Least-squares estimation of a convex function

Amélie Fils-Villetard[1], Armelle Guillou[1], Johan Segers[2]

(1) Paris 6 University, France

(2) Tilburg University, the Netherlands

Main ideas

- Aim: Estimation of a convex function defined on a compact interval

Main ideas

- Aim: Estimation of a convex function defined on a compact interval
- Applications: convex regression function, convex density function, Pickands dependence function,...

Main ideas

- Aim: Estimation of a convex function defined on a compact interval
- Applications: convex regression function, convex density function, Pickands dependence function,...
- Tools: Least-squares method, projection theory

Main ideas

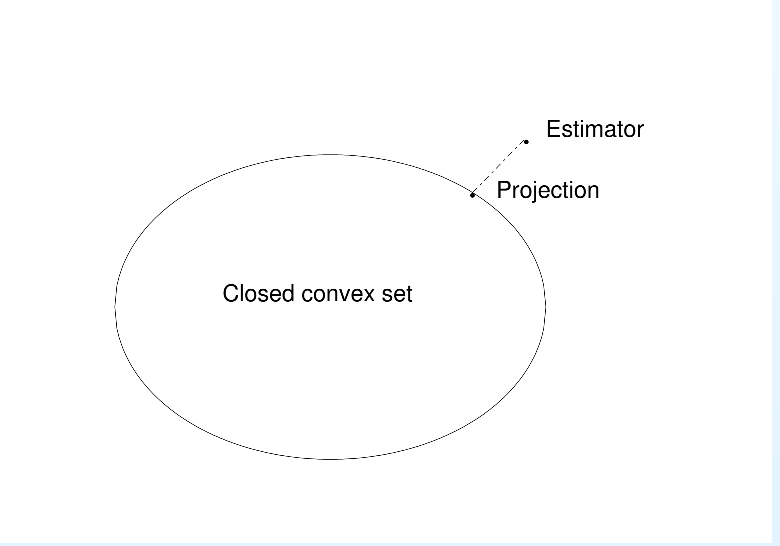
- Aim: Estimation of a convex function defined on a compact interval
- Applications: convex regression function, convex density function, Pickands dependence function,...
- Tools: Least-squares method, projection theory

Our approach:

$\mathcal{F} = \{$convex functions on $[0,1]\}$

Let $f \in \mathcal{F}$.

Given any estimator f_n of f,
consider the projection of f_n on $\mathcal{F}$.



Contents

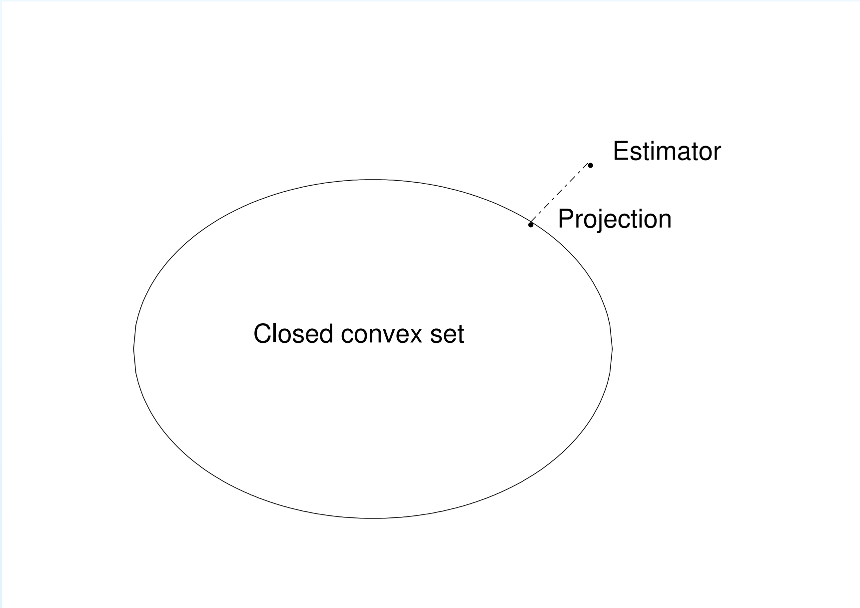
1. Introduction to our method (1)

- $\mathbb{H} = L^2([0,1], d\omega)$

  $\mathcal{F} = \{$all convex functions on $[0,1]\}$

  f_n: any estimator of $f \in \mathcal{F}$

$$\widehat{f}_n^{ls} = \Pi(f_n|\mathcal{F}) = \arg\min_{f\in\mathcal{F}} \int_0^1 \{f_n(\omega) - f(\omega)\}^2 \, d\omega.$$



1. Introduction to our method (2)

- Advantages:
  - Admissible estimator: $\widehat{f}_n^{ls} \in \mathcal{F}$
  - For $f \in \mathcal{F}$:

$$\int_0^1 \{\widehat{f}_n^{ls}(\omega) - f(\omega)\}^2 d\omega \leq \int_0^1 \{f_n(\omega) - f(\omega)\}^2 d\omega$$

  - no smoothing parameter to choose

2. Results on the behavior of our new estimator (1)

$\mathbb{H} = L^2([0,1], d\omega)$.

