Density deconvolution under general assumptions on measurement error distribution

D. Belomestny[1] and A. Goldenshluger [2]

[1]University of Duisburg-Essen and [2]University of Haifa

9th October 2019

Deconvolution problem

Model

Let

$$Y_i = X_i + \varepsilon_i, \quad i = 1, \ldots, n,$$

where $X_1, \ldots, X_n$ are i.i.d. with a density f and $\varepsilon_1, \ldots, \varepsilon_n$ are i.i.d. with a density g.

Convolution

$$f_Y(y) = [f \star g](y) = \int_{-\infty}^{\infty} f(y-x)g(x)\,dx$$

Goal

Deconvolution, that is, estimation of the density f from the observations $Y_1, \ldots, Y_n$.

Fourier approach

Convolution theorem

$$\mathscr{F}[f_Y](u) = \int e^{iux} f_Y(x)\,dx = \mathscr{F}[f](u)\mathscr{F}[g](u)$$

Deconvolution

The estimator for f is usually based on the ratio

$$\mathscr{F}[f](u) = \frac{\mathscr{F}[f_Y](u)}{\mathscr{F}[g](u)},$$

provided $\mathscr{F}[g](u) \neq 0$ for all $u \in \mathbb{R}$.

Problem

In many situations, for example in the case of the uniformly distributed ε_i, the Fourier transform of g has zeros on real line, see Hall and Meister (2007) and Meister (2008).

Bilateral Laplace transform

Definition

For a generic locally integrable function ψ denote

$$\widehat{\psi}(z) = \int_{\mathbb{R}} e^{-zt}\psi(t)\,dt.$$

The convergence region of the above integral will be denoted by

$$\Sigma_\psi = \{z \in \mathbb{C} : \mathrm{Re}(z) \in (\sigma_\psi^-, \sigma_\psi^+)\}, \quad -\infty \le \sigma_\psi^- \le \sigma_\psi^+ \le \infty.$$

Inverse Laplace transform

$$\psi(t) = \frac{1}{2\pi i}\int_{s-i\infty}^{s+i\infty} \widehat{\psi}(z)e^{zt}\,dz$$

Remark

As compared to Fourier approach there is an additional tuning parameter s.

Construction of estimator

1. Let $\sigma_1 < \sigma_2 < \ldots,$ be distinct real parts of zeros of $\widehat{g}$. Denote

$$S_g = \bigcup_{j=1}^{\infty} S_g^{(j)}, \quad S_g^{(j)} = \{z : \sigma_j < \mathrm{Re}(z) < \sigma_{j+1}\}$$

and

$$\check{S}_g = \bigcup_{j=1}^{\infty} \check{S}_g^{(j)}, \quad \check{S}_g^{(j)} = \{z : -\sigma_{j+1} < \mathrm{Re}(z) < -\sigma_j\}.$$

2. Let K be a kernel with bounded support, that is, $\widehat{K}$ is an entire function.

Estimator

Kernel

Given a real number $h > 0$ put

$$L_{s,h}(t) = \frac{1}{2\pi i} \int_{s-i\infty}^{s+i\infty} \frac{\widehat{K}(zh)}{\widehat{g}(-z)} e^{zt} \, dz$$

for any $s \in \check{S}_g$.

Observation

Note that $L_{s,h}(t)$ is the inverse Laplace transform of the function $\frac{\widehat{K}(zh)}{\widehat{g}(-z)}$.

Estimator

$$\widetilde{f}_{s,h}(x_0) = \frac{1}{n} \sum_{i=1}^{n} L_{s,h}(Y_i - x_0)$$

Motivation

Lemma

Suppose that for any $s \in (\varkappa_g^-, 0) \cup (0, \varkappa_g^+)$ the integral in the definition of $L_{s,h}$ is absolutely convergent and

$$\int |L_{s,h}(y - x_0)| f_Y(y)\, dy < \infty,$$

then

$$\int L_{s,h}(y - x_0) f_Y(y)\, dy = \int \frac{1}{h} K\left(\frac{x - x_0}{h}\right) f(x)\, dx.$$

Observation

By making h smaller, we can reduce the bias of $\widetilde{f}_{s,h}(x_0)$, but the variance of $\widetilde{f}_{s,h}(x_0)$ increases with h !

