Department of Statistical and Actuarial Sciences,
The University of Western Ontario,
London, Canada

Advances in Semi-Nonparametric Density Estimation and Shrinkage Regression

Hossein Zareamoghaddam

March 16, 2018

Outline

- Moment-based density estimation techniques are ideally suited for modeling massive data sets. Indeed, once the moments have been evaluated, which is easily achieved even for extremely large data sets, the determination of the estimated density function does not depend on the sample size.
- Once a new set of observations, $x_{n_1+1}, \ldots, x_n$, becomes available in addition to an initial dataset, $x_1, \ldots, x_{n_1}$, there is no need to make use of each of the n_1 original data points since the h^{th} updated moment will then be $\{n_1 m_h + \sum_{i=n_1+1}^{n} x_i^h\}/n$ where m_h denotes the h^{th} sample moment evaluated from the initial data set.

Theorem

The observed values in a sample of size n is uniquely determined by the first n sample moments.

Proof. Let $S = \{x_1, x_2, \ldots, x_n\}$, $M = \{m_1, m_2, \ldots, m_n\}$ and $m_h = \sum_{i=1}^{n} x_i^h/n$. According to the fundamental theorem of algebra, $p(z) = a_0 + a_1 z + \cdots + a_{n-1} z^{n-1} + z^n$ is uniquely defined by its coefficients a_i's and it is also uniquely specified by its n roots x_i's for $i = 1, \ldots, n$. Moreover, given S, the coefficients of $p(x)$ can be expressed in terms of the sequence of moments M via the Newton-Girard identity. Accordingly, a given polynomial of degree n, say $p(x)$, can be represented as follows:

$$\prod_{i=1}^{n}(x - x_i) = \sum_{k=0}^{n}(-1)^{n-k} e_{n-k}\, x^k, \tag{1}$$

where $e_0 = 1$ and

$$e_\ell = \frac{n}{\ell}\sum_{j=1}^{\ell}(-1)^{j-1} e_{\ell-j}\, m_j, \quad \ell = 1, \ldots, n.$$

Thus, given the first n sample moments associated with S, a sample of size n, one can determine the right hand side of (1) whose roots are precisely $\{x_1, x_2, \ldots, x_n\}$. This establishes that S is uniquely specified by M.

Let a density function $f(x)$ be approximated by

$$f_m(x) = f_0(x)\, p_m(x)$$

where

- $f_0(x)$ is an initial (base) density function
- $p_m(x) = \sum_{i=0}^{m} a_i x^i$ is a polynomial adjustment of degree m on the interval (α, β)

  Alternatively, $p_m(x)$ can be expressed as $\sum_{j=0}^{m} \eta_j\, \varphi_j(x)$ where $\varphi_k(x) = \sum_{\ell=0}^{k} \delta_{k,\ell}\, x^\ell$, $k = 0, \ldots, m$, denotes an orthogonal polynomial of order k.

The coefficients a_i's or η_j's are determined by equating the first m moments of $f(x)$ to those of $f_m(x)$.

Definition

$\{\varphi_i(x)\}_{i=0}^{\infty}$ is a sequence of univariate orthogonal polynomials w.r.t. $\omega(x)$ iff $\int_R \omega(x)\varphi_i(x)\varphi_j(x)\mathrm{d}x = \theta_i\,\delta_{ij}$

For example:

- Modified Hermite polynomials:

$$H_n(x) = (-1)^n e^{\frac{x^2}{2}} \frac{\mathrm{d}^n}{\mathrm{d}x^n} e^{-\frac{x^2}{2}}, \; -\infty < x < \infty, \;\; n = 0, 1, 2, \ldots$$

  Recurrence relation:

$$H_{n+1}(x) = xH_n(x) - nH_{n-1}(x),$$

$$H_0(x) = 1, \; H_1(x) = x, \; H_2(x) = x^2 - 1, \; H_3(x) = x^3 - 3x$$

  Orthogonality: $\int_{-\infty}^{\infty} e^{-x^2/2} H_i(x)\, H_j(x)\, \mathrm{d}x = \sqrt{2\pi} j!\, \delta_{ij}.$

- Legendre, Laguerre and Jacobi polynomials

Univariate Density Approximation

Equivalence between making use of standard and orthogonal polynomial adjustments

Let $\tilde{f}_m(x) = f_0(x)\sum_{k=0}^{m} c_k\, x^k$ and $\hat{f}_m(x) = f_0(x)\sum_{j=0}^{m} \eta_j \varphi_j(x)$ where the $\varphi_j(x)$'s are the orthogonal polynomials associated with the base density function $f_0(x)$. Then,

$$
\begin{aligned}
\hat{f}_m(x) &= f_0(x)\sum_{j=0}^{m} \eta_j\, \varphi_j(x) \\
&= f_0(x)\sum_{j=0}^{m} \eta_j \left(\sum_{k=0}^{j} \alpha_k\, x^k\right) \\
&= f_0(x)\sum_{k=0}^{m} \left(\sum_{j=k}^{m} \eta_j\, \alpha_k\right) x^k \qquad \text{as } \sum_{j=0}^{m}\sum_{k=0}^{j} \equiv \sum_{k=0}^{m}\sum_{j=k}^{m} \\
&\equiv f_0(x)\sum_{k=0}^{m} c_k\, x^k \qquad \text{with } c_k = \sum_{j=k}^{m} \eta_j\, \alpha_k \\
&\equiv \tilde{f}_m(x)\,.
\end{aligned}
$$

- The Christoffel–Darboux formula:

$$\sum_{k=0}^{m} \frac{\varphi_k(x)\,\varphi_k(y)}{\theta_k} = \frac{\delta_{m,m}}{\delta_{m+1,m+1}} \frac{\varphi_{m+1}(x)\,\varphi_m(y) - \varphi_m(x)\,\varphi_{m+1}(y)}{\theta_m\,(x-y)},$$

- Kernel functions

$$\mathcal{K}_m(x, x_i) = w(x) \sum_{j=0}^{m} \frac{1}{\theta_j}\,\varphi_j(x_i)\,\varphi_j(x)$$

- Kernel representation of the density estimates

$$\hat{f}_m(x) = \frac{1}{n} \sum_{i=1}^{n} \mathcal{K}_m(x, x_i),$$

Example

Let $X_1 \sim \mathcal{N}(-4, 2)$, $X_2 \sim \mathcal{N}(2, 3)$ and $f(x) = 1/2\left[f_{X_1}(x) + f_{X_2}(x)\right]$.

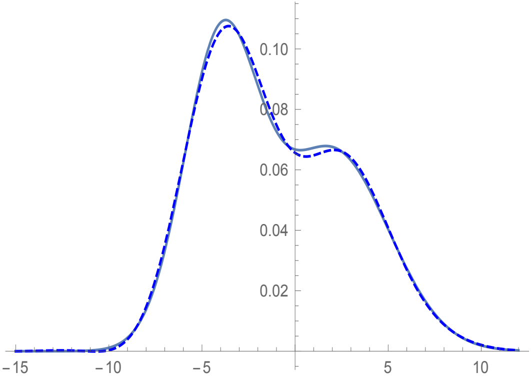


Figure: Exact (solid line) and approximated $\hat{f}_{15}(x)$ (dashed line) density functions

Univariate Density Approximation

The differentiated log density approximation (DLDA)

The differentiated log density approximation (DLDA) methodology is a generalization of the Pearson's frequency curves whereby the derivative of logarithm of a density function $f(x)$ whose support is (α, β) is assumed to be a rational function, that is,

$$\frac{\mathrm{d}}{\mathrm{d}x}\log(f(x)) = \frac{f'(x)}{f(x)} = r(x),$$

where

$$r(x) = \frac{\sum_{i=0}^{\nu} a_i\, x^i}{\sum_{j=0}^{\delta} c_j\, x^j} = \frac{N_\nu(x)}{D_\delta(x)},$$

the coefficients a_i's and c_j's being determined by solving a linear system of equations. Then,

$$f_{\nu,\delta}(x) = \kappa\, e^{\int_\alpha^x r(y)\,\mathrm{d}y}, \qquad \text{where } \kappa \text{ is the normalizing constant.} \tag{2}$$