$\mathcal{F} = \{$all convex functions on $[0,1]\}$

f_n: an estimator of $f \in \mathcal{F}$.

Case 1: Rate of convergence

We only know that, for some positive, bounded sequence λ_n

$$\|f_n - f\|_\infty = \mathcal{O}_P(\lambda_n),$$

Since $\|\hat{f}_n^{ls} - f\|_2 \leq \|f_n - f\|_2 \leq \|f_n - f\|_\infty$, we deduce that

$$\|\hat{f}_n^{ls} - f\|_2 = \mathcal{O}_P(\lambda_n).$$

Examples: convex density or regression functions.

2. Results on the behavior of our new estimator (2)

Case 2: Asymptotic behavior

Let $t_n \to 0$ and $f_n \in \mathbb{H}$ be any estimator of f such that

$$\frac{f_n - f}{t_n} \xrightarrow{d} G \qquad \text{in } \mathbb{H}.$$

Asymptotic behavior of $\dfrac{\Pi(f_n|\mathcal{F}) - f}{t_n}$?

Tangent cone of $\mathcal{F}$ at f

$C_f = \{g \in \mathbb{H} : \|g - \lambda_n(B_n - f)\|_2 \to 0$ for some $\lambda_n \geq 0$ and $B_n \in \mathcal{F}\}$.

C_f is a closed and convex cone.

2. Results on the behavior of our new estimator (2)

- $$\frac{\Pi(f_n|\mathcal{F}) - f}{t_n} = \frac{1}{t_n}\left\{\Pi\left(f + t_n\left(\frac{f_n - f}{t_n}\right)\Big|\mathcal{F}\right) - f\right\}$$

2. Results on the behavior of our new estimator (2)

- $$\frac{\Pi(f_n|\mathcal{F}) - f}{t_n} = \frac{1}{t_n}\left\{\Pi\left(f + t_n\left(\frac{f_n - f}{t_n}\right)\middle|\mathcal{F}\right) - f\right\}$$

- If

$$\begin{array}{rcl} t_n & \longrightarrow & 0 \\ G_n & \stackrel{d}{\longrightarrow} & G \end{array}$$

then $$\frac{\Pi(f + t_n G_n|\mathcal{F}) - f}{t_n} \xrightarrow{d} \Pi(G|C_f)$$

2. Results on the behavior of our new estimator (2)

- $$\frac{\Pi(f_n|\mathcal{F}) - f}{t_n} = \frac{1}{t_n}\left\{\Pi\left(f + t_n\left(\frac{f_n - f}{t_n}\right)\Big|\mathcal{F}\right) - f\right\}$$
- If

$$\begin{array}{ccc} t_n & \longrightarrow & 0 \\ G_n & \xrightarrow{d} & G \end{array}$$

then
$$\frac{\Pi(f + t_n G_n|\mathcal{F}) - f}{t_n} \xrightarrow{d} \Pi(G|C_f)$$

- Conclusion: $$\frac{\Pi(f_n|\mathcal{F}) - f}{t_n} \xrightarrow{d} \Pi(G|C_f) \qquad \text{in } \mathbb{H}.$$

3. Practical study (1)

- Problem: $\Pi(f_n|\mathcal{F})$ can typically not be computed explicitely.

3. Practical study (1)

- Problem: $\Pi(f_n|\mathcal{F})$ can typically not be computed explicitely.
- Solution: Introduction of a subclass $\mathcal{F}_m \subset \mathcal{F}$.

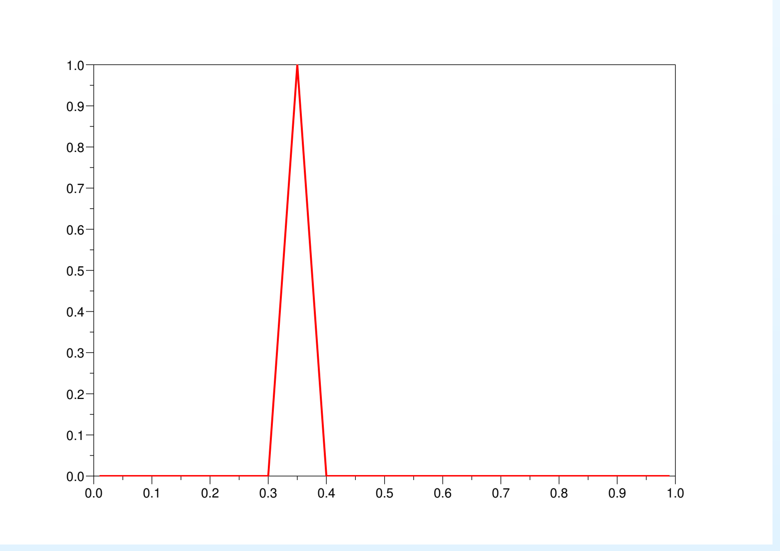
3. Practical study (1)

