Modèles de convolution semi-paramétriques

Session: Modèles bruités avec bruit inconnu ou partiellement connu

C. Matias

CNRS, Laboratoire Statistique & Génome, Évry, FRANCE

http://stat.genopole.cnrs.fr/∼cmatias

Outline

Outline

Convolution models

Classical model

Observations $Y_1, \ldots, Y_n$ i.i.d. such that $\quad Y_k = X_k + \varepsilon_k$,

- X_k i.i.d. with unknown density f,
- ε_k i.i.d. with known density f^ε,
- $\{X_k\}$ and $\{\varepsilon_k\}$ independent.

Observations density: $f^Y(y) = \int f^\varepsilon(y-x) f(x) dx = (f^\varepsilon * f)(y)$.
Corresponding Fourier transforms: $\quad \Phi^Y(u) = \Phi(u)\Phi^\varepsilon(u)$.

Applications

- Mendelsohn & Rice (82): fluorometric data,
- Carroll & Hall (88): nonparametric empirical Bayes pbm (prior estimation for location parameters),
- Errors-in-variable regression models $\begin{cases} Y_k & = X_k + \varepsilon_k \\ Z_k & = r(X_k) + \eta_k \end{cases}$

Known noise distribution = Setup not realistic !

Convolution models

Classical model

Observations $Y_1, \ldots, Y_n$ i.i.d. such that $\quad Y_k = X_k + \varepsilon_k,$

- X_k i.i.d. with unknown density f,
- ε_k i.i.d. with known density f^ε,
- $\{X_k\}$ and $\{\varepsilon_k\}$ independent.

Observations density: $f^Y(y) = \int f^\varepsilon(y-x) f(x)dx = (f^\varepsilon * f)(y)$.
Corresponding Fourier transforms: $\quad \Phi^Y(u) = \Phi(u)\Phi^\varepsilon(u)$.

Applications

- Mendelsohn & Rice (82): fluorometric data,
- Carroll & Hall (88): nonparametric empirical Bayes pbm (prior estimation for location parameters),
- Errors-in-variable regression models $\begin{cases} Y_k & = X_k + \varepsilon_k \\ Z_k & = r(X_k) + \eta_k \end{cases}$

Known noise distribution = Setup not realistic !

Convolution models

Classical model

Observations $Y_1, \dots, Y_n$ i.i.d. such that $\quad Y_k = X_k + \varepsilon_k,$

- X_k i.i.d. with unknown density f,
- ε_k i.i.d. with known density f^ε,
- $\{X_k\}$ and $\{\varepsilon_k\}$ independent.

Observations density: $f^Y(y) = \int f^\varepsilon(y-x) f(x) dx = (f^\varepsilon * f)(y).$
Corresponding Fourier transforms: $\quad \Phi^Y(u) = \Phi(u)\Phi^\varepsilon(u).$

Applications

- Mendelsohn & Rice (82): fluorometric data,
- Carroll & Hall (88): nonparametric empirical Bayes pbm (prior estimation for location parameters),
- Errors-in-variable regression models $\begin{cases} Y_k & = X_k + \varepsilon_k \\ Z_k & = r(X_k) + \eta_k \end{cases}$

Known noise distribution = Setup not realistic !

Convolution models

Alternatives

- Repeated measurements: observe an independent sample of the noise distribution: $\varepsilon'_1, \ldots, \varepsilon'_m$ i.i.d $\sim f^{\varepsilon}$.
- Modelling the noise: semiparametric convolution models.
- Assumptions on the distributions supports.
- $\ldots$

Outline

Semiparametric convolution model

Model

Observations $Y_1, \dots, Y_n$ i.i.d. such that $Y_k = X_k + \varepsilon_k$,

- X_k i.i.d. with unknown density f,
- ε_k i.i.d. with partially known density f^ε,
- $\{X_k\}$ and $\{\varepsilon_k\}$ independent.

Only specific forms of f^ε may be identifiable.

Examples

1) Unknown Gaussian noise variance: $Y_k = X_k + \sigma\varepsilon_k$,, where σ is unknown and $\varepsilon_k \sim \mathcal{N}(0,1)$.Observations density:

$$f^Y(y) = \int \frac{1}{\sigma} f^\varepsilon\left(\frac{y-\theta}{\sigma}\right) f(\theta)d\theta = \left[\frac{1}{\sigma} f^\varepsilon\left(\frac{\cdot}{\sigma}\right) * f\right](y).$$

May be viewed as a continuous mixture model.

Semiparametric convolution model

Model

Observations $Y_1, \ldots, Y_n$ i.i.d. such that $\quad Y_k = X_k + \varepsilon_k,$

- X_k i.i.d. with unknown density f,
- ε_k i.i.d. with partially known density f^ε,
- $\{X_k\}$ and $\{\varepsilon_k\}$ independent.