Assumptions

Assumption (LG)

The Laplace transform $\widehat{g}(z)$ of measurement error distribution exists in a vertical strip $\Sigma_g = \{z \in \mathbb{C} : \sigma_g^- < \mathrm{Re}(z) < \sigma_g^+\}$, $\sigma_g^- < 0 < \sigma_g^+$, and admits the following factorization

$$\widehat{g}(z) = \frac{1}{\widehat{\psi}(z)} \prod_{k=1}^{q} \left(1 - \frac{e^{a_k z}}{\lambda_k}\right)^{m_k},$$

where $\{a_k\}$ are distinct positive real numbers, $\{m_k\}$ are non-negative integer numbers, $|\lambda_k| = 1$, $\forall k = 1, \ldots, q$.

Observation

Assumption (LG) states that $\widehat{g}(z)$ factorizes into a product of two functions: the first function has zeros on the imaginary axis, while the second one does not vanish on $\Sigma_g \setminus \{0\}$. The zeros of $\widehat{g}$ are $z_{k,j} := i(\arg\{\lambda_k\} + 2\pi j)/a_k$

Assumptions

Assumption (PS)

Assume that $\widehat{\psi}(z) \neq 0$ for all $z \in \Sigma_g \backslash \{0\}$, and there exist constants $\omega_0 > 0$, $\gamma > 0$ and $D_1 > 0$, $D_2 > 0$ such that

$$D_1|\omega|^\gamma \leq |\widehat{\psi}(i\omega)| \leq D_2|\omega|^\gamma, \quad \forall |\omega| \geq \omega_0.$$

In addition, for some non-negative integer r and $D_3 > 0$

$$\max_{j=0,\ldots,2r} |\widehat{\psi}^{(j)}(i\omega)| \leq D_3(1 + |\omega|^\gamma), \quad \forall \omega \in \mathbb{R}.$$

Remark

Assumption (PS) is rather standard in density deconvolution problems when it is imposed on $\widehat{g}(i\omega)$: it corresponds to the smooth case. Note however that here it is now imposed on function $\widehat{\psi}(i\omega)$.

Examples

Uniform distribution

Let $\varepsilon \sim U(-\theta, \theta)$; then

$$\widehat{g}(z) = \frac{\sinh(\theta z)}{\theta z} = -\frac{e^{-\theta z}}{2\theta z}(1 - e^{2\theta z}), \quad z \in \mathbb{C}.$$

In this case we have $q = 1$, $m_1 = 1$, $a_1 = 2\theta$, $\lambda_1 = 1$ and

$$\widehat{\psi}(z) = -2\theta z e^{\theta z}$$

- $\widehat{\psi}$ satisfies Assumption (PS) with $\gamma = 1$,
- $\widehat{g}(z)$ has simple zeros on the imaginary axis at $z_k = i\pi k/\theta$, $k = \pm 1, \pm 2, \ldots$
- $S_g = \mathbb{C} \backslash \{z : \mathrm{Re}(z) = 0\}$.

Examples

Convolution of uniform distributions

Consider convolution of distributions $U(-\theta_k, \theta_k)$, $k = 1, \ldots, q$, with distinct parameters $\theta_1, \ldots, \theta_q$, each of multiplicity m_k. In this case

$$\widehat{g}(z) = \prod_{k=1}^{q} \left[\frac{\sinh(\theta_k z)}{\theta_k z} \right]^{m_k} = \frac{\exp\{-z \sum_{k=1}^{q} \theta_k m_k\}}{\prod_{k=1}^{q} (-2\theta_k z)^{m_k}} \prod_{k=1}^{q} (1 - e^{2\theta_k z})^{m_k}, \quad z \in \mathbb{C}.$$

Assumption (LG) holds with $a_k = \theta_k$, $\lambda_k = 1$ for $k = 1, \ldots, q$, and

$$\widehat{\psi}(z) \quad = \quad \prod_{k=1}^{q} (-2\theta_k z)^{m_k} \exp\left\{ z \sum_{k=1}^{q} \theta_k m_k \right\}.$$

Examples

Discrete distributions

Let ε be a discrete random variable taking $2M+1$ values in the set $(j\delta)_{j=-M}^{M}$, $\delta > 0$ with corresponding probabilities $(p_j)_{j=-M}^{M}$. Then

$$\widehat{g}(z) = \sum_{k=-M}^{M} p_k e^{-\delta kz} = e^{-\delta Mz} \sum_{k=0}^{2M} p_{M-k} e^{\delta kz} = e^{-\delta Mz} p_M P(e^{\delta z}),$$

where $P(x) := 1 + \sum_{k=1}^{2M} (p_{M-k}/p_M) x^k$. Let $\lambda_1, \ldots, \lambda_{2M}$ denote the roots of polynomial $P(z)$, then we have

$$\widehat{g}(z) = p_M e^{-\delta Mz} \prod_{k:|\lambda_k|\neq 1} \left(1 - \frac{e^{\delta z}}{\lambda_k}\right) \prod_{k:|\lambda_k|=1} \left(1 - \frac{e^{\delta z}}{\lambda_k}\right).$$

