Outline

1. Introduction

2. The unidimensional case
   - Testing procedure
   - Dense mixtures
   - Sparse mixtures
   - Simulation study

3. The multidimensional contamination problem
   - Testing problem
   - Lower bound
   - Two testing procedures
   - The unbounded case

4. Perspectives
A testing point of view

- We have at our disposal a sample \( X = (X_1, \ldots, X_n) \) of i.i.d random variables having a common density \( f, X_i \in \mathbb{R}^d \).

- Goal: we want to test

\[
H_0 : f \in \mathcal{F}_0 = \{ x \in \mathbb{R}^d \mapsto \phi(x - \mu), \mu \in \mathbb{R}^d \}
\]

against

\[
H_1 : f \in \mathcal{F}_1 = \left\{ x \in \mathbb{R}^d \mapsto (1 - \varepsilon)\phi(x - \mu_1) + \varepsilon\phi(x - \mu_2) ; \right.  \\
\left. \varepsilon \in [0, 1[ , \mu_1, \mu_2 \in \mathbb{R}^d \right\}
\]

where \( \phi(\cdot) \) is a known density.
A testing point of view

We want to

- construct a testing procedure,

- control the first kind error by a fixed level $\alpha$,

- find (optimal) conditions on $(\epsilon, \mu_1, \mu_2)$ for which a second kind error $\beta$ can be achieved.
This question has already been addressed in the literature

- Test based on the likelihood ratio (Garel, 07; Azais et al., 09; ...)
- Modified likelihood ratio test (Chen et al, 01)
- EM approach (Chen and Li, 09)
- Tests based on the empirical characteristic function (Klar and Meintanis, 05)
- Seminal contribution of Y. Ingster (1999)
- The Higher-Critiscism proposed by Donoho and Jin (2004), Cai et al. (11), ...

In these contributions, $d = 1$ and $\mu = \mu_1 = 0$ is a known parameter.
Contributions

- Laurent et al. (2014, Bernoulli):
  - unidimensional case \((d = 1)\)
  - \(\phi(.) = \text{Gaussian density or Laplace density}\)
  - \(\mu, \mu_1, \mu_2\) unknown parameters

- Laurent et al. (preprint):
  - multidimensional case
  - \(\phi(.) = \text{Gaussian density}\)
  - contamination problem: \(\mu = \mu_1 = 0\)

We want to adopt a non-asymptotic point of view
In this talk, we will focus on the Gaussian case
Outline

1. Introduction

2. The unidimensional case
   - Testing procedure
   - Dense mixtures
   - Sparse mixtures
   - Simulation study

3. The multidimensional contamination problem
   - Testing problem
   - Lower bound
   - Two testing procedures
   - The unbounded case

4. Perspectives
Outline

1. Introduction

2. The unidimensional case
   - Testing procedure
   - Dense mixtures
   - Sparse mixtures
   - Simulation study

3. The multidimensional contamination problem
   - Testing problem
   - Lower bound
   - Two testing procedures
   - The unbounded case

4. Perspectives
We want to test:

\[ H_0 : f \in \mathcal{F}_0 = \{ x \in \mathbb{R} \mapsto \phi(x - \mu), \mu \in \mathbb{R} \} \]

against

\[ H_1 : f \in \mathcal{F}_1 = \{ x \in \mathbb{R} \mapsto (1 - \varepsilon)\phi(x - \mu_1) + \varepsilon\phi(x - \mu_2); \varepsilon \in ]0, 1[, \mu_1 < \mu_2 \in \mathbb{R} \} \]
A test based on the order statistics

Let \( X_(1) \leq X_(2) \leq \ldots \leq X_(n) \) be the order statistics.

Idea :
- The spacing of these order statistics are free w.r.t \( \mu \):
  - for some \( k < \ell \in \{1, \ldots, n\} \), \( \mu \) affects the spatial position of \( X_(k) \), but not \( X_(\ell) - X_(k) \).
- The distribution of the variables \( X_(\ell) - X_(k) \) is known under \( H_0 \).
- … and has a different behavior under \( H_1 \), provided \( k \) and \( \ell \) are well-chosen.
A test based on the order statistics

Our test statistics:

$$\Psi_\alpha := \sup_{k \in \mathcal{K}_n} \left\{ \mathbb{1}_{X_{n-k+1} - X_k > q_{\alpha, n}} \right\},$$
A test based on the order statistics

