Stochastic Optimization

Basic Properties and Theory

Alireza Ghaffari-Hadigheh

Azarbaijan Shahid Madani University (ASMU)

hadigheha@azaruniv.ac.ir

Fall 2017

Overview

Formulation

- Basic two-stage stochastic linear program

$$\begin{array}{rcc} \min\ z & = & c^T x + E_\xi[\min q(\omega)^T y(\omega)] \\ s.t. & & Ax = b \\ & & T(\omega)x + Wy(\omega) = h(\omega) \\ & & x \geq 0, y(\omega) \geq 0, \end{array} \quad (1)$$

- c: a known vector in $\mathbb{R}^{n_1}$,
- b: a known vector in $\mathbb{R}^{m_1}$,
- A and W: known matrices of size $m_1 \times n_1$ and $m_2 \times n_2$, respectively,
- W: The recourse matrix, assumed to be fixed.
- $T(\omega)$: $m_2 \times n_1$, $q(\omega) \in \mathbb{R}^{n_2}$, $h(\omega) \in \mathbb{R}^{m_2}$.
- $\xi^T(\omega) = (q(\omega)^T, h(\omega)^T, T_{1\cdot}(\omega), \ldots, T_{m_2\cdot}(\omega))$ with $N = n_2 + m_2 + (m_2 \times n_1)$
- $T_{i\cdot}(\omega)$ is the i -th row of the technology matrix $T(\omega)$.

- Equivalent to (1):

$$\begin{aligned} \min\ z &= c^T x + \mathcal{Q}(x) \\ s.t.\ & Ax = b \\ & T(\omega)x + Wy(\omega) = h(\omega) \\ & x \geq 0, \end{aligned} \tag{2}$$

$$\mathcal{Q}(x) = E_\xi[x, \xi(\omega)] \tag{3}$$

$$Q(x, \xi(\omega)) = \min_y \{q(\omega)^T y | Wy = h(\omega) - T(\omega)x, y \geq 0\}. \tag{4}$$

- When T is nonstochastic:

$$\begin{aligned} \min\ z &= c^T x + \Psi(\chi) \\ s.t.\ & Ax = b \\ & Tx - \chi = 0 \\ & x \geq 0. \end{aligned} \tag{5}$$

  - $\Psi(\chi) = E_\xi \Psi(\chi, \xi(\omega))$
  - $\Psi(\chi, \xi(\omega)) = \min\{q(\omega) Ty | Wy = h(\omega) - \chi, y \geq 0\}$
  - generating an m_2 -dimensional tender $\chi = Tx$ to be bid against the outcomes $h(\omega)$ of the random events.

- $K_1 = \{x | Ax = b, x \geq 0\}$.
- Elementary feasibility set:
  $K_2(\xi) = \{x | y \geq 0 \text{ exists s. t. } W(\omega)y = h(\omega) - T(\omega)x\}$.

Example

$$\begin{array}{ll} \min & 2y_1 + y_2 \\ s.t. & y_1 + 2y_2 \geq \xi_1 - x_1, \\ & y_1 + y_2 \geq \xi_2 - x_1 - x_2, \\ & 0 \leq y_1 \leq 1, 0 \leq y_2 \leq 1. \end{array}$$

Using the upper bounds on y,

- The first constraint implies $\xi_1 - x_1 \leq 3$
- The second one implies $\xi_2 - x_1 - x_2 \leq 2$.
- Thus, $K_2(\xi) = \{x | x_1 \geq \xi_1 - 3, x_1 + x_2 \geq \xi_2 - 2\}$.

As ξ is discrete, we may easily define the second-stage feasibility set as

$$K_2 = \bigcap_{\xi \in \Xi} K_2(\xi).$$

Example

- ξ_1 takes the value 2, 3, 4, $\quad \xi_2$ the values 1, 4, 7
- With some nonspecified probabilities, independently of each other or not,
- $K_2 = \{x | x_1 \geq 1, x_1 + x_2 \geq 5\}$.
- It suffices to know the componentwise maximum of ξ to obtain K_2.
- This set is a polyhedron.

Definition

$\text{pos}\, W = \{t | Wy = t, y \geq 0\}$. Positive hull of W.

Theorem

1. *For a given ξ, the elementary feasibility set $K_2(\xi)$ is a convex polyhedron.*
2. *When ξ is a finite discrete random variable, K_2 is a convex polyhedron.*

Proof.

- Consider some x and ξ such that no $y \geq 0$ exists such that $W(\omega)y = h(\omega) - T(\omega)x$.
- Some x and ξ such that $h(\omega) - T(\omega)x \notin \text{pos}\, W(\omega)$.
- Thus, we have a point, $h(\omega) - T(\omega)x$, which does not belong to a convex set, $\text{pos}\, W(\omega)$.
- Then, there must exist some hyperplane, say $\{x | \sigma^T x = 0\}$, that separates $h(\omega) - T(\omega)x$ from $\text{pos}\, W(\omega)$.
- This hyperplane satisfies $\sigma^T t < 0$ for $t \in \text{pos}\, W(\omega)$ and $\sigma^T (h(\omega) - T(\omega)x) > 0$.
- For one particular ξ, $W(\omega)$ is fixed and there can be only finitely many different such hyperplanes.

The intersection of finitely many convex polyhedra is a convex polyhedron.

- For fixed value of x and ξ, the value $Q(x,\xi)$ of the second-stage program is given by

$$Q(x,\xi) = \min_{y}\{q(\omega)^T y | W(\omega)y = h(\omega) - T(\omega)x, y \geq 0\}, \tag{6}$$

- Difficulties may arise when the mathematical program (6) is unbounded below or infeasible.
- Unboundedness typically results of an ill-defined model and can easily be avoided by adding upper bounds on y.
- Infeasibility is avoided if we only consider $x \in K_2$.
- Thus, for $x \in K_2$, $Q(x,\xi)$ is finite for all ξ and we may define

$$\mathcal{Q}(x) = E_{\xi} Q(x,\xi) = \sum_{k=1}^{K} p_k Q(x,\xi_k)$$

- $k = 1, \ldots, K$ represents the K realizations of ξ.

The deterministic equivalent program $\min\{z(x) = c^T x + \mathcal{Q}(x), x \in K_1 \cap K_2\}$.

Theorem

For a given ξ, the value function $Q(x,\xi)$ is

(a) *a piecewise linear convex function in (h, T) ;*

(b) *a piecewise linear concave function in q ;*

(c) *a piecewise linear convex function in x for all $x \in K_2$.*

When ξ is a finite discrete random variable, $\mathcal{Q}(x)$ is piecewise linear and convex on K_2.

