Robustness for Probabilistic Programs

Jitka Dupačová

Dept. of Probability and Mathematical Statistics, Charles University, Prague, Czech Republic

Groningen, January 27, 2012 (extended)

Outline

- The model
- Survey of achievements – the milestones
  - 1958 – 1970 model building
  - 1971 convexity issue
  - 1978 successful applications, software development
  - 1990 *p*-level efficient points
  - 1990 stability and robustness
- Hot problems
  - applications, software development
  - stability and robustness also for discrete distributions
  - worst-case analysis
  - extensions to FSD constraints
- Main quoted references

The model

We shall deal with probabilistic programs

$$\min_{x \in \mathcal{X}} G_0(x, P) := E_P f_0(x, \omega) \text{ subject to} \tag{1}$$

$$G_j(x, P) \leq 0, j = 1, \ldots, J$$

where $G_j(x, P) \leq 0$ are probabilistic constraints such as

$$P(\omega : g(x, \omega) \leq 0) \geq 1 - \varepsilon, \tag{2}$$

- $\mathcal{X} \subset \mathbb{R}^N$ is fixed nonempty closed set,
- P is known probability distribution of random parameter ω whose support Ω is closed subset of $\mathbb{R}^M$,
- $g : \mathbb{R}^N \times \Omega \to \mathbb{R}^K$ and
- $\varepsilon \in (0, 1)$ fixed, chosen by decision maker or prescribed by regulations.

Individual probabilistic constraints correspond to $K = 1$, for joint probabilistic constraints $K > 1$.

Denote $\mathcal{X}(P)$ set of feasible solutions, $\mathcal{X}^*(P)$ set of optimal solutions, $\varphi(P)$ optimal value of the objective function in (1), (2).

Probabilistic constraints

Probabilistic constraints (2) are reliability type constraints which can be written as

$$H(x,P) := P(\omega : \max_k g_k(x,\omega) \le 0) := P(\omega : \omega \in \mathcal{H}(x)) \ge 1-\varepsilon \quad (3)$$

with $\mathcal{H}(x) := \{y \in \mathbb{R}^M : g_k(x,y) \le 0 \forall k\}$. Constraints in (2) can be equivalently written with expectation of characteristic function

$$\mathcal{X}(P) = \{x \in \mathcal{X} : E_P I_{\mathcal{H}(x)}(\omega)) \ge 1-\varepsilon\} \quad (4)$$

Contrary to common static stochastic programs

$$\min_{x\in\mathcal{X}} E_P f(x,\omega) \quad (5)$$

constraints depend on P.
Moreover, the integrand in (4) is not smooth and the resulting optimization problem is typically nonconvex one. $\rightsquigarrow$ Reasons why are probabilistic programs classified as hard optimization problems.

Survey of achievements – The milestones

- 1958 Charnes, Cooper and Symonds – randomness influences constraints $g_k(x) \le 0$ of a nominal deterministic MP $\rightsquigarrow$ use reliability type individual probabilistic, chance constraints

$$P(g_k(x,\omega) \le 0) \ge 1 - \varepsilon_k;$$

  for each constraint separately. P is a known probability distribution of random parameter ω on $\Omega \subset \mathbb{R}^M$ and probability levels $\varepsilon_k \in (0,1)$ are fixed, chosen by the decision maker.
  Ignores stochastic dependence.
- 1970 – Use (one or more) joint probabilistic constraints

$$P(g_k(x,\omega) \le 0, \; k = 1,\dots,K) \ge 1 - \varepsilon. \quad (6)$$

  Hard problems due to lack of convexity and smoothness.
  Breakthrough Prékopa 1971 $\rightarrow$ it is possible to deal efficiently with problems that belong to one of convex programming subclasses.

α-concave functions

Nonnegative function $f(x)$ defined on convex set $\mathcal{C} \subset \mathbb{R}^N$ is α-concave with $\alpha \in [-\infty, \infty]$ if for all $x, y \in \mathcal{C}$, $\lambda \in [0, 1]$

$$f(\lambda x + (1-\lambda)y) \geq m_\alpha(f(x), f(y), \lambda).$$

Function $m_\alpha : \mathbb{R}_+ \times \mathbb{R}_+ \times [0, 1] \rightarrow \mathbb{R}$ is defined as $m_\alpha(a, b, \lambda) = 0$ if $ab = 0$, and for $a > 0$, $b > 0$, $0 \leq \lambda \leq 1$

$$m_\alpha(a, b, \lambda) = \begin{cases} a^\lambda b^{1-\lambda} & \text{if } \alpha = 0, \quad \text{i.e. } f \text{ log-concave} \\ \max[a, b] & \text{if } \alpha = \infty, \quad \text{i.e. } f \text{ quasi-convex} \\ \min[a, b] & \text{if } \alpha = -\infty, \quad \text{i.e. } f \text{ quasi-concave} \\ (\lambda a^\alpha + (1-\lambda)b^\alpha)^{1/\alpha} & \text{otherwise.} \end{cases}$$

Apply suitable transforms of α-concave distribution functions.

