Multistage Stochastic Programs: Approximations, Bounds and Time Consistency

Georg Pflug
University of Vienna

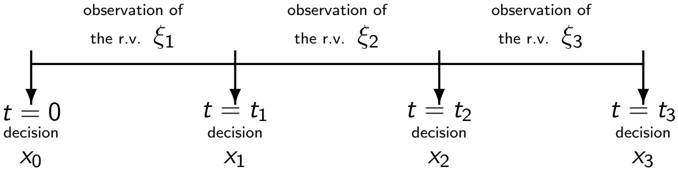
Multistage stochastic optimization problems

Many real decision problems under uncertainty involve several decision stages:

- hydropower storage and generation management
- thermal electricity generation
- portfolio management
- logistics
- asset/liabilty management in insurance

At each time $t = 0, 1, \dots, T-1$ a decision x_t can/must be made. We call the sequence $x = (x_0, x_1, \dots, x_{T-1})$ a *strategy*. The costs of the strategy x is expressed in terms of a cost function, which depends also on some random parameters (the scenario process) $\xi = (\xi_1, \dots, \xi_T)$ defined on some probability space $(\Omega, \mathcal{F}, P)$

$$Q(x_0, \xi_1, x_1, \dots, x_{T-1}, \xi_T).$$



observation of the r.v. ξ_1
observation of the r.v. ξ_2
observation of the r.v. ξ_3

$t = 0$ decision x_0
$t = t_1$ decision x_1
$t = t_2$ decision x_2
$t = t_3$ decision x_3

Decisions can only be made on the basis of the available information. For this reason, we assume that a filtration $\mathfrak{F} = (\mathcal{F}_1, \dots, \mathcal{F}_T = \mathcal{F})$ is defined in $(\Omega, \mathcal{F}, P)$ such that $\xi_t \lhd \mathcal{F}_t$ (ξ_t is measurable w.r.t. $\mathcal{F}_t$).

The Decision Problem

The final objective is to minimize a functional $\mathcal{R}$ of the stochastic cost function, such as the expectation, a quantile or some other functional $\mathcal{R}$

$$(Opt) \left\| \begin{array}{l} \text{Minimize in } x_0, x_1(\xi_1), \ldots, x_{T-1}(\xi_1, \ldots, \xi_{T-1}) : \\ \mathcal{R}[Q(x_0, \xi_1, \ldots, x_{T-1}, \xi_T)] \\ \ \text{s.t. } x \lhd \mathfrak{F} \\ \ \text{and possibly other constraints on } x_0, \ldots, x_{T-1} : x \in \mathbb{X} \end{array} \right.$$

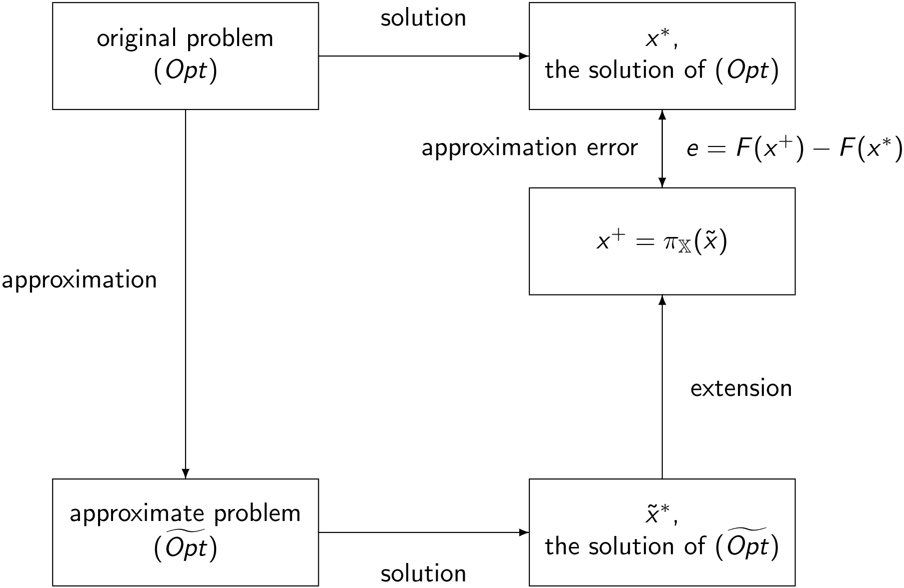
$x \lhd \mathfrak{F}$ means that $x_t \lhd \mathcal{F}_t$, i.e. that the decisions are *nonanticipative*.

