THE INTERPLAY OF M-ESTIMATION & STOCHASTIC PROGRAMMING

Jitka Dupačová

Dept. of Probability and Mathematical Statistics, Charles University, Prague, Czech Republic
www.karlin.mff.cuni.cz/ ∼dupacova

Stochastic Models and Control, Bad Herrenalb, March 31, 2011

Content

Decision problems of SP

SP is extension of linear or nonlinear programming to decision models where coefficients (parameters) are not known with certainty and have been given a probabilistic representation.

One cannot wait until the realizations of parameters are observed,

decision has to be accepted NOW — **decision problems of SP**

Reformulation of optimization problem is needed!

For example the wish is:

Choose the "best" $\mathbf{x} \in \mathcal{X}$ — the set of hard constraints.

The approach:

Quantify random outcome of $\mathbf{x}$ when random parameter ω occurs as $f(\mathbf{x}, \omega)$.

The "best" $\mathbf{x}$ minimizes $E_P f(\mathbf{x}, \omega)$ on $\mathcal{X}$.

Basic SP model

$$\min_{\mathbf{x} \in \mathcal{X}(P)} E_P f(\mathbf{x}, \omega) \tag{1}$$

- known probability distribution P of random parameter ω whose support Ω is a closed subset of $\mathbb{R}^s$;
- given, nonempty, closed set $\mathcal{X}(P) \subset \mathbb{R}^n$ of decisions $\mathbf{x}$; mostly, $\mathcal{X}$ does not depend on P;
- preselected random objective $f : \mathcal{X}(P) \times \Omega \to \mathbb{R}$ or $\mathbb{R} \cup \{+\infty\} \sim$ loss or cost caused by decision $\mathbf{x}$ when scenario ω occurs.
  As a function of ω,
  f is measurable & its expectation $E_P f(\mathbf{x}, \omega)$ exists $\forall x \in \mathcal{X}$.
  Structure of f may be quite complicated (e.g. for multistage problems). For convex $\mathcal{X}$, a frequent assumption is that f is lower semicontinuous and convex wrt. $\mathbf{x}$, i.e., f is *convex normal integrand*

Need to study properties of model (existence of expectation, convexity...)
To solve SP — approximate P by discrete probability distribution, say P_ν, based on sample $\omega_1, \ldots, \omega_\nu$ from P, etc.

Comparison with M-estimates

M-estimate θ^ν of the true parameter θ of P is an optimal solution of an unconstrained optimization problem obtained for a sample $\omega_1, \dots, \omega_\nu$ from probability distribution P and for a suitable fitting function f.

Generic problem formulation

$$\{\theta^\nu\} = \arg\min_\theta \sum_{i=1}^{\nu} f(\theta, \omega_i) \tag{2}$$

When empirical probability distribution P^ν is used to approximate the true unknown probability distribution P in SP (1), expectation of $f(x, \omega)$ is approximated by sample average.

Sample Average Approximation (SAA) of SP

$$\min_{x \in \mathcal{X}} \sum_{i=1}^{\nu} f(x, \omega_i) \tag{3}$$

is solved and corresponding optimal solutions $x^*(P^\nu)$ are accepted as approximation of the true optimal solution $x^*(P)$.

Comparison with M-estimates – cont.

Similarity → one may exploit robustness measures developed in robust estimation to analyze properties of approximate optimal solutions of SP.

In spite of similarities of the two types of problems, there are various differences e.g.

1. Goal for SP – to support decisions; in statistics emphasis is on identification of parameter or distribution;

2. Solutions of stochastic programming problems have to fulfil various constraints and the corresponding set of feasible solutions $\mathcal{X}$ is closed.

3. Contrary to statistical problems where f is a tool for estimating true parameter value, random objective function $f(x, \omega)$ in SP reflects goals of the decision problem and all minimizers of its expectation are equally acceptable. Their uniqueness is not required.

