t \in \mathbb{T}_m^2 = (\mathbb{Z}/m\mathbb{Z})^2 \qquad K = m^2$$

- $G = N * F$
- $N =$ white noise on $\mathbb{T}_m^2$
- $F : \mathbb{T}_m^2 \to \mathbb{R}$ such that $\sum_t F(t)^2 = 1$.

$\Rightarrow$ G stationary Gaussian process on $\mathbb{T}_m^2$, centered, with variance 1.



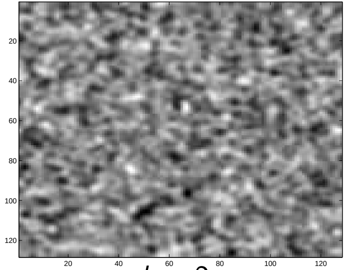
$(b = 20, \ m = 128)$

$$F_b(t) = C_b \exp\left(-d_{\mathbb{T}_m^2}(0,t)^2 / b^2\right)$$

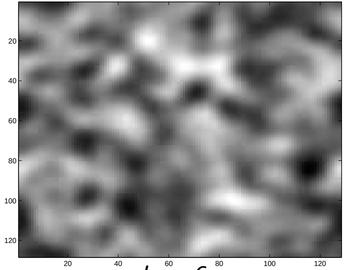
$b =$ bandwidth

correlations increase with b.

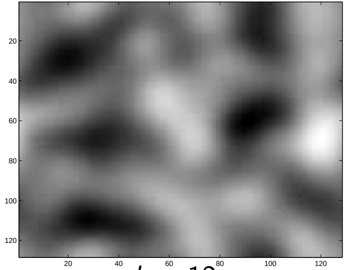
$(G_t)_{t \in \mathbb{T}_m^2}$ with $m = 128$: bandwith $\leftrightarrow$ correlations



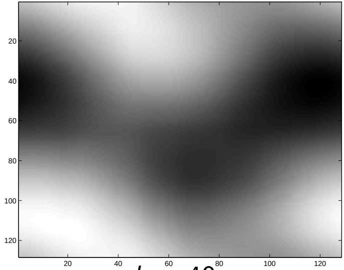
$b = 2$



$b = 6$



$b = 12$



$b = 40$

Multiple Simultaneous Hypothesis Testing

For every k we test : $H_{0,k}$: "$\mu_k = 0$" against $H_{1,k}$: "$\mu_k \neq 0$".

A *multiple testing procedure* rejects :

$$R(\mathbf{Y}) \subset \{1, \dots, K\} .$$

Type I errors measured by the Family Wise Error Rate :

$$\mathrm{FWER}(R) = \mathbb{P}\left(\exists k \in R(\mathbf{Y}) \text{ s.t. } \mu_k = 0\right) .$$

$\Rightarrow$ build a procedure R such that $\mathrm{FWER}(R) \leq \alpha$?

- strong control of the FWER : $\forall \mu \in \mathbb{R}^k$, not only $\mu = 0$
- $|R|$ as large as possible

Multiple Simultaneous Hypothesis Testing

For every k we test : $H_{0,k}$: "$\mu_k = 0$" against $H_{1,k}$: "$\mu_k \neq 0$".

A *multiple testing procedure* rejects :

$$R(\mathbf{Y}) \subset \{1, \ldots, K\}.$$

Type I errors measured by the Family Wise Error Rate :

$$\mathrm{FWER}(R) = \mathbb{P}\left(\exists k \in R(\mathbf{Y}) \text{ s.t. } \mu_k = 0\right).$$

$\Rightarrow$ build a procedure R such that $\mathrm{FWER}(R) \leq \alpha$?

- strong control of the FWER : $\forall \mu \in \mathbb{R}^k$, not only $\mu = 0$
- $|R|$ as large as possible

Multiple Simultaneous Hypothesis Testing

For every k we test : $H_{0,k}$: "$\mu_k = 0$" against $H_{1,k}$: "$\mu_k \neq 0$".

A *multiple testing procedure* rejects :

$$R(\mathbf{Y}) \subset \{1, \ldots, K\}.$$

Type I errors measured by the Family Wise Error Rate :

$$\mathrm{FWER}(R) = \mathbb{P}\left(\exists k \in R(\mathbf{Y}) \text{ s.t. } \mu_k = 0\right).$$

$\Rightarrow$ build a procedure R such that $\mathrm{FWER}(R) \leq \alpha$?

- strong control of the FWER : $\forall \mu \in \mathbb{R}^k$, not only $\mu = 0$
- $|R|$ as large as possible

Multiple Simultaneous Hypothesis Testing

For every k we test : $H_{0,k}$: "$\mu_k = 0$" against $H_{1,k}$: "$\mu_k \neq 0$".

A *multiple testing procedure* rejects :

$$R(\mathbf{Y}) \subset \{1, \dots, K\}.$$

Type I errors measured by the Family Wise Error Rate :

$$\mathrm{FWER}(R) = \mathbb{P}\left(\exists k \in R(\mathbf{Y}) \text{ s.t. } \mu_k = 0\right).$$

$\Rightarrow$ build a procedure R such that $\mathrm{FWER}(R) \leq \alpha$?

- strong control of the FWER : $\forall \mu \in \mathbb{R}^k$, not only $\mu = 0$
- $|R|$ as large as possible

Multiple Simultaneous Hypothesis Testing

For every k we test : $H_{0,k}$: "$\mu_k = 0$" against $H_{1,k}$: "$\mu_k \neq 0$".

A *multiple testing procedure* rejects :

$$R(\mathbf{Y}) \subset \{1, \ldots, K\} .$$

Type I errors measured by the Family Wise Error Rate :

$$\mathrm{FWER}(R) = \mathbb{P}\left(\exists k \in R(\mathbf{Y})\, \text{s.t.}\, \mu_k = 0\right) .$$

$\Rightarrow$ build a procedure R such that $\mathrm{FWER}(R) \leq \alpha$?

- strong control of the FWER : $\forall \mu \in \mathbb{R}^k$, not only $\mu = 0$
- $|R|$ as large as possible

Thresholding

$$R(\mathbf{Y}) = \{k \text{ s.t. } \sqrt{n}|\overline{\mathbf{Y}}_k| > t\},$$

where

- $\overline{\mathbf{Y}} = \frac{1}{n}\sum_{i=1}^{n} Y^i$ empirical mean
- $t = t_\alpha(\mathbf{Y})$ threshold (independent from $k \in \{1, \dots, K\}$).

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

$\Rightarrow$ confidence region *and* control of the FWER

Thresholding

$$R(\mathbf{Y}) = \{k \text{ s.t. } \sqrt{n}|\overline{\mathbf{Y}}_k| > t\},$$

where

- $\overline{\mathbf{Y}} = \frac{1}{n}\sum_{i=1}^{n} Y^i$ empirical mean
- $t = t_\alpha(\mathbf{Y})$ threshold (independent from $k \in \{1, \ldots, K\}$).

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

$\Rightarrow$ confidence region *and* control of the FWER

Bonferroni threshold

Union bound :

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

where $\overline{\Phi}$ is the standard Gaussian upper tail function.

Bonferroni's threshold : $t_\alpha^{\mathrm{Bonf}} = \sigma\overline{\Phi}^{-1}(\alpha/(2K))$.

- deterministic threshold
- too conservative if there are strong correlations between the coordinates Y_k
  ($K \leftrightarrow 1$ if $Y_1 = \cdots = Y_K$)

How to do better ?

Bonferroni threshold

Union bound :

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

where $\overline{\Phi}$ is the standard Gaussian upper tail function.

Bonferroni's threshold : $t_\alpha^{\mathrm{Bonf}} = \sigma \overline{\Phi}^{-1}(\alpha/(2K))$.

- deterministic threshold
- too conservative if there are strong correlations between the coordinates Y_k
  ($K \leftrightarrow 1$ if $Y_1 = \cdots = Y_K$)

How to do better ?

Bonferroni threshold

Union bound :

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

where $\overline{\Phi}$ is the standard Gaussian upper tail function.

Bonferroni's threshold : $t_\alpha^{\mathrm{Bonf}} = \sigma\overline{\Phi}^{-1}(\alpha/(2K))$.