One has

$$f_{\nu,\delta}(x)\sum_{i=0}^{\nu} a_i\, x^i = f'_{\nu,\delta}(x)\sum_{j=0}^{\delta} c_j\, x^j .$$

On multiplying both sides of this equation by x^h and integrating over the support $(\alpha,\ \beta)$, one has

$$\int_\alpha^\beta f_{\nu,\delta}(x)\sum_{i=0}^{\nu} a_i\, x^{i+h}\mathrm{d}x = \int_\alpha^\beta f'_{\nu,\delta}(x)\sum_{j=0}^{\delta} c_j\, x^{j+h}\mathrm{d}x,\ \ h=0,\ 1,\ldots,\ \nu+\delta;$$

then, on interchanging the sum and the integral on each side of the above equation and integrating the left-hand side by parts, one obtains $\nu+\delta+1$ linear equations of the following form:

$$\sum_{i=0}^{\nu} a_i\,\mu(i+h) = \sum_{j=0}^{\delta} c_j\left(f_{\nu,\delta}(\beta)\,\beta^{j+h} - f_{\nu,\delta}(\alpha)\,\alpha^{j+h}\right) - \sum_{j=0}^{\delta} c_j(j+h)\,\mu(j+h-1),\ \ h=0,\ 1,\ldots,\ \nu+\delta.$$

Remark

The degree of the denominator $D_\delta(x)$, that is δ, corresponds to the number of times the density function $f_X(x)$ intersects the abscissa plus the number of points at which this density function is not differentiable. Since a simple polynomial cannot adequately account for abrupt changes in the slope of a function, one needs to include the number of points of non-differentiability in the set of roots of $D_\delta(x)$. As well, increasing the degree of numerator $N_\nu(x)$ generally leads to more accurate approximations.

Lemma

Letting $x > \alpha$, $i \geq 2$ and A, B, a and b be constants. One has

$$\int \frac{A\,y + B}{y^2 + a\,y + b}\mathrm{d}y = \frac{A}{2}\log(y^2 + a\,y + b) + (B - \frac{aA}{2})\int \frac{\mathrm{d}y}{y^2 + a\,y + b}.$$

We consider the case that the equation $y^2 + a\,y + b$ has two real roots of λ_1, λ_2 $(\lambda_1 \neq \lambda_2)$. By setting $m^2 = b - \frac{a^2}{4}$, one has $\lambda_1 = -a/2 + m$ and $\lambda_2 = -a/2 - m$, and

$$\int \frac{A\,y + B}{y^2 + a\,y + b}\mathrm{d}y = \log\left(|y - \lambda_1|^{E_1}|y - \lambda_2|^{E_2}\right)$$

where $E_1 = \frac{A\,\lambda_1 + B}{\lambda_1 - \lambda_2}$ and $E_2 = \frac{A\,\lambda_2 + B}{\lambda_2 - \lambda_1}$.

Univariate Density Approximation

Explicit representation of differentiated log density approximants (DLDA)

Now, letting $\lambda_1, \lambda_2, \ldots, \lambda_\delta$ denote the distinct real roots of the denominator $D_\delta(y)$ of $r(x)$ and letting $Q_{\nu-\delta}(y) = \sum_{i=0}^{\nu-\delta} q_i\, y^i$ and $R(y) = \sum_{i=0}^{\gamma} d_i\, y^i$ as the quotient and the remainder of the rational function

$$r(y) = \frac{N_\nu(y)}{D_\delta(y)} = \frac{\sum_{i=0}^{\nu} a_i\, x^i}{\sum_{j=0}^{\delta} c_j\, x^j} = Q_{\nu-\delta}(y) + \frac{R(y)}{D_\delta(y)}.$$

The representation of the approximate density function is

$$\begin{aligned} f_{\nu,\delta}(x) &= \kappa\, e^{\int_\alpha^x r(y)\,\mathrm{d}y} \\ &= \kappa e^{\sum_{i=0}^{\nu-\delta} q_i\, x^{i+1}/(i+1)} \prod_{k=1}^{\delta} |x-\lambda_k|^{\tau_k}, \quad x \in (\alpha, \beta), \end{aligned}$$

where κ is a normalizing constant and $\tau_k = \dfrac{\sum\limits_{i=0}^{\gamma} d_i\, \lambda_k^i}{\prod\limits_{j\neq k} (\lambda_k - \lambda_j)}$.

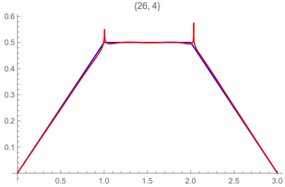
When the DLDA is a differentiable function as is the case for instance when the denominator has two real roots $\lambda_1 < \lambda_2$ which are the end points of the distribution and letting $\theta_j = \frac{q_{j-3}}{j-2}$, $j = 3, \ldots, \nu + 2$, the density estimate can be expressed as

$$f_{\nu,2}(x) = \kappa'(x - \lambda_1)^{\theta_1}(\lambda_2 - x)^{\theta_2}\, e^{\sum_{i=0}^{\nu-1} \theta_{i+3} x^{i+1}}$$
$$= e^{\sum_{j=1}^{\nu+1} p_j(\theta_1, \ldots, \theta_{j+1}) K_j(x) + q(\theta_1, \ldots, \theta_{j+1})} \tag{3}$$

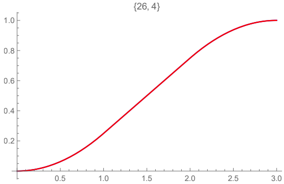
in the notation of Hogg and Craig (1978, p. 366), where $p_j(\theta_1, \ldots, \theta_{j+1}) = \theta_j$ for $j = 1, \ldots, \nu + 1$, $K_1(x) = \log(x - \lambda_1)$, $K_2(x) = \log(\lambda_2 - x)$ and $K_j(x) = x^{j-2}$, $j = 3, \ldots, \nu + 2$. Accordingly, on the basis of a sample of size n,

$$\sum_{i=1}^{n} \log(x_i - \lambda_1),\ \sum_{i=1}^{n} \log(\lambda_2 - x_i),\ \sum_{i=1}^{n} x_i, \ldots, \sum_{i=1}^{n} x_i^{\nu}$$

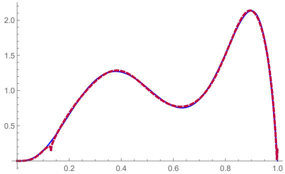
are joint sufficient statistics for $\theta_1, \ldots, \theta_{\nu+2}$, the parameters of the resulting density function.



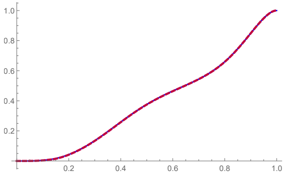
$f(x)$ and $f_{26,4}(x)$



$F(x)$ and $F_{26,4}(x)$



$f(x)$ and $f_{8,2}(x)$



$F(x)$ and $F_{8,2}(x)$

Differentiated log density approximants (DLDA)

Numerical results from 100 Monte Carlo simulations

Table: Average ISE's (SD's in parentheses) for different distributions estimated by applying the DLDA and kernel density estimation (KDE) techniques.

		$n = 500$	$n = 2000$
Beta	DLDA	0.0165 (0.0247)	0.0069 (0.0034)
	KDE	0.0270 (0.0142)	0.0101 (0.0041)
Gamma	DLDA	0.0001 (0.0001)	0.0000 (0.0000)
	KDE	0.0002 (0.0001)	0.0002 (0.0000)
Exponential	DLDA	0.0449 (0.0366)	0.0381 (0.0157)
	KDE	0.0641 (0.0148)	0.0457 (0.0062)
Student t	DLDA	0.0013 (0.0013)	0.0006 (0.0003)
	KDE	0.0020 (0.0011)	0.0006 (0.0003)
Normal	DLDA	0.0006 (0.0004)	0.0002 (0.0001)
	KDE	0.0018 (0.0009)	0.0007 (0.0003)

where $$ISE(\hat{f}) = \int \left(\hat{f}(x) - f(x)\right)^2 \mathrm{d}x.$$

Now, let a target bivariate density function $f(x, y)$ be approximated by

$$f_{p,q}(x, y) = f_0(x, y)\,\lambda_{p,q}(x, y)$$

where $f_0(x, y)$ is a suitable initial density approximant and $\lambda_{p,q}(x, y) = \sum_{i=0}^{p}\sum_{j=0}^{q} c_{i,j}x^i y^j$ is a polynomial adjustment.

