Non-asymptotic detection of two-component mixtures

B. Laurent, C. Marteau and C. Maugis-Rabusseau

Outline

Outline

A testing point of view

- We have at our disposal a sample $\mathcal{X} = (X_1, \dots, X_n)$ 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

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

where $\phi(.)$ 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 α,
- find (optimal) conditions on $(\varepsilon, \mu_1, \mu_2)$ for which a second kind error β can be achieved.

Bibliography

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 ($d = 1$)
  - $\phi(.) =$ Gaussian density or Laplace density
  - μ, μ_1, μ_2 unknown parameters
- Laurent et al. (preprint) :
  - multidimensional case
  - $\phi(.) =$ 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

Outline

Testing problem

- We want to test :

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

against

$$\begin{aligned} H_1 \quad : \quad & f \in \mathcal{F}_1 = \{x \in \mathbb{R} \mapsto (1-\varepsilon)\phi(x-\mu_1) + \varepsilon\phi(x-\mu_2); \\ & \qquad \varepsilon \in]0,1[, \mu_1 < \mu_2 \in \mathbb{R}\} \end{aligned}$$

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 μ:
    for some $k < \ell \in \{1, ..., n\}$, μ 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 ℓ are well-chosen.

A test based on the order statistics

- Our test statistics:

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

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

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

- 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

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

- Our test statistics:

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

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

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

α_n and $q_{\alpha_n,k}$ are approximated (via Monte-Carlo method for instance)

First error rate

- By definition, Ψ_α is a level-α test:

$$\begin{aligned}
\mathbb{P}_{H_0}(\Psi_\alpha = 1) &= \mathbb{P}_{H_0}\left(\sup_{k \in \mathcal{K}_n} \left\{ \mathbb{1}_{X_{(n-k+1)} - X_{(k)} > q_{\alpha_n,k}} \right\} = 1\right) \\
&= \mathbb{P}_{H_0}\left(\exists k \in \mathcal{K}_n; X_{(n-k+1)} - X_{(k)} > q_{\alpha_n,k}\right) \\
&\leq \alpha.
\end{aligned}$$

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

$$\begin{aligned}
&\mathbb{P}_{H_0}\left(\exists k \in \mathcal{K}_n, X_{(n-k+1)} - X_{(k)} > q_{\alpha/|\mathcal{K}_n|,k}\right) \\
\leq &\sum_{k \in \mathcal{K}_n} \mathbb{P}_{H_0}(X_{(n-k+1)} - X_{(k)} > q_{\alpha/|\mathcal{K}_n|,k}), \\
\leq &\sum_{k \in \mathcal{K}_n} \frac{\alpha}{|\mathcal{K}_n|} \leq \alpha.
\end{aligned}$$

Second kind error

The test Ψ_α is a multiple testing procedure.

Note that for any $f \in \mathcal{F}_1$,

$$\begin{aligned}
\mathbb{P}_f(\Psi_\alpha = 0) &= \mathbb{P}_f\left(\sup_{k\in\mathcal{K}_n}\left\{\mathbb{1}_{X_{(n-k+1)}-X_{(k)}>q_{\alpha n,k}}\right\} = 0\right), \\
&= \mathbb{P}_f\left(\bigcap_{k\in\mathcal{K}_n}\left\{\mathbb{1}_{X_{(n-k+1)}-X_{(k)}>q_{\alpha n,k}}\right\} = 0\right), \\
&\leq \inf_{k\in\mathcal{K}_n} \mathbb{P}_f\left(\mathbb{1}_{X_{(n-k+1)}-X_{(k)}>q_{\alpha n,k}} = 0\right),
\end{aligned}$$

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

Outline

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 ε can be very small

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

Outline

"Road map"

- We assume that $0 < \mu_2 - \mu_1 \leq M$ where M is a positive constant
- $\mathcal{F}_1[M] = \{(1-\varepsilon)\phi(.-\mu_1) + \varepsilon\phi(.-\mu_2); 0 < \mu_2 - \mu_1 \leq M\}$
- 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.

Lower bound (Gaussian case)

Proposition

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

$$\inf_{T_\alpha} \sup_{\substack{f \in \mathcal{F}_1[M] \\ \varepsilon(1-\varepsilon)(\mu_2-\mu_1)^2 \geq \rho}} \mathbb{P}_f(T_\alpha = 0) > \beta.$$

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,

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

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

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

Let $\sigma^2 = \mathrm{Var}(\eta_i)$ and ψ_α be the test defined by

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

where $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ and c_α is such that $\mathbb{P}_{H_0}(S_n^2 - \sigma^2 > c_\alpha/\sqrt{n}) \leq \alpha$.

