Stability and Sample-based Approximations of Composite Stochastic Optimization Problems

Darinka Dentcheva

STEVENS INSTITUTE of TECHNOLOGY
THE INNOVATION UNIVERSITY

This work is supported by the Office of Naval Research under grant no. N00014-21-1-2161.

Robustness and Resilience in Stochastic Optimization and Statistical Learning: Mathematical Foundations

May 23, 2022

Outline

Composite risk functionals

Motivation

$$\varrho[X] = \mathbb{E}\Big[f_1\Big(\mathbb{E}\big[f_2\big(\mathbb{E}[\,\cdots f_k(\mathbb{E}[f_{k+1}(X)], X)]\ \cdots, X\big)\big], X\Big)\Big],$$

X is an integrable random vector with values in $\mathcal{X} \subseteq \mathbb{R}^m$ and probability distribution P. $f_j : \mathbb{R}^{m_j} \times \mathbb{R}^m \to \mathbb{R}^{m_{j-1}}$, $j = 1, \ldots, k$, with $m_0 = 1$ and $f_{k+1} : \mathbb{R}^m \to \mathbb{R}^{m_k}$.

Example

The mean-semi-deviation of order $p \geq 1$ for a random variable X representing a loss is

$$\varrho[X] = \mathbb{E}[X] + \kappa\Big[\mathbb{E}\big[\big(\max\{0, X - \mathbb{E}[X]\}\big)^p\big]\Big]^{\frac{1}{p}},$$

where $\kappa \in [0, 1]$. We have $k = 2$, $m = 1$, and

$$\begin{aligned} f_1(\eta_1, x) &= x + \kappa\, \eta_1^{\frac{1}{p}}, \\ f_2(\eta_2, x) &= \big[\max\{0, x - \eta_2\}\big]^p, \\ f_3(x) &= x. \end{aligned}$$

Composite functionals in optimization

Composite functionals

$$\varrho[X] = \mathbb{E}[f_1(\mathbb{E}[f_2(\mathbb{E}[\ldots f_k(\mathbb{E}[f_{k+1}(X)], X)] \ldots, X)], X)]$$

Risk measures representable as optimal values of composite functionals

$$\theta[X] = \min_{u \in U} f_1(u, \mathbb{E}[f_2(u, X)])$$

$$\mathcal{S}[X] = \operatorname{argmin}_{u \in U} f_1(u, \mathbb{E}[f_2(u, X)])$$

where $U \subset \mathbb{R}^d$ is a nonempty closed set.

Optimized composite functionals

$$\vartheta[X] = \min_{u \in U} \varrho(u, X)$$

$$\varrho[u, X] = \mathbb{E}\left[f_1\left(u, \mathbb{E}[f_2(u, \mathbb{E}[\ldots f_k(u, \mathbb{E}[f_{k+1}(u, X)], X)] \ldots, X)], X\right)\right]$$

where $U \subset \mathbb{R}^d$ is a nonempty closed set.

Goal

- asymptotic behavior of sample based optimization problems with composite functionals;
- stability with respect to measure perturbations;
- bias of the estimators.

References

1. D. Dentcheva, Y. Lin, and S. Penev, "Stability and Sample-based Approximations of Composite Stochastic Optimization Problems," to appear in *Operations Research* (2022).
2. D.Dentcheva, Y. Lin, "Bias Reduction in Sample-Based Optimization," *SIAM J. Optimization* 32.1 (2022): 130-151.
3. D. Dentcheva, S. Penev, A. Ruszczyński: Statistical estimation of composite risk functionals and risk optimization problems, *Annals of the Institute of Statistical Mathematics* 2017.

For a measure $Q \in \mathcal{P}(\mathcal{X})$, we define

$$\begin{aligned} \bar{f}_j^Q(u,\eta_j) &= \int_{\mathcal{X}} f_j(u,\eta_j,x)\, Q(dx), \quad j = 1,\cdots,k \\ \bar{f}_{k+1}^Q(u) &= \int_{\mathcal{X}} f_{k+1}(u,x)\, Q(dx) \end{aligned} \tag{1}$$

Assumption: A compact set $\mathcal{U}$ exists such that $S \subset \operatorname{int}\mathcal{U} \subset U$.
Compact convex sets $I_1 \subset \mathbb{R}^{m_1}, \cdots, I_k \subset \mathbb{R}^{m_k}$ are such that

$$\bar{f}_{j+1}^P(\mathcal{U}, I_{j+1}) \subset \operatorname{int}(I_j), \quad \bar{f}_{k+1}^P(\mathcal{U}) \subset \operatorname{int}(I_k);$$

$$I = I_1 \times I_2 \times \cdots I_k, \quad d = m_0 + m_1 + \cdots + m_k.$$

Approach: Embed the functions into the space

$$\mathcal{H} = \mathcal{C}_1(\mathcal{U} \times I_1) \times \mathcal{C}_{m_1}(\mathcal{U} \times I_2) \times \cdots \times \mathcal{C}_{m_{k-1}}(\mathcal{U} \times I_k) \times \mathcal{C}_{m_k}(\mathcal{U})$$

where $\mathcal{C}_{m_{j-1}}$ is the space of $\mathbb{R}^{m_{j-1}}$-valued continuous function on $\mathcal{U} \times I_j$.
We analyze the vector function $\bar{\mathbf{f}}^{(Q^1..Q^{k+1})} \in \mathcal{H}$:

$$\bar{\mathbf{f}}^{(Q^1..Q^{k+1})}(u,\eta) = (\bar{f}_1^{Q^1}(u,\eta_1), \bar{f}_2^{Q^2}(u,\eta_2), \cdots, \bar{f}_k^{Q^k}(u,\eta_k), \bar{f}_{k+1}^{Q^{k+1}}(u))^\top$$

