Density estimation from observations with multiplicative measurement errors

Alexander Goldenshluger

Department of Statistics
University of Haifa
Israel

joint work with

Denis Belomestny

Duisburg-Essen University
Germany

Luminy, Dec 18-22, 2017

Happy Birthday!

 

Till $120 + 120 = 240$!

Outline

- Problem formulation and background
- Preliminaries
  - the Mellin transform
  - identifiability
  - linear functional strategy
- Estimation at a point separated away from zero
  - estimator and risk bounds
- Estimation at zero
  - estimator and risk bounds
- Summary and concluding remarks

Part 1: Problem formulation and background

Problem

- Model: we observe a sample $Y_1, \dots, Y_n$ generated by

$$Y_i = X_i \eta_i, \quad i = 1, \dots, n,$$

  – $X_1, \dots, X_n$ are iid random variables with density f_X;
  – $\eta_1, \dots, \eta_n$ are iid random variables, independent of $X_1, \dots, X_n$, with density g.

- The goal: estimate f_X on the basis of $Y_1, \dots, Y_n$. We consider the problem of estimating f_X at a single given point x_0.

- f_Y is a scale mixture of f_X and g:

$$f_Y(y) = [f_X \star g](y) := \int_{-\infty}^{\infty} \frac{1}{x} f_X(\tfrac{y}{x}) g(x) \mathrm{d}x = \int_{-\infty}^{\infty} \frac{1}{x} g(\tfrac{y}{x}) f_X(x) \mathrm{d}x.$$

Related literature

- Stochastic volatility model: $X_i > 0$, $\eta_i \sim \mathcal{N}(0, 1)$.

  Van Es, Spreij & Van Zanten (2003), Van Es & Spreij (2011),

  Belomestny & Schoemakers (2015)...

- Multiplicative censoring model: $X_i > 0$, $\eta_i \sim U[0, 1]$.

  Vardi (1989), Vardi & Zhang (1992), Andersen & Hansen (2001),

  Belomestny, Comte & Genon-Catalot (2016)...

- Estimation of mixing densities/distributions
  - location models (deconvolution): vast literature...
  - exponential families: Zhang (1990, 95), Loh & Zhang (1997)...
  - demixing of scale mixtures: ???

A naive approach: deconvolution with log-data

- The case of positive r.v.: if $X_i > 0$ and $\eta_i > 0$ then
  - take log–transformation of the data

$$Y_i' = X_i' + \eta_i',\ Y_i' = \ln Y_i,\ X_i' = \ln X_i,\ \eta_i' = \ln \eta_i.$$

  - estimate $f_{X'}(x)$ by a deconvolution technique and transform back using the formula $f_X(x) = \frac{1}{x} f_{X'}(\ln x)$.
- The approach is not applicable when
  - random variables may take negative values
  - estimating at zero: the inverse transformation is not well-defined there.

Research questions

- The problem was studied only for specific distributions of the measurement errors (uniform, normal, Beta), and proposed estimators are designed for these distributions...
- The goal is to provide answers to the following questions:
  - How to construct estimators of f_X in *a principled way* under general assumptions on g?
  - What is the achievable estimation accuracy?
  - What is the accuracy of the deconvolution estimator resulting from the naive approach?

Part 2: Preliminaries

1. The Mellin transform

- The Mellin transform: for a function u supported on $[0, \infty)$

$$\widetilde{u}(z) = \mathcal{M}[u; z] := \int_0^\infty x^{z-1} u(x) \mathrm{d}x,$$

where z takes values in a strip $\Omega_u := \{z \in \mathbb{C} : a < \mathrm{Re}(z) < b\}$

(the strip Ω_u can degenerate to a vertical line).

- If $u(x) = O(x^{-a+\epsilon})$ as $x \to 0+$ and $u(x) = O(x^{-b-\epsilon})$ as $x \to \infty$ then the integral defines an analytic function on Ω_u.

- If u is a probability density then always $\{1 + i\omega, \omega \in \mathbb{R}\} \in \Omega_u$.