Let $n \geq 2$ and $\mathcal{K}_n$ be the subset of $\{1, 2, \ldots, n/2\}$ defined by

$$\mathcal{K}_n = \{2^j, 0 \leq j \leq \left\lfloor \ln_2(n/2) \right\rfloor \}.$$

Our test statistics:

$$\Psi_\alpha := \sup_{k \in \mathcal{K}_n} \left\{ \mathbb{1}_{X_{n-k+1} - X_k > q_{\alpha,n,k}} \right\},$$
A test based on the order statistics

Let \( n \geq 2 \) and \( \mathcal{K}_n \) be the subset of \( \{1, 2, \ldots, n/2\} \) defined by

\[
\mathcal{K}_n = \{2^j, 0 \leq j \leq \lceil \ln_2(n/2) \rceil \}.
\]

Our test statistics:

\[
\Psi_\alpha := \sup_{k \in \mathcal{K}_n} \left\{ \mathbb{1}_{X_{n-k+1} - X(k) > q_{u,k}} \right\},
\]

where

\( q_{u,k} \) is the \((1 - u)\)-quantile of \( X_{n-k+1} - X(k) \) under \( H_0 \) for all \( u \in ]0, 1[ \),

\[
\alpha_n = \sup \left\{ u \in ]0, 1[, \mathbb{P}_{H_0} \left( \exists k \in \mathcal{K}_n, X_{n-k+1} - X(k) > q_{u,k} \right) \leq \alpha \right\}.
\]

\( \alpha_n \) and \( q_{\alpha_n,k} \) are approximated (via Monte-Carlo method for instance)
By definition, $\Psi_\alpha$ is a level-$\alpha$ test:

$$
P_{H_0}(\Psi_\alpha = 1) = P_{H_0}\left(\sup_{k \in \mathcal{K}_n} \{ 1_{X(n-k+1)-X(k) > q_{\alpha, n, k}} \} = 1 \right)
$$

$$
= P_{H_0}(\exists k \in \mathcal{K}_n; X(n-k+1) - X(k) > q_{\alpha, n, k})
\leq \alpha.
$$

Remark: $\frac{\alpha}{|\mathcal{K}_n|} \leq \alpha_n \leq \alpha$.

$$
P_{H_0}(\exists k \in \mathcal{K}_n, X(n-k+1) - X(k) > q_{\alpha/|\mathcal{K}_n|, k})
\leq \sum_{k \in \mathcal{K}_n} P_{H_0}(X(n-k+1) - X(k) > q_{\alpha/|\mathcal{K}_n|, k}),
$$

$$
\leq \sum_{k \in \mathcal{K}_n} \frac{\alpha}{|\mathcal{K}_n|} \leq \alpha.
$$
The test $\psi_\alpha$ is a multiple testing procedure. Note that for any $f \in \mathcal{F}_1$,

$$P_f(\psi_\alpha = 0) = P_f \left( \sup_{k \in \mathcal{K}_n} \left\{ \mathbb{1}_{X_{(n-k+1)} - X(k) > q_{\alpha_n,k}} \right\} = 0 \right),$$

$$= P_f \left( \bigcap_{k \in \mathcal{K}_n} \left\{ \mathbb{1}_{X_{(n-k+1)} - X(k) > q_{\alpha_n,k}} \right\} = 0 \right),$$

$$\leq \inf_{k \in \mathcal{K}_n} P_f \left( \mathbb{1}_{X_{(n-k+1)} - X(k) > q_{\alpha_n,k}} = 0 \right),$$

The second kind error of $\psi_\alpha$ is close to the smallest one in the collection $\mathcal{K}_n$. 