Methodology based on epi-convergence developed to get consistency of $\{x^\nu\}$; their asymptotic normality cannot be expected

Hence, asymptotics for stochastic programming can be exploited for (inequality) constrained M-estimates, such as isotonic regression, etc.

Asymptotic inference

Sample space $(Z, \mathcal{F}, \mu)$ with increasing sequence of σ-fields $(\mathcal{F}^\nu)_{\nu=1}^\infty$ contained in $\mathcal{F}$. Sample path ζ, increasing sample size $\rightsquigarrow$ sequence of $\mathcal{F}^\nu$-measurable probability distributions $\{P^\nu(\bullet, \zeta), \nu = 1, 2, \dots\}$ on $(\Omega, \mathcal{B})$ based on information collected up to ν.
Optimal value $\varphi(P^\nu)$ and optimal solutions $\mathbf{x}^*(P^\nu)$ of approximate stochastic program

$$\min_{\mathbf{x} \in \mathcal{X}} F(\mathbf{x}, P^\nu) := E_{P^\nu} f(\mathbf{x}, \omega) \tag{4}$$

based on $P^\nu(\bullet, \zeta)$ depend on the used sample path ζ. All presented results hold true for almost all sample paths ζ, i.e., μ-a.s.

Probability distribution P^ν can be a wide-sense empirical probability distribution.
(For empirical probability distributions the sample path $\zeta = \{\omega_1, \omega_2, \dots\}$ is obtained by simple random sampling from $(\Omega, \mathcal{B}, P)$, $\mu = P^\infty$ and the empirical stochastic program is (3).)

Classical Consistency Results

- If $P^\nu \to P$ weakly and $f(\mathbf{x}, \bullet)$ is continuous bounded function of $\omega \, \forall \mathbf{x} \in \mathcal{X} \implies$ pointwise convergence of expected value objectives $F(\mathbf{x}, P^\nu) \to F(\mathbf{x}, P) \quad \forall \mathbf{x} \in \mathcal{X}$;
- If $\mathcal{X}$ is compact & convergence of expectations is uniform on $\mathcal{X} \implies$ (μ-a.s.) convergence of optimal values $\varphi(P^\nu) \to \varphi(P)$.
- If, moreover, $f(\bullet, \omega)$ is strictly convex on $\mathcal{X} \implies$ convergence of (unique) optimal solutions $\mathbf{x}^*(P^\nu)$ of $\min_{\mathbf{x} \in \mathcal{X}} F(\mathbf{x}, P^\nu)$ to the unique optimal solution $\mathbf{x}^*(P)$ of problem

$$\min_{\mathbf{x} \in \mathcal{X}} F(\mathbf{x}, P) := E_P f(\mathbf{x}, \omega) \tag{5}$$

  and some rates of convergence.

Notice that merely pointwise convergence of empirical expectations does not imply consistency of optimal values. Convexity of $f(x, \bullet)$ helps (e.g. Haberman).

General consistency result is based on notion of epi-convergence of lower semicontinuous (lsc) functions cf. [D-W]. The main step is to prove that approximate objective functions $F(\mathbf{x}, P^\nu)$ epi-converge to the true objective function in (5), which in turn implies convergence results for optimal values and for sets of optimal solutions.

Epi-convergence

DEFINITION – Epi-convergence.

Sequence of functions $\{u^\nu : \mathbb{R}^n \to \bar{\mathbb{R}}, \nu = 1, \dots\}$ is said to epi-converge to $u : \mathbb{R}^n \to \bar{\mathbb{R}}$ if for all $\mathbf{x} \in \mathbb{R}^n$ the two following properties hold true:

$$\liminf_{\nu \to \infty} u^\nu(\mathbf{x}^\nu) \geq u(\mathbf{x}) \text{ for all } \{\mathbf{x}^\nu\} \to \mathbf{x} \tag{6}$$

and for some $\{\mathbf{x}^\nu\}$ converging to $\mathbf{x}$

$$\limsup_{\nu \to \infty} u^\nu(\mathbf{x}^\nu) \leq u(\mathbf{x}). \tag{7}$$

Pointwise convergence implies condition (7), additional assumptions are needed to get validity of condition (6). Pointwise convergence of lsc. convex functions u, u^ν with int dom$(u) \neq \emptyset$ implies epi-convergence.