- The inversion formula

$$u(x) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} x^{-z} \widetilde{u}(z) \mathrm{d}z, \quad \forall c \in (a, b).$$

2. The Mellin transform

- The Parceval formula: for any c in the common strip of analyticity of $\widetilde{u}(1-z)$ and $\widetilde{v}(z)$

$$\int_0^\infty u(x)v(x)\mathrm{d}x = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\widetilde{u}(1-z)\widetilde{v}(z)\mathrm{d}z.$$

- The Mellin transform of $u \star v$:

$$[\widetilde{u \star v}](z) = \mathcal{M}[u \star v; z] = \mathcal{M}[u; z]\mathcal{M}[v; z].$$

- The bilateral Laplace transform of a function u on $(-\infty, \infty)$

$$\check{u}(z) = \mathcal{L}[u; z] := \int_{-\infty}^{\infty} u(x)e^{-zx}\mathrm{d}x.$$

- The relationship between the Mellin and Laplace transforms

$$\mathcal{M}[u; z] = \int_{-\infty}^{\infty} e^{-zx}u(e^{-x})\mathrm{d}x = \mathcal{L}[w; z], \;\; w(x) := u(e^{-x}).$$

Identifiability

- The issue of identifiability arises since X and η may take negative values.
- Let $g^+(x) := g(x)\mathbf{1}(x > 0)$ and $g^-(x) := g(-x)\mathbf{1}(x > 0)$, and let $\tilde{g}^+(z)$, $\tilde{g}^-(z)$ be the one–sided Mellin transforms defined on

$$\Omega_{g^+} \cap \Omega_{g^-} =: \{z \in \mathbb{C} : a < \mathrm{Re}(z) < b\},$$

$$\Omega_{g^+} \cap \Omega_{g^-} \neq \varnothing \quad \text{since} \quad a < 1 < b.$$

- Lemma: *The probability density f_X is identifiable from $f_Y = f_X \star g$ if and only if $\tilde{g}^+(z) \neq \tilde{g}^-(z)$ almost everywhere on $\Omega_{g^+} \cap \Omega_{g^-}$, or $g(x) \neq g(-x)$ on a set of positive Lebesgue measure.*
  - if g is non–symmetric then f_X is identifiable;
  - if $\mathrm{supp}(g) = [0, \infty)$ then f_X is identifiable.

General idea for estimator construction

- Linear functional strategy

  Find a pair of kernels $K(x, y)$ and $L(x, y)$ such that

  (i) $\int_{-\infty}^{\infty} K(x, y) f_X(y) \mathrm{d}y$ approximates well the value $f_X(x)$ to be recovered;

  (ii) kernel $L(x, y)$ is related to $K(x, y)$ via

$$\int_{-\infty}^{\infty} K(x, y) f_X(y) \mathrm{d}y = \int_{-\infty}^{\infty} L(x, y) f_Y(y) \mathrm{d}y.$$

- Under (i) and (ii)

$$\hat{f}_X(x) = \frac{1}{n} \sum_{i=1}^{n} L(x, Y_i)$$

  is a sensible estimator for $f_X(x)$...

Part 3: Estimation at a point separated away from zero

1. Kernel construction

- Kernel K: for $K : \mathbb{R} \to \mathbb{R}$ and $h > 0$ let

$$K_h(x,y) = \begin{cases} \frac{1}{xh} K\left(\frac{\ln(y/x)}{h}\right), & y/x > 0 \\ 0, & y/x < 0. \end{cases}$$