The coefficients $c_{i,j}$ of the adjustment are determined by equating the joint moments of $f(x, y)$ to those associated with $f_{p,q}(x, y)$.

Bivariate Density Approximation

Let $\mu(k,\ell) = \int\int x^k\, y^\ell\, f(x,\,y)\,\mathrm{d}x\,\mathrm{d}y$ and $m(k,\ell) = \int\int x^k\, y^\ell\, f_0(x,\,y)\,\mathrm{d}x\,\mathrm{d}y$ be the $(k,\ell)^{\mathrm{th}}$ joint moments associated with $f(x,\,y)$ and $f_0(x,\,y)$ respectively; then

$$\begin{aligned}
\mu(k,\ell) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^k y^\ell f_{p,q}(x,\,y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^k y^\ell\, f_0(x,\,y)\,\lambda_{p,q}(x,\,y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^k y^\ell\, f_0(x,\,y)\sum_{i=0}^{p}\sum_{j=0}^{q} c_{i,j}\, x^i y^j\,\mathrm{d}x\,\mathrm{d}y \\
&= \sum_{i=0}^{p}\sum_{j=0}^{q}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} c_{i,j}\, x^{k+i} y^{\ell+j}\, f_0(x,\,y)\,\mathrm{d}x\,\mathrm{d}y, \\
&= \sum_{i=0}^{p}\sum_{j=0}^{q} c_{i,j}\, m(k+i,\,\ell+j),
\end{aligned}$$

for $k = 0,\,\ldots,\,p$ and $\ell = 0,\,\ldots,\,q$.

Bivariate Density Approximation

This yields the following $(p+1)(q+1)$ linear equations:

$$\mu(k,\ell) = \sum_{i=0}^{p}\sum_{j=0}^{q} c_{i,j}\, m(k+i,\, \ell+j),\;\; k = 0,\, 1, \ldots,\, p,\;\; \ell = 0,\, 1, \ldots,\, q.$$

The $c_{i,j}$'s can be obtained by solving the linear system $M\mathbf{c} = \boldsymbol{\mu}$ where $\mathbf{c}$ and $\boldsymbol{\mu}$ are vectors of dimensions $(p+1)(q+1)$ whose $(i(q+1)+(j+1))^{\text{th}}$ component, $c_{i,j}$ and $\mu(i,j)$, appear in the same order for $i = 0,\, 1,\, \ldots,\, p$ and $j = 0,\, 1,\, \ldots,\, q$.

Let $\mathbf{z} = (x, y)'$ and $\Lambda_{2\times 2}$ be a positive definite matrix. The bivariate orthogonal Hermite polynomials of orders r_1 and r_2 $(r = r_1 + r_2)$ associated with a $\mathcal{N}_2(\mathbf{0}, A)$ distribution where $A = \Lambda^{-1}$ can be generated as follows:

$$H_{r_1, r_2}(\mathbf{z}, \Lambda) \equiv (-1)^r exp(\frac{1}{2}\mathbf{z}' A \mathbf{z}) \frac{\partial^r}{\partial x^{r_1} \partial y^{r_2}} exp(-\frac{1}{2}\mathbf{z}' A \mathbf{z})$$

For example, when $\Lambda = \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}$

$$H_{1,1}(\mathbf{z}, \Lambda) = \frac{1}{5} - \frac{3x^2}{25} + \frac{7xy}{25} - \frac{2y^2}{25}$$

$$H_{1,2}(\mathbf{z}, \Lambda) = -\frac{8x}{25} + \frac{3x^3}{125} + \frac{6y}{25} - \frac{13x^2y}{125} + \frac{16xy^2}{125} - \frac{4y^3}{125}$$

$$H_{2,2}(\mathbf{z}, \Lambda) = \frac{8}{25} - \frac{33x^2}{125} + \frac{9x^4}{625} + \frac{52xy}{125} - \frac{42x^3y}{625} - \frac{22y^2}{125} + \frac{61x^2y^2}{625} - \frac{28xy^3}{625} + \frac{4y^4}{625}$$

- Let $\{H_{i,j}(\mathbf{z}) = \sum_{\ell=0}^{j} \sum_{k=0}^{i} \alpha_{ijk\ell} x^k y^\ell\}$ be a bivariate Hermite polynomial; then $H^*_{u,v}(\mathbf{z}) = \sum_{n=0}^{v} \sum_{m=0}^{u} \alpha^*_{uvnm} x^m y^n$ which is called the dual BHP of orders u and v with respect to the bivariate function $\Psi(\mathbf{z})$ defined on $\mathbb{R}^2$, satisfies the equation,

$$\int\int_{\mathbb{R}^2} \Psi(\mathbf{z}) H_{i,j}(\mathbf{z}) H^*_{u,v}(\mathbf{z})\, \mathrm{d}\mathbf{z} = \begin{cases} \theta_{u,v} & \text{if } (i,j) = (u,v) \\ 0 & \text{otherwise} \end{cases}$$

  where $\Psi(\mathbf{z})$ is the PDF of a $\mathcal{N}_2(\mathbf{0},\, A)$ random vector.
- The dual bivariate Hermite polynomials can be obtained as

$$H^*_{u,v}(x,\, y) = E[(x + iX)^u (y + iY)^v].$$

Let $f_0(x, y)$ be a normal base density associated with Hermite orthogonal polynomials, then

$$f_{p,q}(x, y) = f_0(x, y) \sum_{i=0}^{p} \sum_{j=0}^{q} \eta_{i,j} H_{i,j}(x, y).$$

The coefficient η_{ij}'s are obtained by equating $\int\int H^*_{u,v}(x, y) f_{p,q}(x, y)\,\mathrm{d}x\,\mathrm{d}y$ to $\int\int H^*_{u,v}(x, y) f(x, y)\,\mathrm{d}x\,\mathrm{d}y$ for $u = 0, 1, \ldots, p$ and $v = 0, 1, \ldots, q$ which does not require solving any linear system of equations, where $H^*_{u,v}(x, y)$ is the dual Hermite polynomial of $H_{u,v}(x, y)$.

Theorem

For a given base density function, the approximated density functions obtained by utilizing standard polynomial adjustments are mathematically equivalent to those obtained by making use of linear combinations of orthogonal polynomials as adjustments.

Bivariate Density Approximation
Equivalence between standard and orthogonal polynomial adjustments

Noting that $\sum_{i=0}^{p}\sum_{k=0}^{i} \equiv \sum_{k=0}^{p}\sum_{i=k}^{p}$ and $\sum_{j=0}^{q}\sum_{\ell=0}^{j} \equiv \sum_{\ell=0}^{q}\sum_{j=\ell}^{q}$, and letting $\tilde{f}_{p,q}(x,\,y) = f_0(x,\,y)\sum_{k=0}^{p}\sum_{\ell=0}^{q} c_{k,\ell}\,x^k y^\ell$ and $\hat{f}_{p,q}(x,\,y)$ $= f_0(x,\,y)\sum_{i=0}^{p}\sum_{j=0}^{q}\eta_{i,j}H_{i,j}(x,\,y)$, one has

$$
\begin{aligned}
\hat{f}_{p,q}(x,\,y) &= f_0(x,\,y)\sum_{i=0}^{p}\sum_{j=0}^{q}\eta_{i,j}H_{i,j}(x,\,y)\\
&= f_0(x,\,y)\sum_{i=0}^{p}\sum_{j=0}^{q}\eta_{i,j}\sum_{k=0}^{i}\sum_{\ell=0}^{j}\alpha_{ijk\ell}\,x^k y^\ell,\\
&= f_0(x,\,y)\sum_{i=0}^{p}\sum_{k=0}^{i}\sum_{j=0}^{q}\sum_{\ell=0}^{j}\eta_{i,j}\,\alpha_{ijk\ell}\,x^k y^\ell,\\
&= f_0(x,\,y)\sum_{k=0}^{p}\sum_{\ell=0}^{q}\left\{\sum_{i=k}^{p}\sum_{j=\ell}^{q}\eta_{k,\,\ell}\,\alpha_{ijk\ell}\right\}x^k y^\ell\\
&\equiv f_0(x,\,y)\sum_{k=0}^{p}\sum_{\ell=0}^{q}c_{k,\ell}\,x^k y^\ell\\
&= \tilde{f}_{p,q}(x,\,y).
\end{aligned}
$$

