How to Make the Most Out of Evolving Information: Function Identification Using Epi-Splines

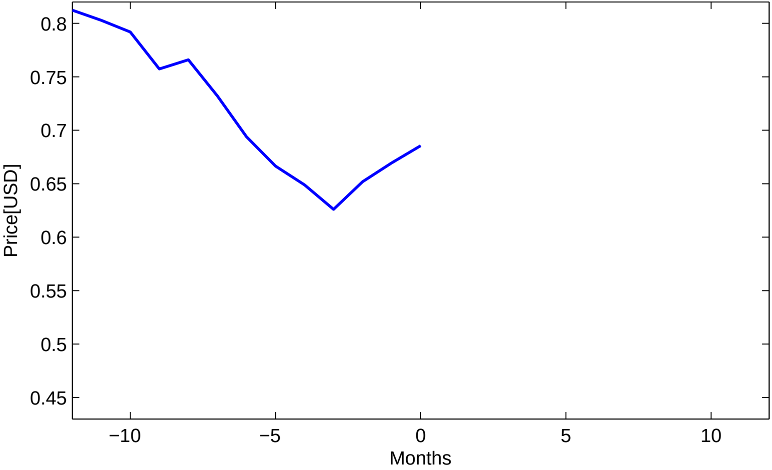
Johannes O. Royset
Operations Research, NPS
in collaboration with L. Bonfiglio, S. Buttrey, S. Brizzolara, J. Carbaugh, I.Rios, D. Singham, K. Spurkel, G. Vernengo, J-P. Watson, R. Wets, D. Woodruff

Smart Energy and Stochastic Optimization, ENPC ParisTech
June 2015

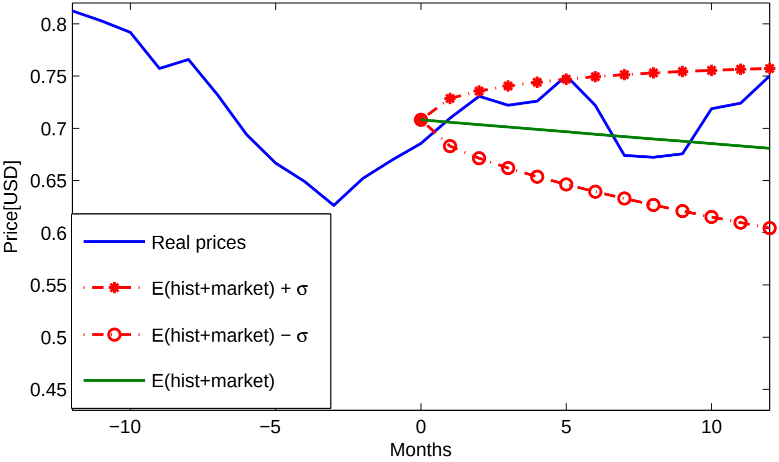
This material is based upon work supported in part by the U.S. Army Research Laboratory and the U.S. Army Research Office under grants 00101-80683, W911NF-10-1-0246 and W911NF-12-1-0273 as well as by DARPA under HR0011-14-1-0060

Goal: estimate, predict, forecast

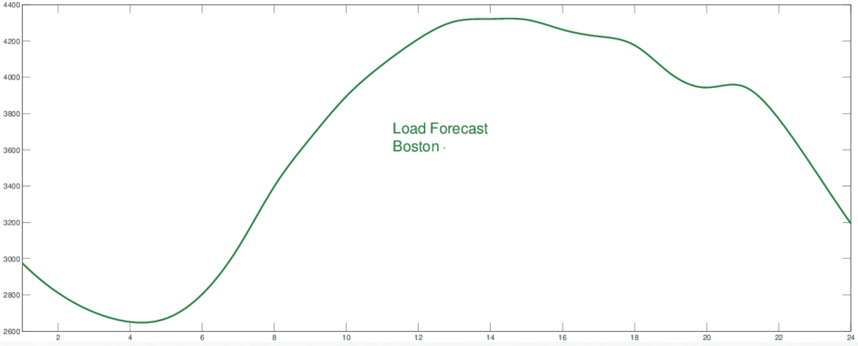
Copper prices



Copper prices

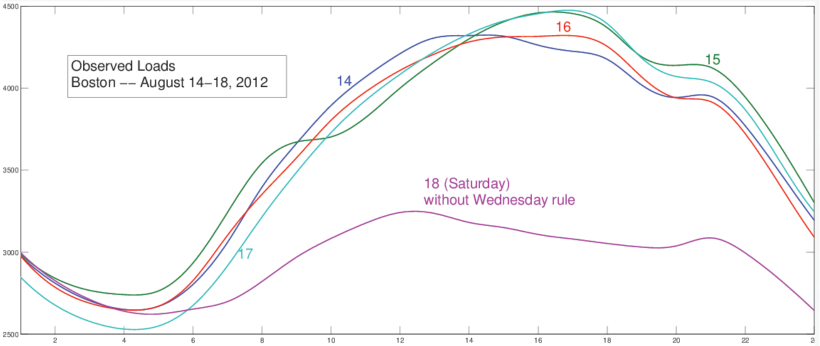


Hourly electricity loads



Hard information: data

- Observations
- Usually scarce, excessive, corrupted, uncertain



Soft information

- Indirect observations, external information
- Knowledge about
  - structures
  - established “laws”
  - physical restrictions
  - shapes, smoothness of curves

Evolving information

data acquisition, information growth

Function Identification Problem

Identify a function that

best represents all available information

Applications of function identification approach

- energy
- natural resources
- financial markets
- image reconstruction
- uncertain quantification
- variograms
- nonparametric regression
- deconvolution
- density estimation

Outline

- Illustrations and formulations (slides 12-33)
- Framework (slides 34-60)
- Epi-splines and approximations (slides 61-77)
- Implementations (slides 78-85)
- Examples (slides 86-124)

Illustrations and formulations

Identifying financial curves

Estimate discount factor curve given instruments $i = 1, 2, ..., I$ and

$$\text{payments} \quad p_0^i, p_1^i, ..., p_{N_i}^i \quad \text{at times} \quad 0 = t_0^i, t_1^i, ..., t_{N_i}^i$$

Identifying financial curves

Estimate discount factor curve given instruments $i = 1, 2, ..., I$ and

$$\text{payments} \quad p_0^i, p_1^i, ..., p_{N_i}^i \quad \text{at times} \quad 0 = t_0^i, t_1^i, ..., t_{N_i}^i$$

Identify nonnegative, nonincreasing function f with $f(0) = 1$

$$\sum_{j=0}^{N_i} f(t_j^i) p_j^i \approx 0 \quad \text{for all } i$$

Identifying financial curves

Estimate discount factor curve given instruments $i = 1, 2, ..., I$ and

$$\text{payments} \quad p_0^i, p_1^i, ..., p_{N_i}^i \quad \text{at times} \quad 0 = t_0^i, t_1^i, ..., t_{N_i}^i$$

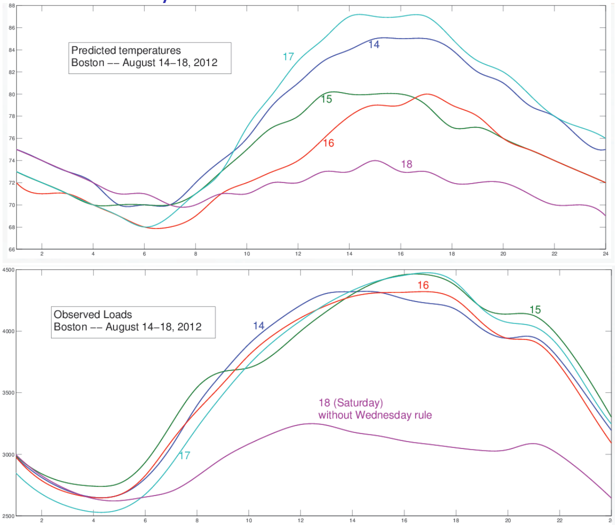
Identify nonnegative, nonincreasing function f with $f(0) = 1$

$$\sum_{j=0}^{N_i} f(t_j^i) p_j^i \approx 0 \quad \text{for all } i$$

Constrained infinite-dimensional optimization problem:

$$\begin{aligned} &\text{minimize} \ \sum_{i=1}^{I} \left| \sum_{j=0}^{N_i} f(t_j^i) p_j^i \right| \\ &\text{such that} \ \ f(0) = 1, \ f \geq 0, \ f' \leq 0 \end{aligned}$$

Day-ahead electricity load forecast



Data: predicted temperature, dew point; observed load

Day-ahead load forecast problem

Using a relevant data and weather forecast for tomorrow, determine tomorrow's load curve and its uncertainty

Day-ahead load forecast problem

Using a relevant data and weather forecast for tomorrow, determine tomorrow's load curve and its uncertainty

- $j = 1, 2, ..., J$: days in data set

Day-ahead load forecast problem

Using a relevant data and weather forecast for tomorrow, determine tomorrow's load curve and its uncertainty

- $j = 1, 2, ..., J$: days in data set
- $h = 1, 2, ..., 24$: hours of the day

Day-ahead load forecast problem

Using a relevant data and weather forecast for tomorrow, determine tomorrow's load curve and its uncertainty

- $j = 1, 2, ..., J$: days in data set
- $h = 1, 2, ..., 24$: hours of the day
- t_h^j: predicted temperature in hour h of day j

Day-ahead load forecast problem

Using a relevant data and weather forecast for tomorrow, determine tomorrow's load curve and its uncertainty

- $j = 1, 2, ..., J$: days in data set
- $h = 1, 2, ..., 24$: hours of the day
- t_h^j: predicted temperature in hour h of day j
- d_h^j: predicted dew point in hour h of day j

Day-ahead load forecast problem

Using a relevant data and weather forecast for tomorrow, determine tomorrow's load curve and its uncertainty

- $j = 1, 2, ..., J$: days in data set
- $h = 1, 2, ..., 24$: hours of the day
- t_h^j: predicted temperature in hour h of day j
- d_h^j: predicted dew point in hour h of day j
- l_h^j: actual observed load in hour h of day j

Day-ahead load forecast problem (cont.)

"Functional" regression model:

$$l_h^j = f^{\mathrm{tmp}}(h) t_h^j + f^{\mathrm{dpt}}(h) d_h^j + e_h^j,$$

- regression "coefficients" f^{tmp} and f^{dpt} are functions of time
- e_h^j: error between the observed and predicted loads

Day-ahead load forecast problem (cont.)

"Functional" regression model:

$$l_h^j = f^{\mathrm{tmp}}(h)t_h^j + f^{\mathrm{dpt}}(h)d_h^j + e_h^j,$$

- regression "coefficients" f^{tmp} and f^{dpt} are functions of time
- e_h^j: error between the observed and predicted loads

Regression problem: find functions f^{tmp} and f^{dpt} on $[0, 24]$ that

$$\text{minimize} \sum_{j=1}^{J} \sum_{h=1}^{24} \left| l_h^j - \left[f^{\mathrm{tmp}}(h)t_h^j + f^{\mathrm{dpt}}(h)d_h^j \right] \right|$$

subject to smoothness, curvature conditions

Day-ahead load forecast problem (cont.)

"Functional" regression model:

$$l_h^j = f^{\mathrm{tmp}}(h) t_h^j + f^{\mathrm{dpt}}(h) d_h^j + e_h^j,$$

- regression "coefficients" f^{tmp} and f^{dpt} are functions of time
- e_h^j: error between the observed and predicted loads

Regression problem: find functions f^{tmp} and f^{dpt} on $[0, 24]$ that

$$\text{minimize } \sum_{j=1}^{J} \sum_{h=1}^{24} \left| l_h^j - \left[f^{\mathrm{tmp}}(h) t_h^j + f^{\mathrm{dpt}}(h) d_h^j \right] \right|$$

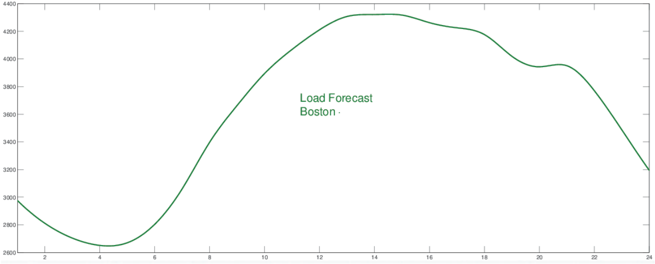
subject to smoothness, curvature conditions

Constrained infinite-dimensional optimization problem

Day-ahead load forecast problem (cont.)

Given temperature t and dew point d forecasts (functions of time) for tomorrow,

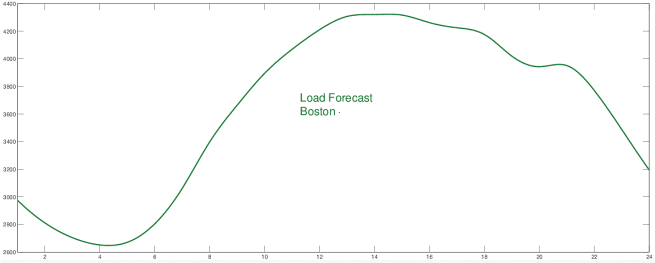
$$\text{load forecast becomes } f^{\mathrm{tmp}}(h)t(h) + f^{\mathrm{dpt}}(h)d(h), \quad h \in [0, 24]$$



Day-ahead load forecast problem (cont.)

