Parameter recovery in two-component contamination mixtures: the $\mathbb{L}^2$ strategy

S. Gadat [1], J. Kahn [2], C. Marteau [3] and Cathy Maugis-Rabusseau [4] .

Mathematical Methods of Modern Statistics
CIRM 2017

1. Université Toulouse I - Toulouse School of Economics
2. Université Toulouse III - IMT
3. Université Lyon I - Institut Camille Jordan
4. INSA de Toulouse - IMT

Outline

Outline

The mixture model

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

In an unsupervised classification context, $f^\star$ can be considered of the form

$$f^\star = \sum_{j=1}^{K} \lambda_j^\star \phi(. - \mu_j^\star),$$

where ϕ is a **known** density, $\lambda_j^\star \in [0, 1]$, $\mu_j^\star \in \mathbb{R}^d$ and K are unknown parameters.

Classical statistical issues

- estimation of the sequences $(\lambda_j^\star)_{j=1\ldots,K}$ and $(\mu_j^\star)_{j=1\ldots,K}$,
- estimation of the component number K (model selection task).

Mixture as an inverse problem

The estimation of the mixture parameters turns to be an inverse (deconvolution) problem. Indeed,

$$X_i = U_i + \epsilon_i, \quad \forall i \in \{1, \dots, n\},$$

where $\epsilon_i \sim \phi$ (error term) and U_i are associated to the discrete measure $G = \sum_{k=1}^{K} \lambda_k^\star \delta_{\mu_k^\star}$. Then,

$$f^\star = G * \phi.$$

In this contact, the 'classical' deconvolution tools are not available.

Two component mixtures

In this talk, we consider the particular contamination case, namely $K = 2$, $\mu_1 = 0$ and $\mu_2 = \mu^\star$:

$$f^\star = f_{\lambda^\star}(x) = (1 - \lambda^\star)\phi(x) + \lambda^\star \phi(x - \mu^\star) \quad \forall x \in \mathbb{R}^d.$$

The X_i can be written

$$X_i = \mu^\star V_i + \epsilon_i, \quad \forall i \in \{1, \ldots, n\},$$

where $V_i \sim Ber(\lambda^\star)$, $\epsilon_i \sim \phi$.

N.B. : Strong analogies with the sequence model

$$y_k = \theta_k + \eta_k, \quad k \in \{1, \ldots, n\},$$

where $\theta_k \in \{0, \theta\}$ and $\mathrm{card}\{k : \ \theta_k \neq 0\} = s \ (\sim n\lambda^\star)$.

References

[1]J-M. Azais, E. Gassiat and C. Mercadier. The likelihood-ratio test for general mixture models with or without structural parameters, *ESAIM Probab. Stat.*, **13**, (2009) 301-327.

[2] T. Cai, X. Jeng and J. Jin. Optimal detection of heterogeneous and heteroscedastic mixture, *J.R. Stat. Soc. Ser. B*, **73**, (2011) 629-662.

[3] D. Donoho and J. Jin. Higher Criticism for detecting sparse heterogeneous mixtures, *Annals of Statistics*, **32**, (2004) 962-994.

[4] B. Garel. Recent asymptotic results in testing for mixtures, *Comput. Statist. Data. Anal.*, **51** (2007) 5295-5304.

[5] Y. Ingster. Minimax detection of a signal for l^n-balls, *Mathematical Methods of Statistics*, **7** (1999), 401-428.

Outline

Goal of this contribution :

- provide an estimation of both $\lambda^\star$ and $\mu^\star$ (ϕ is known).
- handle the case where $\lambda^\star, |\mu^\star| \to 0$ as $n \to +\infty$.
- discussion on the 'direct' and 'inverse' point of views.
- establish lower bounds.

All the results are available in a multivariate setting. For the sake of simplicity, we only consider the case $d = 1$ along this talk.