Only specific forms of f^ε may be identifiable.

Examples

1) Unknown Gaussian noise variance: $Y_k = X_k + \sigma\varepsilon_k$,, where σ is unknown and $\varepsilon_k \sim \mathcal{N}(0,1)$.Observations density:

$$f^Y(y) = \int \frac{1}{\sigma} f^\varepsilon\left(\frac{y-\theta}{\sigma}\right) f(\theta)d\theta = \left[\frac{1}{\sigma} f^\varepsilon\left(\frac{\cdot}{\sigma}\right) * f\right](y).$$

May be viewed as a continuous mixture model.

Semiparametric convolution model

Model

Observations $Y_1, \ldots, Y_n$ i.i.d. such that $\quad Y_k = X_k + \varepsilon_k,$

- X_k i.i.d. with unknown density f,
- ε_k i.i.d. with partially known density f^ε,
- $\{X_k\}$ and $\{\varepsilon_k\}$ independent.

Only specific forms of f^ε may be identifiable.

Examples

1) Unknown Gaussian noise variance: $Y_k = X_k + \sigma\varepsilon_k,$, where σ is unknown and $\varepsilon_k \sim \mathcal{N}(0,1)$. Observations density:

$$f^Y(y) = \int \frac{1}{\sigma} f^\varepsilon\left(\frac{y-\theta}{\sigma}\right) f(\theta) d\theta = \left[\frac{1}{\sigma} f^\varepsilon\left(\frac{\cdot}{\sigma}\right) * f\right](y).$$

May be viewed as a continuous mixture model.

Semiparametric convolution model

Model

Observations $Y_1, \ldots, Y_n$ i.i.d. such that $Y_k = X_k + \varepsilon_k,$

- X_k i.i.d. with unknown density f,
- ε_k i.i.d. with partially known density f^ε,
- $\{X_k\}$ and $\{\varepsilon_k\}$ independent.

Only specific forms of f^ε may be identifiable.

Examples

1) Unknown Gaussian noise variance: $Y_k = X_k + \sigma\varepsilon_k,$, where σ is unknown and $\varepsilon_k \sim \mathcal{N}(0,1)$.Observations density:

$$f^Y(y) = \int \frac{1}{\sigma} f^\varepsilon\left(\frac{y-\theta}{\sigma}\right) f(\theta) d\theta = \left[\frac{1}{\sigma} f^\varepsilon\left(\frac{\cdot}{\sigma}\right) * f\right](y).$$

May be viewed as a continuous mixture model.

Semiparametric convolution model

Examples

2) Unknown scale parameter of a stable noise: ε_k i.i.d. with stable density f^ε and Fourier transform $\Phi^\varepsilon(u) = \exp(-|\sigma u|^s)$ where $s > 0$ is known and σ is unknown.

3) Unknown smoothing parameter of a stable noise: Same context but with $s > 0$ unknown and σ is known.

For those 3 examples, under additional assumptions on the density f of X_k, the model parameters are identifiable.

Aims

- Estimate the finite dimensional parameters (σ or s),
- Use a plug-in technique in the methods for known noise density case,
- Evaluate its impact on estimation/goodness-of-fit testing on f (minimax risk setting).

Semiparametric convolution model

Examples

2) Unknown scale parameter of a stable noise: ε_k i.i.d. with stable density f^ε and Fourier transform $\Phi^\varepsilon(u) = \exp(-|\sigma u|^s)$ where $s > 0$ is known and σ is unknown.
3) Unknown smoothing parameter of a stable noise: Same context but with $s > 0$ unknown and σ is known.

For those 3 examples, under additional assumptions on the density f of X_k, the model parameters are identifiable.

Aims

- Estimate the finite dimensional parameters (σ or s),
- Use a plug-in technique in the methods for known noise density case,
- Evaluate its impact on estimation/goodness-of-fit testing on f (minimax risk setting).

Semiparametric convolution model

Examples

2) Unknown scale parameter of a stable noise: ε_k i.i.d. with stable density f^ε and Fourier transform $\Phi^\varepsilon(u) = \exp(-|\sigma u|^s)$ where $s > 0$ is known and σ is unknown.
3) Unknown smoothing parameter of a stable noise: Same context but with $s > 0$ unknown and σ is known.

For those 3 examples, under additional assumptions on the density f of X_k, the model parameters are identifiable.

Aims

- Estimate the finite dimensional parameters (σ or s),
- Use a plug-in technique in the methods for known noise density case,
- Evaluate its impact on estimation/goodness-of-fit testing on f (minimax risk setting).

Semiparametric convolution model

Examples

2) Unknown scale parameter of a stable noise: ε_k i.i.d. with stable density f^ε and Fourier transform $\Phi^\varepsilon(u) = \exp(-|\sigma u|^s)$ where $s > 0$ is known and σ is unknown.
3) Unknown smoothing parameter of a stable noise: Same context but with $s > 0$ unknown and σ is known.

