Robust Optimization of Uncertain Optimization Problems Affected by Ambiguous Probability Distributions

Aharon Ben-Tal

Technion—Israel Institute of Technology

July 2019, Ghent, Belgium

DATA UNCERTAINTY IN OPTIMIZATION

♣ **Consider a generic optimization problem of the form**

$$\min_x \{f(x;\zeta) : F(x;\zeta) \in \mathbf{K}\}$$

• $x \in \mathbf{R}^n$**: decision vector •** $\zeta \in \mathbf{R}^M$**: data •** $\mathbf{K} \subset \mathbf{R}^m$**: closed convex set**

♠ **More often than not the data** ζ **is** *uncertain* **– not known exactly when problem is solved.**

Sources of data uncertainty:

- **part of the data is measured/estimated** ⇒ *estimation errors*
- **part of the data (e.g., future demands/prices) does not exist when problem is solved** ⇒ *prediction errors*
- **some components of a solution cannot be implemented exactly as computed** ⇒ *implementation errors* **which in many models can be mimicked by appropriate data uncertainty**

Example

Effect of data inaccuracy

Data Uncertainty in Optimization

♣ **Consider a real-world LP program `PILOT4` from the `NETLIB` library (1,000 variables, 410 constraints). The constraint # 372 is:**

$$\begin{aligned}
[a^{\mathrm{n}}]^T x \equiv -15.79081x_{826} - 8.598819x_{827} - 1.88789x_{828} - 1.362417x_{829} - 1.526049x_{830} \\
-0.031883x_{849} - 28.725555x_{850} - 10.792065x_{851} - 0.19004x_{852} - 2.757176x_{853} \\
-12.290832x_{854} + 717.562256x_{855} - 0.057865x_{856} - 3.785417x_{857} - 78.30661x_{858} \\
-122.163055x_{859} - 6.46609x_{860} - 0.48371x_{861} - 0.615264x_{862} - 1.353783x_{863} \\
-84.644257x_{864} - 122.459045x_{865} - 43.15593x_{866} - 1.712592x_{870} - 0.401597x_{871} \\
+x_{880} - 0.946049x_{898} - 0.946049x_{916} \\
\geq b \equiv 23.387405
\end{aligned}$$

♠ **Most of the coefficients are "ugly reals" (like -15.79081 or -84.644257). It is highly unlikely that the corresponding real-life parameters are known to high accuracy, so that the ugly coefficients can be thought of as uncertain – not known exactly.**

The only exception is the coefficient 1 at x_{880} – it perhaps reflects the structure of the problem and might be exact.

$$
\begin{aligned}
[a^{\mathrm{n}}]^T x \equiv -15.79081x_{826} - 8.598819x_{827} - 1.88789x_{828} - 1.362417x_{829} - 1.526049x_{830} \\
-0.031883x_{849} - 28.725555x_{850} - 10.792065x_{851} - 0.19004x_{852} - 2.757176x_{853} \\
-12.290832x_{854} + 717.562256x_{855} - 0.057865x_{856} - 3.785417x_{857} - 78.30661x_{858} \\
-122.163055x_{859} - 6.46609x_{860} - 0.48371x_{861} - 0.615264x_{862} - 1.353783x_{863} \\
-84.644257x_{864} - 122.459045x_{865} - 43.15593x_{866} - 1.712592x_{870} - 0.401597x_{871} \\
+x_{880} - 0.946049x_{898} - 0.946049x_{916} \\
\geq b \equiv 23.387405
\end{aligned}
$$

(?) What happens with the constraint, evaluated at the nominal solution x^{n} as reported by CPLEX, when the accuracy in the uncertain data is 0.1%:

$$
|a_i^{\mathrm{true}} - a_i^{\mathrm{n}}| \leq 0.001|a_i^{\mathrm{n}}| \qquad (*)
$$

- **In the worst case, the constraint can be violated by as much as 450%:**

$$
\min_{a^{\mathrm{true}}} \left\{ [a^{\mathrm{true}}]^T x^{\mathrm{n}} \,|\, a^{\mathrm{true}} \textbf{ satisfies } (*) \right\} - b < -128.2 \approx 4.5|b|.
$$

• **Assuming "random uncertainty":**

$$a_i^{\text{true}} = a_i^{\text{n}} + \epsilon_i |a_i^{\text{n}}|, \quad \epsilon_i \sim \textbf{Uniform}[-0.001, 0.001]$$

and running 1,000 simulations, we come to the results as follows:

Prob$\{V > 0\}$	**Prob**$\{V > 150\%\}$	**Mean**(V)
0.50	**0.18**	**125%**

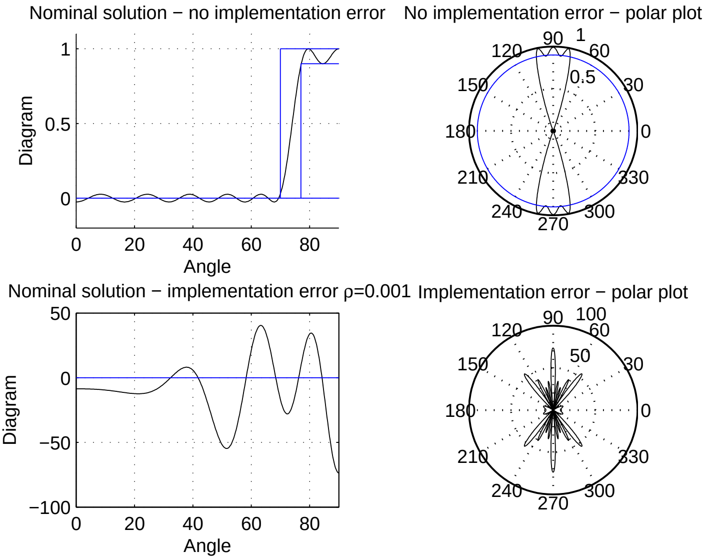
$$V = \max\left[\frac{b-(a^{\text{true}})^T x^{\text{n}}}{|b|}, 0\right]$$

⇒ The nominal solution is highly "unreliable" – small perturbations of (clearly uncertain!) data entries can make the solution heavily infeasible...

