Stochastic optimization
Chance constrained programming

Fabian Bastin
fabian.bastin@umontreal.ca
Université de Montréal – CIRRELT – IVADO

A long story

- Introduced in 1959 by Charnes and Cooper
  `https://dl.acm.org/doi/10.1287/mnsc.6.1.73`
- And also a bit improbable.
- Cooper dropped high-school to support his family, and became a professional boxer.
- Became an accountant for Eric Louis Kohler, met while hitchhiking.
- Kohler financed his bachelor at University of Chicago.
- At 26, he enrolled at Columbia University and finished his coursework and dissertation, but never received his PhD due to its claim that decision making was not a centralized process.

A long story (cont'd)

- The collaboration with Charnes was however successful, with more than 200 publications, and led a successful academic carrer.
- Source: `https://www.informs.org/Explore/History-of-O.R.-Excellence/Biographical-Profiles/Cooper-William-W`

Cooper and Charnes

INFORMS John Von Neumann prize (with Richard J. Duffin)

Mathematical Formulation

$$
\begin{aligned}
\min_{x \in \mathcal{X}} \quad & \mathbb{E}[f(x, \xi)] \\
\text{s.t.} \quad & \mathbb{P}\left(g_i(x, \xi) \leq 0\right) \geq 1 - \alpha_i, \quad i = 1, \ldots, m,
\end{aligned}
$$

where:

- $\xi \in \mathbb{R}^p$ is a random vector with known probability distribution $\mathbb{P}$;
- $x \in \mathbb{R}^n$ is the vector of decision variables;
- $\mathcal{X} \subseteq \mathbb{R}^n$ is a convex deterministic set;
- $f(x, \xi)$ is the cost (or utility) function;
- $g_i(x, \xi)$ are random constraint functions;
- $\alpha_i \in [0, 1]$ specifies the acceptable risk level for constraint i.

Probabilistic constraints

The probabilistic constraint

$$\mathbb{P}\left(g_i(x,\xi) \leq 0\right) \geq 1 - \alpha_i$$

ensures that constraint i is satisfied with probability at least $1 - \alpha_i$. Equivalently, the probability of violation

$$\mathbb{P}\left(g_i(x,\xi) > 0\right) \leq \alpha_i$$

is limited to a user-specified tolerance. Smaller values of α_i produce more conservative, risk-averse solutions.

Individual and joint chance constraints

Two common variants are:

1. **Individual (or separate) chance constraints:** each constraint $g_i(x, \xi) \leq 0$ is required to hold with probability $1 - \alpha_i$;
2. **Joint (or integrated) chance constraints:**

$$\mathbb{P}\left(g_i(x, \xi) \leq 0,\ \forall i\right) \geq 1 - \alpha,$$

which ensures simultaneous satisfaction of all constraints with a single confidence level $1 - \alpha$.

Toy example

Source: J. Linderoth (`https://jlinderoth.github.io/classes/ie495/lecture22.pdf`)

Consider the toy problem

$$\begin{aligned} \min_{x} \ & x_1 + x_2 \\ \text{s.t. } & \xi_1 x_1 + x_2 \geq 7 \\ & \xi_2 x_1 + x_2 \geq 4 \\ & x_1, x_2 \geq 0. \end{aligned}$$

Instead of requiring that a constraint holds for all the scenarios, we can require that the constraint is satisfied with a given (large) probability.

Chance constraints

1. Separate chance constraints

$$P[\xi_1 x_1 + x_2 \geq 7] \geq \alpha_1$$
$$P[\xi_2 x_1 + x_2 \geq 4] \geq \alpha_2$$

