On Statistical Inference for Optimization with Composite Risk Functionals

Darinka Dentcheva

Stevens Institute of Technology Hoboken, New Jersey, USA

joint work with

Yang Lin, Gregory J. Stock

Multi-Stage Stochastic Optimization for Clean Energy Transition

September, 2024 Oaxaca, Mexico

Outline

1. Composite Risk functionals. Definitions and Framework
2. Higher Order Inverse Risk Measures
3. Central Limit Theorem for composite risk functionals
4. Optimized composite risk functionals
5. Multivariate Extension
6. Applications

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 domain $\mathcal{X} \subseteq \mathbb{R}^m$ and probability distribution P. $f_j : \mathbb{R}^{m_j} \times \mathbb{R}^m \to \mathbb{R}^{m_{j-1}}$, $j = 1, \dots, k$, 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 Risk Functionals in Optimization

Composite Functionals

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

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 compact set.

Optimized composite functionals

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

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

- D. Dentcheva, S. Penev, A. Ruszczyński: Statistical estimation of composite risk functionals and risk optimization problems, Annals of the Institute of Statistical Mathematics 2017

The empirical estimator

Given $\{X_i\}_{i\geq 1}$ i.i.d random variables with probability measure P, we denote by P_n the empirical measure: $P_n = \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}$; The empirical estimator has the form:

- Composite Functionals

$$\varrho^{(n)} = \sum_{i_0=1}^{n}\frac{1}{n}\Big[f_1\Big(\sum_{i_1=1}^{n}\frac{1}{n}\big[f_2\big(\sum_{i_2=1}^{n}\frac{1}{n}[\cdots f_k(\sum_{i_k=1}^{n}\frac{1}{n}f_{k+1}(X_{i_k}), X_{i_{k-1}})]\;\cdots, X_{i_1})\big], X_{i_0}\Big)\Big]$$

- Risk Measures Representable as Optimal Values of Composite Functionals

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

Kernel Estimation Introduction

Assumption: The symmetric Kernel $K : \mathbb{R}^d \to \mathbb{R}$ of order $r > 0$ satisfies

k1. $\int_{\mathbb{R}} y_l^j K(y) dy_l = 0$ for $l = 1, \cdots, d$ and $j = 1, \cdots, r-1$.

k2. $\int_{\mathbb{R}^d} |y|^r |K(y)| dy < \infty$.

The smooth empirical measure for bandwidt h_n is defined as

$$P_n * K_{h_n}(x) = \frac{1}{h_n n} \sum_{i=1}^{n} K\Big(\frac{x - X_i}{h_n}\Big).$$

$$\mathbb{E}\big[f(X)\big] = \frac{1}{n} \sum_{i=1}^{n} \int_{\mathbb{R}^d} f(x) K\Big(\frac{x - X_i}{h_n}\Big) \frac{1}{h_n^d}\, dx$$

- Composite functionals

$$\varrho_K^{(n)} = \sum_{i_0=1}^{n} \frac{1}{n}[\int f_1(\sum_{i_1=1}^{n} \frac{1}{n}[\int f_2(\sum_{i=1}^{n} \frac{1}{n}[\cdots$$

$$\int f_k(\sum_{i_k=1}^{n}[\frac{1}{n} \int f_{k+1}(x) \frac{1}{h_n^{m_k}} K(x - X_{i\,k})dx], x) \frac{1}{h_n^{m_{k-1}}} K(x - X_{i\,k-1})dx$$

$$]\cdots, x) \frac{1}{h_n^{m_1}} K(x - X_{i1})dx], x) \frac{1}{h_n^{m_0}} K(x - X_{i0})dx]$$

- Risk Measures Representable as Optimal Values of Composite Functionals

$$\theta_K^{(n)} = \min_{z \in Z} f_1\Big(z, \sum_{i=1}^{n} \frac{1}{n} \int_{\mathbb{R}^d} f_2(z,x) K\big(\frac{x - X_i}{h_n}\big) \frac{1}{h_n^d} dx\Big)$$

Higher-order inverse measures of risk

$$\theta[X] = \min_{z\in\mathbb{R}}\{z + c\|\max(0, X - z)\|_p\}$$

where $p > 1$ and $\|\cdot\|_p$ is the norm in the $\mathcal{L}^p$ space.
We define

$$f_1(z, \eta_1) = z + c\eta^{\frac{1}{p}}$$
$$f_2(z, x) = (\max(0, x - z))^p$$

- The empirical estimator is

$$\theta^{(n)}[X] = \min_{z\in\mathbb{R}}\left\{z + c\left[\frac{1}{n}\sum_{i=1}^{n}\big(\max(0, X_i - z)\big)^p\right]^{\frac{1}{p}}\right\}$$