♣ Among 90 `NETLIB` LP problems,

- **In 19 problems 0.01%-perturbations of "clearly uncertain" data result in more than 5%-violations of (some of) the constraints as evaluated at the nominal optimal solution;**
- **In 13 of these 19 problems, 0.01%-perturbations of "clearly uncertain" data result in more than 50%-violations of the constraints!**

Nominal solution - dream and reality



♠ **With traditional modelling methodology,**

• **"large" data uncertainty is modelled in a stochastic fashion and then processed via Stochastic Programming techniques**

> **<u>Fact:</u>** *In many cases, it is difficult to specify reliably the distribution of uncertain data and/or to process the resulting Stochastic Programming program.*

♠ **The ultimate goal of** *Robust Optimization* **is to take into account data uncertainty already at the modelling stage in order to "immunize" solutions against uncertainty.**

• **In contrast to Stochastic Programming, Robust Optimization does not assume stochastic nature of the uncertain data (although can utilize, to some extent, this nature, if any).**

"NON-ADJUSTABLE" ROBUST OPTIMIZATION: Robust Counterpart of Uncertain Problem

$$\boxed{\min_x \{f(x,\zeta) : F(x,\zeta) \in \mathbf{K}\}} \qquad (\mathbf{U})$$

♣ **The initial ("Non-Adjustable") Robust Optimization paradigm (Soyster '73, B-T&N '97–, El Ghaoui et al. '97–, Bertsimas&Sim '03–,...) is based on the following tacitly accepted assumptions:**

A.1. All decision variables in (U) represent "here and now" decisions which should get specific numerical values as a result of solving the problem and *before* the actual data "reveals itself".

A.2. The uncertain data are "unknown but bounded": one can specify an appropriate (typically, bounded) *uncertainty set* $\mathcal{U} \subset \mathbf{R}^M$ of possible values of the data. The decision maker is fully responsible for consequences of the decisions to be made when, and only when, the actual data is within this set.

A.3. The constraints in (U) are "hard" – we cannot tolerate violations of constraints, even small ones, when the data is in $\mathcal{U}$.

$$\boxed{\begin{array}{c}\min_x \{f(x,\zeta) : F(x,\zeta) \in \mathbf{K}\} \\ \zeta \in \mathcal{U}\end{array}} \qquad (\mathbf{U})$$

♠ **Conclusions:**

• **The only meaningful candidate solutions are the *robust* ones – those which remain feasible whatever be a realization of the data from the uncertainty set:**

$$x \textbf{ robust feasible } \Leftrightarrow \; F(x,\zeta) \in \mathbf{K} \;\; \forall \zeta \in \mathcal{U}$$

• **"Robust optimal" solution to be used is a robust solution with the smallest possible *guaranteed* value of the objective, that is, the optimal solution of the optimization problem**

$$\min_{x,t} \{t : f(x,\zeta) \le t,\; F(x,\zeta) \in \mathbf{K} \;\; \forall \zeta \in \mathcal{U}\} \qquad \textbf{(RC)}$$

called the *Robust Counterpart* of (U).

Optimization Problems with Uncertain Data

♣ **A generic optimization problem is of the form**

$$\min_x \{f(x;\zeta) \mid F(x;\zeta) \le 0\} \qquad (\mathbf{P}_\zeta)$$

- x **is the design vector**
- f, F **are specified by the description of the problem**
- ζ **is a finite-dimensional vector specifying the data.**

Example 1. Linear Programming:

$$\min_x \{c^T x \mid Ax \le b\} \qquad [\zeta = (c, A, b)]$$

Example 2. Convex Quadratic Programming:

$$\min_x \{c^T x \mid x_i^T A_i^T A_i x_i - 2b_i^T x_i + c_i \le 0,\ i = 1, ..., m\} \qquad [\zeta = (c, \{A_i, b_i, c_i\}_{i=1}^m)]$$

Example 3. Conic Quadratic Programming:

$$\min_x \{c^T x \mid \| A_i x - b_i \|_2 \le c_i^T x - d_i,\ i = 1, ..., m\} \qquad [\zeta = (c, \{A_i, b_i, c_i, d_i\}_{i=1}^m)]$$

Example 4. Semidefinite Programming:

$$\min_x \left\{c^T x \mid A_0 + \sum_{j=1}^{\dim x} x_i A_i \succeq 0\right\} \qquad [\zeta = (c, A_0, ..., A_{\dim x})]$$

Robust Linear Programming

We focus on:

$$(a + Bp)^T x \leq \beta, \forall p \in U,$$

where $p \in \mathbb{R}^m$ is the uncertain vector, $B \in \mathbb{R}^{n \times m}$, and U the uncertainty region.

Uncertainty region	U	Robust Counterpart	Tractability
Box	$\|p\|_\infty \leq 1$	$a^T x + \|B^T x\|_1 \leq \beta$	LP
Ball	$\|p\|_2 \leq 1$	$a^T x + \|B^T x\|_2 \leq \beta$	CQP
Polyhedral	$Cp + d \geq 0$	$\begin{cases} a^T x + d^T y \leq \beta \\ C^T y = -B^T x \\ y \geq 0 \end{cases}$	LP
Cone (closed, convex, pointed)	$Cp + d \in K$	$\begin{cases} a^T x + d^T y \leq \beta \\ C^T y = -B^T x \\ y \in K^* \end{cases}$	Conic Opt.

Robust Solution of Optimization Affected by Uncertain Probabilities

Robust Solution of Optimization Problem Affected by Uncertain Probabilities

Consider the robust linear constraint

$$(a + Bp)^T x \leq b\,, \quad x \in I\!R^n, \quad a \in I\!R^n, \quad B \in I\!R^{n\times m} \qquad (0)$$

where $p \in I\!R^m$ is an uncertain probability vector:

$$p \in \Delta_m = \{p \in I\!R^m \mid p^T\ell = 1,\ p \geq 0\} \quad \ell = (1, 1, \ldots, 1)^T.$$

The RC of (0), w.r.t. an uncertainty set U is

$$(a + Bp)^T x \leq b\,, \quad \forall\, p \in U\,.$$

How to construct U so it has the following properties:

(a) U is based on empirical probability estimates (q) obtained from historical data.

(b) U is related to a statistical confidence region (based on asymptotic theory).

(c) U is such that the RC (2) is tractable.

A family of U's that has these properties is

$$U_p = \{p \in \Delta_m \mid I_q(p, q) \leq \rho_p\}$$

where $I_\phi(p, q) = \sum_{i=1}^m q_i \phi\left(\frac{p_i}{q_i}\right)$ is the so-called ϕ-divergence (= distance) between two probability vectors p and q. The function $\phi(t)$ is convex for $t \geq 0$ with $\phi(1) = 0$, $0\phi(a/0) = a \lim_{t\to\infty} \phi(t)/t$ for $a \succ 0$ and $0\phi(0/0) = 0$.

Given some ϕ, the **adjoint** of ϕ is defined for $t \geq 0$ as:

$$\tilde{\phi}(t) = t(\phi)(1/t)$$

which in itself is an admissible divergence function and the following relation holds:

$$I_{\tilde{\phi}}(p, q) = I_\phi(q, p)\,.$$

The RC (2) with $U = U_p$ will turn out to be given in terms of the conjugate function of ϕ:

$$\phi^*(s) = \sum_{t \geq 0} \{st - \phi(t)\},\,.$$

A list of ϕ-divergence with their conjugates and adjoints is given in the following table.