- Kernel L: for $s \in (1-b, 1-a)$ define

$$L_{s,h}(x,y) :=$$

$$\begin{cases} \dfrac{1}{2\pi i|x|} \displaystyle\int_{s-i\infty}^{s+i\infty} \left|\frac{x}{y}\right|^z \frac{\check{K}(zh)\widetilde{g}^+(1-z)}{[\widetilde{g}^+(1-z)]^2 - [\widetilde{g}^-(1-z)]^2} \mathrm{d}z, & y/x > 0, \\ -\dfrac{1}{2\pi i|x|} \displaystyle\int_{s-i\infty}^{s+i\infty} \left|\frac{x}{y}\right|^z \frac{\check{K}(zh)\widetilde{g}^-(1-z)}{[\widetilde{g}^+(1-z)]^2 - [\widetilde{g}^-(1-z)]^2} \mathrm{d}z, & y/x < 0. \end{cases}$$

2. Kernel construction

- Lemma: *Let $K_h(x,y)$ be as above; if the integrals in the definition of $L_{s,h}(x,y)$ are absolutely convergent then*

$$\int_{-\infty}^{\infty} L_{s,h}(x,y) f_Y(y)\mathrm{d}y = \int_{-\infty}^{\infty} K_h(x,t) f_X(t)\mathrm{d}t.$$

- One can always set $s=0$ because $0 \in (1-b, 1-a)$:

$$L_{0,h}(x,y) = \begin{cases} \frac{1}{2\pi|x|}\int_{-\infty}^{\infty}\left|\frac{x}{y}\right|^{i\omega}\frac{\widehat{K}(\omega h)\tilde{g}^{+}(1-i\omega)}{[\tilde{g}^{+}(1-i\omega)]^2-[\tilde{g}^{-}(1-i\omega)]^2}\mathrm{d}\omega, & \frac{y}{x}>0, \\ -\frac{1}{2\pi|x|}\int_{-\infty}^{\infty}\left|\frac{x}{y}\right|^{i\omega}\frac{\widehat{K}(\omega h)\tilde{g}^{-}(1-i\omega)}{[\tilde{g}^{+}(1-i\omega)]^2-[\tilde{g}^{-}(1-i\omega)]^2}\mathrm{d}\omega, & \frac{y}{x}<0, \end{cases}$$

where $\widehat{K}(\omega) := \check{K}(i\omega)$, $\omega \in \mathbb{R}$, is the Fourier transform of K.

- If $\mathrm{supp}(g) \subseteq [0,\infty)$ then $\tilde{g} \equiv \tilde{g}^{+}$, $\tilde{g}^{-} \equiv 0$, and

$$L_h(x,y) := L_{0,h}(x,y) = \frac{1}{2\pi|x|}\int_{-\infty}^{\infty}\left|\frac{x}{y}\right|^{i\omega}\frac{\widehat{K}(\omega h)}{\tilde{g}(1-i\omega)}\mathrm{d}\omega, \quad \frac{y}{x}>0.$$

Connection to deconvolution with log–data

- The log-transformed data

$$Y_i' = X_i' + \eta_i', \;\; \eta_i' = \ln \eta_i, \; X_i' = \ln X_i, \; Y_i' = \ln Y_i.$$

- The standard deconvolution estimator for $f_{X'}(t_0)$ is

$$\hat{f}_{X'}(t_0) = \frac{1}{2\pi n} \sum_{j=1}^{n} \int_{-\infty}^{\infty} \frac{\widehat{K}(\omega h)}{\widehat{g}_{\eta'}(-\omega)} e^{-i\omega(Y_j'-t_0)} \mathrm{d}\omega;$$

hence, by applying the inverse transformation

$$\hat{f}_X(x_0) = \frac{1}{x_0} \hat{f}_{X'}(\ln x_0) = \frac{1}{2\pi x_0 n} \sum_{j=1}^{n} \int_{-\infty}^{\infty} \frac{\widehat{K}(\omega h)}{\widehat{g}_{\eta'}(-\omega)} e^{-i\omega \ln(Y_i/x_0)} \mathrm{d}\omega$$

- The deconvolution estimator is a specific case of our estimator $\frac{1}{n}\sum_{j=1}^{n} L_{s,h}(x_0, Y_j)$ when $s=0$: if $\eta \sim g$ then

$$g_{\eta'}(x) = e^x g(e^x), \quad \widehat{g}_{\eta'}(\omega) = \mathcal{M}[g; 1+i\omega] = \widetilde{g}(1+i\omega).$$

Assumptions on the error density g

- From now on we assume: $X > 0$, $\eta > 0$,

$$\text{supp}(g) \subseteq [0, \infty), \ \Omega_g =: \{z : \text{Re}(z) \in (a, b)\}, \ \text{supp}(f_X) \subseteq [0, \infty).$$

- By the Riemann–Lebesgue lemma, $\widetilde{g}(z) \to 0$ as $|\text{Im}(z)| \to \infty$.