Empirical estimators

Given $\{X_i\}_{i\geq 1}$ i.i.d random variables with measure P, let P_N be the empirical measure.

The empirical estimators

$$\varrho_E^{(N)}[u,X] = \sum_{i_0=1}^{N} \frac{1}{N}\Big[f_1(u, \sum_{i_1=1}^{N} \frac{1}{N}\Big[f_2(u, \sum_{i_2=1}^{N} \frac{1}{N}\Big[\cdots$$

$$f_k\Big(u, \sum_{i_k=1}^{N} \frac{1}{N} f_{k+1}(u, X_{i_k}), X_{i_{k-1}}\Big)\Big]\cdots, X_{i_1})\Big], X_{i_0})\Big]$$

$$\vartheta_E^{(N)}[X] = \min_{u\in U} \; \varrho_E^{(N)}[u,X]$$

$$S_E^{(N)}[X] = \big\{u \in \mathcal{U} : \; \vartheta_E^{(N)} = \varrho_E^{(N)}\big\}$$

We use the entire sample for estimating each expectation

For two sets, $A, B \subset \mathbb{R}^n$, the one-sided distance of A to B is

$$\mathrm{d}(A,B) = \sup_{x\in A} d(x,B) = \sup_{x\in A}\inf_{y\in B}\|x-y\|.$$

The Pompeiu-Hausdorff distance between the sets is defined as

$$\mathrm{D}(A,B) = \max\big\{\mathrm{d}(A,B), \mathrm{d}(B,A)\big\}$$

- Assume $f_j(u,\eta_j,\cdot)$, $j=1,\dots,k$, and $f_{k+1}(u,\cdot)$ are uniformly bounded for all $u\in\mathcal{U}$ and for all $\eta_j\in I_j$ by a P-integrable function $g:\mathbb{R}^m\to\mathbb{R}$.
- Suppose $S_E^{(N)}[X]\neq\emptyset$ for N large enough.

Then $\varrho_E^{(N)}[u,X] \xrightarrow[N\to\infty]{a.s.} \varrho[u,X]$ for every $u\in\mathcal{U}$,
$\vartheta_E^{(N)} \xrightarrow[N\to\infty]{a.s.} \vartheta$, and $\mathrm{d}(S_E^{(N)},S) \xrightarrow[N\to\infty]{a.s.} 0$.
If the true problem has a unique solution, then $\mathrm{D}(S_E^{(N)},S) \xrightarrow[N\to\infty]{a.s.} 0$.

Given a point $\hat{u} \in \mathcal{U}$, we define

$$\bar{f}^P_{k+1}(u) = \int_{\mathcal{X}} f_{k+1}(u,x)\,P(dx), \quad \mu_{k+1} = \bar{f}^P_{k+1}(\hat{u})$$

$$\bar{f}^P_j(u,\eta_j) = \int_{\mathcal{X}} f_j(u,\eta_j,x)\,P(dx), \quad j = 1,\dots,k,$$

$$\mu_j = \bar{f}^P_j(\hat{u},\mu_{j+1}), \quad j = 1,\dots,k;$$

$$\tilde{\mathcal{H}} = \mathcal{C}^{(0,1)}_1(U \times I_1) \times \mathcal{C}^{(0,1)}_{m_1}(U \times I_2) \times \dots \mathcal{C}^{(0,1)}_{m_{k-1}}(U \times I_k) \times \mathcal{C}_{m_k}(U),$$

where $\mathcal{C}^{(0,1)}_{m_{j-1}}(U \times I_j)$ is the space of $\mathbb{R}^{m_{j-1}}$-valued continuous functions on $U \times I_j$, which are differentiable with respect to the second argument with continuous derivatives on $U \times I_j$.

The Jacobian of $f_j(u,\eta_j,x)$ w.r.t. the second argument is denoted $f_j'(u,\eta_j,x)$. For every direction $d \in \tilde{\mathcal{H}}$, we define recursively:

$$\xi_{k+1}(d) = d_{k+1},$$

$$\xi_j(d) = \int_{\mathcal{X}} f_j'(\hat{u},\mu_{j+1},x)\xi_{j+1}(d)\,P(dx) + d_j\big(\mu_{j+1}\big), \quad j = k,k-1,\dots,1.$$

Assumptions

(a1) The optimal solution $\hat{u}$ of the composite problem is unique.

(a2) The functions $f_j(\cdot,\cdot,x)$, $j = 1..k$, and $f_{k+1}(\cdot,x)$ are Lipschitz continuous for every $x \in \mathcal{X}$ with square integrable constants.

