Three different approaches to frontier estimation

Stéphane Girard

INRIA Grenoble Rhône-Alpes & LJK
http://mistis.inrialpes.fr/people/girard/

March, 2014

Joint work with A. Guillou (Université de Strasbourg, France), A. Iouditski (Université de Grenoble, France), L. Menneteau (Université de Montpellier, France), A. Nazin (Institute of Control Sciences, Moscow, Russia) and G. Stupfler (Université Aix-Marseille, France).
Outline

- Very brief overview of the literature:
  - First frontier estimator Geffroy (ISUP, 1964)
  - Piecewise polynomial estimators Härdle, Park, Tsybakov (JMVA, 1995)
- Extreme-value estimators,
- Linear programming estimators,
- High order moments estimators.
Let \((X_i, Y_i), 1 \leq i \leq n\) be \(n\) independent copies of a random pair \((X, Y)\) such that their common distribution has a support

\[ S := \{(x, y) \in \Omega \times \mathbb{R} ; 0 \leq y \leq g(x)\} \]

where

- \(X\) has a density \(f_X\) on the compact subset \(\Omega \subset \mathbb{R}^d\),
- \(Y|X = x\) has a density \(f(.|x)\) on \([0, g(x)]\),
- \(g\) is a positive function, \(g(x) = \sup\{Y|X = x\}\).

We address the problem of the estimation of \(g\), called the frontier of \(S\).
Illustration $\Omega = [0, 1]$
Härdle, Park, Tsybakov (JMVA, 1995) assumed that, for all \((x, y) \in S\),

- \(f_X(x) \geq f_{\text{min}} > 0\),
- \(f(y|x) \geq c(g(x) - y)^\alpha\) where \(c > 0\) and \(\alpha \geq 0\).

Two cases arise:

- If \(\alpha = 0\) then \(f(y|x) \geq c > 0\) for all \(y \in [0, g(x)]\), this is the situation of a “sharp boundary”.
- If \(\alpha > 0\) then we may have \(f(y|x) \to 0\) as \(y \to g(x)\), this is the situation of a “non-sharp boundary”.

Stéphane Girard

Three different approaches to frontier estimation
First frontier estimator **Geffroy (ISUP, 1964)**, based on the extreme-values of the sample:

- Partition of $\Omega = [0, 1]$ into equidistant $k_n$ intervals $I_{n,r}$, $r = 1, \ldots, k_n$,
- Maxima on each bin: $Y_{n,r}^* = \max\{Y_i : X_i \in I_{n,r}\}$,
- Piecewise constant estimator:

$$\hat{g}_n(x) = \sum_{r=1}^{k_n} 1\{x \in I_{n,r}\} Y_{n,r}^*.$$
Illustration: Jeffroy’s estimator

Three different approaches to frontier estimation
Asymptotic behaviour of the $L_1$– distance $\Delta_n := \int_0^1 |\hat{g}_n(x) - g(x)| \, dx$.

**Theorem**

Assume that $g$ is $\gamma$– Lipschitzian $\gamma \in (0, 1]$ and $\alpha = 0$ (sharp boundary). If some conditions on $(k_n)$ hold, then $(n/k_n)(\Delta_n - \beta_n)$ converges in distribution to a Gumbel r.v. with c.d.f $\psi(z) = \exp(-\exp(-\theta z))$ where

$$\theta = \inf_{x \in [0,1]} f_X(x)f(g(x)|x),$$

and $\beta_n$ is the solution of the equation

$$\int_0^1 \exp \left[ \log k_n - \frac{n\beta_n}{k_n} f_X(x)f(g(x)|x) \right] \, dx = 1.$$

The rate of convergence $(n/k_n)$ is (up to a logarithmic factor) $n^{\gamma/(1+\gamma)}$. 
Proposed in Härdle, Park, Tsybakov (JMVA, 1995) to deal with

- sharp or non-sharp boundaries ($\alpha \geq 0$),
- smoother frontiers, \textit{i.e.} for $\gamma > 0$, it is assumed that the $\lfloor \gamma \rfloor$th derivative of the frontier $g$ is $(\gamma - \lfloor \gamma \rfloor)$-Lipschitzian.