$\exists$ extension to nonconvex sets $\mathcal{C}$; important for discrete distributions:

Distribution function F is α-concave on $\mathcal{A} \subset \mathbb{R}^N$ if $x, y, z \in \mathcal{A}$, $\lambda \in (0, 1)$, $z \geq \lambda x + (1-\lambda)y \implies F(\lambda x + (1-\lambda)y) \geq m_\alpha(F(x), F(y), \lambda)$.

Convexity for Probabilistic Constraints

General result about convexity of sets defined by (6), cf. Theorem 4.39 in [S-D-R] – extension of Prékopa's theorem.

Theorem

Assume $g_k : \mathbb{R}^N \times \mathbb{R}^M \to \mathbb{R}\, \forall k$ *quasi-convex,* $\omega \in \mathbb{R}^M$ *random vector with* α*-concave probability distribution, then* $H(x,P) := P(\omega : g_k(x,\omega) \leq 0\, \forall k)$ *is* α*-concave on*

$$\mathcal{D} := \{x \in \mathbb{R}^N : \exists y \in \mathbb{R}^M \text{ s.t. } g_k(x,y) \leq 0\, \forall k\}.$$

PROBLEM: Joint quasi-convexity of $g_k(x,\omega) \longrightarrow$
Theorem is applicable e.g. for $g_k(x,\omega) = -g_k(x) + \omega_k,\ \forall k$ i.e. for separable joint probabilistic constraints.
Prominent standing of separable joint constraints with α-concave distribution of right-hand sides and convex $g(x)$.

Another favorable class – linear ChC with joint normal distribution of coefficients; cf. [Prekopa], [Henrion]. Convexity also for individual probabilistic constraints with radial distributions cf. [Calafiore& El Ghaoui], uniform over a convex set.

Derivatives of probability function $H(x, P)$

Derivatives are expressed as surface or volume integrals, boundary of set $\mathcal{H}(x) := \{y \in \mathbb{R}^M : g_k(x, y) \leq 0 \, \forall k\}$ plays an important role. The first result was due to Raik for case of $\mathcal{H}(x)$ defined by one function $g : \mathbb{R}^N \times \mathbb{R}^M \to \mathbb{R}$. For more general results cf. Marti and Uryasev, their summary is in [S-D-R].

Let $\mathcal{H}(x) = \{y \in \mathbb{R}^M \, : \, g(x, \omega) \leq 0\}$, $g : \mathbb{R}^N \times \mathbb{R}^M \to \mathbb{R}$. Assume that probability distribution P has density $\theta(z)$, its support Ω is closed set with piecewise smooth boundary, g is continuously differentiable and such that

$$\mathrm{bd}\mathcal{H}(x) = \mathcal{S}(x) = \{z \in \Omega \, : \, g(x, z) = 0\},$$

$M - 1$-dimensional surface of $\mathcal{H}(x)$. Let $\nabla_z g(x, z) > 0$, $\|\nabla_x g(x, z)\| > 0$. Then dervative of $H(x, P)$ can be expressed as a surface integral

$$\left(\frac{\partial H(x, P)}{\partial x_i} \right)_{i=1}^{N} = \int_{\mathrm{bd}\mathcal{H}(x) \cap \Omega} \frac{\theta(z)}{\|\nabla_z g(x, z)\|} \nabla_x g(x, z) d\mathcal{S}.$$

Simplification for separable problems $\rightsquigarrow$ derivatives of (continuous) distribution function.

Base for applications and software development

Minimization of objective function subject to constraints (6) is nonlinear programming problem and in principle, known NLP algorithms can be adapted provided that checking feasibility is is easy and the resulting feasible region is convex.

Example – linear constraints with normally distributed rhs.

• Around 1978 – Successful applications, cf. volume [SzTAKI] edited by Prékopa.

In general, even to test feasibility of a candidate solution, i.e. evaluation of $P(g(x,\omega) \leq 0)$ turns to be a rather demanding task. Approximation, simulation and bounding techniques have been developed.

Interest in probabilistic programs with discrete distributions that appear in approximations and/or in problem formulation. Use of integer variables.

Situation is much better for integrated chance constraints, cf. Klein Haneveld: "Remove" the 0-1 characteristic function.

$$P(\omega : g(x,\omega) \geq 0) \geq \varepsilon \Longleftrightarrow E[I(g^{-}(x,\omega))] \leq 1 - \varepsilon$$

relaxed to $E[g(x,\omega)^{-}](= \int_{-\infty}^{0} P(\omega : g(x,\omega) \leq t)dt)$.