(a3) The functions $f_j(u,\cdot,x)$, $j = 1..k$, are continuously differentiable for every $x \in \mathcal{X}$, $u \in \mathcal{U}$; their derivatives are continuous with respect to the first two arguments.

It holds

$$\sqrt{N}\big(\vartheta^{(N)} - \vartheta\big) \xrightarrow[N\to\infty]{d} \xi_1(\hat{u}, W),$$

$W(\cdot) = \big(W_1(\cdot), \ldots, W_k(\cdot), W_{k+1}\big)$ is a zero-mean Brownian process on I; $W_j(\cdot)$ is a Brownian process of dimension m_{j-1} on I_j, $j = 1, \ldots, k$, and W_{k+1} is an m_k-dimensional normal vector.

Covariance structure

The covariance function of $W(\cdot)$ has the form

$$
\begin{aligned}
&\operatorname{cov}\big[W_i(\eta_i), W_j(\eta_j)\big] = \\
&\qquad \int_{\mathcal{X}} \big[f_i(\hat{u}, \eta_i, x) - \bar{f}_i(\hat{u}, \eta_i)\big]\big[f_j(\hat{u}, \eta_j, x) - \bar{f}_j(\hat{u}, \eta_j)\big]^\top P(dx), \\
&\qquad\qquad \eta_i \in I_i,\ \eta_j \in I_j,\ i, j = 1, \dots, k \\
&\operatorname{cov}\big[W_i(\eta_i), W_{k+1}\big] = \\
&\qquad \int_{\mathcal{X}} \big[f_i(\hat{u}, \eta_i, x) - \bar{f}_i(\hat{u}, \eta_i)\big]\big[f_{k+1}(\hat{u}, x) - \bar{f}_{k+1}(\hat{u})\big]^\top P(dx), \\
&\qquad\qquad \eta_i \in I_i,\ i = 1, \dots, k \\
&\operatorname{cov}\big[W_{k+1}, W_{k+1}\big] = \\
&\qquad \int_{\mathcal{X}} \big[f_{k+1}(\hat{u}, x) - \bar{f}_{k+1}(\hat{u})\big]\big[f_{k+1}(\hat{u}, x) - \bar{f}_{k+1}(\hat{u})\big]^\top P(dx).
\end{aligned}
$$

Assumption A

The sequence of measures $\{\mu_N\}$ are independent of P_N, normalized, and satisfying

- μ_N converges weakly to point mass $\delta(0)$ when $N \to \infty$;
- $\int_{\mathbb{R}^m} \|z\| \, d\mu_N(z)$ is finite and $\lim_{r\to\infty} \lim_{N\to\infty} \int_{\|z\|>r} \|z\| \, d\mu_N(z) = 0$.

The smooth estimators for the expectation of a function $g : \mathbb{R}^m \to \mathbb{R}$ based on the sequence $\{\mu_N\}$ is defined as follows:

$$\frac{1}{N} \sum_{i=1}^{N} \int_{\mathbb{R}^m} g(X_i + z) \, d\mu_N(z).$$

A function $g : \mathbb{R}^m \to \mathbb{R}$ admits a modulus of continuity $w : \mathbb{R}_+ \to \mathbb{R}$ if

- $\lim_{t\downarrow 0} w(t) = w(0) = 0$;
- for all $x, z \in \mathbb{R}^m$, it holds $|g(x) - g(x')| \le w(\|x - x'\|)$;

SLLN for smoothed estimators

Let an index set $J \subseteq \{1, 2, \ldots, k+1\}$, and a sequence of measures $\{\mu_N\}$ satisfying Assumption A be given. Assume the following conditions:

- For $j \notin J$, the functions $f_j(u, \eta_j, \cdot)$ as well as $f_{k+1}(u, \cdot)$ for $k+1 \notin J$ are uniformly bounded for all $(u, \eta) \in \mathcal{U} \times I$ by a P-integrable function.
- For $j \in J$, the functions $f_{j,i}(u, \eta_j, \cdot)$, $i = 1, \ldots, m_{j-1}$ for all $(u, \eta_j) \in \mathcal{U} \times I_j$; if $k+1 \in J$, then $f_{k+1,i}(u, \cdot)$, $i = 1, \ldots, m_k$ for all $u \in \mathcal{U}$ admit a modulus of continuity.

Then $\varrho_\mu^{(N,J)}[u, X] \xrightarrow[N\to\infty]{a.s.} \varrho[u, X]$ for every $u \in \mathcal{U}$, $\vartheta_\mu^{(N,J)} \xrightarrow[N\to\infty]{a.s.} \vartheta$, and $\mathrm{d}(S_\mu^{(N,J)}, S) \xrightarrow[N\to\infty]{a.s.} 0$.

Kernel estimators

$$\frac{1}{Nh_N^m}\sum_{i=1}^{N}\int_{\mathbb{R}^m} g(x)K\Big(\frac{x-X_i}{h_N}\Big)\,dx,$$

where $h_N > 0$ is a smoothing parameter such that $\lim_{N\to\infty} h_N = 0$.