The estimator requires a partition $I_{n,r}$, $r = 1, \ldots, k_n$ of $\Omega = [0, 1]$. On the $r$th bin, the estimator is defined as the polynomial of degree $\lfloor \gamma \rfloor$ covering all the points and with smallest surface.

$$\hat{g}_n^{\theta}(x) = \sum_{r=1}^{k_n} \mathbb{I}\{x \in I_{n,r}\} P_{n,r}(x; \theta_{n,r}).$$

$$\theta_{n,r} = \arg \min_{\theta} \int_{I_{n,r}} P_{n,r}(x; \theta) dx \text{ s.t. } P_{n,r}(X_i; \theta) \geq Y_i, \ X_i \in I_{n,r}.$$ 

Note that if $\gamma \in (0, 1]$ then $\lfloor \gamma \rfloor = 0$ and we find back Geffroy's estimator.
Under the above assumptions, and for a well chosen partition, piecewise polynomial estimators have the optimal rate of convergence for the $L_1$-error, that is $n^{\gamma}/(1+(\alpha+1)\gamma)$.

- In the case where $\alpha = 0$ (sharp boundary) and $\gamma \in (0, 1]$, Geffroy’s estimator has the optimal rate of convergence.
- In practice, the estimators are biased downward and discontinuous. The choice of the partition $(k_n)$ is also an issue.
Illustration: Piecewise linear estimator

Three different approaches to frontier estimation
Contributions

- **Extreme-value estimator** (smoothed, bias correction, sharp boundary, pointwise asymptotic normality)
- **Linear programming estimator** (smoothed, no partition of $\Omega$, sharp boundary, strong $L_1$ consistency)
- **High order moments estimator** (smoothed, no partition of $\Omega$, non-sharp boundary, pointwise asymptotic normality, strong $L_\infty$ consistency)
1. Extreme-value estimator

Support $S = \{(x, y) \in \Omega \times \mathbb{R} ; 0 \leq y \leq g(x)\}$ with $\Omega \subset \mathbb{R}^d$.

- **Geffroy’s estimator.**

$$\hat{g}^{(0)}_n(x) = \sum_{r=1}^{k_n} \mathbb{I}\{x \in I_{n,r}\} Y_{n,r}^*.$$  

where $\{I_{n,r}, \ r = 1, \ldots, k_n\}$ is a partition of $\Omega$ and $Y_{n,r}^* = \max\{Y_i : X_i \in I_{n,r}\}$.

- **Bias correction.**
  Assume that $Y|X = x$ is uniformly distributed on $[0, g(x)]$ (sharp boundary).

$$\hat{g}^{(1)}_n(x) = \sum_{r=1}^{k_n} \mathbb{I}\{x \in I_{n,r}\} Y_{n,r}^*(1 + \frac{N_{n,r}^{-1}}{N_{n,r}}),$$

where $N_{n,r}$ is the number of $X_i \in I_{n,r}$. 
Extreme-value estimator

- **Smoothing**

\[
\hat{g}_n^{(2)}(x) = \int_{\mathbb{R}^d} K_{h_n}(x - t)\hat{g}_n^{(1)}(t)dt
\]

where \( K_{h_n}(u) = h_n^{-d}K(u/h_n) \), \( K \) is \( d \)-dimensional density with compact support and \( h_n \) is a smoothing parameter.

Nonparametric regression over the extreme-values of the sample:

\[
\hat{g}_n^{(2)}(x) = \sum_{r=1}^{k_n} \int_{I_{n,r}} K_{h_n}(x - t)dt \ Y_{n,r}^*(1 + N_{n,r}^{-1})
\]

G & Menneteau (JSPI, 2005), Menneteau (ESAIM, 2008)

**Theorem**

Assume that \( g \) is \( \gamma \)-Lipschitzian, \( \gamma \in (0, 1] \). Under some conditions on the \((h_n)\) and \((k_n)\) sequences, for all \((x_1, \ldots, x_p) \subset \Omega\), the random vector

\[
\left\{ nh_n^{d/2}k_n^{-1/2} (\hat{g}_n^{(2)}(x_j) - g(x_j)) : 1 \leq j \leq p \right\}
\]

is asymptotically centred Gaussian with diagonal covariance matrix.
Choosing $h_n \asymp n^{-1/(\gamma+d)}$ and $k_n \asymp n^d/(\gamma+d)$, the rate of convergence is $n^\gamma/(d+\gamma)$, up to logarithmic factors.