- deterministic threshold
- too conservative if there are strong correlations between the coordinates Y_k
  ($K \leftrightarrow 1$ if $Y_1 = \cdots = Y_K$)

How to do better ?

Bonferroni threshold

Union bound :

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

where $\overline{\Phi}$ is the standard Gaussian upper tail function.

Bonferroni's threshold : $t_\alpha^{\mathrm{Bonf}} = \sigma\overline{\Phi}^{-1}(\alpha/(2K))$.

- deterministic threshold
- too conservative if there are strong correlations between the coordinates Y_k
  $(K \leftrightarrow 1 \text{ if } Y_1 = \cdots = Y_K)$

How to do better ?

Ideal threshold

$$\begin{aligned} \mathrm{FWER}(R) &\leq \mathbb{P}(\exists \quad k \text{ s.t. } \quad \sqrt{n}|\overline{\mathbf{Y}}_k - \mu_k| > t) \\ &\leq \mathbb{P}(\sqrt{n} \sup_k |\overline{\mathbf{Y}}_k - \mu_k| > t) \end{aligned}$$

Ideal threshold : $t = q_\alpha^\star$, $1 - \alpha$ quantile of $\mathcal{L}\left(\sqrt{n} \sup |\overline{\mathbf{Y}} - \mu|\right)$.

$q_\alpha^\star$ depends on Σ, unknown (and $K^2 \gg Kn$)

$\Rightarrow q_\alpha^\star$ estimated by **resampling**.

Ideal threshold

$$\begin{aligned}\mathrm{FWER}(R) &\leq \mathbb{P}(\exists \quad k \text{ s.t. } \quad \sqrt{n}|\overline{\mathbf{Y}}_k - \mu_k| > t) \\ &\leq \mathbb{P}(\sqrt{n} \sup_k |\overline{\mathbf{Y}}_k - \mu_k| > t)\end{aligned}$$

Ideal threshold : $t = q^\star_\alpha$, $1 - \alpha$ quantile of $\mathcal{L}\left(\sqrt{n} \sup |\overline{\mathbf{Y}} - \mu|\right)$.

$q^\star_\alpha$ depends on Σ, unknown (and $K^2 \gg Kn$)

$\Rightarrow q^\star_\alpha$ estimated by **resampling**.

Ideal threshold

$$\begin{aligned} \mathrm{FWER}(R) &\le \mathbb{P}(\exists \quad k \,\text{s.t.} \quad \sqrt{n}|\overline{\mathbf{Y}}_k - \mu_k| > t) \\ &\le \mathbb{P}(\sqrt{n} \sup_k |\overline{\mathbf{Y}}_k - \mu_k| > t) \end{aligned}$$

Ideal threshold : $t = q^\star_\alpha$, $1-\alpha$ quantile of $\mathcal{L}\left(\sqrt{n} \sup |\overline{\mathbf{Y}} - \mu|\right)$.

$q^\star_\alpha$ depends on Σ, unknown (and $K^2 \gg Kn$)

$\Rightarrow q^\star_\alpha$ estimated by **resampling**.

Goal

Find a threshold $t_\alpha(\mathbf{Y})$ such that

$$(\mathcal{P}_\alpha): \quad \mathbb{P}(\sqrt{n}||\overline{\mathbf{Y}} - \mu||_\infty > t_\alpha(\mathbf{Y})) \leq \alpha.$$

If $R = \{k \text{ s.t. } \sqrt{n}|\overline{\mathbf{Y}}_k| > t_\alpha(\mathbf{Y})\}$, FWER$(R) \leq \alpha$.

- $(\mathcal{P}_\alpha)$ gives a confidence ball for μ.
- Non-asymptotic : $\forall K, n$
- $t_\alpha(\mathbf{Y})$ should be close to $q_\alpha^\star$

Goal

Find a threshold $t_\alpha(\mathbf{Y})$ such that

$$(\mathcal{P}_\alpha): \quad \mathbb{P}(\sqrt{n}\|\overline{\mathbf{Y}} - \mu\|_\infty > t_\alpha(\mathbf{Y})) \leq \alpha.$$

If $R = \{k \text{ s.t. } \quad \sqrt{n}|\overline{\mathbf{Y}}_k| > t_\alpha(\mathbf{Y})\}$, $\mathrm{FWER}(R) \leq \alpha$.

- $(\mathcal{P}_\alpha)$ gives a confidence ball for μ.
- Non-asymptotic : $\forall K, n$
- $t_\alpha(\mathbf{Y})$ should be close to $q_\alpha^\star$

Goal

Find a threshold $t_\alpha(\mathbf{Y})$ such that

$$(\mathcal{P}_\alpha): \quad \mathbb{P}(\sqrt{n}||\overline{\mathbf{Y}} - \mu||_\infty > t_\alpha(\mathbf{Y})) \leq \alpha.$$

If $R = \{k \text{ s.t. } \quad \sqrt{n}|\overline{\mathbf{Y}}_k| > t_\alpha(\mathbf{Y})\}$, $\mathrm{FWER}(R) \leq \alpha$.

- $(\mathcal{P}_\alpha)$ gives a confidence ball for μ.
- Non-asymptotic : $\forall K, n$
- $t_\alpha(\mathbf{Y})$ should be close to $q_\alpha^\star$

Goal

Find a threshold $t_\alpha(\mathbf{Y})$ such that

$$(\mathcal{P}_\alpha): \quad \mathbb{P}(\sqrt{n}||\overline{\mathbf{Y}} - \mu||_\infty > t_\alpha(\mathbf{Y})) \leq \alpha.$$

If $R = \{k \text{ s.t. } \quad \sqrt{n}|\overline{\mathbf{Y}}_k| > t_\alpha(\mathbf{Y})\}$, $\text{FWER}(R) \leq \alpha$.

- $(\mathcal{P}_\alpha)$ gives a confidence ball for μ.
- Non-asymptotic : $\forall K, n$
- $t_\alpha(\mathbf{Y})$ should be close to $q^\star_\alpha$

Outline

Resampling principle [Efron 1979 ; ...]

Sample $Y^1, \ldots, Y^n \xrightarrow{\textit{resampling}} (W_1, Y^1), \ldots, (W_n, Y^n)$ weighted sample

- Weight vector : $(W_1, \ldots, W_n)$, independent from $\mathbf{Y}$
- "Y^i is kept W_i times in the resample"
- Example : Efron's bootstrap $\Leftrightarrow$ n-sample with replacement $\Leftrightarrow (W_1, \ldots, W_n) \sim \mathcal{M}(n; n^{-1}, \ldots n^{-1})$

Heuristics : $(\text{true distribution}, \text{sample}) \approx (\text{sample}, \text{resample})$

Resampling principle [Efron 1979 ; ...]

Sample $Y^1, \dots, Y^n \xrightarrow{\text{resampling}} (W_1, Y^1), \dots, (W_n, Y^n)$ weighted sample

- Weight vector : $(W_1, \dots, W_n)$, independent from $\mathbf{Y}$
- "Y^i is kept W_i times in the resample"
- Example : Efron's bootstrap $\Leftrightarrow$ n-sample with replacement $\Leftrightarrow (W_1, \dots, W_n) \sim \mathcal{M}(n; n^{-1}, \dots n^{-1})$

Heuristics : $(\text{true distribution}, \text{sample}) \approx (\text{sample}, \text{resample})$

Resampling principle [Efron 1979 ; ...]

Sample $Y^1, \ldots, Y^n \overset{resampling}{\longrightarrow} (W_1, Y^1), \ldots, (W_n, Y^n)$ weighted sample

- Weight vector : $(W_1, \ldots, W_n)$, independent from $\mathbf{Y}$
- "Y^i is kept W_i times in the resample"
- Example : Efron's bootstrap $\Leftrightarrow$ n-sample with replacement $\Leftrightarrow (W_1, \ldots, W_n) \sim \mathcal{M}(n; n^{-1}, \ldots n^{-1})$

Heuristics : $(\text{true distribution}, \text{sample}) \approx (\text{sample}, \text{resample})$

Resampling principle [Efron 1979 ; ...]