For those 3 examples, under additional assumptions on the density f of X_k, the model parameters are identifiable.

Aims

- Estimate the finite dimensional parameters (σ or s),
- Use a plug-in technique in the methods for known noise density case,
- Evaluate its impact on estimation/goodness-of-fit testing on f (minimax risk setting).

Some general results about convolution

- Regularity assumptions are needed. Usually, two classes of regularities

  super smooth (SS): $|\Phi(u)| \sim_{+\infty} c\exp(-\alpha|u|^r)$.
  ex: Gaussian, Cauchy, stable laws, Student, logistic, EVD...
  ordinary smooth (OS): $|\Phi(u)| \sim_{+\infty} c|u|^{-\beta}$.
  ex: χ^2, Gamma, Laplace, Exponential...

- In general, the rates of convergence are slow. Example: SS noise + OS signal = logarithmic rate.
- The smoother is the noise, the lower are the rates of deconvolution.
- For fixed noise regularity, faster rates are obtained for more regular signal densities.

Some general results about convolution

- Regularity assumptions are needed. Usually, two classes of regularities

  super smooth (SS): $|\Phi(u)| \sim_{+\infty} c\exp(-\alpha|u|^r)$.
  ex: Gaussian, Cauchy, stable laws, Student, logistic, EVD...
  ordinary smooth (OS): $|\Phi(u)| \sim_{+\infty} c|u|^{-\beta}$.
  ex: χ^2, Gamma, Laplace, Exponential...

- In general, the rates of convergence are slow. Example: SS noise + OS signal = logarithmic rate.
- The smoother is the noise, the lower are the rates of deconvolution.
- For fixed noise regularity, faster rates are obtained for more regular signal densities.

Some general results about convolution

- Regularity assumptions are needed. Usually, two classes of regularities

  super smooth (SS): $|\Phi(u)| \sim_{+\infty} c\exp(-\alpha|u|^r)$.
  ex: Gaussian, Cauchy, stable laws, Student, logistic, EVD...
  ordinary smooth (OS): $|\Phi(u)| \sim_{+\infty} c|u|^{-\beta}$.
  ex: χ^2, Gamma, Laplace, Exponential...
- In general, the rates of convergence are slow. Example: SS noise + OS signal = logarithmic rate.
- The smoother is the noise, the lower are the rates of deconvolution.
- For fixed noise regularity, faster rates are obtained for more regular signal densities.

Some general results about convolution

- Regularity assumptions are needed. Usually, two classes of regularities

  super smooth (SS): $|\Phi(u)| \sim_{+\infty} c\exp(-\alpha|u|^r)$.
  ex: Gaussian, Cauchy, stable laws, Student, logistic, EVD…
  ordinary smooth (OS): $|\Phi(u)| \sim_{+\infty} c|u|^{-\beta}$.
  ex: χ^2, Gamma, Laplace, Exponential…
- In general, the rates of convergence are slow. Example: SS noise + OS signal = logarithmic rate.
- The smoother is the noise, the lower are the rates of deconvolution.
- For fixed noise regularity, faster rates are obtained for more regular signal densities.

Some general results about convolution

noise / signal	OS	SS
OS	n^{-a}	$(\log n)^{-a}$
SS	$\frac{(\log n)^a}{\sqrt{n}}$	$\exp(-c(\log n)^a), a < 1$

These results exist with adaptive/minimax/optimal versions, for different risks (pointwise, $\mathbb{L}_2, \mathbb{L}_p, \dots$).

Main differences in the semiparametric setting

The parameter may or may not act as a nuisance. We illustrate this in two cases:

The scale parameter: a real nuisance (Butucea, CM)

- Estimation of the parameter is the one who determines the rates.
- Rates for the unknown density f are overall slower than in the case of known noise distribution.
- In particular, lower bounds can not be deduced from the known σ case.

The smoothness parameter: free adaptation (Butucea, CM, Pouet)

- Rates of convergence for the parameter are slow, but overall faster than those for estimating f.
- Thus adaptation with respect to noise smoothness may be achieved with no loss in the convergence rates.

Main differences in the semiparametric setting

The parameter may or may not act as a nuisance. We illustrate this in two cases:

The scale parameter: a real nuisance (Butucea, CM)

- Estimation of the parameter is the one who determines the rates.
- Rates for the unknown density f are overall slower than in the case of known noise distribution.
- In particular, lower bounds can not be deduced from the known σ case.

The smoothness parameter: free adaptation (Butucea, CM, Pouet)

- Rates of convergence for the parameter are slow, but overall faster than those for estimating f.
- Thus adaptation with respect to noise smoothness may be achieved with no loss in the convergence rates.