Given temperature t and dew point d forecasts (functions of time) for tomorrow,

$$\text{load forecast becomes } f^{\mathrm{tmp}}(h)t(h) + f^{\mathrm{dpt}}(h)d(h), \quad h \in [0, 24]$$



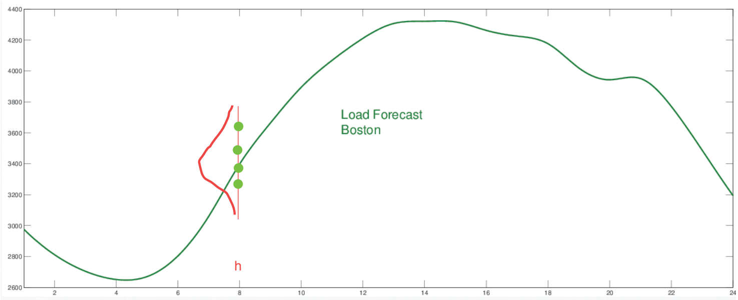
But, what about uncertainty?

Estimation of errors

For hour h:

$$\text{minimized errors } e_h^j = l_h^j - [f^{\mathrm{tmp}}(h)t_h^j + f^{\mathrm{dpt}}(h)d_h^j], \quad j = 1, \ldots, J$$

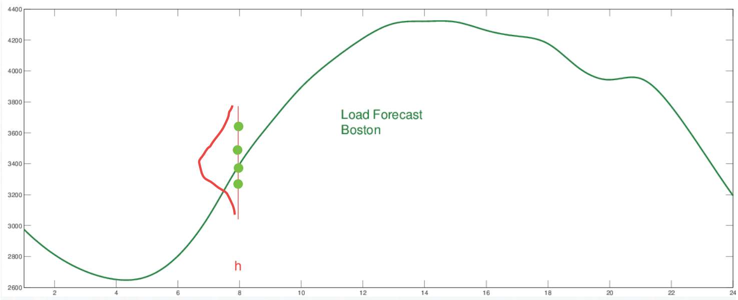
$=$ samples from actual error density



Estimation of errors

For hour h:

$$\text{minimized errors } e_h^j = l_h^j - [f^{\mathrm{tmp}}(h)t_h^j + f^{\mathrm{dpt}}(h)d_h^j], \quad j = 1, \ldots, J$$
$$= \text{ samples from actual error density}$$



Probability density estimation problem with very small data set

Estimation of errors (cont.)

Probability density estimation problem:

Given iid sample $x^1, x^2, ..., x^\nu$,

find a function $h \in \mathcal{H}$ that maximizes $\prod_{i=1}^{\nu} h(x^i)$

Estimation of errors (cont.)

Probability density estimation problem:

Given iid sample $x^1, x^2, ..., x^\nu$,

find a function $h \in \mathcal{H}$ that maximizes $\prod_{i=1}^{\nu} h(x^i)$

$\mathcal{H}$ = constraints on h: $h \geq 0$, $\int h(x)dx = 1$, and more

Estimation of errors (cont.)

Probability density estimation problem:

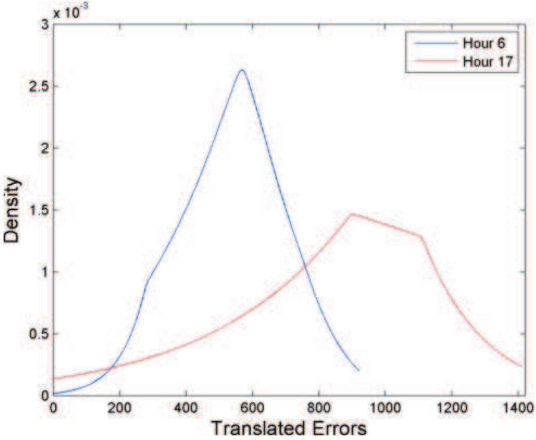
Given iid sample $x^1, x^2, ..., x^\nu$,

find a function $h \in \mathcal{H}$ that maximizes $\prod_{i=1}^{\nu} h(x^i)$

$\mathcal{H}$ = constraints on h: $h \geq 0$, $\int h(x)dx = 1$, and more

Constrained infinite-dimensional optimization problem

Estimation of errors (cont.)



Generating load scenarios

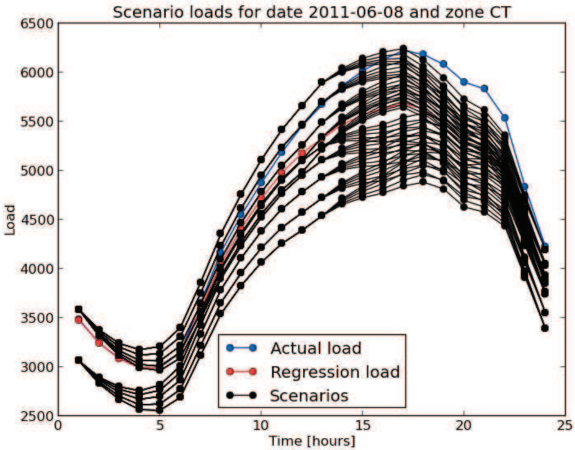
Also need to consider conditioning:

- if load above regression function early, then likely above later
- error data for density estimation reduced further

Generating load scenarios

Also need to consider conditioning:

- if load above regression function early, then likely above later
- error data for density estimation reduced further



Forecast precision (Connecticut 2010-12)

Season	Mean % error
Fall	3.99
Spring	2.73
Summer	4.19
Winter	3.47

Uncertainty quantification (UQ)

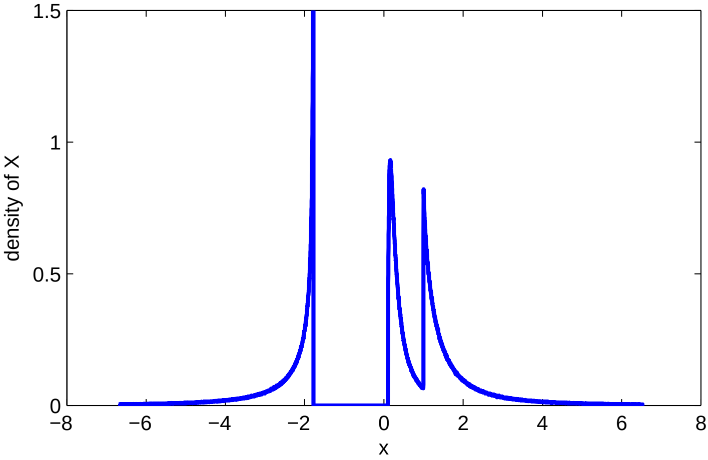
Engineering, biological, physical systems

- Input: random vector V ("known" distribution)
- System function g; implicitly defined e.g. by simulation
- Output: random variable

$$X = g(V)$$

Illustration of UQ challenges

Output amplitude of dynamical system:



How to estimate densities like this with a small sample?

Probability density estimation

Sample $X^1, \ldots, X^\nu$: maximize $\prod_{i=1}^{\nu} h(X^i)$ s.t. $h \in H^\nu \subset \mathcal{H}$

- Sample size might be growing
- Constraint set H^ν might be evolving

Evolving probability density estimation

maximize $\prod_{i=1}^{\nu}(h(X^i))^{1/\nu}$ s.t. $h \in H^\nu \subset \mathcal{H}$

Evolving probability density estimation

maximize $\prod_{i=1}^{\nu}(h(X^i))^{1/\nu}$ s.t. $h \in H^{\nu} \subset \mathcal{H}$

maximize $\frac{1}{\nu}\sum_{i=1}^{\nu} \log h(X^i)$ s.t. $h \in H^{\nu} \subset \mathcal{H}$

Evolving probability density estimation

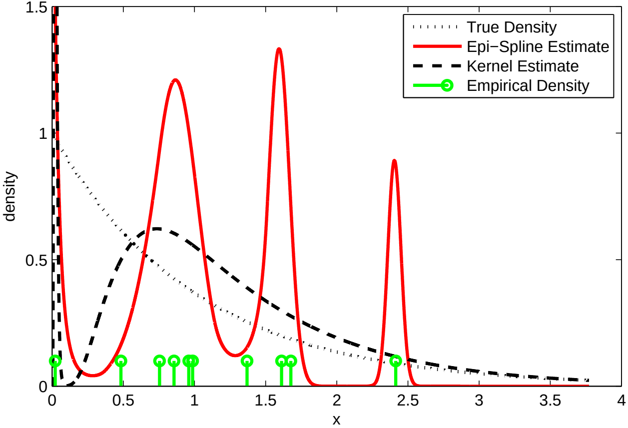
maximize $\prod_{i=1}^{\nu}(h(X^i))^{1/\nu}$ s.t. $h \in H^\nu \subset \mathcal{H}$

maximize $\frac{1}{\nu}\sum_{i=1}^{\nu} \log h(X^i)$ s.t. $h \in H^\nu \subset \mathcal{H}$

actual problem: maximize $E[\log h(X)]$ s.t. $h \in H \subset \mathcal{H}$

Evolving constraints

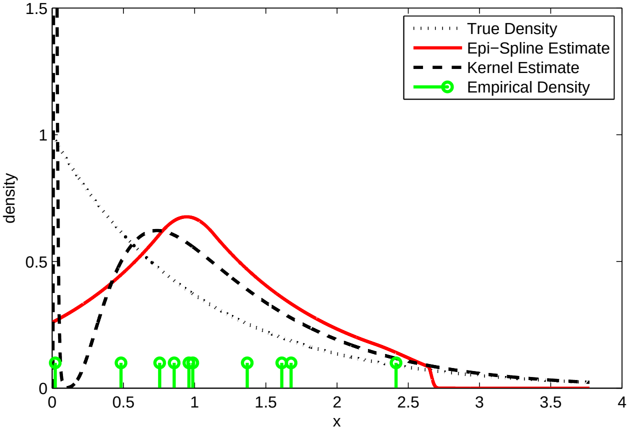
Constraints: nonnegative support, smooth, curvature bound



MSE $= 0.2765$ (epi-spline) and 0.3273 (kernel)

Evolving constraints (cont.)

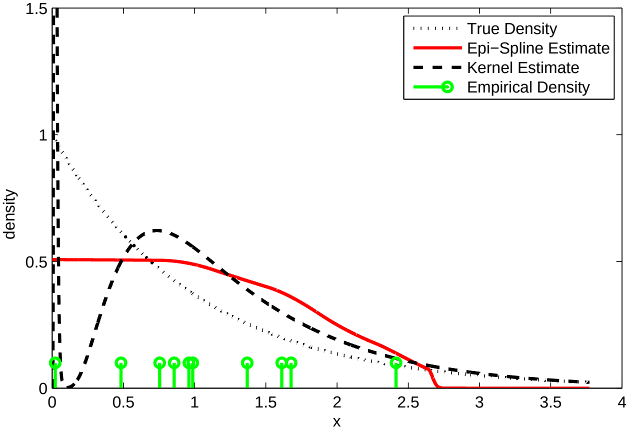
Also log-concave



MSE $= 0.1144$ (epi-spline) and 0.3273 (kernel)

Evolving constraints (cont.)

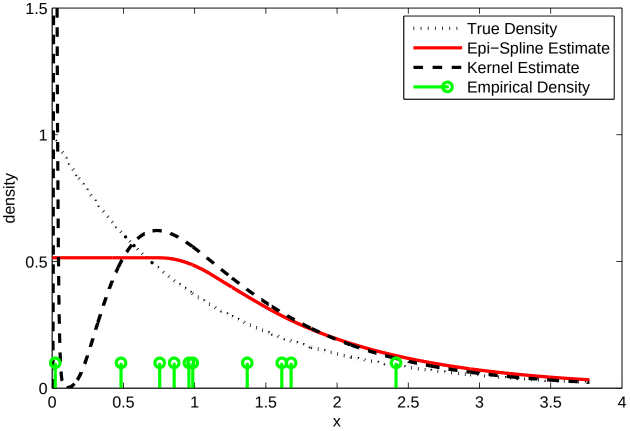
Also nonincreasing



MSE $= 0.0470$ (epi-spline) and 0.3273 (kernel)

Evolving constraints (cont.)

Also slope bounds



MSE $= 0.0416$ (epi-spline) and 0.3273 (kernel)

Take-aways so far

- Challenges in forecasting and estimation are function identification problems
- Day-ahead forecasting of electricity demand involves
  - functional regression to get trend
  - probability density estimation to get errors
- Soft information supplements hard information (data)

Questions?