Example

A mixture of bivariate normal densities. Let

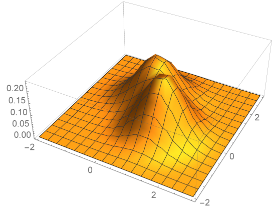
$$\mathbf{Z}_1 \sim \mathcal{N}_2\left(\begin{pmatrix} 1.1 \\ -0.1 \end{pmatrix}, \begin{pmatrix} 0.33 & 0.03 \\ 0.03 & 0.33 \end{pmatrix}\right), \quad \mathbf{Z}_2 \sim \mathcal{N}_2\left(\begin{pmatrix} 0.2 \\ 1.2 \end{pmatrix}, \begin{pmatrix} 0.4 & 0.04 \\ 0.04 & 0.4 \end{pmatrix}\right)$$

and $f(x, y) = \frac{1}{2}(f_{\mathbf{Z}_1}(x, y) + f_{\mathbf{Z}_2}(x, y))$. The base density $f_0(x, y)$ is assumed to have the following distribution:

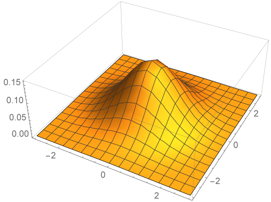
$$\mathcal{N}_2\left(\begin{pmatrix} 0.65 \\ 0.55 \end{pmatrix}, \begin{pmatrix} 0.5675 & -0.2575 \\ -0.2575 & 0.7875 \end{pmatrix}\right),$$

Bivariate Density Approximation

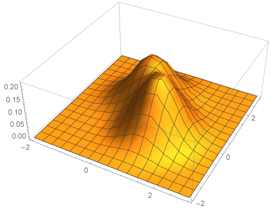
Equivalence between standard and orthogonal polynomial adjustments



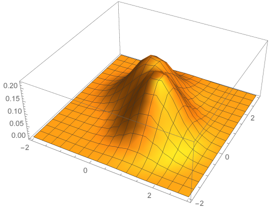
Exact density



Base density



$\tilde{f}_{6,6}(x,\ y)$ (S)



$\hat{f}_{6,6}(x,\ y)$ (O)

Bivariate Density Estimation

Degree selection criteria

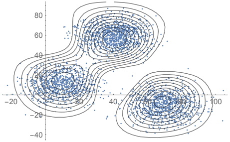
- The $(i,j)^{\mathrm{th}}$ joint sample moments: $\hat{\mu}_{i,j} = \frac{1}{N}\sum_{k=1}^{N} x_k^i\, y_k^j$
- ECDF$(x,\ y)$ denotes the empirical CDF associated with the dataset
- $F_{p,q}(x,\ y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{p,q}(x,y)\mathrm{d}y\mathrm{d}x$ is the estimated CDF

Degree selection criteria: *degree corresponding to a local minimum of SSD values*

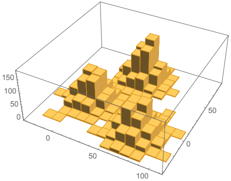
- Univariate: $SSD(p) = \sum_{i=1}^{N}(\mathrm{ECDF}(x_i) - F_p(x_i))^2$
- Bivariate: $SSD(p,q) = \sum_{i=1}^{N}(\mathrm{ECDF}(x_i,\ y_i) - F_{p,q}(x_i,\ y_i))^2$

Bivariate Density Estimation

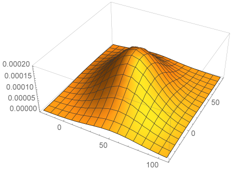
"*xclara*" data set with $N = 3000$ observations (Struyf *et al.*, 1996)



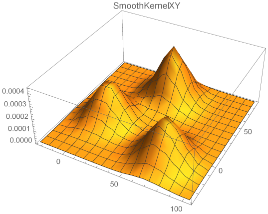
Scatter plot



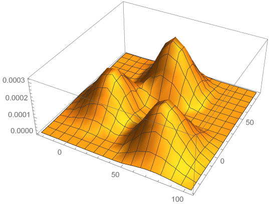
3D Histogram



Base density



KDE

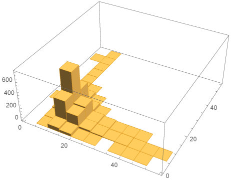


$f_{7,11}(x, y)$

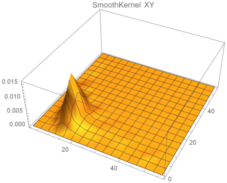
Trivariate Density Estimation

"*CommViolPredUnnormalizedData*" with $N = 2315$ obs. (Redmond, 1990)

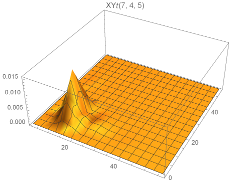
First, **the trivariate density estimate** $f_{7,4,5}(x, y, z)$ is determined. The resulting marginal density functions $f_{7,4,5}(x, y) = \int f_{7,4,5}(x, y, z)\mathrm{d}z$, $f_{7,4,5}(y, z)$ and $f_{7,4,5}(x, z)$ and the corresponding kernel density estimates (KDE) are plotted.



3D Histogram



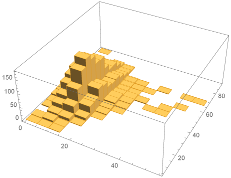
KDE of (x, y)



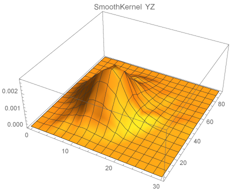
$f_{7,4,5}(x, y)$

Trivariate Density Estimation

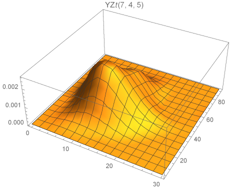
"*CommViolPredUnnormalizedData*" with $N = 2315$ obs. (Redmond, 1990)



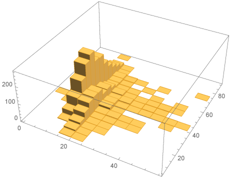
3D Histogram



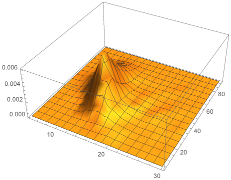
KDE of (y, z)



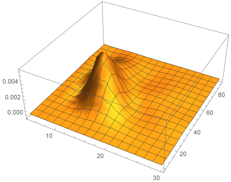
$f_{7,4,5}(y, z)$



3D Histogram



KDE of (x, z)

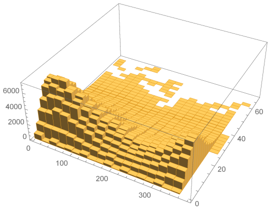


$f_{7,4,5}(x, z)$

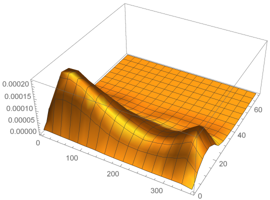
Bivariate Density Estimation

"*Covertype*" data with $N = 581,012$ obs. (Blackard and Denis, 2000)

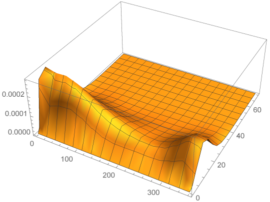
This methodology is applicable to **big data** as it is based on a certain number of joint sample moments, as opposed to the individual observations themselves. Alternatively, it provides a functional representation of the density estimates.