For epi-convergence we have

$$\limsup\{\arg\min u^\nu\} \subset \arg\min u \tag{8}$$

and convergence of optimal values.

Consult Rockafellar & Wets, Variational analysis, Springer, 1998.

Assumptions

a. $\mathcal{X} \subset R^n$ is a nonempty closed set independent of P,

b. $f(\mathbf{x}, \omega)$ is a random lower semicontinuous function, i.e. f is jointly measurable and $f(\bullet, \omega)$ is lower semicontinuous for all $\omega \in \Omega$,

c. $P^\nu \to P$ weakly.

To get *epi-convergence* of expectations $F(\mathbf{x}, P^\nu) \to F(\mathbf{x}, P)$, additional assumptions on convergence of $P^\nu \to P$ and properties of f are needed.

d. continuity of $f(\mathbf{x}, \bullet)$ on Ω,

e. uniform convergence (asymptotic neglibility, tightness) of probability distributions P, P^ν with respect to functions $f(\mathbf{x}, \bullet)\, \forall \mathbf{x} \in \mathcal{X}$; this replaces assumption of bounded integrals $f(\mathbf{x}, \bullet)\, \forall \mathbf{x}$.

f. local (lower) Lipschitz property of $f(\bullet, \omega)$ for all $\omega \in \Omega$; in case of $f(\bullet, \omega)$ convex for all $\omega \in \Omega$, this assumption is not needed.

Consistency Results

Proposition 1. (cf. Theorems 3.7, 3.8 of [D-W]): Under assumptions **a–f**, $F(\mathbf{x}, P)$ is μ-a.s. both epi-limit and pointwise limit of $F(\mathbf{x}, P^\nu)$ for $\nu \longrightarrow \infty$.

Epi-convergence of the objective functions, cf. [R-W], implies the consistency result:

Proposition 2. (cf. Theorem 3.9 of [D-W]): Under assumptions **a–f** we have that μ-a.s.

$$\lim \sup_{\nu \to \infty} \varphi(P^\nu) \le \varphi(P)$$

and any cluster point $\hat{\mathbf{x}}$ of any sequence $\{\mathbf{x}^*(P^\nu), \nu = 1, 2, \dots\}$ of optimal solutions $\mathbf{x}^*(P^\nu) \in \mathcal{X}^*(P^\nu)$ belongs to $\mathcal{X}^*(P)$.

In particular, if there is compact set $\mathcal{D} \subset \mathbb{R}^n$ such that μ-a.s., for $\nu = 1, 2, \dots, \mathcal{X}^*(P^\nu) \cap \mathcal{D} \neq \emptyset$ and $\mathbf{x}^*(P) \in \mathcal{X}^*(P) \cap \mathcal{D} \Longrightarrow \exists$ measurable selection $\mathbf{x}^*(P^\nu)$ of $\mathcal{X}^*(P^\nu)$ such that $\mathbf{x}^*(P) = \lim_{\nu \to \infty} \mathbf{x}^*(P^\nu)$ for μ-almost all ζ and also $\varphi(P) = \lim_{\nu \to \infty} \varphi(P^\nu)$ μ-a.s.