Outline

Framework

Actual problem

- Constrained infinite-dimensional optimization problem

$$\begin{aligned}&\min \psi(f)\\&\text{such that } f \in F \subseteq \mathcal{F} \subseteq \text{ some function space}\end{aligned}$$

Actual problem

- Constrained infinite-dimensional optimization problem

$$
\begin{aligned}
&\min \psi(f) \\
&\text{such that } f \in F \subseteq \mathcal{F} \subseteq \text{ some function space}
\end{aligned}
$$

- criterion ψ (norms, error measures, etc)

Actual problem

- Constrained infinite-dimensional optimization problem

$$\begin{aligned} &\min \psi(f) \\ &\text{such that } f \in F \subseteq \mathcal{F} \subseteq \text{ some function space} \end{aligned}$$

- criterion ψ (norms, error measures, etc)
- feasible set F (shape restrictions, external info)

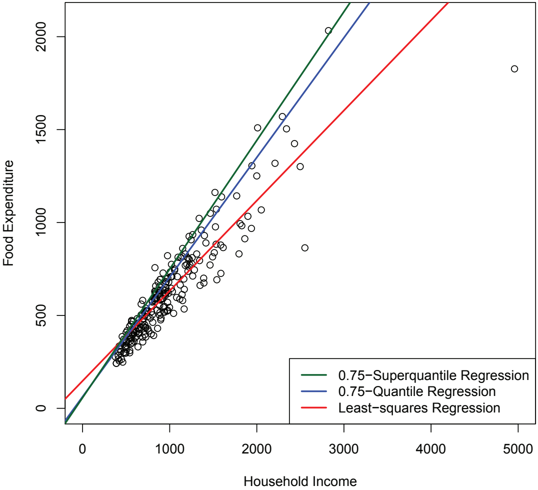
Actual problem

- Constrained infinite-dimensional optimization problem

$$\begin{aligned} &\min \psi(f) \\ &\text{such that } f \in F \subseteq \mathcal{F} \subseteq \text{ some function space} \end{aligned}$$

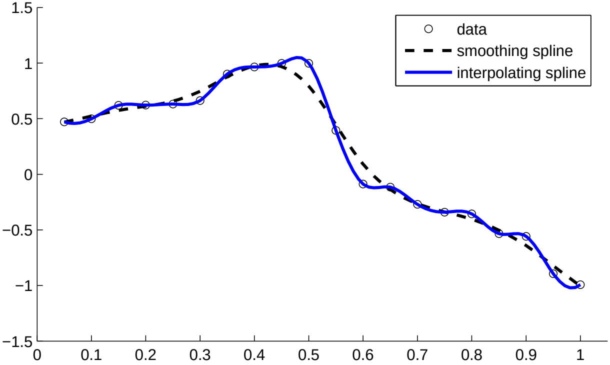
- criterion ψ (norms, error measures, etc)
- feasible set F (shape restrictions, external info)
- subspace of interest $\mathcal{F}$ (continuity, smoothness)

Special case: linear regression



find affine f that $\min \sum_j (y^j - f(x^j))^2$ or other error measures

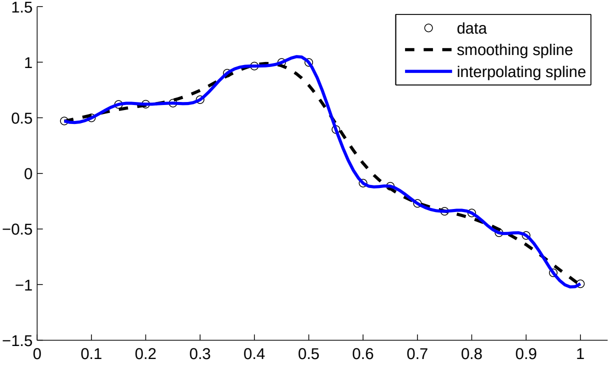
Special case: interpolation



$$\min \|f''\|^2$$

such that $f(x^j) = y^j$ for all j and $f \in$ Sobolev space

Special case: smoothing



$$\min \sum_j (y^j - f(x^j))^2 + \lambda \|f''\|^2$$

such that $f \in$ Sobolev space

Evolving problems

- Evolving infinite-dimensional optimization problems

$$\begin{aligned} &\min \psi^\nu(f) \\ &\text{such that } f \in F^\nu \subseteq \mathcal{F} \subseteq \text{ some function space} \end{aligned}$$

- evolving/approximating criterion ψ^ν
- evolving/approximating feasible set F^ν

Evolving problems

- Evolving infinite-dimensional optimization problems

$$\min \psi^\nu(f)$$
$$\text{such that } f \in F^\nu \subseteq \mathcal{F} \subseteq \text{ some function space}$$

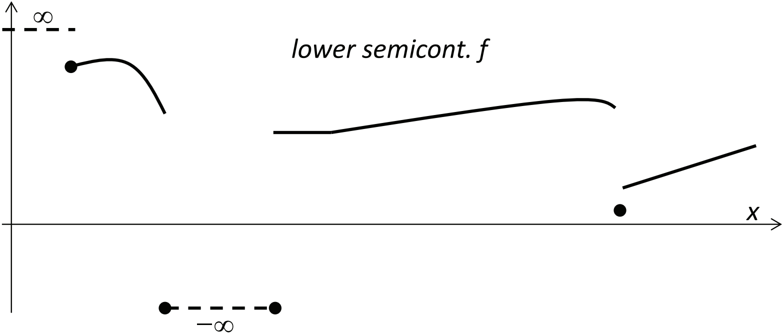
- evolving/approximating criterion ψ^ν
- evolving/approximating feasible set F^ν

Function space

extended real-valued lower semicontinuous functions on $\boldsymbol{R}^n$

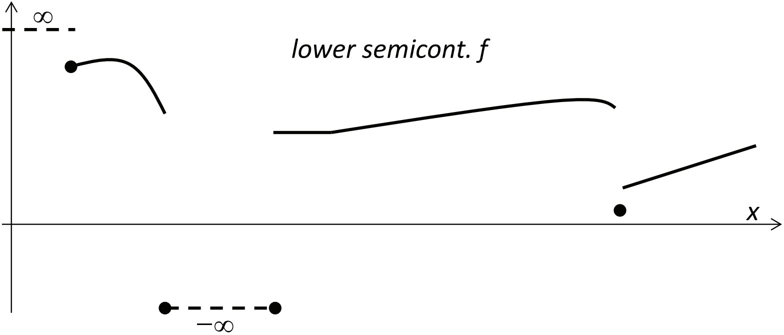
Function space

extended real-valued lower semicontinuous functions on $\mathbb{R}^n$



Function space

extended real-valued lower semicontinuous functions on $\boldsymbol{R}^n$

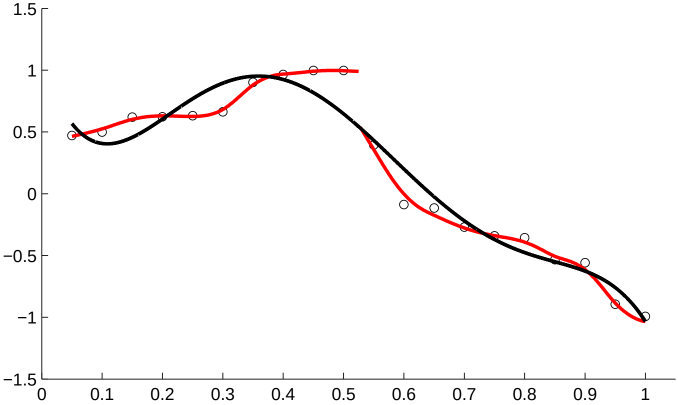


$(\text{lsc-fcns}(\boldsymbol{R}^n), dl)$: complete separable metric, dl = epi-distance

Modeling possibilities with lsc functions

- functions with jumps and high growth
- response surface building with implicit constraints
- system identification with subsequent minimization
- nonlinear transformations requiring $\pm\infty$
- functions with unbounded domain

Jumps



Fits by polynomial and lsc function

Modeling possibilities with lsc functions

- functions with jumps and high growth
- response surface building with implicit constraints
- system identification with subsequent minimization
- nonlinear transformations requiring $\pm\infty$
- functions with unbounded domain

Nonlinear transformations

Recall: probability density estimation with sample $X^1, ..., X^\nu$:

maximize $\frac{1}{\nu}\sum_{i=1}^{\nu} \log h(X^i)$ s.t. $h \in H^\nu \subset \mathcal{H}$ including $h \geq 0$

Nonlinear transformations

Recall: probability density estimation with sample $X^1, ..., X^\nu$:

maximize $\frac{1}{\nu}\sum_{i=1}^{\nu} \log h(X^i)$ s.t. $h \in H^\nu \subset \mathcal{H}$ including $h \geq 0$

Exponential transformation: $h(x) = \exp(-s(x))$

Nonlinear transformations

Recall: probability density estimation with sample $X^1, ..., X^\nu$:

maximize $\frac{1}{\nu}\sum_{i=1}^{\nu} \log h(X^i)$ s.t. $h \in H^\nu \subset \mathcal{H}$ including $h \geq 0$

Exponential transformation: $h(x) = \exp(-s(x))$

minimize $\sum_{i=1}^{\nu} s(X^i)$ s.t. $s \in S^\nu \subset \mathcal{S}$ excluding nonnegativity

Nonlinear transformations

Recall: probability density estimation with sample $X^1, \ldots, X^\nu$:

maximize $\frac{1}{\nu}\sum_{i=1}^{\nu} \log h(X^i)$ s.t. $h \in H^\nu \subset \mathcal{H}$ including $h \geq 0$

Exponential transformation: $h(x) = \exp(-s(x))$

minimize $\sum_{i=1}^{\nu} s(X^i)$ s.t. $s \in S^\nu \subset \mathcal{S}$ excluding nonnegativity

- h log-concave $\iff$ s convex

Nonlinear transformations

Recall: probability density estimation with sample $X^1, ..., X^\nu$:

maximize $\frac{1}{\nu}\sum_{i=1}^{\nu} \log h(X^i)$ s.t. $h \in H^\nu \subset \mathcal{H}$ including $h \geq 0$

Exponential transformation: $h(x) = \exp(-s(x))$

minimize $\sum_{i=1}^{\nu} s(X^i)$ s.t. $s \in S^\nu \subset \mathcal{S}$ excluding nonnegativity

- h log-concave $\iff$ s convex
- If $s = \langle c(\cdot), r\rangle$, then certain expression linear in r

Nonlinear transformations

Recall: probability density estimation with sample $X^1, \ldots, X^\nu$:

maximize $\frac{1}{\nu}\sum_{i=1}^{\nu} \log h(X^i)$ s.t. $h \in H^\nu \subset \mathcal{H}$ including $h \geq 0$

Exponential transformation: $h(x) = \exp(-s(x))$

minimize $\sum_{i=1}^{\nu} s(X^i)$ s.t. $s \in S^\nu \subset \mathcal{S}$ excluding nonnegativity

- h log-concave $\iff$ s convex
- If $s = \langle c(\cdot), r\rangle$, then certain expression linear in r

But need s to take on ∞ for $h = \exp(-s(\cdot))$ to vanish

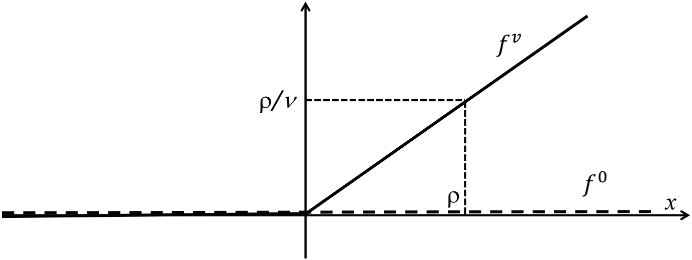
Identifying functions with unbounded domains

Complications for "standard" spaces

Identifying functions with unbounded domains

Complications for "standard" spaces

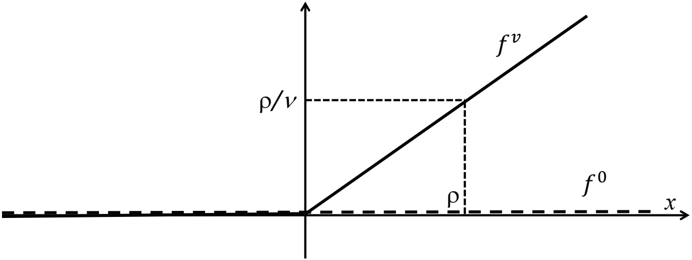
Is f^ν near f^0?



Identifying functions with unbounded domains

Complications for "standard" spaces

Is f^ν near f^0?

