# Frontier estimation via regression on high power-transformed data

**Stéphane Girard**
INRIA Rhône-Alpes, team Mistis
http://mistis.inrialpes.fr/people/girard

Joint work with Pierre Jacob, Montpellier 2 University

# Outline

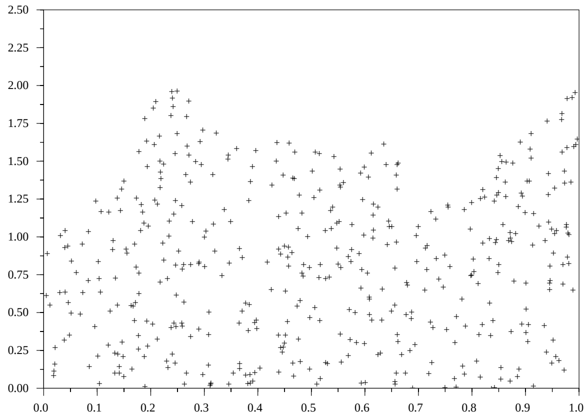
1. Frontier estimation.
2. Basic principle.
3. Theoretical properties.
4. Simulation study.

# 1. Frontier estimation.

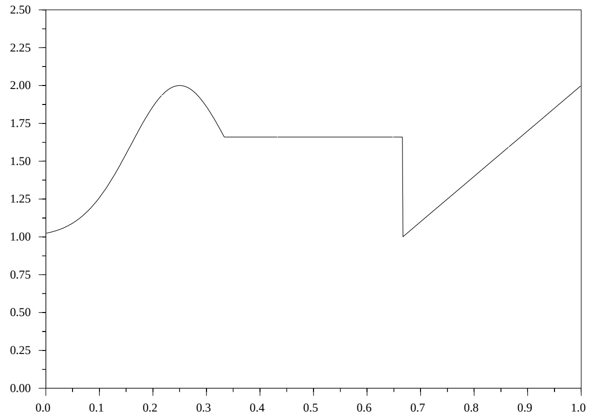
Let $(X_i, Y_i)$, $i = 1, \ldots, n$ be independent copies of a random pair $(X, Y)$ with support $S$ defined by

$$S = \{(x, y) \in E \times \mathbb{R}; 0 \leq y \leq g(x)\}.$$

The unknown function $g : E \longrightarrow \mathbb{R}$ is called the frontier. We address the problem of estimating $g$ in the case $E \subset \mathbb{R}^d$.




$\overset{?}{\longrightarrow}$



Our estimator of the frontier is based on a kernel regression on the power-transformed data. More precisely, the estimator of $g$ is defined for all $x \in \mathbb{R}^d$ by

$$\hat{g}_n(x) = \left( (p+1) \sum_{i=1}^{n} K_h(x - X_i) Y_i^p \Big/ \sum_{i=1}^{n} K_h(x - X_i) \right)^{1/p},$$

where

- $K_h(t) = K(t/h)/h^d$, with $K$ being a probability density function (kernel) on $\mathbb{R}^d$,
- $h = h_n$ a non-random sequence (bandwidth) such that $h \to 0$ as $n \to \infty$,
- $p = p_n$ a non-random sequence such that $p \to \infty$ as $n \to \infty$.

Note that, basing on the same principle, a local polynomial estimator has also been proposed.

## 2. Basic principle.

Let $Y_1, \ldots, Y_n$ be independent random variables from a $U([0,\theta])$ distribution. Consider

$$\bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i \quad \text{and} \quad Y_{n,n} = \max\{Y_1, \ldots, Y_n\}.$$

It is well-known that $2\bar{Y}_n$ and $\frac{n+1}{n}Y_{n,n}$ are two unbiased estimators of $\theta$ with variances

$$\operatorname{var}(2\bar{Y}_n) \propto \frac{1}{n} \quad \text{and} \quad \operatorname{var}\left(\frac{n+1}{n}Y_{n,n}\right) \propto \frac{1}{n^2}.$$