Assumptions:

(k1) The kernel K of order $s > 1$ is a density function with respect to the Lebesgue measure satisfying $\int\limits_{\mathbb{R}^m} y_l^j K(y)dy = 0$ for $l = 1, \dots, m$, $j = 1, \dots, \lfloor s \rfloor$ with $\lfloor s \rfloor$ being the largest integer smaller than s.

(k2) The s-th moment $\int\limits_{\mathbb{R}^m} \|y\|^s K(y)dy$ is finite.

SLLN for kernel-based estimators

Suppose all functions $f_{j,i} : \mathbb{R}^m \to \mathbb{R}$, $j \in J$, $i = 1, \dots, m_{j-1}$, admit a modulus of continuity with respect to the last argument, which does not depend on u or η. Then $\varrho_K^{(N)}[u, X] \xrightarrow[N\to\infty]{a.s.} \varrho[u, X]$ for every $u \in \mathcal{U}$, $\vartheta_K^{(N)} \xrightarrow[N\to\infty]{a.s.} \vartheta$, and $\mathrm{d}(S_K^{(N)}, S) \xrightarrow[N\to\infty]{a.s.} 0$.

Let an index set $J \subseteq \{1, 2, \ldots, k+1\}$.

- The sequence of measures $\{\mu_N\}$ satisfies Assumption A.
- For $j = 1, \ldots, k$, $f_j(u, \eta_j, \cdot)$ and $f_{k+1}(u, \cdot)$ are continuous and uniformly bounded by a P-integrable function $g_j : \mathbb{R}^m \to \mathbb{R}$ for all $u \in \mathcal{U}$ and for all $\eta_j \in I_j$.

Then $\varrho_\mu^{(N,J)}[u, X] \xrightarrow[N\to\infty]{p} \varrho[u, X]$ for every $u \in \mathcal{U}$, $\vartheta_\mu^{(N,J)} \xrightarrow[N\to\infty]{p} \vartheta$, and $\mathrm{d}(S_\mu^{(N,J)}, S) \xrightarrow[N\to\infty]{p} 0$.
Additionally, if the true problem has a unique solution, then $\mathrm{D}(S_\mu^{(N,J)}, S) \xrightarrow[N\to\infty]{p} 0$ as well.

Wavelet-based density estimator

$$\widetilde{d}_N(x) = \sum_{\ell=-\infty}^{\infty} 2^{-j/2}\frac{1}{N}\sum_{i=1}^{N}\phi(2^j X_i - \ell)\}\phi_{j\ell}(x)$$

$$\vartheta_w^{(N)} = \min_{u\in\mathcal{U}} \int_{\mathcal{X}} f(u,x)\tilde{d}_{N,j}(x)\,dx = \min_{u\in\mathcal{U}} \frac{2^j}{N}\sum_{i=1}^{N}\int_{\mathcal{X}} f(u,x)K(2^j X_i, 2^j x)\,dx,$$

with the generalized kernel $K(y,x) = \sum_{\ell\in\mathbb{Z}}\phi(y-\ell)\phi(x-\ell)$.

The function $\phi_{j\ell}(x) = 2^{j/2}\phi(2^j x - \ell)$, where $\phi(x)$ is right-continuous, non-negative, with finite variation, and with compact support in $[-a, a]$, $1/2 \le a < \infty$, and satisfies the conditions

(w1) $\sum_{\ell=-\infty}^{\infty}\phi(x-\ell) = 1$ for all $x \in \mathbb{R}$

(w2) $x - \sum_{\ell=-\infty}^{\infty}\ell\phi(x-\ell) = 0$ for all $x \in \mathbb{R}$.

Order of the optimal resolution level $j^* \propto \log_2 N$, with $j^* = \log_2 N/5$ giving very good performance over a wide variety of density classes.

Probability metrics

The set of all Lipschitz-continuous and bounded functions on $\text{conv}(\mathcal{X})$

$$\mathfrak{F} = \Big\{g : \big|g(x) - g(x')\big| \leq \|x - x'\|, \quad \sup_{x \in \text{conv}(\mathcal{X})} \big|g(x)\big| \leq 1\Big\}.$$

The metric $\beta(Q, Q')$ metrizes the weak convergence on $\mathcal{P}(\mathcal{X})$

$$\beta(Q, Q') = \sup_{g \in \mathfrak{F}} \left| \int_{\mathcal{X}} g(x) dQ(x) - \int_{\mathcal{X}} g(x) dQ'(x) \right|.$$

For a function $w : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\lim_{t \downarrow 0} w(t) = w(0) = 0$, we define

$$\tilde{\mathfrak{F}} = \Big\{g : \text{conv}(\mathcal{X}) \to \mathbb{R} : \big|g(x) - g(x')\big| \leq w\big(\|x - x'\|\big)\Big\}$$

$$\tilde{\beta}(Q, Q') = \sup_{g \in \tilde{\mathfrak{F}}} \left| \int_{\mathcal{X}} g(x) Q(dx) - \int_{\mathcal{X}} g(x) Q'(dx) \right|$$

- $\tilde{\mathfrak{F}}$ consists of real-valued functions that admit joint modulus of continuity.
- If $w(t) = Lt$, then $\mathfrak{F} \subset \tilde{\mathfrak{F}}$, hence, every sequence of measures converging with respect to $\tilde{\beta}$ also converges with respect to β.