Proof.

- To prove convexity in (a) and (c), we just need to prove that $f(b) = \min\{q^T y | Wy = b\}$ is a convex function in b.
- Consider two different vectors, say b_1 and b_2, and some convex combination $b_\lambda = \lambda b_1 + (1-\lambda) b_2, \lambda \in (0,1)$.
- Let y_1^* and y_2^* be some optimal solution of $\min\{q^T y | Wy = b\}$ for $b = b_1$ and $b = b_2$, respectively.
- Then, $\lambda y_1^* + (1-\lambda) y_2^*$ is a feasible solution of $\min\{q^T y | Wy = b_\lambda\}$.
- Let y_λ^* be an optimal solution of this last problem.
- We have

$$\begin{aligned} f(b_\lambda) &= q^T y_\lambda^* \leq q^T (\lambda y_1^* + (1-\lambda) y_2^*) \\ &= \lambda q^T y_1^* + (1-\lambda) q^T y_2^* = \lambda f(b_1) + (1-\lambda) f(b_2) \end{aligned}$$

Proof. (continued...)

- A similar proof can be given to show concavity in q.
- To prove piecewise linearity, observe that solving (6) for given x and ξ amounts to finding some square submatrix $B(\omega)$ of $W(\omega)$, called a basis, such that $y_B = B(\omega)^{-1}(h(\omega) - T(\omega)x$, $y_N = 0$, where y_B is the subvector associated with the columns of B and y_N includes the remaining components of y.
- A basis is feasible if $y_B \geq 0$ and a feasible basis is optimal if $a_B(\omega)^T B(\omega)^{-1} W(\omega) \leq q(\omega)^T$.
- As long as these conditions hold, we have $Q(x, \xi) = qB(\omega)^T B(\omega)^{-1}(h(\omega) - T(\omega)x)$, which is linear in q, h, T and x on a domain defined by the feasibility and optimality conditions.
- Piecewise linearity follows from the existence of finitely many different optimal bases for the second-stage program.

Example

$$\begin{aligned} \min \quad & 2y_1 + y_2 \\ s.t. \quad & y_1 + 2y_2 \geq \xi_1 - x_1, \\ & y_1 + y_2 \geq \xi_2 - x_1 - x_2, \\ & y_1, y_2 \geq 0. \end{aligned}$$

- To reduce the calculations, assume $0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1$.

  (i) If $\xi \leq x_1 + x_2 \Rightarrow y_1 = 0, y_2 = 1 - x_1$;
  (ii) If $\xi > x_1 + x_2 \Rightarrow y_1 = \xi - x_1 - x_2$ and $y_2 = (1 - \xi + x_2)^+$ where $a^+ = \max(a, 0)$

$$Q(x,\xi) = \begin{cases} 1 - x_1 & for \quad 0 \leq \xi < x_1 + x_2, \\ \xi + 1 - 2x_1 - x_2 & for \quad x_1 + x_2 \leq \xi \leq 1 + x_2, \\ 2(\xi - x_1 - x_2) & for \quad 1 + x_2 \leq \xi. \end{cases}$$

- $Q(x,\xi)$ is clearly piecewise linear in x.

Another property when q and T are fixed

- For any $\lambda \geq 0$

$$Q(x, [q, \lambda(h') + Tx, T]) = \lambda Q(x, [q, h' + Tx, T]) \quad (7)$$

- Because a dual optimal solution for $h = h' + Tx$ is also dual feasible for $h = \lambda(h') + Tx$ and complementary with y^* optimal for $h = h' + Tx$.
- Because λy^* is also feasible for $h = \lambda(h') + Tx$, λy^* is optimal for $h = \lambda(h') + Tx$, demonstrating (7).
- This says that $Q(x, [q, h' + Tx, T])$ is a positively homogeneous function of h'.
- From the convexity of $Q(x, [q, h' + Tx, T])$ in $h = h' + Tx$, this function is also sublinear in h'.

ξ is not a discrete random variable

- For fixed value of x and ξ, the value of the second-stage program is, as before, given by (6).

$$Q(x,\xi) = \min_{y}\{q(\omega)^T y | W(\omega)y = h(\omega) - T(\omega)x, y \geq 0\}.$$

- When the mathematical program (6) is unbounded below or infeasible, the value of the second-stage program is defined to be $-\infty$ or $+\infty$, respectively.
- The expected second-stage value is, as given in (3) $\mathcal{Q}(x) = E_{\xi}[x, \xi(\omega)]$
- $+\infty + (-\infty) = +\infty$
    - Conservative attitude, rejecting any first-stage decision that could lead to an undefined recourse action for some realization even if some other realization would induce an infinitely low-cost.
    - Reflects the fact that second-stage programs can easily be bounded by bounding y, while infeasibilities may be inherent to the problem.

Elementary feasibility set

- For any given ξ,

$$\begin{aligned} K_2(\xi) &= \{x | Q(x,\xi) < \infty\} \\ K_2(\xi) &= \{x | y \geq 0 \ exists \ s.\ t.\ W(\omega)y = h(\omega) - T(\omega)x\}. \end{aligned}$$

- Both definitions are equivalent for a given ξ and enjoy the properties of Theorem (Page 9).
- When ξ is not a discrete random variable, we define K_2 in two different ways:

$$K_2 = \{x | \mathcal{Q}(x) < \infty\}$$

or

$$K_2^P = \bigcap_{\xi \in \Xi} K_2(\xi)$$

- K_2^P is said to define the possibility interpretation of the second-stage feasibility set.
- A first-stage decision x belongs to K_2^P if, for all possible values of the random vector ξ, a feasible second-stage decision can be taken.

Two sets, K_2 and K_2^P, can indeed be different when the random variable is a continuous random variable.

Example

- Consider an example where the second stage is defined by

$$Q(x,\xi) = \min_y \{y | \xi y = 1 - x, y \geq 0\}$$

- ξ has a triangular distribution on [0,1], namely, $P(\xi \leq u) = u^2$
- W reduces to a 1×1 matrix and is the only random element.
- For all ξ in (0,1], the optimal y is $\frac{1-x}{\xi}$

$$K_2(\xi) = \{x | x \leq 1\}$$

$$Q(x,\xi) = \frac{1-x}{\xi}, \forall x \leq 1$$

Example

Example (Continued)

- When $\xi = 0$, $K_2(0) = \{x | x = 1\}$.
- for $x \neq 1$, $Q(x, 0)$ should normally be $+\infty$. However, because the probability that $\xi = 0$ is zero, the convention is to take $Q(x, 0) = 0$. This corresponds to defining $0 \times \infty = 0$.
- $K_2^P = \{x|x = 1\} \cap \{x|x \leq 1\} = \{x|x = 1\}$ while

$$\mathcal{Q}(x) = \int_0^1 \frac{1-x}{\xi} \times 2\xi d\xi = 2(1-x), \forall x \leq 1$$

- $K_2 = \{x|x \leq 1\}$ and K_2^P is strictly contained in K_2.