Outline

Existing results

- Likelihood methods : No available analytic expression for mixture models : EM algorithms required.
  - Initialisation of the EM ?
  - Robustness issues for related models (see, e.g., Baraud, Birgé and Sart (2017)).
- Arias -Castro and Verzelen (2016) : estimation and clustering in a multivariate setting. The parameter $\lambda^\star$ is fixed and known $(= 1/2)$. The density ϕ is available (Gaussian ?).
- Butucea and Vandekerkhove (2010) : Estimation of both ϕ (supposed to be symmetric) and the mixture parameters (fixed). Asymptotic normality.

Existing results

- Collier, Comminges and Tsybakov (2017) : Estimation of linear and quadratic functionals in the sequence model. The sparsity parameter $s(\sim n\lambda^\star)$ is assumed to be known.
- Heinrich and Kahn (2015) : Convergence rates with Wasserstein distance when the component number is unknown. The mixture components are fixed with respect to n.
- Bunea et als. (2010) : SPADES and mixture models. Algorithm based on the $\mathbb{L}^2$ distance with a sparsity penalization. The compatibility condition does not allow to handle situations where the mixture parameters are close to each others.

An estimator based on the $\mathbb{L}^2$ distance

For all $(\lambda, \mu) \in [0, 1] \times \mathbb{R}$, define

$$f_{\lambda,\mu} = (1 - \lambda)\phi(.) + \lambda\phi(. - \mu).$$

The term $\|f_{\lambda,\mu} - f^\star\|^2$ can be estimated (with bias) by

$$\|f_{\lambda,\mu}\|^2 - \frac{2}{n}\sum_{i=1}^{n} f_{\lambda,\mu}(X_i).$$

Given a grid $\mathcal{M}$ on μ, we define

$$(\hat{\lambda}, \hat{\mu}) = \arg\min_{(\lambda,\mu)\in[0,1]\times\mathcal{M}} \left[\|f_{\lambda,\mu}\|^2 - \frac{2}{n}\sum_{i=1}^{n} f_{\lambda,\mu}(X_i)\right],$$

and $\hat{f} = f_{\hat{\lambda},\hat{\mu}}$.

An estimator based on the $\mathbb{L}^2$ distance

Our estimation strategy is based on the estimation of the convoluted density $f^\star$ (direct problem). We expect to recover informations on the underlying discrete mixture measure $G^\star$ (inverse problem).

Similar approaches (in different setting) in, e.g.,

- Laurent et al. (2011),
- Lepski (2016),
- Blanchard et al. (2016),

Outline

An estimator based on the $\mathbb{L}^2$ distance

Using classical tools, we can easily establish the following oracle inequality :

Proposition

$$\mathbb{E}\|\hat{f} - f^\star\|^2 \lesssim \inf_{\lambda,\mu\in\mathcal{M}} \|f_{\lambda,\mu} - f^\star\|^2 + \frac{\log^2(|\mathcal{M}|)}{n}.$$

Question : Can we retrieve convergence results from this inequality ?

Minoration of the $\mathbb{L}^2$ distance

Using simple algebra

$$\begin{aligned}&\|\hat{f} - f^\star\|^2 \\ &= \|(1-\hat{\lambda})\phi + \hat{\lambda}\phi(.-\hat{\mu}) - (1-\lambda^\star)\phi - \lambda^\star\phi(.-\mu^\star)\|^2,\end{aligned}$$

Minoration of the $\mathbb{L}^2$ distance

Using simple algebra

$$
\begin{aligned}
&\|\hat{f} - f^\star\|^2 \\
&\quad = \|(1-\hat{\lambda})\phi + \hat{\lambda}\phi(.-\hat{\mu}) - (1-\lambda^\star)\phi - \lambda^\star\phi(.-\mu^\star)\|^2, \\
&\quad = \|(\lambda^\star - \hat{\lambda})\{\phi - \phi(.-\mu^\star)\} + \lambda^\star\{\phi(.-\hat{\mu}) - \phi(.-\mu^\star)\}\|^2,
\end{aligned}
$$