Optimal $L_1$– rate of convergence for sharp boundaries ($\alpha = 0$) and $\gamma$– Lipschitzian frontiers, $\gamma \in (0, 1]$.

The rate of convergence of this extreme-value estimator is no more optimal for smoother frontier functions ($\gamma > 1$). The approximation of $g(x)$ by a constant value $Y_{n,r}^*$ for $x \in I_{n,r}$ is not precise enough.
Illustration: Extreme-value estimator
Contributions

- **Extreme-value estimator** (smoothed, bias correction, sharp boundary, pointwise asymptotic normality)
- **Linear programming estimator** (smoothed, no partition of $\Omega$, sharp boundary, strong $L_1$—consistency)
- **High order moments estimator** (smoothed, no partition of $\Omega$, non-sharp boundary, pointwise asymptotic normality, strong $L_\infty$—consistency)
2. Linear programming estimator

Support $S = \{(x, y) \in [0, 1] \times \mathbb{R} ; 0 \leq y \leq g(x)\}$, where $g$ is $\gamma$-Lipschitzian, $\gamma \in (0, 1]$.

The estimator is a linear combination of kernel functions:

$$\hat{g}_n(x) = \sum_{i=1}^{n} \alpha_i K_{h_n} (x - X_i).$$

The coefficients $(\alpha_i)_{i=1,...,n}$ are obtained by minimizing the surface of the estimated support:

$$\min \int_{\mathbb{R}} \hat{g}_n(x) dx = \min \sum_{i=1}^{n} \alpha_i,$$

under the following constraints: for all $i = 1, \ldots, n$

- $\hat{g}_n(X_i) \geq Y_i$ (the sample is below the estimated frontier)
- $\alpha_i \geq 0$ (the estimated frontier function is positive)
- $|\hat{g}'_n(X_i)| \leq c_0 h_n^{\gamma-1}$ (Lipschitz constraint)

Linear Programming (LP) problem.
Remark 1. Assume that $Y|X = x$ is uniformly distributed on $[0, g(x)]$

- Joint distribution of the sample $\Sigma_n = (X_i, Y_i)_{i=1,...,n}$:

$$P(\Sigma_n | g) = \prod_{i=1}^{n} \frac{g(X_i)}{C_g} \cdot \frac{1}{g(X_i)} \mathbb{I}\{0 \leq Y_i \leq g(X_i)\},$$

with $C_g = \int_{\mathbb{R}} g(x)dx$.

- Log-likelihood. Since $C_{\hat{g}_n} = \sum_{i=1}^{n} \alpha_i$, we have

$$L(\alpha) = \log P(\Sigma_n | \hat{g}_n) = -n \log \sum_{i=1}^{n} \alpha_i + \sum_{i=1}^{n} \log \mathbb{I}\{Y_i \leq \hat{g}_n(X_i)\}.$$ 

The (LP) problem can be read as the maximization of the log-likelihood under the additional constraints $|\hat{g}_n'(X_i)| \leq c_0 h_n^{\gamma-1}$, $i = 1, \ldots, n$. 

Stéphane Girard
Three different approaches to frontier estimation
Illustration: Linear programming estimator

(LP): Linear optimisation problem under linear constraints.

- Efficient algorithms,
- The solution is sparse: only few $\alpha_i \neq 0$ (triangles), not the same points as $\hat{g}_n(X_i) = Y_i$ (squares).
Theorem

Assume that $g$ is $\gamma$–Lipschitzian, $\gamma \in (0, 1]$. Under some conditions on the $(h_n)$ sequence (namely $h_n \asymp (\log n / n)^{1/(\gamma+1)}$)

$$\Delta_n := \int_0^1 |\hat{g}_n(x) - g(x)| \, dx = O\left(\left(\frac{\log n}{n}\right)^{\gamma/(1+\gamma)}\right),$$

almost surely.