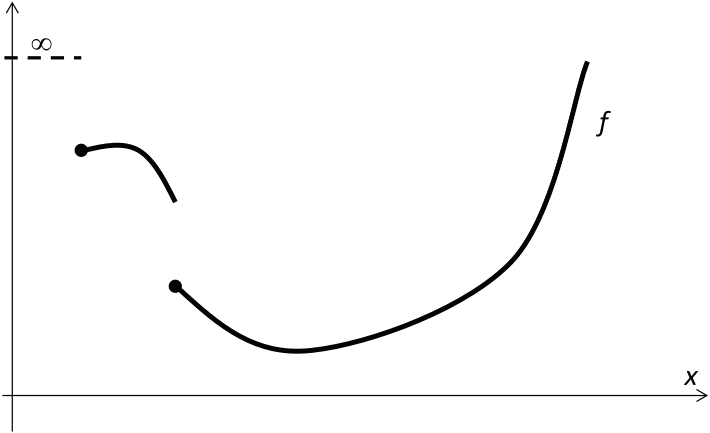


Not in $\mathcal{L}^p$-norm, but epi-distance $\leq 1/\nu$

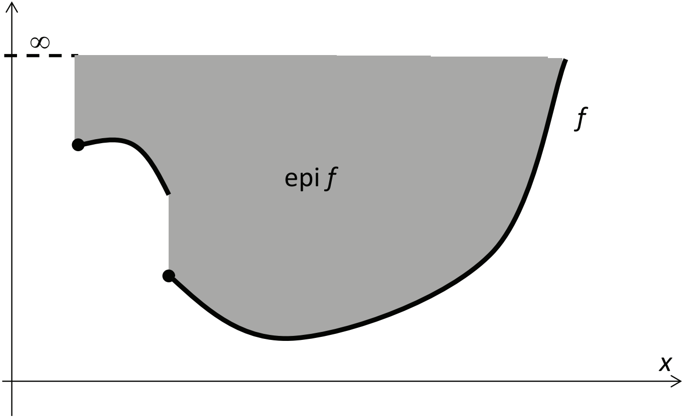
Modeling possibilities with lsc functions

- functions with jumps and high growth
- response surface building with implicit constraints
- system identification with subsequent minimization
- nonlinear transformations requiring $\pm\infty$
- functions with unbounded domain

Epi-graph



Epi-graph



Geometric view of lsc functions

$$f \in \mathrm{lsc\text{-}fcns}(\boldsymbol{R}^n) \Longleftrightarrow \mathrm{epi}\ f \text{ nonempty closed subset of } \boldsymbol{R}^{n+1}$$

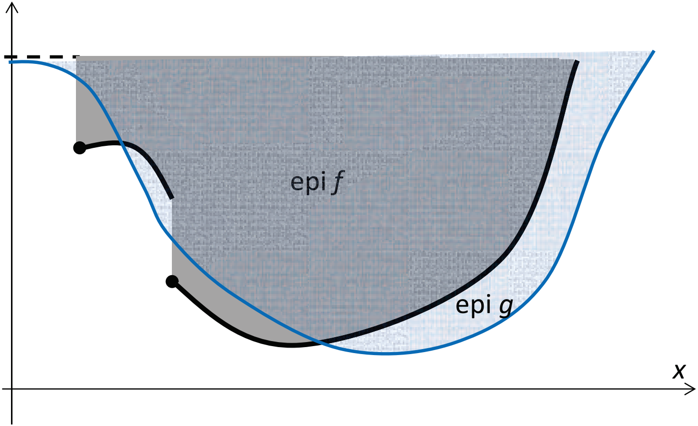
Geometric view of lsc functions

$$f \in \text{lsc-fcns}(\boldsymbol{R}^n) \Longleftrightarrow \text{epi } f \text{ nonempty closed subset of } \boldsymbol{R}^{n+1}$$

Notation:

- $\rho\boldsymbol{B} = \boldsymbol{B}(0,\rho) =$ origin-centered ball with radius ρ
- $d(y,S) = \inf_{y' \in S} \|y - y'\| =$ distance between y and S

Distances between epi-graphs

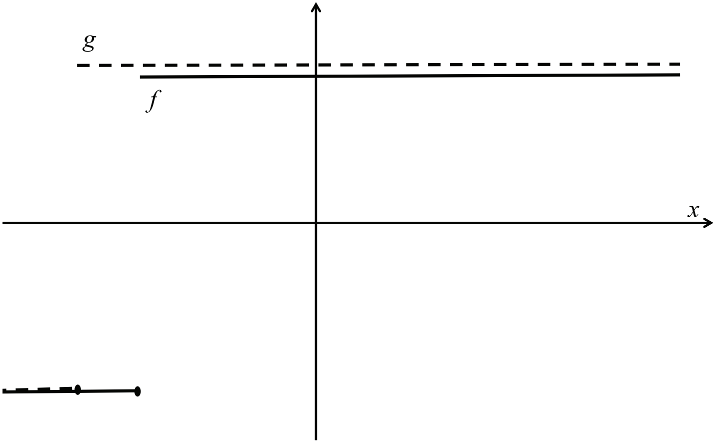


Distances between epi-graphs (cont.)

$$\rho\text{-}\textit{epi-distance} = dl_\rho(f,g) = \sup_{\bar{x}\in\rho\boldsymbol{B}} |d(\bar{x}, \mathrm{epi}\ f) - d(\bar{x}, \mathrm{epi}\ g)|, \quad \rho \geq 0$$

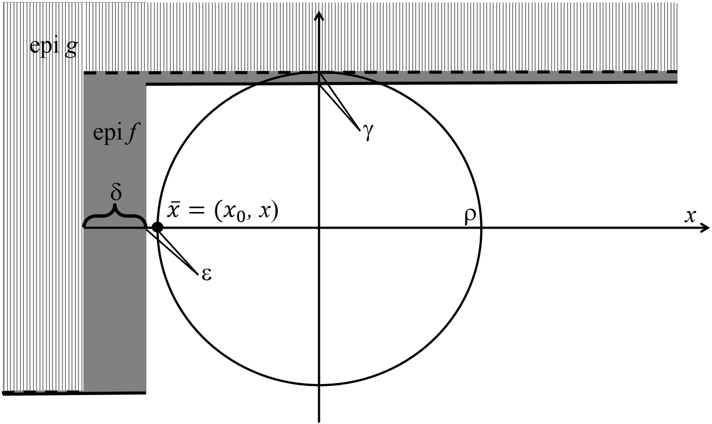
Distances between epi-graphs (cont.)

$$\rho\textit{-epi-distance} \;= dl_\rho(f,g) = \sup_{\bar{x}\in\rho\boldsymbol{B}} |d(\bar{x},\mathrm{epi}\ f) - d(\bar{x},\mathrm{epi}\ g)|,\quad \rho \geq 0$$



Distances between epi-graphs (cont.)

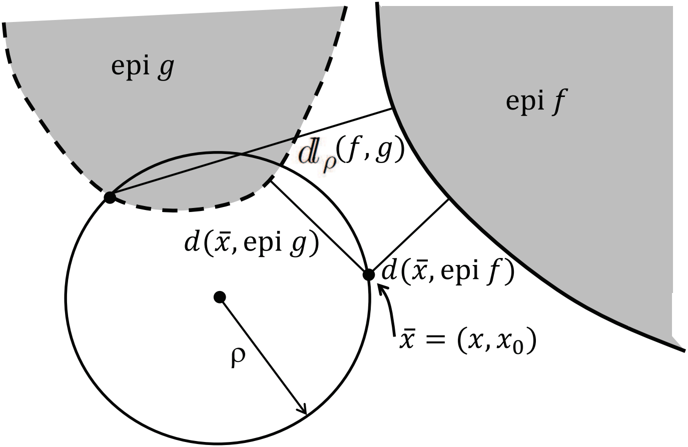
$$\rho\text{-}\textit{epi-distance} = dl_\rho(f,g) = \sup_{\bar{x}\in\rho\boldsymbol{B}} |d(\bar{x},\mathrm{epi}\ f) - d(\bar{x},\mathrm{epi}\ g)|, \quad \rho \geq 0$$



$$dl_\rho(f,g) = \delta$$

Distances between epi-graphs (cont.)

$$\textit{ρ-epi-distance} \ = dl_\rho(f,g) = \sup_{\bar{x}\in\rho\boldsymbol{B}} |d(\bar{x}, \mathrm{epi}\ f) - d(\bar{x}, \mathrm{epi}\ g)|, \quad \rho \geq 0$$



Metric on the lsc functions

ρ-epi-distance dl_ρ is a pseudo-metric on $\mathrm{lsc\text{-}fcns}(\boldsymbol{R}^n)$

$$\text{epi-distance} = dl(f,g) = \int_0^\infty dl_\rho(f,g)e^{-\rho}d\rho$$

dl is a metric on $\mathrm{lsc\text{-}fcns}(\boldsymbol{R}^n)$

The epi-distance induces the epi-topology on $\mathrm{lsc\text{-}fcns}(\boldsymbol{R}^n)$
$(\mathrm{lsc\text{-}fcns}(\boldsymbol{R}^n), dl)$ is a complete separable metric space (Polish)

Epi-convergence

$f^\nu \in \text{lsc-fcns}(\boldsymbol{R}^n)$ epi-converges to f if $dl(f^\nu, f) \to 0$

Actual problem

- Constrained infinite-dimensional optimization problem

$$\begin{aligned} &\min \psi(f) \\ &\text{such that } f \in F \subseteq \mathcal{F} \subseteq \mathrm{lsc\text{-}fcns}(\boldsymbol{R}^n) \end{aligned}$$

- criterion ψ (norms, error measures, etc)
- feasible set F (shape restrictions, external info)
- subspace of interest $\mathcal{F}$ (continuity, smoothness)

Take-aways in this section

- Actual problem:
  - find best extended real-valued lower semicontinuous function
  - captures regression, curve fitting, interpolation, density estimation etc:
  - allows functions on $\boldsymbol{R}^n$, jumps, transformation requiring $\pm\infty$
- Evolving problem due to changes in information and approximations
- Theory about lsc functions:
  - metric space under epi-distance
  - epi-distance $=$ distance between epi-graphs

Questions?

Outline

- Illustrations and formulations (slides 12-33)
- Framework (slides 34-60)
- Epi-splines and approximations (slides 61-77)
- Implementations (slides 78-85)
- Examples (slides 86-124)

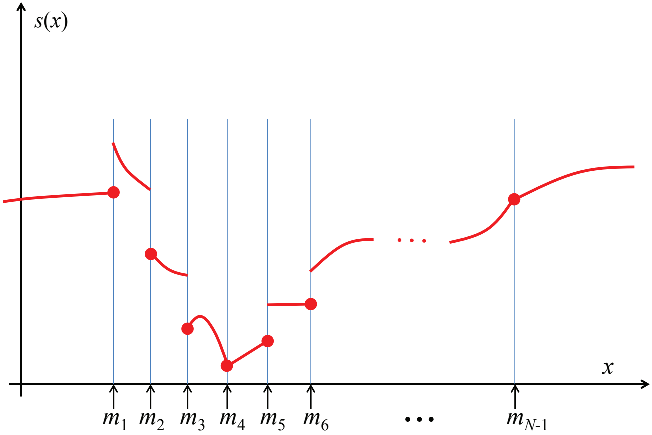
Epi-splines and approximations

Epi-splines = piecewise polynomial + lower semicont.

Piecewise polynomial not automatic; must be constructed

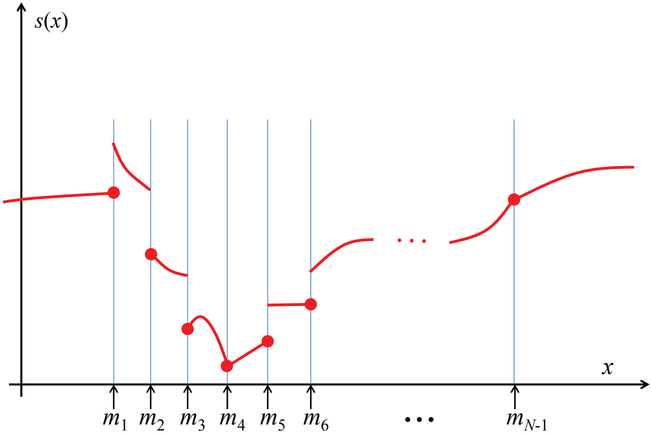
Epi-splines = piecewise polynomial + lower semicont.

Piecewise polynomial not automatic; must be constructed



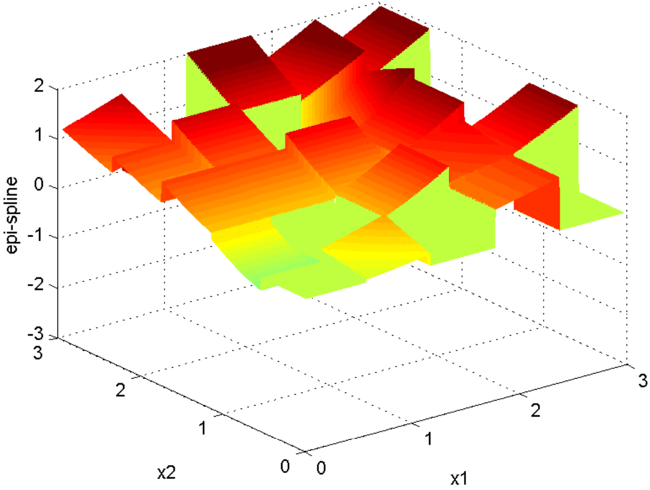
Epi-splines = piecewise polynomial + lower semicont.