Divergence	$\phi(t)$	$\phi(t), t \geq 0$[a]	$I_\phi(p,q)$	$\phi^*(s)$	$\widetilde{\phi}(t)$	**RCP**
Kullback-Leibler	$\phi_{kl}(t)$	$t\log t - t + 1$	$\sum p_i \log\left(\frac{p_i}{q_i}\right)$	$e^s - 1$	$\phi_b(t)$	S.C.
Burg entropy	$\phi_b(t)$	$-\log t + t - 1$	$\sum q_i \log\left(\frac{q_i}{p_i}\right)$	$-\log(1-s),\ s<1$	$\phi_{kl}(t)$	S.C.
J-divergence	$\phi_j(t)$	$(t-1)\log t$	$\sum (p_i - q_i)\log\left(\frac{p_i}{q_i}\right)$	no closed form	$\phi_j(t)$	S.C.
χ^2-distance	$\phi_c(t)$	$\frac{1}{t}(t-1)^2$	$\sum \frac{(p_i-q_i)^2}{p_i}$	$2 - 2\sqrt{1-s},\ s<1$	$\phi_{mc}(t)$	CQP
Modified χ^2-distance	$\phi_{mc}(t)$	$(t-1)^2$	$\sum \frac{(p_i-q_i)^2}{q_i}$	$\begin{cases} -1, & s<-2 \\ s+s^2/4, & s\geq -2 \end{cases}$	$\phi_c(t)$	CQP
Hellinger distance	$\phi_h(t)$	$(\sqrt{t}-1)^2$	$\sum(\sqrt{p_i}-\sqrt{q_i})^2$	$\frac{s}{1-s},\ s<1$	$\phi_h(t)$	CQP
χ divergence of order $\theta>1$	$\phi^\theta_{ca}(t)$	$\lvert t-1\rvert^\theta$	$\sum q_i \lvert 1-\frac{p_i}{q_i}\rvert^\theta$	$s + (\theta-1)\left(\frac{\lvert s\rvert}{\theta}\right)^{\frac{\theta}{\theta-1}}$	$t^{1-\theta}\phi^\theta_{ca}(t)$	CQP
Variation distance	$\phi_v(t)$	$\lvert t-1\rvert$	$\sum \lvert p_i - q_i\rvert$	$\begin{cases} -1, & s\leq -1 \\ s, & -1\leq s\leq 1 \end{cases}$	$\phi_v(t)$	LP
Cressie and Read	$\phi^\theta_{cr}(t)$	$\frac{1-\theta+\theta t-t^\theta}{\theta(1-\theta)}, \quad \theta\neq 0,1$[b]	$\frac{1}{\theta(1-\theta)}(1-\sum p_i^\theta q_i^{1-\theta})$	$\frac{1}{\theta}(1-s(1-\theta))^{\frac{\theta}{\theta-1}} - \frac{1}{\theta}$ $s < \frac{1}{1-\theta}$	$\phi^{1-\theta}_{cr}(t)$	CQP

Table 2: Some ϕ-divergence examples, with their conjugates and adjoints. The last column indicates the tractability of (1); S.C. means "admits self-concordant barrier".

[a] $\phi(t) = \infty$, for $t < 0$

[b] Note that $\phi^1_{cr}(t) = \phi_b(t)$ and $\phi^0_{cr}(t) = \phi_{kl}(t)$.

The next result (based on L. Pardo's *Statistical Inference Based on Divergence Measures*, 2006) establishes the statistical interpretation of the uncertainty setup.

Let $q_N = (q_{1,N}, q_{2,N}, \ldots, q_{m,N})^T$ be an m-dimensional vector of sampled frequencies of m scenarios based on the random sample, $Z_1, \ldots, Z_N$. Assume the ϕ is a divergence function, twice differential in the neighborhood of 1, with $\phi''(1) > 0$.

Theorem 1

(i) The statistics $\frac{2N}{\phi''(1)}\, I_\phi(p, q_N)$ has asymptotically $(N \to \infty)$ a x^2_{m-1} distribution.

(ii) The uncertainty set

$$U_p = \left\{ p \in \Delta_m : I_\phi(p, q_N) \leq \rho_\phi = \frac{\phi''(1)}{2N}\, x^2_{m-1,\,1-\alpha} \right\}$$

(where $x^2_{m-1,\,1-\alpha}$ is the $(1-\alpha)$ percentile of the x^2_{m-1} distribution)
is an $(1-\alpha)$ confidence set for p.

Robust Counterpart with ϕ-Divergence Uncertainty

Consider the following robust linear constraint:

$$(a + Bp)^T x \leq \beta \,, \quad \forall\, p \in U \,, \tag{1}$$

where $x \in I\!R^n$ is the optimization vector, $a \in I\!R^n$, $B \in I\!R^{n\times m}$, and $\beta \in I\!R$ are given parameters, $p \in I\!R^m$ is the uncertain parameter, and

$$U = \{p \in I\!R^m \mid p \geq 0,\ Cp \leq d,\ I_\phi(p,q) \leq \rho\} \,, \tag{2}$$

where $q \in I\!R^m$ (with $q \geq 0$), $\rho > 0$, $d \in I\!R^k$, and $C \in I\!R^{k\times m}$ are given.

Theorem 2 A vector $x \in I\!R^n$ satisfies (1) with uncertainty region U given by (2) such that $q \in U$ if and only if there exist $\eta \in I\!R^k$ and $\lambda \in I\!R$ such that (x, λ, η) satisfies

$$\begin{cases} a^T x + d^T \eta + \rho\lambda + \lambda \sum_i q_i \phi^* \left(\frac{b_i^T x - c_i^T \eta}{\lambda} \right) \leq \beta \\ \eta \geq 0k,\ \lambda \geq 0 \,, \end{cases} \tag{3}$$

where b_i and c_i are the i-th columns of B and C, respectively, and ϕ^* is the conjugate function of ϕ (with $0\phi^*\left(\frac{s}{0}\right) := 0$ if $s \leq 0$ and $0\phi^*\left(\frac{s}{0}\right) := +\infty$ if $s > 0$).

Numerical Illustration: Multi-item Newsvendor Example

As a numerical illustration, we consider a multi-item newsvendor problem. This problem deals with optimizing the inventory of several items which can only be sold in one period. Due to the uncertain demand, this newsvendor can face both unsold items or unmet demand. The unsold items will return a loss, and unmet demand generates a cost of lost sales. For each item j, we define the purchase cost c_j, the selling price v_j, the salvage value of unsold items s_j, and the cost of lost sales l_j. Furthermore, we denote γ for the budget that is available for the purchase of the items.