- Optimal $L_1$–rate of convergence for sharp boundaries ($\alpha = 0$), $d = 1$ and $\gamma$–Lipschitzian frontiers, $\gamma \in (0, 1]$, up to the logarithmic factor.
- Extension to $\gamma > 1$ should be possible.
Linear programming estimator

**Sketch of the proof:** Lower bound.

**Lemma**

\[ \hat{g}_n(x) \geq g(x) - O(h^\gamma) \text{ a.e.} \]

**Proof:** There exists a.e. a point \((X_i, Y_i)\) close to \((x, g(x))\) i.e. such that \(|x - X_i| \leq c_1 h_n\) and \(0 \leq g(X_i) - Y_i \leq c_2 h_\gamma^n\). Then,

\[
\begin{align*}
g(x) - \hat{g}_n(x) &= [g(x) - g(X_i)] \\
&+ [g(X_i) - Y_i] \\
&+ [Y_i - \hat{g}_n(X_i)] \\
&+ [\hat{g}_n(X_i) - \hat{g}_n(x)].
\end{align*}
\]

The terms are respectively controlled:

i) \(|g(x) - g(X_i)| \leq c_3 h_\gamma^n\) : the frontier is \(\gamma\)-Lipschitzian,

ii) \(0 \leq g(X_i) - Y_i \leq c_2 h_\gamma^n\) : choice of the point,

iii) \(Y_i - \hat{g}_n(X_i) \leq 0\) : the point is below the estimated frontier,

iv) \(|\hat{g}_n(X_i) - \hat{g}_n(x)| \leq c_0 h_\gamma^{\gamma-1} c_1 h_n\) : Lipschitz constraint.
Sketch of the proof: Upper bound.

Lemma

There exist a solution \( \tilde{g}_n \) to (LP) such that

\[
\int_0^1 \tilde{g}_n(x) dx \leq \int_0^1 g(x) dx + c_4 h_n^{\gamma} \text{ a.e.}
\]

Proof: The idea is to consider

\[
\tilde{\alpha}_{i,n} = \int_{X_{i-1,n}}^{X_{i+1,n}} (g(x) + c_4 h_n^{\gamma}) dx
\]

and show that

\[
\tilde{g}_n(x) = \sum_{i=1}^{n} \tilde{\alpha}_{i,n} K_{h_n} (x - X_{i,n})
\]

satisfies the constraints of (LP).
Contributions

- **Extreme-value estimator** (smoothed, bias correction, sharp boundary, pointwise asymptotic normality)
- **Linear programming estimator** (smoothed, no partition of $\Omega$, sharp boundary, strong $L_1$—consistency)
- **High order moments estimator** (smoothed, no partition of $\Omega$, non-sharp boundary, pointwise asymptotic normality, strong $L_\infty$—consistency)
High order moments estimator

- Support $S = \{(x, y) \in \Omega \times \mathbb{R} ; 0 \leq y \leq g(x)\}$ with $\Omega \subset \mathbb{R}^d$.
- Conditional survival function of $Y$ given $X = x$

$$F(y | x) = (1 - y/g(x))^{\alpha(x)+1}, \ \forall x \in \Omega, \ \forall y \in [0, g(x)],$$

where $\alpha(x) \geq -1$ (sharp or non-sharp boundary).
- Conditional moments: $\forall p \geq 1$,

$$\mu_p(x) := \mathbb{E}(Y^p | X = x).$$

Then, for all $p \geq 1$ and $\theta > 1$,

$$\frac{1}{g(x)} = \frac{1}{(\theta - 1)p} \left[ (\theta p + 1) \frac{\mu_{\theta p}(x)}{\mu_{\theta p+1}(x)} - (p + 1) \frac{\mu_p(x)}{\mu_{p+1}(x)} \right].$$
1. Estimate $\mu_p(x)$ by a kernel estimator

$$\hat{\mu}_p(x) := \frac{\sum_{i=1}^n Y_i^p K_{h_n}(x - X_i)}{\sum_{i=1}^n K_{h_n}(x - X_i)}.$$  

The bandwidth $h_n$ selects the $X_i$’s close to $x$.