Examples

Discrete distributions

Therefore

$$\widehat{\psi}(z) = \frac{e^{\delta M z}}{p_M \prod_{k:|\lambda_k|\neq 1}(1 - e^{\delta z}/\lambda_k)}, \quad \delta^{-1}\ln(\lambda_-) < \mathrm{Re}(z) < \delta^{-1}\ln(\lambda_+),$$

where $\lambda_- := \max\{|\lambda_k| : |\lambda_k| < 1\}$, and $\lambda_+ := \min\{|\lambda_k| : |\lambda_k| > 1\}$.

- If $\varepsilon \sim \mathrm{Bern}(1/2)$ then

$$\widehat{g}(z) = \frac{1}{2}(1 + e^z)$$

  and (LG) holds with $q = 1$, $a_1 = 1$, $\lambda_1 = -1$, $m_1 = 1$, and $\widehat{\psi}(z) = 2$.
- If $\varepsilon \sim \mathrm{Bin}(m, 1/2)$, then

$$\widehat{g}(z) = 2^{-m}(1 + e^z)^m$$

  and (LG) holds with $q = 1$, $a_1 = 1$, $\lambda_1 = -1$, $m_1 = m$, and $\widehat{\psi}(z) = 2^m$.

Examples

Convolution of uniform and a smooth density

Let φ be a probability density having the Laplace transform $\widehat{\varphi}$ in a strip $\Sigma_\varphi = \{z : \sigma_\varphi^- < \mathrm{Re}(z) < \sigma_\varphi^+\}$ satisfying $|\widehat{\varphi}(z)| \neq 0$, $\forall z \in \Sigma_\varphi$. Assume that

$$|\widehat{\varphi}(i\omega)| \asymp |\omega|^{-\gamma}$$

for some $\gamma > 0$ as $|\omega| \to \infty$; Let g be a convolution of the uniform density on $[-\theta, \theta]$ and φ; then

$$\widehat{g}(z) = \frac{\sinh(\theta z)}{\theta z}\widehat{\varphi}(z) = -\frac{e^{-\theta z}\widehat{\varphi}(z)}{2\theta z}(1 - e^{2\theta z}), \;\; \sigma_\varphi^- < \mathrm{Re}(z) < \sigma_\varphi^+,$$

and (LG) obviously holds with $\widehat{\psi}(z) = -2\theta z e^{\theta z}/\widehat{\varphi}(z)$.

Kernel representation

Under Assumption (LG) the kernel $L_{s,h}$ has a representation

$$L_{s,h}(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\widehat{K}((s+i\omega)h)\,\widehat{\psi}(-s-i\omega)}{\prod_{k=1}^{q}[1-e^{-a_k(s+i\omega)}/\lambda_k]^{m_k}}\, e^{(s+i\omega)t}\mathrm{d}\omega$$

for all $s \in (\varkappa_g^-, 0) \cup (0, \varkappa_g^+)$.

Remark

Note that for any $s \in (\varkappa_g^-, 0) \cup (0, \varkappa_g^+)$ the denominator of the integrand in the representation does not vanish.

Observation

It is convenient to represent $L_{s,h}(t)$ as an infinite series.

Kernel representation

Uniform error density

Let K be continuosly differentiable on $\mathbb{R}$ such that $\widehat{K}(z)$ exists for all $z \in \mathbb{C}$, and

$$\int_{-\infty}^{\infty} |\widehat{K}(s+i\omega)|\,|\omega| \mathrm{d}\omega < \infty, \quad \forall s \in \mathbb{R}.$$

Then

$$L_{s,h}(t) = \begin{cases} L_{+,h}(t), & s > 0, \\ L_{-,h}(t), & s < 0, \end{cases}$$

where

$$L_{+,h}(t) := \frac{2\theta}{h^2} \sum_{j=0}^{\infty} K'\left(\frac{t-\theta(2j+1)}{h}\right)$$

and

$$L_{-,h}(t) := -\frac{2\theta}{h^2} \sum_{j=0}^{\infty} K'\left(\frac{t+\theta(2j+1)}{h}\right).$$

Kernel representation

Convolution of uniform distributions

In this case the corresponding kernel is

$$L_{s,h}(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\widehat{K}((s+i\omega)h)[\theta(s+i\omega)]^m}{[\sinh(\theta(s+i\omega))]^m} e^{(s+i\omega)t} \mathrm{d}\omega, \quad s \neq 0$$

and it has a representation

$$L_{s,h}(t) = \frac{1}{h^{m+1}} \sum_{j=0}^{\infty} C_{j,m} K^{(m)}\left(\frac{t-\theta(m+2j)}{h}\right),$$

where

$$C_{j,m} = \binom{j+m-1}{m-1}$$

is the number of weak compositions of j into m parts.