Item (j)	1	2	3	4	5	6	7	8	9	10	11	12
c	4	5	6	4	5	6	4	5	6	4	5	6
v	6	8	9	5	9	8	6	8	9	6.5	7	8
s	2	2.5	1.5	1.5	2.5	2	2.5	1.5	2	2	1.5	1
l	4	3	5	4	3.5	4.5	3.5	3	5	3.5	3	5
$q_{1,N}^{(j)}$	0.375	0.250	0.375	0.127	0.958	0.158	0.485	0.142	0.679	0.392	0.171	0.046
$q_{2,N}^{(j)}$	0.375	0.250	0.250	0.786	0.007	0.813	0.472	0.658	0.079	0.351	0.484	0.231
$q_{3,N}^{(j)}$	0.250	0.500	0.375	0.087	0.035	0.029	0.043	0.200	0.242	0.257	0.345	0.723

Table 1. Parameter values for the multi-item newsvendor example

We assume that demand for all items is a random variable that can take on m values, denoted as d_i, $i = 1, \ldots, m$. We denote $p_i^{(j)}$ for the unknown probability that the demand for item j equals d_i, and we let the uncertainty region for $p^{(j)} = (p_1^{(j)}, \ldots, p_m^{(j)})^T$ be given by

$$U^{(j)} := \left\{ p^{(j)} \in I\!R^m \mid p^{(j)} \geq 0,\ (p^{(j)})^T e = 1,\ I_\phi\left(p^{(j)}, q_N^{(j)}\right) \leq \rho \right\}\,, \tag{6}$$

where $q_N^{(j)}$ represents the sample-based estimated probability distribution for item j.

Denote by Q_j the order quantity for item j. We consider a multi-item newsvendor problem in which he maximizes the expected profit on the least profitable item:

$$\max_Q \min_j \sum_i p_i^{(j)} \bar{r}_j(Q_j, i)$$

where $\bar{r}_j(Q_j, i) = v_j \min(d_i, Q_j) + s_j(Q_j - d_i)^+ - l_j(d_i - Q_j)^+ - cQ_j$. The robust version of this problem can be stated as:

$$\begin{aligned} &\max \|z\|_\infty \\ &\quad \text{s.t.} - c_j Q_j + \sum_i p_i^{(j)} f_{i,j}(Q_j) \geq z_j, \quad \forall\, j,\ \forall\, p^{(j)} \in U^{(j)} \\ &\qquad \sum_j c_j Q_j \leq \gamma\,, \end{aligned}$$

with

$$f_{i,j}(Q_j) = v_j \min\{d_i, Q_j\} + s_j \max\{0, q_j - d_i\} - l_j \max\{0, d_i - Q_j\}\,.$$

From our previous result, the RC is given by

$$
\begin{aligned}
&\max \|z\|_\infty \\
&\text{s.t.} - c_j Q_j - \eta_j - \lambda_j \rho - \lambda_j \sum_i q_{i,N}^{(j)} \phi^* \left(\frac{-f_{i,j}(Q_j) - \eta_j}{\lambda_j} \right) \geq z_j,\ \forall\, j \\
&\quad \sum_j c_j Q_j \leq \gamma \\
&\quad \lambda \geq 0\,.
\end{aligned}
$$

Our numerical results apply to the case with $n = 12$ different items, and $m = 3$ scenarios for the demand for each item: low demand (4), medium demand (8), and high demand (10), denoted as $d_1 = 4$, $d_2 = 8$, and $d_3 = 10$, respectively. The parameter values of the revenue functions, as well as the values of $q_{i,N}^{(j)}$, are given in Table 1. Furthermore, the budget is set at $\gamma = 1000$.

We solve the RCP for the Cressie and Read ϕ-divergence function with $\theta = 0.5$. For both ϕ-divergence functions, we consider the case where $\rho = \rho^a$ is the test statistic and the case where $\rho = \rho^c$ is the test statistic $\frac{2N}{\phi''(1)} I_\phi(p, q_N)$. In each case, the confidence level is set at $\alpha = 0.05$, and we determine the robust optimal solutions for different sample sizes $N = 10, 20, \ldots, 1000$.

Using the solutions of the RCP problems and the solution of the non-robust problem (i.e., assuming that Q_N is the true probability vector), we make several comparisons. First, we compare the performance of the robust versus the non-robust solutions for the different values of the sample size N.

To make comparisons, we proceed as follows. First, we sample 10,000 hypothetically true p-vectors. Next, for each sampled probability vector p, we calculate the value of the objective function for the non-robust as well as for the robust optimal solutions. We then compare the performance of the different solutions by determining the mean and the range (i.e., the minimum and the maximum value) of the objective values corresponding to the sampled p-vectors. The p-vectors are sampled such that approximately 95% of the sample satisfies $I_{\phi_{me}}(p, q_N) \leq \bar{\rho} := \rho_{\phi_{me}}$, where ϕ_{me} denotes the modified χ^2-divergence. The results show that the mean return of the objective values for the robust solution is higher than the mean for the non-robust solution. Moreover, the dispersion of objective values for the robust solution is significantly smaller than the range of objective values for the non-robust solution. In particular, the robust solution avoids substantial losses.

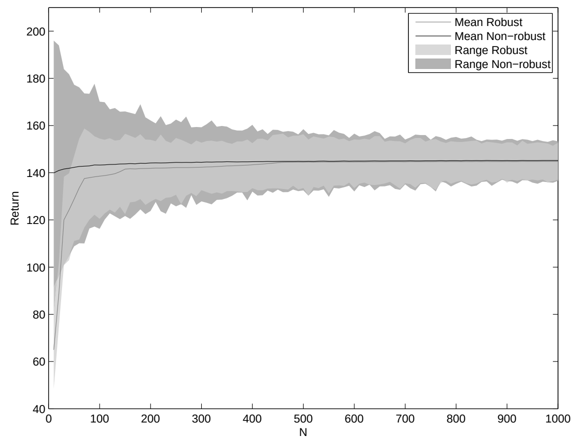


Figure 1: Cressie-Read for $\theta = 0.5$, and ρ^c_ϕ, and the 1-norm.

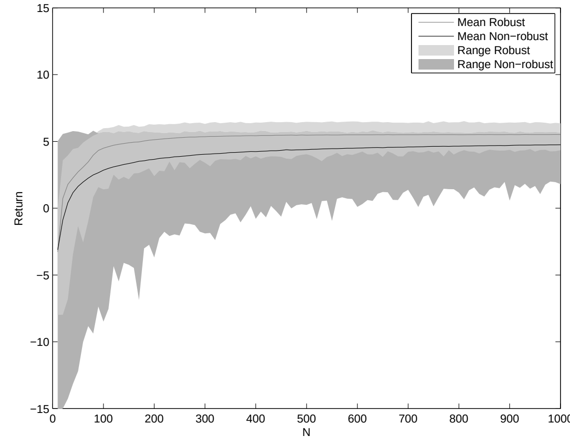


Figure 2: Cressie-Read for $\theta = 0.5$, and ρ^c_ϕ, and the ∞-norm.

dispersion of objective values for the robust solution is significantly smaller than the range of objective values for the non-robust solution for the ∞-norm. In particular, the robust solution avoids substantial losses.