A point is not in K_2^P as soon as it is infeasible for some ξ value, regardless of the distribution of ξ, while K_2 does not consider infeasibilities occurring with zero probability.

This kind of difficulty rarely occurs for programs with a fixed W matrix. It never occurs when the random vector satisfies some conditions.

Proposition 3.

If ξ has finite second moments, then

$$P(\omega | Q(x,\xi) < \infty) = 1 \text{ implies } \mathcal{Q}(x) < \infty.$$

See (Walkup and Wets [1967]).

Theorem

For a stochastic program with fixed recourse where ξ has finite second moments, the sets K_2 and K_2^P coincide.

Proof.

- First consider $x \in K_2^P$.
- This implies $Q(x,\xi) < \infty$ with probability one, so that, by Proposition 3, $\mathcal{Q}(x)$ is bounded above and $x \in K_2$.
- Now, consider $x \in K_2$.
- It follows that $\{\xi | Q(x,\xi) < \infty\}$ is a set of measure one.
- Observe that $Q(x,\xi) < \infty$ is equivalent to $h(\omega) - T(\omega)x \in \operatorname{pos} W$ and that $h(\omega) - T(\omega)x$ is a linear function of ξ, and $\{\xi \in \Sigma | Q(x,\xi) < \infty\}$ is a closed subset of Σ of measure one, for any set Σ of measure one.
- In particular, $\{\xi \in \Xi | Q(x,\xi) < \infty\}$ is a closed subset of Ξ having measure one.
- By definition of Ξ, this set can only be Ξ itself, so that $\{\xi | Q(x,\xi) < \infty\}\Xi$ and therefore $x \in K_2^P$.

□

- A third definition of the second-stage feasibility set: $\{x | Q(x, \xi) < \infty \text{with probability one}\}$.
- For problems with fixed recourse where ξ has finite second moments, this set also coincides with K_2 and K_2^P.
- In the following, we simply speak of K_2, the second-stage feasibility set.

Theorem

When W is fixed and ξ has finite second moments:

(a) *K_2 is closed and convex.*

(b) *If T is fixed, K_2 is polyhedral.*

(c) *Let Ξ_T be the support of the distribution of T. If $h(\xi)$ and $T(\xi)$ are independent and Ξ_T is polyhedral, then K_2 is polyhedral.*

Proof.

- The proof of (a) is elementary under the possibility representation of K_2.
- If T is fixed, $x \in K_2$ if and only if $h(\xi) - Tx \in \text{pos}\, W$ for all $\xi \in \Xi_h$, where Ξ_h is the support of the distribution of $h(\xi)$.
- Consider some x and ξ s.t. $h(\xi) - Tx \notin \text{pos}\, W$. Then there must exist some hyperplane, say $\{x | \sigma^T x = 0\}$ that separates $h(\xi) - Tx$ from $\text{pos}\, W$.
- This hyperplane must satisfy $\sigma^T t \leq 0$ for $t \in \text{pos}\, W$ and $\sigma^T(h(\xi) - Tx) > 0$.
- Because W is fixed, there need only be finitely many different such hyperplanes, so that $h(\xi) - Tx \in \text{pos}\, W$ is equivalent to $W^*(h(\xi) - Tx) \leq 0$ for some matrix W^*.

Proof (Continued)

- This matrix, called the polar matrix of W, is obtained by choosing some minimal set of separating hyperplanes.
- The set is minimal if removing any hyperplane would no longer guarantee the equivalence between $h(\xi) - Tx \in \text{pos}\, W$ and $W^*(h(\xi) - Tx) \leq 0$ for all x and ξ in Ξ_h.
- It follows that $x \in K_2$ if and only if $W^*(h(\xi) - Tx) \leq 0$ for all $\xi in \Xi$.
- This can still be an infinite system of linear inequalities due to $h(\xi)$.
- We may, however, replace this system by

$$(W^* T)_{i.} x \geq u_i^* = \sup_{h(\xi) \in \Xi h} W_{i.}^* h(\xi), i = 1, \ldots, I, \tag{8}$$

where $W_{i.}^*$ is the i -th row of W^* and I is the finite number of rows of W^*.

Proof (Continued)

- If for some i, u_i^* is unbounded, then the problem is infeasible and the result in (b) is trivially satisfied.
- If, for all i, $u_i^* < \infty$, then the system (8) constitutes a finite system of linear inequalities defining the polyhedron $K_2 = \{x | W^* Tx \geq u^*\}$ where u^* is the vector whose i- th component is u_i^*. This proves (b).
- When T is stochastic, a relation similar to (8) holds, which, unless Ξ_T is finite, defines an infinite system of inequalities.
- Whenever Ξ_T is polyhedral, (c) can be proved by working on the extremal elements of Ξ_T. (See Wets [1974, Corollary 4.13].)

$Q(x, \xi)$ it is not $-\infty$

Theorem

For a stochastic program with fixed recourse where ξ has finite second moments,

(a) *$\mathcal{Q}(x)$ is a Lipschitzian convex function and is finite on K_2.*

(b) *If $F(\xi)$ is an absolutely continuous distribution, $\mathcal{Q}(x)$ is differentiable on $\mathrm{ri}\ K_2$.*

Proof.

Convexity and finiteness in (a) are immediate. A proof of the Lipschitz condition can be found in Wets [1972] or Kall [1976]. □

Remarks

- When the random variables are appropriately described by a finite distribution,
  - The constraint set K_2 is defined by the possibility interpretation and is easily seen to be polyhedral.
  - The second-stage recourse function $\mathcal{Q}(x)$ is piecewise linear and convex on K_2.
- When the random variables cannot be described by a finite distribution,
  - They can usually be associated with some probability density.
  - Many common probability densities are absolutely continuous and have finite second moments; so, the constraints set definitions K_2 and K_2^P coincide and the second-stage value function $\mathcal{Q}(x)$ is differentiable and convex.
  - Classical nonlinear programming techniques could then be applied.

- In general, one can only compute $\mathcal{Q}(x)$ by numerical integration of $Q(x, \xi)$, for a given value of x.
- Most nonlinear techniques would also require the gradients of $\mathcal{Q}(x)$, which in turn require numerical integration.
- We come to the conclusion that numerical integration, as of today, produces an effective computational method only when the random vector is of small dimensionality.
- As a consequence, the practical solution of stochastic programs having continuous random variables is, in general, a difficult problem.
- One line of approach is to approximate the random variable by a discrete one and let the discretization be finer and finer, hoping that the solutions of the successive problems with discrete random variables will converge to the optimal solution of the problem with a continuous random variable.