Piecewise polynomial not automatic; must be constructed

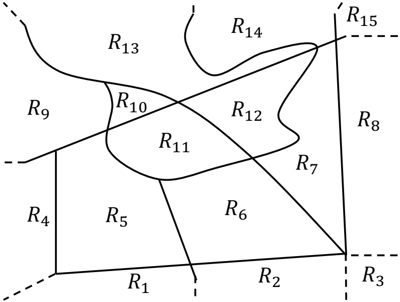


Optimization over new class of piecewise polynomial functions

n-dim epi-splines



n-dim epi-splines (cont.)

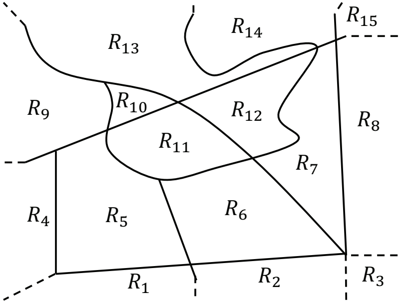


Epi-spline s of order p defined on $\boldsymbol{R}^n$ with partition $\mathcal{R} = \{R_k\}_{k=1}^{N}$ is a real-valued function that

on each R_k, $k = 1, ..., N$, is polynomial of total degree p and

for every $x \in \boldsymbol{R}^n$, has $s(x) = \liminf_{x' \to x} s(x')$

n-dim epi-splines (cont.)



Epi-spline s of order p defined on $\boldsymbol{R}^n$ with partition $\mathcal{R} = \{R_k\}_{k=1}^{N}$ is a real-valued function that

on each R_k, $k = 1, ..., N$, is polynomial of total degree p and

for every $x \in \boldsymbol{R}^n$, has $s(x) = \liminf_{x' \to x} s(x')$

$\text{e-spl}_n^p(\mathcal{R})$ = epi-splines of order p on $\boldsymbol{R}^n$ with partition $\mathcal{R}$

Evolving and approximating problems

$$\min \psi^\nu(s)$$
$$\text{such that } s \in S^\nu = F^\nu \cap \mathcal{S}^\nu$$

Evolving and approximating problems

$$\min \psi^\nu(s)$$
$$\text{such that } s \in S^\nu = F^\nu \cap \mathcal{S}^\nu$$

- subspace of interest $\mathcal{S}^\nu \subseteq \text{e-spl}_n^{p^\nu}(\mathcal{R}^\nu) \cap \mathcal{F}$

Evolving and approximating problems

$$\min \psi^\nu(s)$$
$$\text{such that } s \in S^\nu = F^\nu \cap \mathcal{S}^\nu$$

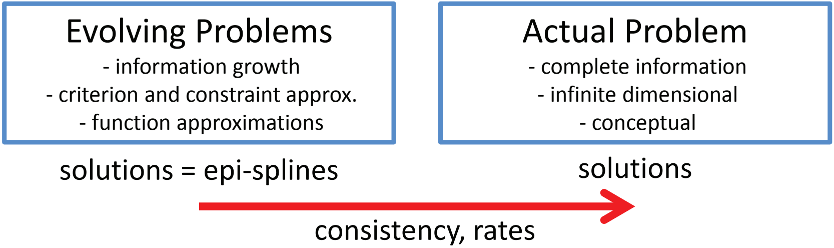
- subspace of interest $\mathcal{S}^\nu \subseteq \text{e-spl}_n^{p^\nu}(\mathcal{R}^\nu) \cap \mathcal{F}$
- evolving/approximate feasible set F^ν

Evolving and approximating problems

$$\min \psi^\nu(s)$$
$$\text{such that } s \in S^\nu = F^\nu \cap \mathcal{S}^\nu$$

- subspace of interest $\mathcal{S}^\nu \subseteq \mathrm{e\text{-}spl}_n^{p^\nu}(\mathcal{R}^\nu) \cap \mathcal{F}$
- evolving/approximate feasible set F^ν
- evolving/approximate criterion ψ^ν

Function identification framework



Convergence of evolving problems

Epi-convergence of functionals on a metric space:

Convergence of evolving problems

Epi-convergence of functionals on a metric space:

Definition: $\{\psi^\nu : S^\nu \to \boldsymbol{R}\}_{\nu=1}^{\infty}$ epi-converge to $\psi : F \to \boldsymbol{R}$ if and only if

(i) for every $s^\nu \to f \in \mathrm{lsc\text{-}fcns}(\boldsymbol{R}^n)$, with $s^\nu \in S^\nu$, we have $\liminf \psi^\nu(s^\nu) \geq \psi(f)$ if $f \in F$ and $\psi^\nu(s^\nu) \to \infty$ otherwise;

(ii) for every $f \in F$, there exists $\{s^\nu\}_{\nu=1}^{\infty}$, with $s^\nu \in S^\nu$, such that $s^\nu \to f$ and $\limsup \psi^\nu(s^\nu) \leq \psi(f)$.

Convergence of evolving problems

Epi-convergence of functionals on a metric space:

Definition: $\{\psi^\nu : S^\nu \to \boldsymbol{R}\}_{\nu=1}^\infty$ epi-converge to $\psi : F \to \boldsymbol{R}$ if and only if

(i) for every $s^\nu \to f \in \mathrm{lsc\text{-}fcns}(\boldsymbol{R}^n)$, with $s^\nu \in S^\nu$, we have $\liminf \psi^\nu(s^\nu) \geq \psi(f)$ if $f \in F$ and $\psi^\nu(s^\nu) \to \infty$ otherwise;

(ii) for every $f \in F$, there exists $\{s^\nu\}_{\nu=1}^\infty$, with $s^\nu \in S^\nu$, such that $s^\nu \to f$ and $\limsup \psi^\nu(s^\nu) \leq \psi(f)$.

Equivalent to previous definition

Convergence of evolving problems

Epi-convergence of functionals on a metric space:

Definition: $\{\psi^\nu : S^\nu \to \boldsymbol{R}\}_{\nu=1}^{\infty}$ epi-converge to $\psi : F \to \boldsymbol{R}$ if and only if

(i) for every $s^\nu \to f \in \mathrm{lsc\text{-}fcns}(\boldsymbol{R}^n)$, with $s^\nu \in S^\nu$, we have $\liminf \psi^\nu(s^\nu) \geq \psi(f)$ if $f \in F$ and $\psi^\nu(s^\nu) \to \infty$ otherwise;

(ii) for every $f \in F$, there exists $\{s^\nu\}_{\nu=1}^{\infty}$, with $s^\nu \in S^\nu$, such that $s^\nu \to f$ and $\limsup \psi^\nu(s^\nu) \leq \psi(f)$.

Equivalent to previous definition

Also say that the evolving optimization problems epi-converge to the actual problem.

Consequence of epi-convergence

If evolving problems epi-converge to the actual problem, then

$$\limsup \left(\inf_{s \in S^\nu} \psi^\nu(s) \right) \leq \inf_{f \in F} \psi(f).$$

Consequence of epi-convergence

If evolving problems epi-converge to the actual problem, then

$$\limsup \left(\inf_{s \in S^\nu} \psi^\nu(s) \right) \leq \inf_{f \in F} \psi(f).$$

Moreover, if s^k are optimal for the evolving problems $\min_{s \in S^{\nu_k}} \psi^{\nu_k}(s)$ and $s^k \to f^0$, then f^0 is optimal for the actual problem and

$$\lim_{k \to \infty} \inf_{s \in S^{\nu_k}} \psi^{\nu_k}(s) = \inf_{f \in F} \psi(f) = \psi(f^0).$$

When will evolving problems epi-converge to actual?

$$\min \psi(f) \quad \text{such that } f \in F \subseteq \mathcal{F} \subseteq \text{lsc-fcns}(\boldsymbol{R}^n)$$

$$\min \psi^\nu(s) \quad \text{such that } s \in S^\nu = F^\nu \cap \mathcal{S}^\nu$$

Theorem: A sufficient condition for epi-convergence is

- ψ^ν converges continuously to ψ relative to $\mathcal{F}$
- Epi-splines dense in lsc functions
- F^ν set-converges to F
- F is solid

Epi-splines dense in lsc functions

$\{\mathcal{R}^\nu\}_{\nu=1}^\infty$, with $\mathcal{R}^\nu = \{R_1^\nu, ..., R_{N^\nu}^\nu\}$, is an infinite refinement if

for every $x \in \boldsymbol{R}^n$ and $\epsilon > 0$, there exists a positive integer $\bar{\nu}$ s.t. $R_k^\nu \subset \boldsymbol{B}(x, \epsilon)$ for every $\nu \geq \bar{\nu}$ and k satisfying $x \in \mathrm{cl}\, R_k^\nu$.

Epi-splines dense in lsc functions

$\{\mathcal{R}^\nu\}_{\nu=1}^\infty$, with $\mathcal{R}^\nu = \{R_1^\nu, ..., R_{N^\nu}^\nu\}$, is an infinite refinement if

for every $x \in \boldsymbol{R}^n$ and $\epsilon > 0$, there exists a positive integer $\bar{\nu}$ s.t.
$R_k^\nu \subset \boldsymbol{B}(x, \epsilon)$ for every $\nu \geq \bar{\nu}$ and k satisfying $x \in \mathrm{cl}\, R_k^\nu$.

Theorem: For any $p = 0, 1, 2, ...,$ and infinite refinement $\{\mathcal{R}^\nu\}_{\nu=1}^\infty$,

$$\bigcup_{\nu=1}^{\infty} \text{e-spl}_n^p(\mathcal{R}^\nu) \text{ is dense in } \text{lsc-fcns}(\boldsymbol{R}^n)$$

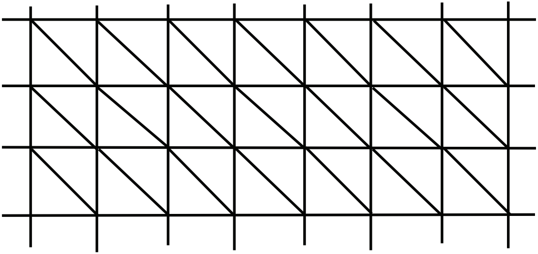
Dense in continuous functions under simplex partitioning

Definition: A simplex S in $\boldsymbol{R}^n$ is the convex hull of $n+1$ points $x^0, x^1, ..., x^n \in \boldsymbol{R}^n$, with $x^1 - x^0$, $x^2 - x^0$, ..., $x^n - x^0$ linearly independent.

Dense in continuous functions under simplex partitioning

Definition: A simplex S in $\boldsymbol{R}^n$ is the convex hull of $n+1$ points $x^0, x^1, ..., x^n \in \boldsymbol{R}^n$, with $x^1 - x^0$, $x^2 - x^0$, ..., $x^n - x^0$ linearly independent.

Definition: A partition $R_1, R_2, ..., R_N$ of $\boldsymbol{R}^n$ is a simplex partition of $\boldsymbol{R}^n$ if $\mathrm{cl}\, R_1$, ..., $\mathrm{cl}\, R_N$ are "mostly" simplexes.



Dense in cnts fcns under simplex partitioning (cont.)

Theorem: For any $p = 1, 2, ..,$ and $\{\mathcal{R}^\nu\}_{\nu=1}^{\infty}$, an infinite refinement of $\boldsymbol{R}^n$ consisting of simplex partitions of $\boldsymbol{R}^n$,

$$\bigcup_{\nu=1}^{\infty} \text{e-spl}_n^p(\mathcal{R}^\nu) \cap \mathcal{C}^0(\boldsymbol{R}^n) \text{ is dense in } \mathcal{C}^0(\boldsymbol{R}^n).$$

Rates in univariate probability density estimation

For convex and finite-dimensional estimation problem:

Theorem: For any "correct" soft information,

$$\nu^{1/2} d_{KL}(h^0 || h_\epsilon^\nu) = O_p(1) \text{ for some } h_\epsilon^\nu = \text{ near-optimal solution}$$

Decomposition

Theorem: For every $s \in \text{e-spl}^p_n(\mathcal{R})$, with $n > p \geq 1$ and $\mathcal{R} = \{R_k\}_{k=1}^N$, there exist $q_{k,i} \in \text{poly}^p(\boldsymbol{R}^{n-1})$, $i = 1, 2, ..., n$ and $k = 1, 2, ..., N$, such that

$$s(x) = \sum_{i=1}^{n} q_{k,i}(x_{-i}), \text{ for all } x \in R_k.$$

an n-dim epi-spline is the sum of n, $(n-1)$-dim epi-splines

Take-aways in this section

- Epi-splines are piecewise polynomials defined on arbitrary partition of $\boldsymbol{R}^n$
- Only lower-semicontinuity required (not smoothness)
- Dense in space of extended real-valued lower-semicontinuous functions under epi-distance (i.e., epi-splines can approximate every lsc function to an arbitrary accuracy)
- Solutions of evolving problem (in terms of epi-splines) are approximate solutions of actual problem

Questions?