- The kernel estimator is

$$\theta_K^{(n)}[X] = \min_{z\in\mathbb{R}}\left\{z + c\left[\frac{1}{n}\sum_{i=1}^{n}\int(\max(0, x - z))^p\frac{1}{h_n}K\Big(\frac{x - X_i}{h_n}\Big)dx\right]^{\frac{1}{p}}\right\}$$

Framework

We define:

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

$$\mu_{k+1} = \int_{\mathcal{X}} f_{k+1}(x)\, P(dx), \quad \mu_j = \bar{f}_j(\mu_{j+1}), \quad j = 1, \dots, k.$$

$I_j \subset \mathbb{R}^{m_j}$ are compact sets such that $\mu_{j+1} \in \text{int}(I_j)$, $j = 1, \dots, k$.
$\mathcal{H} = \mathcal{C}_1(I_1) \times \mathcal{C}_{m_1}(I_2) \times \dots \mathcal{C}_{m_{l-1}}(I_k) \times \mathbb{R}^{m_k}$, where $\mathcal{C}_{m_{j-1}}(I_j)$ is the space of continuous functions on I_j with values in $\mathbb{R}^{m_{j-1}}$.
Hadamard directional derivatives of $f_j(\cdot, x)$ at μ_{j+1} in directions ζ_{j+1}:

$$f_j'(\mu_{j+1}, x; \zeta_{j+1}) = \lim_{\substack{t \downarrow 0 \\ s \to \zeta_{j+1}}} \frac{1}{t}\big[f_j(\mu_{j+1} + ts, x) - f_j(\mu_{j+1}, x)\big].$$

For every direction $d = (d_1, \dots, d_k, d_{k+1}) \in \mathcal{H}$, we define recursively the sequence of vectors:

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

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

Strong Law of the Large Numbers for Composite Risk Functionals

Let $\mathcal{F} = \{f_\theta : \theta \in \Theta\}$ be a collection of functions $f_\theta(x) = f(\theta, x)$:

- $f_\theta(\cdot)$ are measurable and bounded by integrable envolope function;
- the index set I is compact metric set;
- $f(\cdot, x)$ is continuous for any x.

Then $\mathcal{F}$ is Glinvenko-Cantelli class, i.e. $\sup_{\theta\in\Theta} |P_n f - Pf| \xrightarrow{a.s.} 0$

Assumptions for the estimated composite functional:

a1. The functions $f_j(\eta_j, \cdot)$, $f_{k+1}(\cdot)$ are measurable, continuous and uniformly bounded for all $\eta_j \in I_j$ $j = 1, \cdots, k$ by a measurable function.

a2. For all $x \in \mathcal{X}$, $f_j(\cdot, x)$, are continuous on I_j.

a3. The bandwidth $h_n \to 0$ when $n \to \infty$.

Then $\varrho_K^{(n)}(X) \xrightarrow{a.s.} \varrho$.

SSLN for composite risk functionas as optimal values

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

Assumptions:

b1. The function $f_1(\cdot, \cdot)$ is continuous.

b2. The functions $f_1(\eta, \cdot)$ and $f_2(z, \cdot)$ are measurable and uniformly bounded for all $\eta \in I$ and all $z \in Z$ by a measurable function.

b3. Let the bandwidth $h_n \to 0$ when $n \to \infty$.

Then the SLLN for the optimal value holds, i.e. $\theta_K^{(n)} \xrightarrow{a.s.} \theta$ as $n \to \infty$.
Additionally, $\mathbb{D}(\mathcal{S}_K^{(n)}, \mathcal{S}) \to 0$ $w.p.1.$ as $n \to \infty$.
If the optimal solution is unique, then $\mathbb{H}(\mathcal{S}_K^{(n)}, \mathcal{S}) \to 0$.

The distance $\mathbb{D}(A, B)$ denotes the deviation of the set A from set B, and $\mathbb{H}(A, B)$ stands for the Hausdorff distance between sets A and B.

SLLN for optimized composite risk functionals

$$\vartheta(X) = \min_{u\in U} \varrho(u, X) \quad \mathcal{S}(X) = \operatorname{argmin}_{u\in U}\varrho(u, X)$$

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

Assumptions:

c1. The functions f_j, $j = 1, \cdots, k$ are continuous;

c2. The functions $f_j(u, \eta, \cdot)$ are measurable and uniformly bounded for all $\eta_j \in I_j$ by a measurable function.

c3. The bandwidth $h_n \to 0$ when $n \to \infty$.

Then $\vartheta_K^{(n)} \xrightarrow{a.s.} \varrho$. If the optimal solution is unique, then $\mathbb{D}(\mathcal{S}_K^{(n)}, \mathcal{S}) \to 0$.

