Applications des estimateurs d’Excess-Mass

Mathilde Mougeot

ParisX, LPMA Universities

joint work with C. Butucea, K. Tribouley
Outline

- Excess mass estimation
- Performances through Numerical results (Butucea et al, EJS, 2007)
- Application: estimation of the number of modes
Introduction

- $X_1, \ldots, X_n$ be i.i.d. observations in $\mathbb{R}^d$, $d \geq 1$ having distribution function $F$, density function $f$.

- $E_f(\nu) = \int_{\mathbb{R}^d} (f(x) - \nu) 1\{(f(x)-\nu)>0\} \, dx$, $\forall \nu \in \mathbb{R}$

$f$ is $d$ multi variate density  

Excess Mass
Previous works

- Dip test of multimodality, Hartigan & Hartigan (’85)
- Excess mass estimates and tests for multi modality, Muller & Sawitzki, (’91)
- Estimating density contours clusters, Polonik, (’95)
- Non parametric estimation of density level sets, Tsybakov, (’97).
- Plug-in estimator, Rigollet & Vert (2006)
Excess mass estimation

Our approach:

- We relate the problem to estimating integrated functionals:
  \[ \mathcal{E}_f(\nu) = \int \Phi_\nu(f), \quad \Phi_\nu(u) = (|u| - \nu)_+ \]
- Estimation procedure: (Lepski et al ’99)
  1. Approximation of the functional \( \Phi_\nu \) by \( A_N \Phi_\nu \)
  2. Estimation of the density \( f \) by \( \hat{f} \)
  3. Plug-in: \( \hat{f} \) in \( A_N \Phi_\nu \)
  4. Estimator of the Excess mass
Approximation of the functionnal $\phi_\lambda$ (step1)

- $(\mathcal{H}) f$ density on $[-1, 1]^d$, $f < \rho < 1$,
  $\Phi_\nu(u) = (|u| - \nu)_+ : [-1, 1] \to [0, 1]$

- Approximation by Fourier series:
  \[ A_N \Phi_\nu(u) \overset{\text{def}}{=} c_0(\nu) + \sum_{k=1}^{N} (c_k(\nu) \cos(\pi ku) + b_k(\nu) \sin(\pi ku)) \]

  where the Fourier coefficients are defined and computed:
  \[
  c_0(\nu) = \frac{\langle 1, \Phi_\nu \rangle}{2} = \frac{(1 - \nu)^2}{2} \\
  c_k(\nu) = \frac{\langle \cos(\pi k \cdot), \Phi_\nu \rangle}{\pi^2 k^2} = \frac{2}{\pi^2 k^2} (\cos(\pi k) - \cos(\pi k \nu)) \\
  b_k(\nu) = \frac{\langle \sin(\pi k \cdot), \Phi_\nu \rangle}{\pi^2 k^2} = 0.
  \]
Estimation of the density $f$ (step 2)

Non parametric estimation of density $f$ (step 2)

- Wavelet Estimation of $f$. scaling function $\phi$ and associated wavelet function $\Psi$

- $f = \sum_k \alpha_{j,k} \Phi_{j,k} + \sum_j \sum_k \beta_{j',k} \Psi_{j',k}$

- $X_1, \ldots, X_n$ be i.i.d. observations in $\mathbb{R}^d$, $d \geq 1$ having density function $f$

  $\hat{f}_j(t) \overset{\text{def}}{=} \sum_k \hat{\alpha}_{jk} \Phi_{jk}(t)$ with $\hat{\alpha}_{jk} = \frac{1}{n} \sum_{i=1}^n \phi_{jk}(X_i)$
Estimation of $A_N \Phi_\nu(f(t))$:

- At level $j$, $A_{N,j}(t) =$
  $c_0(\nu) + \sum_{k=1}^{N} c_k(\nu) \exp(\pi^2 k^2 \lambda_j(t)^2 / 2) \cos(\pi k \hat{f}_j(t))$.