- Problem: $\Pi(f_n|\mathcal{F})$ can typically not be computed explicitely.
- Solution: Introduction of a subclass $\mathcal{F}_m \subset \mathcal{F}$.
- $\mathcal{F}_m = \{f \in \mathcal{F}$ piecewise linear with knots on $0, 1/m, \ldots, 1\}$

$$\mathcal{F}_m = \left\{ f = \sum_{k=0}^{m} \lambda_k f_{km} : (\lambda_0, ..., \lambda_m) \in \Lambda_m \right\},$$

where $f_{0m}, ..., f_{mm} \in \mathbb{H}$ and $\Lambda_m \subset \mathbb{R}^{m+1}$.

Function f_{km}:



3. Practical study (2)

- Problem: Find $\quad \arg\min_{h \in \mathcal{F}_m} \int_0^1 \{B(\omega) - h(\omega)\}^2 d\omega.$

  Let $h \in \mathcal{F}_m$: $\qquad h = \sum_{k=0}^{m} \lambda_k f_{km}.$

  So, we have to minimize the quadratic function
  $$\mathbf{\Lambda}^T \int_0^1 \mathbf{f}(\omega)\mathbf{f}^T(\omega) d\omega\, \mathbf{\Lambda} - 2\mathbf{\Lambda}^T \int_0^1 \mathbf{f}(\omega) B(\omega) d\omega,$$
  in $\mathbf{\Lambda} = (\lambda_0, \ldots, \lambda_m)^T \in \mathbb{R}^{m+1}$, where $\mathbf{f} = (f_{0m}, \ldots, f_{mm})^T$, under linear equality and inequality constraints.

3. Practical study (2)

- Problem: Find $\displaystyle \arg\min_{h\in\mathcal{F}_m} \int_0^1 \{B(\omega)-h(\omega)\}^2 d\omega.$

  Let $h \in \mathcal{F}_m$: $\displaystyle h = \sum_{k=0}^{m} \lambda_k f_{km}.$

  So, we have to minimize the quadratic function

  $$\mathbf{\Lambda}^T \int_0^1 \mathbf{f}(\omega)\mathbf{f}^T(\omega)d\omega\, \mathbf{\Lambda} - 2\mathbf{\Lambda}^T \int_0^1 \mathbf{f}(\omega)B(\omega)d\omega,$$

  in $\mathbf{\Lambda} = (\lambda_0, \dots, \lambda_m)^T \in \mathbb{R}^{m+1}$, where $\mathbf{f} = (f_{0m}, \dots, f_{mm})^T$, under linear equality and inequality constraints.

- Final estimator: $\displaystyle \widehat{f}_n^{ls} = \sum_{k=0}^{m} \hat{\lambda}_k f_{km}$

  with $\hat{\Lambda} = (\hat{\lambda}_0, ..., \hat{\lambda}_m)$ the solution of the quadratic program.

3. Practical study (3)

- Asymptotic behavior of $\dfrac{\Pi(f_n|\mathcal{F}_m) - f}{t_n}$?

$$\frac{\Pi(f_n|\mathcal{F}_m) - f}{t_n} = \frac{\Pi(f_n|\mathcal{F}_m) - \Pi(f_n|\mathcal{F})}{t_n} + \frac{\Pi(f_n|\mathcal{F}) - f}{t_n}$$

3. Practical study (3)

- Asymptotic behavior of $\dfrac{\Pi(f_n|\mathcal{F}_m) - f}{t_n}$?

$$\frac{\Pi(f_n|\mathcal{F}_m) - f}{t_n} = \frac{\Pi(f_n|\mathcal{F}_m) - \Pi(f_n|\mathcal{F})}{t_n} + \frac{\Pi(f_n|\mathcal{F}) - f}{t_n}$$

- Define the Haussdorff distance between $\mathcal{F}$ and $\mathcal{F}_m$:

$$\delta_m = \sup_{f \in \mathcal{F}} \inf_{f_m \in \mathcal{F}_m} \|f - f_m\|_2.$$

$$\Big\|\Pi(r|\mathcal{F}_m) - \Pi(r|\mathcal{F})\Big\|_2 \leq \Big[\delta_m\Big\{2\Big\|r - \Pi(r|\mathcal{F})\Big\|_2 + \delta_m\Big\}\Big]^{1/2}.$$

3. Practical study (3)

- Asymptotic behavior of $\dfrac{\Pi(f_n|\mathcal{F}_m) - f}{t_n}$?