Sample $Y^1, \ldots, Y^n \stackrel{resampling}{\longrightarrow} (W_1, Y^1), \ldots, (W_n, Y^n)$ weighted sample

- Weight vector : $(W_1, \ldots, W_n)$, independent from $\mathbf{Y}$
- "Y^i is kept W_i times in the resample"
- Example : Efron's bootstrap $\Leftrightarrow$ n-sample with replacement
$\Leftrightarrow (W_1, \ldots, W_n) \sim \mathcal{M}(n; n^{-1}, \ldots n^{-1})$

Heuristics : $(\text{true distribution}, \text{sample}) \approx (\text{sample}, \text{resample})$

Examples of resampling weights

- Efron's bootstrap : $(W_1, \dots, W_n) \sim \mathcal{M}(n; n^{-1}, \dots n^{-1})$
- Rademacher : W_i iid $\sim \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$
- Random hold-out : $W_i = 2\mathbb{1}_{i \in I}$, I uniform over subsets of $\{1, \dots, n\}$ of cardinality $n/2$
- Leave-one-out : $W_i = \frac{n}{n-1}\mathbb{1}_{i \neq J}$, $J \sim \mathcal{U}(\{1, \dots, n\})$
- V-fold cross-validation : $W_i = \frac{V}{V-1}\mathbb{1}_{i \notin B_J}$, $J \sim \mathcal{U}(\{1, \dots, V\})$ for some partition $B_1, \dots, B_V$ of $\{1, \dots, n\}$

Examples of resampling weights

- Efron's bootstrap : $(W_1, \ldots, W_n) \sim \mathcal{M}(n; n^{-1}, \ldots n^{-1})$
- Rademacher : W_i iid $\sim \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$
- Random hold-out : $W_i = 2\mathbb{1}_{i \in I}$, I uniform over subsets of $\{1, \ldots, n\}$ of cardinality $n/2$
- Leave-one-out : $W_i = \frac{n}{n-1}\mathbb{1}_{i \neq J}$, $J \sim \mathcal{U}(\{1, \ldots, n\})$
- V-fold cross-validation : $W_i = \frac{V}{V-1}\mathbb{1}_{i \notin B_J}$, $J \sim \mathcal{U}(\{1, \ldots, V\})$ for some partition $B_1, \ldots, B_V$ of $\{1, \ldots, n\}$

Examples of resampling weights

- Efron's bootstrap : $(W_1, \ldots, W_n) \sim \mathcal{M}(n; n^{-1}, \ldots n^{-1})$
- Rademacher : W_i iid $\sim \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$
- Random hold-out : $W_i = 2\mathbb{1}_{i \in I}$, I uniform over subsets of $\{1, \ldots, n\}$ of cardinality $n/2$
- Leave-one-out : $W_i = \frac{n}{n-1}\mathbb{1}_{i \neq J}$, $J \sim \mathcal{U}(\{1, \ldots, n\})$
- V-fold cross-validation : $W_i = \frac{V}{V-1}\mathbb{1}_{i \notin B_J}$, $J \sim \mathcal{U}(\{1, \ldots, V\})$ for some partition $B_1, \ldots, B_V$ of $\{1, \ldots, n\}$

Examples of resampling weights

- Efron's bootstrap : $(W_1, \ldots, W_n) \sim \mathcal{M}(n; n^{-1}, \ldots n^{-1})$
- Rademacher : W_i iid $\sim \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$
- Random hold-out : $W_i = 2\mathbb{1}_{i \in I}$, I uniform over subsets of $\{1, \ldots, n\}$ of cardinality $n/2$
- Leave-one-out : $W_i = \frac{n}{n-1}\mathbb{1}_{i \neq J}$, $J \sim \mathcal{U}(\{1, \ldots, n\})$
- V-fold cross-validation : $W_i = \frac{V}{V-1}\mathbb{1}_{i \notin B_J}$, $J \sim \mathcal{U}(\{1, \ldots, V\})$ for some partition $B_1, \ldots, B_V$ of $\{1, \ldots, n\}$

Examples of resampling weights

- Efron's bootstrap : $(W_1, \dots, W_n) \sim \mathcal{M}(n; n^{-1}, \dots n^{-1})$
- Rademacher : W_i iid $\sim \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$
- Random hold-out : $W_i = 2\mathbb{1}_{i \in I}$, I uniform over subsets of $\{1, \dots, n\}$ of cardinality $n/2$
- Leave-one-out : $W_i = \frac{n}{n-1}\mathbb{1}_{i \neq J}$, $J \sim \mathcal{U}(\{1, \dots, n\})$
- V-fold cross-validation : $W_i = \frac{V}{V-1}\mathbb{1}_{i \notin B_J}$, $J \sim \mathcal{U}(\{1, \dots, V\})$ for some partition $B_1, \dots, B_V$ of $\{1, \dots, n\}$

Quantile method

Ideal threshold : $q_\alpha^\star$, $1-\alpha$ quantile of $\mathcal{L}\left(\sqrt{n}\|\overline{\mathbf{Y}}-\mu\|_\infty\right)$

$\Rightarrow$ Resampling estimate of $q_\alpha^\star$:
$q_\alpha^{\text{quant}}(\mathbf{Y})$, $1-\alpha$ quantile of $\mathcal{L}\left(\sqrt{n}\|\overline{\mathbf{Y}}_W - \overline{W}\,\overline{\mathbf{Y}}\|_\infty|\mathbf{Y}\right)$

$$\overline{\mathbf{Y}}_W := \frac{1}{n}\sum_{i=1}^n W_i Y^i \qquad \text{Resampling empirical mean}$$

$$\overline{W} := \frac{1}{n}\sum_{i=1}^n W_i$$

$q_\alpha^{\text{quant}}(\mathbf{Y})$ depends only on $\mathbf{Y}$, so it can be computed with real data

Quantile method

Ideal threshold : $q_\alpha^\star$, $1-\alpha$ quantile of $\mathcal{L}\left(\sqrt{n}\|\overline{\mathbf{Y}}-\mu\|_\infty\right)$

$\Rightarrow$ Resampling estimate of $q_\alpha^\star$:
$q_\alpha^{\text{quant}}(\mathbf{Y})$, $1-\alpha$ quantile of $\mathcal{L}\left(\sqrt{n}\|\overline{\mathbf{Y}}_W-\overline{W}\,\overline{\mathbf{Y}}\|_\infty|\mathbf{Y}\right)$

$$\overline{\mathbf{Y}}_W := \frac{1}{n}\sum_{i=1}^n W_i Y^i$$ Resampling empirical mean

$$\overline{W} := \frac{1}{n}\sum_{i=1}^n W_i$$

$q_\alpha^{\text{quant}}(\mathbf{Y})$ depends only on $\mathbf{Y}$, so it can be computed with real data

Quantile method

Ideal threshold : $q_\alpha^\star$, $1-\alpha$ quantile of $\mathcal{L}\left(\sqrt{n}\|\overline{\mathbf{Y}}-\mu\|_\infty\right)$

$\Rightarrow$ Resampling estimate of $q_\alpha^\star$:
$q_\alpha^{\text{quant}}(\mathbf{Y})$, $1-\alpha$ quantile of $\mathcal{L}\left(\sqrt{n}\|\overline{\mathbf{Y}}_W-\overline{W}\,\overline{\mathbf{Y}}\|_\infty|\mathbf{Y}\right)$

$$\overline{\mathbf{Y}}_W := \frac{1}{n}\sum_{i=1}^n W_i Y^i$$ Resampling empirical mean

$$\overline{W} := \frac{1}{n}\sum_{i=1}^n W_i$$

$q_\alpha^{\text{quant}}(\mathbf{Y})$ depends only on $\mathbf{Y}$, so it can be computed with real data

Concentration method

- $\|\overline{\mathbf{Y}} - \mu\|_\infty$ concentrates around its expectation, standard-deviation $\leq \sigma n^{-1/2}$
- Estimate $\mathbb{E}\left[\|\overline{\mathbf{Y}} - \mu\|_\infty\right]$ by resampling

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

Works if expectations ($\propto \sqrt{\log(K)}$) are larger than fluctuations ($\propto n^{-1/2}$)

Concentration method

- $\|\overline{\mathbf{Y}} - \mu\|_\infty$ concentrates around its expectation, standard-deviation $\leq \sigma n^{-1/2}$
- Estimate $\mathbb{E}\left[\|\overline{\mathbf{Y}} - \mu\|_\infty\right]$ by resampling