Similarly, introducing for all $p \geq 1$,

$$\bar{Y_n^p} = \frac{1}{n}\sum_{i=1}^{n} Y_i^p,$$

the random variable $(p+1)\bar{Y_n^p}$ is an unbiased estimator of $\theta^p$ with variance

$$\operatorname{var}((p+1)\bar{Y_n^p}) \propto \frac{p^2}{n(2p+1)}.$$

Consider the new estimator of $\theta$ defined by

$$\hat{\theta}_n = ((p+1)\bar{Y_n^p})^{1/p} = \left(\frac{p+1}{n}\sum_{i=1}^{n} Y_i^p\right)^{1/p}.$$

Then, if $p \to \infty$ with $p/n \to 0$, one has the convergence in distribution

$$\sqrt{n(2p+1)}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, \theta^2),$$

which can be compared to the classical result

$$n\left(\frac{n+1}{n}Y_{n,n} - \theta\right) \xrightarrow{d} \mathcal{EVD}.$$

- Both estimators have (almost) same asymptotical variances,
- In the conditional case, $\hat{\theta}_n$ is easier to implement than $\frac{n+1}{n}Y_{n,n}$ since it does not require the extraction of the conditional maxima.

Back to the conditional case, if $Y$ given $X = x$ follows a $\mathcal{U}([0, g(x)])$ distribution, then

$$r_n(x) = \mathbb{E}\left((p+1)Y^p | X = x\right) = g^p(x),$$

and our estimator

$$\hat{g}_n(x) = \left( (p+1) \sum_{i=1}^{n} K_h(x - X_i) Y_i^p \Bigg/ \sum_{i=1}^{n} K_h(x - X_i) \right)^{1/p}$$

can be interpreted as

$$\hat{g}_n(x) = \hat{r}_n^{1/p}(x)$$

where $\hat{r}_n(x)$ is the classical kernel estimator for the conditional expectation

$$\hat{r}_n(x) = (p+1) \sum_{i=1}^{n} K_h(x - X_i) Y_i^p \Bigg/ \sum_{i=1}^{n} K_h(x - X_i).$$

## 3. Theoretical properties.

### Assumptions

**(A.1)**: The frontier $g$ is $\alpha$-Lipschitz and the $X_i$'s cdf $f$ is $\beta$-Lipschitz, with $0 < \alpha \leq \beta \leq 1$,

**(A.2)**: $0 < g_{\min} \leq g(x)$, $\forall x \in \mathbb{R}^d$,

**(A.3)**: $f(x) \leq f_{\max} < \infty$, $\forall x \in \mathbb{R}^d$,

**(A.4)**: $K$ is a Lipschitzian pdf on $\mathbb{R}^d$, with support included in $B$, the unit ball of $\mathbb{R}^d$.

**(A.5)**: $Y$ given $X = x$ is uniformly distributed on $[0, g(x)]$.

The frontier estimator can be expanded as

$$\hat{g}_n(x) = \left( (p+1) \sum_{i=1}^{n} K_h(x - X_i) Y_i^p \Bigg/ \sum_{i=1}^{n} K_h(x - X_i) \right)^{1/p} = \left( \frac{\hat{\varphi}_n(x)}{\hat{f}_n(x)} \right)^{1/p}$$

where:

- $\hat{f}_n(x)$ is the classical kernel estimator

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^{n} K_h(x - X_i)$$

  of $f(x)$ the $X_i$'s probability density function.

- $\hat{\varphi}_n(x)$ is the classical kernel estimator

$$\hat{\varphi}_n(x) = \frac{p+1}{n} \sum_{i=1}^{n} K_h(x - X_i) Y_i^p$$

  of $\varphi_n(x) = f(x) r_n(x)$.