3D Histogram



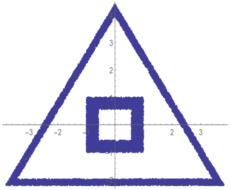
KDE



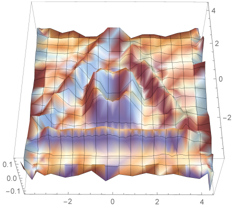
$f_{7,3}(x, y)$

Bivariate Density Estimation

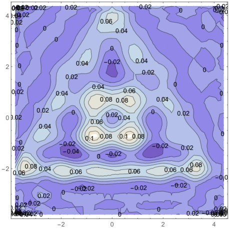
Making use of Legendre polynomials



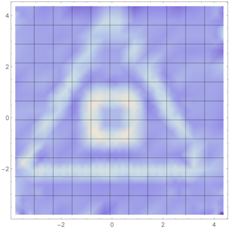
Scatter plot



$f_{40,40}(x, y)$



Contour plot



Density projection

Example

Consider the following nonparametric regression model

$$y_i = m(x_i) + \epsilon_i, \; i = 1, \dots, n \tag{4}$$

where

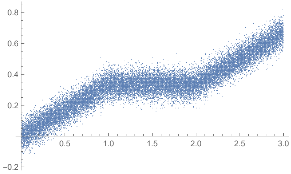
$$m(x) = \begin{cases} \frac{x}{3}, & \text{if } 0 \leq x < 1 \\ \frac{1}{3}, & \text{if } 1 \leq x < 2 \\ \frac{x-1}{3}, & \text{if } 2 \leq x \leq 3 \end{cases}$$

and ϵ_i's follows normal distribution $\mathcal{N}(0, 0.05)$ with $N = 20000$ points generated randomly from (4).

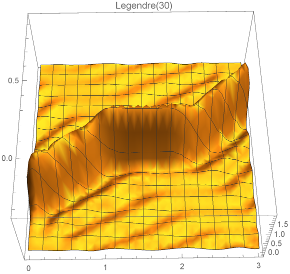
The resulting density estimate was obtained by making use of Legendre polynomials of orders at most 30.

Bivariate Density Estimation

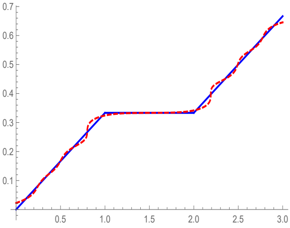
Application to nonparametric regression models



Scatter plot



$f_{30,30}(x, y)$



$\hat{m}(x)$ and $m(x)$

Letting $(w_i,\ z_i)$, $i = 1, \ldots, n$, be a dataset with sample mean $(\bar{w},\ \bar{z})$ and sample covariance matrix V. The standardized data is obtained as follows:

$$\begin{pmatrix} x_i \\ y_i \end{pmatrix} = \hat{V}^{-1/2} \begin{pmatrix} w_i - \bar{w} \\ z_i - \bar{z} \end{pmatrix}.$$

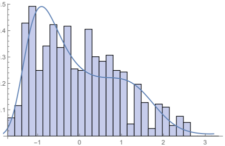
The x_i's and the y_i's are then uncorrelated, and we let

$$f_{\nu_1,\delta_1,\nu_2,\delta_2,p}(x,\ y) = f_{\nu_1,\delta_1}(x) f_{\nu_2,\delta_2}(y)\, \lambda_p(x, y),$$

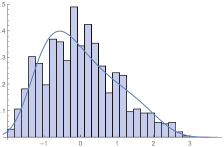
where $f_{\nu_1,\delta_1}(x)$ and $f_{\nu_2,\delta_2}(y)$ denote the estimated marginal density functions for the standardized vector $(X,\ Y)'$ and $\lambda_p(x, y)$ is a bivariate polynomial adjustment of order p in each variable.

Bivariate differentiated log density estimation

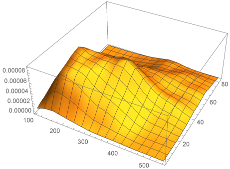
“concrete" data set with $N = 1030$ observations (Yeh, 1998)



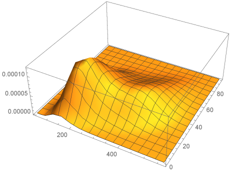
$f_4(x)$ and histogram



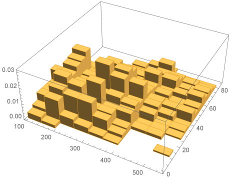
$f_3(y)$ and histogram



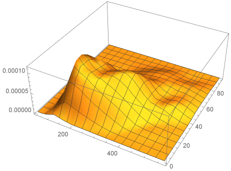
Bivariate KDE



Base PDF $\phi(w, z)$



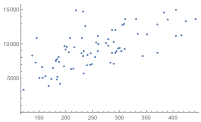
Bivariate histogram



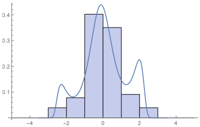
$f_{4,3,6}(w, z)$

Bivariate differentiated log density estimation

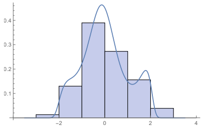
“Flood" data set with $N = 77$ observations collected in the Madawaska Basin, Quebec, 1990-1995



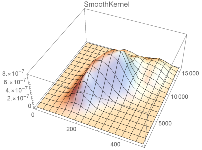
Scatterplot



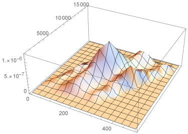
$f_{5,2}(x)$ and histogram



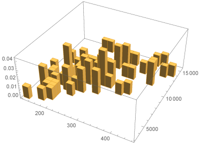
$f_{5,2}(y)$ and histogram



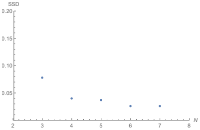
Bivariate KDE



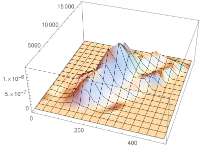
Base PDF $\phi(w, z)$



Bivariate histogram



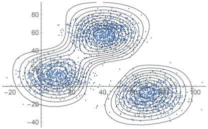
$SSD(p)$ vs p



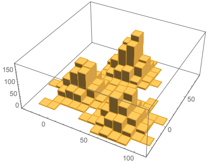
$g_{5,2,5,2,6}(w, z)$

Bivariate Density Estimation

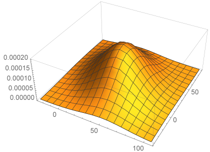
"*xclara*" data set with $N = 3000$ observations (Struyf *et al.*, 1996)



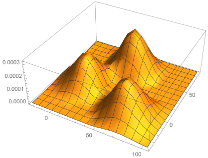
Scatter plot



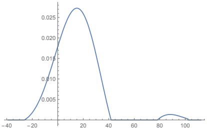
3D Histogram



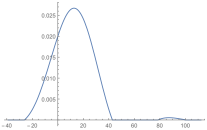
Base density



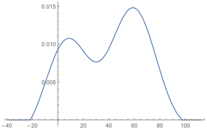
$f_{7,11}(x, y)$



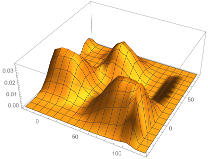
$f_{7,11}(y|x = -15)$



$f_{7,11}(y|x = 4)$



$f_{7,11}(y|x = 25)$



$f_{7,11}(y|x)$

A semi-nonparametric regression model
The model

Consider the n semi-nonparametric regression equations

$$y_{it} = \alpha_i + \beta_i\, m(x_{it}) + \epsilon_{it}, \qquad i = 1, \ldots, n \;\; t = 1, \ldots, T$$

where y_{it} is the t^{th} response on the i^{th} individual, with the corresponding explanatory variable x_{it}.

The unknown coefficients α_i and β_i are the <u>location</u> and <u>scale</u> parameters associated with i^{th} model, $m(\cdot)$ is the <u>common nonparametric</u> component, and the ϵ_{it}'s are the error terms which are assumed to be independently and identically distributed with mean zero and constant variance σ^2.