A1. The class

$$\mathcal{F} = \{f(\eta, x) = [f_1(\eta_1, x), f_2(\eta_2, x), \cdots, f_k(\eta_k, x), f_{k+1}(x)]^\top : \\ \eta = (\eta_1, \ldots, \eta_k) \in I\}$$

is a subset of a translation invariant class $\tilde{\mathcal{F}}$,
i.e., $f(\eta, \cdot + y) \in \tilde{\mathcal{F}}$ for all $y \in \mathbb{R}^d$.

A2. $\{\mu_n\}_{n=1}^{\infty}$ is a proper approximated convolution identity: sequence of finite signed Borel measures which converge weakly to the point mass δ_0, and for every $a > 0$, $\lim_{n\to\infty} |\mu_n|\big(\mathbb{R}^M \setminus [-a, a]^M\big) = 0$

- $\mu_n(\mathbb{R}^M) = 1$;
- for all n, for all $f \in \mathcal{F}$, $f(\eta, \cdot + y)$ is μ_n – integrable;
- $\int_{\mathbb{R}^M} \|f(\eta, \cdot - y)\|_{2,\mathbb{P}} d\mu_n(y) < \infty$ for all $f(\eta, \cdot) \in \mathcal{F}$.

A3. $\sup_{\mathcal{F}} \mathbb{E} \left(\int_{\mathbb{R}^M} (f(\eta, X + y) - f(\eta, X)) d\mu_n(y) \right)^2 \xrightarrow{n\to\infty} 0$

A4. $\sup_{\mathcal{F}} \sqrt{n} \left| \mathbb{E} \int_{\mathbb{R}^M} (f(\eta, X + y) - f(\eta, X)) d\mu_n(y) \right| \xrightarrow{n\to\infty} 0$

Central Limit Theorems for composite risk functionals

Assume A1. A3. and A4.

A'2. The symmetric kernel K is of order two or higher and satisfies regularity assumptions and the bandwidth $h_n \to 0$ when $n \to \infty$;

A5. For all $x \in \mathcal{X}$, the functions $f_j(\cdot, x)$, $j = 1, \dots, k$, are Lipschitz continuous with square-integrable Lipschitz constant and Hadamard directionally differentiable;

Then $\sqrt{n}\big[\varrho^{(n)} - \varrho\big] \xrightarrow{\mathcal{D}} \xi_1(W)$, where $W(\cdot) = \big(W_1(\cdot), \dots, W_k(\cdot), W_{k+1}\big)$ is a zero-mean Brownian process on I; $W_j(\cdot)$ is a Brownian process of dimension m_{j-1} on I_j, $j = 1, \dots, k$, and W_{k+1} is an m_k-dimensional normal vector.

$$\mathrm{cov}\left[W_i(\eta_i), W_j(\eta_j)\right] = \int_{\mathcal{X}} \left[f_i(\eta_i, x) - \bar{f}_i(\eta_i)\right]\left[f_j(\eta_j, x) - \bar{f}_j(\eta_j)\right]^\top P(dx),$$

$$\eta_i \in I_i,\ \eta_j \in I_j,\ i, j = 1, \dots, k,$$

$$\mathrm{cov}\left[W_i(\eta_i), W_{k+1}\right] = \int_{\mathcal{X}} \left[f_i(\eta_i, x) - \bar{f}_i(\eta_i)\right]\left[f_{k+1}(x) - \mu_{k+1}\right]^\top P(dx),$$

$$\eta_i \in I_i,\ i = 1, \dots, k,$$

$$\mathrm{cov}\left[W_{k+1}, W_{k+1}\right] = \int_{\mathcal{X}} \left[f_{k+1}(x) - \mu_{k+1}\right]\left[f_{k+1}(x) - \mu_{k+1}\right]^\top P(dx).$$

Semi-deviations continued

The limiting random variable

$$\xi_1(W) = V_1 + \frac{\kappa}{p}\Big(\mathbb{E}\big\{\big[\max\{0, X - \mathbb{E}[X]\}\big]^p\big\}\Big)^{\frac{1-p}{p}} \times \Big(V_2 - p\mathbb{E}\big\{\big[\max\{0, X - \mathbb{E}[X]\}\big]^{p-1}\big\} V_1\Big).$$

Here V_1 and V_2 are normal random variables ($V_2 = W_2(\mathbb{E}[X])$) and

$$\mathrm{Var}\big[V_1\big] = \mathrm{Var}[X],$$

$$\mathrm{Var}[V_2] = \mathbb{E}\Big\{\Big(\big[\max\{0, X - \mathbb{E}[X]\}\big]^p - \mathbb{E}\big(\big[\max\{0, X - \mathbb{E}[X]\}\big]^p\big)\Big)^2\Big\},$$

$$\mathrm{cov}[V_2, V_1] = \mathbb{E}\Big\{\Big(\big[\max\{0, X - \mathbb{E}[X]\}\big]^p - \mathbb{E}\big(\big[\max\{0, X - \mathbb{E}[X]\}\big]^p\big)\Big)\Big(X - \mathbb{E}[X]\Big)\Big\}.$$

If $p = 1$ the limit distribution may be obtained in a different way.