Comments

- For convex function $f(\bullet, \omega)$, convex $\mathcal{X}$ and for empirical probability distributions P^ν epi-convergence of $F(\mathbf{x}, P^\nu)$ to $F(\mathbf{x}, P)$ follows from the strong law of large numbers for sums of random closed sets and the consistency result can be extended from R^n to reflexive Banach spaces, cf. King & Wets.
- The approach based on epi-convergence allows for application to problems where both the integrand and the probability distribution are approximated.
- $\exists$ results valid under various assumptions about approximate probability distributions P^ν, e.g. Korf & Wets;
- Important generalization to discontinuous integrands $f(\mathbf{x}, \bullet)$. In such cases, uniform integrability is not sufficient for semicontinuity of integrals $F(\mathbf{x}, P^\nu)$. A suitable additional condition is that probability of the set of discontinuity points of $f(\mathbf{x}, \bullet)$ for the true problem is zero; cf. application to approximated integer stochastic programs or probabilistic programs.
- Epi-consistency employed for qualitative studies of SP (consistency)

Example – Constrained nonlinear L_1 regression

We have $\omega = \{\omega^0, \omega^1, \ldots, \omega^m\}$, $f(\theta, \omega) = |\omega^0 - g(\theta, \omega^1, \ldots, \omega^m)|$, $\mathcal{S} \subset \mathbb{R}^n$, $\mathcal{S} \neq \emptyset$ a closed set of admissible parameter values.

$$\theta^\nu \in \arg\min_{\theta \in \mathcal{S}} \sum_{i=1}^{\nu} f(\theta, \omega_i) := F(\theta, P^\nu) \qquad (9)$$

ASSUMPTIONS to get **a – f**:
$g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^1$ continuous, locally Lipschitz wrt. θ with an integrable Lipschitz constant;
Convergence of empirical probability distributions P^ν to P is f-tight. (OK e.g. when Ω is compact).

THEN any cluster point $\hat{\theta}$ of any sequence $\{\theta(P^\nu), \nu = 1, 2, \ldots\}$ such that $\theta(P^\nu) \in \arg\min_{\theta \in \mathcal{S}} F(\theta, P^\nu)$ almost surely belongs to $\arg\min_{\theta \in \mathcal{S}} F(\theta, P)$, etc. as in Proposition 2.

Existence of optimal solutions of (9) is guaranteed if $F(\mathbf{x}, P^\nu)$ is infcompact, i.e. $\{\mathbf{x} \in \mathbb{R}^n : F(\mathbf{x}, P^\nu) \leq \alpha\}$ is bounded for all α.
Sufficient condition is $f(\mathbf{x}, \omega)$ infcompact with positive probability, e.g. for a realization of ω, cf. [D-W].

Constrained robust nonlinear regression

Composite random objective function, say

$$\phi(\omega^0 - g(\theta, \omega^1, \dots, \omega^m))$$

which appears in (9) at the place of $f(\theta, \omega)$.

Even with polyhedral set $\mathcal{S}$ and convex piecewise quadratic-linear function ϕ, the resulting optimization problem is rather challenging both from the viewpoint of optimization and also of sensitivity analysis, for nonlinear regression one cannot rely on convexity, etc.

S.M.Robinson proposes a general method which can be applied to local stability analysis (and a Newton method for computing optimal solutions of similar problems).
Idea: transform the problem into one with convex random objective and nonlinear constraints, say

$$\min \sum_{i=1}^{\nu} \phi(\xi_i) \quad \text{subject to}$$

$$\xi_i = \omega_i^0 - g(\theta, \omega_i^1, \dots, \omega_i^m), \forall i, \theta \in \mathcal{S}.$$

Another problem – sample dependent constraints!

Asymptotic distributions – Optimal value

Derived under assumption that consistency holds true.
For empirical stochastic program (3) asymptotic normality of optimal value $\varphi(P^\nu)$ can be proved under relatively weak assumptions, e.g. compact $\mathcal{X} \neq \emptyset$, unique true optimal solution $\mathbf{x}^*(P)$ and $f(\bullet, \omega)$ Lipschitz continuous $\forall \omega$, with finite expectation $E\{f(\mathbf{x}^*(P), \omega)\}^2$; consult Shapiro $\rightsquigarrow$ approximate confidence intervals for the true optimal value.