Outline

Implementations

Implementation considerations

- Selection of epi-spline order and composition $h = \exp(-s)$
- Selection of partition: go fine!
- Implementation of criterion functional and soft info

Implementation considerations

- Selection of epi-spline order and composition $h = \exp(-s)$
- Selection of partition: go fine!
- Implementation of criterion functional and soft info

Provide details for one-dimensional epi-splines of order 1

Focusing on criteria and constraints for probability densities

Implementation of criteria functionals

Probability densities $h \in \mathrm{e\text{-}spl}^1(m)$, with mesh $m = \{m_0, m_1, ..., m_N\}$,

$-\infty < m_0 < m_1 < ... < m_N < \infty$

Implementation of criteria functionals

Probability densities $h \in \text{e-spl}^1(m)$, with mesh $m = \{m_0, m_1, ..., m_N\}$,

$-\infty < m_0 < m_1 < ... < m_N < \infty$

First-order epi-splines (piecewise linear):
$h(x) = a_0^k + a^k x$ for $x \in (m_{k-1}, m_k)$

Implementation of criteria functionals (cont.)

Log-likelihood for observations $x^1, ..., x^\nu$:

$$\log \prod_{i=1}^{\nu} h(x^i) = \sum_{i=1}^{\nu} \log(a_0^{k_i} + a^{k_i} x^i), \text{ where } k_i \text{ such that } x^i \in (m_{k_i - 1}, m_{k_i})$$

Implementation of criteria functionals (cont.)

Log-likelihood for observations $x^1, ..., x^\nu$:

$$\log \prod_{i=1}^{\nu} h(x^i) = \sum_{i=1}^{\nu} \log(a_0^{k_i} + a^{k_i} x^i), \text{ where } k_i \text{ such that } x^i \in (m_{k_i - 1}, m_{k_i})$$

Entropy:

$$-\int h(x) \log h(x) dx = -\sum_{k=1}^{N} \int_{m_{k-1}}^{m_k} (a_0^k + a^k x) \log(a_0^k + a^k x)$$

Implementation of criteria functionals (cont.)

Log-likelihood for observations $x^1, ..., x^\nu$:

$$\log \prod_{i=1}^{\nu} h(x^i) = \sum_{i=1}^{\nu} \log(a_0^{k_i} + a^{k_i} x^i), \text{ where } k_i \text{ such that } x^i \in (m_{k_i - 1}, m_{k_i})$$

Entropy:

$$-\int h(x) \log h(x) dx = -\sum_{k=1}^{N} \int_{m_{k-1}}^{m_k} (a_0^k + a^k x) \log(a_0^k + a^k x)$$

Both concave in epi-parameters a_0^k and a^k, $k = 1, ..., N$

Soft information

Nonnegativity: $h \geq 0$ if $a_0^k + a^k m_{k-1} \geq 0$ and $a_0^k + a^k m_k \geq 0$, $k = 1, ..., N$

Soft information

Nonnegativity: $h \geq 0$ if $a_0^k + a^k m_{k-1} \geq 0$ and $a_0^k + a^k m_k \geq 0$, $k = 1, ..., N$

Integrate to one:

$$\int h(x)dx = \sum_{k=1}^{N} a_0^k (m_k - m_{k-1}) + \sum_{k=1}^{N} \frac{a^k}{2} (m_k^2 - m_{k-1}^2) = 1$$

Soft information

Nonnegativity: $h \geq 0$ if $a_0^k + a^k m_{k-1} \geq 0$ and $a_0^k + a^k m_k \geq 0$, $k = 1, ..., N$

Integrate to one:

$$\int h(x)dx = \sum_{k=1}^{N} a_0^k (m_k - m_{k-1}) + \sum_{k=1}^{N} \frac{a^k}{2} (m_k^2 - m_{k-1}^2) = 1$$

Continuity: $a_0^k + a^k m_k = a_0^{k+1} + a^{k+1} m_k$ for $k = 1, ..., N-1$.

Soft information

Nonnegativity: $h \geq 0$ if $a_0^k + a^k m_{k-1} \geq 0$ and $a_0^k + a^k m_k \geq 0$, $k = 1, ..., N$

Integrate to one:

$$\int h(x)dx = \sum_{k=1}^{N} a_0^k (m_k - m_{k-1}) + \sum_{k=1}^{N} \frac{a^k}{2}(m_k^2 - m_{k-1}^2) = 1$$

Continuity: $a_0^k + a^k m_k = a_0^{k+1} + a^{k+1} m_k$ for $k = 1, ..., N-1$.

Log-concavity, convexity, monotonicity

Soft information (cont.)

Deconvolution $Y = X + W$: Given densities h_W and h_Y,

$$h_Y(y) = \int h(x) h_W(y-x)dx = \sum_{k=1}^{N} a_0^k \int_{m_{k-1}}^{m_k} h_W(y-x)dx + \sum_{k=1}^{N} a^k \int_{m_{k-1}}^{m_k} x h_W(y-x)dx \text{ for all } y$$

Soft information (cont.)

Deconvolution $Y = X + W$: Given densities h_W and h_Y,

$$h_Y(y) = \int h(x) h_W(y-x)dx = \sum_{k=1}^{N} a_0^k \int_{m_{k-1}}^{m_k} h_W(y-x)dx$$
$$+ \sum_{k=1}^{N} a^k \int_{m_{k-1}}^{m_k} x h_W(y-x)dx \text{ for all } y$$

Inverse problem $Y = g(X)$: Given $v_q = \int y^q h_Y(y)dy$, $q = 1, 2, \ldots$

$$v_q = \int [g(x)]^q h(x)dx \approx \sum_{k=1}^{N} \sum_{j=1}^{M^k} w^{jk} [g(x^{jk})]^q (a_0^k + a^k x^{jk})$$

Soft information (cont.)

Deconvolution $Y = X + W$: Given densities h_W and h_Y,

$$h_Y(y) = \int h(x) h_W(y-x)dx = \sum_{k=1}^{N} a_0^k \int_{m_{k-1}}^{m_k} h_W(y-x)dx + \sum_{k=1}^{N} a^k \int_{m_{k-1}}^{m_k} x h_W(y-x)dx \text{ for all } y$$

Inverse problem $Y = g(X)$: Given $v_q = \int y^q h_Y(y)dy$, $q = 1, 2, ...$

$$v_q = \int [g(x)]^q h(x)dx \approx \sum_{k=1}^{N} \sum_{j=1}^{M^k} w^{jk} [g(x^{jk})]^q (a_0^k + a^k x^{jk})$$

Convex optimization problems

Software, references

- Papers and tutorials: http://faculty.nps.edu/joroyset
- Software for univariate probability density estimation
  - Matlab toolbox http://faculty.nps.edu/joroyset/XSPL.html
  - R toolbox (S. Buttrey) http://faculty.nps.edu/sebuttre/home/Software/expepi/index.html

Outline

- Illustrations and formulations (slides 12-33)
- Framework (slides 34-60)
- Epi-splines and approximations (slides 61-77)
- Implementations (slides 78-85)
- Examples (slides 86-124)

Examples: forecasting and fitting

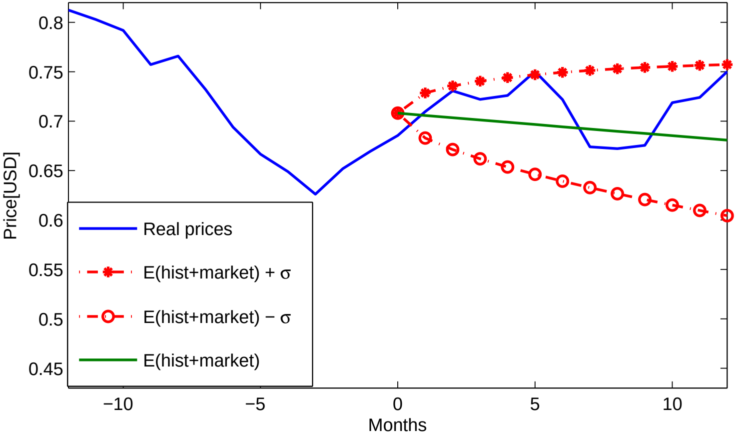
Copper price forecast

Stochastic differential equation: estimate drift, volatility

Steps:

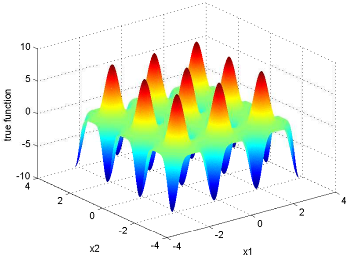
- identify discount factor curve using epi-splines and futures market information $\Longrightarrow$ future spot prices
- identify drift using epi-splines, historical data, future spot prices, etc
- identify volatility using epi-splines and observed/estimated errors

Copper price forecast (cont.)



Surface reconstruction

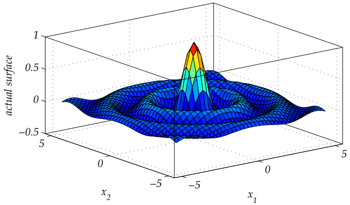
$f(x) = (\cos(\pi x_1) + \cos(\pi x_2))^3$ on $[-3,3]^2$
continuity; second-order epi-splines; $N = 400$; 900 uniform points

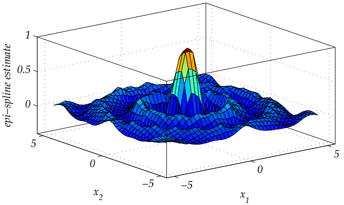


Actual function and epi-spline approximation

Surface reconstruction

$f(x) = \sin(\pi\|x\|)/(\pi\|x\|)$ for $x \in [-5,5]^2 \setminus \{0\}$, $f(0) = 1$
cont. diff.; second-order epi-splines; $N = 225$; 600 random points





Actual function and epi-spline approximation

Examples: density estimation

Examples of probability density estimation

Find a density h that maximizes log-likelihood of observations and satisfies soft information

Examples: earthquake losses, queuing, robustness, deconvolution, UQ, mixture, bivariate densities

Earthquake losses

Earthquake losses, Vancouver Region*

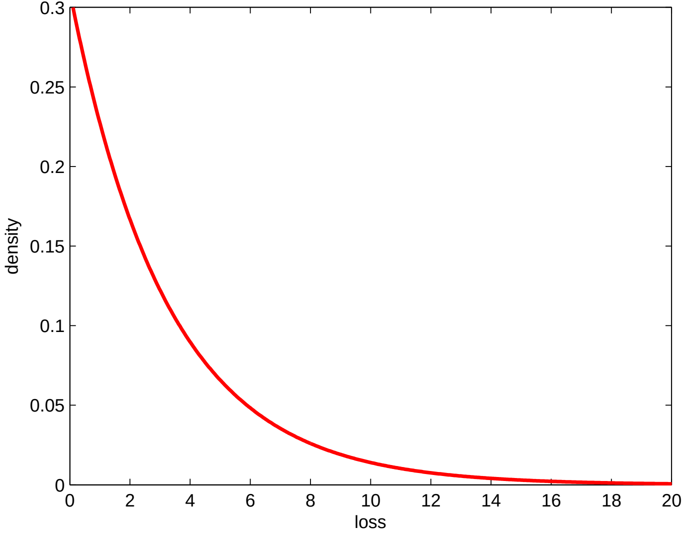
Comprehensive damage model (284 random variables)

100,000 simulations of 50-year loss

*Data from Mahsuli, 2012; see also Mahsuli & Haukaas, 2013

Earthquake losses (cont.)

Probability density of 50-year loss (billion CAD)

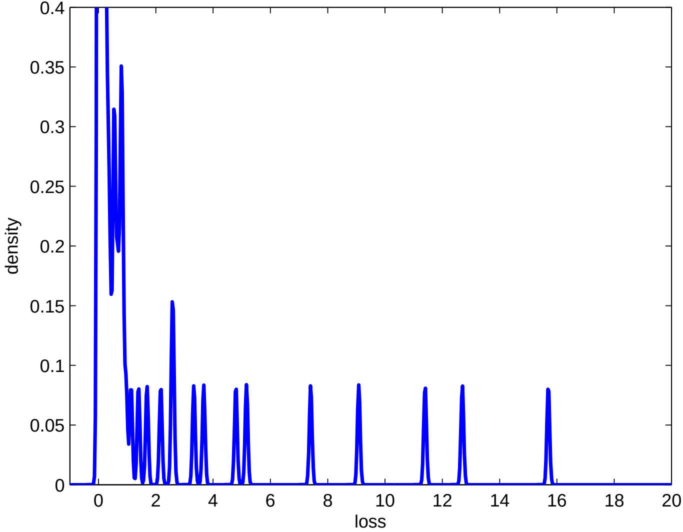


Use fewer simulations?

Using 100 simulations only and kernel estimator

Use fewer simulations?