2. Joint (integrated) chance constraint

$$P[\xi_1 x_1 + x_2 \geq 7 \cap \xi_2 x_1 + x_2 \geq 4] \geq \alpha$$

Example: joint chance constraints

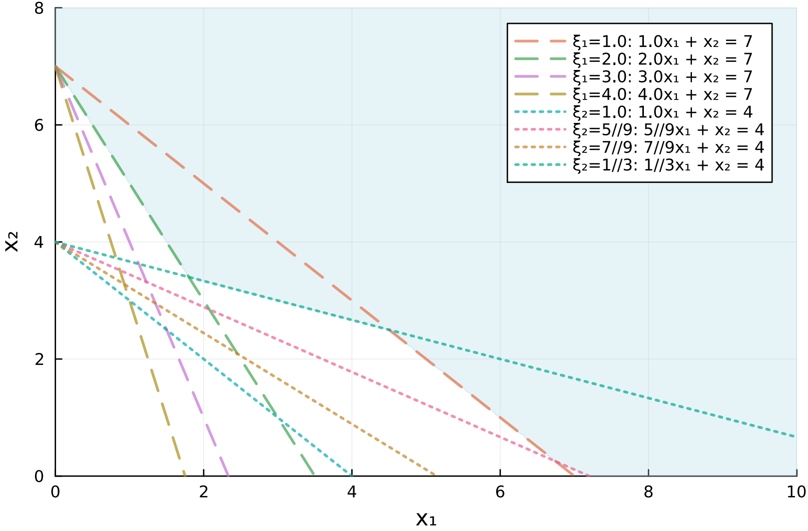
$$P[(\xi_1, \xi_2) = (1, 1)] = 0.1 \tag{1}$$
$$P[(\xi_1, \xi_2) = (2, 5/9)] = 0.4 \tag{2}$$
$$P[(\xi_1, \xi_2) = (3, 7/9)] = 0.4 \tag{3}$$
$$P[(\xi_1, \xi_2) = (4, 1/3)] = 0.1 \tag{4}$$

Assume that $\alpha \in (0.8, 0.9]$, and we have the joint constraint

$$P[\xi_1 x_1 + x_2 \geq 7 \cap \xi_2 x_1 + x_2 \geq 4] \geq \alpha$$

We then have to satisfy constraints (2) and (3) and either (1) or (4).

Example: frontiers of constraints



Feasible set

Feasible set:

$$K_1(\alpha) = \{x \mid P[g_i(x, \xi) \le 0] \ge 1 - \alpha_i,\ i, \dots, m\}$$

$K_1(\alpha)$ is not necessarily convex, even if $g_i(x, \xi)$, $i, \dots, m$, are linear in x.

Quasi-concavity

Definition (Quasi-Concavity)

A function $f : S \to \mathbb{R}$, where $S \subseteq \mathbb{R}^n$ is a convex set, is **quasi-concave** if for all $\mathbf{x}, \mathbf{y} \in S$ and all $\lambda \in [0, 1]$:

$$f(\lambda \mathbf{x} + (1 - \lambda)\mathbf{y}) \geq \min\{f(\mathbf{x}), f(\mathbf{y})\}$$

Definition

The probability measure P defined on the Borel sets of $\mathbb{R}^n$ is said to be quasi-concave, if for any convex subsets A, B of $\mathbb{R}^n$, and any $\lambda \in [0, 1]$,

$$P[(1 - \lambda)A + \lambda B] \geq \min\{P[A], P[B]\}.$$

(See Prékopa (2003), “Probabilistic Programming”, Chapter 5 in “Stochastic Programming”)

Applications in optimization

Theorem (Convexity of Upper Level Sets)

Let $f : S \to \mathbb{R}$ be a quasi-concave function, where $S \subseteq \mathbb{R}^n$ is a convex set. Then for any $\alpha \in \mathbb{R}$, the upper level set

$$U_\alpha = \{\boldsymbol{x} \in S \mid f(\boldsymbol{x}) \geq \alpha\}$$

is convex.

Corollary (Feasible Set Convexity)

If f is quasi-concave, then the constraint:

$$f(\boldsymbol{x}) \geq \alpha$$

defines a ***convex feasible set*** *for any $\alpha \in \mathbb{R}$.*

Applications in stochastic programming

For chance-constrained programming problems of the form:

$$P[g(\mathbf{x}, \xi) \leq 0] \geq 1 - \alpha,$$

if $P[g(\mathbf{x}, \xi) \leq 0]$ is quasi-concave in $\boldsymbol{x}$, then $\{\boldsymbol{x} \mid P[g(\boldsymbol{x}, \xi) \leq 0] \geq 1 - \alpha\}$ is **convex**, making the optimization problem **tractable.**

Log-concavity

Definition (Log-Concave Function)

A function $f : S \to \mathbb{R}_+$, where $S \subseteq \mathbb{R}^n$ is a convex set, is **log-concave** if for all $\boldsymbol{x}, \boldsymbol{y} \in S$ and $\lambda \in [0, 1]$:

$$f(\lambda \mathbf{x} + (1 - \lambda)\mathbf{y}) \geq f(\mathbf{x})^\lambda f(\mathbf{y})^{1-\lambda}$$

Theorem

If f is log-concave and $f > 0$, then f is quasi-concave.