$$\frac{\Pi(f_n|\mathcal{F}_m) - f}{t_n} = \frac{\Pi(f_n|\mathcal{F}_m) - \Pi(f_n|\mathcal{F})}{t_n} + \frac{\Pi(f_n|\mathcal{F}) - f}{t_n}$$

- Define the Haussdorff distance between $\mathcal{F}$ and $\mathcal{F}_m$:

$$\delta_m = \sup_{f \in \mathcal{F}} \inf_{f_m \in \mathcal{F}_m} \|f - f_m\|_2.$$

$$\Big\|\Pi(r|\mathcal{F}_m) - \Pi(r|\mathcal{F})\Big\|_2 \leq \Big[\delta_m \Big\{2\Big\|r - \Pi(r|\mathcal{F})\Big\|_2 + \delta_m\Big\}\Big]^{1/2}.$$

- If $\delta_m = o(t_n)$, then $\|\Pi(f_n|\mathcal{F}_m) - \Pi(f_n|\mathcal{F})\|_2 = o_p(t_n)$.

3. Practical study (3)

- Asymptotic behavior of $\dfrac{\Pi(f_n|\mathcal{F}_m) - f}{t_n}$?

$$\frac{\Pi(f_n|\mathcal{F}_m) - f}{t_n} = \frac{\Pi(f_n|\mathcal{F}_m) - \Pi(f_n|\mathcal{F})}{t_n} + \frac{\Pi(f_n|\mathcal{F}) - f}{t_n}$$

- Define the Haussdorff distance between $\mathcal{F}$ and $\mathcal{F}_m$:

$$\delta_m = \sup_{f \in \mathcal{F}} \inf_{f_m \in \mathcal{F}_m} \|f - f_m\|_2.$$

$$\Big\|\Pi(r|\mathcal{F}_m) - \Pi(r|\mathcal{F})\Big\|_2 \leq \Big[\delta_m \Big\{2\Big\|r - \Pi(r|\mathcal{F})\Big\|_2 + \delta_m\Big\}\Big]^{1/2}.$$

- If $\delta_m = o(t_n)$, then $\|\Pi(f_n|\mathcal{F}_m) - \Pi(f_n|\mathcal{F})\|_2 = o_p(t_n)$.

- Conclusion: $\dfrac{\Pi(f_n|\mathcal{F}_m) - f}{t_n} \xrightarrow{d} \Pi(G|C_f)$ in $\mathbb{H}$.

4. Application to Pickands dependence function (1)

Copula $(Y_{i1}, Y_{i2})_{i=1,\ldots,n} \overset{iid}{\sim} F$, with margins F_1, F_2

Copula $C : [0,1]^2 \to [0,1]$

$$F(x_1, x_2) = C(F_1(x_1), F_2(x_2)) \qquad \forall (x_1, x_2) \in \mathbb{R}^2$$

If F_1 and F_2 are both continuous, C is uniquely defined by

$$C(u_1, u_2) = F(F_1^{-1}(u_1), F_2^{-1}(u_2)) \qquad 0 \leq u_1, u_2 \leq 1$$

where $F_i^{\leftarrow}(p) = \inf\{u \in \mathbb{R} | F_i(u) \geq p\} \qquad p \in (0,1)$

4. Application to Pickands dependence function (2)

Bivariate extreme-value copula

$$(Y_{i1}, Y_{i2})_{i=1,\ldots,n} \overset{iid}{\sim} \text{ GEV distribution}$$

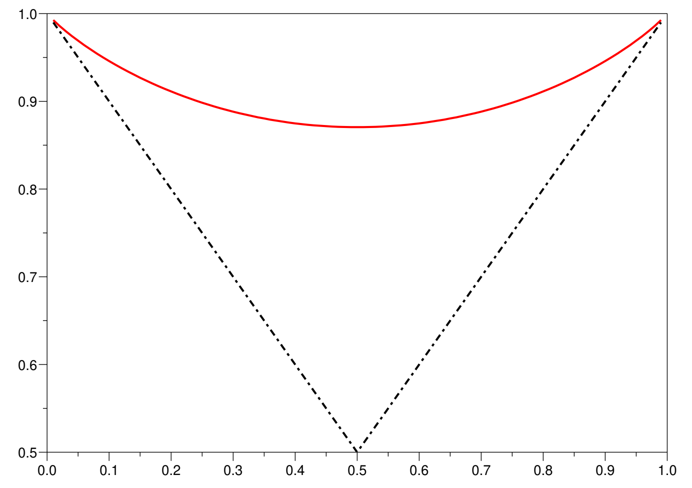
For $0 \le u, v \le 1$,

$$C(u,v) = \exp\left\{ \log(uv) A\left(\frac{\log u}{\log(uv)} \right) \right\},$$

with A a Pickands dependence function.

