# Optimisation et processus dynamiques en apprentissage et dans les problémes inverses

## Fondazione Mediaterraneo, Sestri Levante (GE), Italia

Giuseppe Carlo Calafiore

Dipartimento di Automatica e Informatica
Politecnico di Torino – ITALY

10 Septembre, 2014

# Mixed-Integer Random Convex Programs

## Polya's Urns and the Distribution of the Violation Probability

*A dichromatic Polya's urn model consists of an initial urn containing a total of $c$ balls, $b$ of which are, say, black and the remaining ones are white. At the initial step, we pick uniformly at random one ball from the urn, we look at its color, and then we put back the ball in the urn, along with a new ball having the same color of the picked ball, thus forming a new urn with $c+1$ elements. Then, we repeat the same operation on the new urn, and so on, iterating the procedure indefinitely.*
*The Polya's urn process is formed by looking at the sequence of indicator variables describing the color (say, 1 for black and 0 for white) of the new balls inserted in the urn at steps $i = 1, 2, \ldots$*

---

# Random $M$-convex programs

- Let $\delta \in \Delta$ denote a vector of random parameters, with $\Delta \subseteq \mathbb{R}^\ell$, and let $\mathbb{P}$ be a probability measure on $\Delta$.
- Let $M = \mathbb{R}^n \times \mathbb{Z}^d$, the mixed integer domain; let $\kappa = n + d$.
- Let $x \in M$ be a design variable, and consider a family of functions $f(x, \delta) : (\mathbb{R}^\kappa \times \Delta) \to \mathbb{R}$ defining the design constraints and parameterized by $\delta$.
- We assume that $f(x, \delta)$ is convex and lsc in the variable $x$, for any fixed $\delta$. Specifically, for a given design vector $x$ and realization $\delta$ of the random parameter $\delta$, the design constraint is satisfied if $f(x, \delta) \le 0$.
- Define

$$\Omega_N \doteq (\delta^{(1)}, \ldots, \delta^{(N)}) \in \Delta^N, \tag{1}$$

  where $N \ge 1$ is an integer, and $\delta^{(i)} \in \Delta$, $i = 1, \ldots, N$, are independent random variables identically distributed (iid) according to $\mathbb{P}$, where $\Delta^N = \Delta \times \Delta \cdots \Delta$ ($N$ times). Let $\mathbb{P}^N$ denote the product probability measure on $\Delta^N$.

# Random $M$-convex programs

- To each $\delta^{(j)}$ we associate a constraint function

$$f_j(x) \doteq f(x, \delta^{(j)}), \quad j = 1, \ldots, N.$$

  Thus, to each randomly extracted $\Omega_N$ there correspond $N$ random constraints $f_j(x)$, $j = 1, \ldots, N$.

- We define a random $M$-convex optimization problem (RMCP) as:

$$P[\Omega_N]: \qquad J(\Omega_N) = \begin{array}{ll} \min_{x \in \mathcal{X}} & f_0(x) \\ \text{s.t.} & f_j(x) \leq 0, \; j = 1, \ldots, N. \end{array} \tag{2}$$

- $J(\Omega_N)$ is a random variable.

# The random violation probability

- The *violation probability* of $P[\Omega_N]$ is defined as

$$V_N = V_N(\Omega_N) \doteq \mathbb{P}\{\delta \in \Delta : J(\Omega_N, \delta) > J(\Omega_N)\}. \quad (3)$$

  The complement of $V_N$, $1 - V_N$, usually called the *reliability*, gives a measure of "probabilistic robustness" for the optimal value $J(\Omega_N)$ with respect to the uncertainty parameter $\delta$, since it quantifies the probability with which this optimal value remains unchanged when a new, unseen, uncertainty instance is added to the problem constraints.

- $V_N$ is itself a random variable with support in $[0, 1]$, since it depends on the random extraction of the $N$-multisample $\Omega_N$.