Risk functionals as optimal values

Composite risk functional of the higher order risk measures

$$\theta[X] = \min_{u \in U} f_1\big(u, \mathbb{E}[f_2(u, X)]\big), \quad U \subset \mathbb{R}^d \text{ is a nonempty compact set.}$$

The class $\mathcal{F} = \{f(u,x) = [f_1(u,x), f_2(u,x)]^\top : u \in U\}$
Assume that A1, A2. (or A'2), A3, A4 are satisfied.

C1. The function $f_1(u, \cdot)$ is differentiable $\forall u \in U$, and both $f_1(\cdot, \cdot)$ and its derivative w.r.t. the second argument, $\nabla f_1(\cdot, \cdot)$, are jointly continuous;
C2. $f_2(\cdot, x)$ is Lipschitz-continuous with a square-integrable Lipschitz constant.

Then $\sqrt{n}\big[\theta_K^{(n)} - \theta\big] \xrightarrow{\mathcal{D}} \min_{u \in \hat{U}} \big\langle \nabla f_1\big(u, \mathbb{E}[f_2(u, X)]\big), W(u)\big\rangle$, where
$W(u)$ is a zero-mean Brownian process on U with the covariance function

$$\operatorname{cov}\big[W(u'), W(u'')\big] = \int_{\mathcal{X}} \big(f_2(u', x) - \mathbb{E}[f_2(u', X)]\big)\big(f_2(u'', x) - \mathbb{E}[f_2(u'', X)]\big)^\top P(dx).$$

Optimized Composite Risk Functionals

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

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

Assumptions in addition to A1–A4:

D1. The optimal solution $\hat{u}$ of this problem is unique;

D2. Compact sets $I_1, \dots, I_k$ are selected so that $\text{int}(I_k) \supset \bar{f}_{k+1}(U)$, and $\text{int}(I_j) \supset \bar{f}_{j+1}(U, I_{j+1})$, $j = 1, \dots, k-1$.

D3. The functions $f_j(\cdot, \cdot, x)$, $j = 1, \dots, k$, and $f_{k+1}(\cdot, x)$ are Lipschitz continuous for every $x \in \mathcal{X}$ with square integrable Lipschitz constants.

D4. The functions $f_j(u, \cdot, x)$, $j = 1, \dots, k$, are continuously differentiable for every $x \in \mathcal{X}$, $u \in U$; their derivatives are continuous with respect to the first two arguments.

Central Limit Theorem for the optimized risk functional

It holds

$$\sqrt{n}\big[\vartheta^{(n)} - \vartheta\big] \xrightarrow{\mathcal{D}} \xi_1(\hat{u}, W),$$

$W(\cdot) = \big(W_1(\cdot), \dots, W_k(\cdot), W_{k+1}\big)$ is a zero-mean Brownian process on $I = I_1 \times I_2 \times \cdots \times I_k$; $W_j(\cdot)$ is a Brownian process of dimension m_{j-1} on I_j, $j = 1, \dots, k$, and W_{k+1} is an m_k-dimensional normal vector. The covariance function of $W(\cdot)$ has the form

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

$$\operatorname{cov}\big[W_i(\eta_i), W_{k+1}\big] = \int_{\mathcal{X}} \big[f_i(\hat{u}, \eta_i, x) - \bar{f}_i(\hat{u}, \eta_i)\big]\big[f_{k+1}(\hat{u}, x) - \bar{f}_{k+1}(\hat{u})\big]^\top P(dx),$$
$$\eta_i \in I_i,\ \ i = 1, \dots, k$$

$$\operatorname{cov}\big[W_{k+1}, W_{k+1}\big] = \int_{\mathcal{X}} \big[f_{k+1}(\hat{u}, x) - \bar{f}_{k+1}(\hat{u})\big]\big[f_{k+1}(\hat{u}, x) - \bar{f}_{k+1}(\hat{u})\big]^\top P(dx).$$

Optimization problems with mean–semideviation

$$\min_{u\in U} \varrho(\varphi(u,X)) = \mathbb{E}(\varphi(u,X)) + \kappa\Big(\mathbb{E}\big[(\varphi(u,X) - \mathbb{E}[\varphi(u,X)])^p\big]\Big)^{\frac{1}{p}}\Big],$$

where $\varphi : \mathbb{R}^d \times \mathcal{X} \to \mathbb{R}$. We have

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

and

$$\begin{aligned} \bar{f}_1(\eta_1,u) &= \kappa\eta_1^{\frac{1}{p}} + \mathbb{E}[\varphi(u,X)], \\ \bar{f}_2(\eta_2,u) &= \mathbb{E}\big\{\big[\max\{0,\varphi(u,X)-\eta_2\}\big]^p\big\}, \\ \bar{f}_3(u) &= \mathbb{E}[\varphi(u,X)]. \end{aligned}$$

We assume that $p > 1$ and $\hat{u}$ is the unique solution of the problem.