Approximations

In order to numerically solve the multiperiod stochastic optimization problem, the stochastic process (ξ_t) must be approximated by a simple stochastic process $\tilde{\xi}_t$, which takes only a small number of values. Likewise the filtration $\mathfrak{F}$ must be approximated by a smaller one $\tilde{\mathfrak{F}}$ such that $\sigma(\tilde{\xi}) \subseteq \tilde{\mathfrak{F}}$.

$$\tilde{F}(\tilde{x}_1, \ldots, \tilde{x}_{T-1}) = \mathcal{R}[Q(\tilde{x}_0, \tilde{\xi}_1, \tilde{x}_1, \ldots, \tilde{x}_{T-1}, \tilde{\xi}_T)]$$

$$(\widetilde{Opt}) \left\| \begin{array}{l} \text{Minimize in } \tilde{x}_0, x_1(\tilde{\xi}_1), \ldots, \tilde{x}_{T-1}(\tilde{\xi}_1, \ldots, \tilde{\xi}_{T-1}) : \\ \mathcal{R}[Q(\tilde{x}_0, \tilde{\xi}_1, \ldots, \tilde{x}_{T-1}, \tilde{\xi}_T)] \\ \text{ s.t. } \tilde{x} \lhd \tilde{\mathfrak{F}} \\ \text{ and possibly other constraints} \tilde{x} \in \tilde{\mathbb{X}}. \end{array} \right.$$



original problem
(Opt)

solution

x^*,
the solution of (Opt)

approximation error

$e = F(x^+) - F(x^*)$

$x^+ = \pi_{\mathbb{X}}(\tilde{x})$

approximation

extension

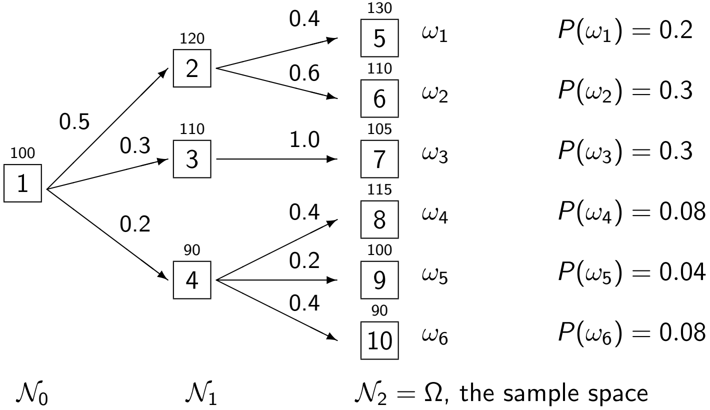
approximate problem
$(\widetilde{Opt})$

solution

$\tilde{x}^*$,
the solution of $(\widetilde{Opt})$

A valuated tree

Scenario trees are valuated trees: The nodes are valuated with the scenario process values, the arcs are valuated with the conditional probabilities.



An exemplary finite tree process $\nu = (\nu_0, \nu_1, \nu_2)$ with nodes $\mathcal{N} = \{1, \ldots 10\}$ and leaves $\mathcal{N}_2 = \{5, \ldots 10\}$ at $T = 2$ stages. The

Distances for Multistage Stochastic Optimization

The Kantorovich/Wasserstein distance.