- Our main goal is to characterize the cumulative distribution function $F_N$ of $V_N$, where

$$F_N(v) \doteq \mathbb{P}^N\{\Omega_N : V_N(\Omega_N) \leq v\}, \quad v \in [0, 1]. \quad (4)$$

# Support constraints and Helly's dimension

- We say that a "constraint" $\delta^{(i)} \in \Omega_N$ is a *support constraint* for problem $P[\Omega_N]$ if
$$J(\Omega_N \setminus \delta^{(i)}) < J(\Omega_N)$$
- The *Helly's dimension* of the domain $\mathbb{R}^n \times \mathbb{Z}^d$ is
$$h = (n+1)2^d$$

**Theorem**

*For any $N$, the number of support constraints of problem $P[\Omega_N]$ is $\leq \nu$ w.p. 1, where $\nu = h - 1$ if $P[\Omega_N]$ is feasible w.p. one, and $\nu = h$ in general.*

# The random number of support constraints

- Consider a multi-sample $\Omega_{N+1}$ of the $\delta^{(i)}$s of cardinality $N+1$:

$$\Omega_{N+1} \doteq \{\delta^{(1)}, \ldots, \delta^{(N)}, \delta^{(N+1)}\} = \{\Omega_N, \delta^{(N+1)}\},$$

  and define

$$\Upsilon_{N+1} \doteq \text{the number of support constraints of problem } P[\Omega_{N+1}]. \qquad (5)$$

- $\Upsilon_{N+1}$ is a random integer, since its value depends on the random extractions in $\Omega_{N+1}$.
- However, we have from Theorem 1 that $\Upsilon_{N+1} \leq h$ w.p. one, since no problem $P[\cdot]$ can have more than $h$ support constraints.
- Therefore, all we know about $\Upsilon_{N+1}$ is that it is a random integer with support in the set $\{0, 1, \ldots, h\}$.

# Violation indicator variables and de Finetti's representation

- Consider an infinite sequence of iid extractions of the $\delta^{(j)}$s:

$$\delta^{(1)}, \ldots, \delta^{(N)}, \delta^{(N+1)}, \delta^{(N+2)}, \ldots \tag{6}$$

- For some given integer $N \geq 1$, let $\Omega_N = \{\delta^{(1)}, \ldots, \delta^{(N)}\}$, and consider the infinite sequence of $\{0,1\}$-random variables

$$Z_i \doteq \begin{cases} 1 & \text{if } J(\Omega_N, \delta^{(N+i)}) > J(\Omega_N) \\ 0 & \text{otherwise} \end{cases}, \quad i = 1, 2, \ldots \tag{7}$$

  In words, $Z_i = 1$ if $\delta^{(N+i)}$ is a support constraint of $P[\Omega_N, \delta^{(N+i)}]$, and it is zero otherwise.

- Let $V_N = V_N(\Omega_N)$ be defined as in (3). Then, since the extractions $\delta^{(N+i)}$, $i = 1, \ldots$, are iid, it follows that each $Z_i$ has probability $V_N$ of being equal to one, and probability $1 - V_N$ of being equal to zero, and that this probability is independent of $i$, i.e.,

$$Z_i = \begin{cases} 1 & \text{w.p. } V_N \\ 0 & \text{w.p. } 1 - V_N \end{cases}, \quad i = 1, 2, \ldots \tag{8}$$

# Violation indicator variables and de Finetti's representation

- The $Z_i$s form what is known as a *conditionally iid Bernoulli sequence*, meaning that, conditional on $V_N = v$, the $Z_i$s are independent Bernoulli variables with "success" probability $v$.
- For any integer $r \geq 1$ and any given $z_i \in \{0, 1\}$, $i = 1, \dots, r$, it thus holds that

$$\mathbb{P}^{N+r}\{Z_i = z_i,\ i = 1, \dots, r \mid V_N = v\} = v^{\sum_{i=1}^r z_i}(1 - v)^{r - \sum_{i=1}^r z_i}. \tag{9}$$

- Now, integrating (9) with respect to the probability measure $F_N$ of $V_N$, we obtain

$$\begin{aligned} &\mathbb{P}^{N+r}\{Z_i = z_i,\ i = 1, \dots, r\} \\ &\quad = \int_0^1 \mathbb{P}^{N+r}\{Z_i = z_i,\ i = 1, \dots, r \mid V_N = v\} \mathrm{d}F_N(v) \\ &\quad = \int_0^1 v^{\sum_{i=1}^r z_i}(1 - v)^{r - \sum_{i=1}^r z_i} \mathrm{d}F_N(v), \end{aligned} \tag{10}$$