Convergence rates

Class of densities

For $A > 0$, $\beta > 0$ define a class $\mathscr{H}_{x_0}(A, \beta)$ of functions f such that

$$|f^{(\lfloor\beta\rfloor)}(x) - f^{(\lfloor\beta\rfloor)}(x')| \leq A|x - x'|^{\beta - \lfloor\beta\rfloor}, \quad \forall x, x' \in (x_0 - d, x_0 + d),$$

where $\lfloor\beta\rfloor = \max\{k \in \mathbb{N} \cup \{0\} : k < \beta\}$.

Convergence rates

Kernel K

The kernel K is supported in $[-1, 1]$ and fulfils the following conditions.

(K1) For a fixed positive integer m_0

$$\int_{-1}^{1} K(t)\mathrm{d}t = 1, \int_{-1}^{1} t^j K(t)\mathrm{d}t = 0, \; j = 1, 2, \dots, m.$$

(K2) For a positive integer r kernel K is r times continuously differentiable on $\mathbb{R}$ and

$$\max_{t \in [-1,1]} |K^{(j)}(t)| \leq C_K < \infty, \quad \forall j = 0, 1, \dots, r.$$

Convergence rates

Theorem

Assume that $f \in \mathscr{H}_{x_0}(A,\beta)$, and let (K1)-(K2) hold with $m_0 \geq \beta + 1$ and $r = 1$. Let $\tilde{f}_*(x_0)$ be the estimator associated with kernels L_{+,h_*} if $x_0 \geq 0$ and L_{-,h_*} if $x_0 < 0$, and with bandwidth $h_* := (\theta A^{-2} n^{-1})^{1/(2\beta+3)}$. Then

$$\limsup_{n\to\infty} \left\{ \varphi_n^{-1} \mathscr{R}_n[\tilde{f}_*; \mathscr{H}_{x_0}(A,\beta)] \right\} \leq C, \;\; \varphi_n := A^{3/(2\beta+3)} \left(\frac{\theta}{n}\right)^{\beta/(2\beta+3)},$$

where C may depend on β and $\bar{x}$ only. Here

$$\mathscr{R}_n[\tilde{f}; \mathscr{F}] = \sup_{f\in\mathscr{F}} \left[\mathbb{E}_f |\tilde{f}(x_0) - f(x_0)|^2 \right]^{1/2}.$$

General setting

Let N be a natural number, and denote

$$\mathscr{L}_N := \left\{ a^T j = \sum_{k=1}^{q} a_k j_k : j = (j_1, \dots, j_q) \in \{0, 1, \dots, N\}^q \right\}.$$

The estimator of $f(x_0)$ is defined as follows

$$\tilde{f}_{s,h}^{(N)}(x_0) := \frac{1}{n} \sum_{j=1}^{n} L_{s,h}^{(N)}(Y_j - x_0), \;\; s \in (\varkappa_g^-, 0) \cup (0, \varkappa_g^+),$$

where

$$L_{s,h}^{(N)}(t) := \begin{cases} L_{+,h}^{(N)}(t), & s \in (0, \varkappa_g^+), \\ L_{-,h}^{(N)}(t), & s \in (-\varkappa_g^-, 0), \end{cases}$$

and

$$L_{+,h}^{(N)}(t) := \sum_{\ell \in \mathscr{L}_N} \mathscr{C}_\ell^+ R_h(t - \ell), \;\; L_{-,h}^{(N)}(t) := \sum_{\ell \in \mathscr{L}_N} \mathscr{C}_\ell^- R_h(t + \ell).$$

General setting

Here

$$R_h(t) := \frac{1}{2\pi} \int_{-\infty}^{\infty} \widehat{K}(i\omega h) \widehat{\psi}(-i\omega) e^{i\omega t} \mathrm{d}\omega,$$

$$\mathscr{C}_\ell^+ := \sum_{j : a^T j = \ell} \left[\prod_{k=1}^{q} C_{j_k, m_k} \lambda_k^{-j_k} \right]$$

and

$$\mathscr{C}_\ell^- := \sum_{j : a^T j = \ell} \left[\prod_{k=1}^{q} (-1)^{m_k} C_{j_k, m_k} \lambda_k^{j_k + m_k} \right].$$