Let $L(h)$ be the Lipschitz constant of the function h:

$$L(h) = \sup\{ \frac{|h(u) - h(v)|}{d(u,v)} : u \neq v \}.$$

The Kantorovich distance.

$$\mathrm{d}_1(P, \tilde{P}) = \sup\{ \int h \, dP - \int \, h d\tilde{P} : L(h) \leq 1 \}.$$

Theorem (Kantorovich-Rubinstein). Dual version of Kantorovich-distance:

$$\begin{aligned} \mathrm{d}_1(P, \tilde{P}) \quad = \quad & \inf\{\mathbb{E}(\mathrm{d}(X,Y) : (X,Y) \text{ is a bivariate r.v. with} \\ & \text{given marginal distributions } P \text{ and } \tilde{P}\}. \end{aligned}$$

Generalization: The Wasserstein-distance of order r

$$\begin{aligned} \mathrm{d}_r(P, \tilde{P}) \quad = \quad & \inf\{ \left(\int \mathrm{d}(u,v)^r \, d\pi(u,v) \right)^{1/r} : \pi \text{ is a probability distribution} \\ & \text{on } \Xi \times \tilde{\Xi} \text{ with given marginal distributions } P \text{ and } \tilde{P}\}. \end{aligned}$$

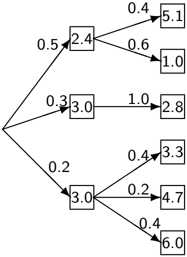
Closedness in Wasserstein distance implies closedness in various other aspects

Assume that $X \sim P$ and $\tilde{X} \sim \tilde{P}$. Then

1. $\left|\mathbb{E}|X|^p - \mathbb{E}|\tilde{X}|^p\right| \leq$
   $p \cdot \mathrm{d}_r\left(P, \tilde{P}\right) \cdot \max\left\{\mathbb{E}^{\frac{r-1}{r}}\left[|X|^{r\cdot\frac{p-1}{r-1}}\right], \mathbb{E}^{\frac{r-1}{r}}\left[|\tilde{X}|^{r\cdot\frac{p-1}{r-1}}\right]\right\}$,
2. $|\mathbb{E}(X^p) - \mathbb{E}(X^p)| \leq$
   $p \cdot \mathrm{d}_r\left(P, \tilde{P}\right) \cdot \max\left\{\mathbb{E}^{\frac{r-1}{r}}\left[|X|^{r\cdot\frac{p-1}{r-1}}\right], \mathbb{E}^{\frac{r-1}{r}}\left[|\tilde{X}|^{r\cdot\frac{p-1}{r-1}}\right]\right\}$ for p an integer,
3. $\left|\mathbb{E}X^2 - \mathbb{E}\tilde{X}^2\right| \leq 2 \cdot \mathrm{d}_2\left(P, \tilde{P}\right) \cdot \max\left\{\mathbb{E}^{\frac{1}{2}}\left[X^2\right], \mathbb{E}^{\frac{1}{2}}\left[\tilde{X}^2\right]\right\}$,
4. $\left|\mathbb{E}|X|^r - \mathbb{E}|\tilde{X}|^r\right| \leq r \cdot \mathrm{d}_r\left(P, \tilde{P}\right) \cdot \max\left\{\mathbb{E}^{\frac{r-1}{r}}\left[|X|^r\right], \mathbb{E}^{\frac{r-1}{r}}\left[|\tilde{X}|^r\right]\right\}$ and
5. $\left|\mathbb{E}|X|^p - \mathbb{E}|\tilde{X}|^p\right| \leq$
   $p \cdot \mathrm{d}_2\left(P, \tilde{P}\right) \cdot \max\left\{\mathbb{E}^{\frac{1}{2}}\left[|X|^{2(p-1)}\right], \mathbb{E}^{\frac{1}{2}}\left[|\tilde{X}|^{2(p-1)}\right]\right\}$,

where $p \geq 1$ and $r > 1$.