Second kind error
In the sequel, two kinds of alternatives are considered:

- **the dense regime**: $0 < \mu_2 - \mu_1 \leq M$ and $\varepsilon > \frac{C}{\sqrt{n}}$

- **the sparse regime**: $\mu_2 - \mu_1$ can be large (asymptotic point of view)
  ... such $\varepsilon$ can be very small

**Goal**: Find optimal conditions on $(\varepsilon, \mu_1, \mu_2)$ for the both regimes.
Outline

1 Introduction

2 The unidimensional case
   - Testing procedure
   - Dense mixtures
   - Sparse mixtures
   - Simulation study

3 The multidimensional contamination problem
   - Testing problem
   - Lower bound
   - Two testing procedures
   - The unbounded case

4 Perspectives
We assume that $0 < \mu_2 - \mu_1 \leq M$ where $M$ is a positive constant

$$\mathcal{F}_1[M] = \{(1 - \epsilon)\phi(\cdot - \mu_1) + \epsilon\phi(\cdot - \mu_2); 0 < \mu_2 - \mu_1 \leq M\}$$

In this regime,

- establish a lower bound (Gaussian case),
- validate this bound with a test based on the variance,
- prove that our testing procedure is optimal.
Proposition

Let $\alpha, \beta \in ]0, 1[$ and $M > 0$. There exists $C = C(\alpha, \beta, M) > 0$ such that for all $\rho < \frac{C}{\sqrt{n}}$,

$$\inf_{T_\alpha} \sup_{f \in F_1[M]} \mathbb{P}_f(T_\alpha = 0) > \beta.$$ 

Remarks:

- Testing is not possible if $\varepsilon(1 - \varepsilon)(\mu_2 - \mu_1)^2 < C/\sqrt{n}$.

- In the "contamination problem", the separate condition is different: $\varepsilon(\mu_2 - \mu_1) \geq C/\sqrt{n}$.

- Non-asymptotic result.
Upper bound - Test based on the variance

Under $H_1$,

$$X_i = (\mu_2 - \mu_1) V_i + \eta_i, \quad \forall i \in \{1 \ldots n\},$$

where $V_i \sim B(\varepsilon)$ and $\eta_i \sim \phi(\cdot - \mu_1).$

$$\text{Var}(X_i) = \text{Var}(\eta_i) + \varepsilon(1 - \varepsilon)(\mu_2 - \mu_1)^2.$$

Let $\sigma^2 = \text{Var}(\eta_i)$ and $\psi_\alpha$ be the test defined by

$$\psi_\alpha = 1\{S_n^2 > \sigma^2 + c_\alpha / \sqrt{n}\},$$

where $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ and $c_\alpha$ is such that

$$P_{H_0}(S_n^2 - \sigma^2 > c_\alpha / \sqrt{n}) \leq \alpha.$$

By definition, $\psi_\alpha$ is a level-$\alpha$ test.
Upper bound - Test based on the variance

For any $f \in \mathcal{F}_1[M]$, 

\[
\mathbb{P}_f(\psi_\alpha = 0) = \mathbb{P}_f(S_n^2 \leq \sigma^2 + c_\alpha/\sqrt{n}),
\]

\[
= \mathbb{P}_f(S_n^2 - \mathbb{E}[S_n^2] \leq c_\alpha/\sqrt{n} - \varepsilon(1 - \varepsilon)(\mu_2 - \mu_1)^2),
\]

\[
\leq \mathbb{P}_f \left( \left| S_n^2 - \mathbb{E}[S_n^2] \right| \geq \varepsilon(1 - \varepsilon)(\mu_2 - \mu_1)^2 - c_\alpha/\sqrt{n} \right),
\]

\[
\leq \frac{\text{Var}(S_n^2)}{[\varepsilon(1 - \varepsilon)(\mu_2 - \mu_1)^2 - c_\alpha/\sqrt{n}]^2}.
\]

In particular, if $\text{Var}(S_n^2) \leq C/n$, we have 

\[
\mathbb{P}_f(\psi_\alpha = 0) \leq \beta,
\]

as soon as 

\[
\varepsilon(1 - \varepsilon)(\mu_2 - \mu_1)^2 > \frac{C_{\alpha,\beta}}{\sqrt{n}}.
\]
Proposition

Let $\alpha \in ]0, 1[$ and $\beta \in ]0, 1 - \alpha[$. Assume that the density function $\phi$ has a finite fourth moment: $\int_{\mathbb{R}} x^4 \phi(x) \, dx \leq B$. There exists a positive constant $C(\alpha, \beta, M, B)$ such that if

$$\rho \geq C(\alpha, \beta, M, B)/\sqrt{n},$$

then

$$\sup_{f \in \mathcal{F}_1[M]} \mathbb{P}_f(\psi_\alpha = 0) \leq \beta.$$
Proposition