# Violation indicator variables and de Finetti's representation

- Variables $Z_i$, $i = 1, \ldots, r$, are thus not independent, but they are *exchangeable*, meaning that for any integer $r \geq 1$ and for any permutation $j_1, \ldots, j_r$ of the indices $1, \ldots, r$, it holds that

$$\mathbb{P}^{N+r}\{Z_i = z_i,\, i = 1, \ldots, r\} = \mathbb{P}^{N+r}\{Z_i = z_{j_i},\, i = 1, \ldots, r\},$$

  for any given sequence $z_i \in \{0, 1\}$, $i = 1, \ldots, r$.

- This is due to the fact that the probability in (10) depends only on $\sum_{i=1}^{r} z_i$, a quantity which is invariant to permutations of the addends.
- Thus, equation (10) is nothing but the de Finetti representation of the probability law of the exchangeable sequence $\{Z_i\}_{i=1,2,\ldots}$, where $V_N$ is the directing random measure, and $F_N$ is its distribution.
- Our objective is to determine the cumulative probability distribution $F_N$ of $V_N$ (or a useful bound on it), by exploiting the special properties of the $J$ function.

# The distribution $F_N$ of the violation probability $V_N$

- Let beta$(\alpha, \beta)$ denote a beta density function with parameters $\alpha > 0$, $\beta > 0$:

$$\text{beta}(\alpha, \beta; t) = \frac{1}{B(\alpha, \beta)} t^{\alpha-1}(1-t)^{\beta-1},$$

  where

$$B(\alpha, \beta) \doteq \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)},$$

  and $\Gamma$ is the Gamma function.

- Also, denote by $\mathrm{Fbeta}(\alpha, \beta)$ the cumulative distribution function of the beta$(\alpha, \beta)$ density:

$$\mathrm{Fbeta}(\alpha, \beta; t) = \int_0^t \text{beta}(\alpha, \beta; \vartheta)\mathrm{d}\vartheta, \quad t \in [0, 1].$$

  $\mathrm{Fbeta}(\alpha, \beta; t)$ is the regularized incomplete beta function.

- A standard result establishes that, for $\alpha, \beta$ integers, it holds that

$$\mathrm{Fbeta}(\alpha, \beta; t) = \sum_{i=\alpha}^{\alpha+\beta-1} \binom{\alpha+\beta-1}{i} t^i (1-t)^{\alpha+\beta-1-i}, \quad t \in [0, 1].$$

# The distribution $F_N$ of the violation probability $V_N$

Main result

## Theorem

*Let $h = (n+1)2^d$ be the Helly's dimension of the mixed-integer domain $\mathbb{R}^n \times \mathbb{Z}^d$, and let $F_N$ be the cumulative probability distribution of the random violation probability $V_N$ defined in (3).*
*Then, for any $N \geq h$, it holds that*

$$F_N(v) \geq \mathrm{Fbeta}(h, N+1-h; v) = \sum_{i=h}^{N} \binom{N}{i} v^i (1-v)^{N-i}, \quad v \in [0,1]. \qquad (11)$$

...Yes, but what about Polya's urns?

# Constructing a Polya's urn

Consider $\Omega_{N+r}$ as a random "urn" containing "balls" $\delta^{(i)}$, $i = 1, \ldots, N + r$.

Coloring balls in the initial urn $\Omega_{N+1}$

- For any initial random urn $\Omega_{N+1}$, we assign colors to the balls according to the following method.
- We do a first pass over the elements of the initial urn, and color a ball black if it is a support constraint of $P[\Omega_{N+1}]$, and white otherwise. In this way, we paint black in the first pass a number $\Upsilon_{N+1}^{(1)}$ of balls. If $\Upsilon_{N+1}^{(1)} < h$, then we do a second pass, and paint in black additional $h - \Upsilon_{N+1}^{(1)}$ balls, chosen uniformly at random among those balls that were white after the first pass. In this way, the initial urn $\Omega_{N+1}$ will always contain exactly $h$ black balls.
- Moreover, if we select uniformly at random an integer $i \in \{1, \ldots, N + 1\}$, then the (total) probability that ball $\delta^{(i)}$ is black is $h/(N + 1)$, for all $i$.