• Around 1990 – use of p-level efficient points (Prékopa, Sen and others)

Discrete distributions

Consider now finitely discrete distribution

$$P(\omega = \omega^s) = p_s,\ p_s > 0,\ s = 1, \ldots, S,\ \sum_s p_s = 1 \qquad (7)$$

$\mathcal{X}$ compact and (vector) functions $g(\bullet, \omega^s)$ continuous for all s. Rewrite constraints (6) as follows:
For each $s \in \{1, \ldots, S\}$ introduce binary variable z^s such that $z^s = 0$ guarantees that $g(x, \omega^s) \le 0\, \forall x \in \mathcal{X}$ and K-dimensional vector M_s whose components are sufficiently large. For deterministic objective function G_0 the problem is

$$\text{minimize } G_0(x) \text{ subject to}$$

$$g(x, \omega^s) - M_s z^s \le 0,\ s = 1, \ldots, S \qquad (8)$$

$$\sum_s p_s z^s \le \varepsilon, \qquad (9)$$

$$x \in \mathcal{X},\ z^s \in \{0, 1\}\, \forall s.$$

For convex G_0 and convex $g(\bullet, \omega^s)\, \forall s \rightarrow$ convex mixed integer program. Large size problem, various approaches to solving it efficiently have been elaborated.

p-efficient points

For separable functions $g(x, \omega)$ constraints can be expressed by means of distribution function F, $F(g(x)) \geq 1 - \varepsilon$. New solution techniques:

Definition

Point $v \in \mathbb{R}^m$ is p-efficient point of distribution function F if $F(v) \geq p$ and there is no $y \leq v$, $y \neq v$ such that $F(y) \geq p$.

The p-efficient points are minimal elements of the p-level set

$$\mathcal{Z}_p = \{z \in \mathbb{R}^M \ : \ P(\omega \leq z) \geq p\}$$

of distribution function F for value p. Constraints $F(g(x)) \geq p$ can be written as $g(x) \in \mathcal{Z}_p$. $\rightsquigarrow$

$$\min_x \{G_0(x) \ : \ g(x) \in \mathcal{Z}_p, \ x \in \mathcal{X}\}, \qquad (10)$$

$\mathcal{Z}_p \neq \emptyset$, closed for all $p \in (0, 1)$, mostly nonconvex.
If v^j, $j \in J$ are all p-efficient points of F, we get representation

$$\mathcal{Z}_p = \bigcup_{j \in J} K_j, \ K_j = \{v^j\} + \mathbb{R}^M_+$$

and constraints $g(x) \in \mathcal{Z}_p$ in (10) can be written as

$$g(x) \geq z \geq \sum_j \lambda_j v^j, \ \lambda_j \in \{0, 1\}, \ \sum_j \lambda_j \geq 1. \qquad (11)$$

p-efficient points – cont.

For finitely discrete distributions (7) $\exists$ FINITE No. of p-efficient points and $\mathcal{Z}_p$ is convex for $p > 1 - \min_s p_s$. Constraints in (10) are then

$$g(x) \geq \omega^s,\ s = 1, \ldots, S,\ x \in \mathcal{X};$$

cf. permanently feasible solutions.

For finite J – solve J problems separately for each p-level point v^j to get

$$\phi_j = \min\{G_0(x) : g(x) \geq v^j,\ x \in \mathcal{X}\},$$

the optimal value equals $\varphi(P) = \min_j \phi_j$. For convex G_0, concave vector function g and convex $\mathcal{X}$, one solves then many convex programs.

Another possibility – allow in disjunctive constraints (11) $\lambda_j \in [0,1]\,\forall j, \sum \lambda_j = 1$ and solve relaxed convex problem which provides a lower bound for $\varphi(P)$.
We have conv$\mathcal{Z}_p = conv\{v^j,\ j \in J\} + \mathbb{R}^M_+$, closed and its extreme points are contained in $\mathcal{Z}_p \rightsquigarrow$ convex hull problem with $\mathcal{Z}_p$ in (10) replaced by conv$\mathcal{Z}_p$.

Discrete separable problems

For α-concave distribution F the set $\mathcal{Z}_p$ is convex. Extreme points of $\mathcal{Z}_p$ are contained in set of p-efficient points and for finitely discrete distributions on $\mathbb{Z}^M$ set of p-efficient points is nonempty, finite and (10) can be replaced by

$$\min_{x,\lambda,z} G_0(x) \text{ subject to } x \in \mathcal{X},\ \lambda_j \geq 0, z \in \mathbb{Z}^M$$

$$g(x) \geq z,\ z \geq \sum_{j \in J} \lambda_j v^j,\ \sum_{j \in J} \lambda_j = 1,$$

cf. [S-D-R], Prékopa, Dentcheva, Ruszczyńsky and others. For solving it, knowledge of p-efficient points is crucial.

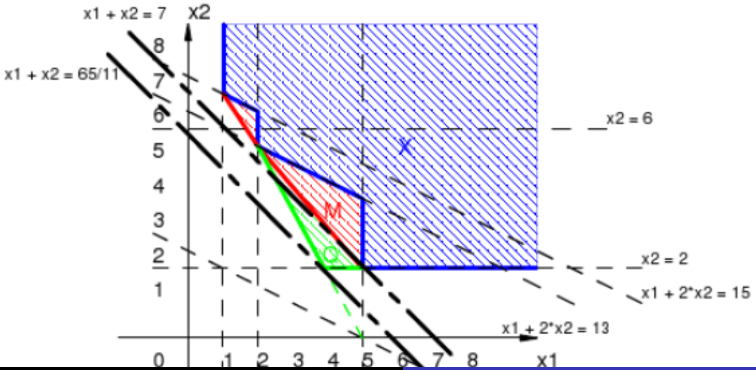
$\exists$ Numerical techniques which combine generation of p-efficient points and optimization.