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

Works if expectations ($\propto \sqrt{\log(K)}$) are larger than fluctuations ($\propto n^{-1/2}$)

Concentration method

- $\|\overline{\mathbf{Y}} - \mu\|_\infty$ concentrates around its expectation, standard-deviation $\leq \sigma n^{-1/2}$
- Estimate $\mathbb{E}\left[\|\overline{\mathbf{Y}} - \mu\|_\infty\right]$ by resampling

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

Works if expectations ($\propto \sqrt{\log(K)}$) are larger than fluctuations ($\propto n^{-1/2}$)

Concentration method

- $\|\overline{\mathbf{Y}} - \mu\|_\infty$ concentrates around its expectation, standard-deviation $\leq \sigma n^{-1/2}$
- Estimate $\mathbb{E}\left[\|\overline{\mathbf{Y}} - \mu\|_\infty\right]$ by resampling

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

Works if expectations ($\propto \sqrt{\log(K)}$) are larger than fluctuations ($\propto n^{-1/2}$)

Results from empirical process theory

$$(\overline{\mathbf{Y}}_W - \overline{W}\,\overline{\mathbf{Y}}) = \frac{1}{n}\sum_{i=1}^{n}(W_i - \overline{W})Y^i = \overline{\mathbf{Y}}_{W-\overline{W}}$$

$\Rightarrow$ Empirical bootstrap process

- Asymptotically (K fixed, $n \to \infty$): many results, e.g. [van der Vaart and Wellner 1996]
  $\Rightarrow$ both methods are asymptotically valid.
- Non-asymptotic results (learning theory, bounded case) : Rademacher complexities [Koltchinskii 01], [Bartlett, Boucheron and Lugosi 02]
  Bootstrap penalization [Fromont 05].

Results from empirical process theory

$$(\overline{\mathbf{Y}}_W - \overline{W}\,\overline{\mathbf{Y}}) = \frac{1}{n}\sum_{i=1}^{n}(W_i - \overline{W})Y^i = \overline{\mathbf{Y}}_{W-\overline{W}}$$

$\Rightarrow$ Empirical bootstrap process

- Asymptotically (K fixed, $n \to \infty$): many results, e.g. [van der Vaart and Wellner 1996]
  $\Rightarrow$ both methods are asymptotically valid.
- Non-asymptotic results (learning theory, bounded case) : Rademacher complexities [Koltchinskii 01], [Bartlett, Boucheron and Lugosi 02]
  Bootstrap penalization [Fromont 05].

Results from empirical process theory

$$(\overline{\mathbf{Y}}_W - \overline{W}\,\overline{\mathbf{Y}}) = \frac{1}{n}\sum_{i=1}^{n}(W_i - \overline{W})Y^i = \overline{\mathbf{Y}}_{W-\overline{W}}$$

$\Rightarrow$ Empirical bootstrap process

- Asymptotically (K fixed, $n \to \infty$): many results, e.g. [van der Vaart and Wellner 1996]
  $\Rightarrow$ both methods are asymptotically valid.
- Non-asymptotic results (learning theory, bounded case) :
  Rademacher complexities [Koltchinskii 01], [Bartlett, Boucheron and Lugosi 02]
  Bootstrap penalization [Fromont 05].

Classical procedures and existing results

- Parametric statistics : no, because $K \gg n$.
- Asymptotic results : not valid, because $K \gg n$.
- Randomization tests : do not work if $H_{0,k} = \{\mu_k \leq 0\}$.
  And do not give confidence regions.
- Holm's procedure : unefficient with strong correlations

Classical procedures and existing results

- Parametric statistics : no, because $K \gg n$.
- Asymptotic results : not valid, because $K \gg n$.

- Randomization tests : do not work if $H_{0,k} = \{\mu_k \leq 0\}$.
  And do not give confidence regions.

- Holm's procedure : unefficient with strong correlations

Classical procedures and existing results

- Parametric statistics : no, because $K \gg n$.
- Asymptotic results : not valid, because $K \gg n$.
- Randomization tests : do not work if $H_{0,k} = \{\mu_k \leq 0\}$.
  And do not give confidence regions.
- Holm's procedure : unefficient with strong correlations

Concentration method : three ideas

$$\|\overline{\mathbf{Y}} - \mu\|_\infty \simeq \mathbb{E}\left[\|\overline{\mathbf{Y}} - \mu\|_\infty\right]$$

$$\mathbb{E}\left[\|\overline{\mathbf{Y}} - \mu\|_\infty\right] \propto \mathbb{E}\left[\|\overline{\mathbf{Y}}_W - \overline{W}\,\overline{\mathbf{Y}}\|_\infty\right]$$

$$\mathbb{E}\left[\|\overline{\mathbf{Y}}_W - \overline{W}\,\overline{\mathbf{Y}}\|_\infty | \mathbf{Y}\right] \simeq \mathbb{E}\left[\|\overline{\mathbf{Y}}_W - \overline{W}\,\overline{\mathbf{Y}}\|_\infty\right]$$

$$\Rightarrow q_\alpha^{\text{conc}}(\mathbf{Y}) \propto \sqrt{n}\mathbb{E}\left[\|\overline{\mathbf{Y}}_W - \overline{W}\,\overline{\mathbf{Y}}\|_\infty | \mathbf{Y}\right] + \text{remainders}(\sigma, n, \alpha)$$

Concentration method : three ideas

$$\|\overline{\mathbf{Y}} - \mu\|_\infty \simeq \mathbb{E}\left[\|\overline{\mathbf{Y}} - \mu\|_\infty\right]$$

$$\mathbb{E}\left[\|\overline{\mathbf{Y}} - \mu\|_\infty\right] \propto \mathbb{E}\left[\|\overline{\mathbf{Y}}_W - \overline{W}\,\overline{\mathbf{Y}}\|_\infty\right]$$

$$\mathbb{E}\left[\|\overline{\mathbf{Y}}_W - \overline{W}\,\overline{\mathbf{Y}}\|_\infty | \mathbf{Y}\right] \simeq \mathbb{E}\left[\|\overline{\mathbf{Y}}_W - \overline{W}\,\overline{\mathbf{Y}}\|_\infty\right]$$

$$\Rightarrow q_\alpha^{\mathrm{conc}}(\mathbf{Y}) \propto \sqrt{n}\mathbb{E}\left[\|\overline{\mathbf{Y}}_W - \overline{W}\,\overline{\mathbf{Y}}\|_\infty | \mathbf{Y}\right] + \mathrm{remainders}(\sigma, n, \alpha)$$

Concentration method : three ideas

$$\|\overline{\mathbf{Y}} - \mu\|_\infty \simeq \mathbb{E}\left[\|\overline{\mathbf{Y}} - \mu\|_\infty\right]$$

$$\mathbb{E}\left[\|\overline{\mathbf{Y}} - \mu\|_\infty\right] \propto \mathbb{E}\left[\|\overline{\mathbf{Y}}_W - \overline{W}\,\overline{\mathbf{Y}}\|_\infty\right]$$

$$\mathbb{E}\left[\|\overline{\mathbf{Y}}_W - \overline{W}\,\overline{\mathbf{Y}}\|_\infty | \mathbf{Y}\right] \simeq \mathbb{E}\left[\|\overline{\mathbf{Y}}_W - \overline{W}\,\overline{\mathbf{Y}}\|_\infty\right]$$

$$\Rightarrow q_\alpha^{\text{conc}}(\mathbf{Y}) \propto \sqrt{n}\mathbb{E}\left[\|\overline{\mathbf{Y}}_W - \overline{W}\,\overline{\mathbf{Y}}\|_\infty | \mathbf{Y}\right] + \text{remainders}(\sigma, n, \alpha)$$

First result

Theorem

W exchangeable (Efron, Rademacher, Random hold-out, Leave-one-out). For every $\alpha \in (0;1)$,

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

satisfies

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

with $\sigma^2 := \max_k \mathrm{var}(Y_k^1)$, and

$$B_W := \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^n (W_i - \overline{W})^2\right)^{1/2} > 0 \quad \text{et} \quad C_W := \left[(n/(n-1))\mathbb{E}(W_1 - \overline{W})^2\right]^{1/2}$$