# Constructing a Polya's urn

## Coloring other balls

- Next, we consider coloring the additional balls $\delta^{(N+i)}$, for $i = 2, \ldots$
- if $\delta^{(N+i)}$ is a support constraint of $P[\Omega_N, \delta^{(N+i)}]$, we label it black. If it is not a support constraint, we label it black anyways, with probability $(h - \Upsilon_{N+1}^{(i)})/(N+1)$ (where $\Upsilon_{N+1}^{(i)}$ is the number of support constraints of $P[\Omega_N, \delta^{(N+i)}]$), or white otherwise.
- Since the total probability of $\delta^{(N+i)}$ being a support constraint of $P[\Omega_N, \delta^{(N+i)}]$ is $\Upsilon_{N+1}^{(i)}/(N+1)$, we see that the above procedure assigns color black to $\delta^{(N+i)}$ w.p. $\Upsilon_{N+1}^{(i)}/(N+1) + (h - \Upsilon_{N+1}^{(i)})/(N+1) = h/(N+1)$.
- Moreover, it is easy also to see that if we select a uniform random integer $i \in \{1, \ldots, N+r\}$, then the (total) probability that the $i$-th element $\delta^{(i)}$ in urn $\Omega_{N+r}$ is black is $h/(N+1)$, for any $i$ and $r$.

# Constructing a Polya's urn

- This process thus behaves as a classical Polya's urn, in which the initial urn contains $h$ black balls and $N+1-h$ white balls and, at each step $r = 1, \ldots$, a ball is picked at random from urn $\Omega_{N+r}$, its color is observed, and then the ball is put back in the urn, together with an additional ball of the same color, thus forming the new urn $\Omega_{N+r+1}$, etc.
- Let us denote with $B_i$, $i = 1, \ldots$, the indicator variables of this auxiliary process, where $B_i = 1$ if the $i$-th ball is black, and $B_i = 0$ otherwise.
- By the known properties of the Polya's urn, the $B_i$s form an exchangeable sequence of conditionally iid Bernoulli variables. By de Finetti's representation theorem, this sequence admits a directing random measure, that we name $\tilde{V}$, and whose distribution we denote by $\tilde{F}$.

# The latent distribution of the Polya's urn

- According to known results about the Polya's urn process (see, e.g., Feller), we have that if the initial urn $\Omega_{N+1}$ contains $h$ black balls (and hence $N+1-h$ white balls), then the probability that the urn $\Omega_{N+1+r}$, contains exactly $h+r$ black balls (i.e., the probability that all $\delta^{(N+1+k)}$ are labeled black, for $k=1,\ldots,r$) is given by

$$\frac{\Gamma(h+r)\Gamma(N+1)}{\Gamma(N+1+r)\Gamma(h)}. \tag{12}$$

- Now, conditional on $\tilde{V}=v$, the probability that $\delta^{(N+1+k)}$ is labeled black is $v$ (independent of $k$), therefore the probability (conditional on $\tilde{V}=v$) that all $\delta^{(N+1+k)}$, $k=1,\ldots,r$, are labeled black is $v^r$, and hence the (unconditional) probability of this last event (which coincides with (12)) is

$$\int_0^1 v^r \mathrm{d}\tilde{F}(v),$$

that is, it is the the raw moment of order $r$ of the r.v. $\tilde{V}$, thus

$$\mathbb{E}\{\tilde{V}^r\} = \frac{\Gamma(h+r)\Gamma(N+1)}{\Gamma(N+1+r)\Gamma(h)}, \quad r=0,1,\ldots$$

# The latent distribution of the Polya's urn

- This sequence of moments forms a *completely monotone* sequence, hence it identifies uniquely the distribution $\tilde{F}$ which generates these moments.
- Direct inspection then shows that $\tilde{F}$ must be the cumulative of a beta$(h, N+1-h)$ density. Thus

$$\tilde{F}(v) = \text{Fbeta}(h, N+1-h; v), \quad v \in [0,1].$$

...Ok, but what is the relation with our original violation process?