- if $\epsilon \sim \mathcal{N}(0, \lambda^2)$, then
  $e^{\pi^2 k^2 \lambda^2 / 2} E[\cos(\pi k(f(t) + \epsilon))] = \cos(\pi kf(t))$
Estimation of the Excess mass (step 4)

Estimation of \( \mathcal{E}(\nu) \) by

- \( \lambda^2 \) should be estimated by \( \hat{\lambda} \)

\[
\hat{\mathcal{E}}(\nu) = \int_{\mathbb{K}} \hat{A}_{N,j}(t) dt \\
= \sum_{k=0}^{N} c_k(\nu) \int_{\mathbb{K}} K_{k,j,t} \cos\left(\pi k \hat{f}_j(t)\right) dt \\
\text{with } K_{k,j,t} = \exp\left(\frac{\pi^2 k^2}{2} \min\left\{\hat{\lambda}^2_j(t), \gamma \frac{2jd}{n}\right\}\right)
\]
Upper bounds and convergence properties

Besov type smoothness condition for \( f \) related to the wavelet expansion for the density

**Theorem**

for \( n \) large enough

\[
\sup_{f \in \mathcal{F}(m^*) \cap b_{p,q}^s(L)} (n \log n)^{\frac{s}{2s+d}} E_f |\hat{E}^*(\lambda) - \mathcal{E}(\lambda)| \leq C,
\]

where

- \( 2j^* = (n \log n)^{\frac{1}{2s+d}} \), \( N^* = (C_0 n \log n)^{s/(2s+d)} \)
- \( C_0 > 0 \) is a constant smaller than \( \min\{2s, d\} \)
- with \( s > 0 \), \( 1 \leq p \leq \infty \), \( 1 \leq q \leq \infty \), \( L, D, m^* > 0 \) and \( 0 < \rho < 1 \).
Set of studied densities

(a): standard gaussian; (b): mixture of gaussian and uniform; (c): mixture of gauss., laplace; (d): mixture of gauss. with isolated spoke.
Algorithm

Parameters Estimations:

- \( \hat{N} = (C_0 n \log n)^{\frac{s}{d+2s}}, \quad C_0 = d. \)
- \( \hat{h} = (n \log n)^{\frac{1}{d+2s}} \)
- Bootstrap estimation: \( \hat{\lambda}^2(x) \)
  \[
  \hat{E}^*(\nu) = \sum_{k=0}^N c_k(\nu) \int_{\mathbb{K}} \exp \left( \frac{\pi^2 k^2}{2} \hat{\lambda}^2(x) \right) \cos \left( \pi k \hat{f}_n(x) \right) dx.
  \]
  estimation procedure.
- A sequence \( \nu_1, \ldots, \nu_{100} = 1 \) is considered.

Performances

- 20 MonteCarlo simulations
- comparison with the Plug in estimator
### Numerical results

<table>
<thead>
<tr>
<th>( f )</th>
<th>( n )</th>
<th>( \tilde{E}_2^{PI} / \tilde{E}_2^* )</th>
<th>( \tilde{p}_2 )</th>
<th>( \tilde{E}<em>\infty^{PI} / \tilde{E}</em>\infty^* )</th>
<th>( \tilde{p}_\infty )</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>100</td>
<td>0.93</td>
<td>0.45</td>
<td>0.92</td>
<td>0.45</td>
</tr>
<tr>
<td>a</td>
<td>1000</td>
<td>1.19</td>
<td>0.70</td>
<td>1.10</td>
<td>0.75</td>
</tr>
<tr>
<td>a</td>
<td>10000</td>
<td>1.32</td>
<td>0.55</td>
<td>1.02</td>
<td>0.60</td>
</tr>
<tr>
<td>b</td>
<td>100</td>
<td>0.66</td>
<td>0.20</td>
<td>0.75</td>
<td>0.30</td>
</tr>
<tr>
<td>b</td>
<td>1000</td>
<td>1.71</td>
<td>0.90</td>
<td>1.51</td>
<td>0.95</td>
</tr>
<tr>
<td>b</td>
<td>10000</td>
<td>2.06</td>
<td>1.00</td>
<td>1.49</td>
<td>1.00</td>
</tr>
<tr>
<td>c</td>
<td>100</td>
<td>0.51</td>
<td>0.15</td>
<td>0.84</td>
<td>0.30</td>
</tr>
<tr>
<td>c</td>
<td>1000</td>
<td>2.17</td>
<td>0.80</td>
<td>1.37</td>
<td>0.80</td>
</tr>
<tr>
<td>c</td>
<td>10000</td>
<td>3.64</td>
<td>1.00</td>
<td>1.83</td>
<td>1.00</td>
</tr>
<tr>
<td>d</td>
<td>100</td>
<td>1.76</td>
<td>1.00</td>
<td>1.38</td>
<td>1.00</td>
</tr>
<tr>
<td>d</td>
<td>1000</td>
<td>2.82</td>
<td>1.00</td>
<td>1.66</td>
<td>1.00</td>
</tr>
<tr>
<td>d</td>
<td>10000</td>
<td>3.91</td>
<td>1.00</td>
<td>1.95</td>
<td>1.00</td>
</tr>
</tbody>
</table>