- Assumption [G1] (polynomial decay)

  For $\sigma \in (a, b)$ there exist positive ω_0, c_0, $B_2 > B_1$, γ such that

  – $\min_{|\omega| \leqslant \omega_0} |\widetilde{g}(\sigma + i\omega)| \geqslant c_0 > 0$

  – $B_1 |\omega|^{-\gamma} \leqslant |\widetilde{g}(\sigma + i\omega)| \leqslant B_2 |\omega|^{-\gamma}, \quad |\omega| \geqslant \omega_0.$

- [G1] stipulates that $|\widetilde{g}(z)|$ does not have zeros on the line $\{\sigma + i\omega, \omega \in \mathbb{R}\}$, and polynomially decreasing as $|\omega| \to \infty$.

Examples

- Example 1 (Beta distribution): let $g(x) = (\nu+1)x^{\nu}/\theta^{\nu+1}$, $x \in (0,\theta)$, $\nu > -1$. Then $a = -\nu$, $b = \infty$, and [G1] holds with $\gamma = 1$

$$|\widetilde{g}(\sigma + i\omega)| = \frac{\theta^{\sigma-1}(\nu+1)}{\sqrt{(\nu+\sigma)^2+\omega^2}}, \quad \forall \sigma \in (-\nu, \infty).$$

- Example 2 (Pareto distribution): let $g(x) = (\nu-1)\theta^{\nu-1}/x^{\nu}$, $x > \theta$, $\theta > 0$ and $\nu > 1$. Then $a = -\infty$, $b = \nu$, and [G1] holds with $\gamma = 1$,

$$|\widetilde{g}(\sigma + i\omega)| = \frac{(\nu-1)\theta^{\sigma-1}}{\sqrt{(\nu-\sigma)^2+\omega^2}}, \quad \forall \sigma \in (-\infty, \nu).$$

- Example 3: the product of two independent uniform on $[0,1]$ random variables has density $g(x) = \ln(1/x)$, $x \in (0,1)$. Here $\widetilde{g}(z) = 1/z^2$, and [G1] is satisfied with $\gamma = 2$.

Estimator

- Kernel K: Let $K : \mathbb{R} \to \mathbb{R}$ satisfy the following conditions:

  (i) $\mathrm{supp}(K) = [-1, 1]$,

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

  (ii) K is q times cont. differentiable, $\max_{j=0,\dots,q} \|K^{(j)}\|_\infty \leqslant C_K$.

- The estimator: for $x_0 > 0$, $s \in (1-b, 1-a)$ and $h > 0$ define

$$\begin{aligned} \hat{f}_{s,h}(x_0) &= \frac{1}{n}\sum_{j=1}^{n} L_{s,h}(x_0, Y_j) \\ &= \frac{1}{2\pi n x_0}\sum_{j=1}^{n}\int_{-\infty}^{\infty}\left(\frac{x_0}{Y_j}\right)^{s+i\omega}\frac{\check{K}((s+i\omega)h)}{\widetilde{g}(1-s-i\omega)}\mathrm{d}\omega. \end{aligned}$$

- If $\widetilde{g}(z)$ does not have zeros for all $z \in \Omega_g$ then the integral does not depend on s; otherwise there is a dependence.