Concerning the effect of N, the effect of using ρ^c_ϕ versus ρ^a_ϕ, and the differences between the two ϕ-divergence measures, we observe the following:

Effect of N. Because 95 percent of the sampled p-vectors needs to satisfy $I_{\phi_{mc}}(p, q_N) \leq \overline{\rho}$, and because $\overline{\rho}$ is decreasing in N, the range of the expected returns becomes smaller as N increases. However, because 5 percent of the sampled p-vectors does not need to satisfy $I_{\phi_{mc}}(p, q_N) \leq \overline{\rho}$, the range does not converge to a single value.

Effect of ρ^c_ϕ versus ρ^a_ϕ. With regard to the differences between the robust solutions in case ρ^a_ϕ is used (i.e., the uncertainty region is based on the approximate test statistic) and when ρ^c_ϕ is used (i.e., the uncertainty region is based on the corrected test statistic), we observe that there are significant differences only for relatively small values for N. This occurs of course since the effect of the correction becomes smaller as N increases.

Comparison of different ϕ-divergence measures. The different ϕ-divergence measures lead to different optimal quantities, but the structure of the solutions is similar. The mean expected utility as well as the range of the expected utilities over the sampled p-vectors is similar for the two ϕ-divergence measures.

7 Concluding remarks

In this paper we have shown that the robust counterpart of linear and nonlinear optimization problems with uncertainty regions defined by ϕ-divergence distance measures can be reformulated as tractable optimization problems. Thus, these uncertainty regions are useful alternatives to uncertainty regions considered in the existing literature, particularly so when the uncertainty is associated with probabilities. In this latter case, we have shown that uncertainty regions based on ϕ-divergence test statistics have a natural interpretation in terms of statistical confidence sets. This allows for an approach that is fully data-driven.

Our approach also has other applications. For example, ϕ-divergence distances can be directly

Distributionally Robust Optimization

Two types of constraints:

- worst-case expected feasibility constraints:

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}} f(\mathbf{x},\mathbf{z}) \leq 0, \tag{WC-EF}$$

- worst-case chance constraints:

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left(f(\mathbf{x},\mathbf{z}) > 0\right) \leq \epsilon. \tag{WC-CC}$$

(WC-EF) is used to construct *safe approximations* of (WC-CC).

Scarf's (1958) Newsvendor Problem

The newsvendor buys x newspapers with cost \$c each ($0 < c < 1$).

He sells them with price \$1. The demand for newspapers is d,

a random variable with partially known probability ($\wp \in P$).

The DRO is then

$$\boxed{\max_x \sup_{\wp \in P} E_\wp \min(x, d) - cx}$$

Scarf assumed that

$$P = \{\wp : E_\wp d = \mu\,,\; E_\wp (d - \mu)^2 = \sigma^2\}$$

and obtained an exact solution to the inner sup problem.

Ambiguity set $\mathcal{P}$

Ambiguity set $\mathcal{P}$ should be such that it is possible to obtain good, computationally tractable upper bounds on

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}} f(\mathbf{x}, \mathbf{z})$$

Most frequently, $\mathcal{P}$ consists of $\mathbb{P}$ with known:

- mean
- (co)variance matrix
- possibly, higher order moment information

Major works: Scarf (1958), Dupačová (1977), Birge and Wets (1987), Birge and Dulá (1991), Gallego (1992), Gallego, Ryan & Simchi-Levi (2001), Delage and Ye (2010), Wiesemann et al. (2014) and many others...

Forgotten result of Ben-Tal and Hochman (1972)

An exact upper bound when the dispersion measure is the mean absolute deviation (MAD).

Theorem

Assume that a one-dimensional random variable z has support included in $[a, b]$ *and its mean and mean absolute deviation are* μ *and* d*:*

$$\mathcal{P} = \{\mathbb{P} : \quad supp(z) \subseteq [a, b], \quad \mathbb{E}_{\mathbb{P}} z = \mu, \quad \mathbb{E}_{\mathbb{P}} |z - \mu| = d\}.$$

Then, for any convex function $g : \mathbb{R} \to \mathbb{R}$ *it holds that*

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} g(z) = p_1 g(a) + p_2 g(\mu) + p_3 g(b),$$

where $p_1 = \frac{d}{2(\mu - a)}, \quad p_3 = \frac{d}{2(b - \mu)}, \quad p_2 = 1 - p_1 - p_3.$

Generalization to multiple dimensions

The result of Ben-Tal and Hochman (1972) generalizes to multidimensional $\mathbf{z}$ with independent components.

$$\mathcal{P} = \{\mathbb{P} : \text{supp}(z_i) \subseteq [a_i, b_i], \quad \mathbb{E}_{\mathbb{P}} z_i = \mu_i, \quad \mathbb{E}_{\mathbb{P}} |z_i - \mu_i| = d_i, \quad z_i \perp z_j\}.$$

Independence implies that the worst-case distribution is a product of the per-component worst-case distributions.

For each convex $g(\cdot)$ it holds that

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} g(\mathbf{z}) = \sum_{\alpha \in \{1,2,3\}^{n_{\mathbf{z}}}} \left(\prod_{i=1}^{n_{\mathbf{z}}} p^i_{\alpha_i} \right) g(\tau^1_{\alpha_1}, \ldots, \tau^{n_{\mathbf{z}}}_{\alpha_{n_{\mathbf{z}}}})$$

where $p^i_{\alpha_i}$ and $\tau^i_{\alpha_i}$ depend only on a_i, b_i, μ_i, and d_i (not on $g(\cdot)$).

Lower bound result

Ben-Tal and Hochman (1972) provide also an exact formula for the lower bound on the expectation if additionally, it is known that $\mathbb{P}(z \geq \mu) = \beta$:

$$\mathcal{P}_\beta = \{\mathbb{P} : \quad \mathbb{P} \in \mathcal{P}, \quad \mathbb{P}(z \geq \mu) = \beta\}.$$

Then, for any convex function $g : \mathbb{R} \to \mathbb{R}$ it holds that

$$\inf_{\mathbb{P} \in \mathcal{P}_\beta} \mathbb{E}_\mathbb{P} g(z) = \beta g\left(\mu + \frac{d}{2\beta}\right) + (1-\beta) g\left(\mu - \frac{d}{2(1-\beta)}\right).$$

Chance Constraints

$$p(w) \equiv \text{Prob}\left\{ w_0 + \sum_{\ell=1}^{d} z_\ell w_\ell \geq 0 \right\} \geq 1 - \epsilon \qquad \text{(C)}$$

- In general, (C) can be difficult to process:
  - The feasible set X of (C) can be nonconvex, which makes it problematic to optimize under the constraint.
  - Even when convex, X can be "computationally intractable":

    Let $z \sim$ Uniform$([0.1]^d)$. In this case, X is convex (Lagoa et al., 2005); however, *unless* $P = NP$, *there is no algorithm capable to compute* $p(w)$ *within accuracy* δ *in time polynomial in the size of the (rational) data* w *and in* $\ln(1/\delta)$ (L. Khachiyan, 1989).