There exists a positive constant $C_{\alpha,\beta,M} > 0$ such that, if

$$\rho \geq C(\alpha, \beta, M) \sqrt{\frac{\ln \ln(n)}{n}},$$

then

$$\sup_{f \in F_1[M], \epsilon(1-\epsilon)(\mu_2-\mu_1)^2 \geq \rho} \mathbb{P}_f(\psi_\alpha = 0) \leq \beta.$$

Remarks:

- The proof is based on the control of deviations of the order statistics and the associated quantiles.
- This log log term is due to the multiple (adaptive) testing procedure.
An asymptotic study

The asymptotic dense regime in the Gaussian setting:

\[ \varepsilon \sim n^{-\delta} \quad \text{and} \quad \mu_2 - \mu_1 \sim n^{-r} \quad n \to +\infty \]

with \( 0 < \delta \leq \frac{1}{2} \) and \( 0 < r < \frac{1}{2} \).

**Corollary**

The detection boundary in the dense regime is \( r^*(\delta) = \frac{1}{4} - \frac{\delta}{2} \):

the detection is possible when \( r < r^*(\delta) = \frac{1}{4} - \frac{\delta}{2} \) and impossible if \( r > r^*(\delta) \).

Remark : in the "contamination problem"

\[ r^*(\delta) = \frac{1}{2} - \delta \]
1 Introduction

2 The unidimensional case
   - Testing procedure
   - Dense mixtures
   - Sparse mixtures
   - Simulation study

3 The multidimensional contamination problem
   - Testing problem
   - Lower bound
   - Two testing procedures
   - The unbounded case

4 Perspectives
The asymptotic sparse regime:

\[ \varepsilon \sim n^{-\delta} \quad \text{and} \quad \mu_2 - \mu_1 \sim \sqrt{2r \ln(n)} \]

with \( \frac{1}{2} < \delta < 1 \) and \( 0 < r < 1 \).

"\( \varepsilon \ll \frac{1}{\sqrt{n}} \) and \( \mu_2 - \mu_1 \to +\infty \) when \( n \to +\infty \)."
Proposition

We assume that $r > r^*(\delta)$ with

$$r^*(\delta) = \begin{cases} 
\delta - \frac{1}{2} & \text{if } \frac{1}{2} < \delta < \frac{3}{4} \\
(1 - \sqrt{1 - \delta})^2 & \text{if } \frac{3}{4} \leq \delta < 1
\end{cases}.$$ 

Then, setting $f(.) = (1 - \varepsilon)\phi(. - \mu_1) + \varepsilon\phi(. - \mu_2)$, we have, for $n$ large enough,

$$\mathbb{P}_{f}(\psi_{\alpha} = 0) \leq \beta.$$ 

In the sparse regime, we exactly recover the separation boundaries that are already known in the contamination problem.
The variance test for sparse mixtures

For any \( f = (1 - \varepsilon)\phi(\cdot - \mu_1) + \varepsilon\phi(\cdot - \mu_2), \)

\[
\text{Var}_f(X_i) = \text{Var}_\phi(X_i) + \varepsilon(1 - \varepsilon)(\mu_1 - \mu_2)^2.
\]

For both Gaussian and Laplace mixtures,

\[
\text{Var}_f(X_i) - \text{Var}_\phi(X_i) = \varepsilon(1 - \varepsilon)(\mu_1 - \mu_2)^2 \ll \frac{1}{\sqrt{n}}, \text{ as } n \to +\infty.
\]

Since the variance is estimated at a parametric "rate" \( 1/\sqrt{n} \), the test \( \psi_\alpha \) will fail in this setting.
Outline

1. Introduction

2. The unidimensional case
   - Testing procedure
   - Dense mixtures
   - Sparse mixtures
   - Simulation study

3. The multidimensional contamination problem
   - Testing problem
   - Lower bound
   - Two testing procedures
   - The unbounded case

4. Perspectives
Simulation study

Our testing procedure is compared with the adaptations of

- Kolmogorov-Smirnov test: \( \hat{\psi}_{KS,\alpha} = 1_{T_{KS} > \hat{q}_{KS,\alpha}} \) where

\[
\hat{T}_{KS} = \sup_{x \in \mathbb{R}} \sqrt{n} |F_n(x) - \Phi_G(x - \bar{X})|
\]