Main differences in the semiparametric setting

The parameter may or may not act as a nuisance. We illustrate this in two cases:

The scale parameter: a real nuisance (Butucea, CM)

- Estimation of the parameter is the one who determines the rates.
- Rates for the unknown density f are overall slower than in the case of known noise distribution.
- In particular, lower bounds can not be deduced from the known σ case.

The smoothness parameter: free adaptation (Butucea, CM, Pouet)

- Rates of convergence for the parameter are slow, but overall faster than those for estimating f.
- Thus adaptation with respect to noise smoothness may be achieved with no loss in the convergence rates.

Outline

Parameter estimation: scale parameter case

Model: $Y_k = X_k + \sigma\varepsilon_k$, where σ unknown and

Assumptions

- SS noise: $b\exp(-|u|^s) \leq |\Phi^\varepsilon(u)| \leq B\exp(-|u|^s)$, for large enough $|u|$, s known,
- $\exists r \in (0;s), \alpha > 0$ such that $|\Phi(u)| \geq c\exp(-\alpha|u|^r)$, for large enough $|u|$.

Estimation

Observe that for $u > 0$, the function

$$|F(\tau,u)| = |\Phi^Y(u)|e^{(\tau u)^s} = |\Phi(u)|e^{(\tau^s-\sigma^s)u^s} \underset{u\to\infty}{\longrightarrow} \begin{cases} 0 & \text{if } \tau \leq \sigma \\ +\infty & \text{if } \tau > \sigma \end{cases}$$

Estimate F by $\widehat{F}_n(\tau,u) = \hat{\Phi}_n^Y(u)e^{(\tau u)^s}$. Let $(u_n) \nearrow +\infty$ and

$$\widehat{\sigma}_n = \inf\{\tau, \tau > 0, |\widehat{F}_n(\tau,u_n)| \geq 1\}.$$

Parameter estimation: scale parameter case

Model: $Y_k = X_k + \sigma\varepsilon_k$, where σ unknown and

Assumptions

- SS noise: $b\exp(-|u|^s) \leq |\Phi^\varepsilon(u)| \leq B\exp(-|u|^s)$, for large enough $|u|$, s known,
- $\exists r \in (0; s), \alpha > 0$ such that $|\Phi(u)| \geq c\exp(-\alpha|u|^r)$, for large enough $|u|$.

Identifiability

We have $\Phi^Y(u) = \Phi(u)\Phi^\varepsilon(\sigma u)$ and thus $\lim_{|u|\to\infty} \frac{\log|\Phi^Y(u)|}{|u|^s} = -\sigma^s$.

Estimation

Observe that for $u > 0$, the function

$$|F(\tau, u)| = |\Phi^Y(u)|e^{(\tau u)^s} = |\Phi(u)|e^{(\tau^s - \sigma^s)u^s} \underset{u\to\infty}{\longrightarrow} \begin{cases} 0 & \text{if } \tau \leq \sigma \\ +\infty & \text{if } \tau > \sigma \end{cases}$$

Parameter estimation: scale parameter case

Model: $Y_k = X_k + \sigma\varepsilon_k$, where σ unknown and

Assumptions

- SS noise: $b\exp(-|u|^s) \leq |\Phi^\varepsilon(u)| \leq B\exp(-|u|^s)$, for large enough $|u|$, s known,
- $\exists r \in (0;s), \alpha > 0$ such that $|\Phi(u)| \geq c\exp(-\alpha|u|^r)$, for large enough $|u|$.

Estimation

Observe that for $u > 0$, the function

$$|F(\tau,u)| = |\Phi^Y(u)|e^{(\tau u)^s} = |\Phi(u)|e^{(\tau^s-\sigma^s)u^s} \underset{u\to\infty}{\longrightarrow} \begin{cases} 0 & \text{if } \tau \leq \sigma \\ +\infty & \text{if } \tau > \sigma \end{cases}$$

Estimate F by $\widehat{F}_n(\tau,u) = \hat{\Phi}_n^Y(u)e^{(\tau u)^s}$. Let $(u_n) \nearrow +\infty$ and

$$\widehat{\sigma}_n = \inf\{\tau, \tau > 0, |\widehat{F}_n(\tau,u_n)| \geq 1\}.$$

Parameter estimation: scale parameter case

Convergence results (Butucea, CM)

- The previous estimator is consistent.
- When the signal is SS, rate of convergence $= O((\log n)^{r/s-1})$.
- When the signal is OS, rate of convergence $= O\left(\frac{\log\log n}{\log n}\right)$.
- Those rates of convergence are minimax and lower than the classical rate for estimating f.

Parameter estimation: smoothing parameter case

Model: $Y_k = X_k + \varepsilon_k$, where

Assumptions

- SS stable noise: $\Phi^{\varepsilon}(u) = \exp(-|\sigma u|^s)$ and σ known, s unknown,
- $\exists A, \beta' > 0$ such that $|\Phi(u)| \geq A|u|^{-\beta'}$, for large enough $|u|$.