Example

There are 3 p-efficient points of a discrete distribution F of ω, $(1, 15, 6), (2, 13, 2), (5, 5, 2)$, in

$$\min\{x_1 + x_2 \,:\, x_1 \geq \omega_1, x_1 + 2x_2 \geq \omega_2, x_2 \geq \omega_3\}.$$

The blue set of feasible solutions of (10) is nonconvex, its convex hull means an extension for red set M, for relaxed problem green set is added. p-efficient points appear as right-hand sides of the parallel lines of constraints.



Approximations and solution techniques

Besides of simplification to individual probabilistic constraints with carefully chosen thresholds, ∃ number of approaches that propose approximation problems which are convex and yield feasible or likely to be feasible solutions to the original probabilistic program; balance between tractability and feasibility.

Use Monte Carlo technique – replace P by the empirical distribution P^ν and solve the "sample" based probabilistic program – Sample Average Approximation $\rightsquigarrow$ probabilistic program, with different, discrete probability distribution. Also ε is replaced by a different risk level.

∃ asymptotic results, including rates of convergence, estimates of required sample size; cf. Ahmed, Luedtke, Nemirovski, Pagoncelli, Shapiro and others.

Another idea, cf. [Calafiore, Campi] is to select finite number of randomly chosen ω^s, $s = 1, \ldots, S$ and solve deterministic nonlinear program with constraints

$$g_k(x, \omega^s) \leq 0,\ k = 1, \ldots, K,\ s = 1, \ldots, S.$$

Convexity of original problem implies convexity of this sampled program and when sufficient number of samples is used, the obtained solution fails to satisfy only small portion of original constraints.

Approximations and solution techniques – cont.

Approaches to solve numerically the "big-M" problem for discrete distributions depend on structure of (8). (The origin of scenarions is not that important.)

- Using p-efficient points for separable linear problems [Prékopa et al]

For nonlinear nonseparable problems

- Lagrangian relaxation wrt. constraint (9) provides a scenario decomposable problem that can be solved by means of Progressive Hedging Algorithm, cf. [WWW]
- Various upper and lower bounding schemes based on Lagrangian relaxation or other valid inequalities
- Branch & bound techniques, e.g. [Ruszczynski]

Robustness properties

To solve complex probabilistic programs one tends to simplify or reformulate the model (e.g. replace joint probability constraints by individual ones), to approximate the probability distribution, etc. These approximations and simplifications ask for development of suitable validation techniques and for stability and robustness tests.

$\exists$ qualitative stability results wrt. ε, g, $\mathcal{X}$, e.g. [Henrion]

Moreover, probability distribution P itself is not known completely $\longrightarrow$
$\exists$ two sources of uncertainty and errors
& one wishes to get solutions reliable enough to support sensible decisions.

• Since 1990 – various stability results developed mainly for special convex classes of probability distributions, cf. [Römisch in R-S] and references therein.

Stability results and robustness wrt. P

Key result on stability for

$$\min_x G_0(x, P) \text{ s.t. } x \in \mathcal{X},\ P(g_k(x,\omega) \le 0,\ k = 1, \dots, K) \ge 1 - \varepsilon \quad (12)$$

is Theorem 5 of [Römisch in R-S]. It provides upper semicontinuity of set of optimal solutions and local Lipschitz property of optimal value function if the objective function $G_0(x, P)$ is Lipschitz continuous on $\mathcal{X} \cap \mathrm{cl}U$ and a metric regularity holds at each optimal solution of unperturbed problem.
Basic assumption: $\mathcal{X}^*(P) \neq \emptyset$ and belongs to open bounded set $\mathcal{U}$.

Definition

Consider arbitrary mapping $\mathcal{S} : \mathcal{X} \to \mathcal{Y}$, $(x, y) \in \mathrm{gph}\mathcal{S}$. Then $\mathcal{S}$ is metrically regular around $(\bar{x}, \bar{y}) \in \mathrm{gph}\mathcal{S}$ if $\exists \kappa > 0$ such that

$$d(x, \mathcal{S}^{-1}(y)) \le \kappa d(y, \mathcal{S}(x))\, \forall x \in \mathcal{N}(\bar{x}),\ y \in \mathcal{N}(\bar{y}). \quad (13)$$

(cf. Theorem 9.43 of [R-W].)

Metric regularity is related with continuity of constraint set $\mathcal{X}(P)$ when some perturbation is considered. Proper choice of probability distance is important and requirements concern mainly the unperturbed problem.