Functional class

For $A > 0$, $\alpha > 0$ f belongs to a class $\mathscr{H}_\alpha(A)$ of functions with

$$|f^{(\lfloor\alpha\rfloor)}(x) - f^{(\lfloor\alpha\rfloor)}(x')| \leq A|x - x'|^{\alpha - \lfloor\alpha\rfloor}, \quad \forall x, x' \in \mathbb{R},$$

where $\lfloor\alpha\rfloor = \max\{k \in \mathbb{N} \cup \{0\} : k < \alpha\}$.

Let p be a non–negative integer, $B > 0$, and let $\gamma > 0$ be a constant appearing in Assumption (PS). Assume that f belongs to a class $\mathscr{T}_p(B)$ of densities satisfying

$$\int_{-\infty}^{\infty} |x|^p f(x) \mathrm{d}x \leq B.$$

Error bounds

Let

$$H_{N,j}(\omega) := \begin{cases} \sum_{\ell \in \mathscr{L}_N^*} \mathscr{C}_\ell^+ e^{-i\omega(x_0+\ell)}(x_0+\ell)^{-j}, & x_0 \geq 0, \\ \sum_{\ell \in \mathscr{L}_N^*} \mathscr{C}_\ell^- e^{-i\omega(x_0-\ell)}(x_0-\ell)^{-j}, & x_0 < 0 \end{cases}$$

with $\mathscr{L}_N^* := \mathscr{L}_N \backslash \{0\}$. Let $0 < h < a_{\min} := \min\{a_1, \ldots, a_q\}$ and $r := \max\{j \in \mathbb{N} \cup \{0\} : 2j \leq p\}$; then

$$\mathscr{R}_n\big[\tilde{f}_h^{(N)}; \mathscr{F}_{\alpha,p}\big] \;\leq\; C_1 \left\{ Ah^\alpha + \frac{B}{h(x_0 + a_{\min} N)^p} \right\}$$

$$+\; \frac{C_2 B}{n} \left\{ \frac{1}{h^{2\gamma+1}} + \sum_{l=0}^{r} h^l \int_{-\infty}^{\infty} \left| \widehat{K}^{(l)}(i\omega h) \widehat{\psi}^{(r-l)}(-i\omega) H_{N,r}(\omega) \right|^2 \mathrm{d}\omega \right\}$$

with $\mathscr{F}_{\alpha,p} = \mathscr{T}_p(B) \cap \mathscr{H}_\alpha(A)$.

Error bounds

Assumption

Assume that

$$\sum_{\ell \in \mathscr{L} \setminus \{0\}} \max\{|\mathscr{C}_\ell^+|, |\mathscr{C}_\ell^-|\} \, \ell^{-\nu} \leq C_0 < \infty$$

for some $C_0 > 0$ and $\nu > 1$.

Theorem

Assume that $f \in \mathscr{F}_{\alpha,p}(A,B)$ with $p \geq 2\nu$. Let $h = \left[B(A^2 n)^{-1}\right]^{1/(2\alpha+2\gamma+1)}$ and $N \geq \left(A^{-2\gamma+1} B^{2\gamma+\alpha} n^{\alpha+1}\right)^{1/p(2\alpha+2\gamma+1)}$. Then for large enough n one has

$$\mathscr{R}_{n,\Delta_{x_0}}\left[\tilde{f}_{h_*}^{(N)}; \mathscr{F}_{\alpha,p}(A,B)\right] \leq C_1 A^{\frac{2\gamma+1}{2\alpha+2\gamma+1}} \left(Bn^{-1}\right)^{\frac{\alpha}{2\alpha+2\gamma+1}},$$

where C_1 may depend on α, and p only.

Minimax convergence rates

Observation

It is well known that under standard assumptions *in the smooth case*

$$|\widehat{g}(i\omega)| \neq 0, \quad \widehat{g}(i\omega) \asymp |\omega|^{-\gamma}, \quad |\omega| \to \infty$$

the minimax rate of convergence is

$$\mathscr{R}_n^*[\mathscr{H}_\alpha(A)] \asymp n^{\alpha/(2\alpha+2\gamma+1)}.$$

References

Belomestny, D. and Goldenshluger, J. (2019).
Deconvolution via bilateral Laplace transform.