Measure Perturbations

Given a sequence of measures $\{Q_N^j\}$, $j = 1, \ldots k+1$, $N \in \mathbb{N}$.

Approximate (measure-perturbed) problem

$$\varrho^{(Q_N^1..Q_N^{k+1})}[u, X] = \bar{f}_1^{Q_N^1}\Big(u, \bar{f}_2^{Q_N^2}\big(u, \cdots \bar{f}_k^{Q_N^k}(u, \bar{f}_{k+1}^{Q_N^{k+1}}(u)) \cdots\big)\Big)$$

$$\vartheta^{(Q_N^1..Q_N^{k+1})} = \min_{u \in U} \varrho^{(Q_N^1..Q_N^{k+1})}[u, X]$$

$$S^{(Q_N^1..Q_N^{k+1})} = \Big\{u \in \mathcal{U} : \; \varrho^{(Q_N^1..Q_N^{k+1})}[u, X] = \vartheta^{(Q_N^1..Q_N^{k+1})}\Big\}$$

Assume that $Q_N^j \xrightarrow[N\to\infty]{w} P$, $j = 1, \dots k+1$ and $S^{(Q_N^1 .. Q_N^{k+1})} \neq \emptyset$ for N large enough. Suppose one of the following conditions:

(a) the functions $f_j(u, \eta_j, \cdot)$ and $f_{k+1}(u, \cdot)$ belong to $\mathfrak{F}$ for all $(u, \eta) \in U \times I$, $j = 1, \dots, k$;

(b) the functions $f_j(u, \eta_j, \cdot)$ and $f_{k+1}(u, \cdot)$ belong to $\tilde{\mathfrak{F}}$ for all $(u, \eta) \in U \times I$, $j = 1, \dots, k$ and the sequences of measures Q_N^j satisfy uniform integrability condition, e.g.,

$$\lim_{N\to\infty} \int_{\mathcal{X}} \|x\| \; Q_N^j(dx) = \int_{\mathcal{X}} \|x\| \; P(dx) \; < \infty$$

Then $\varrho^{(Q_N^1 .. Q_N^{k+1})}[u, X] \xrightarrow[N\to\infty]{} \varrho[u, X]$ for every $u \in \mathcal{U}$, $\vartheta^{(Q_N^1 .. Q_N^{k+1})} \xrightarrow[N\to\infty]{} \vartheta$, and $\mathrm{d}(S^{(Q_N^1 .. Q_N^{k+1})}, S) \xrightarrow[N\to\infty]{} 0$.

Additionally, if S is a singleton, then the Pompeiu–Hausdorff distance $\mathrm{D}\big(S^{(Q_N^1 .. Q_N^{k+1})}, S\big)$ converges to zero.

Basic Single-Layer Problem and Bias

$$\theta = \min_{u \in \mathcal{U}} \mathbb{E}\big[F(u, X)\big].$$

An i.i.d. sample $X_1, X_2, \ldots, X_N$ of the probability measure P is given.

Approximations

- Sample Average Approximation $\theta_{SAA}^{(N)} = \min_{u \in U} \frac{1}{N} \sum_{i=1}^{N} F(u, X_i)$
- Kernel-based Approximation $\theta_{K}^{(N)} = \min_{u \in U} \frac{1}{N} \sum_{i=1}^{N} \int_{\mathbb{R}^m} F(u, z) K\Big(\frac{z - X_i}{h_N}\Big)$

Downward bias

$$\theta = \min_{u \in U} \mathbb{E}[F(u, X)] = \min_{u \in U} \mathbb{E}\Big[\frac{1}{N} \sum_{i=1}^{N} F(u, X_i)\Big]$$

$$\geq \mathbb{E}\Big[\min_{u \in U} \frac{1}{N} \sum_{i=1}^{N} F(u, X_i)\Big] = \mathbb{E}[\theta_{SAA}].$$

Smoothing reduces the bias

Assume that

- the function $F(\hat{u}, \cdot)$ admits a modulus of continuity of Hölder type.
- $F(u, \cdot)$ is convex for any fixed $u \in U$,
- the kernel K satisfies (k1)-(k2).