Special case: $K_1 \subseteq K_2$

- Every solution x that satisfies the first-period constraints, $Ax = b$, also has a feasible completion in the second stage.
- The stochastic program has relatively complete recourse.
- Although relatively complete recourse is very useful in practice and in many of the theoretical results that follow, it may be difficult to identify because it requires some knowledge of the sets K_1 and K_2.
- A special type of relatively complete recourse may be identified from the structure of W.
- This form, called complete recourse, holds when there exists $y \geq 0$ such that $Wy = t$ for all $t \in \mathbb{R}^{m_2}$.

- Complete recourse is also represented by $\operatorname{pos} W = \mathbb{R}^{m_2}$
- It says that W contains a positive linear basis of $\mathbb{R}^{m_2}$.
- Complete recourse is often added to a model to ensure that no outcome can produce infeasible results.
- With most practical problems, this should be the case.
- In some instances, complete recourse may not be apparent.
- A special type of complete recourse offers additional computational advantages to stochastic programming solutions.
- It is called simple recourse.
- For a simple recourse problem, $W = [I, -I]$, y is divided correspondingly as (y^+, y^-), and $q = (q^+, q^-)$.
- In this case, the optimal values of $y_i^+(\omega), y_i^-(\omega)$ are determined purely by the sign of $h_i(\omega) - T_{i.}(\omega)x$ provided that $q_i^+ + q_i^- \geq 0$ with probability one.

Theorem

Suppose the two-stage stochastic program in (1) is feasible and has simple recourse and that ξ has finite second moments. Then $\mathcal{Q}(x)$ is finite if and only if $q_i^+ + q_i^- \geq 0$ with probability one.

Proof.

- If $q_i^+(\omega) + q_i^-(\omega) < 0$ for $\omega \in \Omega_1$ where $P(\Omega_1) > 0$, then, for any feasible x in (1), for all $\omega \in \Omega_1$ where $h_i(\omega) - T_{i.}(\omega)x > 0$, let $y_i^+(\omega) = h - i(\omega) - T_{i.}(\omega)x + u$, $y_i^-(\omega) = u$. By letting $u \to \infty$, $Q(x,\omega) \to -\infty$. A similar argument applies if $h_i(\omega) - T_{i.}(\omega)x \leq 0$, so $\mathcal{Q}(x)$ is not finite.
- If $q_i^+ + q_i^- \geq 0$ with probability one, then
$$Q(x,\omega) = \sum_{i=1}^{m_2}(q_i^+(\omega)(h_i(\omega) - T_{i.}(\omega)x)^+ + q_i^-(\omega)(-h_i(\omega) + T_{i.}(\omega)x)^+)$$
which is finite for all ω. Using Proposition 2, we obtain the result.

□

Assumption: $q_i^+ + q_i^- \geq 0$ with probability one.

- $\mathcal{Q}(x) = \sum_{i=1}^{m_2} \mathcal{Q}_i(x)$,
- $\mathcal{Q}_i(x) = E_\omega[Q_i(x, \xi(\omega))]$,
- $Q_i(x, \xi(\omega)) = q_i^+(\omega)(h_i(\omega) - T_{i.}(\omega)x)^+ + q_i^-(\omega)(-h_i(\omega) + T_{i.}(\omega)x)^+$.
- When q and T are fixed, this characterization of $\mathcal{Q}$ allows its expression as a separable function in the remaining random components h_i.
- Often, in this case, $T_{i.}x$ is substituted with χ_i and Ψ is substituted for ξ so that $\mathcal{Q}(x) = \Psi(\chi)$.
- We then obtain $\Psi(\chi) = \sum_{i=1}^{m_2} \Psi_i(\chi_i)$ where $\Psi_i(\chi_i) = E_{h_i}[\psi_i(\chi_i, h_i)]$ and $\psi_i(\chi_i, h_i) = q_i^+(h_i - \chi_i)^+ + q_i^-(-h_i + \chi_i)^+$.
- We, however, continue to use $\mathcal{Q}(x)$ to maintain consistency with our previous results.

- We can define the objective function even further.
- In this case, let h_i have an associated distribution function F_i, mean value $\overline{h}_i$, and let $q_i = q_i^+ + q_i^-$.
- We can then write

$$\mathcal{Q}_i(x) = q_i^+\overline{h}_i - (q_i^+ - q_i F_i(T_{i.}x))T_{i.}x - q_i \int_{h_i \leq T_{i.}x} h_i dF_i(h_i). \quad (9)$$

- Of particular importance in optimization is the subdifferential of this function, which has the following simple form:

$$\partial\mathcal{Q}_i(x) = \{\pi(T_{i.})^T \mid -q_i^+ + q_i F_i^-(T_{i.}x) \leq \pi \leq -q_i^+ + q_i F_i^+(T_{i.}x)\}, \quad (10)$$

- $F_i^-(h) = \lim t \uparrow hF_i(t)$ and $F_i^+(h) = \lim_{t\downarrow h} F_i(t) = F_i(h)$.

Optimality conditions for stochastic programs

- Special conditions that can apply to stochastic programs
- Stochastic programs may differ from other mathematical programs.
- Additional assumptions that guarantee necessary and sufficient conditions for two-stage stochastic linear programs.

1. When is a solution to (2) attainable?
2. What form do the optimality conditions take and how can they be simplified?
3. What types of dual problems can be formulated to accompany (2) and do they obtain bounds on optimal values?
4. How stable is an optimal solution to (2) to changes in the parameters and distributions?

- Assumptions: ξ has finite second moments, and $\mathcal{Q}$ is Lipschitzian.
- Then, we can apply a direct subgradient result.
- Question: Whether the solution of (2) can indeed be obtained, i.e., whether the optimal objective value is finite and attained by some value of x.

Example

Find

$$\inf\{E_\xi[y+(\xi)] | y^+(\xi), y^-(\xi) \geq 0, x + y^+(\xi) - y^-(\xi) = \xi, a.s.\}, \qquad (11)$$

where ξ is, for example, negative exponentially distributed on $[0, \infty)$. For any finite value of x, (11) has a positive value, but the infimum over x is zero.

almost surely (denoted a.s.), i.e., for all $\omega \in \Omega$ except perhaps for sets with zero probability.