Trees are nested distributions



$$\mathbb{P} = \left[\begin{array}{ccc} 0.2 & 0.3 & 0.5 \\ \hline 3.0 & 3.0 & 2.4 \\ \left[\begin{array}{ccc} 0.4 & 0.2 & 0.4 \\ \hline 6.0 & 4.7 & 3.3 \end{array}\right] & \left[\begin{array}{c} 1.0 \\ \hline 2.8 \end{array}\right] & \left[\begin{array}{cc} 0.6 & 0.4 \\ \hline 1.0 & 5.1 \end{array}\right] \end{array}\right]$$

Distances between trees as nested distributions

Definition. For two nested distributions $\mathbb{P} \sim (\Xi, \mathcal{F}, P, \xi)$, $\tilde{\mathbb{P}} \sim \left(\tilde{\Xi}, \tilde{\mathcal{F}}, \tilde{P}, \tilde{\xi}\right)$ and a distance function d on $\mathbb{R}^m$ the *nested distance of order* $r \geq 1$ – denoted $\mathrm{dl}_r\left(\mathbb{P}, \tilde{\mathbb{P}}\right)$ – is the optimal value of the optimization problem

$$\begin{aligned} \underset{(\text{in } \pi)}{\text{minimize}} \quad & \left(\int d\left(\xi(\omega), \tilde{\xi}(\tilde{\omega})\right)^r \pi\left(\mathrm{d}\omega, \mathrm{d}\tilde{\omega}\right)\right)^{\frac{1}{r}} \\ \text{subject to} \quad & \pi\left(M \times \tilde{\Xi} \mid \mathcal{F}_t \otimes \tilde{\mathcal{F}}_t\right) = P\left(M \mid \mathcal{F}_t\right) \qquad \left(M \in \mathcal{F}_T\right) \\ & \pi\left(\Xi \times N \mid \mathcal{F}_t \otimes \tilde{\mathcal{F}}_t\right) = \tilde{P}\left(N \mid \tilde{\mathcal{F}}_t\right) \qquad \left(N \in \tilde{\mathcal{F}}_T\right) \end{aligned} \tag{1}$$

where the infimum in (1) is among all bivariate probability measures $\pi \in \mathcal{P}\left(\Omega \times \Omega'\right)$, which are measures on the product sigma algebra $\mathcal{F}_T \otimes \tilde{\mathcal{F}}_T$. We will refer to the nested distance also as *process distance*, or *multistage distance*. The nested distance dl_2 (order $r = 2$), with d a weighted Euclidean distance is referred to as *quadratic nested distance*.

How to calculate the nested distance

The nested distance between discrete trees can be calculated by solving the a linear program

$$\begin{array}{rl} \text{minimize} & \\ (\text{in } \pi) & \sum_{i,j} \pi_{i,j} \cdot d_{i,j}^r \\ \text{subject to} & \sum_{j \succ n} \pi\left(i, j \middle| m, n\right) = P\left(i \middle| m\right) \quad (m \prec i,\ n), \\ & \sum_{i \succ m} \pi\left(i, j \middle| m, n\right) = \tilde{P}\left(j \middle| n\right) \quad (n \prec j,\ m), \\ & \pi_{i,j} \geq 0 \text{ and } \sum_{i,j} \pi_{i,j} = 1, \end{array}$$

where again $\pi_{i,j}$ is a matrix defined on the leave nodes ($i \in \mathcal{N}_T$, $j \in \mathcal{N}'_T$) and $m \in \mathcal{N}_t$, $n \in \mathcal{N}'_t$ are arbitrary nodes. The conditional probabilities $\pi\left(i, j \middle| m, n\right)$ are given by

$$\pi\left(i, j \middle| m, n\right) = \frac{\pi_{i,j}}{\sum_{i' \succ m, j' \succ n} \pi_{i',j'}}.$$