For inference based on these approximate confidence intervals one should realize that empirical optimal value $\varphi(P^\nu)$ has a one-directional bias in the sense that

$$E_\mu \varphi(P^\nu) \leq \varphi(P).$$

Asymptotic confidence interval for this lower bound on the true optimal value $\varphi(P)$ can be obtained from CLT.

Asymptotic distribution of optimal solutions/ M-estimates.

In presence of constraints asymptotic normality of empirical optimal solutions $\mathbf{x}^*(P^\nu)$ cannot be expected even when all solution sets $\mathcal{X}^*(P)$ and $\mathcal{X}^*(P^\nu)\forall\nu$ are singletons. Intuitive sufficient conditions – true solution $\mathbf{x}^*(P) \in \text{int}\mathcal{X}$ or $\exists$ affine approximation of $\mathcal{X}$ at $\mathbf{x}^*(P)$. Idea of Chernoff (1954).

It is possible to prove that under reasonable assumptions, asymptotic distribution of optimal solutions $\mathbf{x}^*(P^\nu)$ is conically normal being projection of normal distribution on a convex cone. Asymptotic normality can appear only when the problem reduces on a neighborhood of the true optimal solution $\mathbf{x}^*(P)$ to an unconstrained one; cf. conditions for constrained linear L_1 regression (convex polyhedral $\mathcal{S}$, strict complementarity conditions valid at the true parameter value, positive densities) or convex $f(\bullet, \omega)$ and linear equality constraints (e.g. Niemiro).

General tool — generalized δ-method.
It requires certain differentiability property of optimal solution map $\mathbf{x}^*$ at P, and a suitable version of CLT.

Consult Shapiro for various results on SAA problem, i.e. for empirical distributions P^ν.

Comments

• CLT for $\nabla_x F(\mathbf{x}, P^\nu)$ is obtained, e.g., for $f(\bullet, \omega)$ convex $C^{1,1}$-functions for all ω, with square integrable Lipschitz constants, with a finite nonsingular variance matrix $\mathbf{V} = \text{var}[\nabla_x f_0(\mathbf{x}^*(P), \omega)]$, a finite expectation $E\|\nabla_x f_0(\mathbf{x}^*(P), \omega)\|^2$ and for empirical probability distributions P^ν. Assumption of empirical probability distributions can be relaxed.

• Differentiability assumptions concerning $f(\mathbf{x}, \omega)$ or $F(\mathbf{x}, P)$ restrict applicability of these results. To an extent, they can be relaxed.

• In case of explicit equality constraints one exploits mostly classical Lagrangian approach; e.g.Aitkinson & Silvey (1958).

• $\exists$ Extensions to smooth or convex inequalities, to wide-sense empirical programs. The key property is again validity of a version of CLT for (generalized) gradients of approximate objective functions.

Contributiors from SP: King, Korf, Pflug, Rockafellar, Römisch, Shapiro, Vogel, Wets, J.D., etc.
Contributiors from Statistics: Chernoff, Aitkinson - Silvey, Haberman, Niemiro, etc.

Measures of Robustness

- Influence function for optimal value can be obtained without using analytical form of solutions.
- Finite gross error sensitivity can hold true only under special requirements concerning contamination.
- Local shift sensitivity is related to the Lipschitz property of $f(x, \bullet)$.
- Gross-error breakdown type of characteristics appears e.g. in local stability analysis of the optimal value for probability dependent set of feasible solutions $\mathcal{X}(P)$ in (1).
- For SP Fisher consistency is not expected. ($T(F_\theta) = \theta, \ \forall \theta \in \Theta, \approx$ the model estimator assymptotically measures the right quantity.

APPLICATIONS – mainly for investment problems $\sim$ stress testing

- Resistance with respect to additional scenarios for bond portfolio management.
- Local stability results for risk measures, for Markowitz model and for mean-risk models.
- Global stability results for the optimal performance of CVaR minimizing portfolio, see Example from [3].