- When (C) is difficult to process "as it is", one can look for a *safe tractable approximation of (C) — a computationally tractable convex set* U_ϵ *such that* $U_\epsilon \subset X \equiv \{w : p(w) \geq \epsilon\}$.

Probabilistic Guarantees via RO

$$f_0(x) + \sum_{l=1}^{d} z_l f_l(x) \leq 0\,. \qquad (1)$$

Assumption

$z_1, z_2, \ldots, z_d$ independent rv's

$z_l \sim \mathbf{P}_l \in \mathcal{P}_l$ (compact all prob. dist. in $\mathcal{P}_l$ has common support $= [-1, 1]$).

Definition A vector x satisfying, for a given $0 < z < 1$:

$$\Pr\{f_0(x) + \Sigma z_l f_l(x) \leq 0\} \geq 1 - \epsilon \quad \text{(chance constraint)} \qquad (2)$$

provides a *safe approximation* of (1).

Challenge Find uncertainty set for z, U_ϵ s.t. the Robust Counterpart of (1):

$$f_0(x) + \Sigma z_l f_l(x) \leq 0,\ \forall\, z \in U_\epsilon \qquad (3)$$

is a safe approximation of (1), i.e., every x satisfying (3) satisfies the CC (2).

THEOREM

Consider the uncertainty set:

$$U_\epsilon = B \cap (M + E)$$

where

$$B = \{u \in I\!R^d \mid \|u\|_\infty \leq 1\}$$

$$M = \{u \mid \mu_l^- \leq u_l \leq \mu_l^+,\ l = 1, \ldots, d\} \qquad (1)$$

$$E = \{u \mid \Sigma u_l^2/\sigma_l^2 \leq 2\log(1/\epsilon)\}$$

and where μ_l^-, μ_l^+ and σ_l are such that

$$A_l(y) \leq \max(\mu_l^- y,\ \mu_l^+ y)\ +\ \frac{\sigma_l^2}{2}\, y_l^2, \quad \forall\, l = 1, \ldots, d$$

and

$$A_l(y) = \max_{P_l \in \mathcal{P}_l} \log\left(\int \exp(ys) dP_l(s)\right).$$

Then, a vector x satisfying

$$f_0(x) + \Sigma z_j f_j(x) \leq 0, \quad \forall\, z \in U_\epsilon$$

satisfies the cc inequality:

$$Pr\{f_0(x) + \Sigma z_j f_j(x) \leq 0\} \geq 1 - \epsilon$$

Safe approximations

As such, (WC-CC) is intractable and we need a *safe approximation* - a computationally tractable set $\mathcal{S}$ of deterministic constraints such that

$$\mathbf{x} \text{ feasible for } \mathcal{S} \Rightarrow \mathbf{x} \text{ feasible for (WC-CC)}$$

How to construct safe approximations?
The crucial step is a construction of an upper bound on the moment generating function (MGF) of $\mathbf{z}$ (Ben-Tal et al. (2009)):

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}} \exp(\mathbf{w}^T \mathbf{z}).$$

.

Recall: For each convex $g(\cdot)$ it holds that

$$\sup_{\mathbb{P}\in\mathcal{P}} E_{\mathbb{P}} g(z) = \sum_{\alpha\in\{1.2.3\}^n} \left(\prod_{i=1}^{n} p^i_{\alpha_i}\right) g\left(\tau^1_{\alpha^1}, \ldots, \tau^{n_z}_{\alpha_{n_z}}\right)$$

where $p^i_{\alpha_i}$ and $\tau^i_{\alpha_i}$ depend only on a_i, b_i, μ_i, and d_i (not on $g(\cdot)$).

This formula has 3^n terms!

However:

$$\begin{aligned}
&\sup_{\mathbb{P}\in P} \log\left(E_{\mathbb{P}} \exp(w^T z)\right) = \sup_{\mathbb{P}\in\mathcal{P}} \log\left(E_{\mathbb{P}}\left(e^{w_1 z_1+\cdots+w_n z_n}\right)\right) \\
&= \sup_{\mathbb{P}\in\mathcal{P}} \log\left(E_{\mathbb{P}} \prod_{i=1}^{n} e^{w_i z_i}\right) = \text{ due to } z_i\text{'s being independent} \\
&= \sup_{\mathbb{P}\in\mathcal{P}} \log\left(\prod_{i=1}^{n} E\, e^{w_i z_i}\right) = \sup_{\mathbb{P}\in\mathcal{P}} \sum_{i=1}^{n} \left(\log E\, e^{w_i z_i}\right).
\end{aligned}$$

So here we need to apply the (B-H) upper (lower) bound separately to each on the n one-variable convex functions $E\, e^{w_i z_i}$!

Setting

We assume w.l.o.g. that $\text{supp}(z_i) \in [-1, 1]$, $\mathbb{E}_{\mathbb{P}} z_i = 0$ and $\mathbb{E}_{\mathbb{P}} |z_i - 0| = d$.

Consider the (WC-CC):

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left(\mathbf{a}^T(\mathbf{z})\mathbf{x} > b(\mathbf{z}) \right) \leq \epsilon,$$

where

$$[\mathbf{a}(\mathbf{z}); b(\mathbf{z})] = [\mathbf{a}^0; b^0] + \sum_{i=1}^{n_\mathbf{z}} z_i [\mathbf{a}^i; b^i].$$

MGF with our distributional assumptions

We know exactly the worst-case value of the MGF (not just an upper bound):

$$\begin{aligned}
\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}} \exp(\mathbf{w}^T\mathbf{z}) &= \prod_{i=1}^{n_z} \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}} \exp(w_i z_i) \\
&= \prod_{i=1}^{n_z} \left(\frac{d}{2}\exp(-w_i) + 1 - d + \frac{d}{2}\exp(w_i) \right) \\
&= \prod_{i=1}^{n_z} (d\cosh(w_i) + 1 - d)
\end{aligned}$$

Using this fact, we are able to construct three safe approximations of increasing tightness and increasing complexity.

An example of a safe approximation

Theorem

Let

$$[\mathbf{a}(\mathbf{z}); b(\mathbf{z})] = [\mathbf{a}^0; b^0] + \sum_{i=1}^{n_{\mathbf{z}}} z_i [\mathbf{a}^i; b^i].$$

If there exists $\alpha > 0$ *such that* $(\mathbf{x}, \alpha)$ *satisfies the constraint*

$$(\mathbf{a}^0)^T \mathbf{x} - b_0 + \alpha \log \left(\sum_{i=1}^{n_{\mathbf{z}}} \left(d_i \cosh \left(\frac{(\mathbf{a}^i)^T \mathbf{x} - b^i}{\alpha} \right) + 1 - d_i \right) \right) + \alpha \log(1/\epsilon) \leq 0,$$

then $\mathbf{x}$ *satisfies the (WC-CC):* $\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\left(\mathbf{a}^T(\mathbf{z})\mathbf{x} > b(\mathbf{z}) \right) \leq \epsilon.$

The approximating constraint is convex in $(\mathbf{x}, \alpha)$!