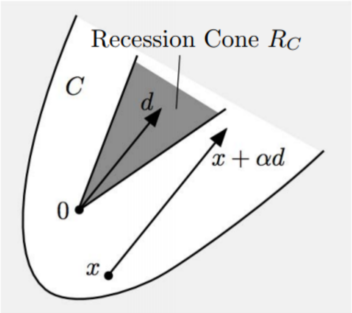
Recession cone, Recession value

definition

The recession cone (rc),

$$\{v|u+\lambda v \in S, \forall\lambda \geq 0 \text{ and } u \in S\}$$

when applied to a set, S, and the recession value, $sup_{x\in \text{dom } f}(f(x+v)-f(x))$when applied to a proper convex function, f.



Sufficient conditions to guarantee that a solution to (2) exists

Theorem

Suppose that the random elements ξ have finite second moments and one of the following:
(a) the feasible region K is bounded; or
(b) the recourse function $\mathcal{Q}$ is eventually linear in all recession directions of K, i.e., $\mathcal{Q}(x+\lambda v) = \mathcal{Q}(x+\overline{\lambda}v) + (\lambda - \overline{\lambda})rc\mathcal{Q}(v)$ for some $\overline{\lambda} \geq 0$ (dependent on x), all $\lambda \geq \overline{\lambda}$, and some constant recession value, $rc\mathcal{Q}(v)$, for all v such that $x + \lambda v \in K$ for all $x \in K$ and $\lambda \geq 0$.
Then, if problem (2) has a finite optimal value, it is attained for some $x \in \mathbb{R}^n$.

Note

As shown in Wets [1974], if T is fixed and Ξ is compact, the condition in (b) is obtained.

- Assumption: An optimal solution can be attained (we would expect in most practical situations).
- The general deterministic equivalent form gives us the following result in terms of Karush-Kuhn-Tucker conditions.

Theorem

Suppose (2) has a finite optimal value. A solution $x^* \in K_1$, is optimal in (2) if and only if there exists some $\lambda^* \in \mathbb{R}^{m_1}, \mu^* \in \mathbb{R}_+^{n_1}, \mu^* T x^* = 0$, such that,

$$-c + A^T \lambda^* + \mu^* \in \partial \mathcal{Q}(x^*). \quad (12)$$

This result can be combined with our previous results on simple recourse functions to obtain specific conditions for that problem as follows.

Corollary

Suppose (1) has simple recourse and a finite optimal value. Then $x^* \in K_1$ is optimal in (2) corresponding to this problem if and only if there exists some $\lambda^* \in \mathbb{R}^{m_1}$, $\mu^* \in \mathbb{R}^{n_1}_{+}$, ${\mu^*}^T x^* = 0, \pi_i^*$ such that $-(q_i^+ - q_i F_i^-(T_{i.}x^*)) \le \pi_i^* \le -(q_i^+ - q_i F_i^+(T_{i.}x^*))$ and

$$-c + A^T \lambda^* + \mu^* - (\pi^*)^T T = 0. \tag{13}$$

- Inclusion (12) suggests that a subgradient method or other nondifferentiable optimization procedure may be used to solve (2).
- Finite realizations of the random vector lead to equivalent linear programs (although of large scale),
- Absolutely continuous distributions lead to a differentiable recourse function $\mathcal{Q}$.
- if $\mathcal{Q}$ is differentiable, we can replace $\partial\mathcal{Q}(x^*)$ with $\nabla\mathcal{Q}(x^*)$ to obtain:

$$c + \nabla\mathcal{Q}(x^*) = A^T\lambda^* + \mu^* \tag{14}$$

- Possible algorithms based on convex minimization subject to linear constraints are then admissible.

- The main practical possibilities for solutions of (2): either large-scale linear programming or smooth nonlinear optimization.
- The chief difficulty is in characterizing $\partial \mathcal{Q}$ because even evaluating this function is difficult.
- This evaluation is decomposable into subgradients of the recourse function for each realization of ξ , which form the subdifferential set $\partial Q(x, \xi(\omega))$, where we interpret the subgradient elements as being defined with respect to the decision variables x .

Theorem

If $x \in K$, then

$$\partial \mathcal{Q}(x) = E_\omega \partial Q(x, \xi(\omega)) + N(K_2, x), \tag{15}$$

where $N(K_2, x) = \{v | v^T y \leq 0, \forall y \text{ such that } x + y \in K_2\}$, the normal cone to K_2 at x .

If the problem has relatively complete recourse, then, for any y such that $x + y \in K_1$, we must also have $x + y \in K_2$. Hence, $N(K_2, x) \subset N(K_1, x) = \{v | v = A^T \lambda + \mu, \mu T x = 0, \mu \geq 0\}$.

Corollary

If (2) has relatively complete recourse, a solution x^* is optimal in (2) if and only if there exists some $\lambda^* \mathbb{R}^{m_1}, \mu^* \mathbb{R}^{n_1}_+, {\mu^*}^T x^* = 0$, such that

$$-c + A^T \lambda^* + \mu^* E_\omega \partial Q(x, \xi(\omega)). \tag{16}$$

(Exercise 10):

$$\begin{aligned} E_\omega \partial Q(x, \xi(\omega)) &= \{-E[\pi \mathbf{T}] | \pi^T W \leq \mathbf{q}^T, \\ &\pi^T(\mathbf{h} - \mathbf{T}x) \geq (\pi')^T(h - Tx), \forall (\pi')^T W \leq \mathbf{q}^T a.s.\} \end{aligned} \quad (17)$$

An equivalent dual program to (2)

Under the relatively complete recourse assumption can be obtained (Exercise 11) by solving the following maximization problem:

$$\begin{aligned} \max v = \quad & b^T \lambda + E_\omega[h(\omega)^T \pi(\omega)] \\ s.t. \quad & A^T \lambda + E_\omega[T(\omega)^T \pi(\omega)] \leq c, \\ & W^T \pi(\omega) \leq q(\omega), \quad a.s. \end{aligned} \quad (18)$$

Definition

The optimal solution set is stable, if it changes continuously in some sense when parameters of the problem change continuously.

Main Result

Stability is achieved (i.e., some optimal solution of an original problem is close to some optimal solution of a perturbed problem) if problem (2) has complete recourse and the set of recourse problem dual solutions, $\{\pi | \pi^T W \leq q(\omega)^T\}$, is nonempty with probability one.

In general, we wish to have a different x, y pair for every realization of the random outcomes. We then wish to restrict the x decisions to be the same for almost all outcomes.