Contamination technique in SP I.

Qualitative and quantitative stability analysis for solutions of SP with respect to changes of probability distribution P comes out from general results of parametric programming; cf. Römisch and references therein.

Perturbations of P modeled via contamination $\rightsquigarrow$ less general but applicable results; general parameter P reduces to real scalar parameter λ. Local properties of optimal solutions and of optimal value can be obtained and are related with influence function approach in M-estimation; cf. J.D. [2].
Directly applicable global results can be proved for optimal value.

Contamination means to model perturbed probability distribution P as

$$P_\lambda = (1-\lambda)P + \lambda Q, \quad 0 \le \lambda \le 1, \tag{10}$$

i.e. to use probability distribution P contaminated by probability distribution Q. For fixed probability distributions P, Q, denote by $F(\mathbf{x}, P_\lambda)$ objective function in (5) computed for contaminated distribution and by

$$\varphi(\lambda) := \min_{\mathbf{x} \in \mathcal{X}} F(\mathbf{x}, P_\lambda)$$

the optimal value function.

Contamination technique in SP II.

Due to (10), expectation $F(\mathbf{x}, P_\lambda)$ is linear in parameter λ,

$$\frac{d}{d\lambda} F(\mathbf{x}, P_\lambda) = F(\mathbf{x}, Q) - F(\mathbf{x}, P)$$

and classical approaches of parametric programming can be exploited.

E.g. if (5) is convex program with fixed set $\mathcal{X}$ of feasible solutions, if $\mathcal{X}^*(P) \neq \emptyset$ and compact, $\mathcal{X}^*(Q) \neq \emptyset$ then optimal value $\varphi(\lambda)$ is finite concave function on $[0, 1]$, continuous at 0, with derivative at $\lambda = 0^+$

$$\varphi'(0^+) = \min_{\mathbf{x} \in \mathcal{X}^*(P)} F(\mathbf{x}, Q) - \varphi(0). \quad (11)$$

Bounds on optimal value $\varphi(\lambda)$ for an arbitrary $\lambda \in [0, 1]$ follow by properties of concave functions:

$$(1 - \lambda)\varphi(0) + \lambda\varphi(1) \leq \varphi(\lambda) \leq \varphi(0) + \lambda\varphi'(0^+) \quad \forall \lambda \in [0, 1] \quad (12)$$

which is a global robustness result.

Contamination technique in SP III.

Upper bound of derivative in (12) is $F(\mathbf{x}(P), Q) - \varphi(0)$ where $\mathbf{x}(P)$ is an arbitrary optimal solution of initial problem (5) obtained for probability distribution P. If the optimal solution is unique, this upper bound is attained.

Evaluation of bounds in (12) requires solution of another stochastic program of type (5) for new distribution Q to get $\varphi(1)$ and evaluation of expectation $F(\mathbf{x}(P), Q)$ at an already known optimal solution $\mathbf{x}(P)$ of initial problem (5) but for the contaminating distribution Q.

This approach provides global bounds on $\varphi(P_\lambda)$ also when $F(\mathbf{x}, P)$ is convex in $\mathbf{x}$ and concave in P.

Analytic form of results is not required – possible use for IF of complicated constrained estimates.

$\exists$ generalizations to nonconvex problems (1) and extensions to the set $\mathcal{X}$ dependent on P; the last results, however, are of a local character only.

Example – Stress testing for CVaR I.