Robustness analysis via Contamination

was developed and applied for $\mathcal{X}(P)$ independent of P and for expectation type objective $G_0(x, P)$.
Assume that such SP was solved for P, finite optimal value $\varphi(P)$.
Changes in probability distribution P are modeled using contaminated distributions

$$P(t) := (1-t)P + tQ, \; t \in [0,1]$$

with Q another *fixed* probability distribution such that $\varphi(Q)$ is finite.
Via contamination, robustness analysis wrt. changes in P gets reduced to much simpler analysis of parametric program with scalar parameter t.
Objective function is linear in $P \Longrightarrow$
$G_0(x,t) := G_0(x, P(t)) = (1-t)G_0(x,P) + tG_0(x,Q)$ is linear wrt. t
$\Longrightarrow$ optimal value function $\varphi(t) := \min_{x\in\mathcal{X}} G_0(x,t)$ is concave on $[0,1]$
$\Longrightarrow$ continuity and existence of directional derivatives in $(0,1)$.
Continuity at $t=0$ is property related with stability for SP. In general, one needs set of optimal solutions $\mathcal{X}^*(P) \neq \emptyset$, bounded.
Concave $\varphi(t) \Longrightarrow$ global contamination bounds

$$\varphi(0) + t\varphi'(0^+) \geq \varphi(t) \geq (1-t)\varphi(0) + t\varphi(1), \; t \in [0,1]. \qquad (14)$$

They quantify change in optimal value due to considered perturbations.
Choice of Q?

Application to stress testing for scenario-based SP

Stochastic program for risk management or for portfolio optimization solved for fixed set of scenarios ω^s, $s = 1, \dots, S$ $\longrightarrow$
P – probability distribution concentrated at ω^s, $s = 1, \dots, S$ with probabilities p_s. Reformulation needed:

$$\min_{x \in \mathcal{X}} G_0(x, P) := \sum_s p_s u^s(x)$$

with *fixed* set of first-stage feasible solutions (initial investments) and with convex performance measures u dependent on scenarios (covers static, two-stage, multistage SP).

ORIGIN OF SCENARIOS?

Inclusion of other out-of-sample or *stress* scenarios – another discrete probability distribution Q carried by out-of-sample or stress scenarios indexed by $\sigma = 1, \dots, S'$, with probabilities q_σ. Contaminated probability distribution $P(t)$ is carried by *pooled* sample of $S + S'$ scenarios that occur with probabilities

$$(1-t)p_1, \dots, (1-t)p_S, tq_1, \dots, tq_{S'}.$$

Bounds for the optimal value $\varphi(P(t))$ of problem based on pooled sample – test of robustness, stress test; cf. [ALM], [D-P]

Contamination bounds – constraints dependent on P

Denote $\mathcal{X}(t) = \{x \in \mathcal{X} : G(x, P(t)) \leq 0\}$, $\varphi(t)$, $\mathcal{X}^*(t)$ the set of feasible solutions, the optimal value and the set of optimal solutions of contaminated problem

$$\text{minimize } G_0(x, P(t)) \text{ on the set } \mathcal{X}(P(t)). \tag{15}$$

The task: to construct computable lower and upper bounds for $\varphi(t)$ & exploit them for robustness analysis with respect to inclusion of additional scenarios etc.

New problems – $\varphi(t)$ is no more concave in t.

$\exists$ formulas for directional derivative $\varphi(0^+)$ based on Lagrange function $L(x, u, t) = G_0(x, P(t)) + uG(x, P(t))$ for contaminated problem.

Generic form

$$\varphi'(0^+) = \min_{x \in \mathcal{X}^*(0)} \max_{u \in \mathcal{U}^*(x,0)} \frac{\partial}{\partial t} L(x, u, 0).$$

Thanks to the assumed structure of perturbations lower bound can be derived for $G(x, P)$ linear (or concave) with respect to P without any smoothness or convexity assumptions with respect to x,

However, to get at least local upper bound (14) means to get $\mathcal{X}(t)$ fixed for t small enough. To this purpose, various assumptions are needed.

Lower bound

1. One constraint dependent on P and objective G_0 *independent* of P: For contaminated probability distribution $P(t)$ we get

$$\min_{x \in \mathcal{X}} G_0(x) \text{ subject to } G(x,t) := G(x, P(t)) \le 0 \tag{16}$$

Theorem

Let $G(x, \bullet)$ in (16) be linear (or concave) function of $t \in [0,1]$ and optimal value function

$$\varphi(t) := \min_{x \in \mathcal{X}} G_0(x) \text{ subject to } G(x,t) \le 0$$

be finite on $[0,1]$. Then $\varphi(t)$ is quasiconcave in $t \in [0,1]$ with lower bound

$$\varphi(t) \ge \min\{\varphi(1), \varphi(0)\}. \tag{17}$$

2. Objective function G_0 also depends on probability distribution

$$\min_{x \in \mathcal{X}} G_0(x,t) := G_0(x, P(t)) \text{ subject to } G(x,t) \le 0. \tag{18}$$

For $G_0(x,P)$ linear (or concave) in P, lower bound can be obtained by application of (17) separately to $G_0(x,P)$ and $G_0(x,Q)$.
Notice that no convexity assumptions with respect to x were needed.