Optimization problems with mean–semideviation continued

We set $\mu_3 = \mathbb{E}[\varphi(\hat{u}, X)]$.
Then $\mu_2 = \mathbb{E}\big\{\big[\max\{0, \varphi(\hat{u}, X) - \mathbb{E}[\varphi(\hat{u}, X)]\}\big]^p\big\}$ and $\mu_1 = \varrho(X)$.
The limiting element

$$\xi_2(d) = \bar{f}_2'(\mu_3, \hat{u}; d_3) + d_2(\mu_3) = -p\mathbb{E}\big\{\big[\max\{0, \varphi(\hat{u}, X) - \mu_3\}\big]^{p-1}\big\} d_3 + d_2(\mu_3),$$

$$\xi_1(d) = \bar{f}_1'\big(\mu_2, \hat{u}; \xi_2(d)\big) + d_1\big(\mu_2\big) = \frac{\kappa}{p}\mu_2^{\frac{1}{p}-1}\xi_2(d) + d_1\big(\mu_2\big).$$

The limiting element is

$$V_1 + \frac{\kappa}{p}\Big(\mathbb{E}\big\{\big[\max\{0, \varphi(\hat{u}, X) - \mathbb{E}[\varphi(\hat{u}, X)]\}\big]^p\big\}\Big)^{\frac{1-p}{p}} \times \Big(V_2 - p\mathbb{E}\big\{\big[\max\{0, \varphi(\hat{u}, X) - \mathbb{E}[\varphi(\hat{u}, X)]\}\big]^{p-1}\big\} V_1\Big).$$

The normal random variables V_1 and V_2 have variances:

$$\mathrm{Var}(V_1) = \mathrm{Var}\big(\varphi(\hat{u}, X)\big);$$

$$\mathrm{Var}(V_2) = \mathbb{E}\Big\{\Big(\big[\max\{0, \varphi(\hat{u}, X) - \mathbb{E}[\varphi(\hat{u}, X)]\}\big]^p - \mathbb{E}\big(\big[\max\{0, \varphi(\hat{u}, X) - \mathbb{E}[\varphi(\hat{u}, X)]\}\big]^p\big)\Big)\Big(\varphi(\hat{u}, X) - \mathbb{E}[\varphi(\hat{u}, X)]\Big)\Big\}.$$

Numerical comparison of the two estimators

Consider the higher order inverse risk measure

$$\theta(X) = \min_{u\in\mathbb{R}}\{u + c\|\max(0, X - u)\|_p\}$$

Uniform kernel function $K(u) = \frac{1}{2h_n}$ with support on $|u| \leq h_n$.

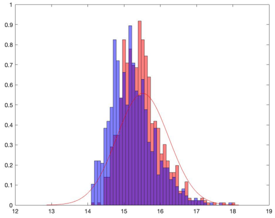
For $p > 1$, the kernel estimator $\theta_K^{(n)}$ has the form

$$\min_{u\in\mathbb{R}}\left\{u + c\Big(\frac{1}{n}\sum_{i=1}^{n}\int(\max(0, x-u))^p\frac{1}{2h_n}\mathbb{I}_{(|x-X_i|\leq h_n)}\,dx\Big)^{\frac{1}{p}}\right\}$$

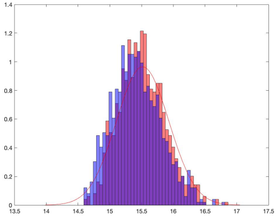
$$= \min_{u\in\mathbb{R}}\left\{u + c\Big(\sum_{i=1}^{n}\frac{1}{2n(p+1)h_n}\big[(\max(0, h_n + X_i - u))^{p+1} - (\max(0, -h_n + X_i - u))^{p+1}\big]\Big)^{\frac{1}{p}}\right\}$$

We use a sample X_i, $i = 1, \ldots, n$ from $X \sim N(10, 3)$ and set $p = 2$.
Recall that BIAS=$\mathbb{E}(\theta^{(n)}) - \theta$ and VARIANCE=$\mathbb{E}[\theta^{(n)} - \mathbb{E}(\theta^{(n)})]^2$.