Then constants $L > 0, \alpha > 0$ and $h_N^* > 0$ exist, such that for $h_N \in (0, h_N^*)$ the following relations hold:

$$
\begin{aligned}
&|\mathbb{E}[\theta_K^{(N)}] - \theta| \leq |\mathbb{E}[\theta_{SAA}^{(N)}] - \theta| \\
&\mathbb{E}|\theta_K^{(N)} - \mathbb{E}[\theta_K^{(N)}]| \leq \mathbb{E}|\theta_{SAA}^{(N)} - \mathbb{E}[\theta_{SAA}^{(N)}]| + Lh_N^\alpha \\
&\left(\mathbb{E}[(\theta_K^{(N)} - \mathbb{E}[\theta_K^{(N)}])^2]\right)^{\frac{1}{2}} \leq \left(\mathbb{E}(\theta_{SAA}^{(N)} - \mathbb{E}[\theta_{SAA}^{(N)}])^2\right)^{\frac{1}{2}} + Lh_N^\alpha \\
&\left(\mathbb{E}\Big[(\theta_K^{(N)} - \theta)^2\Big]\right)^{\frac{1}{2}} \leq \left(\mathbb{E}\Big[(\theta_{SAA}^{(N)} - \theta)^2\Big]\right)^{\frac{1}{2}} + Lh_N^\alpha.
\end{aligned}
$$

Applications: LASSO problem

Consider an outcome Y and explanatory variables comprised in an m-dimensional vector X. The objective of LASSO is to solve

$$\min_{\beta_0,\beta\in\mathbb{R}^m} \sum_{i=1}^{N} \left(y_i - \beta_0 - \beta^\top X^i\right)^2 \text{ subject to } \sum_{j=1}^{m} |\beta_j| \le t.$$

The smoothed (with respect to the data) objective function

$$\sum_{i=1}^{N} \int (\tilde{\beta}^\top \tilde{x})^2 K\Big(\frac{\tilde{X}^i - \tilde{x}}{h_N}\Big) \frac{1}{h_N^{m+1}}\, d\tilde{x},$$

where $\tilde{\beta} = (\beta, -1, \beta_0)$ and $\tilde{x} = (x, y, 1)$ with $\tilde{X}^i = (X^i, Y^i, 1)$, $i = 1..N$. The Bias Reduction Theorem apply.

Corollary Smoothing with the normal kernel with covariance A is equivalent to

$$\min_{\beta\in\mathbb{R}^m} \sum_{i=1}^{N} \left(y_i - \beta^\top X^i\right)^2 + h_N^2 \|(-1, \beta)\|_A^2 \text{ subject to } \sum_{j=1}^{m} |\beta_j| \le t.$$

Given a weight $\kappa \in (0, 1)$ for the Average Value-at-Risk and a probability level $\beta \in (0, 1)$, we define

$$F(u, \eta, x) = -\kappa \langle u, x \rangle + (1 - \kappa)\Big(-\eta + \frac{1}{\beta} \max \big\{0, \eta - \langle u, x \rangle \big\}\Big),$$

$$U = \{u \in \mathbb{R}^m : \sum_{i=1}^{m} u_i = K,\ l_i \leq u_i \leq b_i\}$$

The AVaR_β-portfolio optimization problem

$$\min_{u \in U, \eta \in \mathbb{R}} \mathbb{E}\big[F(u, \eta, X)\big].$$

The modulus of continuity is $w(t) = c(\kappa + (1 - \kappa)/\beta)t$.

The Consistency Theorem and the Bias Reduction Theorem apply.

Bias in composite optimization

$$\vartheta = \min_{u \in U} f_1(u, \mathbb{E}[f_2(u, X)])$$

$$\vartheta_E^{(N)} = \min_{u \in U} f_1\left(u, \frac{1}{N}\sum_{i=1}^{N} f_2(u, X_i)\right)$$

where $U \subset \mathbb{R}^n$ is a nonempty compact set.

Downward bias

If the function $f_1(u, \cdot)$ is concave for all $u \in U$, then $\mathbb{E}[\vartheta_E^{(N)}] \leq \vartheta$.

Higher-order measures of risk

$$\varrho[Y] = \min_{z \in \mathbb{R}} \left\{-z + \frac{1}{\alpha}\left[\mathbb{E}(\max(0, z - Y)^q)\right]^{1/q}\right\}.$$

$$\varrho[Y] = \min_{z\in\mathbb{R}} \left\{ -z + \frac{1}{\alpha} [\mathbb{E}(\max(0, z - Y)^q)]^{1/q} \right\}.$$

Given a weight $\kappa \in (0, 1)$ for and a parameter $\beta \in (0, 1)$, we define

$$f_1(u, \eta, x) = (1-\kappa)\eta_1 + \kappa\big(- u_0 + \frac{1}{\alpha}\eta_2^{1/q}\big),$$

$$f_2(u, x) = \begin{pmatrix} -\langle x, u\rangle \\ [\max(0, u_0 - \langle x, u\rangle)]^q \end{pmatrix},$$

$$U = \Big\{ u \in \mathbb{R}^m : \sum_{i=1}^{m} u_i = K,\ l_i \le u_i \le b_i \Big\}$$

Optimization problem using the higher order risk measure

$$\min_{u\in U} \mathbb{E}\Big[f_1\Big(u, \mathbb{E}[f_2(u, z, X)], X\Big)\Big].$$

The Consistency Theorem and the Bias Reduction Theorem apply.

Smoothing reduces negative bias in composite optimization

Given an index set $J \subset 1, \ldots k+1$, let $l = \max\{j : j \in J\}$. Assume that

- for all $j \in J$, the functions f_j are convex with respect to the last argument and the functions $f_j(u, \cdot, x)$, $j = 1, \ldots \ell - 1$ are monotonically non-decreasing;
- for each $j \in J$, the function f_j has a Hölder modulus of continuity $w_j^x(t) = \ell_j^x t^{\beta_j}$ with respect to the last argument and the function $f_{j-1}(u, \cdot, x)$, $j > 1$, has a modulus of continuity $w_j^\eta(t) = \ell_j t^{\alpha_j}$.
- The order of the kernel is at least $\max_{j \in J} \beta_j$.