Upper bound

To derive upper bound for optimal value of contaminated problem with probability dependent constraints we assume that $G(x, t)$ is linear in t on interval $[0, 1]$.

1. Assume first that for an optimal solution $x(0)$ of (18), the constraint is not active, i.e. $G(x(0), 0) < 0$. Then there exists $t_0 > 0$ such that $G(x(0), t) \leq 0 \rightarrow$ trivial local upper bound

$$\varphi(t) \leq G_0(x(0), t) = (1 - t)\varphi(0) + tG_0(x(0), 1)\, \forall t \in [0, t_0]. \qquad (19)$$

2. Assume that $x(0) \in \mathcal{X}^*(0) \cap \mathcal{X}(1)$. Then $x(0) \in \mathcal{X}(t)\, \forall t \in [0, 1] \Rightarrow \varphi(t) \leq G_0(x(0), t) = (1 - t)\varphi(0) + tG_0(x(0), 1)$, i.e. (19) is upper bound valid for all $t \in [0, 1]$.

Notice that trivial upper bound (19) holds true without any convexity or smoothness assumptions and for arbitrary distribution Q. For G_0 independent of t, it is in agreement with quasiconcavity of $\varphi(t)$.

3.Upper bound $\varphi(t)$ can be constructed whenever $\exists$ feasible solution $\hat{x} \in \mathcal{X}(P_t) \rightarrow \varphi(t) \leq G_0(\hat{x}, t)$.

Upper bound – cont.

Direct search for $\hat{x} \in \mathcal{X}$ which satisfies constraints

$$G_j(x,0) \le 0 \,\forall j \text{ and } G_j(x,1) \le 0 \,\forall j$$

may be manageable, namely, when $Q = \delta_{\omega^*}$ is degenerated probability distribution: It means to augment $\mathcal{X}$ by deterministic constraints $g_k(x,\omega^*) \le 0,\ k \in K_j,\ j = 1,\dots,J$.

For problems with one joint probability constraint one may solve

$$\min_{x\in\mathcal{X}} G(x,1) \text{ subject to } G(x,0) \le 0.$$

The above ideas do not exploit parametric form of constraints in $\mathcal{X}(P_t)$. Instead – for problems with one joint probabilistic constraint solve

$$\min_{x\in\mathcal{X}} [(1-t)G(x,0) + tG(x,1)] \tag{20}$$

for increasing values of $t \rightsquigarrow \hat{x} \in \mathcal{X}(P_t)$ and upper bound $\varphi(t) \le G_0(\hat{x}, t)$.

Illustrative example

In jointly constrained probabilistic program of [P-A-S]

$$\begin{aligned} &\min x_1 + x_2 \text{ subject to} \\ &P(\omega_1 x_1 + x_2 \geq 7, \omega_2 x_1 + x_2 \geq 4) \geq 1 - \varepsilon, \\ &x_1 \geq 0,\ x_2 \geq 0 \end{aligned} \tag{21}$$

random components (ω_1, ω_2) are independent and have uniform distributions on intervals $[1, 4]$ and $[1/3, 1]$. It is convex program. Independence $\rightarrow$ explicit form of optimal solution can be obtained directly: $x_1^*(P) \doteq 3.6735$, $x_2^*(P) \doteq 2.7755$ and $\varphi(P) \doteq 6.4480$ for $\varepsilon = .05$.

To stress distribution P we choose extremal scenario $(\omega_1^*, \omega_2^*) = (1.02, 0.34)$. Optimal solution $x_1^*(P)$, $x_2^*(P)$ is infeasible for $t = 1$, $x_1^*(Q) \doteq 4.4118$, $x_2^*(Q) \doteq 2.5000$ and $\varphi(Q) \doteq 6.9118$.

Hence, for all $0 \leq t \leq 1$ lower bound (17) for $\varphi(t)$ is $\varphi(P) \doteq 6.4480$.

Solution $\hat{x}_1 = 4.4725$, $\hat{x}_2 = 2.4994$ of "upper bound problem" (20) for $t = 0$ is feasible for all contaminated problems $(7.0614 \geq 7, 4.02 > 4)$.

Then, $6.9719 = \hat{x}_1 + \hat{x}_2$ is upper bound for $\varphi(t)\, \forall t$.

Local upper bound via NLP stability results

If for all optimal solutions $G(x(0), 0) = 0$ and $G(x(0), 1) > 0$, a detailed analysis is needed. Second order analysis as done in [D-K] offers such possibility. However, we want to avoid using higher order differentiability.

For differentiable functions G_j in (1) – (2) properties of $\mathcal{X}(t) = \mathcal{X}(P_t)$ for small t follow from results of cf. [Robinson], [B-S]. Linear independence condition at $x^*(0)$ implies that $x^*(0)$ is nondegenerate point, vector $u^*(0)$ of Lagrange multipliers is unique and problem (15) can be locally reduced to one with a *fixed set of feasible solutions:*

$$\min_z G_0(T(z, t), t) \text{ on a set } \mathcal{C} \tag{22}$$

where $T(z, t)$ is continuously differentiable and $T(0, 0) = x^*(0)$. However, linearity of objective function with respect to t gets lost. This can be compared to situation described in detail in Example 1 of [B-D] for stochastic linear program with individual probabilistic constraints and random right-hand sides ω_k. Using quantiles of marginal probability distributions, problem can be cast into linear program with dual feasible set independent of P. But quantiles of contaminated marginal probability distributions – parameter dependent coefficients in dual objective function – are not linear in t.