By definition, ψ_α is a level-α test.

Upper bound - Test based on the variance

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

$$\begin{aligned}
\mathbb{P}_f(\psi_\alpha = 0) &= \mathbb{P}_f(S_n^2 \leq \sigma^2 + c_\alpha/\sqrt{n}), \\
&= \mathbb{P}_f(S_n^2 - \mathbb{E}[S_n^2] \leq c_\alpha/\sqrt{n} - \varepsilon(1-\varepsilon)(\mu_2 - \mu_1)^2), \\
&\leq \mathbb{P}_f\left(\left|S_n^2 - \mathbb{E}[S_n^2]\right| \geq \varepsilon(1-\varepsilon)(\mu_2 - \mu_1)^2 - c_\alpha/\sqrt{n}\right), \\
&\leq \frac{\mathrm{Var}(S_n^2)}{[\varepsilon(1-\varepsilon)(\mu_2-\mu_1)^2 - c_\alpha/\sqrt{n}]^2}.
\end{aligned}$$

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

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

as soon as

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

Upper bound - Test based on the variance

Proposition

Let $\alpha \in]0,1[$ and $\beta \in]0, 1-\alpha[$. Assume that the density function ϕ 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

$$\sup_{\substack{f \in \mathcal{F}_1[M] \\ \varepsilon(1-\varepsilon)(\mu_2-\mu_1)^2 \geq \rho}} \mathbb{P}_f(\psi_\alpha = 0) \leq \beta.$$

Upper bound - our testing procedure (Ψ_α)

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

$$\sup_{\substack{f\in\mathcal{F}_1[M] \\ \varepsilon(1-\varepsilon)(\mu_2-\mu_1)^2\geq\rho}} \mathbb{P}_f(\Psi_\alpha = 0) \leq \beta.$$

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:

$$\varepsilon \underset{n\to+\infty}{\sim} n^{-\delta} \text{ and } \mu_2 - \mu_1 \underset{n\to+\infty}{\sim} n^{-r}$$

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"

$$r^*(\delta) = \frac{1}{2} - \delta$$

Outline

Sparse Gaussian mixtures - Asymptotic study

- The asymptotic sparse regime:

$$\varepsilon \underset{n\to+\infty}{\sim} n^{-\delta} \text{ and } \mu_2 - \mu_1 \underset{n\to+\infty}{\sim} \sqrt{2r\ln(n)}$$

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

$$"\varepsilon \ll \frac{1}{\sqrt{n}} \text{ and } \mu_2 - \mu_1 \to +\infty \text{when } n \to +\infty."$$

Sparse Gaussian mixtures - Asymptotic study

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 $f = (1-\varepsilon)\phi(. - \mu_1) + \varepsilon\phi(. - \mu_2)$,

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

For both Gaussian and Laplace mixtures,

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

Since the variance is estimated at a parametric "rate" $1/\sqrt{n}$, the test ψ_α will fail in this setting

Outline

Simulation study

Our testing procedure is compared with the adaptations of

- Kolmogorov-Smirnov test: $\widehat{\psi}_{KS,\alpha} = \mathbb{1}_{\hat{T}_{KS} > \hat{q}_{KS,\alpha}}$ where

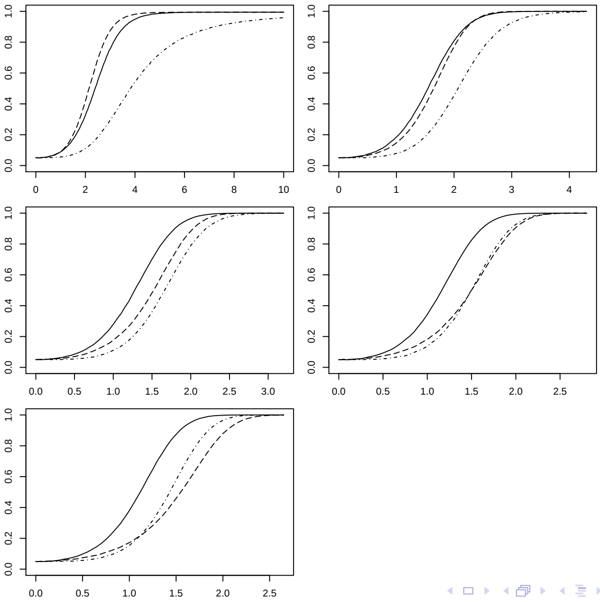
$$\hat{T}_{KS} = \sup_{x \in \mathbb{R}} \sqrt{n} |F_n(x) - \Phi_G(x - \bar{X})|$$