Sketch of the proof

- Expectations :

$$\mathbb{E}\left[\|\overline{\mathbf{Y}} - \mu\|_\infty\right] = B_W^{-1}\mathbb{E}\left[\|\overline{\mathbf{Y}}_W - \overline{W}\,\overline{\mathbf{Y}}\|_\infty\right]$$

- Gaussian concentration theorem for $\|\overline{\mathbf{Y}} - \mu\|_\infty$: standard deviation $\leq \sigma n^{-1/2}$
- Gaussian concentration theorem for $\mathbb{E}\left[\|\overline{\mathbf{Y}}_W - \overline{W}\,\overline{\mathbf{Y}}\|_\infty\right]$: standard deviation $\leq C_W \sigma n^{-1}$

Sketch of the proof

- Expectations :

$$\mathbb{E}\left[\|\overline{\mathbf{Y}} - \mu\|_\infty\right] = B_W^{-1}\mathbb{E}\left[\|\overline{\mathbf{Y}}_W - \overline{W}\,\overline{\mathbf{Y}}\|_\infty\right]$$

- Gaussian concentration theorem for $\|\overline{\mathbf{Y}} - \mu\|_\infty$: standard deviation $\leq \sigma n^{-1/2}$
- Gaussian concentration theorem for $\mathbb{E}\left[\|\overline{\mathbf{Y}}_W - \overline{W}\,\overline{\mathbf{Y}}\|_\infty\right]$: standard deviation $\leq C_W \sigma n^{-1}$

Sketch of the proof

- Expectations :

$$\mathbb{E}\left[\|\overline{\mathbf{Y}} - \mu\|_\infty\right] = B_W^{-1}\mathbb{E}\left[\|\overline{\mathbf{Y}}_W - \overline{W}\,\overline{\mathbf{Y}}\|_\infty\right]$$

- Gaussian concentration theorem for $\|\overline{\mathbf{Y}} - \mu\|_\infty$: standard deviation $\leq \sigma n^{-1/2}$
- Gaussian concentration theorem for $\mathbb{E}\left[\|\overline{\mathbf{Y}}_W - \overline{W}\,\overline{\mathbf{Y}}\|_\infty\right]$: standard deviation $\leq C_W \sigma n^{-1}$

Remarks

- B_W and C_W are independent from K and easy to compute. In most cases, $C_W B_W^{-1} = \mathcal{O}(1)$.
- $\lVert\cdot\rVert_\infty$ can be replaced by $\lVert\cdot\rVert_p$, $p \geq 1$, or by $\sup_k (\cdot)_+$ $\Rightarrow$ different shapes for the confidence regions.
- True for any exchangeable weight vector. Can be generalized to V-fold cross-validation weights (with $C_W B_W^{-1} \approx \sqrt{n/V}$)
- Can be extended (with larger constants) to symmetric bounded variables.
- Almost deterministic threshold.

Remarks

- B_W and C_W are independent from K and easy to compute. In most cases, $C_W B_W^{-1} = \mathcal{O}(1)$.
- $\|\cdot\|_\infty$ can be replaced by $\|\cdot\|_p$, $p \geq 1$, or by $\sup_k (\cdot)_+$ $\Rightarrow$ different shapes for the confidence regions.
- True for any exchangeable weight vector. Can be generalized to V-fold cross-validation weights (with $C_W B_W^{-1} \approx \sqrt{n/V}$)
- Can be extended (with larger constants) to symmetric bounded variables.
- Almost deterministic threshold.

Remarks

- B_W and C_W are independent from K and easy to compute.
  In most cases, $C_W B_W^{-1} = \mathcal{O}(1)$.
- $\|\cdot\|_\infty$ can be replaced by $\|\cdot\|_p$, $p \geq 1$, or by $\sup_k (\cdot)_+$
  $\Rightarrow$ different shapes for the confidence regions.
- True for any exchangeable weight vector.
  Can be generalized to V-fold cross-validation weights (with $C_W B_W^{-1} \approx \sqrt{n/V}$)
- Can be extended (with larger constants) to symmetric bounded variables.
- Almost deterministic threshold.

Remarks

- B_W and C_W are independent from K and easy to compute. In most cases, $C_W B_W^{-1} = \mathcal{O}(1)$.
- $\|\cdot\|_\infty$ can be replaced by $\|\cdot\|_p$, $p \geq 1$, or by $\sup_k (\cdot)_+$ $\Rightarrow$ different shapes for the confidence regions.
- True for any exchangeable weight vector. Can be generalized to V-fold cross-validation weights (with $C_W B_W^{-1} \approx \sqrt{n/V}$)
- Can be extended (with larger constants) to symmetric bounded variables.
- Almost deterministic threshold.

Remarks

- B_W and C_W are independent from K and easy to compute. In most cases, $C_W B_W^{-1} = \mathcal{O}(1)$.
- $\|\cdot\|_\infty$ can be replaced by $\|\cdot\|_p$, $p \geq 1$, or by $\sup_k (\cdot)_+$ $\Rightarrow$ different shapes for the confidence regions.
- True for any exchangeable weight vector. Can be generalized to V-fold cross-validation weights (with $C_W B_W^{-1} \approx \sqrt{n/V}$)
- Can be extended (with larger constants) to symmetric bounded variables.
- Almost deterministic threshold.

Choice of the weights

Accuracy : ratio C_W/B_W in the deviation term

Complexity when computing $\mathbb{E}[\cdot|\overline{\mathbf{Y}}]$: cardinal of the support of $W=(W_i)_i$

Weight	Accuracy C_W/B_W	Complexity
Efron	$\leq \frac{1}{2}(1-\frac{1}{n})^{-n} \xrightarrow[n\to\infty]{} \frac{e}{2}$	n^n
Rademacher	$\leq (1-n^{-1/2})^{-1} \xrightarrow[n\to\infty]{} 1$	2^n
R. h.-o. $(n/2)$	$= \sqrt{\frac{n}{n-1}} \xrightarrow[n\to\infty]{} 1$	$\binom{n}{n/2} \propto n^{-1/2}2^n$
Leave-one-out	$= \sqrt{\frac{n}{n-1}} \xrightarrow[n\to\infty]{} 1$	n
regular V-fold c.-v.	$= \sqrt{\frac{n}{V-1}}$	V

For all exchangable weights, $C_W/B_W \geq \sqrt{n/(n-1)}$.

$\Rightarrow$ **Leave-one-out** and **regular V-fold c.-v.** seem good (V to be chosen).

Choice of the weights

Accuracy : ratio C_W / B_W in the deviation term
Complexity when computing $\mathbb{E}[\cdot | \overline{\mathbf{Y}}]$: cardinal of the support of $W = (W_i)_i$

Weight	Accuracy C_W / B_W	Complexity
Efron	$\leq \frac{1}{2}(1 - \frac{1}{n})^{-n} \xrightarrow[n\to\infty]{} \frac{e}{2}$	n^n
Rademacher	$\leq (1 - n^{-1/2})^{-1} \xrightarrow[n\to\infty]{} 1$	2^n
R. h.-o. $(n/2)$	$= \sqrt{\frac{n}{n-1}} \xrightarrow[n\to\infty]{} 1$	$\binom{n}{n/2} \propto n^{-1/2} 2^n$
Leave-one-out	$= \sqrt{\frac{n}{n-1}} \xrightarrow[n\to\infty]{} 1$	n
regular V-fold c.-v.	$= \sqrt{\frac{n}{V-1}}$	V

For all exchangable weights, $C_W / B_W \geq \sqrt{n/(n-1)}$.
$\Rightarrow$ **Leave-one-out** and **regular V-fold c.-v.** seem good (V to be chosen).

Choice of the weights

Accuracy : ratio C_W/B_W in the deviation term
Complexity when computing $\mathbb{E}[\cdot|\overline{\mathbf{Y}}]$: cardinal of the support of $W=(W_i)_i$

Weight	Accuracy C_W/B_W	Complexity
Efron	$\leq \frac{1}{2}(1-\frac{1}{n})^{-n} \xrightarrow[n\to\infty]{} \frac{e}{2}$	n^n
Rademacher	$\leq (1-n^{-1/2})^{-1} \xrightarrow[n\to\infty]{} 1$	2^n
R. h.-o. $(n/2)$	$= \sqrt{\frac{n}{n-1}} \xrightarrow[n\to\infty]{} 1$	$\binom{n}{n/2} \propto n^{-1/2}2^n$
Leave-one-out	$= \sqrt{\frac{n}{n-1}} \xrightarrow[n\to\infty]{} 1$	n
regular V-fold c.-v.	$= \sqrt{\frac{n}{V-1}}$	V

For all exchangable weights, $C_W/B_W \geq \sqrt{n/(n-1)}$.
$\Rightarrow$ **Leave-one-out** and **regular V-fold c.-v.** seem good (V to be chosen).