VaR/CVaR – Basic formulas
$\mathbf{x} \in \mathcal{X} \subset R^n$ decision vector, $\omega \in \Omega \subset R^m$ random
$g(\mathbf{x}, \omega)$ loss if $\mathbf{x}$ selected, ω appeared
P probability distribution of ω, $\alpha \in (0, 1)$

Distribution function of loss $P\{\omega : g(\mathbf{x}, \omega) \leq k\} := G(\mathbf{x}, P; k)$

Value at Risk $\sim$ quantile of distribution $G(\mathbf{x}, P; k)$:

$$\mathrm{VaR}_\alpha(\mathbf{x}, P) = \min\{k \in R : G(\mathbf{x}, P; k) \geq \alpha\}$$

or

$$\mathrm{VaR}^+_\alpha(\mathbf{x}, P) = \inf\{k \in R : G(\mathbf{x}, P; k) > \alpha\}$$

Conditional Value at Risk—CVaR_α is mean of the α-tail distribution G_α of $g(\mathbf{x}, \omega)$ defined as

$$G_\alpha(\mathbf{x}, P; k) = 0 \quad \text{for } k < \mathrm{VaR}_\alpha(\mathbf{x}, P)$$

$$G_\alpha(\mathbf{x}, P; k) = \frac{G(\mathbf{x}, P; k) - \alpha}{1 - \alpha} \quad \text{for } k \geq \mathrm{VaR}_\alpha(\mathbf{x}, P).$$

Example – Stress testing for CVaR II.

ASSUME $E_P|g(\mathbf{x},\omega)| < \infty \, \forall \mathbf{x} \in \mathcal{X}$ and put

$$\Phi_\alpha(\mathbf{x},\psi,P) = \psi + \frac{1}{1-\alpha} E_P(g(\mathbf{x},\omega) - \psi)^+$$

FUNDAMENTAL MINIMIZATION FORMULA (Rockafellar & Uryasev)
As a function of ψ, $\Phi_\alpha(\mathbf{x},\psi,P)$ is finite and convex (hence *continuous*) with

$$\min_\psi \Phi_\alpha(\mathbf{x},\psi,P) = \mathrm{CVaR}_\alpha(\mathbf{x},P)$$

$$\arg\min_\psi \Phi_\alpha(\mathbf{x},\psi,P) = [\mathrm{VaR}_\alpha(\mathbf{x},P), \mathrm{VaR}_\alpha^+(\mathbf{x},P)]$$

Minimum of CVaR

$$\varphi_C(P) := \min_{\mathbf{x}\in\mathcal{X}} \mathrm{CVaR}_\alpha(\mathbf{x},P) = \min_{\psi,\mathbf{x}\in\mathcal{X}} \Phi_\alpha(P,\mathbf{x},\psi). \qquad (13)$$

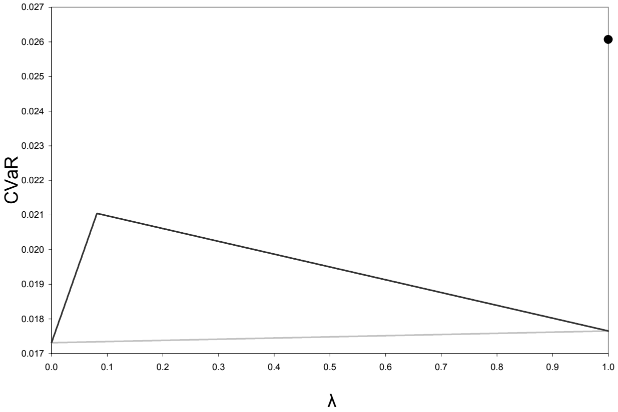
ASSUME $g(\mathbf{x},\omega)$ convex in $\mathbf{x}$, $\mathcal{X}$ compact convex independent of P.
DENOTE $\psi^*(P), \mathbf{x}_C^*(P)$ an optimal solution of (13).

Then directional derivative used in contamination bounds is

$$\varphi_C'(0^+) \le \Phi_\alpha(\mathbf{x}_C^*(P), \psi^*(P), Q) - \varphi_C(P).$$

Numerical example [3]

Scenario-based risk management problem, P is carried by scenarios ω^s, $s = 1, \ldots, S$ with probabilities p_s. DATA $\rightarrow$ monthly losses in EUR; P, Q each carried by 5184 equiprobable scenarios, $\alpha = 0.99$ and $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^n : \sum_i x_i = 1,\ 0 \le x_i \le 0.3\, \forall i\}$.