Jointly Convex Constraint Function

Definition

A constraint function $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is **jointly convex** in $(\boldsymbol{x}, \xi)$ if for all $(\boldsymbol{x}_1, \xi_1), (\boldsymbol{x}_2, \xi_2) \in \mathbb{R}^n \times \mathbb{R}^m$ and all $\lambda \in [0, 1]$:

$$g(\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2, \lambda \xi_1 + (1-\lambda)\xi_2) \leq \lambda g(\boldsymbol{x}_1, \xi_1) + (1-\lambda) g(\boldsymbol{x}_2, \xi_2)$$

Prékopa's fundamental theorem

Theorem

Let $\boldsymbol{\xi} \in \mathbb{R}^m$ be a random vector with log-concave probability distribution, and let $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ be quasi-convex in $(\mathbf{x}, \boldsymbol{\xi})$. Then the probability function

$$\phi(\mathbf{x}) = P[g(\mathbf{x}, \xi) \leq 0]$$

is log-concave (and hence quasi-concave) in $\mathbf{x}$.

(Prékopa, A. (1973). On logarithmic concave measures and functions. Acta Scientiarum Mathematicarum 34, 335–343)

Examples of jointly quasi-convex functions

Example (Affine in $(\mathbf{x}, \xi)$)

If $g(\mathbf{x}, \xi) = \mathbf{a}^\top \mathbf{x} + \mathbf{b}^\top \xi + c$, where $\mathbf{a} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^m$, $c \in \mathbb{R}$, then g is jointly quasi-convex (in fact, jointly convex).

Example (Maximum)

The maximum of quasi-convex functions is quasi-convex.

Common log-concave distributions

$$\text{Normal:} \quad f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$$\text{Exponential:} \quad f(x) = \lambda \exp(-\lambda x), \quad x \geq 0$$

$$\text{Uniform:} \quad f(x) = \begin{cases} 1/\mu(S), & x \in S \\ 0 & \text{otherwise,} \end{cases}$$

where $\mu(S)$ is the measure of S

$$\text{Gamma (shape} \geq 1\text{):} \quad f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x > 0, \alpha \geq 1$$

$$\text{Beta } (\alpha \geq 1, \beta \geq 1)\text{:} \quad f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}, \quad x \in [0,1]$$

$$\text{Multivariate normal:} \quad f(x) = \frac{1}{\sqrt{(2\pi)^n/2 \det(\Sigma)}} e^{-\frac{1}{2}(x-\mu)'\Sigma(x-\mu)}$$

Quasi-concave probability distributions

If you have such a density, you can

- use Lagrangian techniques
- use a reduced-gradient technique (see Kall & Wallace, Section 4.1)

Joint constraints

Theorem (Prékopa's Theorem for Joint Constraints)

If

1. ξ *has a log-concave probability distribution*
2. *Each* $g_i(\boldsymbol{x}, \xi)$ *is jointly quasi-convex in* $(\boldsymbol{x}, \xi)$
3. *The set* $\{(\boldsymbol{x}, \xi) \mid g_i(\boldsymbol{x}, \xi) \leq 0,\ i = 1, \ldots, m\}$ *is convex*

Then the function

$$\phi(\mathbf{x}) = P\left[g_i(\mathbf{x}, \xi) \leq 0,\ i = 1, \ldots, m\right]$$

is log-concave (hence quasi-concave) in $\mathbf{x}$*, and thus the feasible set* $\{\mathbf{x} \mid \phi(\mathbf{x}) \geq 1 - \alpha\}$ *is convex.*

Maximum reformulation

For joint constraints $P[g_j(\mathbf{x},\xi) \leq 0,\ \forall j] \geq 1-\alpha$, define:

$$h(\mathbf{x},\xi) = \max_{j=1,\ldots,m} g_j(\mathbf{x},\xi)$$

Then the constraint becomes:

$$P[h(\mathbf{x},\xi) \leq 0] \geq 1-\alpha$$

Further reading

For a more complete analysis of convexity in chance-constrained programming, see Section 4.2 in Lectures notes on stochastic programming (3rd edition).