- 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 $\hat{\psi}_{HC,\alpha} = \mathbb{1}_{\widehat{HC} > \hat{q}_{HC,\alpha}}$ with

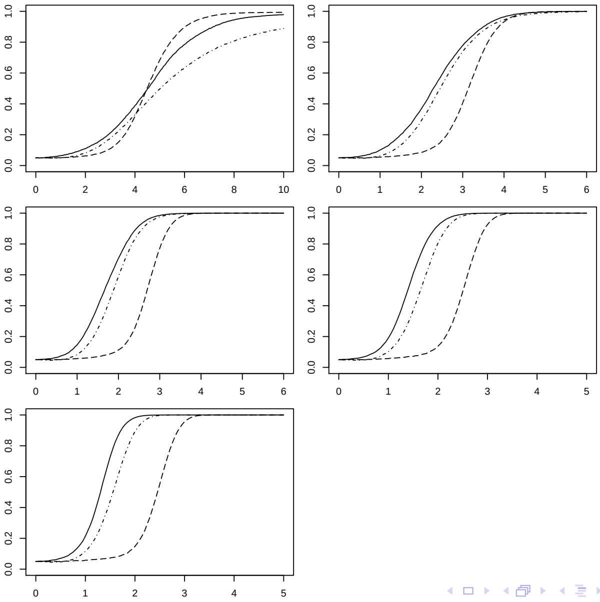
$$\widehat{HC} = \max_{1 \leq i \leq n} \frac{\sqrt{n}\left(\frac{i}{n} - \hat{p}_{(i)}\right)}{\sqrt{\hat{p}_{(i)}(1 - \hat{p}_{(i)})}}.$$

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



Simulation study - Laplace case



Outline

Outline

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 $\|\mu\| \leq M$.

Outline

A lower bound

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

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

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

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

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

See e.g, Ingster (1999) or Baraud (2002) for more details.

Lower bound

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 $\rho < C\, d^{\frac{1}{4}}/\sqrt{n}$,

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

Testing is impossible if $\varepsilon\|\mu\| < \frac{C\, d^{\frac{1}{4}}}{\sqrt{n}}$.

Outline

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

Proposition

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

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

where v_α 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

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

then

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

Second testing procedure ($\Psi_{2,\alpha}$)

- The sample X is splitted in 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

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

then

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

Outline

Results when μ 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 $\mathcal{C}(\alpha, \beta)$, only depending on α, β and $n_0 \in \mathbb{N}^*$ such that, for $n \geq n_0$ and for all $f = f_{(\varepsilon,\mu)} \in \mathcal{F}$ satisfying $\varepsilon \geq \mathcal{C}(\alpha, \beta)\frac{\ln\ln(n)}{n}$ 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

$$\mathbb{P}_f(\Psi_{1,\alpha/2} \vee \Psi_{2,\alpha/2} = 0) \leq \beta.$$

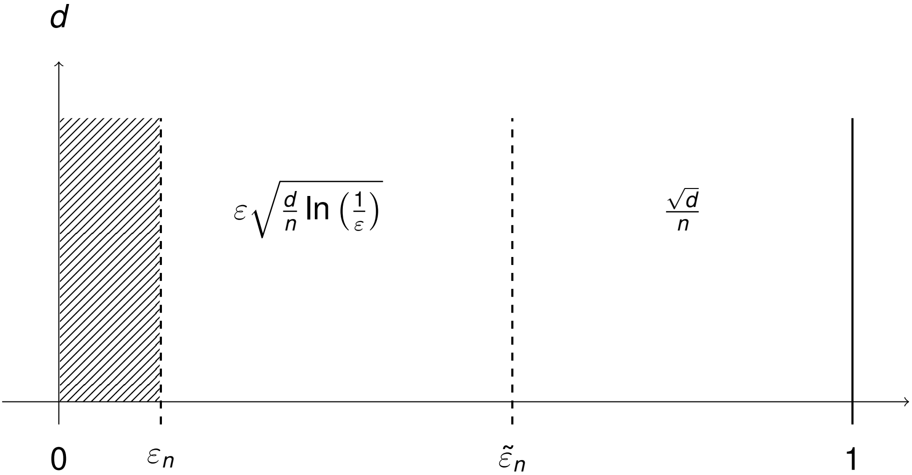


Figure: Summary of the separation condition on $\varepsilon^2\|\mu\|^2$ for the test $\Psi_{1,\alpha/2} \vee \Psi_{2,\alpha/2}$, where $\varepsilon_n = \ln\ln(n)/n$ and $\tilde{\varepsilon}_n = \inf\left\{\varepsilon \in]0,1[: \varepsilon^2 \ln(1/\varepsilon) > \frac{1}{n}\right\}$