Then constants $L > 0, \alpha > 0$ and $h_N^* > 0$ exist, such that for all $h_N \in (0, h_N^*)$

$$\left|\mathbb{E}[\vartheta_K^{(N,J)}] - \vartheta\right| \leq \left|\mathbb{E}[\vartheta_E^{(N)}] - \vartheta\right|.$$

$$\mathbb{E}\left|\vartheta_K^{(N,J)} - \mathbb{E}[\vartheta_K^{(N,J)}]\right| \leq \mathbb{E}\left|\vartheta_E^{(N)} - \mathbb{E}[\vartheta_E^{(N)}]\right| + Lh_N^\alpha,$$

$$\left(\mathbb{E}[(\vartheta_K^{(N,J)} - \mathbb{E}[\vartheta_K^{(N,J)}])^2]\right)^{\frac{1}{2}} \leq \left(\mathbb{E}(\vartheta_E^{(N)} - \mathbb{E}[\vartheta_E^{(N)})^2\right)^{\frac{1}{2}} + Lh_N^\alpha$$

$$\left(\mathbb{E}\left[\left(\vartheta_K^{(N,J)} - \vartheta\right)^2\right]\right)^{\frac{1}{2}} \leq \left(\mathbb{E}\left[\left(\vartheta_E^{(N)} - \vartheta\right)^2\right]\right)^{\frac{1}{2}} + Lh_N^\alpha.$$

We estimate the risk measure

$$\varrho[X] = \min_{u\in\mathbb{R}} \left\{ u + \frac{1}{\alpha}[\mathbb{E}(\max(0, X-u)^q)]^{1/q} \right\},$$

with parameters $q = 1$ or $q = 2$ and $\alpha = 0.05$ or $\alpha = 0.2$

- For the kernel estimator, we have experimented with the Gaussian kernel, the uniform kernel $K(x) = \frac{1}{2h_N}$ with support on $|x| \leq h_N$, and the Epanechnikov kernel $K(x) = \frac{3}{4}(1-x^2)$ on the support: $|x| \leq h_N$.
- First sequence of experiments, we have X_i, $i = 1, \cdots, N$ from a normal distribution $\mathcal{N}(10, 3)$ and have selected values of $\alpha = 0.05$ and $\alpha = 0.2$. The optimal $u^* = 14.5048$ is determined by numerical integration resulting in the “true" $\vartheta_0 = 15.5163$ as the estimated risk. The bandwidth recommended for a kernel density estimator for the normal distribution is $1.06\hat{\sigma}N^{-\frac{1}{5}}$ with $\hat{\sigma}$ being the sample variance.
- In a second series of experiments, we used the t distribution with various degrees of freedom ν such as 6, 8 and 60, with the data shifted to have the same mean of 10 as the normal simulated data before.

Bias

N \ h_N	M	$1.06\hat{\sigma}N^{-\frac{1}{5}}$	0.5	0.35	0.2	0.05	SAA
100	5000	0.0489	-0.0105	-0.0395	-0.0601	-0.071	-0.0719
200	5000	0.0495	0.018	-0.0091	-0.0277	-0.0374	-0.0382
500	5000	0.0412	0.0353	0.0095	-0.0076	-0.0159	-0.0166

Variance

N \ h_N	M	$1.06\hat{\sigma}N^{-\frac{1}{5}}$	0.5	0.35	0.2	0.05	SAA
100	5000	0.1713	0.1723	0.177	0.1804	0.1824	0.1826
200	5000	0.087	0.0866	0.089	0.0908	0.0917	0.0918
500	5000	0.0364	0.0358	0.0368	0.0375	0.0379	0.0379

Numerical Results for $AVaR_{0.05}$, Epanechnikov kernel

Bias

N \ h_N	M	$1.06\hat{\sigma}N^{-\frac{1}{5}}$	0.5	0.35	0.2	0.05	SAA
100	500	-0.0135	-0.0496	-0.0679	-0.0807	-0.0873	-0.0878
200	500	0.0118	-0.0074	-0.024	-0.0354	-0.0414	-0.0419
500	500	0.0181	0.0146	-0.0009	-0.0112	-0.016	-0.0163

Variance

N \ h_N	M	$1.06\hat{\sigma}N^{-\frac{1}{5}}$	0.5	0.35	0.2	0.05	SAA
100	500	0.1832	0.1763	0.1792	0.1816	0.183	0.1832
200	500	0.0889	0.0887	0.0902	0.0912	0.0917	0.0918
500	500	0.0398	0.0394	0.04	0.0404	0.0406	0.0406