- This says that the decision, $(x(\omega), y(\omega))$, is a function (with suitable properties) on Ω.
- We restrict this to some space, X, of measurable functions on Ω, for example, the p-integrable functions, $\mathcal{L}_p(\Omega, \mathcal{B}, \mu; \mathbb{R}^n)$, for some $1 \leq p \leq \infty$.
- The general version of (2) is

$$\inf_{(x(\omega),y(\omega))\in X} \int_{\Omega} (c^T x(\omega) + q(\omega)^T y(\omega))\mu(d\omega)$$

$$\begin{array}{llcll} s.t. & Ax(\omega) & = & b, & a.s., \\ & E_{\Omega}(x(\omega)) - x(\omega) & = & 0, & a.s., \\ & T(\omega)x(\omega) + Wy(\omega) & = & h(\omega), & a.s., \\ & x(\omega), y(\omega) & \geq & 0, & a.s. \end{array} \quad (19)$$

- Problem (19) is equivalent to (2) if, for example, X is the space of essentially bounded functions on Ω and K is bounded for (2).
- The two formulations are not necessarily the same.
- The only difference in optimality conditions of (19) from those of (12) is that we include explicit multipliers for the nonanticipativity constraints.
- For continuous distributions, these multipliers may have a difficult representation unless (19) has relatively complete recourse.
- The difficulty is that we cannot guarantee boundedness of the multipliers and may not be able to obtain an integrable function to represent them.
- This difficulty is caused when future constraints restrict the set of feasible solutions at the first stage.

- For finite distributions, (19) is an implementable problem structure that is used in several algorithms.
- In this case, with K possible realizations of ξ with probabilities $p^k, k = 1, \dots, K$, the problem becomes:

$$\inf_{\substack{(x^h, y^k),\\ k=1,\dots,K}} \sum_{k=1}^{K} p^k (c^T x^k + (q^k)^T y^k)$$

$$\begin{array}{lrcll} s.t. & Ax^k & = & b, & k = 1, \dots, K, \\ & \sum_{j \neq k} p^j x^j + (1 - p^j) x^k & = & 0, & k = 1, \dots, K, \\ & T^k x^k + W y^k & = & h^k, & k = 1, \dots, K, \\ & x^k, y^k & \geq & 0, & k = 1, \dots, K \end{array} \tag{20}$$

- Problem (20) is almost completely decomposes into K separate problems for the K realizations.
- The only links are in the second set of constraints that impose nonanticipativity.

- Probabilistic, or chance, constraints take the form:

$$P\{A^i(\omega)x \geq h^i(\omega)\} \geq \alpha^i \tag{21}$$

- $0 < \alpha^i < 1$ and $i = 1, \ldots, I$ is an index of the constraints that must hold jointly.
- We can model these constraints in a general expectational form $E_\omega(f^i(\omega, x(\omega))) \geq \alpha^i$ where f^i is an indicator of $\{\omega | Ai(\omega)x \geq hi(\omega)\}$.
- The objective is often an expectational functional (the E-model), or the variance of some result (the V-model) or the probability of some occurrence (such as satisfying the constraints) (the P-model).
- Another variation includes an objective that is a quantile of a random function

- The goal is to determine deterministic equivalents and their properties.
- To maintain consistency, let

$$K_1^i(\alpha^i) = \{x | P(A^i(\omega)x \geq hi(\omega)) \geq \alpha^i\}, \qquad (22)$$

- $0 < \alpha^i \leq 1$ and $\bigcap_i K_1^i(1) = K_1$.
- $K_1^i(\alpha^i)$ need not be convex or even connected.

Example

- $\Omega = \{\omega_1, \omega_2\}, P[\omega_1] = P[\omega_2] = \frac{1}{2}$

$$A^i(\omega_1) = A^i(\omega_2) = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, h^i(\omega_1) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}, h^i(\omega_2) = \begin{pmatrix} 2 \\ -3 \end{pmatrix}, \quad (23)$$

- For $0 < \alpha^i \leq \frac{1}{2}, , K_1^i(\alpha^i) = [0, 1] \cup [2, 3]$.

- When each i corresponds to a distinct linear constraint and A^i is a fixed row vector, obtaining a deterministic equivalent of (22) is fairly straightforward.
- In this case, $P(A^i x \geq h^i(\omega)) = F^i(A^i x)$, where F^i is the distribution function of h^i.
- Hence, $K_1^i(\alpha^i) = \{x | F^i(A^i x) \geq \alpha^i\}$, which immediately yields a deterministic equivalent form.
- In general, the constraints must hold jointly so that the set I is a singleton.
- This situation corresponds to requiring an α-confidence interval that x is feasible.

- One of the main results in probabilistic constraints is that, in the joint constraint case, a large class of probability measures on $h(\omega)$ (for A fixed) leads to convex and closed $K_1(\alpha)$.

Definition

A probability measure P is in this class of quasi-concave measures if for any convex measurable sets U and V and any $0 \leq \lambda \leq 1$,

$$P((1-\lambda)U + \lambda V) \geq \min\{P(U), P(V)\}. \quad (24)$$

Theorem

Suppose A is fixed and h has an associated quasi-concave probability measure P . Then $K_1(\alpha)$ is a closed convex set for $0 \le \alpha \le 1$.

Theorem

If f is the density of a continuous probability distribution in $\mathbb{R}^m$ and $f^{-\frac{1}{m}}$ is convex on $\mathbb{R}^m$, then the probability measure

$$P(B) = \int_B f(x)dx,$$

defined for all Borel sets B in $\mathbb{R}^m$ is quasi-concave.

- This result states that any density of the form $f(x) = e^{-l(x)}$ for some convex function l yields a quasi-concave probability measure.
- These measures includethe multivariate normal, beta, and Dirichlet distributions and are logarithmically concave (because, for

$$0 \leq \lambda \leq 1, P((1-\lambda)U + \lambda V) \geq P(U)^{\lambda} P(V)^{1-\lambda}$$

for all Borel sets U and V).
- These distributions lead to computable deterministic equivalents as, for example, in the following theorem.

Theorem

Suppose A is fixed and the components $h_i, i = 1, \dots, m_1$, of h are stochastically independent random variables with logarithmically concave probability measures, P_i , and distribution functions, F_i, then $K_1(\alpha) = \{x | \sum_{i=1}^{m_1} \ln(F^i(A^{i\cdot}x)) \geq \ln \alpha\}$ and is convex.

Theorem

If $A_{1.}, \dots, A_{n_1.}$, h have a joint normal distribution with a common covariance structure, a matrix C , such that $E[(A_{i.} - E(A_{i.}))(A_{j.} - E(A_{j.}))^T] = r_{ij}C$ for i, j in $1, \dots, n_1$, and $E[(A_{i.} - E(A_{i.}))(h - E(h))] = s_iC$ for $i = 1, \dots, n-1$, where r_{ij} and s_i are constants for all i and j , then $K_1(\alpha)$ is convex for $\alpha \geq \frac{1}{2}$.