Perturbed discrete distributions

Consider separable probabilistic program with finitely discrete distribution (7). Let $\mathcal{Z}$ contain all possible atoms of ω. The p-level set $\mathcal{Z}_p$ and p-efficient points depend on $p, \mathcal{Z}$ and on probabilities of atoms of P. With fixed p and $\mathcal{Z}$, scenario probabilities p_s are the ingredients which may be perturbed.

When $F(v^j) > p$ for all p-efficient points $v^j, j \in J$, sufficiently small changes in probabilities do not influence the set of p-efficient points and the set $\mathcal{Z}_p$ persists:
For example, to increase probability of $z^* \in \text{int}\mathcal{Z}_p$ means to contaminate P by δ_{z^*}. Set $\mathcal{Z}_p$ does not change if $0 \leq t \leq t_1$,
$t_1 = 1 - p[\min_j F(v_j)]^{-1}$. A similar result holds true if $z^* \notin \mathcal{Z}_p$.
When $F(v^*) = p$ for p-level point $v^j = v^*$ of P, then $t_1 = 0$. Increasing probability of v^* means that contaminated $F_t(v^*) > p$. To keep other $F_t(v^j) \geq p$ contamination parameter is limited to $0 \leq t \leq t_2$,
$t_2 = 1 - p[\min_{j \neq *} F(v_j)]^{-1}$. See [D-P] for a similar procedure applied to VaR contamination.

Possible extension of set $\mathcal{Z}$ can be treated by contamination technique, too: Start with an augmented set $\tilde{\mathcal{Z}} = \mathcal{Z} \cup \{z^*\}$ for which probability of z^* is 0 and apply the above idea to increase its probability.

Various limitations for robustness analysis including contamination technique

- Metric regularity needed in robustness analysis
  Bonnans-Shapiro, Rockafellar-Wets, Mordukhovich
  Henrion, Schultz, Römisch
  directional regularity is OK for contamination; choice of Q?
- Convexity of $\mathcal{X}(P)$
  Prekopa, Borel, ...
- Differentiability of probabilistic constraints
  Marti, Uryasev, Raik
- Discrete distributions
  Using properties of p-efficient points $\rightsquigarrow$ $\exists$ good chances for separable constraints

Importance of lower bound – no assumptions needed

Open possibilities: Try to apply other types of bounds e.g.

$$|\varphi(P) - \varphi(P_t)| \leq Ld(P, P_t) \text{ for } t \text{ small enough}$$

or lower bounds based on duality.

Alternative approaches

General lower bound, but limited possibilities to construct local upper contamination bounds for nonconvex probabilistic programs when differentiability cannot be guaranteed; trivial upper bounds are exception. Indirect approach was suggested in [B-D]: Apply contamination technique to penalty reformulation of probabilistic program. Then set of feasible solutions does not depend on P and for approximate problem, global bounds (14) follow. See Example 4 of [B-D] for numerical results. Bounds for optimal value of probabilistic program follow by worst-case analysis with respect to a whole set $\mathcal{P}$ of considered probability distributions, e.g. [P-W]. It means to hedge against all probability distributions belonging to chosen ambiguity set $\mathcal{P}$ and to solve

$$\min_{x \in \mathcal{X}} \max_{P \in \mathcal{P}} G_0(x, P) \text{ subject to} \tag{23}$$

$$P(\omega : g(x, \omega) \leq 0) \geq 1 - \varepsilon \quad \forall P \in \mathcal{P}. \tag{24}$$

or subject to

$$\min_{P \in \mathcal{P}} P(\omega : g(x, \omega) \leq 0) \geq 1 - \varepsilon. \tag{25}$$

Worst-case analysis I.

Problem (23)–(24) or (25) need not be more complicated than the underlying probabilistic program. Its tractability depends on function $g(x, \omega)$ and on choice of $\mathcal{P}$. Bounding expectations, in our case – bounding probabilities $P(\omega : g(x, \omega) \leq 0)$ in (25), for fixed x and for P belonging to convex class of probability distributions has got long tradition for scalar random variables, say, $\eta := g(x, \omega)$ and for P defined by known moment values. Additional information e.g. symmetry and/or unimodality can be incorporated by transformation or by duality arguments [Popescu MOR]. Given mean M and variance σ^2 of η the best upper bounds on $P(\eta \geq a)$ inclue one-sided Chebyshev inequality.

Distribution	$a > M$	$a \leq M$
arbitrary	$\frac{\sigma^2}{\sigma^2+(M-a)^2}$	1
symmetric	$1/2 \min\{1, \frac{\sigma^2}{(M-a)^2}\}$	1
symmetric unimodal	$1/2 \min\{1, \frac{4\sigma^2}{9(M-a)^2}\}$	1

Table: Optimal $M - \sigma^2$ upper bounds for $P(\eta \geq a)$.