(a): standard gaussian; (b): mixture of gaussian and uniform; (c): mixture of 2 gaussian and laplace; (d): mixture of gaussian with isolated spoke.
Set of studied 2D densities

(A): 2D Gaussian. (B): mixture of 2D gaussian and uniform.
(C): mixture of two 2D gaussian. (D): Mixture of three 2D gaussian.
### Numerical results

<table>
<thead>
<tr>
<th>$f$</th>
<th>$n$</th>
<th>$\tilde{E}_2^{PI}/\tilde{E}_2^*$</th>
<th>$\tilde{p}_2$</th>
<th>$\tilde{E}<em>\infty/\tilde{E}</em>\infty^*$</th>
<th>$\tilde{p}_\infty$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>400</td>
<td>1.94</td>
<td>0.95</td>
<td>1.31</td>
<td>0.95</td>
</tr>
<tr>
<td>A</td>
<td>1000</td>
<td>2.41</td>
<td>1.00</td>
<td>1.54</td>
<td>1.00</td>
</tr>
<tr>
<td>A</td>
<td>10000</td>
<td>2.41</td>
<td>1.00</td>
<td>1.58</td>
<td>1.00</td>
</tr>
<tr>
<td>B</td>
<td>400</td>
<td>2.25</td>
<td>1.00</td>
<td>1.38</td>
<td>1.00</td>
</tr>
<tr>
<td>B</td>
<td>1000</td>
<td>2.53</td>
<td>1.00</td>
<td>1.44</td>
<td>1.00</td>
</tr>
<tr>
<td>B</td>
<td>10000</td>
<td>2.39</td>
<td>1.00</td>
<td>1.45</td>
<td>1.00</td>
</tr>
<tr>
<td>C</td>
<td>400</td>
<td>2.13</td>
<td>1.00</td>
<td>1.40</td>
<td>1.00</td>
</tr>
<tr>
<td>C</td>
<td>1000</td>
<td>2.09</td>
<td>1.00</td>
<td>1.47</td>
<td>1.00</td>
</tr>
<tr>
<td>C</td>
<td>10000</td>
<td>2.83</td>
<td>1.00</td>
<td>1.71</td>
<td>1.00</td>
</tr>
<tr>
<td>D</td>
<td>400</td>
<td>1.03</td>
<td>0.55</td>
<td>1.02</td>
<td>0.55</td>
</tr>
<tr>
<td>D</td>
<td>1000</td>
<td>1.04</td>
<td>0.85</td>
<td>1.03</td>
<td>0.85</td>
</tr>
<tr>
<td>D</td>
<td>10000</td>
<td>1.06</td>
<td>1.00</td>
<td>1.04</td>
<td>1.00</td>
</tr>
</tbody>
</table>

(A): 2D Gaussian. (B): mixture of 2D gaussian and uniform. (C): mixture of two 2D gaussian. (D): Mixture of three 2D gaussian.
Applications: Excess Mass and number of modes

Local extrema of $f(x)$  \[ \eta = 3 \text{ extrema} \]
\[ \triangleright M = 2 \text{ modes} \]

$\alpha$-Cusps of $\mathcal{E}$  \[ \text{Cusp} = 4 \]
\[ \triangleright M = \frac{\text{Cusp}}{2} \]

- **Definition $\alpha$-Cusp:** A regular function $\Phi$ admits a cusp at $\nu_0$ if there exists $\alpha > 0$ such that: $|\Phi(\nu_0 + h) - \Phi(\nu_0)| \geq C|h|^{\alpha}$ when $h \to 0$ and $C$ positive constant.