The main approximation result

Let $\mathcal{Q}_L$ be the family of all real valued cost functions $Q(x_0, y_1, x_1, \dots, x_{T-1}, y_T)$, defined on $\mathbb{X}_0 \times \mathbb{R}^{n_1} \times \mathbb{X}_1 \times \cdots \times \mathbb{X}_{T-1} \times \mathbb{R}^{n_T}$ such that

- $x = (x_0, \dots, x_{T-1}) \mapsto Q(x_0, y_1, x_1, \dots, x_{T-1}, y_T)$ is convex for fixed $y = (y_1, \dots, y_T)$ and
- $y_t \mapsto Q(x_0, y_1, x_1, \dots, x_{t_1}, y_T)$ is Lipschitz with Lipschitz constant L for fixed x.

Consider the optimization problem $(Opt(\mathbb{P}))$

$$v_Q(\mathbb{P}) := \min\{\mathbb{E}_P[Q(x_0, \xi_1, x_1, \dots, x_{T-1}, \xi_T)] : x \lhd \mathfrak{F}, x \in \mathbb{X}\},$$

where $\mathbb{X}$ is a convex set and $\mathbb{P}$ is the nested distribution of the scenario process.
An approximative problem $(Opt(\tilde{\mathbb{P}}))$ is given by

$$v_Q(\tilde{\mathbb{P}}) := \min\{\mathbb{E}_{\tilde{P}}[Q(x_0, \tilde{\xi}_1, x_1, \dots, x_{T-1}, \tilde{\xi}_T)] : x \lhd \tilde{\mathfrak{F}}, x \in \mathbb{X}\},$$

where $\tilde{\mathbb{P}}$ is the nested distribution of the approximative scenario process.

Theorem. For Q in $\mathcal{Q}_L$

$$|v_Q(\mathbb{P}) - v_Q(\tilde{\mathbb{P}})| \leq L \cdot \mathsf{dl}(\mathbb{P}, \tilde{\mathbb{P}}).$$

Remarks.

- The bound is sharp: Let $\mathbb{P}$ and $\tilde{\mathbb{P}}$ be two nested distributions on $[\Xi, \mathsf{dl}]$. Then there exists a cost function $Q(\cdot) \in \mathcal{H}_1$ such that

  $$v_Q(\mathbb{P}) - v_Q(\tilde{\mathbb{P}}) = \mathsf{dl}(\mathbb{P}, \tilde{\mathbb{P}}).$$

- The inequality

  $$|v_Q(\mathbb{P}) - v_Q(\tilde{\mathbb{P}})| \leq L \cdot d(\mathbb{P}, \tilde{\mathbb{P}}),$$

  where d is the multivariate Kantorovich distance, does NOT hold.

Distortion functionals

Let G_Y be the distribution function of Y. Then the distortion functional $\mathcal{R}_\sigma$ with distortion density σ is defined as

$$\mathcal{R}_\sigma(Y) = \int_0^1 \sigma(u) G_Y^{-1}(u)\, du$$

A special example is the average value-at-risk, which has distortion density

$$\sigma_\alpha(u) = \begin{cases} 0 & u < \alpha \\ \frac{1}{1-\alpha} & u \geq \alpha \end{cases}$$

An extension of the main result

Theorem. Let $\mathcal{R}_\sigma$ be a distortion risk functional with bounded distortion, $\sigma \in L^\infty$.
Consider the optimization problem $(Opt(\mathbb{P}))$

$$v_{Q,\mathcal{R}_\sigma}(\mathbb{P}) := \min\{\mathcal{R}_{\sigma,\mathbb{P}}[Q(x_0, \xi_1, x_1, \ldots, x_{T-1}, \xi_T)] : x \lhd \mathfrak{F}, x \in \mathbb{X}\},$$

where $\mathbb{X}$ is a convex set and $\mathbb{P}$ is the nested distribution of the scenario process.
An approximative problem $(Opt(\tilde{\mathbb{P}}))$ is given by