An other testing procedure

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

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

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

$t_{n,d,k,\alpha} = d + 2\sqrt{d\ x_{n,k,\alpha}} + 2\ x_{n,k,\alpha}$ 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 α, β 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

$$\mathbb{P}_f(\Psi_{1,\frac{\alpha}{2}} \vee \Psi_{4,\frac{\alpha}{2}} = 0) \leq \beta.$$

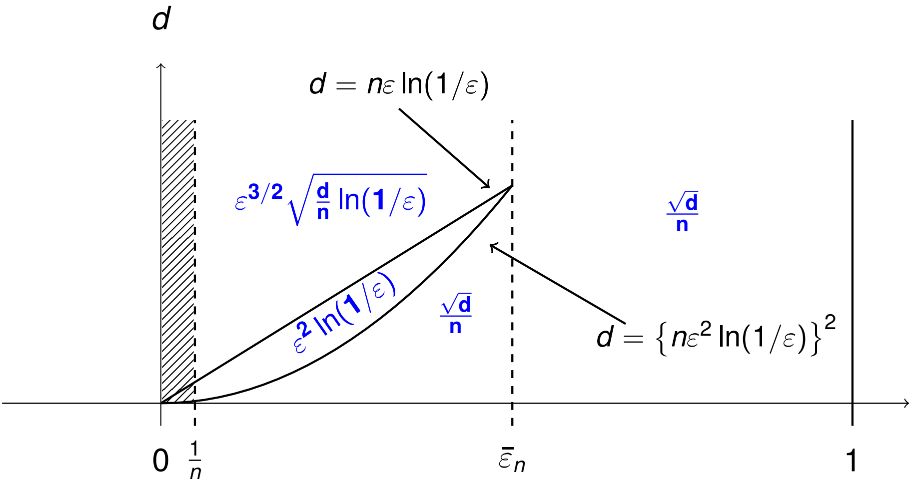


Figure: Summary of the separation condition on $\varepsilon^2\|\mu\|^2$ for the test $\Psi_{1,\alpha/2} \vee \Psi_{4,\alpha/2}$, where $\bar{\varepsilon}_n = \inf\{\varepsilon \in]0,1[; n\varepsilon^3 \ln(1/\varepsilon) \geq 1\}$

Outline

Perspectives

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

References I

Azaïs, J.-M., Gassiat, É., and Mercadier, C. (2009).
The likelihood ratio test for general mixture models with or without structural parameter.
ESAIM Probab. Stat., 13:301–327.

Cai, T. T., Jeng, X. J., and Jin, J. (2011).
Optimal detection of heterogeneous and heteroscedastic mixtures.
J. R. Stat. Soc. Ser. B Stat. Methodol., 73(5):629–662.

Chen, H., Chen, J., and Kalbfleisch, J. D. (2001).
A modified likelihood ratio test for homogeneity in finite mixture models.
Journal of the Royal Statistical Society. Series B (Statistical Methodology), 63(1):pp. 19–29.

Chen, J. and Li, P. (2009).
Hypothesis test for normal mixture models: the EM approach.
Ann. Statist., 37(5A):2523–2542.

Chernoff, H. and Lander, E. (1995).
Asymptotic distribution of the likelihood ratio test that a mixture of two binomials is a single binomial.
J. Statist. Plann. Inference, 43(1-2):19–40.

Dacunha-Castelle, D. and Gassiat, E. (1999).
Testing the order of a model using locally conic parametrization: population mixtures and stationary ARMA processes.
Ann. Statist., 27(4):1178–1209.

References II

Donoho, D. and Jin, J. (2004).
Higher criticism for detecting sparse heterogeneous mixtures.
Ann. Statist., 32(3):962–994.

Garel, B. (2007).
Recent asymptotic results in testing for mixtures.
Comput. Statist. Data Anal., 51(11):5295–5304.

Klar, B. and Meintanis, S. G. (2005).
Tests for normal mixtures based on the empirical characteristic function.
Comput. Statist. Data Anal., 49(1):227–242.