Using 100 simulations only and kernel estimator



Using function identification approach

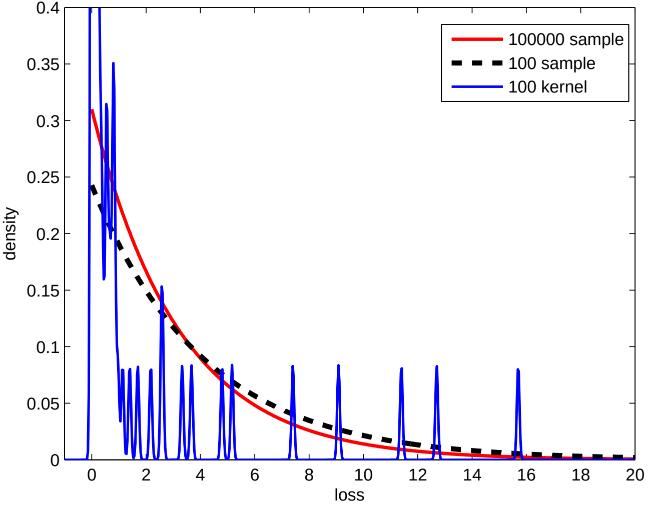
Max likelihood of 100 simulations; second-order exp. epi-splines

Eng. knowledge: nonincreasing, smooth, nonnegative support

Using function identification approach

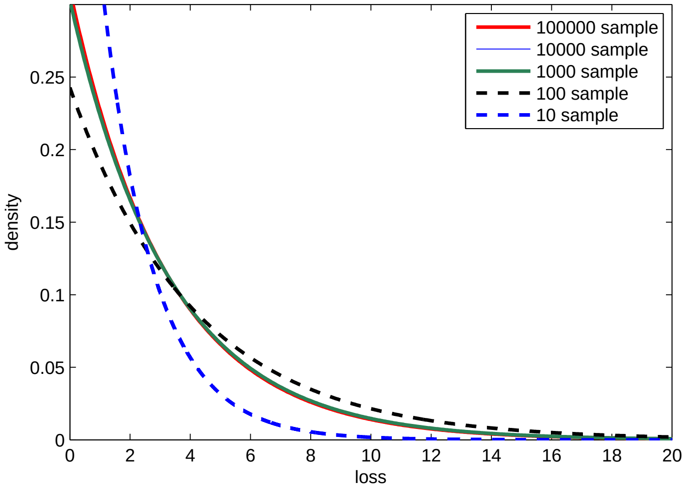
Max likelihood of 100 simulations; second-order exp. epi-splines

Eng. knowledge: nonincreasing, smooth, nonnegative support



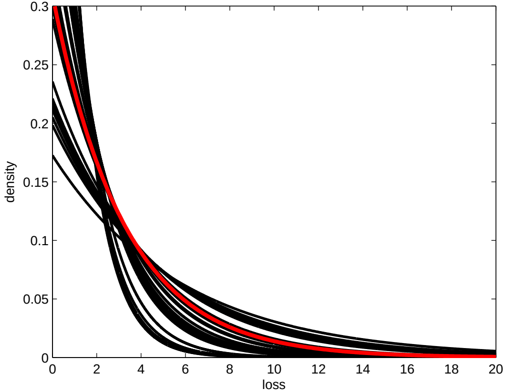
Using function identification approach (cont.)

Varying number of simulations; same soft information



Using function identification approach (cont.)

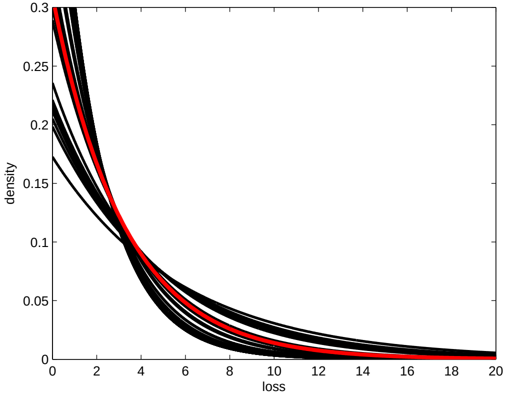
30 meta-replications of 100 simulations



	100,000 sim.	100 sim.
mean	3.2	3.1($\pm$1.3)
average 10% highest	10.6	10.3($\pm$4.4)

Using function identification approach (cont.)

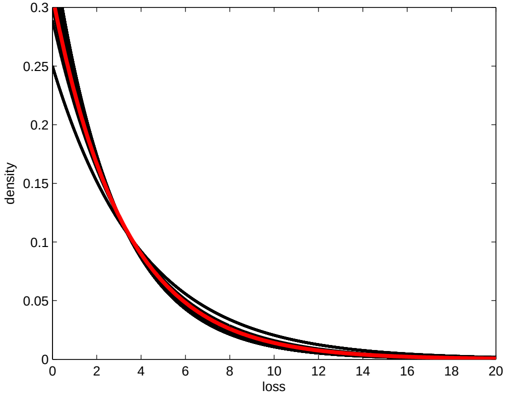
Also pointwise Fisher info.: $h'(y)/h(y) \in [-0.5, -0.1]$



	100,000 sim.	100 sim.
mean	3.2	3.2($\pm$1.2)
average 10% highest	10.6	10.6($\pm$4.0)

Using function identification approach (cont.)

Also pointwise Fisher info.: $h'(x)/h(x) \in [-0.35, -0.25]$



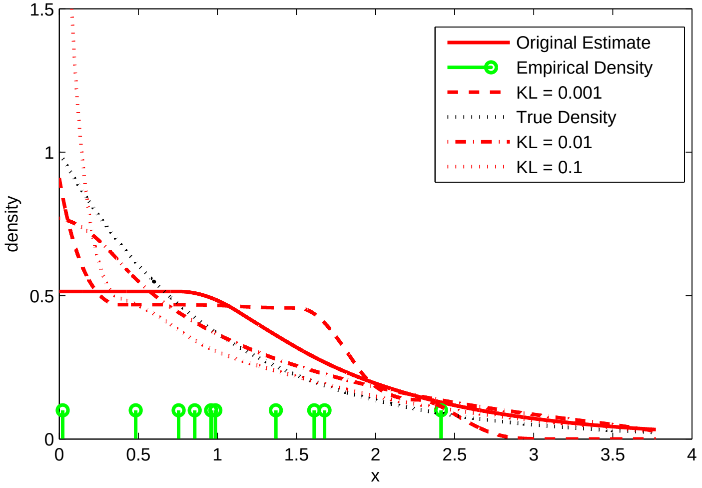
	100,000 sim.	100 sim.
mean	3.2	$3.3(\pm 0.5)$
average 10% highest	10.6	$10.9(\pm 1.7)$

Building robustness

Diversity of estimates using KL-divergence

Returning to exponential density

Continuously diff., nonincreasing, nonnegative support

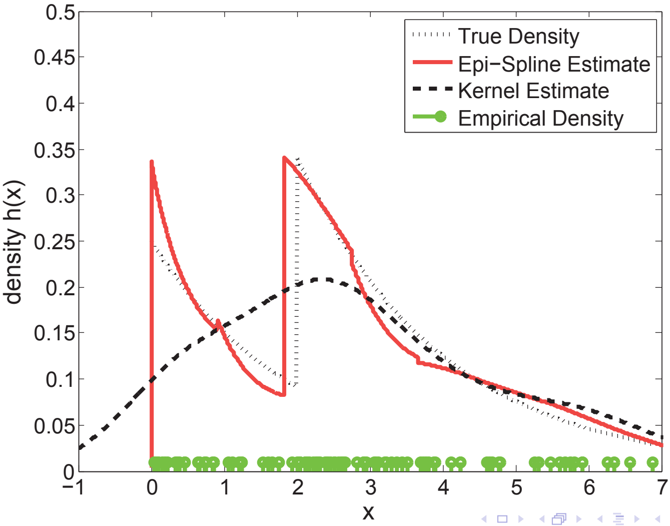


Queuing

M/M/1; 50% of customers delayed for fixed time

X = customer time-in-service; 100 obs.; exp. epi-spline

Soft info: lsc, $X \geq 0$, pointwise Fisher, unimodal upper tail



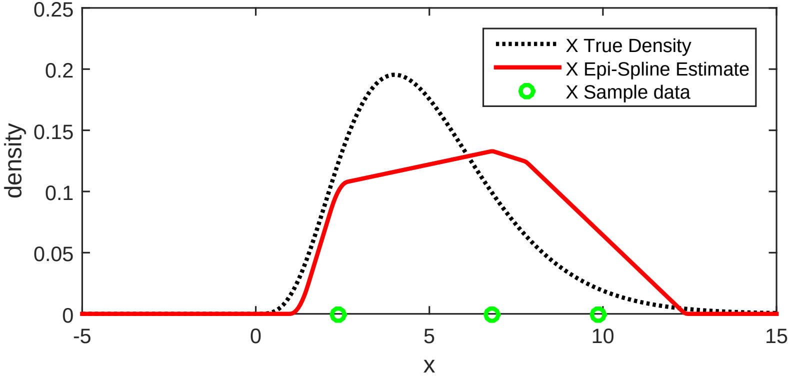
Deconvolution

True density Gamma(5,1)

First-order epi-spline on $[0, 23]$, $N = 1000$

Three sample points

Contin., unimodal, convex tails, bounds on gradient jumps



Also noisy observations

5000 observations of $Y = X + W$

W independent normal noise; mean 0, stdev 3.2

Also noisy observations

5000 observations of $Y = X + W$

W independent normal noise; mean 0, stdev 3.2

Estimate h_Y separately

Also noisy observations

5000 observations of $Y = X + W$

W independent normal noise; mean 0, stdev 3.2

Estimate h_Y separately

$|h_Y(y) - \int h(x) h_W(y-x) dx| \leq 0.005$ for 101 y points

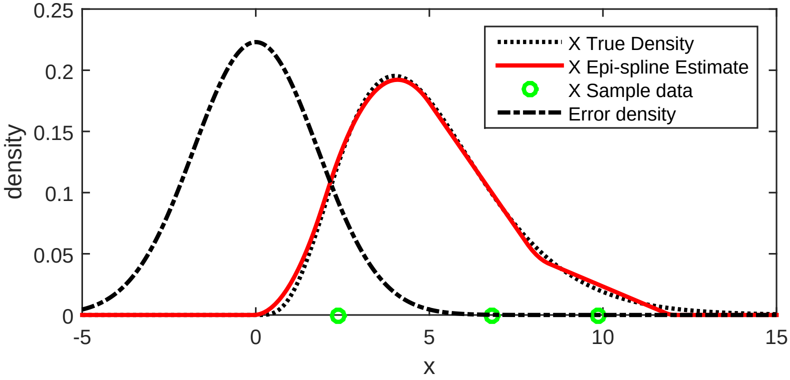
Also noisy observations

5000 observations of $Y = X + W$

W independent normal noise; mean 0, stdev 3.2

Estimate h_Y separately

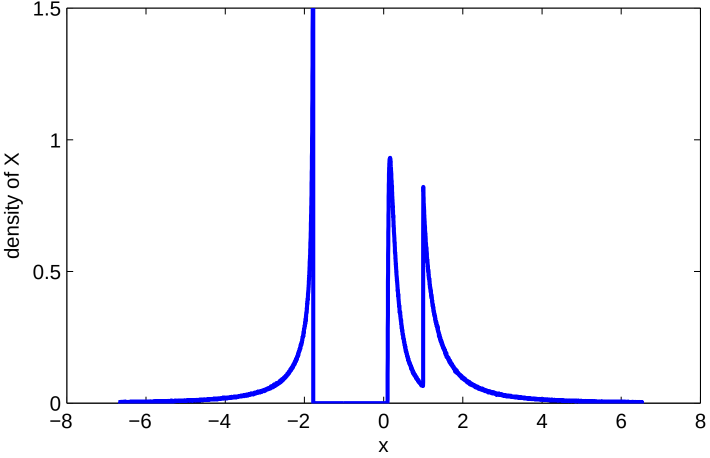
$|h_Y(y) - \int h(x)h_W(y-x)dx| \leq 0.005$ for 101 y points