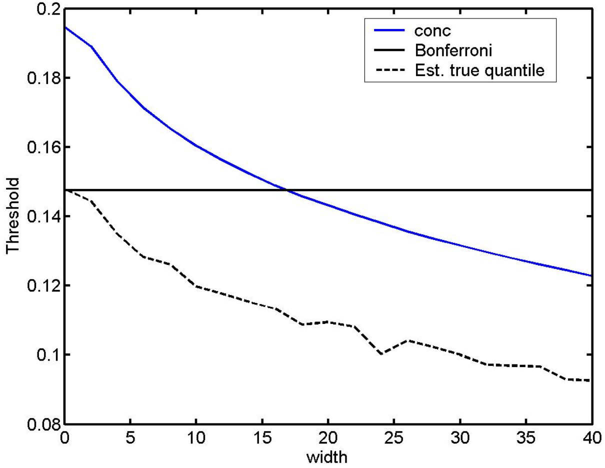
Choice of the weights

Accuracy : ratio C_W/B_W in the deviation term
Complexity when computing $\mathbb{E}[\cdot|\overline{\mathbf{Y}}]$: cardinal of the support of $W = (W_i)_i$

Weight	Accuracy C_W/B_W	Complexity
Efron	$\leq \frac{1}{2}(1-\frac{1}{n})^{-n} \xrightarrow[n\to\infty]{} \frac{e}{2}$	n^n
Rademacher	$\leq (1-n^{-1/2})^{-1} \xrightarrow[n\to\infty]{} 1$	2^n
R. h.-o. $(n/2)$	$= \sqrt{\frac{n}{n-1}} \xrightarrow[n\to\infty]{} 1$	$\binom{n}{n/2} \propto n^{-1/2}2^n$
Leave-one-out	$= \sqrt{\frac{n}{n-1}} \xrightarrow[n\to\infty]{} 1$	n
regular V-fold c.-v.	$= \sqrt{\frac{n}{V-1}}$	V

For all exchangable weights, $C_W/B_W \geq \sqrt{n/(n-1)}$.
$\Rightarrow$ **Leave-one-out** and **regular V-fold c.-v.** seem good (V to be chosen).

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



Second concentration threshold

If $q_\alpha^{\text{conc},1}(\mathbf{Y})$ was constant, we would take $\min\left(q_\alpha^{\text{conc},1}(\mathbf{Y}), t_\alpha^{\mathrm{Bonf}}\right)$ as threshold.

Theorem

Same assumptions as Thm. 1. Then, $\forall \alpha, \delta \in (0;1)$, the threshold $q_\alpha^{\text{conc},2}(\mathbf{Y})$ equal to

$$\min\left(t_{\alpha(1-\delta)}^{Bonf}, \frac{\mathbb{E}\left[\sqrt{n}\|\overline{\mathbf{Y}}_{W-\overline{W}}\|_\infty \middle| \mathbf{Y}\right]}{B_W} + \sigma \overline{\Phi}^{-1}(\alpha(1-\delta)/2) + \frac{\sigma C_W}{\sqrt{n} B_W} \overline{\Phi}^{-1}(\alpha\delta/2) \right)$$

satisfies

$$\mathbb{P}\left(\sqrt{n}\|\overline{\mathbf{Y}} - \mu\|_\infty > q_\alpha^{conc,2}(\mathbf{Y})\right) \leq \alpha$$

$\Rightarrow q_\alpha^{\text{conc},2}(\mathbf{Y})$ always better than $q_\alpha^{\text{conc},1}(\mathbf{Y})$ and Bonferroni threshold (up to δ)

Second concentration threshold

If $q_\alpha^{\text{conc},1}(\mathbf{Y})$ was constant, we would take $\min\left(q_\alpha^{\text{conc},1}(\mathbf{Y}), t_\alpha^{\mathrm{Bonf}}\right)$ as threshold.

Theorem

Same assumptions as Thm. 1. Then, $\forall \alpha, \delta \in (0;1)$, *the threshold* $q_\alpha^{conc,2}(\mathbf{Y})$ *equal to*

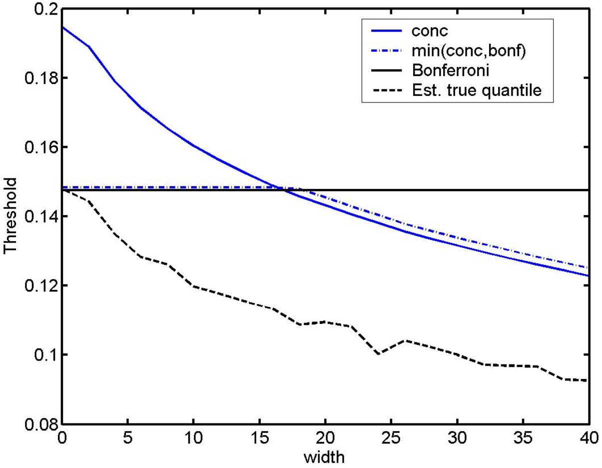
$$\min\left(t^{Bonf}_{\alpha(1-\delta)}, \frac{\mathbb{E}\left[\sqrt{n}\|\overline{\mathbf{Y}}_{W-\overline{W}}\|_\infty \middle| \mathbf{Y}\right]}{B_W} + \sigma\overline{\Phi}^{-1}(\alpha(1-\delta)/2) + \frac{\sigma C_W}{\sqrt{n}B_W}\overline{\Phi}^{-1}(\alpha\delta/2)\right)$$

satisfies

$$\mathbb{P}\left(\sqrt{n}\|\overline{\mathbf{Y}} - \mu\|_\infty > q_\alpha^{conc,2}(\mathbf{Y})\right) \leq \alpha$$

$\Rightarrow q_\alpha^{\text{conc},2}(\mathbf{Y})$ always better than $q_\alpha^{\text{conc},1}(\mathbf{Y})$ and Bonferroni threshold (up to δ)

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



Method

- Rademacher weights : W_i i.i.d. $\sim \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$
- Resampling heuristics suggests that $q_\alpha^{\text{quant}}(\mathbf{Y})$, the $(1-\alpha)$ quantile of
$$\mathcal{L}\left(\sqrt{n}\|\overline{\mathbf{Y}}_W - \overline{W}\,\overline{\mathbf{Y}}\|_\infty \middle| \mathbf{Y}\right)$$
should satisfy $\mathbb{P}\left(\|\overline{\mathbf{Y}} - \mu\|_\infty > q_\alpha^{\text{quant}}(\mathbf{Y})\right) \leq \alpha$.

$$\begin{aligned} q_\alpha^{\text{quant}}(\mathbf{Y}) &= \inf\left\{x \middle| \mathbb{P}_W(\sqrt{n}\|\overline{\mathbf{Y}} - \mu\|_\infty > x) \leq \alpha\right\} \\ &= \inf\left\{x \middle| 2^{-n} \sum_{w \in \{-1,1\}^n} \mathbf{1}\left[\sqrt{n}\left\|\frac{1}{n}\sum_{i=1}^n w_i(Y^i - \overline{\mathbf{Y}})\right\|_\infty > x\right] \leq \alpha\right\} \end{aligned}$$

Theorem

Theorem

Let $\alpha, \delta, \gamma \in (0,1)$ *and* f *a non-negative threshold with FWER bounded by* $\alpha\gamma/2$ *:*

$$\mathbb{P}\left(\sqrt{n}\|\overline{\mathbf{Y}} - \mu\| > f(\mathbf{Y})\right) \leq \frac{\alpha\gamma}{2}$$

Then,

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

has a FWER bounded by α *:*

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

Remarks

- Uses only the symmetry of Y around its mean
- The threshold f only appears in a second-order term.
- Gaussian case $\Rightarrow$ three thresholds :
  take f among $t^{\text{Bonf}}_{\alpha\gamma/2}$, $q^{\text{conc},1}_{\alpha\gamma/2}$ and $q^{\text{conc},2}_{\alpha\gamma/2}$.
- In simulation experiments, f is almost unnecessary.