Parameter estimation: smoothing parameter case

Model: $Y_k = X_k + \varepsilon_k$, where

Assumptions

- SS stable noise: $\Phi^{\varepsilon}(u) = \exp(-|\sigma u|^s)$ and σ known, s unknown,
- $\exists A, \beta' > 0$ such that $|\Phi(u)| \geq A|u|^{-\beta'}$, for large enough $|u|$.

Identifiability of (f, s)

Assume $\Phi_1^Y = \Phi_2^Y$, where $\Phi_i^Y(u) = \Phi_i(u)e^{-|\sigma_i u|^{s_i}}, i = 1, 2$ and $s_1 \leq s_2$. Then we get

Parameter estimation: smoothing parameter case

Model: $Y_k = X_k + \varepsilon_k$, where

Assumptions

- SS stable noise: $\Phi^{\varepsilon}(u) = \exp(-|\sigma u|^s)$ and σ known, s unknown,
- $\exists A, \beta' > 0$ such that $|\Phi(u)| \geq A|u|^{-\beta'}$, for large enough $|u|$.

Identifiability of (f, s)

Assume $\Phi_1^Y = \Phi_2^Y$, where $\Phi_i^Y(u) = \Phi_i(u)e^{-|\sigma_i u|^{s_i}}, i = 1, 2$ and $s_1 \leq s_2$. Then we get

$$\log|\Phi_1(u)| - |\sigma_1 u|^{s_1} = \log|\Phi_2(u)| - |\sigma_2 u|^{s_2}$$

Parameter estimation: smoothing parameter case

Model: $Y_k = X_k + \varepsilon_k$, where

Assumptions

- SS stable noise: $\Phi^{\varepsilon}(u) = \exp(-|\sigma u|^s)$ and σ known, s unknown,
- $\exists A, \beta' > 0$ such that $|\Phi(u)| \geq A|u|^{-\beta'}$, for large enough $|u|$.

Identifiability of (f, s)

Assume $\Phi_1^Y = \Phi_2^Y$, where $\Phi_i^Y(u) = \Phi_i(u)e^{-|\sigma_i u|^{s_i}}, i = 1, 2$ and $s_1 \leq s_2$. Then we get

$$|u|^{-s_1} \log|\Phi_1(u)| - \sigma_1^{s_1} = |u|^{-s_1} \log|\Phi_2(u)| - \sigma_2^{s_2}|u|^{s_2 - s_1}$$

Parameter estimation: smoothing parameter case

Model: $Y_k = X_k + \varepsilon_k$, where

Assumptions

- SS stable noise: $\Phi^\varepsilon(u) = \exp(-|\sigma u|^s)$ and σ known, s unknown,
- $\exists A, \beta' > 0$ such that $|\Phi(u)| \geq A|u|^{-\beta'}$, for large enough $|u|$.

Identifiability of (f, s)

Assume $\Phi_1^Y = \Phi_2^Y$, where $\Phi_i^Y(u) = \Phi_i(u)e^{-|\sigma_i u|^{s_i}}, i = 1, 2$ and $s_1 \leq s_2$. Then we get

$$\lim_{|u|\to\infty} |u|^{-s_1} \log|\Phi_1(u)| - \sigma_1^{s_1} = \lim_{|u|\to\infty} |u|^{-s_1} \log|\Phi_2(u)| - \sigma_2^{s_2}|u|^{s_2 - s_1}$$

which implies $s_1 = s_2$, $\sigma_1 = \sigma_2$ and then $\Phi_1 = \Phi_2$.

Parameter estimation: smoothing parameter case

Model: $Y_k = X_k + \varepsilon_k$, where

Assumptions

- SS stable noise: $\Phi^{\varepsilon}(u) = \exp(-|\sigma u|^s)$ and σ known, s unknown,
- $\exists A, \beta' > 0$ such that $|\Phi(u)| \geq A|u|^{-\beta'}$, for large enough $|u|$.

Estimation in $[\underline{s}; \bar{s}]$

- Construct a grid $\mathcal{S}_n = \{\underline{s} = s_1 < s_2 < \cdots < s_N = \bar{s}\}$
- Note that for large enough $|u|$, there exists some k s.t.

$$[q_{\beta'}\Phi^k](u) \leq |\Phi^Y(u)| \leq \Phi^k(u)$$

  where $q_{\beta'}(u) = A|u|^{-\beta'}$ and $\Phi^k(u) = \exp(-\gamma|u|^{s_k})$.
- Let $u_n \to \infty$ and select $\hat{s}_n =$ index k on the grid $\mathcal{S}_n$ such that $|\hat{\Phi}^Y(u_n)|$ is closest to the interval $[[q_{\beta'}\Phi^k](u_n); \Phi^k(u_n)]$.