Ma, W., Feng, Y., Chen, K. and Ying, Z. (2015). Functional and parametric estimation in a semi and nonparametric model with application to mass spectrometry data, The International Journal of Biostatistics, 11(2), 285–303.

1. Set $\alpha_1 = 0$ and $\beta_1 = 1$. The initial kernel estimate of $m(\cdot)$ is obtained from $(x_{1t}, y_{1t}),\ t = 1, ..., T$ as $\tilde{m}(x_{it}) = \frac{\sum_{t=1}^{T} \omega_{1t}(x) y_{1t}}{\sum_{t=1}^{T} \omega_{1t}(x)}$.
2. On replacing $m(\cdot)$ by $\tilde{m}(\cdot)$, the parameters $(\alpha_i, \beta_i), i = 2, ..., n$, are estimated from $y_{it} = \alpha_i + \beta_i \tilde{m}(x_{it}) + \epsilon_{it}$ as
$$\hat{\beta}_i = \frac{\sum_{t=1}^{T} [\tilde{m}(x_{it}) - \bar{\tilde{m}}(x_{i.})] y_{it}}{\sum_{t=1}^{T} [\tilde{m}(x_{it}) - \bar{\tilde{m}}(x_{i.})]^2}, \quad \hat{\alpha}_i = \bar{y}_{i.} - \hat{\beta}_i \bar{\tilde{m}}(x_{i.}).$$
3. Given the estimates $\hat{\alpha}_i$ and $\hat{\beta}_i$, the estimate for the function $m(\cdot)$ is updated as
$$\hat{m}(x) = \frac{\sum_{i=1}^{n} \sum_{t=1}^{T} \omega_{it}^*(x) y_{it}^*}{\sum_{i=1}^{n} \sum_{t=1}^{T} \omega_{it}^*(x)}.$$
4. Repeat Steps 2 and 3 until convergence is observed.

- Once $m(\cdot)$ is identified, the problem reduces to solving multiple simple regression models.
- $m(\cdot)$ is regarded as the nuisance component whereas $\tilde{m}(\cdot)$ is viewed as its consistent estimator.

The objective consists in improving the accuracy of the parametric estimators under the assumption “$H_0 : \beta_1 = \ldots = \beta_n = \beta_0$" (or $H_0 : \boldsymbol{\beta} = \beta_0 \mathbf{1}_n$), where β_0 is unknown.

The *unrestricted* (full model) estimates of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are obtained by applying the least squares approach as follows:

$$\hat{\boldsymbol{\beta}}^U = (\hat{\beta}_1^U, \ldots, \hat{\beta}_n^U)' \qquad \text{and} \qquad \hat{\boldsymbol{\alpha}}^U = \bar{\mathbf{y}} - T_n \hat{\boldsymbol{\beta}}^U, \tag{5}$$

where $\bar{\mathbf{y}} = (\bar{y}_{1\cdot}, \ldots, \bar{y}_{n\cdot})'$, $\hat{\beta}_i^U = [\mathbf{m}_i'\mathbf{y}_i - \frac{1}{T}(\mathbf{m}_i'\mathbf{1}_T)(\mathbf{y}_i'\mathbf{1}_T)]/(T\,Q_i)$, $T_n = \mathrm{Diag}(\bar{m}_{1\cdot}, \ldots, \bar{m}_{n\cdot})$, $T\,Q_i = \mathbf{m}_i'\mathbf{m}_i - \frac{1}{T}(\mathbf{m}_i'\mathbf{1}_T)^2$, $\bar{m}_{i\cdot} = \frac{1}{T}(\mathbf{m}_i'\mathbf{1}_T)$ and $\bar{y}_{i\cdot} = \frac{1}{T}(\mathbf{y}_i'\mathbf{1}_T)$.
Furthermore, the unbiased estimator of σ^2 is

$$s^2 = \frac{1}{nT - 2n} \sum_{i=1}^{n} \|\mathbf{y}_i - \hat{\boldsymbol{\alpha}}_i^U \mathbf{1}_T - \hat{\boldsymbol{\beta}}_i^U \mathbf{m}_i\|_2^2, \tag{6}$$

where $\|\cdot\|_2$ denotes the L_2-norm.

Under the null hypothesis, i.e., $H_0: \boldsymbol{\beta} = \beta_0 \mathbf{1}_n$, the *restricted* estimators of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are

$$\hat{\boldsymbol{\beta}}^R = \frac{\mathbf{1}_n \mathbf{1}_n' D_{22} \hat{\boldsymbol{\beta}}^U}{nTQ} \qquad \text{and} \qquad \hat{\boldsymbol{\alpha}}^R = \hat{\boldsymbol{\alpha}}^U + T_n H \hat{\boldsymbol{\beta}}^U, \tag{7}$$

where $H = I_n - [\mathbf{1}_n \mathbf{1}_n' D_{22}]/(nTQ)$ with $nQ = \sum_{i=1}^n Q_i$ and $D_{22}^{-1} = \mathrm{Diag}(TQ_1, \ldots, TQ_n)$.

Theorem

Under some regularity conditions, then given $m(\cdot)$, one has

$$\begin{pmatrix} \hat{\boldsymbol{\alpha}}^U - \boldsymbol{\alpha} \\ \hat{\boldsymbol{\beta}}^U - \boldsymbol{\beta} \end{pmatrix} \sim \mathcal{N}_{2n}\left\{ \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \sigma_1^2 \begin{pmatrix} D_{11} & -T_n D_{22} \\ -T_n D_{22} & D_{22} \end{pmatrix} \right\},$$

$$\begin{pmatrix} \hat{\boldsymbol{\alpha}}^R - \boldsymbol{\alpha} \\ \hat{\boldsymbol{\beta}}^R - \boldsymbol{\beta} \end{pmatrix} \sim \mathcal{N}_{2n}\left\{ \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \sigma_1^2 \begin{pmatrix} D_{11}^* & D_{12}^* \\ D_{21}^* & D_{22} \end{pmatrix} \right\},$$

$$\begin{pmatrix} \hat{\boldsymbol{\beta}}^U - \boldsymbol{\beta} \\ \hat{\boldsymbol{\beta}}^U - \hat{\boldsymbol{\beta}}^R \end{pmatrix} \sim \mathcal{N}_{2n}\left\{ \begin{pmatrix} \mathbf{0} \\ H\boldsymbol{\beta} \end{pmatrix}, \sigma_1^2 \begin{pmatrix} D_{22} & HD_{22} \\ D_{22}H' & HD_{22} \end{pmatrix} \right\},$$

$$\begin{pmatrix} \hat{\boldsymbol{\beta}}^R - \beta_0 \mathbf{1}_n \\ \hat{\boldsymbol{\beta}}^U - \hat{\boldsymbol{\beta}}^R \end{pmatrix} \sim \mathcal{N}_{2n}\left\{ \begin{pmatrix} (\bar{\beta} - \beta_0)\mathbf{1}_n \\ H\boldsymbol{\beta} \end{pmatrix}, \sigma_1^2 \begin{pmatrix} \frac{\mathbf{1}_n \mathbf{1}_n'}{nQ} & O \\ O & H D_{22} \end{pmatrix} \right\}$$

where $\bar{\beta}\mathbf{1}_n = \mathbf{1}_n\mathbf{1}_n' D_{22}^{-1}\boldsymbol{\beta}/(nQ)$, $D_{11} = (N)^{-1} + T_n D_{22} T_n$, $N = \mathrm{Diag}(T, \ldots, T)$, $D_{11}^* = (N)^{-1} + T_n \mathbf{1}_n \mathbf{1}_n' T_n/(nQ)$ and $D_{12}^* = \frac{1}{nQ}\mathbf{1}_n\mathbf{1}_n' T_n$ and $D_{21}^* = {D_{12}^*}'$.