## 3.1. Preliminary result.

The properties of $\hat{f}_n(x)$ are well-known:

$$\begin{aligned}
\mathbb{E}\left(\frac{\hat{f}_n(x)}{f(x)}\right) &= 1 + O(h^\alpha), \\
\text{var}\left(\frac{\hat{f}_n(x)}{f(x)}\right) &= O(1/nh^d).
\end{aligned}$$

Let us focus on $\hat{\varphi}_n(x)$:

**Lemma 1** *Under* **(A.1)**–**(A.5)**, *if* $ph^\alpha \to 0$, *then for all* $x \in \mathbb{R}^d$

$$\begin{aligned}
\mathbb{E}\left(\frac{\hat{\varphi}_n(x)}{\varphi_n(x)}\right) &= 1 + O(ph^\alpha), \\
var\left(\frac{\hat{\varphi}_n(x)}{\varphi_n(x)}\right) &= \frac{1}{nh^d}\frac{(p+1)^2}{2p+1}\int_B K^2(s)ds\frac{1}{f(x)}\left[1 + o(1)\right].
\end{aligned}$$

In view of these results, the asymptotic behavior of $\hat{g}_n(x)$ is driven by $\hat{\varphi}_n(x)$.

## 3.2. Asymptotic normality.

**Theorem 1** *Suppose that* $nph^{d+2\alpha} \to 0$ *and* $p/(nh^d) \to 0$. *Let us define*

$$\sigma_n^{-1}(x) = ((2p+1)nh^d)^{1/2} \left( \frac{f(x)}{\int_B K^2(t)dt} \right)^{1/2}.$$

*Then, under* **(A.1)**–**(A.5)**, *for all* $x \in \mathbb{R}^d$,

$$\sigma_n^{-1}(x) \left( \frac{\widehat{g}_n(x)}{g(x)} - 1 \right) \xrightarrow{d} \mathcal{N}(0,1).$$

One can choose $h = n^{-1/(d+\alpha)}$ and $p = \varepsilon_n n^{\alpha/(d+\alpha)}$, where $(\varepsilon_n)$ is a sequence tending to zero arbitrarily slowly. These choices yield

$$\sigma_n^{-1}(x) = \varepsilon_n^{1/2} n^{\alpha/(d+\alpha)} \left( \frac{2f(x)}{\int_B K^2(t)dt} \right)^{1/2} (1+o(1)),$$

which is the optimal speed (up to the $\varepsilon_n$ factor) for estimating $\alpha-$ Lipschitzian $d-$ dimensional frontiers.

### 3.3. Complete convergence.

Although, in the definition of $\hat{g}_n(x)$, the normalizing term $(p+1)^{1/p}$ is specially designed for the case where $Y$ given $X = x$ is uniformly distributed on $[0, g(x)]$, it can be shown that $\hat{g}_n(x)$ is completely convergent to $g$ without assumption neither on the distribution of $X$ nor on the distribution of $Y$ given $X = x$.

**Theorem 2** *Suppose* **(A.1)**–**(A.4)** *hold and* $nh^d / \log n \to \infty$. *Then* $\widehat{g}_n(x)$ *converges completely to* $g(x)$ *for all* $x \in \mathbb{R}^d$ *such that* $f(x) > 0$.

## 4. Numerical experiments.

Here, we limit ourselves to unidimensional random variables $X$ $(d = 1)$ with compact support $E = [0, 1]$. Besides, $Y$ given $X = x$ is distributed on $[0, g(x)]$ such that

$$\mathbb{P}(Y > y | X = x) = \left(1 - \frac{y}{g(x)}\right)^{\gamma},$$

with $\gamma > 0$. This conditional survival distribution function belongs to the Weibull domain of attraction, with extreme value index $-\gamma$.

- The case $\gamma = 1$ corresponds to the situation where $Y$ given $X = x$ is uniformly distributed on $[0, g(x)]$.
- The larger $\gamma$ is, the smaller the probability $\mathbb{P}(Y > y|X = x)$ is, when $y$ is close to the frontier $g(x)$.