- Higher Criticism (Donoho and Jin, 04)
Let \( \hat{p}_i = \mathbb{P}(Z - \bar{X} > X_i) \) where \( Z \sim \mathcal{N}(0, 1) \) for all \( i \in \{1, \ldots, n\} \) and \( \hat{p}(1) \leq \hat{p}(2) \leq \ldots \leq \hat{p}(n) \). The level-\( \alpha \) test function is

\[
\hat{\psi}_{HC,\alpha} = 1_{HC > \hat{q}_{HC,\alpha}}
\]
with

\[
\hat{HC} = \max_{1 \leq i \leq n} \frac{\sqrt{n} \left( \frac{i}{n} - \hat{p}(i) \right)}{\sqrt{\hat{p}(i)(1 - \hat{p}(i))}}.
\]

A Monte-Carlo procedure is considered with \( N = 100000 \) samples of size \( n = 100 \) for a Gaussian mixture with \( \varepsilon \in \{0.05, 0.15, 0.25, 0.35\} \), \( \mu_1 = 0 \) and \( \mu_2 \in [0, 10] \).
Simulation study - Gaussian case

Figure: Power function of the three considered testing procedures (continuous line for our test $\Psi_{\alpha}$, dashed line for Higher Criticism and dotted line for the Kolmogorov-Smirnov test) according to $\mu$, for $\varepsilon = 0.05$ (top-left), 0.15 (top right), 0.25 (middle left) and 0.35 (middle right).
Simulation study - Laplace case

Figure: Power function of the three considered testing procedures (continuous line for our test $\Psi_{\alpha}$, dashed line for Higher Criticism and dotted line for the Kolmogorov-Smirnov test) according to $\mu$, for $\epsilon = 0.05$ (top-left), $0.15$ (top right), $0.25$ (middle left) and $0.35$ (middle right).
1. Introduction

2. The unidimensional case
   - Testing procedure
   - Dense mixtures
   - Sparse mixtures
   - Simulation study

3. The multidimensional contamination problem
   - Testing problem
   - Lower bound
   - Two testing procedures
   - The unbounded case

4. Perspectives
Outline

1. Introduction

2. The unidimensional case
   - Testing procedure
   - Dense mixtures
   - Sparse mixtures
   - Simulation study

3. The multidimensional contamination problem
   - Testing problem
   - Lower bound
   - Two testing procedures
   - The unbounded case

4. Perspectives
Testing problem

- Let $(X_1, \ldots, X_n)$ i.i.d $d$-dimensional random vectors with density $f$

- Let $\phi(.)$ be the density function of the standard Gaussian distribution $\mathcal{N}_d(0_d, I_d)$.

- We want to test

  \[ H_0 : f = \phi \text{ against } H_1 : f \in \mathcal{F}_1 \]

  where

  \[ \mathcal{F}_1 = \{ x \in \mathbb{R}^d \mapsto (1 - \varepsilon)\phi(x) + \varepsilon\phi(x - \mu); \varepsilon \in ]0, 1[ , \mu \in \mathbb{R}^d \} \]

- Dense regime: $\varepsilon > C/\sqrt{n}$ and $\|\mu\| \leq M$. 
1 Introduction

2 The unidimensional case
   - Testing procedure
   - Dense mixtures
   - Sparse mixtures
   - Simulation study

3 The multidimensional contamination problem
   - Testing problem
   - Lower bound
   - Two testing procedures
   - The unbounded case

4 Perspectives
A lower bound

Let $\mathcal{F} \subset \mathcal{F}_1$ a subset of alternatives, and $\pi$ a probability measure on $\mathcal{F}$. Then,

$$\inf_{\psi} \sup_{f \in \mathcal{F}} \mathbb{P}_f(\psi_\alpha = 0) \geq 1 - \alpha - \frac{1}{2} \sqrt{\mathbb{E}_{H_0}[L_\pi^2(X)]} - 1,$$

where $L_\pi^2(X)$ the likelihood ratio $d\mathbb{P}_\pi / d\mathbb{P}_0$ and the infimum is taken over all $\alpha$-level tests.

In particular, for some appropriate constant $\eta(\alpha, \beta)$,

$$\mathbb{E}_{H_0}[L_\pi^2(X)] \leq \eta(\alpha, \beta) \implies \inf_{\psi} \sup_{f \in \mathcal{F}} \mathbb{P}_f(\psi_\alpha = 0) \geq \beta.$$

See e.g, Ingster (1999) or Baraud (2002) for more details.
Let $\mathcal{F}_1[M] = \{f(.) = (1 - \varepsilon)\phi(.) + \varepsilon\phi(.) - \mu); \varepsilon \in ]0, 1[, \|\mu\| \leq M\}$. 