In order to test $H_0 : \boldsymbol{\beta} = \beta_0 \mathbf{1}_n$, the likelihood ratio test statistic,

$$\mathcal{L}_n = \frac{\hat{\boldsymbol{\beta}}^{U'} H' D_{22}^{-1} H \hat{\boldsymbol{\beta}}^{U}}{(n-1)s^2},$$

is considered.

Theorem

Under the null hypothesis, $\mathcal{L}_n$ follows a central $F(n-1, T-2n)$ distribution, while under the alternative hypothesis, $H_a : \boldsymbol{\beta} \neq \beta_0 \mathbf{1}_n$, $\mathcal{L}_n$ follows the noncentral $F_{\Delta^2/2}(n-1, T-2n)$ distribution where $\Delta^2 = \hat{\boldsymbol{\beta}}^{U'} H' D_{22}^{-1} H \hat{\boldsymbol{\beta}}^{U} / \sigma^2$.

$$\text{Shrinkage:} \begin{cases} \hat{\boldsymbol{\beta}}^{S} &= \hat{\boldsymbol{\beta}}^{U} - c\,\mathcal{L}_n^{-1} H\hat{\boldsymbol{\beta}}^{U} \\ \hat{\boldsymbol{\alpha}}^{S} &= \hat{\boldsymbol{\alpha}}^{U} + c\,\mathcal{L}_n^{-1} T_n H\hat{\boldsymbol{\beta}}^{U}, \end{cases}$$

$$\text{Positive-Shrinkage:} \begin{cases} \hat{\boldsymbol{\alpha}}^{PS} &= \hat{\boldsymbol{\alpha}}^{U} + \{1 - (1 - c\,\mathcal{L}_n^{-1})\,\mathcal{I}(\mathcal{L}_n > c)\}\, T_n H\hat{\boldsymbol{\beta}}^{U}, \\ \hat{\boldsymbol{\beta}}^{PS} &= \hat{\boldsymbol{\beta}}^{U} - (1 - c\,\mathcal{L}_n^{-1})\,\mathcal{I}(\mathcal{L}_n > c) H\hat{\boldsymbol{\beta}}^{U}, \end{cases}$$

$$\text{Pretest:} \begin{cases} \hat{\boldsymbol{\beta}}^{PT} &= \hat{\boldsymbol{\beta}}^{U} - H\hat{\boldsymbol{\beta}}^{U}\,\mathcal{I}(\mathcal{L}_n < c_\alpha), \\ \hat{\boldsymbol{\alpha}}^{PT} &= \hat{\boldsymbol{\alpha}}^{U} + T_n H\hat{\boldsymbol{\beta}}^{U}\,\mathcal{I}(\mathcal{L}_n < c_\alpha), \end{cases}$$

where $c = \frac{(n-3)m}{m+2}$ with $m = n(T-2)$, $\mathcal{I}(\cdot)$ is the indicator function , and c_α (the critical point) is the the upper α-quantile of the central F distribution with $n-1$ and m degrees of freedom.

Theorem

Assuming that all the required conditions are satisfied, the bias associated with intercept and slope estimators are

$$\mathbf{b}(\hat{\beta}^U) = \mathbf{0}$$

$$\mathbf{b}(\hat{\beta}^R) = -H\beta$$

$$\mathbf{b}(\hat{\beta}^S) = -c(n-1)H\beta E[\chi^{-2}_{n+1}(\Delta^2)]$$

$$\mathbf{b}(\hat{\beta}^{PS}) = -H\beta\{G_{n+1,m}(c_1;\Delta^2) + c_1 E[F^{-1}_{n+1,m}(\Delta^2)\mathcal{I}(F_{n+1,m}(\Delta^2) > c_1)]\}$$

$$\mathbf{b}(\hat{\beta}^{PT}) = -H\beta\, G_{n+1,m}(\ell_\alpha;\Delta^2)$$

$$\mathbf{b}(\hat{\alpha}^U) = \mathbf{0}$$

$$\mathbf{b}(\hat{\alpha}^R) = T_n H\beta$$

$$\mathbf{b}(\hat{\alpha}^S) = c(n-1)T_n H\beta E[\chi^{-2}_{n+1}(\Delta^2)]$$

$$\mathbf{b}(\hat{\alpha}^{PS}) = c_1 T_n H\beta\,\{E[F^{-1}_{n+1,m}(\Delta^2)] - E[F^{-1}_{n+1,m}(\Delta^2)\mathcal{I}(F_{n+1,m}(\Delta^2) > c_1)]\} + T_n H\beta\, G_{n+1,m}(c_1;\Delta^2)$$

$$\mathbf{b}(\hat{\alpha}^{PT}) = T_n H\beta\, G_{n+1,m}(\ell_\alpha;\Delta^2)$$

where $\ell_\alpha = \frac{n-1}{n+1}F_{n-1,m}(\alpha)$, $\Delta^2 = \hat{\beta}^{U'} H' D_{22}^{-1} H\hat{\beta}^U/\sigma^2$, $c_1 = \frac{c(n-1)}{n+1}$, $\mathbf{b}(\hat{\beta}^*) = E(\hat{\beta}^* - \beta \mid M)$ *and* $\mathbf{b}(\hat{\alpha}^*) = E(\hat{\alpha}^* - \alpha \mid M)$.

Theorem

Assuming that all the required conditions are satisfied, the MSE associated with intercept and slope estimators are

$$\begin{cases} MSE(\hat{\boldsymbol\beta}^U) = & \sigma^2 D_{22} \\ MSE(\hat{\boldsymbol\alpha}^U) = & \sigma^2 D_{11} \end{cases} \qquad \begin{cases} MSE(\hat{\boldsymbol\beta}^R) = & \frac{\sigma^2 \mathbf{1}_n \mathbf{1}_n'}{TQ} + H\boldsymbol\beta\boldsymbol\beta' H' \\ MSE(\hat{\boldsymbol\alpha}^R) = & \sigma^2 D_{11}^* + T_n H\boldsymbol\beta\boldsymbol\beta' H' T_n' \end{cases}$$

$$\begin{cases} MSE(\hat{\boldsymbol\beta}^S) = \sigma^2 D_{22} - c(n-1)\sigma^2 HD_{22}\{2E[\chi_{n+1}^{-2}(\Delta^2)] - (n-3)E[\chi_{n+1}^{-4}(\Delta^2)]\} + c(n^2-1)(H\boldsymbol\beta\boldsymbol\beta' H')E[\chi_{n+3}^{-4}(\Delta^2)] \\ MSE(\hat{\boldsymbol\alpha}^S) = \sigma^2 D_{11} - c(n-1)\sigma^2 T_n HD_{22} H' T_n'\{2\Delta^2 E[\chi_{n+1}^{-2}(\Delta^2)] - (n-3)E[\chi_{n+1}^{-4}(\Delta^2)]\} \\ \qquad +c(n^2-1)(T_n H\boldsymbol\beta\boldsymbol\beta' H' T_n')E[\chi_{n+3}^{-4}(\Delta^2)] \end{cases}$$

$$\begin{cases} MSE(\hat{\boldsymbol\beta}^{PS}) = MSE(\hat{\boldsymbol\beta}^S) - (\sigma^2 HD_{22} - 2H\boldsymbol\beta\boldsymbol\beta' H')\, E[(1 - c_1 F_{n+1,m}^{-1}(\Delta^2))\mathcal{I}(F_{n+1,m}(\Delta^2) < c_1)] \\ \qquad -H\boldsymbol\beta\boldsymbol\beta' H' E[(1 - c_2 F_{n+3,m}^{-1}(\Delta^2))\mathcal{I}(F_{n+3,m}(\Delta^2) < c_2)] \\ MSE(\hat{\boldsymbol\alpha}^{PS}) = MSE(\hat{\boldsymbol\alpha}^S) - (\sigma^2 T_n HD_{22} H' T_n' - 2T_n H\boldsymbol\beta\boldsymbol\beta' H' T_n')\, E[(1 - c_1 F_{n+1,m}^{-1}(\Delta^2))\mathcal{I}(F_{n+1,m}(\Delta^2) < c_1)] \\ \qquad -T_n H\boldsymbol\beta\boldsymbol\beta' H' T_n' E[(1 - c_2 F_{n+3,m}^{-1}(\Delta^2))\mathcal{I}(F_{n+3,m}(\Delta^2) < c_2)] \end{cases}$$