Non-zero components of optimal solutions

Quantity	Value
$\phi_C(P)$	0.01731
$\phi_C(Q)$	0.01765
$\Phi_\alpha(\mathbf{x}_C^*(P), \psi^*(P), Q)$	0.06309
$\Phi_\alpha(\mathbf{x}_C^*(Q), \psi^*(Q), P)$	0.02135
$x_1^*(P)$	0.1288
$x_7^*(P)$	0.2003
$x_9^*(P)$	0.3
$x_{10}^*(P)$	0.2647
$x_{11}^*(P)$	0.1062
$\psi^*(P)$	0.01365
$x_5^*(Q)$	0.1
$x_7^*(Q)$	0.3
$x_9^*(Q)$	0.3
$x_{10}^*(Q)$	0.3
$\psi^*(Q)$	0.01588
$\mathrm{VaR}(\mathbf{x}_C^*(P), Q)$	0.02607

New problems

• Set of feasible solutions in SP or constraints in constrained M-estimation depend on P,

cf. probabilistic constraints.

• Additional difficulties for nonconvex constraints;

$\exists$ some results for VaR in [3], also Shapiro [16].

References from abstract

[1] J. Dupačová (1987), Asymptotic properties of restricted L1-estimates of regression, in: Y. Dodge (ed.), *Statistical Data Analysis based on the L1 norm and Related Methods.* North-Holland, Amsterdam, pp. 263–274.

[2] J. Dupačová (1990), Stability and sensitivity analysis in stochastic programming. *Ann Oper Res* **27**, 115-142.

[3] J. Dupačová, J. Polívka, (2007), Stress testing for VaR and CVaR. *Quantitative Finance* **7**, 411–421.

[4] J. Dupačová, R. J.-B. Wets (1988), Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems. *Ann Stat* **16**, 1517–1549.

[5] A. Shapiro (2000), On the asymptotics of constrained local M-estimators. *Ann Stat* **28**, 948–960.

[6] A. Shapiro (2009), Statistical inference, Chapter 5 in: A. Shapiro, D. Dentcheva, A. Ruszczyński, *Lectures on Stochastic Programming.* MPS-SIAM Series on Optimization, SIAM, Philadelphia, pp. 155–252.

[7] R. J.-B. Wets (1991), Constrained estimation: Consistency and asymptotics. *Appl. Stoch. Models and Data Analysis* **7**, 17–32 and references therein

Additional references

[8] J. Dupačová (1991), On non-normal asymptotic behavior of optimal solutions for stochastic programming problems and on related problems of mathematical ststistics. *Kybernetika* **27**, 38–52.

[9] S. J. Haberman (1989), Concavity and estimation. *Ann Stat* **17**, 1631–1661.

[10] P. J. Huber (1981), *Robust Statistics*. Wiley-Interscience, Hoboken.

[11] A. J. King, R. T. Rockafellar (1993), Asymptotic theory for solutions in statistical estimation and stochastic programming. *Math. Oper. Res.* **18**, 178–162.

[12] W. Niemiro (1992), Asymptotics for M-estimators defined by convex minimization. *Ann Stat* **20**, 1514–1533.

[13] S. M. Robinson (2010),Equations on monotone graphs. *Math. Progr.*

[14] R. T. Rockafellar and S. Uryasev (2002), Conditional value-at-risk for general loss distributions. *J. Banking & Finance* **26**, 1443–1471.

[15] R. T. Rockafellar and J. R-B. Wets (1998), *Variational Analysis*. Springer.

[16] W. Römisch (2003), Stability of stochastic programming. In: A. Ruszczynski, A. Shapiro (Eds.), *Stochastic Programming*, Handbooks in OR & MS **10**. Elsevier, pp. 483–554.