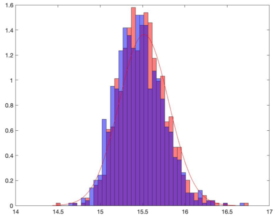
Numerical comparison for normal random variables



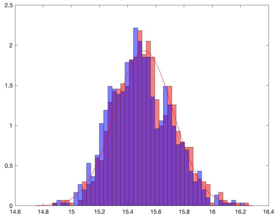
(a) m1000n500p2



(b) m1000n1500p2

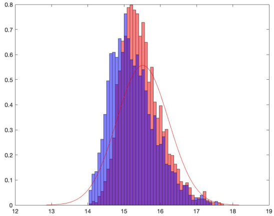


(c) m1000n3000p2

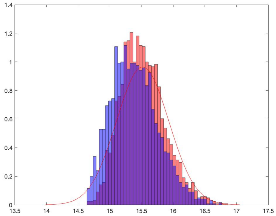


(d) m1000n6000p2

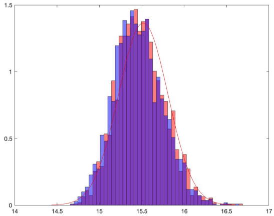
Numerical comparison for normal random variables



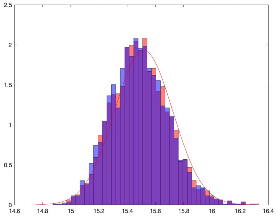
(e) m2500n500p2



(f) m2500n1500p2



(g) m2500n3000p2



(h) m2500n6000p2

The bias and the variance of the estimators

n	m	K-bias	K-variance	E-bias	E-variance
500	1000	-0.0058	0.2930	-0.2981	0.3816
1500	1000	-0.0241	0.1462	-0.1080	0.1640
3000	1000	-0.0318	0.0795	-0.0456	0.0811
6000	1000	-0.0230	0.0406	-0.0262	0.0408
500	2500	-0.1370	0.2913	-0.3278	0.3561
1500	2500	-0.0405	0.1388	-0.1173	0.1538
3000	2500	-0.0383	0.0789	-0.0600	0.0812
6000	2500	-0.0222	0.0426	-0.0254	0.0428

Conclusion

- Better performance when data size is small by kernel estimation.
- Reduce the bias by kernel estimation.
- Bandwidth of the kernel could be optimized.

The effect of the order and heavier tails

We consider p as parameter, X is normal distribution $N(10, 3)$, $n = 2000$ and $m = 2500$.

p	K-bias	K-variance	E-bias	E-variance
1	-0.0019	0.0094	-0.0040	0.0094
1.50	-0.0156	0.0337	-0.0184	0.0338
2.00	-0.0871	0.1207	-0.0910	0.1214
2.5	-0.3481	0.3053	-0.3617	0.3118

T-distribution degrees of freedom 60,8,6,4, mean = 10,
$n = 4000$, $m = 2500$, $p = 2.00$

df	K-bias	K-variance	E-bias	E-variance
60	-0.0260	0.0256	-0.0313	0.0259
8	-0.0815	0.2255	-0.0841	0.2260
6	-0.1464	0.4984	-0.1484	0.4989
4	0.4484	2.6820	0.4496	2.6827

Stochastic Dominance Relation

$$
\begin{aligned}
X \succeq_{(2)} Y &\Leftrightarrow \mathbb{E}[(\eta - X)_+] \le \mathbb{E}[(\eta - Y)_+] \\
&\Leftrightarrow \int_0^\alpha F^{(-1)}(X;t)\,dt \ge \int_0^\alpha F^{(-1)}(Y;t)\,dt \quad \forall \alpha \in [0,1] \\
&\Leftrightarrow \varrho[X] \le \varrho[Y] \quad \forall \varrho \text{ law-invariant coherent risk measures.}
\end{aligned}
$$

kth degree Stochastic Dominance (kSD), $k \ge 2$

$$
X \succeq_{(k)} Y \quad \Leftrightarrow \quad \| \max\big(0, \eta - X\big) \|_{k-1}^{k-1} \le \| \max\big(0, \eta - Y\big) \|_{k-1}^{k-1}
$$

Test for Stochastic Dominance or order 1 or 2

For $X, Y \in \mathcal{L}_k(\Omega, \mathcal{F}, P)$ with $k \geq 1$, we consider the hypothesis

$$H_0 : \ \varrho[X] \leq \varrho[Y] \quad \text{versus } H_a : \ \varrho[X] > \varrho[Y],$$

where the risk functional ϱ is law-invariant coherent measure of risk. Rejecting H_0 implies that X does not dominate Y in orders $m = 1$ or $m = 2$.

Step 0. Set $i = 1$.

Step 1. Select p_i uniformly distributed in $[0, 1]$ and test the hypothesis at level of significance α

$$H_0 : \ \text{AVaR}_{p_i}[X] \leq \text{AVaR}_{p_i}[Y]$$

Step 2. If H_0 is rejected, reject the hypothesis $X \succeq_{(2)} Y$ and stop. If $i = N$ accept $X \succeq_{(2)} Y$, otherwise increase i by one and go to Step 1.