Direct use of known results based on moment problem is limited because convexity of integrands $I_{\mathcal{H}(x)}(\omega))$ in ω is rare.

Worst-case analysis II.

Multidimensional extensions are demanding, are based on duality arguments; cf. [Bertsimas&Popescu], [Popescu MOR]. For P with given mean, covariance matrix and polyhedral or elliptic support (23)–(24) can be solved via semidefinite optimization. $\exists$ Results also for using higher order moments.

Application simplifies if $g(x,\omega) = x^\top \omega$ and $E_P\omega = \mu$, $\mathrm{var}\omega = \Sigma$ are prescribed [Popescu]. Then for fixed x, $E_P\eta = \mu^\top x$ and variance of η is $x^\top \Sigma x \rightsquigarrow$ known one-dimensional results.

Main field of interest – Robust portfolio optimization under VaR constraints such as (24) or (25) with $\mu - \Sigma$ class of probability distributions.

Other favorable classes $\mathcal{P}$:

- Finite list of proposed "scenarios" $(\mu - \Sigma)^s$ & ideas of multiobjective optimization;
- Relative entropy neighborhood of nondegenerate Gaussian $N(\mu, \Sigma)$ distribution [ElGhaoui et al.];
- Prokhorov neighborhood of nominal P cf. [Erdogan & Iyengar] or Kantorovich distance cf. [P-W].

See also Zhu & Fukushima, Fabozzi et al. Results depend on input information. Their stability should be studied similarly as in [Optim] or

FSD constraints - added complexity

Wish - observe stochastic ordering of outcomes. Concept of stochastic ordering was introduced in statistics already in the 1940-ties and is known as the first order stochastic dominance (FSD):

Definition

We say that a random variable X dominates a random variable Y in the first order if $P(X \leq u) \leq P(Y \leq u) \quad \forall u \in \mathbb{R}$.

In financial applications constraints based on the FSD allow us to incorporate random benchmarks (defined on the same probability space) instead of fixed thresholds.

HOWEVER FSD constraints are expressed in general as continuum of probabilistic constraints. Reduction possible for discrete distributions.

The second order stochastic dominance (SSD) constraints are easier to deal with. For comparison of portfolios, SSD constraints can be formulated as continuum of CVaR constraints. Moreover, for discrete probability distributions reduction to finite number of constraints is possible.

See Dentcheva, Henrion, Ruszczyński, Schultz, Kopa,...

General basic quotations

[B-S] Bonnans, Shapiro (2000) *Perturbation Analysis of Optimization Problems*, Springer

[C-D] Calafiore, Dabbene eds. (2006) *Probabilistic and Randomized Methods for Design under Uncertainty*, Springer
Chapters by Calafiore and Campi, Dentcheva, Nemirovski and Shapiro, and references therein.

[SzTAKI] Prékopa ed. (1978) Studies in Applied Stochastic Programming, MTASzTAKI Budapest

[R-W] Rockafellar, Wets (1998) *Variational Analysis*, Springer

[R-S] Ruszczyński, Shapiro eds. (2003) *Stochastic Programming*. Handbooks in OR & MS
Chapters by Prékopa, Römisch, Shapiro and references therein.

[S-D-R] Shapiro, Dentcheva, Ruszczyński (2009) *Lectures on Stochastic Programming*, MPS-SIAM; Chapter 4 and references therein.

Main quoted references

[B-D] Branda, Dupačová (2012) Approximation and contamination bounds for probabilistic programs. *Ann. Oper. Res.* 193, 3–19

[D] Dupačová (1996) Scenario based stochastic programs: Resistance with respect to sample. *Ann. Oper. Res.* 64:21–38

[IFIP]Dupačová (2011) Robustness analysis of stochastic programs with joint probabilistic constraints, submitted to Proceedings of IFIP TC7, Springer

[D-K] Dupačová, Kopa (2011) Robustness for risk constrained stochastic programs. *Ann. Oper. Res.* Online

[D-P] Dupačová, Polívka (2007) Stress testing for VaR and CVaR. *Quantitative Finance* 7:411–421

[P-A-S] Pagoncelli, Ahmed, Shapiro (2009) Sample average method for chance constrained programming: Theory and applications. *JOTA* 142, 399–416

[P-W] Pflug, Wozabal (2007) Ambiguity in portfolio selection. *Quantitative Finance* 7:435–442

[Popescu] Popescu (2007) Robust mean-covariance solutions for stochastic optimization.*Oper. Res.* 55:98–112

[Robinson] Robinson (1986) Local structure of feasible sets in nonlinear programming Part III: Stability and sensitivity. *Math. Progr. Study* 30:45–66

[WWW] Watson, Woodruff, Wets (2010) Scalable Heuristics for a Class of Chance Constrained Stochastic Programs. *INFORMS J. on Computing* 22:543–554