$$v_{Q,\mathcal{R}}(\tilde{\mathbb{P}}) := \min\{\mathcal{R}_{\sigma,\tilde{\mathbb{P}}}[Q(x_0, \tilde{\xi}_1, x_1, \ldots, x_{T-1}, \tilde{\xi}_T)] : x \lhd \tilde{\mathfrak{F}}, x \in \mathbb{X}\},$$

where $\tilde{\mathbb{P}}$ is the nested distribution of the approximative scenario process.
Then

$$|v_{Q,\mathcal{R}_\sigma}(\mathbb{P}) - v_{Q,\mathcal{R}_\sigma}(\tilde{\mathbb{P}})| \leq L \cdot \|\sigma\|_\infty \cdot \mathsf{dl}_1\left(\mathbb{P}, \tilde{\mathbb{P}}\right).$$

Dynamic decomposability and Bellmann's principle

We now maximize an utility functional $\mathcal{U}$ of a profit variable.

$$\mathcal{U}(\text{Profit}) = -\mathcal{R}(-\text{Profit}) = -\mathcal{R}(\text{Loss}).$$

The standard multiperiod maximization problem is

$$\max\{\mathcal{U}[H(x_0, \xi_1, \ldots, x_{T-1}, \xi_T)] : x_t \lhd \mathcal{F}_t, x_t \in \mathbb{X}_t(x_{0:t-1}, \xi_{1:t})\} \qquad (2)$$

where $\mathcal{U}$ is an utility functional and H is a profit function. The problem is dynamically decomposable, if there exist functions H_t and functionals $\mathcal{U}_t$ such that (2) is equivalent to

$$\max_{x_0 \in \mathbb{X}_0} \left(H_0(x_0) + \max_{x_1 \in \mathbb{X}(x_0, \xi_1)} \mathcal{U}_1 \left(H_1(x_{0:1}, \xi_1) + \ldots \right.\right.$$
$$\left.\left. \ldots \max_{x_{T-1} \in \mathbb{X}(x_{0:T-2}, \xi_{1:T-1})} \mathcal{U}_{T-1}(H_{T-1}(x_{0:T-1}, \xi_{0:T})) \right) \right).$$

The time-consistency principle

If the optimal decision sequence is implemented, but only up to time t, and at time t the problem is resolved for the remaining times (keeping the past decisions fixed), then the optimal solution of this subproblem should coincide with that of the original problem.

Time decomposability for dynamic stochastic problems

If a stochastic problem is decomposable in time, then a Bellmann principle holds, the solution is time-consistent and can be found by backward induction.
If the probability functional is the expectation $\mathcal{U} = \mathbb{E}$ and the only measurability constraint is $x_t \lhd \mathcal{F}_t$, then time decomposability holds.
Time decomposability may not hold, if

- the functional is not the expectation
- other measurability conditions are in place, (e.g. $x_t \lhd \mathcal{F}_s$ for $s < t$).

Probability functionals

Let the random variable Y have distribution function $G_Y(u) = P\{Y \leq u\}$ and quantile function $\mathbb{V}@\mathrm{R}_p(Y) = \inf\{u : G_Y(u) \geq p\}$. We define

- the *Average Value-at-Risk* (measures acceptability or utility of profits)
  $\mathbb{A}\mathrm{V@R}(Y) = \frac{1}{\alpha}\int_0^\alpha \mathbb{V}@\mathrm{R}_p(Y)\,dp$
  $\mathbb{A}\mathrm{V@R}(Y) = \inf\{\mathbb{E}(Y\,Z) : 0 \leq Z \leq 1/\alpha; \mathbb{E}(Z) = 1\}$
- the *upper Average Value-at-Risk* (measures risk of costs)
  $\mathrm{U}\mathbb{A}\mathrm{V@R}(Y) = \frac{1}{1-\alpha}\int_\alpha^1 \mathbb{V}@\mathrm{R}_p(Y)\,dp$
- a distortion functional $\int_0^1 \mathbb{V}@\mathrm{R}_p(Y)h(p)\,dp$
- the entropic functional $\frac{-1}{\gamma}\mathbb{E}[\exp(-\gamma Y)]$