# A stochastic dominance relation

- Let us now compare the two processes generated, respectively, by the "true" process $Z_i$, and by the auxiliary Polya's process $B_i$ constructed above.
- It is easy to see that, by the very way in which the auxiliary process is defined, any $\delta^{(N+i)}$ for which

$$J(\Omega_N, \delta^{(N+i)}) > J(\Omega_N),$$

  i.e., for which $Z_i = 1$ (see the definition of $Z_i$ in (7)) is labeled black in the Polya's urn process.
- Thus, for any $i = 1, \ldots$, we have that $Z_i = 1$ implies that $B_i = 1$, i.e.

$$B_i \geq Z_i, \quad \forall i = 1, \ldots$$

- Considering then the averages

$$E_B(k) \doteq \frac{1}{k} \sum_{i=1}^{k} B_i, \quad E_Z(k) \doteq \frac{1}{k} \sum_{i=1}^{k} Z_i,$$

  we have that

$$E_B(k) \geq E_Z(k), \quad \forall k = 1, \ldots \tag{13}$$

# A stochastic dominance relation

- But for any exchangeable sequence it holds that the empirical averages of the indicator variables converge almost surely (a.s.) to the respective directing random measures, hence, for $k \to \infty$, $E_B(k)$ converges a.s. to $\tilde{V}$, and $E_Z(k)$ converges a.s. to $V_N$.
- Thus, using (13), we have that

$$\tilde{V} \geq V_N \quad \text{w.p. } 1,$$

which is a state-wise stochastic dominance condition on the r.v. $\tilde{V}$ and $V_N$.

- From this latter condition it immediately follows that, for any $v \in [0, 1]$,

$$\tilde{V} \leq v \quad \Rightarrow \quad V_N \leq v \quad \text{w.p. } 1,$$

whence

$$\mathbb{P}^N\{V_N \leq v\} \geq \mathbb{P}^N\{\tilde{V} \leq v\} = \mathrm{Fbeta}(h, N - h + 1),$$

which conclude the proof of Theorem 2. □

# Explicit bound on $N$

- In typical applications of RCP theory, one wants to know how large $N$ should be in order to guarantee that the violation probability $V_N$ can be larger than some given (typically small) $v \in (0,1)$ only with probability smaller than some given (typically very small) level $\gamma \in (0,1)$.
- Using (11), we have that $\mathbb{P}^N\{V_N > v\} = 1 - F_N(v) \leq \sum_{i=0}^{h-1} \binom{N}{i} v^i (1-v)^{N-i}$.
- Using quite standard techniques for bounding the lower Binomial tail, one can find conditions on $N$ that guarantee that $\sum_{i=0}^{h-1} \binom{N}{i} v^i (1-v)^{N-i} \leq \gamma$.
- We obtain that the above condition is satisfied for integers $N$ such that

$$N \geq \frac{2}{v}\left(\ln \gamma^{-1} + h - 1\right).$$

Usually, $\gamma$ is set to a very small number, e.g., $\gamma = 10^{-6}$, or lower, so that the event $V_N \leq v$ happens with "practical certainty."

# Conclusions

- A totally new approach to RCP and scenario optimization theory.
- On the one hand, we removed all technical assumptions that plagued existing approaches (uniqueness of solution, nondegeneracy, etc.). On the other hand, the result extend fully from continuous domain to the mixed-integer one.
- The Polya's urn paradigm sheds new light on the inner stochastic mechanism governing random convex (and mixed-integer convex) programs.
- Perspective: possible solution of an open problem related to so-called *scenario problems with violated constraints* (not discussed here), with strict connections to chance-constrained optimization?

Refs.:


1. G. Calafiore, "Random Convex Programs," *SIAM J. OPT.*, 2010.
2. G. Calafiore, "Mixed-Integer Random Convex Programs — Polya's Urns and the Distribution of the Violation Probability," *Submitted*, 2014.