Remarks

- Uses only the symmetry of Y around its mean
- The threshold f only appears in a second-order term.
- Gaussian case $\Rightarrow$ three thresholds :
  take f among $t^{\mathrm{Bonf}}_{\alpha\gamma/2}$, $q^{\mathrm{conc},1}_{\alpha\gamma/2}$ and $q^{\mathrm{conc},2}_{\alpha\gamma/2}$.
- In simulation experiments, f is almost unnecessary.

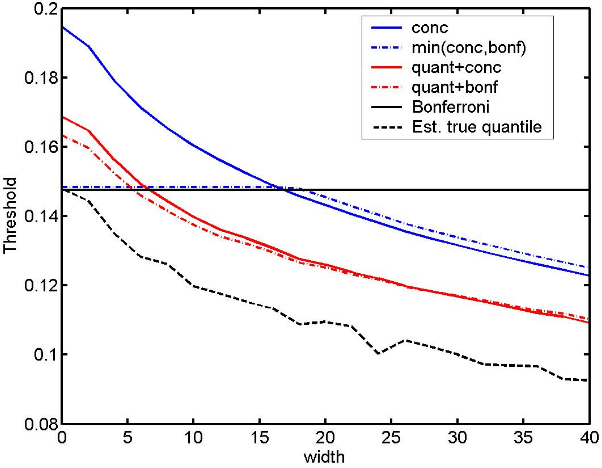
Remarks

- Uses only the symmetry of Y around its mean
- The threshold f only appears in a second-order term.
- Gaussian case $\Rightarrow$ three thresholds :
  take f among $t^{\mathrm{Bonf}}_{\alpha\gamma/2}$, $q^{\mathrm{conc},1}_{\alpha\gamma/2}$ and $q^{\mathrm{conc},2}_{\alpha\gamma/2}$.
- In simulation experiments, f is almost unnecessary.

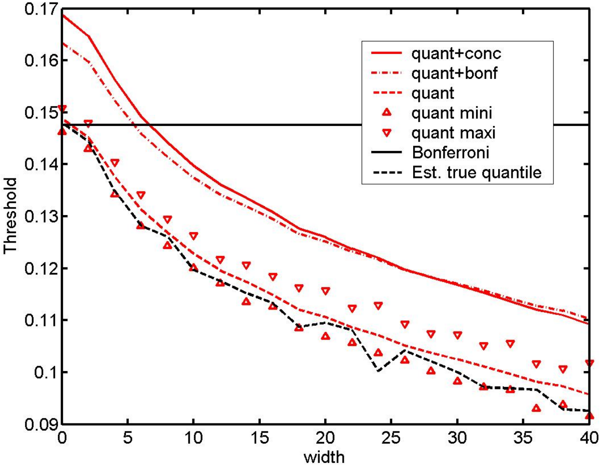
Remarks

- Uses only the symmetry of Y around its mean
- The threshold f only appears in a second-order term.
- Gaussian case $\Rightarrow$ three thresholds :
  take f among $t^{\text{Bonf}}_{\alpha\gamma/2}$, $q^{\text{conc},1}_{\alpha\gamma/2}$ and $q^{\text{conc},2}_{\alpha\gamma/2}$.
- In simulation experiments, f is almost unnecessary.

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

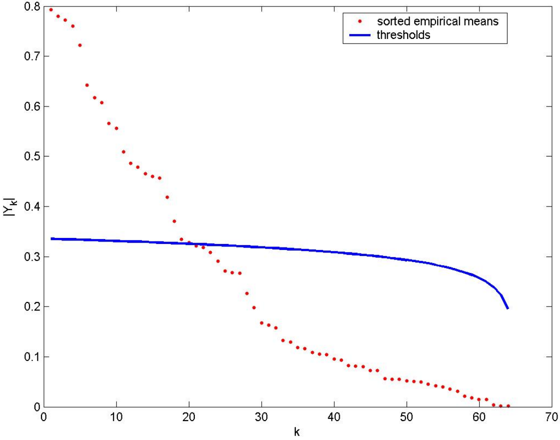


Simulations : without the additive term ?



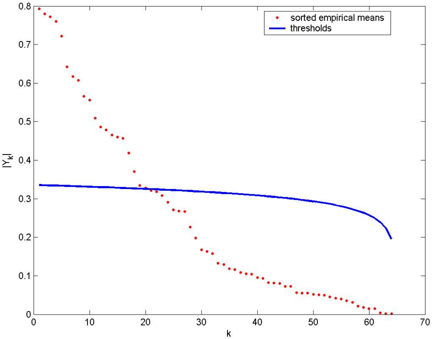
Step-down procedure [Holm 1979 ; Westfall and Young 1993 ; Romano and Wolf 2005]

$$\sup_{1 \le k \le K} \left\{ \overline{\mathbf{Y}}_k - \mu_k \right\} \ge \sup_{\mu_k = 0} \left\{ \overline{\mathbf{Y}}_k - \mu_k \right\}$$



Step-down procedure

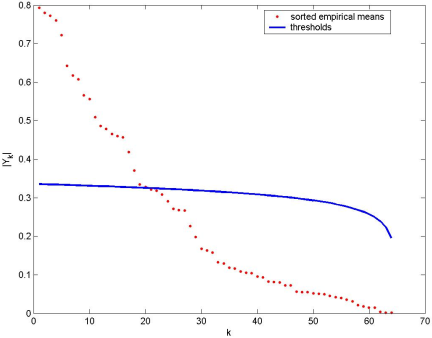
- Reorder the coordinates : $\left|\overline{\mathbf{Y}}_{\sigma(1)}\right| \geq \cdots \geq \left|\overline{\mathbf{Y}}_{\sigma(K)}\right|$
- Define the thresholds $t_k = t(\mathbf{Y}_{\sigma(k)}, \ldots, \mathbf{Y}_{\sigma(K)})$ for $k = 1, \ldots, K$
- Define $\widehat{k} = \max\left\{k \text{ s.t. } \forall k' \leq k, \overline{\mathbf{Y}}_{\sigma(k')} > t_{k'}(\mathbf{Y})\right\}$
- Reject $H_{0,k}$ for all $k \leq \widehat{k}$



$\Rightarrow$ this procedure has a FWER controlled by α if each t_k has (use that $t_{\mathcal{K}} = t((\mathbf{Y}_k)_{k\in\mathcal{K}})$ is a non-decreasing function of $\mathcal{K}$).

Step-down procedure

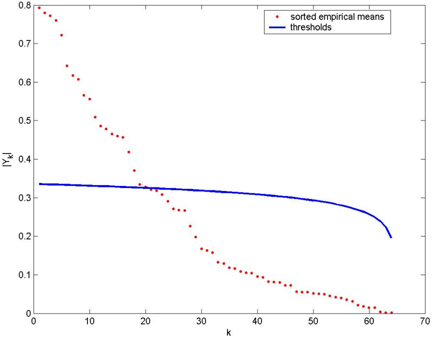
- Reorder the coordinates : $\left|\overline{\mathbf{Y}}_{\sigma(1)}\right| \geq \cdots \geq \left|\overline{\mathbf{Y}}_{\sigma(K)}\right|$
- Define the thresholds $t_k = t(\mathbf{Y}_{\sigma(k)}, \ldots, \mathbf{Y}_{\sigma(K)})$ for $k = 1, \ldots, K$
- Define $\widehat{k} = \max\left\{k \text{ s.t. } \forall k' \leq k, \overline{\mathbf{Y}}_{\sigma(k')} > t_{k'}(\mathbf{Y})\right\}$
- Reject $H_{0,k}$ for all $k \leq \widehat{k}$



$\Rightarrow$ this procedure has a FWER controlled by α if each t_k has (use that $t_{\mathcal{K}} = t((\mathbf{Y}_k)_{k \in \mathcal{K}})$ is a non-decreasing function of $\mathcal{K}$).