Uncertainty quantification

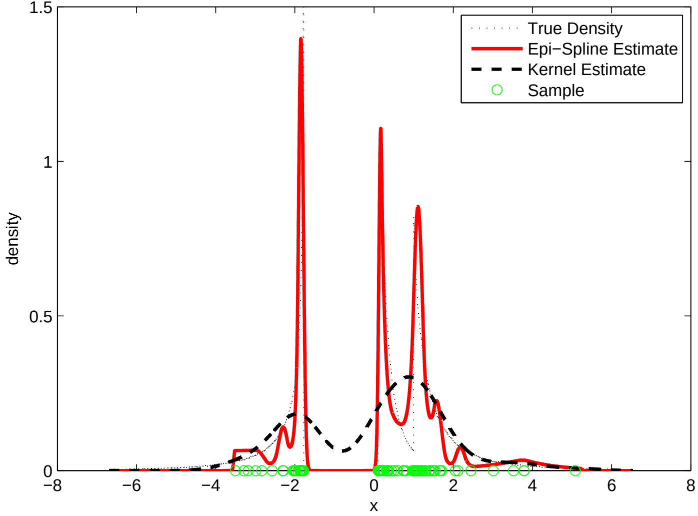
Dynamical system

Recall the true density of the amplitude



Density of amplitude

Sample size 100; cont. diff.; unimodal tails, exp. epi-splines



Gradient information

Gradient information for bijective $g : \boldsymbol{R} \to \boldsymbol{R}$
Recall: If $X = g(V)$, then

$$h(x) = h_V(g^{-1}(x))/|g'(g^{-1}(x))|$$

Gradient information

Gradient information for bijective $g : \boldsymbol{R} \to \boldsymbol{R}$
Recall: If $X = g(V)$, then

$$h(x) = h_V(g^{-1}(x))/|g'(g^{-1}(x))|$$

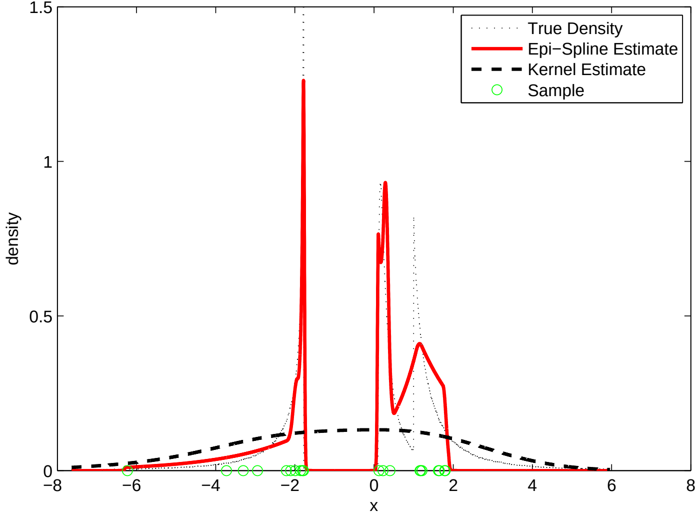
Present context *without* a bijection and data $x^i = g(v^i), g'(v^i)$:

$$h(x^i) \geq \frac{h_V(v^i)}{|g'(v^i)|}$$

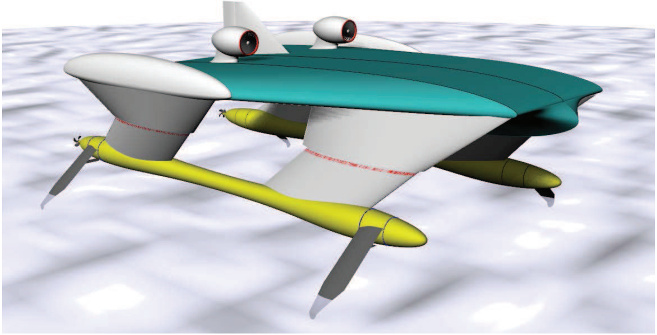
Value of pdf bounded from below at x^i

Gradient information (cont.)

Sample size 20



Fluid dynamics



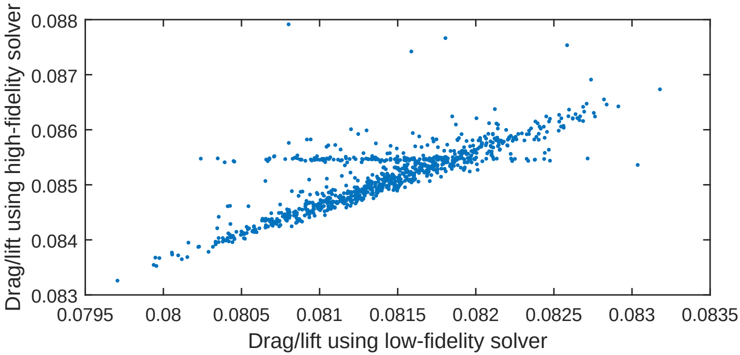
Drag/lift estimates:

High-fidelity RANSE solves (each 4 hours on 8 cores)

Low-fidelity potential flow solves (each 5 sec on 1 core)

Predicting high-fidelity performance

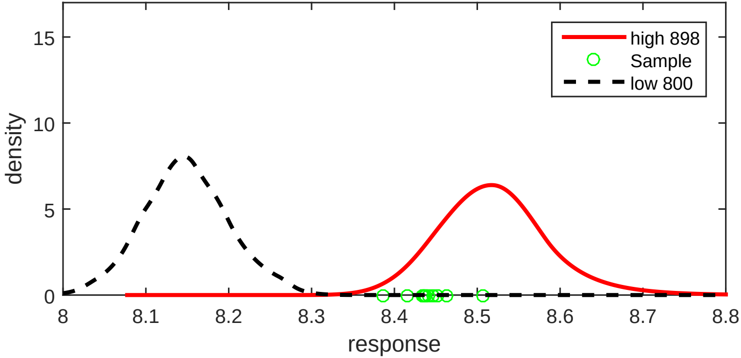
898 high- and low-fidelity solves



Learn from low-fidelity solves and avoid (many) high-fidelity solves

Density using high or low solves

Exponential epi-splines of second order; mesh $N = 50$
Soft info: log-concavity and bounds on second-order derivatives



Estimating conditional error

X =high-fidelity; Y =low-fidelity

$h_X(x) = \int h_{X|Y}(x|y)h_Y(y)dy$

Estimating conditional error

X =high-fidelity; Y =low-fidelity

$h_X(x) = \int h_{X|Y}(x|y)h_Y(y)dy$

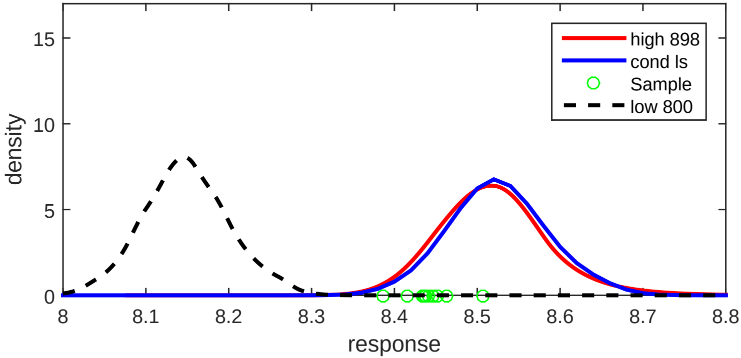
Normal linear least-squares regression model on training data of size 50 $\rightarrow$ $h_{X|Y}(x|y)$

Estimating conditional error

X =high-fidelity; Y =low-fidelity

$h_X(x) = \int h_{X|Y}(x|y) h_Y(y) dy$

Normal linear least-squares regression model on training data of size 50 $\rightarrow \quad h_{X|Y}(x|y)$



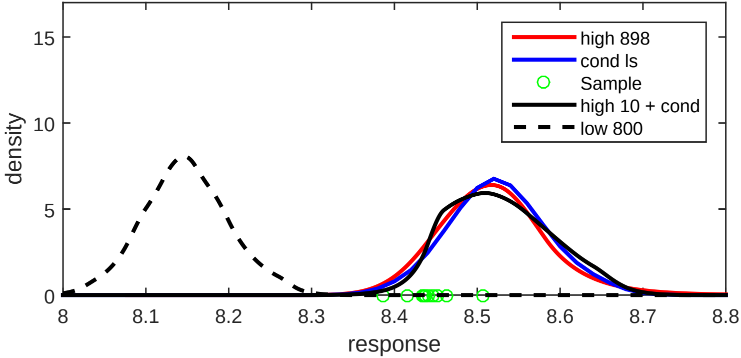
Information fusion

Max likelihood using 10 high-fidelity simulations

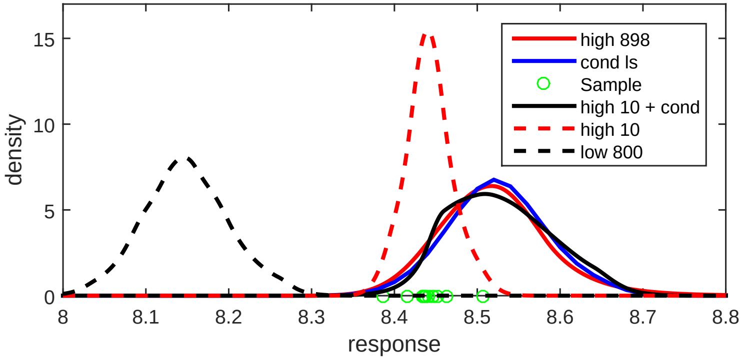
$0.5 h_X(x) \le \int h_{X|Y}(x|y) h_Y(y) dy \le 1.5 h_X(x)$

Information fusion

Max likelihood using 10 high-fidelity simulations
$0.5 h_X(x) \leq \int h_{X|Y}(x|y) h_Y(y) dy \leq 1.5 h_X(x)$



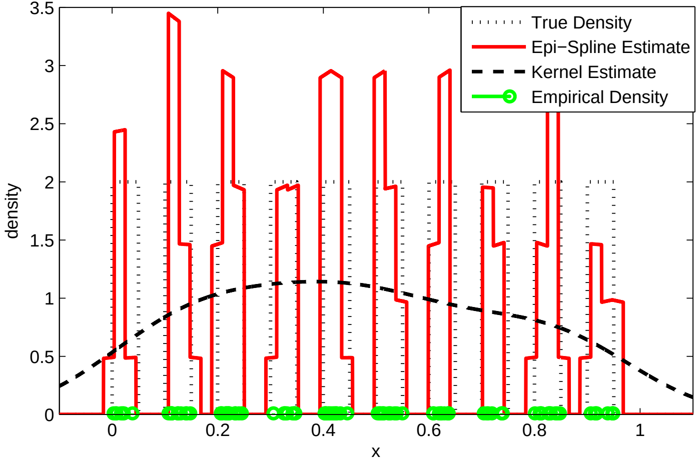
Only 10 high-fidelity



Uniform mixture density

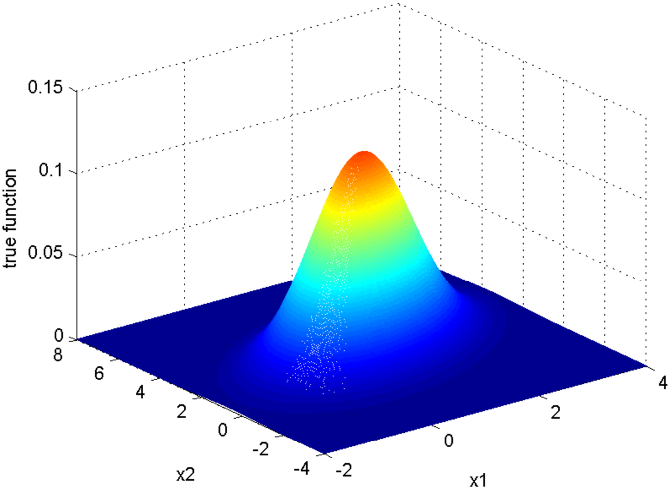
Uniform mixture density

Sample size 100; lsc, slope constraints; exp. epi-splines

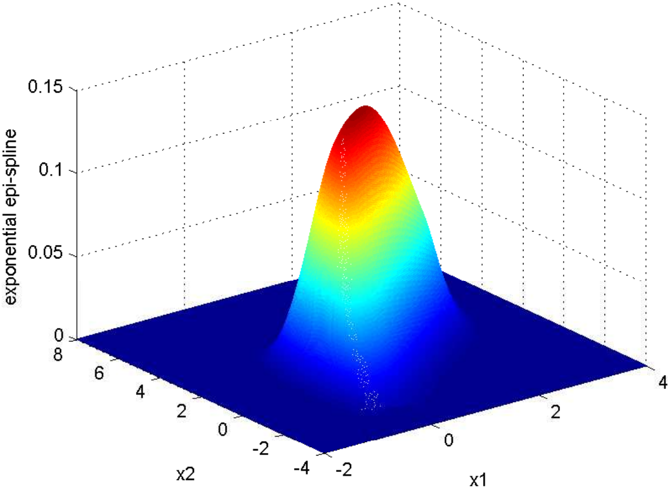


Bivariate normal density

Bivariate normal probability density



Bivariate normal probability density



Curvature, log-concave, 25 sample points, exp. epi-spline

Conclusions

- Function identification problems: rich class

Conclusions

- Function identification problems: rich class
- lsc functions provide modeling flexibility

Conclusions

- Function identification problems: rich class
- lsc functions provide modeling flexibility
- Epi-convergence allows evolution of info./approx.

Conclusions

- Function identification problems: rich class
- lsc functions provide modeling flexibility
- Epi-convergence allows evolution of info./approx.
- Epi-splines central approximation tools

References

- Singham, Royset, & Wets, Density estimation of simulation output using exponential epi-splines
- Royset, Sukumar, & Wets, Uncertainty quantification using exponential epi-splines
- Royset & Wets, From data to assessments and decisions: epi-spline technology
- Royset & Wets, Fusion of hard and soft information in nonparametric density estimation
- Royset & Wets, Multivariate epi-splines and evolving function identification problems
- Feng, Rios, Ryan, Spurkel, Watson, Wets & Woodruff, Toward scalable stochastic unit commitment - Part 1
- Rios, Wets & Woodruff, Multi-period forecasting with limited information and scenario generation with limited data
- Royset & Wets, On function identification problems