4. Application to Pickands dependence function (3)

Pickands dependence function A



- Properties
  - Convexity and continuity
  - $\max(t, 1-t) \le A(t) \le 1$
- Special cases
  - $A(t) = 1$: independence
  - $A(t) = \max(t, 1-t)$: total positive dependence
- Notation : $\mathcal{A} = \{$all Pickands dependence functions$\}$

4. Application to Pickands dependence function (4)

Some estimators for A

$(Y_{i1}, Y_{i2})_{i=1,..,n}$ i.i.d. random vectors, with margins F_1 and F_2

$X_{ij} = -\log F_j(Y_{ij})$ standard exponential margins

- Hall and Tajvidi (2000)

$$A_{HT}(u) = n \left\{ \sum_{i=1}^{n} \min \left(\frac{\hat{X}_i^{(1)}}{1-u}, \frac{\hat{X}_i^{(2)}}{u} \right) \right\}^{-1}$$

where

$$\hat{X}_i^{(j)} = X_{ij} \frac{n}{\sum_{l=1}^{n} X_{lj}}$$

4. Application to Pickands dependence function (5)

- Capéraà, Fougères and Genest (1997)

$$\log \widetilde{A}_n(t) = p(t) \log A_n^0(t) + \{1 - p(t)\} \log A_n^1(t)$$

$$A_n^0(t) = \exp\left\{\int_0^t \frac{G_n(z) - z}{z(1-z)} dz\right\}, \quad A_n^1(t) = \exp\left\{-\int_t^1 \frac{G_n(z) - z}{z(1-z)} dz\right\}$$

where

$$(U_i, V_i) := (F_1(Y_{i1}), F_2(Y_{i2}))$$

$$Z_i = \log(U_i)/\log(U_i V_i)_{i=1,\ldots,n} \stackrel{iid}{\sim} G$$

For a bounded function p on [0,1]:

$$\sqrt{n}(\log \widetilde{A}_n - \log A) \xrightarrow{d} \mathbb{G}^{CFG}$$

4. Application to Pickands dependence function (6)

Asymptotic results on $\mathcal{A}$

We propose a reformulation of CFG's estimator: For $p(t) = t$

$$\log \widetilde{A}_{n,C}(u) = \frac{1}{n}\sum_{i=1}^{n} \log \max(uX_{i1}, (1-u)X_{i2}) - u\frac{1}{n}\sum_{i=1}^{n} \log X_{i1} - (1-u)\frac{1}{n}\sum_{i=1}^{n} \log X_{i2}$$

Using a result from Deheuvels (1991) and the delta-method:

$$\sqrt{n}(\widetilde{A}_{n,C} - A) \xrightarrow{d} A\,\mathbb{G}^{CFG}.$$

$$\Longrightarrow \qquad \sqrt{n}\left(\Pi(\widetilde{A}_{n,C}|\mathcal{A}) - A\right) \xrightarrow{d} \Pi(A\,\mathbb{G}^{CFG}|C_A).$$

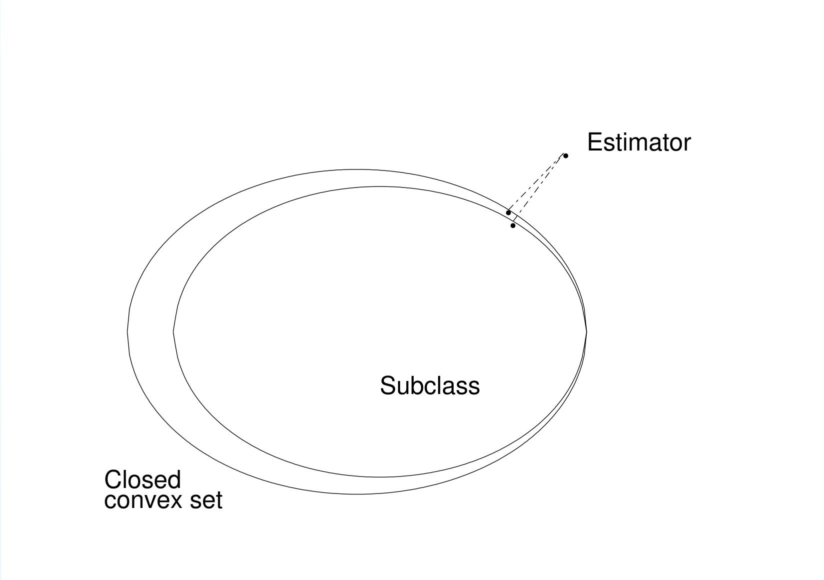
4. Application to Pickands dependence function (7)

Asymptotic results on $\mathcal{A}_m$

$\mathcal{A}_m = \{A \in \mathcal{A}$ piecewise linear with knots on $0, 1/m, \ldots, 1\} \subset \mathcal{A}$

What about $\Pi(\widetilde{A}_{n,C}|\mathcal{A}_m)$?

Here, $\delta_m = O(m^{-5/4})$.