Theorem

Suppose that $m_1 = 1, h_1 = 0$, and $A_{1.}$ has mean $\overline{A}_{1.}$ and covariance matrix C_1, then $K_1(\alpha) = \{x | \overline{A}_{1.} - \phi^{-1}(\alpha)\sqrt{x^T C_1 x} \geq 0\}$, where ϕ is the standard normal distribution function.

$$\begin{aligned} \min \quad & c^T x \\ s.t. \quad & Ax = b, \\ & P_i[T_{i.}x \geq h_i] \geq \alpha_i, i = 1, \ldots, m_2, \\ & x \geq 0, \end{aligned} \tag{25}$$

where P_i is the probability measure of h_i and F_i is the distribution function for h_i.

- For the deterministic equivalent to (25), we just let $F_i(h_i^*) = \alpha_i$

$$\begin{aligned} \min \quad & c^T x \\ s.t. \quad & Ax = b, \\ & T_{i.}x \geq h_i^*, i = 1, \ldots, m_2, \\ & x \geq 0, \end{aligned} \tag{26}$$

- Suppose we solve (26) and obtain an optimal x^* and optimal dual solution $\{\lambda^*, \pi^*\}$, where $c^T x^* = b^T \lambda^* + {h^*}^T \pi^*$.
- If $\pi_i^* = 0$, let $q_i^+ = 0$ and, if $\pi_i^* > 0$, let $q_i^+ = \frac{\pi_i^*}{1-\alpha_i}$.

- An equivalent stochastic program with simple recourse to (25)

$$
\begin{aligned}
\min \quad & c^T x + E_{\mathbf{h}}[q^+ y^+] \\
s.t. \quad & Ax = b, \\
& T_{i.}x + y_i^+ - y_i^- = \mathbf{h}_i^*, i = 1, \ldots, m_2, \\
& x, y_i^+, y_i^- \geq 0,
\end{aligned}
\tag{27}
$$

- For problems (25) and (27) to be equivalent, we mean that any x^* optimal in (25) corresponds to some (x^*, y^{*+}) optimal in (27) for a suitable definition of q^+ and that any (x^*, y^{*+}) optimal in (27) corresponds to x^* optimal in (25)for a suitable definition of α_i .

Theorem

For the q_i^+ defined as a function of some optimal π^* for the dual to (25), if x^* is optimal in (25), there exists $y^{*+} \geq 0$ a.s. such that (x^*, y^{*+}) is optimal in (25).

- If $\xi(\omega)$ is a discrete random variable, there exists a finite number of scenarios which correspond to the realizations of ξ . They are represented as $\xi_1, \xi_2, \ldots, \xi_K$. Scenario k has a probability p_k with

$$\sum_{k=1}^{K} p_k = 1.$$

- Scenarios can be obtained through experts opinions.
- Another typical way to get scenarios is when the information over the random variables comes from historical data.
- The distribution of the random vector is then known as the empirical distribution.

- Assume we have a constraint of the form

$$P\{g(x, y(\omega), \xi(\omega)) \leq 0\} \geq \alpha. \tag{28}$$

- It is a joint probabilistic constraint:
  - $g(.) \leq 0$ may contain several constraints under a vector representation.
  - This includes classical cases such as $g(x, y(\omega), \xi(\omega)) = h(\omega) - Ax$.
  - This also includes cases where the probabilistic constraint depends on the recourse actions.

$$g(x, y(\omega), \xi(\omega)) = h(\omega) - T(\omega)x - W(\omega)y(\omega).$$

Definition

Indicator function $\eta(a) = 0$ if $a \leq 0$ and 1 if at least one component of a is strictly positive.

- The probabilistic constraint is equivalent to

$$\sum_{k=1}^{K} p_k \eta(g(x, y_k, \xi_k)) \leq 1 - \alpha. \tag{29}$$

- The left-hand side of (29) sums up the probability of the scenarios for which $g(.) \leq 0$ is violated.
- Assume that for each scenario k , an upper bound vector u_k can be found such that $g(x, y_k, \xi_k) \leq u_k$ for all feasible x, y_k .
- Then, (29) can be transformed into

$$\begin{aligned} \sum_{k=1}^{K} p_k w_k &\leq 1 - \alpha, && (30) \\ g(x, y_k, \xi_k) &\leq u_k w_k, k = 1, \ldots, K, && (31) \\ w_k \in \{0, 1\} & \quad , k = 1, \ldots, K. && (32) \end{aligned}$$

- The binary variable w_k plays the role of the indicator function.
  - When $g(x, y_k, \xi_k) \leq 0$, w_k takes the value 0 .
  - When at least one component of $g(x, y_k, \xi_k)$ is strictly positive, then $w_k = 1$ and scenario k contributes p_k to the left-hand side in (30).
- The joint probabilistic constraint (28) with a discrete random variable is transformed into a mixed integer programming (MIP) formulation.
- When $g(.)$ is linear, the stochastic program with probabilistic constraint is transformed into a mixed integer linear program (MILP) and can be solved using your favorite MILP solver.

Example

Find the numbers x_1 and x_2 of seats in first and business class for a plane of 200 seats. Assume a joint probabilistic constraint

$$P(x_1 \geq \xi_F, x_1 + x_2 \geq \xi_F + \xi_B) \geq 0.95, \tag{33}$$

where ξ_F and ξ_B represents the weekdays demands in first and business class. This corresponds to the classical case where $g(x, y(\omega), \xi(\omega)) = h(\omega) - Ax$, with $h(\omega)T = (\xi_F, \xi_F + \xi_B)$ and $A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$

- Assume the random variables (ξ_F, ξ_B) are given by the empirical data of last year. (These data must correspond to the number of calls and not to the number of passengers, which may depend on the acceptance policy at that time).
- This creates an empirical distribution of 260 pairs (ξ_F, ξ_B) for each weekday of last year. Each of the 260 pairs is a scenario of probability $\frac{1}{260}$.

Example

Continued...

- we need an upper bound on $\xi_F - x_1$ and on $\xi_F + \xi_B - x_1 - x_2$ for each k .
- Here, it suffices to take ξ_F and $\xi_F + \xi_B$, respectively.
- As an illustration, if scenario k has demands (14,32) in first and business, then the two corresponding constraints in (31) are

$$
\begin{aligned}
14 - x_1 &\leq 14w_k, \\
46 - x_1 - x_2 &\leq 46w_k
\end{aligned}
$$

- Thus, (33) is formulated using 260 binary variables w_k s, one constraint (30) and 520 constraints in (31).
- To put it in more general terms, (33) is reformulated using K extra binary variables and $2K + 1$ extra constraints.