$$\mathbb{A}\mathrm{V@R}_0(Y) = \text{essinf }(Y) \qquad \mathbb{A}\mathrm{V@R}_1(Y) = \mathbb{E}(Y)$$

$$\mathrm{U}\mathbb{A}\mathrm{V@R}_0(Y) = \mathbb{E}(Y) \qquad \mathrm{U}\mathbb{A}\mathrm{V@R}_1(Y) = \text{esssup }(Y)$$

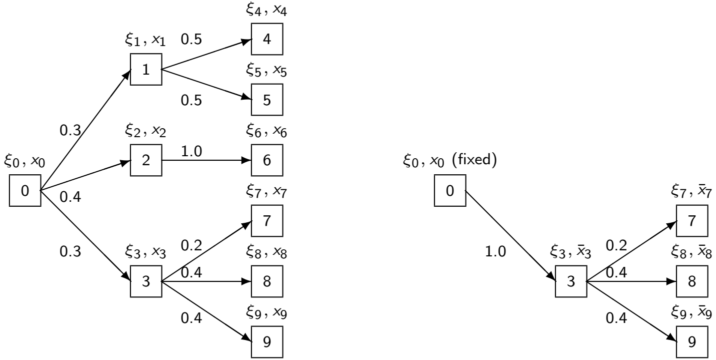
Conditional risk and utility (acceptability) functionals

We consider a probability space $(\Omega, \mathcal{F}, P)$. Let $\mathcal{F}_1$ be a σ-field contained in $\mathcal{F}$. A mapping $\mathcal{U}(\cdot|\mathcal{F}_1) : L_p(\mathcal{F}) \to L_{p'}(\mathcal{F}_1)$ is called *conditional utility mapping* (with observable information $\mathcal{F}_1$) if the following conditions are satisfied for all Y, $\lambda \in [0,1]$:

- **predictable translation-equivariance**. $\mathcal{U}(Y + Y_1|\mathcal{F}_1) = \mathcal{U}(Y|\mathcal{F}_1) + Y_1$, if $Y_1 \lhd \mathcal{F}_1$
- **concavity** $\mathcal{U}(\lambda Y + (1-\lambda)\tilde{Y}|\mathcal{F}_1) \geq \lambda \mathcal{U}(Y|\mathcal{F}_1) + (1-\lambda)\mathcal{U}(\tilde{Y}|\mathcal{F}_1)$,
- **monotonicity** $Y \leq \tilde{Y}$ implies $\mathcal{U}(Y|\mathcal{F}_1) \leq \mathcal{U}(\tilde{Y}|\mathcal{F}_1)$

The negative $\mathcal{R}(Y|\mathcal{F}_1) := -\mathcal{U}(Y|\mathcal{F}_1)$ is called a conditional risk functional.
Let $\mathcal{F}_0 = (\Omega, \emptyset)$ be the trivial σ-algebra. Then $\mathcal{U}(\cdot|\mathcal{F}_1)$ is an unconditional utility functional.

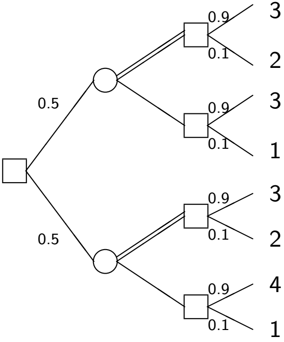
ξ_i: values of the scenario process
x_i: optimal decisions
$\boxed{i}$: node numbers



A full problem and the conditional problem "given node 3". The decision problem is time-consistent, if $x_i = \bar{x}_i$, for all nodes, which are in the subtree of the conditioning node.