Parameter estimation: smoothing parameter case

Convergence results (Butucea, CM, Pouet)

- The previous estimator is consistent.
- Its rate of convergence is logarithmic but faster than the classical rate for estimating f.
- This rate of convergence is minimax.

Outline

Plug-in estimator for f

Classical deconvolution estimator

$$\widehat{f}_n(x) = \frac{1}{nh}\sum_{j=1}^{n} \tilde{K}\left(\frac{Y_i - x}{h}\right) \quad \text{where } \Phi^{\tilde{K}}(u) = \frac{\Phi^K(u)}{\Phi^\varepsilon(u/h)}.$$

Deconvolution estimator when scale parameter unknown

$$\widehat{f}_{n,\widehat{\sigma}}(x) = \frac{1}{nh\widehat{\sigma}}\sum_{j=1}^{n} \tilde{K}\left(\frac{Y_i - x}{h\widehat{\sigma}}\right) \quad \text{where } \Phi^{\tilde{K}}(u) = \frac{\Phi^K(u)}{\Phi^\varepsilon(u/h)}.$$

Deconvolution estimator when smoothing parameter unknown

$$\hat{f}_n(x) = \frac{1}{n\hat{h}_n}\sum_{j=1}^{n} \hat{K}_n\left(\frac{Y_j - x}{\hat{h}_n}\right), \; \Phi^{\hat{K}_n}(u) = \Phi^K(u)\exp\left\{\left(\frac{|u|}{\hat{h}_n}\right)^{\hat{s}_n}\right\}$$

$$\text{and } \hat{h}_n = \left(\frac{\log n}{2} - \frac{\bar{\beta} - \hat{s}_n + 1/2}{\hat{s}_n}\log\log n\right)^{-1/\hat{s}_n}.$$

Plug-in estimator for f

Classical deconvolution estimator

$$\widehat{f}_n(x) = \frac{1}{nh} \sum_{j=1}^{n} \tilde{K}\left(\frac{Y_i - x}{h}\right) \quad \text{where } \Phi^{\tilde{K}}(u) = \frac{\Phi^K(u)}{\Phi^\varepsilon(u/h)}.$$

Deconvolution estimator when scale parameter unknown

$$\widehat{f}_{n,\widehat{\sigma}}(x) = \frac{1}{nh\widehat{\sigma}} \sum_{j=1}^{n} \tilde{K}\left(\frac{Y_i - x}{h\widehat{\sigma}}\right) \quad \text{where } \Phi^{\tilde{K}}(u) = \frac{\Phi^K(u)}{\Phi^\varepsilon(u/h)}.$$

Difficulty: Kernel estimator with random bandwidth $h\widehat{\sigma}$.
Solution: Moments bounds for empirical processes.

Deconvolution estimator when smoothing parameter unknown

$$\hat{f}_n(x) = \frac{1}{n\hat{h}_n} \sum_{j=1}^{n} \hat{K}_n\left(\frac{Y_j - x}{\hat{h}_n}\right), \; \Phi^{\hat{K}_n}(u) = \Phi^K(u) \exp\left\{\left(\frac{|u|}{\hat{h}_n}\right)^{\hat{s}_n}\right\}$$

$$\text{and } \hat{h}_n = \left(\frac{\log n}{2} - \frac{\bar{\beta} - \hat{s}_n + 1/2}{\hat{s}_n} \log\log n\right)^{-1/\hat{s}_n}.$$

Plug-in estimator for f

Classical deconvolution estimator

$$\widehat{f}_n(x) = \frac{1}{nh}\sum_{j=1}^{n} \tilde{K}\left(\frac{Y_i - x}{h}\right) \quad \text{where } \Phi^{\tilde{K}}(u) = \frac{\Phi^{K}(u)}{\Phi^{\varepsilon}(u/h)}.$$

Deconvolution estimator when scale parameter unknown

$$\widehat{f}_{n,\widehat{\sigma}}(x) = \frac{1}{nh\widehat{\sigma}}\sum_{j=1}^{n} \tilde{K}\left(\frac{Y_i - x}{h\widehat{\sigma}}\right) \quad \text{where } \Phi^{\tilde{K}}(u) = \frac{\Phi^{K}(u)}{\Phi^{\varepsilon}(u/h)}.$$

Deconvolution estimator when smoothing parameter unknown

$$\hat{f}_n(x) = \frac{1}{n\hat{h}_n}\sum_{j=1}^{n} \hat{K}_n\left(\frac{Y_j - x}{\hat{h}_n}\right), \; \Phi^{\hat{K}_n}(u) = \Phi^{K}(u)\exp\left\{\left(\frac{|u|}{\hat{h}_n}\right)^{\hat{s}_n}\right\}$$

$$\text{and } \hat{h}_n = \left(\frac{\log n}{2} - \frac{\bar{\beta} - \hat{s}_n + 1/2}{\hat{s}_n}\log\log n\right)^{-1/\hat{s}_n}.$$

Estimation of f: performances of $\hat{f}_n$

Unknown scale parameter case (Butucea, CM)

- The rate of convergence for $\hat{f}_n$ is the same as for $\hat{\sigma}_n$.
- When the signal is SS, rate of convergence $= O((\log n)^{r/s-1})$.
- When the signal is OS, rate of convergence $= O\left(\frac{\log\log n}{\log n}\right)$.
- Those rates of convergence are minimax and lower than the classical rates for estimating f.