**Proposition**

Let $\alpha, \beta \in ]0, 1[$ and $M > 0$. There exists $C = C(\alpha, \beta, M) > 0$ such that for all $\rho < C \frac{d^4}{\sqrt{n}}$,

$$\inf_{T_\alpha} \sup_{f \in \mathcal{F}_1[M]} \inf_{\varepsilon \|\mu\| \geq \rho} \mathbb{P}_f(T_\alpha = 0) > \beta.$$ 

Testing is impossible if $\varepsilon \|\mu\| < \frac{C \frac{d^4}{\sqrt{n}}}{\sqrt{n}}$. 
Outline

1. Introduction

2. The unidimensional case
   - Testing procedure
   - Dense mixtures
   - Sparse mixtures
   - Simulation study

3. The multidimensional contamination problem
   - Testing problem
   - Lower bound
   - Two testing procedures
   - The unbounded case

4. Perspectives
First testing procedure ($\Psi_{1,\alpha}$)

**Proposition**

Let $\alpha \in ]0, 1[$. Let the level-$\alpha$ test

$$\Psi_{1,\alpha} = \mathbb{1}_{\|\sqrt{n}\bar{X}_n\|^2 > \nu_\alpha}$$

where $\nu_\alpha$ is the $(1 - \alpha)$ quantile of $\chi^2(d)$ and $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$.

Let $\beta \in ]0, 1 - \alpha[$ and $M > 0$. Then, there exists a positive constant $C(\alpha, \beta, M)$ such that, if

$$\rho \geq C(\alpha, \beta, M) \frac{d^{1/4}}{\sqrt{n}}$$

then

$$\sup_{f \in \mathcal{F}_1[M] \atop \varepsilon \|\mu\| \geq \rho} \mathbb{P}_f (\Psi_{1,\alpha} = 0) \leq \beta.$$
Second testing procedure ($\Psi_{2,\alpha}$)

- The sample $X$ is split into two different parts:
  \[ A = (A_1, \ldots, A_n) \text{ and } Y = (Y_1, \ldots, Y_n). \]

- Let $v_n = \bar{A}_n/\|\bar{A}_n\|$ where $\bar{A}_n = \frac{1}{n} \sum_{i=1}^{n} A_i$.

- Let $Z_i = \langle Y_i, v_n \rangle$ for all $i \in \{1, \ldots, n\}$ and $Z(1) \leq \cdots \leq Z(n)$.

- Conditionally to $A$,
  - the $Z_i$ are i.i.d standard Gaussian random variables under $H_0$.
  - $Z_i \sim (1 - \varepsilon) \mathcal{N}(0, 1) + \varepsilon \mathcal{N}(\mu, v_n)$ under $H_1$

- The testing procedure:
  \[ \Psi_{2,\alpha} = \sup_{k \in \mathcal{K}_n} \mathbb{1}_{Z(n-k+1) > q_{\alpha, n, k}}. \]
Second testing procedure \((\Psi_{2,\alpha})\)

**Proposition**

Let \(\beta \in ]0, 1 - \alpha[\) and \(M > 0\). Then, there exists a positive constant \(C(\alpha, \beta, M)\) such that, if

\[
\rho \geq C(\alpha, \beta, M)d_4^{1/4} \sqrt{\frac{\ln \ln(n)}{n}}
\]

then

\[
\sup_{f \in \mathcal{F}_1[M], \epsilon \|\mu\| \geq \rho} \mathbb{P}_f (\Psi_{2,\alpha} = 0) \leq \beta.
\]
1 Introduction

2 The unidimensional case
   • Testing procedure
   • Dense mixtures
   • Sparse mixtures
   • Simulation study

3 The multidimensional contamination problem
   • Testing problem
   • Lower bound
   • Two testing procedures
   • The unbounded case