The type I error of this test is asymptotically bounded by α and does not exceed α' for any $\alpha' > \alpha$ and N sufficiently large.

Test for Stochastic Dominance of order $k \geq 2$

For $X, Y \in \mathcal{L}_k(\Omega, \mathcal{F}, P)$ with $k \geq 2$, we consider

$$\theta_k(X) = \min_{z \in \mathbb{R}} \left\{ c \big\| \max(0, z - X) \big\|_k - z \right\}$$

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

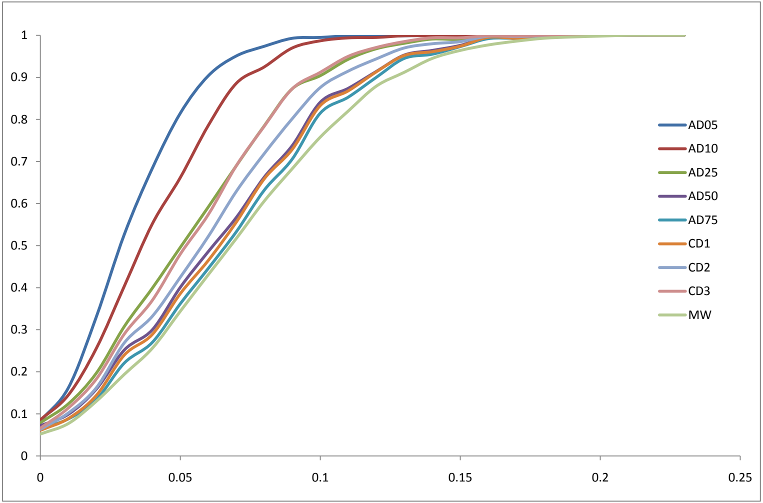
Consistency with higher-order stochastic dominance

If $X \succeq_{(k+1)} Y$ then $\theta_k(X) \leq \theta_k(Y)$ as well as $\varrho_k(X) \leq \varrho_k(Y)$.

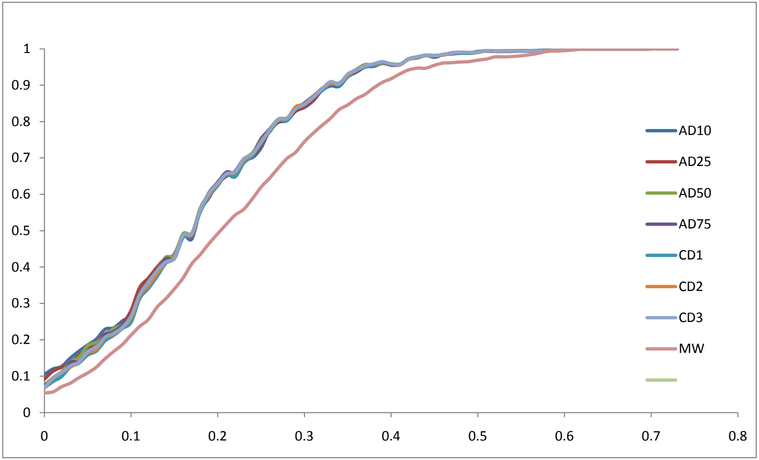
$$H_0 : \ \varrho[X] \leq \varrho[Y] \quad \text{versus } H_a : \ \varrho_k[X] > \varrho_k[Y],$$
$$H_0 : \ \theta_k[X] \leq \theta_k[Y] \quad \text{versus } H_a : \ \theta_k[X] > \theta_k[Y].$$

Rejecting H_0 implies that X does not dominate Y in orders $m = 1, \ldots, k+1$.

Power Comparison: Unif(0,1) vs. Unif(ε, $1+\varepsilon$)



Power Comparison: Gamma(2,1) vs. Gamma(2/(1-ε),1-ε)



Multivariate extension

Consider ℓ random variables X^i, $i = 1, \ldots, \ell$. and ℓ composite risk functionals for them

$$\varrho_i(X^i) = \mathbb{E}\left[f_1^i\left(\mathbb{E}\left[f_2^i\left(\mathbb{E}\left[\cdots f_{k_i}^i\left(\mathbb{E}\left[f_{k_i+1}\left(X^i\right)\right], X^i\right)\right]\cdots, X^i\right)\right], X^i\right)\right]$$

Without loss of generallity, we may assume the same level of nesting.