The scale parameter is thus a nuisance which limits the performances of estimation of f.

Unknown smoothness parameter case (Butucea, CM, Pouet)

- It is possible to estimate f when s is unknown with the classical rates of convergence for deconvolution.
- Such a plug-in procedure is then automatically minimax and adaptive w.r.t s.

Estimation of f: performances of $\hat{f}_n$

Unknown scale parameter case (Butucea, CM)

- The rate of convergence for $\hat{f}_n$ is the same as for $\hat{\sigma}_n$.
- When the signal is SS, rate of convergence $= O((\log n)^{r/s-1})$.
- When the signal is OS, rate of convergence $= O\left(\frac{\log\log n}{\log n}\right)$.
- Those rates of convergence are minimax and lower than the classical rates for estimating f.

The scale parameter is thus a nuisance which limits the performances of estimation of f.

Unknown smoothness parameter case (Butucea, CM, Pouet)

- It is possible to estimate f when s is unknown with the classical rates of convergence for deconvolution.
- Such a plug-in procedure is then automatically minimax and adaptive w.r.t s.

Nonparametric goodness-of-fit testing for f

Framework (unknown smoothness parameter)

- OS Signal belongs to Sobolev class $\mathcal{S}(\beta, L) = \left\{ f : \mathbb{R} \to \mathbb{R}_+, \int f = 1, \frac{1}{2\pi} \int |\Phi(u)|^2 |u|^{2\beta} du \leq L \right\}$,
- SS noise with unknown smoothness parameter s.
- We want to test $H_0 : f = f_0$ versus $H_1(\mathcal{C}, \Psi_n) : f \in \cup_{\beta \in [\underline{\beta}, \bar{\beta}]} \{ f \in \mathcal{S}(\beta, L) \text{ and } \psi_{n,\beta}^{-2} \|f - f_0\|_2^2 \geq \mathcal{C} \}$.

Remarks

- We test $f = f_0$ rather than $f^Y = f_0^Y$.
- We consider alternatives expressed in $\mathbb{L}_2$-norm, thus the problem is strongly related to estimation of $\int (f - f_0)^2$.

Nonparametric goodness-of-fit testing for f

Framework (unknown smoothness parameter)

- OS Signal belongs to Sobolev class $\mathcal{S}(\beta, L) = \left\{ f : \mathbb{R} \to \mathbb{R}_+, \int f = 1, \frac{1}{2\pi} \int |\Phi(u)|^2 |u|^{2\beta} du \leq L \right\}$,
- SS noise with unknown smoothness parameter s.
- We want to test $H_0 : f = f_0$ versus $H_1(\mathcal{C}, \Psi_n) : f \in \cup_{\beta \in [\underline{\beta}, \bar{\beta}]} \{ f \in \mathcal{S}(\beta, L)$ and $\psi_{n,\beta}^{-2} \| f - f_0 \|_2^2 \geq \mathcal{C} \}$.

Remarks

- We test $f = f_0$ rather than $f^Y = f_0^Y$.
- We consider alternatives expressed in $\mathbb{L}_2$-norm, thus the problem is strongly related to estimation of $\int (f - f_0)^2$.

Nonparametric goodness-of-fit testing for f

Framework (unknown smoothness parameter)

- OS Signal belongs to Sobolev class $\mathcal{S}(\beta, L) = \left\{ f : \mathbb{R} \to \mathbb{R}_+, \int f = 1, \frac{1}{2\pi} \int |\Phi(u)|^2 |u|^{2\beta} du \leq L \right\}$,
- SS noise with unknown smoothness parameter s.
- We want to test $H_0 : f = f_0$ versus $H_1(\mathcal{C}, \Psi_n) : f \in \cup_{\beta \in [\underline{\beta}, \bar{\beta}]} \{ f \in \mathcal{S}(\beta, L)$ and $\psi_{n,\beta}^{-2} \| f - f_0 \|_2^2 \geq \mathcal{C} \}$.

Remarks

- We test $f = f_0$ rather than $f^Y = f_0^Y$.
- We consider alternatives expressed in $\mathbb{L}_2$-norm, thus the problem is strongly related to estimation of $\int (f - f_0)^2$.