- **Methodological point:**
  - keep height of extremum
  - lose of localization.
Applications: Excess Mass and number of modes

Local extrema of $f(x)$  \( \alpha \)-Cusps of $\mathcal{E}$

\[
\eta = 3 \text{ extrema} \quad \text{Cusp} = 4 \quad \triangleright M = \frac{\text{Cusp}}{2}
\]

- Definition \( \alpha \)-Cusp: A regular function $\Phi$ admits a cusp at $\nu_0$ if there exists $\alpha > 0$ such that: $|\Phi(\nu_0 + h) - \Phi(\nu_0)| \geq C|h|^\alpha$ when $h \to 0$ and $C$ positive constant.

- Methodological point:
  - keep height of extremum
  - lose of localization.
Applications: Excess Mass and number of modes

Local extrema of $f(x)$  \[\alpha\]-Cusps of $\mathcal{E}$

\[\eta = 3 \text{ extrema} \quad Cusp = 4\]
\[\triangleright M = 2 \text{ modes} \quad \triangleright M = \frac{Cusp}{2}\]

- **Definition** $\alpha$-Cusp: A regular function $\Phi$ admits a cusp at $\nu_0$ if there exists $\alpha > 0$ such that: $|\Phi(\nu_0 + h) - \Phi(\nu_0)| \geq C|h|^\alpha$ when $h \to 0$ and $C$ positive constant.

- **Methodological point:**
  - keep height of extremum
  - lose of localization.

...
Applications: Excess Mass and number of modes

Local extrema of \( f(x) \)  \( \alpha \)-Cusps of \( E \)

\[ \eta = 3 \text{ extrema} \]
\[ M = 2 \text{ modes} \]
\[ Cusp = 4 \]
\[ M = \frac{Cusp}{2} \]

- **Definition \( \alpha \)-Cusp:** A regular function \( \Phi \) admits a cusp at \( \nu_0 \) if there exists \( \alpha > 0 \) such that:
  \[ |\Phi(\nu_0 + h) - \Phi(\nu_0)| \geq C|h|^{\alpha} \]
  when \( h \to 0 \) and \( C \) positive constant.

- **Methodological point:**
  - keep height of extremum
  - lose of localization.
Wavelets decomposition

\[ \mathcal{E} = \sum_k \alpha_{j_0,k} \Phi_{j_0,k} + \sum_{j_0} \sum_k \beta_{j,k} \Psi_{j,k} \]

\( j_0, j \) denotes frequencies
\( j_0 \) denotes localization parameter
\( \Phi \): mother wavelet, \( \Psi \): father wavelet

- \( \mathcal{E} = \text{LowFrequencyContent} \oplus \text{High-Frequency-Details} \)
- Wavelet Decomposition are used to detect \( \mathcal{E} \) singularities
Wavelet variation coefficients

- $\Delta_{j,k} = \beta_{j,k} - \beta_{j,k-1}$
  Wavelet coefficients differences (Raimondo, 1998)

- **Hypothesis:** $\mathcal{E}(\lambda)$ is $r$-regular anywhere excepted at $\lambda_0$ where an $\alpha$-Cusp exists.

- **Properties of $\Delta_{j,k}$:**
  There exist two positive constant $c_1$ and $c_2$
  
  \[
  |\Delta_{j,k}| \leq c_1 2^{-j(r+1/2)} \quad k \notin \{k_0 - 1, k_0, k_0 + 1\}
  \]
  
  \[
  |\Delta_{j,k}| \geq c_2 2^{-j(r+1/2)} \quad k = k_0
  \]

- But: $r, \alpha$ unknown...and practically many different $\alpha$-Cusps...
Automatic Sharp points detection on \( \mathcal{E} \)

- **H**: Distribution of \( \Delta_{j,k} \sim \mathcal{N}(0, \sigma^2) \)
- Universal Threshold (Donoho et al. 1994)
  \[
  T = \sigma_j \sqrt{2 \log(n_j)} \quad n_j = \#\{\beta_{j,k}, k\}, \sigma_j^2 \text{ at level } j
  \]

- Adapated Threshold \( T^* = \sigma_j^* \sqrt{2 \log(n)} \) (\( \sigma_j^* \) adjusted var.)
Deterministic Densities (d=1)