$$\begin{cases} MSE(\hat{\boldsymbol\beta}^{PT}) = \sigma^2 D_{22} - \sigma^2 HD_{22} G_{n+1,m}(\ell_\alpha; \Delta^2) + H\boldsymbol\beta\boldsymbol\beta' H' \{2G_{n+1,m}(\ell_\alpha; \Delta^2) - G_{n+3,m}(\ell_\alpha^*; \Delta^2)\} \\ MSE(\hat{\boldsymbol\alpha}^{PT}) = \sigma^2 D_{11} - \sigma^2 (D_{11} - D_{11}^*) G_{n+1,m}(\ell_\alpha; \Delta^2) + T_n H\boldsymbol\beta\boldsymbol\beta' H' T_n' \{2G_{n+1,m}(\ell_\alpha; \Delta^2) - G_{n+3,m}(\ell_\alpha^*; \Delta^2)\} \end{cases}$$

where $\ell_\alpha^* = \frac{n-1}{n+3} F_{n-1,m}(\alpha)$, $c_2 = \frac{c(n-1)}{n+3}$, $D_{11}^* = N^{-1} + \frac{T_n' \mathbf{1}_n \mathbf{1}_n' T_n}{TQ}$,
$MSE(\hat{\boldsymbol\beta}^*) = E((\hat{\boldsymbol\beta}^* - \boldsymbol\beta)(\hat{\boldsymbol\beta}^* - \boldsymbol\beta)' \mid M)$ *and* $MSE(\hat{\boldsymbol\alpha}^*) = E((\hat{\boldsymbol\alpha}^* - \boldsymbol\alpha)(\hat{\boldsymbol\alpha}^* - \boldsymbol\alpha)' \mid M)$.

The "relative error" (RE) which is utilized as a measure of accuracy is

$$RE(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*) = \frac{Er(\boldsymbol{\alpha}^U, \boldsymbol{\beta}^U)}{Er(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)},$$

where

$$Er(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*) = \sum_{i=1}^{n} (\hat{\alpha}_i^* - \alpha_i)^2 + (\hat{\beta}_i^* - \beta_i)^2,$$

the α_i's and β_i's being the true values for $i = 1, \ldots, n$.

- A larger relative error represents an improvement in accuracy over $(\boldsymbol{\alpha}^U, \boldsymbol{\beta}^U)$
- Simulation studies were carried out under both H_0 and H_a for different values of Δ where $\Delta = \|\boldsymbol{\beta} - \beta_0 \mathbf{1}_n\|_2^2$, Δ being equal to zero under the null hypothesis.

Example

Let

- $\boldsymbol{\alpha} = (0, 0.2, 0.5, 0.7, 1, 0, 0.2, 0.5, 0.7, 1, \ldots)'$
- $\boldsymbol{\beta} = (1, 1, \ldots, 1)'$
- $m(x)$ be $\sin(x)$
- the error terms be independently and identically distributed from $\mathcal{N}(0, 0.25)$
- the x_{it} values randomly generated from $U(0, 20)$
- the values of (h, h^*) set to be $(0.1, 0.1)$ on the basis of a 5-fold cross-validation

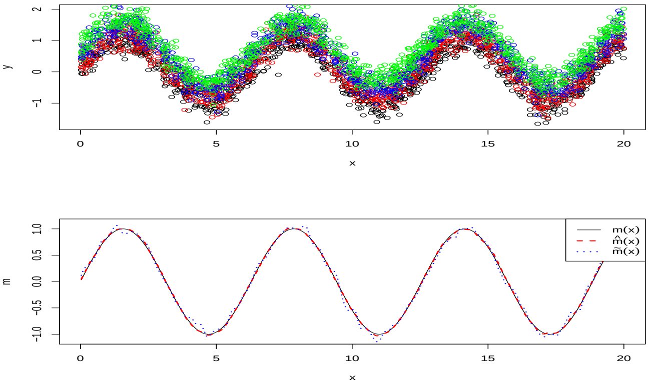


Figure: Scatter plots of the first four sets of simulated values (top panel); the actual function $m(x) = \sin(x)$ and the estimates $\tilde{m}(x)$ and $\hat{m}(x)$ (bottom panel).

Table: $RE(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$ for certain values of n and T ($\Delta = 0$)

n	T	$RE(\boldsymbol{\alpha}^R, \boldsymbol{\beta}^R)$	$RE(\boldsymbol{\alpha}^{PS}, \boldsymbol{\beta}^{PS})$	$RE(\boldsymbol{\alpha}^{PT}, \boldsymbol{\beta}^{PT})$
10	500	1.7389	1.4152	1.5031
10	1000	1.7440	1.5841	1.7440
10	1500	1.5923	1.5495	1.5923
10	2000	1.9465	1.6802	1.9465
10	2500	1.9358	1.5848	1.9358
10	3000	1.4921	1.3002	1.4921
15	2500	1.6641	1.6045	1.6641
20	2500	1.4713	1.4408	1.4713
25	2500	1.9156	1.7880	1.9156
30	2500	1.6359	1.5931	1.6359
35	2500	1.5651	1.4930	1.5651

Table: $RE(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$ for certain $\Delta \geq 0$ with $n = 20$, $T = 1500$

Δ	$RE(\boldsymbol{\alpha}^R, \boldsymbol{\beta}^R)$	$RE(\boldsymbol{\alpha}^{PS}, \boldsymbol{\beta}^{PS})$	$RE(\boldsymbol{\alpha}^{PT}, \boldsymbol{\beta}^{PT})$
0.00	1.99974	1.90482	1.99974
0.05	0.85712	1.24353	0.97805
0.10	0.36683	1.09256	1.00000
0.15	0.19461	1.05267	1.00000
0.20	0.11837	1.03648	1.00000
0.30	0.05624	1.02329	1.00000
0.40	0.03259	1.01790	1.00000
0.50	0.02124	1.01506	1.00000
0.60	0.01495	1.01332	1.00000
0.70	0.01111	1.01214	1.00000
0.80	0.00860	1.01130	1.00000
0.90	0.00687	1.01065	1.00000
1.00	0.00562	1.01013	1.00000

Stein-type location and scale estimators

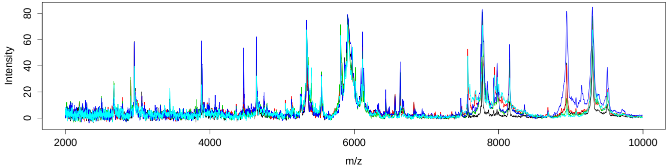
Application to a mass spectrometry data ($n = 35$ and $T = 21000$)

Shrinkage estimation techniques are now applied to an actual mass spectrometry data set, namely, the SELDI-TOF data set collected from a study on liver cancer patients conducted at Chengzheng Hospital Shanghai. The accuracy of the various estimates is measured by evaluating their mean squared error (MSE):

$$MSE^* = \frac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it} - \hat{\alpha}_i^* - \hat{\beta}_i^* \hat{m}(x_{it})^*)^2.$$

Application to mass spectrometry data

Data set: 21000 observations of 35 individuals



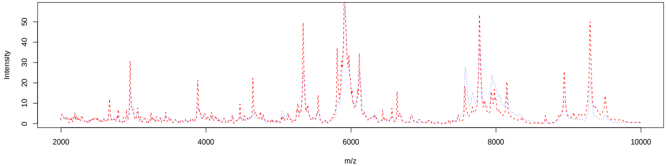


Figure: The observations on five individuals randomly selected from the mass spectrometry data set (top panel); the estimates $\tilde{m}(x)$ in blue and $\hat{m}(x)$ in red (bottom panel).

Application to mass spectrometry data

Data set: 21000 observations of 35 individuals

Table: MSE values obtained from various techniques applied to a mass spectrometry data set.

	Unrestricted	Restricted	Positive Shrinkage	Pretest
MSE	13.48971	17.95069	12.38967	13.48971

Thank you!