Minoration of the $\mathbb{L}^2$ distance

Using simple algebra

$$
\begin{aligned}
&\|\hat{f} - f^\star\|^2 \\
&= \|(1-\hat{\lambda})\phi + \hat{\lambda}\phi(.-\hat{\mu}) - (1-\lambda^\star)\phi - \lambda^\star\phi(.-\mu^\star)\|^2, \\
&= \|(\lambda^\star - \hat{\lambda})\{\phi - \phi(.-\mu^\star)\} + \lambda^\star\{\phi(.-\hat{\mu}) - \phi(.-\mu^\star)\}\|^2, \\
&= (\lambda^\star - \hat{\lambda})^2\|\phi - \phi(.-\hat{\mu})\|^2 + (\lambda^\star)^2\|\phi(.-\hat{\mu}) - \phi(.-\mu^\star)\|^2
\end{aligned}
$$

Minoration of the $\mathbb{L}^2$ distance

Using simple algebra

$$
\begin{aligned}
&\|\hat{f} - f^\star\|^2 \\
&\quad = \quad \|(1-\hat{\lambda})\phi + \hat{\lambda}\phi(.-\hat{\mu}) - (1-\lambda^\star)\phi - \lambda^\star\phi(.-\mu^\star)\|^2, \\
&\quad = \quad \|(\lambda^\star - \hat{\lambda})\{\phi - \phi(.-\mu^\star)\} + \lambda^\star\{\phi(.-\hat{\mu}) - \phi(.-\mu^\star)\}\|^2, \\
&\quad = \quad (\lambda^\star - \hat{\lambda})^2\|\phi - \phi(.-\hat{\mu})\|^2 + (\lambda^\star)^2\|\phi(.-\hat{\mu}) - \phi(.-\mu^\star)\|^2 \\
&\qquad\qquad\quad +2(\lambda^\star - \hat{\lambda})\langle\phi - \phi_{\hat{\mu}}, \phi_{\hat{\mu}} - \phi_\mu\rangle, \\
&\quad \gtrsim \quad ??
\end{aligned}
$$

Question : How can we handle the scalar product in the last equality ?

Assumptions

Assumption $\mathcal{H}_S$ **:** $\phi \in C^3(\mathbb{R}) \cap L^2(\mathbb{R})$

Assumption $\mathcal{H}_{Lip}$ **:** There exists $g \in L^2(\mathbb{R})$ s.t.

$$|\phi(x) - \phi(x-\mu)| \leq |\mu| g(x) \quad \forall x \in \mathbb{R},\ \forall \mu \in [-\mu_{max}; \mu_{\text{max}}],$$

where

$$J := \int_{\mathbb{R}} g^2(x) \phi^{-1}(x) dx < +\infty.$$

Assumptions satisfies by, e.g., Gaussian, Cauchy, Laplace (with slight modification), ...

Consequence

Proposition

If the shape ϕ satisfies $\mathcal{H}_S$ and $\mathcal{H}_{Lip}$, then, for all $a, b \in \mathbb{R}$

$$|\langle \phi - \phi_a, \phi_{a+b} - \phi_a \rangle| \leq \|\phi - \phi_a\| \|\phi_{a+b} - \phi_a\| (1 - c\|\phi - \phi_{a+b}\|).$$

for some positive constant c.

Remark : The classical Cauchy-Schwarz inequality provides $c = 0$. It is improved if $a + b$ is 'far away' from 0 ('correlation' property).

Consequence

Using the previous inequality with $a = \hat{\mu}$ and $b = \mu^\star - \hat{\mu}$, we get

$$\begin{aligned}
&\|\hat{f} - f^\star\|^2 \\
&\quad = (\lambda^\star - \hat{\lambda})^2 \|\phi - \phi(. - \hat{\mu})\|^2 + (\lambda^\star)^2 \|\phi(. - \hat{\mu}) - \phi(. - \mu^\star)\|^2 \\
&\qquad\qquad + 2(\lambda^\star - \hat{\lambda}) \langle \phi - \phi_{\hat{\mu}}, \phi_{\hat{\mu}} - \phi_{\mu} \rangle,
\end{aligned}$$