Example

Consider the farmer problem.

- The example was built assuming a discrete random variable with only three scenarios: good, fair, and bad.
- This number can easily be extended either in a similar manner or by taking past observations of the yields.
- We now assume K scenarios, each consisting of a vector of three yields.
- The farmer finds it inappropriate to purchase large quantities of wheat and/or corn.
- He considers it excessive to purchase more than a total of 20 T. Owing to the uncertainty of mother nature, he allows for a 20% probability of excessive purchases.

Example

Continued...

- Thus, his probabilistic constraint is

$$P(y_1(\omega) + y_2(\omega) \leq 20) \geq 0.80 \qquad (34)$$

- $y_1(\omega)$ and $y_2(\omega)$ are the purchases of wheat and corn, respectively.
- Here is a case where the probabilistic constraint depends on the recourse actions under the general form $g(x, y(\omega), \xi(\omega)) = h(\omega) - T(\omega)x - W(\omega)y(\omega)$.
- To obtain (31), start from the representation of the constraint under scenario k as $-20 + y_1^k + y_2^k \leq 0$,where y_1^k and y_2^k represent the purchase of wheat and corn under scenario k .

Example

Continued...

- From Table 1 in Section 1.1, the total requirement of wheat and corn is 440 .
- The upper bound to form (31) is the value 420 , so that a single constraint of the form

$$y_1^k + y_2^k \leq 20 + 420 w_k \tag{35}$$

(If $y_1^k + y_2^k \leq 20$, then w_k is 0 ; otherwise, $w_k = 1$ and the constraint imposes no limit on the purchase of wheat and corn as the total cannot exceed 440 .)

- The recourse problem with K scenarios and the extra probabilistic constraint (34) is reformulated as an MILP with K extra binary variables and $K + 1$ extra constraints.

- For large values of K , the MILP may become difficult to solve.
- This is due to the structure of (31).
- It is indeed a weak constraint on w_k .
- To see this, consider the example of (35).
- Suppose that the total purchase under scenario k is 30 .
- Then (35) is equivalent to $420w_k \geq 10$, or $w_k \geq 0.0238$.
- As (35) is the only constraint on $w - k$, integrality can only be recovered through branching.
- The MILP solver will have to branch on all nonzero binaries, and none of them is likely to be spontaneously 1 .
- Moreover, after some w_k s are fixed by branching, additional w_ks may become fractional and require extra branching.

Definition

A valid inequality is a linear constraint added to the original formulation, which does not eliminate any integer solution but eliminates fractional solutions.

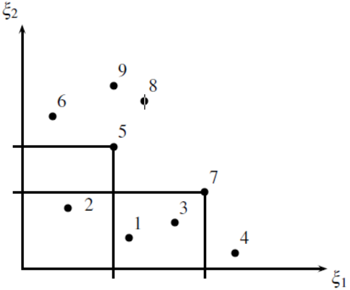
- A valid inequality provides a reformulation of the problem that contains fewer fractional solutions but the same integer solutions.
- To illustrate valid inequalities, we use the example of constraint (34) and its reformulation (35).
- As the probabilistic constraint only depends on corn and wheat, we may restrict our attention for this analysis to the first two components of the random vector.

Definition

Scenario k dominates scenario j if $\xi_k \geq \xi_j$, where the inequality must hold componentwise.

- In the current farmer example, if scenario k dominates scenario j , the yields of wheat and corn are higher in scenario k .
- It follows that the purchases of both products can only be smaller under scenario k .
- Hence, $w_k \leq w_j$.
- A first set of potential valid inequalities is $w_k \leq w_j$ for all pairs of scenarios such that $\xi_k \geq \xi_j$.

- We now illustrate the valid inequalities in the farmer problem with the extra probabilistic constraint (34)
- Imagine the farmer is able to collect 25 scenarios, each having probability 0.04 . (He may obtain them in a cooperative fashion with some fellow farmers or get them from an agricultural research institute.)
- Assume that the first 9 scenarios (restricted to wheats and corns yields) are as follows: (2.25,2.4) , (2.1,2.6) , (2.4,2.5) , (2.6,2.3) , (2.2,3) , (2,3.4) , (2.5,2.7) , (2.3,3.6) , (2.2,3.7) .
- Assume also that, for all other scenarios, $P(A_k) > 0.8$; hence, $w_k = 0$.



There are several dominance relations: $\xi_3 \geq \xi_1, \xi_5 \geq \xi_2, \xi_7 \geq \xi_2, \xi_7 \geq \xi_3, \xi_8 \geq \xi_1, \xi_8 \geq \xi_5, \xi_8 \geq \xi_6, \xi_9 \geq \xi_5, \xi_9 \geq \xi_6$, implying valid inequalities $w_3 \leq w_1, w_5 \leq w_2, w_7 \leq w_2, w_7 \leq w_3, w_8 \leq w_1, w_8 \leq w_5, w_8 \leq w_6, w_9 \leq w_5, w_9 \leq w_6$.

- Dominance sets A_k can be visualized by drawing an horizontal and a vertical half-line from k .
- A_5 and A_7 are illustrated in the Figure 1.
- $A_5 = \{2, 5\}$ with $P(A_5) = 0.08$ and $A_7 = \{1, 2, 3, 7\}$ with $P(A_7) = 0.16$.
- Even if $P(A_5) + P(A_7) > 0.2$, Scenarios 5 and 7 do not constitute a cover as $P(A_5 \cup A_7) = 0.2$.
- Scenarios 3 and 9 have similar probabilities and constitute a cover: $A_3 = \{1, 3\}$ with $P(A_3) = 0.08$, $A_9 = \{2, 5, 6, 9\}$ with $P(A_9) = 0.16$ and $P(A_3 \cup A_9) = 0.24$.
- Thus $w_3 + w_9 \leq 1$ is a valid inequality.

- This example shows that covers based on the dominance sets A_k are difficult to find as probabilities do not sum over sets that may intersect.
- Only minimal covers are of interest.
- As an example, $\{1, 3, 9\}$ is a cover but it is not minimal as removing $\{1\}$ still forms a cover.
- There are several other minimal covers in this example: $\{1, 4, 9\}, \{3, 4, 5, 6\}, \{3, 8\}, \{4, 5, 7\}, \{4, 6, 7\}, \{4, 8\}$,$\{7, 8\}, \{7, 9\}, \{8, 9\}$.
- In general, the MILP only adds minimal covers if they are violated by the current fractional point.