Antenna array (Ben-Tal and Nemirovski (2002))

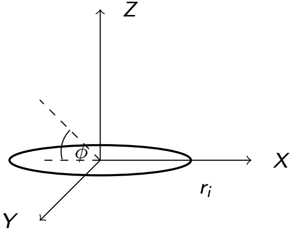
1. We consider an optimization problem with 40 circular antennas.
2. Each antenna has its diagram $D_i(\phi)$ - a plot of intensity of signal sent to different directions.
3. The diagram of the set of 40 antennas is the sum of their diagrams .

$$D(\phi) = \sum_{i=1}^{n} x_i D_i(\phi)$$

4. To the i-th antenna we can send a different amount of power x_i.
5. **Objective:** Set the x_i's in such a way that the diagram has the desired shape.

Application - antenna array optimization

Consider a circular antenna:



Energy sent in angle ϕ is characterized by *diagram*

Diagram of a single antenna:

$$D_i(\phi) = \frac{1}{2} \int_0^{2\pi} \cos\left(\frac{2\pi i}{40} \cos(\phi) \cos(\theta)\right) d\theta$$

Diagram of n antennas

$$D(\phi) = \sum_{i=1}^{n} x_i D_i(\phi)$$

x_i - power assigned to antenna i

Objective: construct $D(\phi)$ as close as possible to the desired $D^*(\phi)$ using the antennas available.

Antenna array (Ben-Tal and Nemirovski (2002))

Problem conditions:

- for $77^\circ < \phi \leq 90^\circ$ the diagram is nearly uniform:

$$0.9 \leq \sum_{i=1}^{n} x_i D_i(\phi) \leq 1, \quad 77^\circ < \phi \leq 90^\circ$$

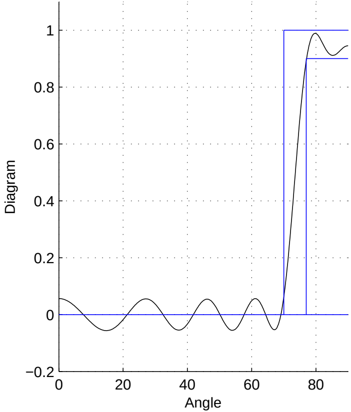
- for $70^\circ < \phi \leq 77^\circ$ the diagram is bounded:

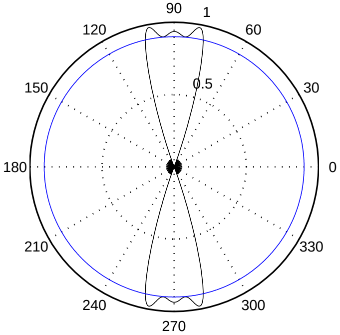
$$-1 \leq \sum_{i=1}^{n} x_i D_i(\phi) \leq 1, \quad 70^\circ < \phi \leq 77^\circ$$

- we minimize the maximum absolute diagram value over $0^\circ < \phi \leq 70^\circ$:

$$\min \max_{0^\circ < \phi \leq 70^\circ} \left| \sum_{i=1}^{n} x_i D_i(\phi) \right|$$

Desired diagram graphically





Implementation error

Typically, decisions x_i suffer from implementation error z_i:

$$x_i \mapsto \widetilde{x}_i = (1 + \rho z_i) x_i$$

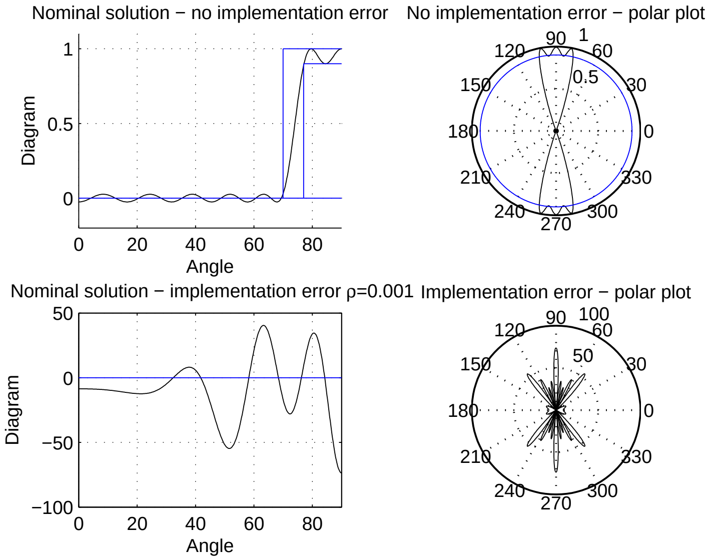
We want each constraint to hold with probability at least $1 - \epsilon$ for all $\mathbb{P} \in \mathcal{P}$, for example:

$$\mathbb{P}\left(\sum_{i=1}^{n} x_i (1 + \rho z_i) D_i(\phi) \leq 1 \right) \geq 1 - \epsilon, \quad 77^\circ < \phi \leq 90^\circ, \quad \forall \mathbb{P} \in \mathcal{P}$$

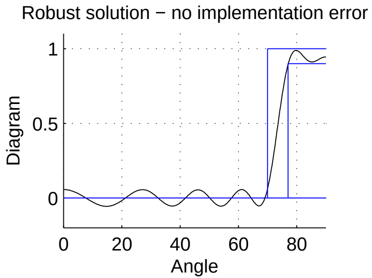
Two solutions:

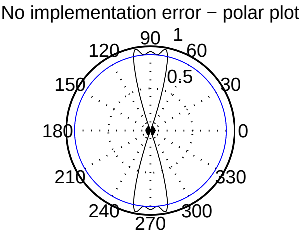
- nominal: no implementation error
- robust: $\rho = 0.001$ and $\epsilon = 0.001$.