Nonparametric goodness-of-fit testing for f

Approach

- (Upper-bound) $\forall \epsilon \in (0;1)$, exhibit $\Delta_n^\star$ s.t. $\exists \mathcal{C}^0 > 0$, with $\forall \mathcal{C} > \mathcal{C}^0$,

$$\overline{\lim_{n\to\infty}} \sup_{s\in[\underline{s},\bar{s}]} \left\{ \mathbb{P}_{f_0,s}[\Delta_n^\star = 1] + \sup_{f\in H_1(\mathcal{C},\Psi_n)} \mathbb{P}_{f,s}[\Delta_n^\star = 0] \right\} \leq \epsilon.$$

- (Lower bound) $\exists \mathcal{C}_0 > 0$ s.t. $\forall 0 < \mathcal{C} < \mathcal{C}_0$,

$$\underline{\lim}_{n\to\infty} \inf_{\Delta_n} \sup_{s\in[\underline{s},\bar{s}]} \left\{ \mathbb{P}_{f_0,s}[\Delta_n = 1] + \sup_{f\in H_1(\mathcal{C},\Psi_n)} \mathbb{P}_{f,s}[\Delta_n = 0] \right\} \geq \epsilon,$$

where the infimum is taken over all test statistics Δ_n.

Nonparametric goodness-of-fit testing for f

Approach

- (Upper-bound) $\forall \epsilon \in (0;1)$, exhibit $\Delta_n^\star$ s.t. $\exists \mathcal{C}^0 > 0$, with $\forall \mathcal{C} > \mathcal{C}^0$,

$$\overline{\lim_{n\to\infty}} \sup_{s\in[\underline{s},\bar{s}]} \left\{ \mathbb{P}_{f_0,s}[\Delta_n^\star = 1] + \sup_{f\in H_1(\mathcal{C},\Psi_n)} \mathbb{P}_{f,s}[\Delta_n^\star = 0] \right\} \leq \epsilon.$$

- (Lower bound) $\exists \mathcal{C}_0 > 0$ s.t. $\forall 0 < \mathcal{C} < \mathcal{C}_0$,

$$\underline{\lim}_{n\to\infty} \inf_{\Delta_n} \sup_{s\in[\underline{s},\bar{s}]} \left\{ \mathbb{P}_{f_0,s}[\Delta_n = 1] + \sup_{f\in H_1(\mathcal{C},\Psi_n)} \mathbb{P}_{f,s}[\Delta_n = 0] \right\} \geq \epsilon,$$

where the infimum is taken over all test statistics Δ_n.

Goodness-of-fit test: procedure

The test statistic

Define

$$\hat{T}_n^0 = \frac{2}{n(n-1)} \sum_{1 \le k < j \le n} < \frac{1}{\hat{h}_n} \hat{K}_n\Big(\frac{\cdot - Y_k}{\hat{h}_n}\Big) - f_0 \ , \ \frac{1}{\hat{h}_n} \hat{K}_n\Big(\frac{\cdot - Y_j}{\hat{h}_n}\Big) - f_0 >$$

and

$$\Delta_n^\star = \begin{cases} 1 & \text{if } |\hat{T}_n^0| \hat{t}_n^{-2} > \mathcal{C}^\star \\ 0 & \text{otherwise,} \end{cases}$$

for some constant $\mathcal{C}^\star > 0$ and a **random** threshold $\hat{t}_n^2$ to be specified.

Goodness-of-fit-test: results

Theorem (Butucea, CM, Pouet)

For any $f_0 \in \mathcal{S}(\bar\beta, L)$*, choose*

$$\hat{t}_n^2 = \left(\frac{\log n}{2}\right)^{-2\bar\beta/\hat{s}_n} \quad ; \quad \hat{h}_n = \left(\frac{\log n}{2} - \frac{2\bar\beta}{\hat{s}_n}\log\log n\right)^{-1/\hat{s}_n}$$

and any large enough positive constant $\mathcal{C}^\star$*. The testing procedure satisfies the testing upper-bound for any* $\epsilon \in (0,1)$ *with testing rate*

$$\Psi_n = \{\psi_{n,\beta}\}_{\beta \in [\underline{\beta}, \bar\beta]} \text{ given by } \psi_{n,\beta} = \left(\frac{\log n}{2}\right)^{-\beta/s}.$$

Moreover, if $f_0 \in \mathcal{S}(\bar\beta, cL)$ *for some* $0 < c < 1$ *and if Assumption* **(T)** *holds, then this testing rate is asymptotically adaptive optimal over the family of classes* $\{\mathcal{S}(\beta, L), \beta \in [\underline{\beta}; \bar\beta]\}$ *and for any* $s \in [\underline{s}; \bar{s}]$ *(i.e. the testing lower-bound holds).*

Take home message

Assumptions on the noise distribution have a strong impact on the quality of the estimators.

Outline

Other setups

Dependent observations

- Works by C. Lacour in the HMM context.
- See the following talk by N. Hilgert.

Exotic spaces

Sphere. See the following talk by T. M. Pham Ngoc.