Multivariate CLT

Setting $Y = (X^1, \ldots, X^\ell)^\top$ and assuming analogous conditions, we have

$$\sqrt{n}(\varrho^n - \varrho(Y)) = \sqrt{n}\left[\begin{pmatrix}\varrho_1^{(n)} \\ \vdots \\ \varrho_\ell^{(n)}\end{pmatrix} - \begin{pmatrix}\varrho_1 \\ \vdots \\ \varrho_\ell\end{pmatrix}\right] \xrightarrow{\mathcal{D}} \xi_1(W).$$

For a vector $a \in \mathbb{R}^\ell$

$$\sqrt{n}\, a^\top \left[\begin{pmatrix}\varrho_1^{(n)} \\ \vdots \\ \varrho_\ell^{(n)}\end{pmatrix} - \begin{pmatrix}\varrho_1 \\ \vdots \\ \varrho_\ell\end{pmatrix}\right] \to a^\top \xi_1(W)$$

where $W(\cdot) = (W_1(\cdot), \ldots, W(\cdot)_{k+1})$ is a zero-mean Brownian process of appropriate dimension.

Application to Portfolio

Efficiency

Given a set of random variables $\mathcal{Q}$, a random variable $X \in \mathcal{Q}$ is efficient under $\succeq$ if there is no $Y \in \mathcal{Q}$ such that Y strictly dominates X.

Consider a family of random variables

$$\mathcal{Q} = \left\{ X(u) = u^\top R : u \in \mathbb{R}^m, u^\top 1 = \gamma \right\},$$

where the random vector R comprises the return rates of a basket of m securities and u represents a feasible portfolio and u denotes the investment allocation.

Multivariate central limit formula applied to $\varrho[u_1^\top R] - \varrho[u_2^\top R]$ allows for statistical testing of

- the relation of risk of two given portfolios;
- efficiency of a given portfolio.

A Portfolio Optimization Problem

Assumption $R = (R_1, R_2, \dots, R_m)$ has eliptical distribution with expectation μ and covariance matrix Σ.

The lower semi-deviation of second order is

$$\left(\mathbb{E}\left[\max(0, \mathbb{E}(u^\top R) - u^\top R)^2\right]\right)^{\frac{1}{2}} = \frac{1}{\sqrt{2}}\sqrt{u^\top \Sigma u}.$$

The mean-semi-deviation optimization problem becomes

The mean-standard-deviation model

$$\max_{u \in \mathbb{R}^m} \; u^\top \mu - \kappa\sqrt{u^\top \Sigma u} \quad \text{s.t.} \quad u^\top 1 = \gamma \tag{1}$$

Here 1 is the m-dimensional vector of ones, $\gamma \in \mathbb{R}$.

Lower Bound for the Cost of Risk κ

Theorem

If problem (1) has an optimal solution, then

$$\begin{aligned} \kappa^2 &> \mu^\top \Sigma^{-1} \mu - \frac{(\mu^\top \Sigma^{-1} 1)^2}{1^\top \Sigma^{-1} 1} \quad \text{for any } \gamma \neq 0 \\ \kappa^2 &= -\frac{(\mu^\top \Sigma^{-1} 1)^2}{1^\top \Sigma^{-1} 1} \quad \text{for } \gamma = 0 \end{aligned} \tag{2}$$

Denoting the lower bound by κ_0, problem (1) always has a solution for $\kappa = 0$ but has no solution for $\kappa \in (0, \kappa_0]$.

Example

For $\kappa = 0.5$, the existence of optimal solution requires

$$\left(\mu^T \Sigma^{-1} 1\right)^2 > 1^T \Sigma^{-1} 1 \left(\mu^T \Sigma^{-1} \mu - 0.25\right). \tag{3}$$

For 10,000 observations of return data for 200 securities we computed $\bar{r} S^{-1} 1 = 300.0096$, $1 S^{-1} 1 = 11065445$, and $\bar{r} S^{-1} \bar{r} = 1.4994$. Substitution of these values into (3) contradicts the inequality. Substitution into (2) gives $\hat{\kappa}_0 = 1.2212$. The estimate problem has no solution for any $0 < \kappa \leq 1.2212$.

Bounded short positions

Mean-standard-deviation model with bounded allocations

$$\max u^\top \mu - \kappa\sqrt{u^\top \Sigma u} \quad \text{s.t.} \quad u^\top 1 = \gamma \quad u \geq \ell. \tag{4}$$

Results

The optimal value ϑ, the optimal solution $\hat{u}$ and the optimal Lagrange multipliers α and λ satisfy the specific optimality conditions implying

$$\kappa^2 > \langle \mu + \lambda, \Sigma^{-1}(\mu + \lambda)\rangle - \frac{\langle \mu + \lambda, \Sigma^{-1} 1\rangle^2}{1^\top \Sigma^{-1} 1} \quad \text{for any } \gamma \neq 0$$

$$\kappa^2 = \langle \mu + \lambda, \Sigma^{-1}(\mu + \lambda)\rangle - \frac{\langle \mu + \lambda, \Sigma^{-1} 1\rangle^2}{1^\top \Sigma^{-1} 1} \quad \text{for } \gamma = 0.$$