Linear inequalities

Assume that the chance constraint can be written in the form

$$P[T(\boldsymbol{\xi})x \geq h(\boldsymbol{\xi})] \geq \alpha.$$

Theorem

Suppose $T(\xi) = T$ is fixed, and $h(\boldsymbol{\xi})$ has a quasi-concave probability measure P. Then $K_1(\alpha)$ is convex for $0 \leq \alpha \leq 1$.

Single constraint: easy case

- The situation in the single constraint case is sometimes simple.
- Suppose again that $T_i(\xi) = T_i$ is constant. Then

$$P[T_i x \geq h_i(\xi)] = F(T_i x) \geq \alpha$$

so the deterministic equivalent is

$$T_i x \geq F^{-1}(\alpha)$$

…linear constraint! The resulting problem is still linear. We have simply relaxed the contraint.

- Recall that the inverse of the cdf is defined as

$$F^{-1}(\alpha) = \min\{x : F(x) \geq \alpha\}.$$

Other "solvable" cases

Let $h(\xi) = h$ be fixed, $T(\xi) = \mathrm{diag}(\xi_1, \xi_2, \ldots, \xi_n)$, with $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ a multivariate normal distribution with mean $\mu = (\mu_1, \mu_2, \ldots, \mu_n)$ and variance-covariance matrix Σ. Then

$$\frac{\sum_{i=1}^{n} \xi_i x_i - \mu^T x}{\sqrt{x^T \Sigma x}} \sim N(0, 1),$$

and

$$K_1(\alpha) = \{x \mid \mu^T x \geq h + \Phi^{-1}(\alpha)\sqrt{x^T \Sigma x}\},$$

where Φ is the standard normal cdf.

$K_1(\alpha)$ is a convex set for $\alpha \geq 0.5$.

It is possible to express it as a second order cone constraint:

$$\|\Sigma^{1/2} x\|_2 \leq \frac{1}{\Phi^{-1}(\alpha)}(\mu' x - h)$$

Second-order cone programming

A second-order cone program (SOCP) is a convex optimization problem of the form

$$\begin{aligned} \min_x \ & f^T x \\ \text{s.t.} \ & \|A_i x + b_i\|_2 \leq c_i^T x + d_i, \ i = 1, \ldots, m \\ & Fx = g \end{aligned}$$

where $x \in \mathbb{R}^n$, $f, c_i \in \mathbb{R}^n$, $A_i \in \mathbb{R}^{n_i \times n}$, $b_i \in \mathbb{R}^{n_i}$, $d_i \in \mathbb{R}$, $F \in \mathbb{R}^{p \times n}$, and $g \in \mathbb{R}^p$.

SOCPs can be solved by interior point methods.

Example: robust portfolio optimization

(Taken from S. Boyd and J. Linderoth)

- Suppose we want to invest in n assets, providing random return rates $\beta_1, \beta_2, \ldots, \beta_n$.
- $\boldsymbol{\beta} \sim N(\boldsymbol{\mu}, \Sigma)$.
- x: total amount to invest.
- Suppose that we want to ensure a return of at least b. We cannot guarantee it all the time, but we want it to occur most of the time.

Example: robust portfolio optimization (cont'd)

Let $x_i \geq 0$ the part of portfolio to invest in asset i. Constraints:

$$P\left[\sum_{i=1}^{n} \beta_i x_i \geq b\right] \geq \alpha \tag{5}$$

$$\sum_{i=1}^{n} x_i \leq x$$

$$x_i \geq 0, \; i = 1, \ldots, n.$$

(5) can be rewritten as

$$\mu^T x - \Phi^{-1}(\alpha)\sqrt{x^T \Sigma x} \geq b.$$

If $b < 0$, (5) is also known as Value at Risk constraint (Ruszczyński and Shapiro, 2003).