Let $m = m_n$ such that $\dfrac{n^{2/5}}{m} \to 0$,

$$\|\Pi(\widetilde{A}_{n,C}|\mathcal{A}_m) - \Pi(\widetilde{A}_{n,C}|\mathcal{A})\| = o(n^{-1/2})$$

$$\Longrightarrow \qquad \sqrt{n}\left(\Pi(\widetilde{A}_{n,C}|\mathcal{A}_m) - A\right) \xrightarrow{d} \Pi(A\,\mathbb{G}^{CFG}|C_A)$$

4. Application to Pickands dependence function (8)

Simulation study

- Logistic model: $A(w) = \left[(1-w)^{1/\alpha} + w^{1/\alpha}\right]^{\alpha} \qquad 0 < \alpha \leq 1$

  - independence: $\alpha = 1$
  - complete dependence: $\alpha \to 0$

- Asymmetric logistic model: For $0 \leq \psi_j \leq 1$, $0 < \alpha \leq 1$

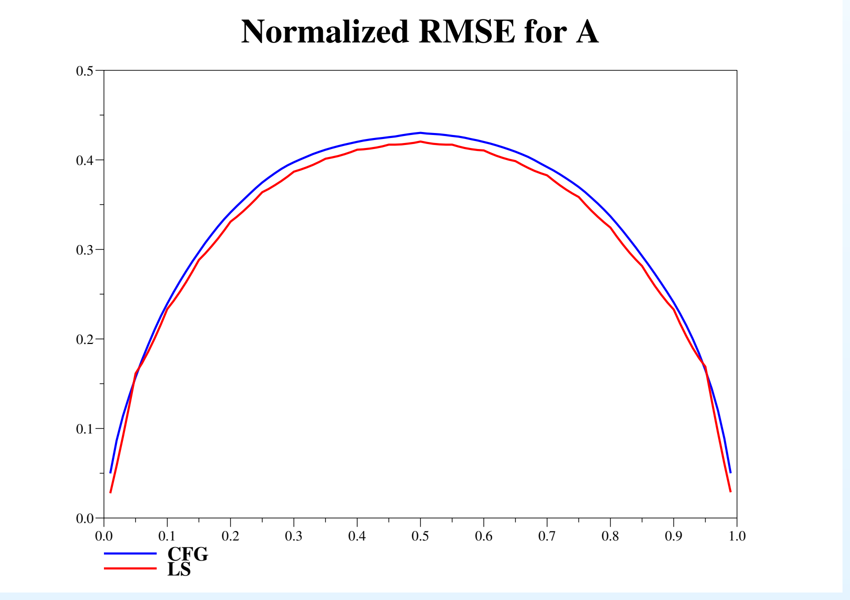
$$A(w) = (1-\psi_1)(1-w)+(1-\psi_2)w+\Big((\psi_1(1-w))^{1/\alpha}+(\psi_2 w)^{1/\alpha}\Big)^{\alpha}$$

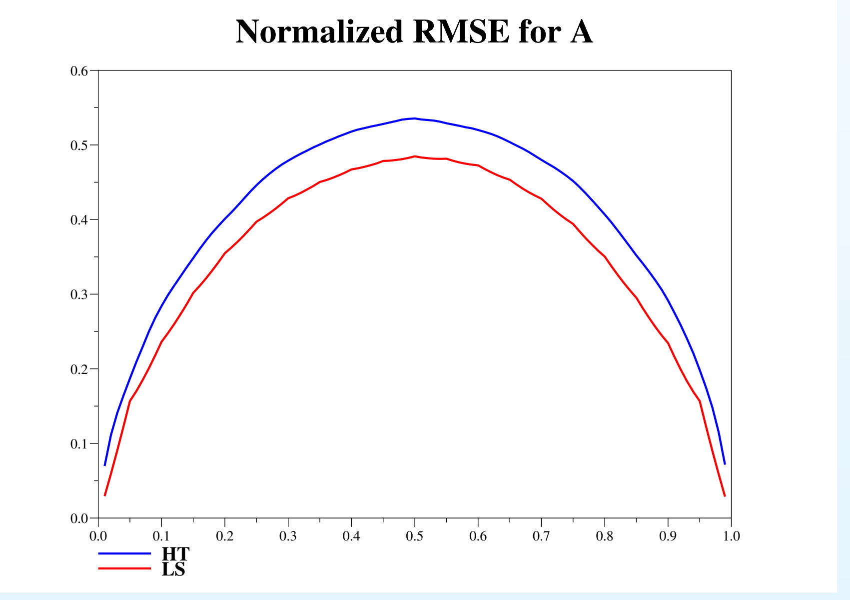
- Normalized root mean squared error for A

$$\left(\frac{n}{500}\sum_{i=1}^{500}\left\{\hat{A}_n^{(i)}(w) - A(w)\right\}\right)^{1/2}$$