Functional class and pointwise risk

- The local Hölder functional class $\mathscr{H}_{x_0,r}(A,\beta)$

  Let $r > 1$, $A > 0$, $\beta > 0$; we say that density f belongs to $\mathscr{H}_{x_0,r} := \mathscr{H}_{x_0,r}(A,\beta)$ if for $\ell := \max\{k \in \mathbb{N}_0 : k < \beta\}$

$$\max_{k=1,\dots,\ell} |f^{(k)}(x)| \leqslant A, \quad \forall x \in [r^{-1}x_0, rx_0],$$

$$|f^{(\ell)}(x) - f^{(\ell)}(x')| \leqslant A|x - x'|^{\beta-\ell}, \quad \forall x, x' \in [r^{-1}x_0, rx_0].$$

- Remark: condition $f \in \mathscr{H}_{x_0,r}(A,\beta)$ does not put any restrictions on the behavior of f at zero and at infinity...

- The risk of an estimator $\hat{f}$ and the minimax risk:

$$\mathrm{Risk}_{x_0}[\hat{f};\mathscr{H}] = \sup_{f_X \in \mathscr{H}} \left[\mathbb{E}_{f_X}|\hat{f}(x_0) - f_X(x_0)|^2\right]^{1/2}$$

$$\mathrm{Risk}^*_{x_0}[\mathscr{H}] := \inf_{\hat{f}} \mathrm{Risk}_{x_0}[\hat{f};\mathscr{H}]$$

Bounds on the risk

- Theorem 1: *Let [G1] hold with* $\sigma = 1$, $\gamma > 1/2$, *and* $\hat{f}_{h_*}$ *be associated with* $m \geqslant \lfloor \beta \rfloor + 1$, $q > \gamma + 1$, $s = 0$, *and*

$$h = h_* := C_1 [A^2 x_0^{2\beta+2} n]^{-1/(2\beta+2\gamma+1)}.$$

If $x_0 \geqslant C_2$ *then for large* n

$$\mathrm{Risk}_{x_0}[\hat{f}_{h_*}; \mathscr{H}_{x_0,r}] \;\leqslant\; C_3 \, \overbrace{A^{\frac{2\gamma+1}{2\beta+2\gamma+1}} \left(x_0^{2\gamma-1} n^{-1} \right)^{\frac{\beta}{2\beta+2\gamma+1}}}^{\phi_n} .$$

- Theorem 2: *Let* $x_0 \geqslant C_4 > 0$, *[G1] holds with* $\sigma = 1$, $\gamma > 1/2$, *and* $|\widetilde{g}'(1 + i\omega)| \leqslant B|\omega|^{-\gamma}$, $\forall |\omega| \geqslant \omega_0$. *Then*

$$\liminf_{n\to\infty} \left\{ \phi_n^{-1} \mathrm{Risk}^*_{x_0}[\mathscr{H}_{x_0,r}] \right\} \;\geqslant\; C_5.$$

Remarks

- The estimator $\hat{f}_{h_*}$ is rate optimal $\Rightarrow$ the deconvolution estimator based on the log–transformed data is rate optimal.
- The accuracy deteriorates for large x_0; e.g., if $\eta \sim U(0,1)$ then the minimax risk is proportional to $x_0^{\beta/(2\beta+3)}$.
- The Mellin transform $\tilde{f}_X$ of $f_X \in \mathscr{H}_{x_0,r}$ is guaranteed to exist *only on the line* $\{1 + i\omega, \omega \in \mathbb{R}\}$. This is essential in Theorems 1 and 2: the lower bound is achieved on

$$f_X^{(0)}(x) = \frac{1}{\pi x[1 + \ln^2(x/x_0)]}, \quad f_X^{(1)}(x) = f_X^{(0)}(x) + \frac{\theta}{x}\psi\Big(\frac{\ln(x/x_0)}{h}\Big).$$

- One can improve dependence of the risk on x_0 under additional conditions on f_X at zero and/or at infinity.