Consequence

Using the previous inequality with $a = \hat{\mu}$ and $b = \mu^\star - \hat{\mu}$, we get

$$
\begin{aligned}
&\|\hat{f} - f^\star\|^2 \\
&\quad = \quad (\lambda^\star - \hat{\lambda})^2 \|\phi - \phi(. - \hat{\mu})\|^2 + (\lambda^\star)^2 \|\phi(. - \hat{\mu}) - \phi(. - \mu^\star)\|^2 \\
&\qquad\qquad +2(\lambda^\star - \hat{\lambda})\langle \phi - \phi_{\hat{\mu}}, \phi_{\hat{\mu}} - \phi_{\mu} \rangle, \\
&\quad \geq \quad (\lambda^\star - \hat{\lambda})^2 \|\phi - \phi(. - \hat{\mu})\|^2 + (\lambda^\star)^2 \|\phi(. - \hat{\mu}) - \phi(. - \mu^\star)\|^2 \\
&\qquad\quad -2|\lambda^\star - \hat{\lambda}|\lambda^\star \|\phi - \phi_{\hat{\mu}}\| \|\phi_{\hat{\mu}} - \phi_{\mu^\star}\| (1 - c\|\phi - \phi_{\mu^\star}\|),
\end{aligned}
$$

Consequence

Using the previous inequality with $a = \hat{\mu}$ and $b = \mu^\star - \hat{\mu}$, we get

$$
\begin{aligned}
&\|\hat{f} - f^\star\|^2 \\
&= (\lambda^\star - \hat{\lambda})^2 \|\phi - \phi(. - \hat{\mu})\|^2 + (\lambda^\star)^2 \|\phi(. - \hat{\mu}) - \phi(. - \mu^\star)\|^2 \\
&\qquad +2(\lambda^\star - \hat{\lambda})\langle \phi - \phi_{\hat{\mu}}, \phi_{\hat{\mu}} - \phi_\mu \rangle, \\
&\geq (\lambda^\star - \hat{\lambda})^2 \|\phi - \phi(. - \hat{\mu})\|^2 + (\lambda^\star)^2 \|\phi(. - \hat{\mu}) - \phi(. - \mu^\star)\|^2 \\
&\quad -2|\lambda^\star - \hat{\lambda}|\lambda^\star \|\phi - \phi_{\hat{\mu}}\| \|\phi_{\hat{\mu}} - \phi_{\mu^\star}\| (1 - c\|\phi - \phi_{\mu^\star}\|), \\
&\gtrsim (\lambda^\star - \hat{\lambda})^2 \|\phi - \phi_{\hat{\mu}}\|^2 \|\phi - \phi_{\mu^\star}\|^2 + (\lambda^\star)^2 \|\phi_{\hat{\mu}} - \phi_{\mu^\star}\|^2 \|\phi - \phi_{\mu^\star}\|^2.
\end{aligned}
$$

Consequence

Gathering the previous result and the oracle inequality obtained few slides ago, we get, with an appropriate choice for the grid $\mathcal{M}$

$$(\lambda^\star - \hat{\lambda})^2 \|\phi - \phi_{\hat{\mu}}\|^2 \|\phi - \phi_{\mu^\star}\|^2 + (\lambda^\star)^2 \|\phi_{\hat{\mu}} - \phi_{\mu^\star}\|^2 \|\phi - \phi_{\mu^\star}\|^2 \lesssim \frac{\log^2(n)}{n}.$$

The Gaussian case :

$$\|\phi_{\mu_1} - \phi_{\mu_2}\|^2 = \|\phi\|^2 (1 - e^{-\frac{(\mu_1 - \mu_2)^2}{4}}) \quad \forall \mu_1, \mu_2 \in \mathbb{R}.$$

In particular,

$$\mathbb{E}\left[(\lambda^\star - \hat{\lambda})^2 (\mu^\star)^4 + (\lambda^\star)^2 (\mu^\star - \hat{\mu})^2 (\mu^\star)^2\right] \lesssim \frac{\log^2(n)}{n}.$$