4. Application to Pickands dependence function (9)

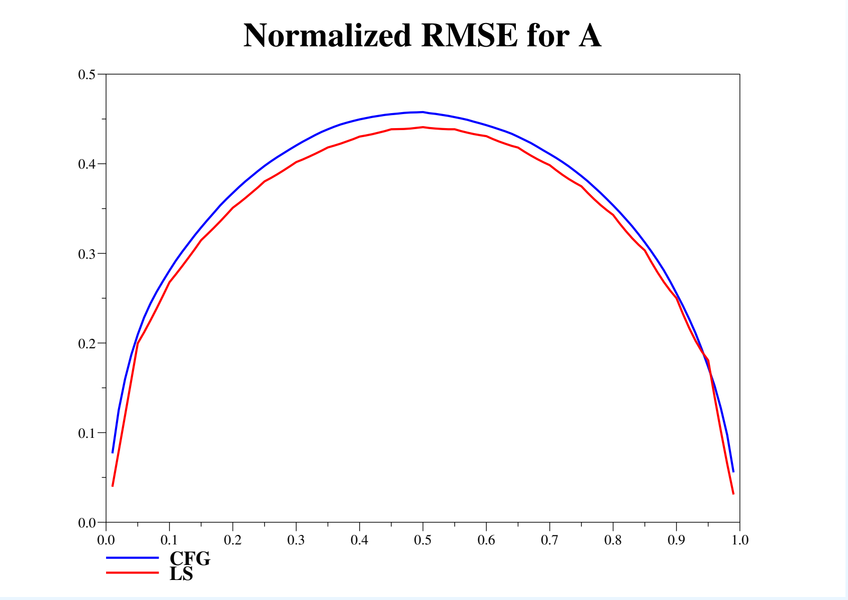
- Simulation of 500 samples of size n=100.
- Choice of m=20 basis functions.

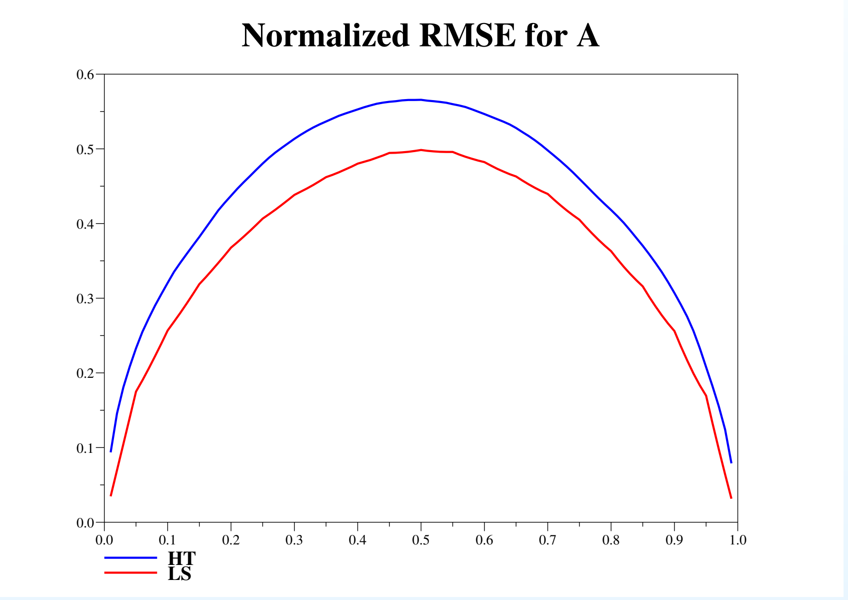




Normalized RMSE. Logistic model with $\alpha = 0.9$

4. Application to Pickands dependence function (10)





Normalized RMSE. Asymmetric logistic model with $\alpha = 0.7$, $\psi_1 = 0.5$, $\psi_2 = 1$

Conclusion

Idea: Estimation of a convex function.

New estimator:

$$\widehat{f}_n^{ls} = \Pi(f_n|\mathcal{F}) = \arg\min_{f \in \mathcal{F}} \int_0^1 \{f_n(\omega) - f(\omega)\}^2 d\omega.$$

Advantages:

- admissible
- smaller integrated squared error
- no smoothing parameter to choose

Let $f_n \in \mathbb{H}$ such that $\dfrac{f_n - f}{t_n} \xrightarrow{d} \mathbb{G}$ with $t_n \to 0$, then

$$\frac{\Pi(f_n|\mathcal{F}) - f}{t_n} \xrightarrow{d} \Pi(\mathbb{G}|C_f) \qquad \text{in } \mathbb{H}.$$

Extension of this work

$\longrightarrow$ Estimation of a function with shape constraints

- Regression function: estimation of $r(x) = E(Y|X = x)$.