2. To deal with the more general situation

$$F(y \mid x) = (1 - y/g(x))^{\alpha(x)+1} \ell(x, (1 - y/g(x))^{-1}),$$

where $\ell(x, .)$ is a slowly-varying function at infinity, $p$ is replaced with a sequence $p_n \to \infty$. The high power $p_n$ gives more weight to the $Y_i$’s close to $g(x)$. 

Three different approaches to frontier estimation
We further assume a Hall model for the slowly-varying function: $\ell$ is supposed to be bounded on $\Omega \times [1, \infty)$ and

$$
\ell(x, z) = C(x) + D(x) z^{-\beta(x)} (1 + \delta(x, z))
$$

where all functions $C$, $D$ and $\beta$ are Lipschitzian. Moreover, for all $x \in \Omega$, $\delta(x, z) \to 0$ as $z \to \infty$.

**Theorem**

*Let $x \in \Omega$ such that $f_X(x) > 0$. Then, under some conditions on the $(h_n)$ and $(p_n)$ sequences,*

$$
v_n(x) = n^{1/2} h_n^{d/2} p_n^{(1-\alpha(x))/2} \left( \frac{\hat{g}_n(x)}{g(x)} - 1 \right) \xrightarrow{d} \mathcal{N} \left( 0, \frac{\|K\|_2^2 \, V(\alpha(x), \theta)}{f_X(x) \, C(x)} \right)
$$

*G, Guillou & Stupfler (JMVA, 2013)*
In the case of \(\gamma-\) Lipschizian frontier, choosing \(h_n \asymp n^{-1/(\gamma(\alpha(x)+1)+d)}\) and \(p_n \asymp n^{\gamma/(\gamma(\alpha(x)+1)+d)}\) yields a the rate of convergence is \(n^{\gamma/(\gamma(\alpha(x)+1)+d)}\), up to logarithmic factors.

Optimal \(L_1\) rate of convergence for sharp/non-sharp boundaries (\(\alpha(x) \geq 0\)) and \(\gamma-\) Lipschitzian frontiers, \(\gamma \in (0, 1]\).

Compared to Härdle, Park, Tsybakov (JMVA, 1995), the case of “super-sharp” boundaries is also possible: \(-1 < \alpha(x) < 0\). In this case, \(f(y|x) \rightarrow \infty\) as \(y \rightarrow g(x)\).
The estimation of the conditional tail-index $\alpha(x)$ is possible with similar techniques:

$$\alpha_n(x) = (p_n + 1) \left( \hat{g}_n(x) \frac{\hat{\mu}_{p_n}(x)}{\hat{\mu}_{p_n+1}(x)} - 1 \right)$$

An uniform almost sure consistency result is also available

G, Guillou & Stupfler (ESAIM, 2014):

**Theorem**

$$\sup_{x \in \Omega} |\hat{g}_n(x) - g(x)| = O \left( n^{-\gamma/((\bar{\alpha}+1)+d)} \right),$$

where $\bar{\alpha} = \sup_{x \in \Omega} \alpha(x)$. 

Three different approaches to frontier estimation
$Y|X = x$ is beta distributed. Best (left) and worst (right) results obtained over 500 replications.
Conclusion

Contributions

- **Extreme-value estimator** (smoothed, bias correction, sharp boundary, pointwise asymptotic normality)
- **Linear programming estimator** (smoothed, no partition of $\Omega$, sharp boundary, strong $L_1$– consistency)
- **High order moments estimator** (smoothed, no partition of $\Omega$, non-sharp boundary, pointwise asymptotic normality, strong $L_\infty$– consistency)

Further work

- Arbitrary smoothness ($\gamma > 1$),
- Adaptive choice of the tuning parameters (bandwidth, ...).