Improvements and choice of s

- Functional class: for $\alpha > 0$, $M > 0$ define

$$\mathscr{F} = \mathscr{F}_{\alpha,M}(A,\beta) := \mathscr{H}_{x_0,r}(A,\beta) \cap \left\{ f : \int_0^\infty x^{2\alpha} f(x)\mathrm{d}x \leqslant M \right\}.$$

- Theorem 3: *For $\epsilon > 0$ let $s_* := \max\{-\alpha, \frac{1}{2}(1-b)+\epsilon\}$. Let [G1] hold with $\sigma = 1 - s_*$, $\gamma > \frac{1}{2}$. Let $\hat{f}_{s_*,h_*}(x_0)$ be associated with*

$$h = h_* := C_1\big[M^{-1}A^2 x_0^{2\beta+2-2s_*} n\big]^{-1/(2\beta+2\gamma+1)}.$$

If $x_0 \geqslant C_2$ then

$$\mathrm{Risk}_{x_0}[\hat{f}_{s_*,h_*};\mathscr{F}] \leqslant C_3 A^{\frac{2\gamma+1}{2\beta+2\gamma+1}} \left(M x_0^{2\gamma-1+2s_*} n^{-1}\right)^{\frac{\beta}{2\beta+2\gamma+1}} .$$

- Note that $s_* < 0$. If $\eta \sim U(0,1)$ and $\alpha = 1$ then $s_* = -1$, and the bound on the risk is proportional to $x_0^{-\beta/(2\beta+3)}$.

Part 4: Estimation at zero

Estimation at zero

- Functional class: $\mathscr{H}_r(A,\beta)$ is the local Hölder class as $\mathscr{H}_{x_0,r}(A,\beta)$ but now on the interval $[0,r]$. Define also

$$\bar{\mathscr{H}}_r = \bar{\mathscr{H}}_r(A,\beta,M) := \mathscr{H}_r(A,\beta) \cap \{f : \|f\|_\infty \leqslant M\}.$$

- The problem is trivial if

$$I_g := \int_0^\infty \frac{g(x)}{x}\mathrm{d}x < \infty \quad \Leftrightarrow \quad 0 \in \Omega_g.$$

Here f_Y is bounded at the origin, and $f_Y(0) = f_X(0)I_g$. Then $f_X(0)$ can estimated with standard rate $n^{-\beta/(2\beta+1)}$.

- The condition $0 \in \Omega_g$ is restrictive: the only interesting case

$$I_g := \int_0^\infty \frac{g(x)}{x}\mathrm{d}x = \infty \quad \Leftrightarrow \quad 0 \notin \Omega_g.$$

Kernel construction and estimator

- For $s \geqslant 0$ we construct a function $K_s : \mathbb{R} \to \mathbb{R}$ s.t.

(i) $\mathrm{supp}(K_s) = [0, \infty)$,

$$\int_0^\infty K_s(t)\mathrm{d}t = 1, \quad \int_0^\infty t^k K_s(t)\mathrm{d}t = 0, \quad k = 1, \dots, m;$$

(ii) $|\widetilde{K}_s(s + i\omega)| \leqslant c_1(s,m)\exp\{-\omega^2/2\}$, $\forall \omega$.

- The kernels K and L: $K_{s,h}(x) := (1/h)K_s(x/h)$, and

$$L_{s,h}(y) := \frac{1}{2\pi i}\int_{s-i\infty}^{s+i\infty} \frac{\widetilde{K}_{s,h}(z)}{\widetilde{g}(1-z)}\mathrm{d}z$$
$$= \frac{1}{2\pi h^{1-s} y^s}\int_{-\infty}^{\infty}\left(\frac{h}{y}\right)^{i\omega}\frac{\widetilde{K}_s(s+i\omega)}{\widetilde{g}(1-s-i\omega)}\mathrm{d}\omega.$$

- The estimator:

$$\hat{f}_h(0) = \frac{1}{n}\sum_{i=1}^n L_{s,h}(Y_i).$$

Assumptions on g

- Assumption [G2]: for some $p \in [0,1)$, $q \geqslant 0$ and $\delta \in (0,1)$

$$c_0 x^{-p}[\ln(1/x)]^q \leqslant g(x) \leqslant C_0 x^{-p}[\ln(1/x)]^q, \quad x \in [0,\delta).$$