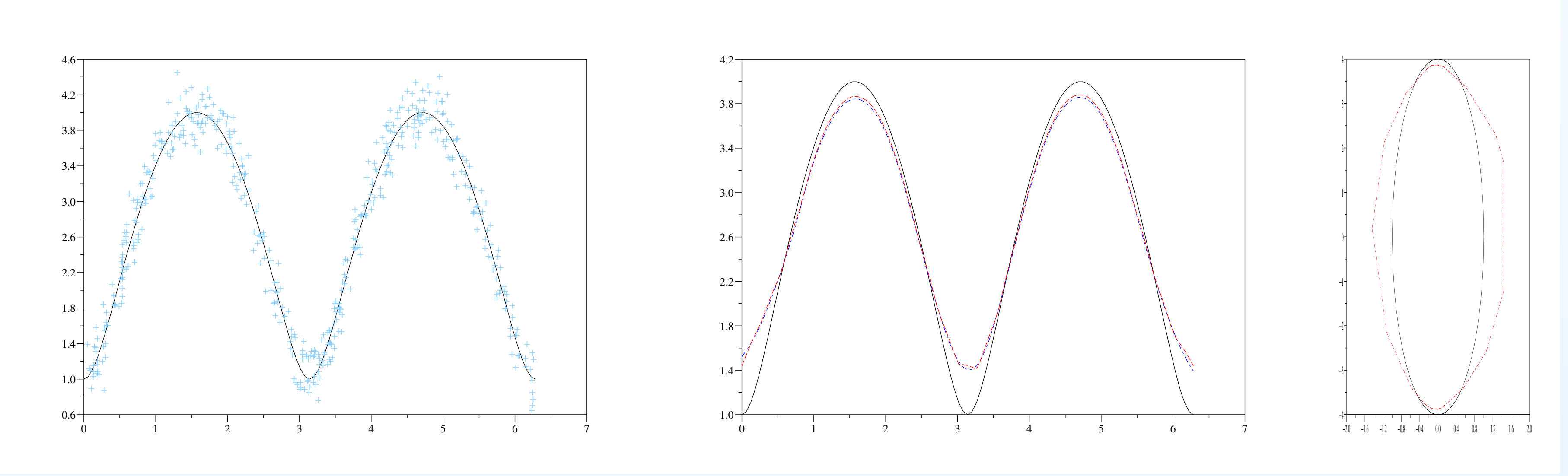
  Hypothesis on r: monotonicity or convexity.

- Density function

  Hypothesis: monotonicity or convexity.

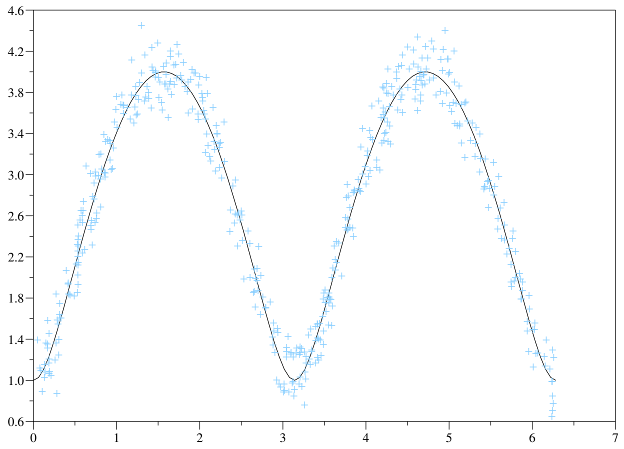
Extension of this work

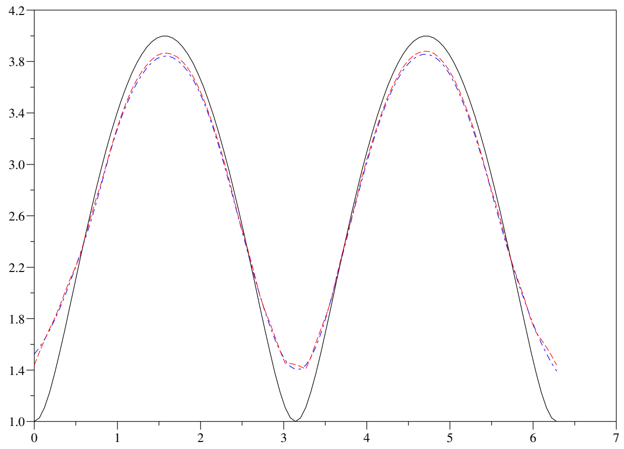
- Support function of a convex set

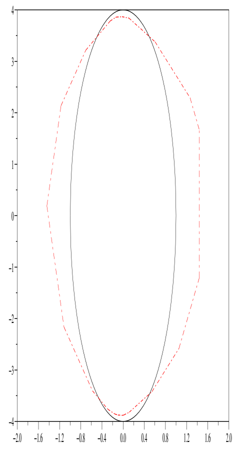


Extension of this work

- Support function of a convex set







- ...