The following kernel is chosen

$$K(t) = \cos^2(\pi t/2)\mathbf{1}\{t \in [-1, 1]\},$$

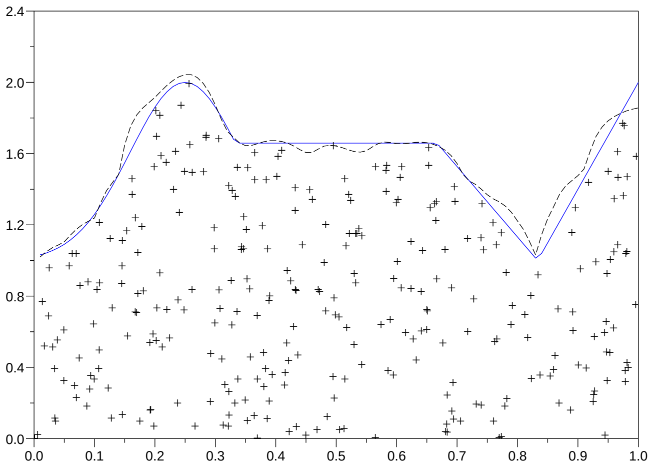
with associated bandwidth $h = 4\hat{\sigma}(X)n^{-1/2}$ and with $p = n^{1/2}$.

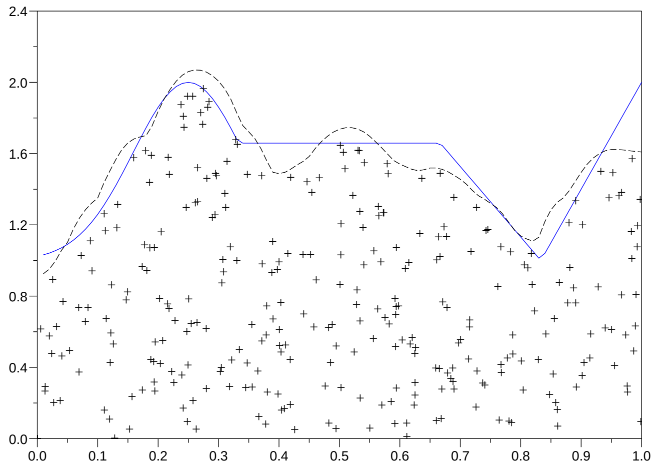
- The dependence of these sequences with respect to $n$ is chosen according to Theorem 1 with $\alpha = d = 1$.
- The multiplicative constant $4\hat{\sigma}(X)$ in $h$ is chosen heuristically.
- The dependence with respect to the standard-deviation of $X$ is inspired from the density estimation case.
- The scale factor 4 was chosen on the basis of intensive simulations.

The experiment involves four steps:

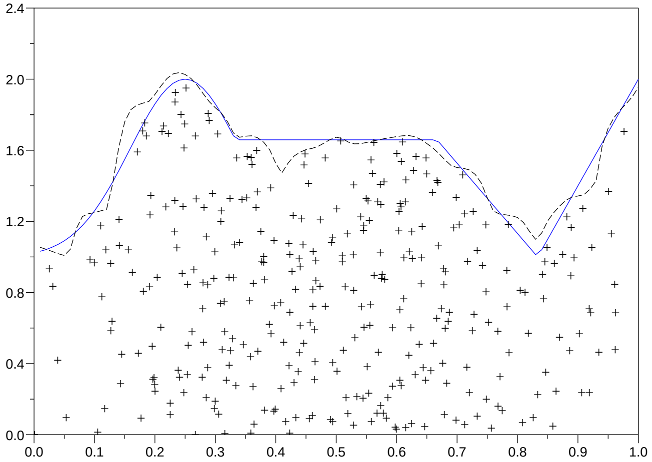
- First, $m = 500$ replications of the sample are simulated.
- For each of the $m$ previous set of points, the frontier estimator $\hat{g}_n$ is computed.
- The $m$ associated $L_1$ distances to $g$ are evaluated on a grid.
- The smallest and largest $L_1$ errors are recorded.