Example: robust portfolio optimization (cont'd)

We can also interpret x_i as proportion of the portfolio (position of asset i), by normalizing $\|x\|_1$ to 1. b is now the minimum return rate of the portfolio and x is the portfolio allocation.

We can add some constraints on the x_i to ensure diversification. We summarize them by requiring $x \in \mathcal{C}$.

A complete program can now be expressed as

$$\begin{aligned} \max_x \; & E[\beta^T x] = \mu^T x \\ \text{s.t.} \; & P\left[\beta^T x \geq b\right] \geq \alpha \\ & \sum_{i=1}^{n} x_i = 1 \\ & x \in \mathcal{C} \end{aligned}$$

Example: loss constraint

Setting b to 0 means that we want to ensure that we will no suffer from loss with some probability. Typicially, α is set to 0.9, 0.95, 0.99,...

The chanced-constraint can also be expressed as

$$P\left[\beta^T x \leq 0\right] \leq 1 - \alpha = \gamma.$$

We can also allow the sale of some parts of the portfolio by allowing some x_i to be negative.

Numerical illustration

(Taken from S. Boyd – `http://ee364a.stanford.edu/lectures/chance_constr.pdf`)

$n = 10$ assets, $\alpha = 0.95$, $\gamma = 0.05$, $\mathcal{C} = \{x | x \succeq -0.1\}$

Compare

- optimal portfolio
- optimal portfolio without loss risk constraint
- uniform portfolio $(1/n)\mathbf{1}$

portfolio	$E[\beta^T x]$	$P[\beta^T x \leq 0])$
optimal	7.51	5.0%
w/o loss constraint	10.66	20.3%
uniform	3.41	18.9%

Short selling case

Let's ignore the non-negativity constraints and consider the program (Ruszczyński and Shapiro, 2003)

$$\begin{aligned} &\min_x \; -\mu^T x \\ &\text{s.t. } \Phi^{-1}(\alpha)\sqrt{x^T \Sigma x} - \mu^T x + b \leq 0. \end{aligned}$$

The Lagrangian is

$$\begin{aligned} L(x,\lambda) &= -\mu^T x + \lambda\left(\Phi^{-1}(\alpha)\sqrt{x^T\Sigma x} - \mu^T x + b\right) \\ &= -(1+\lambda)\mu^T x + \lambda\Phi^{-1}(\alpha)\sqrt{x^T\Sigma x} + \lambda b \end{aligned}$$

Short selling case: KKT conditions

$$
\begin{aligned}
&\frac{dL(x,\lambda)}{dx} = 0 \\
&\Phi^{-1}(\alpha)\sqrt{x^T\Sigma x} - \mu^T x + b \leq 0 \\
&\lambda\left(\Phi^{-1}(\alpha)\sqrt{x^T\Sigma x} - \mu^T x + b\right) = 0 \\
&x \geq 0,\ \lambda \geq 0.
\end{aligned}
$$

Short selling case: solving the KKT conditions

We have

$$\frac{dL(x,\lambda)}{dx} = -(1+\lambda)\mu + \frac{\lambda\Phi^{-1}(\alpha)\Sigma x}{\sqrt{x^T\Sigma x}}$$

If $\lambda = 0$,

$$\frac{dL(x,\lambda)}{dx} = 0 \Rightarrow \mu = 0.$$

Thus, wlog, we assume $\lambda \neq 0$. Therefore

$$\Phi^{-1}(\alpha)\sqrt{x^T\Sigma x} - \mu^T x + b = 0$$

Short selling case: no risk-free asset

(Ruszczyński and Shapiro, 2003) Assume Σ nonsingular and define

$$\rho = \sqrt{\mu^T \Sigma^{-1} \mu}$$

We can show

$$\begin{cases} \text{unbounded problem} & \text{if } \rho \geq \Phi^{-1}(\alpha); \\ x^* = \frac{b}{\rho(\Phi^{-1}(\alpha) - \rho)} \Sigma^{-1} \mu & \text{if } \rho < \Phi^{-1}(\alpha). \end{cases}$$