Step-down procedure

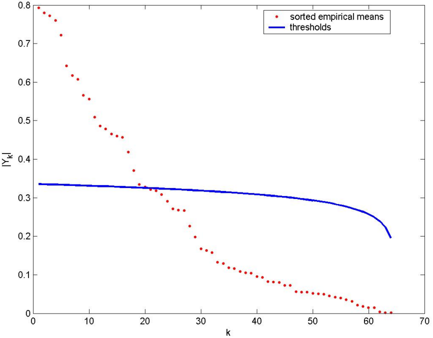
- Reorder the coordinates : $\left|\overline{\mathbf{Y}}_{\sigma(1)}\right| \geq \cdots \geq \left|\overline{\mathbf{Y}}_{\sigma(K)}\right|$
- Define the thresholds $t_k = t(\mathbf{Y}_{\sigma(k)}, \ldots, \mathbf{Y}_{\sigma(K)})$ for $k = 1, \ldots, K$
- Define $\widehat{k} = \max\left\{ k \text{ s.t. } \forall k' \leq k, \overline{\mathbf{Y}}_{\sigma(k')} > t_{k'}(\mathbf{Y}) \right\}$
- Reject $H_{0,k}$ for all $k \leq \widehat{k}$



$\Rightarrow$ this procedure has a FWER controlled by α if each t_k has (use that $t_{\mathcal{K}} = t((\mathbf{Y}_k)_{k\in\mathcal{K}})$ is a non-decreasing function of $\mathcal{K}$).

Step-down procedure

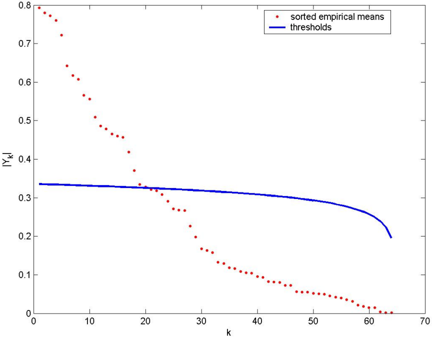
- Reorder the coordinates : $\left|\overline{\mathbf{Y}}_{\sigma(1)}\right| \geq \cdots \geq \left|\overline{\mathbf{Y}}_{\sigma(K)}\right|$
- Define the thresholds $t_k = t(\mathbf{Y}_{\sigma(k)}, \ldots, \mathbf{Y}_{\sigma(K)})$ for $k = 1, \ldots, K$
- Define $\widehat{k} = \max\left\{k \text{ s.t. } \forall k' \leq k, \overline{\mathbf{Y}}_{\sigma(k')} > t_{k'}(\mathbf{Y})\right\}$
- Reject $H_{0,k}$ for all $k \leq \widehat{k}$



$\Rightarrow$ this procedure has a FWER controlled by α if each t_k has (use that $t_{\mathcal{K}} = t((\mathbf{Y}_k)_{k\in\mathcal{K}})$ is a non-decreasing function of $\mathcal{K}$).

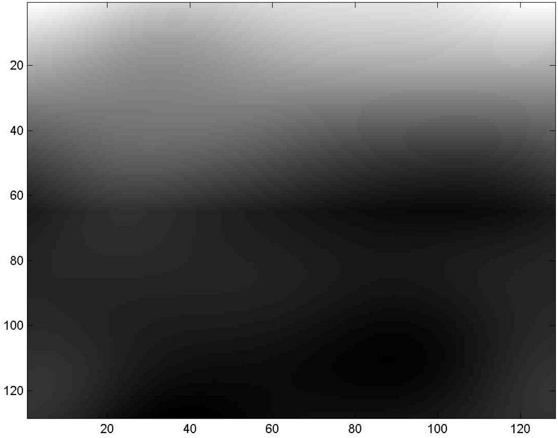
Step-down procedure

- Reorder the coordinates : $\left|\overline{\mathbf{Y}}_{\sigma(1)}\right| \geq \cdots \geq \left|\overline{\mathbf{Y}}_{\sigma(K)}\right|$
- Define the thresholds $t_k = t(\mathbf{Y}_{\sigma(k)}, \ldots, \mathbf{Y}_{\sigma(K)})$ for $k = 1, \ldots, K$
- Define $\widehat{k} = \max\left\{k \text{ s.t. } \forall k' \leq k, \overline{\mathbf{Y}}_{\sigma(k')} > t_{k'}(\mathbf{Y})\right\}$
- Reject $H_{0,k}$ for all $k \leq \widehat{k}$

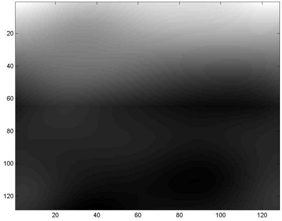


$\Rightarrow$ this procedure has a FWER controlled by α if each t_k has (use that $t_{\mathcal{K}} = t((\mathbf{Y}_k)_{k \in \mathcal{K}})$ is a non-decreasing function of $\mathcal{K}$).

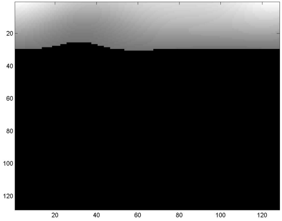
Simulations : with non-zero means, $0 \le \mu_k \le 0.29$, $b = 24$



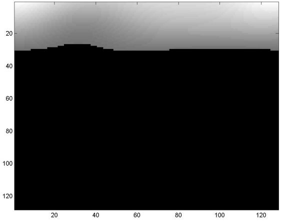
Numerical simulation : $0 \leq \mu_k \leq 0.29,\ b = 24$



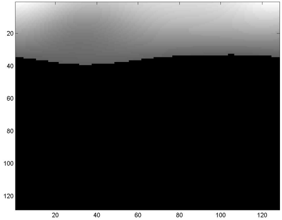
Empirical means



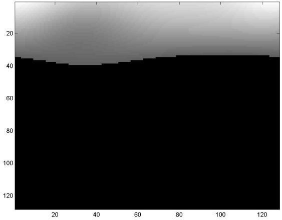
Bonf. : $t = 0.148$



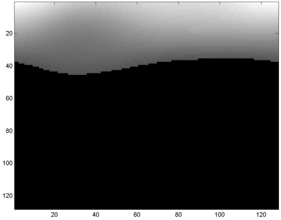
Holm : $t = 0.146$



Quant+Bonf : $t = 0.123$



Uncentered quant. : $t = 0.122$ (=S-d Q+B)



Quant. : $t = 0.106$

Conclusions

Two multiple testing procedures, with resampling techniques :

- concentration method (almost deterministic threshold)
- quantile method, with symmetrization techniques

- FWER controlled by α
- non-asymptotic
- better than Bonferroni if there are enough correlations
- step-down procedures

Conclusions

Two multiple testing procedures, with resampling techniques :

- concentration method (almost deterministic threshold)
- quantile method, with symmetrization techniques

- FWER controlled by α
- non-asymptotic
- better than Bonferroni if there are enough correlations
- step-down procedures

Conclusions

Two multiple testing procedures, with resampling techniques :

- concentration method (almost deterministic threshold)
- quantile method, with symmetrization techniques

- FWER controlled by α
- non-asymptotic
- better than Bonferroni if there are enough correlations
- step-down procedures

Conclusions

Two multiple testing procedures, with resampling techniques :

- concentration method (almost deterministic threshold)
- quantile method, with symmetrization techniques

- FWER controlled by α
- non-asymptotic
- better than Bonferroni if there are enough correlations
- step-down procedures

Conclusions

Two multiple testing procedures, with resampling techniques :

- concentration method (almost deterministic threshold)
- quantile method, with symmetrization techniques

- FWER controlled by α
- non-asymptotic
- better than Bonferroni if there are enough correlations
- step-down procedures

Future works

- FDR instead of FWER ?
- Theoretical study of power (with regular μ) ?

- Self-contained result for quantiles (without f) ?
- Quantiles with other weights ?
- Quantiles without the symmetry assumption ?

- Real-data sets ?

Future works

- FDR instead of FWER ?
- Theoretical study of power (with regular μ) ?

- Self-contained result for quantiles (without f) ?
- Quantiles with other weights ?
- Quantiles without the symmetry assumption ?

- Real-data sets ?

Future works

- FDR instead of FWER ?
- Theoretical study of power (with regular μ) ?

- Self-contained result for quantiles (without f) ?
- Quantiles with other weights ?
- Quantiles without the symmetry assumption ?

- Real-data sets ?

Thank you for your attention !