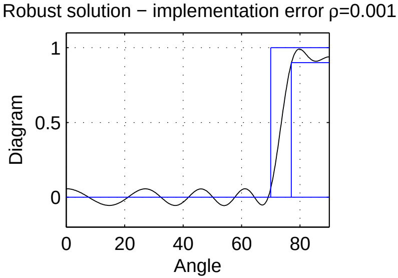
Nominal solution - dream and reality

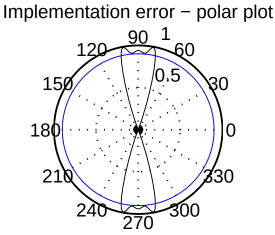


Robust solution - dream and reality









$$(1) \qquad a(\zeta)^T x \le b(\zeta)$$

$$a(\zeta) = a^0 + \sum_{\ell=1}^{L} \zeta_\ell a^\ell, \quad b(\zeta) = b^0 + \sum_{\ell=1}^{L} \zeta_\ell b^\ell$$

$$\zeta_1, \ldots, \zeta_L \ \text{ i.i.d.}, \ \ E(\zeta_\ell) = 0\,, \quad |\zeta_\ell| \le 1$$

$$(\mathrm{CC})_\varepsilon \ \boxed{\mathrm{Prob}_\zeta\left(a(\zeta)^T x \le b(\zeta)\right) \ge 1 - \varepsilon}$$

Let $\mathcal{U}_\Omega = \left\{\zeta \in \mathbb{R}^L \mid \|\zeta\|_2 \le \Omega\right\}$.

Consider the RC of (1) w.r.t. $\mathcal{U}_\Omega$:

$$a(\zeta)^T x \le b(\zeta) \quad \forall\, \zeta \in \mathcal{U}_\Omega$$

which we already know is equivalent to

$$(\mathrm{RC})_\Omega \ \boxed{(a^0)^T x + \Omega\sqrt{\textstyle\sum_{\ell=1}^{L}\left((a^\ell)^T x - b_\ell\right)^2} \le b^0}$$

Theorem 1 *If x solves $(\mathrm{RC})_\Omega$ with $\Omega \ge \sqrt{2\log(1/\varepsilon)}$, then x solves $(\mathrm{CC})_\varepsilon$*

$$\mathrm{OR:} \quad \left\{ \begin{array}{l} x \ \textit{solves}\ (\mathrm{RC})_\Omega \quad \textit{then } x \textit{ solves} \\ (\mathrm{CC})_\varepsilon \quad \textit{with} \quad \varepsilon < e^{-\Omega^2/2} \end{array} \right.$$

e.g., $\Omega = 7.44$, $1 - \varepsilon = 1 - 10^{-12}$.

Illustration: Single-Period Portfolio Selection

There are 200 assets. Asset #200 ("money in the bank") has yearly return $r_{200} = 1.05$ and zero variability. The yearly returns r_ℓ, $\ell = 1, \ldots, 199$ of the remaining assets are independent random variables taking values in the segments $[\mu_\ell - \sigma_\ell, \mu_\ell + \sigma_\ell]$ with expected values μ_ℓ; here

$$\mu_\ell = 1.05 + 0.3\frac{(200-\ell)}{199},\ \sigma_\ell = 0.05 + 0.6\frac{(200-\ell)}{199},\ \ell = 1, \ldots, 199\,.$$

The goal is to distribute \$1 between the assets in order to maximize the return of the resulting portfolio, the required risk level being $\varepsilon = 0.5\%$.

We want to solve the uncertain LO problem

$$\max_{y,t}\left\{t : \sum_{t=1}^{199} r_\ell y_\ell + r_{200} y_{200} - t \geq 0, \sum_{\ell=0}^{200} y_\ell = 1,\ y_\ell \geq 0\,\forall\,\ell\right\},$$

where y_ℓ is the capital to be invested into asset $\#\ell$.

The uncertain data are the returns r_ℓ, $\ell = 1, \dots, 199$; their natural parameterization is

$$r_\ell = \mu_\ell + \sigma_\ell \zeta_\ell \,,$$

where ζ_ℓ, $\ell = 1, \dots, 199$, are independent random perturbations with zero mean varying in the segments $[-1, 1]$. Setting $x = [y; -t] \in \mathbb{R}^{201}$, the problem becomes

$$\left\{ \begin{array}{lll} \text{minimize} & x_{201} & \\ \text{subject to} & & \\ (a) & \left[a^0 + \sum_{\ell=1}^{199} \zeta_\ell a^\ell\right]^T x - \left[b^0 + \sum_{\ell=1}^{199} \zeta_\ell b^\ell\right] \leq 0 & \\ (b) & \sum_{j=1}^{200} x_\ell = 1 & \\ (c) & x_\ell \geq 0, \;\; \ell = 1, \dots, 200 & \end{array} \right. \qquad (4)$$

where

$$\begin{array}{rcl} a^0 & = & [-\mu_1; -\mu_2; \dots; -\mu_{199}; -r_{200}; -1]; a^\ell = \sigma_\ell \cdot [0_{\ell-1,1}; 1; 0_{201-\ell,1}], \;\; \ell = 1, \dots, 199; \\ b^\ell & = & 0, \;\; \ell = 0, 1, \dots, 199 \,. \end{array}$$

The only uncertain constraint in the problem is the linear inequality (a). We consider 3 perturbation sets along with the associated robust counterparts of problem (4).

1. *Box RC* which ignores the information on the stochastic nature of the perturbations affecting the uncertain inequality and uses the only fact that these perturbations vary in $[-1, 1]$. The underlying perturbation set $\mathcal{Z}$ for (a) is $\{\zeta : \|\zeta\|_\infty \leq 1\}$;
2. *Ball-Box* with the safety parameter $\Omega = \sqrt{2\ln(1/\varepsilon)} = 3.255$, which ensures that the optimal solution of the associated RC (a CQ prob.) satisfies (a) with probability at least $1 - \varepsilon = 0.995$. The underlying perturbation set $\mathcal{Z}$ for (a) is $\{\zeta : \|\zeta\|_\infty \leq 1\},\ \|\zeta\|_2 \leq 3.255\}$;

Results

Box RC. The associated RC is the LP

$$\max_{y,t}\left\{t:\begin{array}{l}\sum\limits_{\ell=1}^{199}(\mu_\ell-\sigma_\ell)y_\ell+1.05y_{200}\geq t\\ \sum\limits_{\ell=1}^{200}y_\ell=1,\ y\geq 0\end{array}\right\};$$

as it should be expected, this is nothing but the instance of our uncertain problem corresponding to the worst possible values $r_\ell=\mu_\ell-\sigma_\ell,\ \ \ell=1,\ldots,199$, of the uncertain returns. Since these values are less than the guaranteed return for money, the robust optimal solution prescribes to keep our initial capital in the bank with guaranteed yearly return 1.05.

Ball-Box RC. The associated RC is the conic quadratic problem

$$\max_{y,z,w,t} \left\{ t : \begin{array}{l} \sum_{\ell=1}^{199} (\mu_\ell y_\ell + 1.05 y_{200} - \sum_{\ell=1}^{199} |z_\ell| - 3.255 \sqrt{\sum_{\ell=1}^{199} w_\ell^2} \geq t \\ z_\ell + w_\ell = y_\ell, \quad \ell = 1, \ldots, 199, \ \sum_{\ell=1}^{200} y_\ell = 1, \ y \geq 0 \end{array} \right\} .$$

The robust optimal value is 1.1200, meaning 12.0% profit with risk as low as $\varepsilon = 0.5\%$.