4 Perspectives
Results when $\mu$ is unbounded

**Theorem**

Let $\alpha, \beta \in ]0, 1[$ be fixed and, $\Psi_{1,\alpha}$ and $\Psi_{2,\alpha}$ be the both previous tests. Then, there exists a positive constant $C(\alpha, \beta)$, only depending on $\alpha, \beta$ and $n_0 \in \mathbb{N}^*$ such that, for $n \geq n_0$ and for all $f = f(\varepsilon, \mu) \in \mathcal{F}$ satisfying

$$
\varepsilon \geq C(\alpha, \beta) \frac{\ln \ln (n)}{n}
$$

and

$$
\varepsilon^2 \|\mu\|^2 \geq C(\alpha, \beta) \left[ \left( \frac{\sqrt{d}}{n} \right) \wedge \left\{ \varepsilon \sqrt{\frac{d}{n} \ln \left( \frac{1}{\varepsilon} \right)} \right\} \right],
$$

we have

$$
P_f(\Psi_{1,\alpha/2} \lor \Psi_{2,\alpha/2} = 0) \leq \beta.
$$
Figure: Summary of the separation condition on $\varepsilon^2 \|\mu\|^2$ for the test $\Psi_{1,\alpha/2} \lor \Psi_{2,\alpha/2}$, where $\varepsilon_n = \ln \ln(n)/n$ and $\tilde{\varepsilon}_n = \inf \{\varepsilon \in ]0, 1[ : \varepsilon^2 \ln(1/\varepsilon) > \frac{1}{n}\}$.
An other testing procedure

$$\Psi_{4,\alpha} = \sup_{U \in \mathcal{U}} 1_{T_U > t_{n,d,|U|,\alpha}}$$

where $\mathcal{U}$ denotes the set of the nonempty subsets of $\{1, \ldots, n\}$, $|U|$ denotes the cardinality of $U$,

$$T_U = \frac{1}{|U|} \left\| \sum_{i \in U} X_i \right\|^2,$$

$$t_{n,d,k,\alpha} = d + 2\sqrt{d} x_{n,k,\alpha} + 2 x_{n,k,\alpha} \text{ and } x_{n,k,\alpha} = k \ln(en/k) + \ln(n/\alpha).$$
An other testing procedure

**Theorem**

Let $\alpha, \beta \in ]0, 1[$ be fixed. Let $\Psi_{1, \alpha}$ and $\Psi_{4, \alpha}$ be the both previous tests. There exists a positive constant $C(\alpha, \beta)$ only depending on $\alpha, \beta$ such that, for all $f = f(\varepsilon, \mu) \in \mathcal{F}$ which fulfills $n\varepsilon \geq \frac{8}{\beta}$ and

$$
\varepsilon^2 \|\mu\|^2 \geq C(\alpha, \beta) \left[ \left( \frac{\sqrt{d}}{n} \right) \wedge \left\{ \varepsilon^2 \ln \left( \frac{1}{\varepsilon} \right) + \varepsilon^{3/2} \sqrt{\frac{d}{n} \ln \left( \frac{1}{\varepsilon} \right)} \right\} \right], \quad (1)
$$

we have

$$
P_f(\Psi_{1, \alpha/2} \lor \Psi_{4, \alpha/2} = 0) \leq \beta.
$$
Figure: Summary of the separation condition on $\varepsilon^2 \|\mu\|_2^2$ for the test $\Psi_{1,\alpha/2} \lor \Psi_{4,\alpha/2}$, where $\overline{\varepsilon}_n = \inf\{\varepsilon \in [0, 1]; n\varepsilon^3 \ln(1/\varepsilon) \geq 1\}$

$d = n\varepsilon \ln(1/\varepsilon)$

$d = \{n\varepsilon^2 \ln(1/\varepsilon)\}^2$

$\varepsilon^{3/2} \sqrt{\frac{d}{n}} \ln(1/\varepsilon)$

$\varepsilon^2 \ln(1/\varepsilon)$

$\frac{\sqrt{d}}{n}$
Outline

1. Introduction

2. The unidimensional case
   - Testing procedure
   - Dense mixtures
   - Sparse mixtures
   - Simulation study

3. The multidimensional contamination problem
   - Testing problem
   - Lower bound
   - Two testing procedures
   - The unbounded case

4. Perspectives
- Lower bound when $\|\mu\|$ is unbounded?
- Testing procedure in the sparse regime?
- Consider a more general test problem in the multidimensional context
- ...


References II

Higher criticism for detecting sparse heterogeneous mixtures.

Recent asymptotic results in testing for mixtures.

Tests for normal mixtures based on the empirical characteristic function.