- [G2] holds for $U(0,1)$ and $\exp(1)$ with $p = 0$ and $q = 0$.
- Assumption [G3]: for $1-\sigma \in \Omega_g$ one has

$$\max_{j=1,2} \int_{-\infty}^{\infty} \left(\frac{e^{-\omega^2/2}}{|\widetilde{g}(1-\sigma+i\omega)|} \right)^j \mathrm{d}\omega \leqslant C_1 < \infty$$

$$\int_{-\infty}^{\infty} \left| \frac{\mathrm{d}^l}{\mathrm{d}\omega^l} \left(\frac{e^{-\omega^2/2}}{|\widetilde{g}(1-\sigma+i\omega)|} \right) \right|^2 \mathrm{d}\omega \leqslant C_2 < \infty,$$

  where $l := [(q+1)/2]$, q is given in [G2].
- [G2] specifies behavior of g at zero; [G3] is a weak condition...

Bounds on the risk

- Theorem 4: (Upper bound). *Suppose [G2] and [G3] with* $\sigma = s_* := \frac{1}{2}(1-p)$. *Let* $\hat{f}_{s_*,h_*}(0)$ *be associated with* $s = s_*$ *and*

$$h = h_* := \left[\frac{M(\ln n)^{q+\varkappa}}{A^2 n}\right]^{\frac{1}{2\beta+1+p}}, \quad \varkappa = \begin{cases} 0, & p \in (0,1), \\ 1, & p = 0. \end{cases}$$

Then for large n

$$\mathrm{Risk}_0[\hat{f}_{s_*,h_*}; \bar{\mathscr{H}}_r] \leqslant C_3 M^{\frac{\beta}{2\beta+p+1}} \overbrace{A^{\frac{1+p}{2\beta+1+p}} \left[(\ln n)^{q+\varkappa} n^{-1}\right]^{\frac{\beta}{2\beta+p+1}}}^{\varphi_n} .$$

- Theorem 5: (Lower bound). Suppose [G2] and $\int_0^\infty x g^2(x)(\ln x)^2 \mathrm{d}x \leqslant C_4 < \infty$. Then

$$\liminf_{n\to\infty} \left\{\varphi_n^{-1} \mathrm{Risk}_0^*[\bar{\mathscr{H}}_r]\right\} \geqslant C_5.$$

Remarks

- The estimator $\hat{f}_{s_*,h_*}$ is optimal in order.
- If $p = 0$ then the rate of convergence is only by a $\ln n$–factor worse than the standard rate. This is the case, e.g., for $\eta \sim U(0,1)$ and $\eta \sim \exp(1)$ where the optimal rate is $(\ln n/n)^{\beta/(2\beta+1)}$.
- In contrast to estimating at $x_0 \neq 0$, the rate of convergence is completely determined by local behavior of g near the origin.

1. Concluding remarks

- Two completely different settings:
  - (A): estimating f_X at a point separated from zero;
  - (B): estimating f_X at zero.
- (A): the accuracy depends on x_0 and is determined by
  - the local smoothness of f_X in a vicinity of x_0;
  - the rate at which $|\widetilde{g}(z)|$ decreases as $|\mathrm{Im}(z)| \to \infty$ along a line which is a global characteristic of g.
- (B): the accuracy determined by
  - local smoothness of f_X in a vicinity of 0;
  - local behavior of g in a vicinity of 0.

2. Concluding remarks

- Implementation

  In many cases kernel $L_{s,h}$ can be computed explicitly, and the estimators can be easily implemented.

- “Super-smooth” densities

  The case of exponential decaying $|\tilde{g}(\sigma + i\omega)|$ for $x_0 > 0$ can be treated similarly. Here the rates are slow, logarithmic in n, and the improvements with respect to dependence on x_0 are of the second order smallness.

- Adaptation with respect to unknown smoothness seems to be straightforward, though technical...