<table>
<thead>
<tr>
<th>K</th>
<th>$M_{TD}$ (sd)</th>
<th>$M_{T^*}$ (sd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.83 (0)</td>
<td>1.0 (0)</td>
</tr>
<tr>
<td>2</td>
<td>1.44 (0.36)</td>
<td>1.88 (0.43)</td>
</tr>
<tr>
<td>3</td>
<td>1.80 (0.33)</td>
<td>2.72 (0.25)</td>
</tr>
<tr>
<td>4</td>
<td>2.42 (0.28)</td>
<td>3.64 (0.24)</td>
</tr>
</tbody>
</table>

Mode Computation:
▷ Number of "real" modes are often less than the number of designed densities.
▷ Mode with same height are confounded in $\mathcal{E}(\lambda)$
Deterministic Densities ($d=1$)

<table>
<thead>
<tr>
<th>$K$</th>
<th>$M_{T_D} \text{ (sd)}$</th>
<th>$M_{T^*} \text{ (sd)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.83 (0)</td>
<td>1.0 (0)</td>
</tr>
<tr>
<td>2</td>
<td>1.44 (0.36)</td>
<td>1.88 (0.43)</td>
</tr>
<tr>
<td>3</td>
<td>1.80 (0.33)</td>
<td>2.72 (0.25)</td>
</tr>
<tr>
<td>4</td>
<td>2.42 (0.28)</td>
<td>3.64 (0.24)</td>
</tr>
</tbody>
</table>

Mode Computation:

▷ Number of ”real” modes are often less than the number of designed densities.
▷ Mode with same height are confounded in $E(\lambda)$
**Deterministic Densities (d=1)**

<table>
<thead>
<tr>
<th>K</th>
<th>$M_{TD}$ (sd)</th>
<th>$M_{TD^*}$ (sd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.83 (0)</td>
<td>1.0 (0)</td>
</tr>
<tr>
<td>2</td>
<td>1.44 (0.36)</td>
<td>1.88 (0.43)</td>
</tr>
<tr>
<td>3</td>
<td>1.80 (0.33)</td>
<td>2.72 (0.25)</td>
</tr>
<tr>
<td>4</td>
<td>2.42 (0.28)</td>
<td>3.64 (0.24)</td>
</tr>
</tbody>
</table>

**Mode Computation:**
- Number of "real" modes are often less than the number of designed densities.
- Mode with same height are confounded in $E(\lambda)$
Random Samples d=1

<table>
<thead>
<tr>
<th>K</th>
<th>$\hat{M}_T$ (sd)</th>
<th>$\hat{M}_{T^*}$ (sd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.98 (0.29)</td>
<td>1.40 (0.33)</td>
</tr>
<tr>
<td>2</td>
<td>1.56 (0.42)</td>
<td>1.90 (0.34)</td>
</tr>
<tr>
<td>3</td>
<td>2.18 (0.53)</td>
<td>2.81 (0.17)</td>
</tr>
<tr>
<td>4</td>
<td>2.66 (0.39)</td>
<td>3.37 (0.27)</td>
</tr>
</tbody>
</table>

Mode Estimation:
▷ Limits of density estimation by fixed Kernel (R procedure)
▷ Mode of same height are confounded in $\mathcal{E}(\lambda)$ and not detected
Random Samples $d=1$

<table>
<thead>
<tr>
<th>$K$</th>
<th>$\hat{M}_T$ (sd)</th>
<th>$\hat{M}_{T^*}$ (sd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.98 (0.29)</td>
<td>1.40 (0.33)</td>
</tr>
<tr>
<td>2</td>
<td>1.56 (0.42)</td>
<td>1.90 (0.34)</td>
</tr>
<tr>
<td>3</td>
<td>2.18 (0.53)</td>
<td>2.81 (0.17)</td>
</tr>
<tr>
<td>4</td>
<td>2.66 (0.39)</td>
<td>3.37 (0.27)</td>
</tr>
</tbody>
</table>

**Mode Estimation:**
- Limits of density estimation by fixed Kernel (R procedure)
- Mode of same height are confounded in $E(\lambda)$ and not detected
Conclusion

- Excess mass functional estimation
- Performances on numerical results
- Estimation of the number of modes