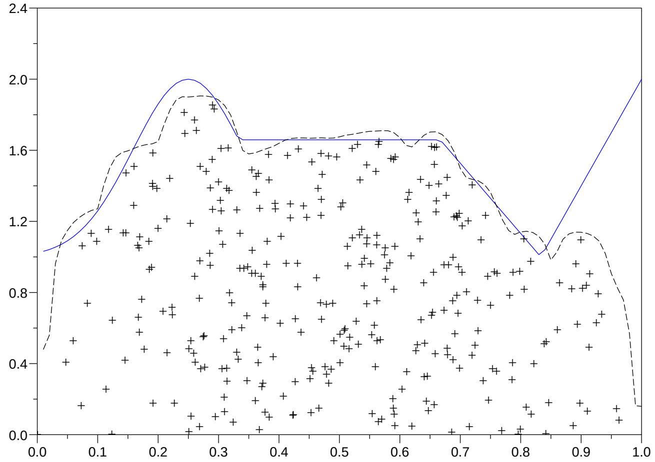
The best situation (i.e. the estimation corresponding to the smallest $L_1$ error) and the worst situation (i.e. the estimation corresponding to the largest $L_1$ error) are represented.



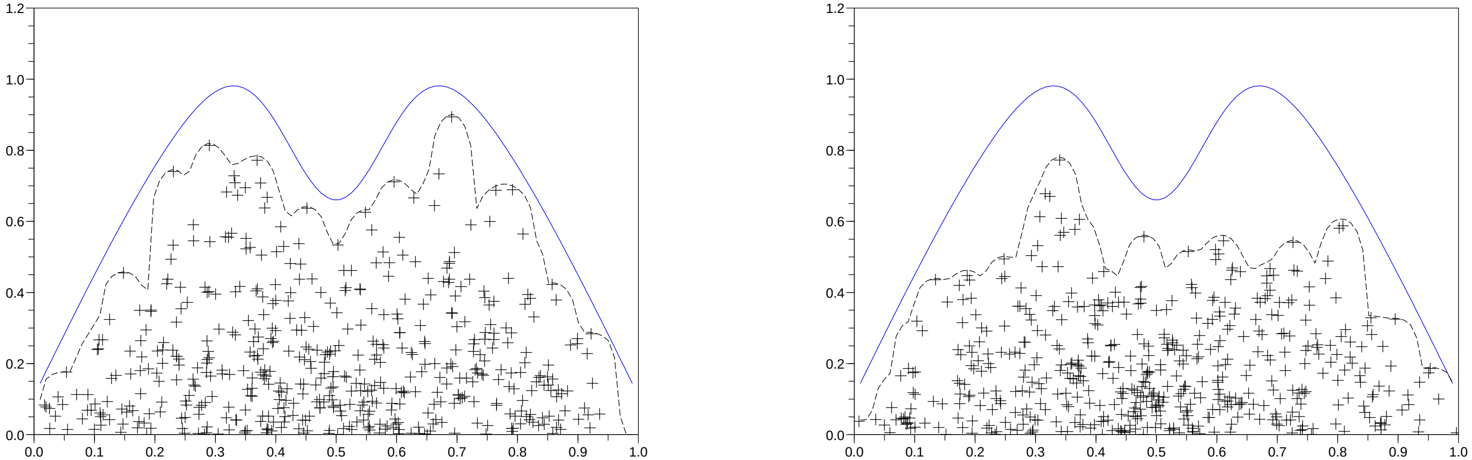




The frontier (blue) and its estimation (black). Left: Best situation, Right: Worst situation. The sample size is $n = 300$, $X$ is uniformly distributed on $[0, 1]$ and $\gamma = 1$.




The frontier (blue) and its estimation (black). Left: Best situation, Right: Worst situation. The sample size is $n = 300$, $X$ is Beta$(2, 2)$ distributed on $[0, 1]$ and $\gamma = 1$.




The frontier (blue) and its estimation (black). Left: Best situation, Right: Worst situation. The sample size is $n = 500$, $X$ is Beta$(2, 2)$ distributed on $[0, 1]$ and $\gamma = 3$.