Numerical Results for $\mathrm{AVaR}_{0.2}$, uniform kernel

Bias

N \ h_N	M	$1.06\hat{\sigma}N^{-\frac{1}{5}}$	0.5	0.35	0.2	0.05	SAA
100	500	0.0324	-0.0057	-0.0237	-0.0359	-0.042	-0.0425
200	500	0.0448	0.0243	0.0071	-0.0043	-0.0098	-0.0103
500	500	0.0328	0.0289	0.0118	0.0007	-0.0045	-0.0163

Variance

N \ h_N	M	$1.06\hat{\sigma}N^{-\frac{1}{5}}$	0.5	0.35	0.2	0.05	SAA
100	500	0.0701	0.0684	0.0693	0.0699	0.0702	0.0702
200	500	0.0342	0.0332	0.0336	0.0339	0.034	0.034
500	500	0.0153	0.0149	0.0151	0.0152	0.0153	0.0153

Numerical Results for $AVaR_{0.2}$, Epanechnikov kernel

Bias

N \ h	M	$1.06\hat{\sigma}N^{-\frac{1}{5}}$	0.5	0.35	0.2	0.05	SAA
100	500	0.0035	-0.0198	-0.0303	-0.0384	-0.0422	-0.0425
200	500	0.0233	0.0111	0.0111	-0.0066	-0.01	-0.0103
500	500	0.0179	0.0155	0.0052	-0.0015	-0.0046	-0.0163

Variance

N \ h	M	$1.06\hat{\sigma}N^{-\frac{1}{5}}$	0.5	0.35	0.2	0.05	SAA
100	500	0.0701	0.0691	0.0693	0.07	0.0702	0.0702
200	500	0.0341	0.0336	0.0336	0.0339	0.034	0.034
500	500	0.0153	0.0151	0.0152	0.0153	0.0153	0.0153

Numerical results for $\mathcal{N}(10, 3)$, $q = 2$, and $\alpha = 0.05$

Uniform Kernel

N	Kernel Bias	Empirical Bias	Kernel Variance	Empirical Variance
100	-0.6095	-1.1896	0.5893	0.5754
200	-0.3930	-0.7891	0.5132	0.5350
500	-0.1655	-0.3236	0.3482	0.4099

Epanechnikov Kernel

N	Kernel Bias	Empirical Bias	Kernel Variance	Empirical Variance
100	-0.7254	-1.1896	0.5813	0.5754
200	-0.4852	-0.7891	0.5168	0.5350
500	-0.2164	-0.3236	0.3641	0.4099

Gaussian Kernel

N	Kernel Bias	Empirical Bias	Kernel Variance	Empirical Variance
100	-0.6095	-1.1896	0.5893	0.5754
200	-0.3930	-0.7891	0.5132	0.5350
500	-0.1655	-0.3236	0.3482	0.4099

Numerical results for $\mathcal{N}(10, 3)$, $q = 2$, and $\alpha = 0.05$

Wavelet-based estimator

N	Wavelet Bias	Empirical Bias	Wavelet Variance	Empirical Variance
100	-0.6430	-1.1668	0.6054	0.6375
200	-0.3728	-0.7677	0.4879	0.5382
500	-0.1016	-0.2996	0.2842	0.3525

Numerical results for the t distribution

Uniform kernel

N	dg	Kernel Bias	Empirical Bias	Kernel Variance	Empirical Variance
100	6	-1.9800	-2.1343	1.3440	1.3150
200	6	-1.4528	-1.5649	1.5973	1.5886
500	6	-0.7694	-0.7952	1.6350	1.6624
100	8	-1.4044	-1.5452	1.2057	1.1805
200	8	-0.9433	-1.0468	1.2299	1.2207
500	8	-0.4875	-0.5126	1.0281	1.0460
100	60	-0.6193	-0.7367	0.2529	0.2457
200	60	-0.3776	-0.4642	0.2168	0.2158
500	60	-0.1513	-0.1789	0.1687	0.1768

Epanechnikov kernel

N	dg	Kernel Bias	Empirical Bias	Kernel Variance	Empirical Variance
100	6	-2.0119	-2.1343	1.3370	1.3150
200	6	-1.4790	-1.5649	1.5954	1.5886
500	6	-0.7782	-0.7952	1.6435	1.6624
100	8	-1.4336	-1.5452	1.1996	1.1805
200	8	-0.9675	-1.0468	1.2299	1.2207
500	8	-0.4960	-0.5126	1.0336	1.0460
100	60	-0.6436	-0.7367	0.2510	0.2457
200	60	-0.3979	-0.4642	0.2166	0.2158
500	60	-0.1606	-0.1789	0.1710	0.1768

Wavelet-based estimator

N	df	Wavelet Bias	Empirical Bias	Wavelet Variance	Empirical Variance
100	6	-1.6239	-2.1477	1.3681	1.4114
200	6	-1.2090	-1.5892	1.4265	1.4979
500	6	-0.5870	-0.7453	1.9387	2.1290
100	8	-1.0266	-1.5532	0.9092	0.9622
200	8	-0.6814	-1.0694	0.9434	1.0175
500	8	-0.3029	-0.480	1.0214	1.1519
100	60	-0.2176	-0.7692	0.2182	0.2506
200	60	-0.0788	-0.5058	0.1745	0.2171
500	60	0.0490	-0.2092	0.0935	0.1366