Generalization

A more general form is

$$\begin{aligned} \min_x \ & h(x) \\ \text{s.t. } & P[g_1(x,\xi) \le 0, \ldots, g_r(x,\xi) \le 0] \ge \alpha \\ & h_1(x) \le 0, \ldots, h_m(x) \le 0. \end{aligned}$$

or

$$\begin{aligned} \min_x \ & h(x) \\ \text{s.t. } & \mathbb{E}\left[\mathcal{I}_{(0,\infty)}\left(g_1(x,\xi) \le 0, \ldots, g_r(x,\xi) \le 0\right)\right] \ge \alpha \\ & h_1(x) \le 0, \ldots, h_m(x) \le 0, \end{aligned}$$

where

$$\mathcal{I}_{(0,\infty)}(t) = \begin{cases} 1 & \text{if } t \le 0, \\ 0 & \text{otherwise.} \end{cases}$$

Solution methods for the general case

- Usually very hard.
- Use a bounding approximation or sample average approximation (SAA).
- We will discuss about it in more details when introducing Monte Carlo techniques.

Probabilistic programming

Source: András Prékopa (2003), “Probabilistic Programming”, Chapter 5 in “Stochastic Programming”, A. Ruszczyński and A. Shapiro (editors), Elsevier.

- Sometimes we only want to maximize a probability.
- General form:

$$\max_x P[g_1(x,\xi) \le 0, \ldots, g_r(x,\xi) \le 0]$$

$$\text{subject to } h_1(x) \le 0, \ldots, h_m(x) \le 0.$$

Measures of violation

- A chance constraint allows constraint violation with some probability.
- The violation can be large.
- It is often desirable to avoid too large violations.
- Can we penalize the violation?

Value at Risk

Source: `https://web.stanford.edu/class/ee364a/lectures/chance_constr.pdf`

Value-at-risk of random variable Z, at level η:

$$\mathrm{VaR}(Z;\eta) = \inf\{\gamma \mid P[Z \le \gamma] \ge \eta\}$$

Therefore, the value-at-risk is simply the inverse of the cdf evaluated at η!

$$\mathrm{VaR}(Z;\eta) = F_Z^{-1}(\eta).$$

Conditional Value at Risk

$$\mathrm{CVaR}(Z;\eta) = \inf_{\beta} \left(\beta + \frac{1}{1-\eta} \mathbb{E}\left[(Z-\beta)_+ \right] \right).$$

Assume that the distribution of Z is continuous.

Solution β^* obtained by solving

$$0 = \frac{d}{d\beta} \left(\beta + \frac{1}{1-\eta} \mathbb{E}\left[(Z-\beta)_+ \right] \right) = 1 - \frac{1}{1-\eta} P[Z \geq \beta],$$

leading to

$$\begin{aligned} & P[Z \geq \beta] = 1-\eta \\ \Leftrightarrow \quad & P[Z \leq \beta] = \eta = \mathrm{VaR}(Z;\eta). \end{aligned}$$

Expected shortfall

Conditional tail expectation (or expected shortfall)

$$
\begin{aligned}
\mathbb{E}[z \mid z \geq \beta^*] &= \mathbb{E}[\beta^* + (z - \beta^*) \mid z \geq \beta^*] \\
&= \beta^* + \frac{\mathbb{E}\left[(z - \beta^*)_+\right]}{P[z \geq \beta^*]} \\
&= \mathrm{CVaR}(z; \eta)
\end{aligned}
$$

- Can be added to the objective.
- Can be used as a constraint: *conditional expectation constraint*

$$
\mathbb{E}[z \mid z \geq \beta^*] \leq d.
$$

Integrated chance constraints

- Consider the stochastic constraints

$$g_i(x, \xi) \leq 0, \; 1, \ldots, r.$$

- Integrated chance constraint:

$$\mathbb{E}\left[\max_i (g_i(x, \xi))_+\right] \leq d.$$

For more details, see Chapter 6, Willem K. Klein Haneveld, Maarten H. van der Vlerk, Ward Romeijnders (2020), "Stochastic Programming - Modeling Decision Problems Under Uncertainty", Springer.