The Gaussian case

$$\mathbb{E}\left[(\lambda^\star - \hat{\lambda})^2(\mu^\star)^4 + (\lambda^\star)^2(\mu^\star - \hat{\mu})^2(\mu^\star)^2\right] \lesssim \frac{\log^2(n)}{n}.$$

In particular

$$\mathbb{E}\left[(\lambda^\star)^2(\mu^\star)^2(\mu^\star - \hat{\mu})^2\right] \lesssim \frac{\log^2(n)}{n},$$

or equivalently

$$\mathbb{E}\left[\left(\frac{\hat{\mu}}{\mu^\star} - 1\right)^2\right] \lesssim \frac{\log^2(n)}{n(\lambda^\star)^2(\mu^\star)^2}.$$

The Gaussian case

$$\mathbb{E}\left[\left(\frac{\hat{\mu}}{\mu^\star} - 1\right)^2\right] \lesssim \frac{\log^2(n)}{n(\lambda^\star)^2(\mu^\star)^2}.$$

In particular, we have a consistant estimation as soon as

$$n(\lambda^\star)^2(\mu^\star)^4 >> 1 \quad \Leftrightarrow \quad \lambda^\star|\mu^\star|^2 >> \frac{1}{\sqrt{n}}.$$

In a similar setting, Donoho and Jin (2004) test

$$H_0 : \left| \begin{array}{l} \lambda^\star = 0 \\ \mu^\star = 0 \end{array} \right. \quad \text{against} \quad H_1 : \ \lambda^\star|\mu^\star| \gtrsim \frac{1}{\sqrt{n}}$$

In some sense, the detection problem is easier.

Outline

Additional result

The $\mathbb{L}^2$ distance between two mixture densities $f_{\lambda_1,\mu_1}, f_{\lambda_2,\mu_2}$ can be related to the Wasserstein distance between the discrete mixture distribution $G_{\lambda_1,\mu_1}, G_{\lambda_2,\mu_2}$.

The Wasserstein $\mathbb{L}^p$-transportation distances between two probability measures m_1 and m_2 on Ω are defined as

$$W_p^p(m_1, m_2) := \inf_{\pi \in \Pi(m_1, m_2)} \int \|x - y\|_p^p d\pi(x, y),$$

where $\Pi(m_1, m_2)$ is the set of probability measures on $\Omega \times \Omega$ having marginals m_1 and m_2.

Additional result

Proposition

For any density that satisfies $(\mathbf{H}_S)$ and $(\mathbf{H}_{Lip})$, there exists a constant c_ϕ such that

$$\|f_{\lambda_1,\mu_1} - f_{\lambda_2,\mu_2}\| \geq c_\phi W_2^2(G_1, G_2),$$

for all $\lambda_1, \lambda_2 \in (0,1)$, $\mu_1, \mu_2 \in [-M; M]$, where

$$G_1 = (1-\lambda_1)\delta_0 + \lambda_1\delta_{\mu_1} \quad \text{and} \quad G_1 = (1-\lambda_1)\delta_0 + \lambda_1\delta_{\mu_1}.$$

In some sense, the direct problem allows to recover informations on the inverse problem.

Conclusion

In order to complete this discussion, it is possible to obtain corresponding lower bounds (not trivial : the loss is not symmetric).

Possible outcomes

- Higher number (unknown) of components ?
- Unknown shape ϕ ?

S. Gadat, J. Kahn, C. Marteau and C. Maugis-Rabusseau. Parameter recovery in two-component contamination mixtures : the $\mathbb{L}^2$ strategy. arXiv :1604.00306.

Parameter recovery in two-component contamination mixtures: the $\mathbb{L}^2$ strategy

S. Gadat [5], J. Kahn [6], C. Marteau [7] and Cathy Maugis-Rabusseau [8] .

Mathematical Methods of Modern Statistics
CIRM 2017

5. Université Toulouse I - Toulouse School of Economics
6. Université Toulouse III - IMT
7. Université Lyon I - Institut Camille Jordan
8. INSA de Toulouse - IMT