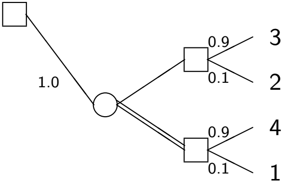
Time inconsistency appears in a natural way in stochastic risk-adverse optimality problems. We want to find

$$\max \mathbb{E}(Y) + 0.5\mathbb{A}\mathrm{V@R}_{0.05}(Y).$$



double line = optimal decision

The conditional problem given the first node:



Time consistent probability functionals

We consider a probability space $(\Omega, \mathcal{F}, P)$ and a filtration $\mathfrak{F} \in \mathcal{F}$. Let $\mathcal{U}_2(\cdot|\mathcal{F}_1)$ be a conditional acceptability-type mapping and let

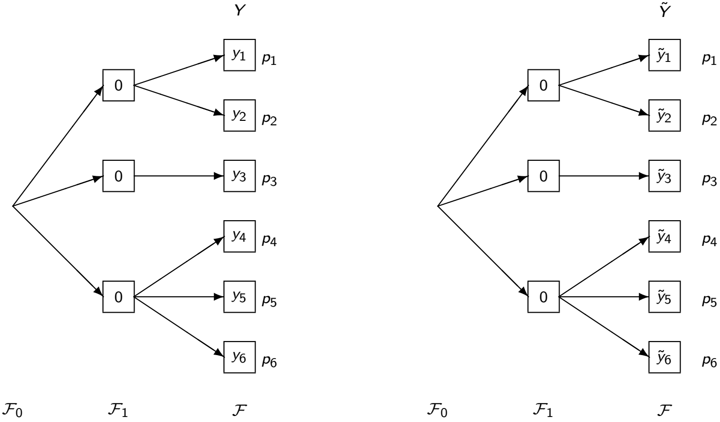
$$\mathcal{U}_1(\cdot)$$

be an unconditional acceptability measure. Typically, but not necessarily, $\mathcal{U}_1$ is the unconditional counterpart of $\mathcal{U}_2(\cdot|\mathcal{F}_1)$.
Definition. (Artzner at al. 2007). The pair $\mathcal{U}_1(\cdot)$, $\mathcal{U}_2(\cdot|\mathcal{F}_1)$ is called *time consistent*, if for all $Y, \tilde{Y} \in \mathcal{Y}$ the implication

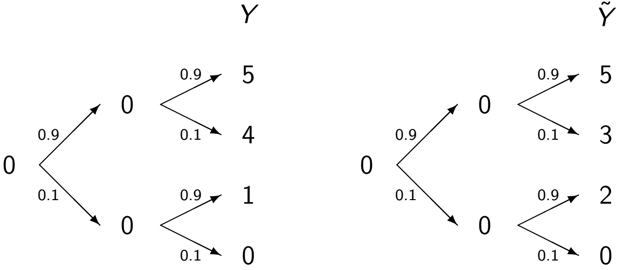
$$\mathcal{U}_2(Y|\mathcal{F}_1) \leq \mathcal{U}_2(\tilde{Y}|\mathcal{F}_1)\, a.s. \Longrightarrow \mathcal{U}_1(Y) \leq \mathcal{U}_1(\tilde{Y})$$

holds.

Illustration



⚡ $\mathbb{A}$V@R is not time-consistent.



$$\mathbb{A}\mathsf{V@R}_{0.1}(Y|\mathcal{F}_1) = (4;0) \geq (3;0) = \mathbb{A}\mathsf{V@R}_{0.1}(\tilde{Y}|\mathcal{F}_1)$$

while

$$\mathbb{A}\mathsf{V@R}_{0.1}(Y) = 0.9 < 1.8 = \mathbb{A}\mathsf{V@R}_{0.1}(\tilde{Y}).$$

Definition. A pair $\mathcal{U}_1(\cdot)$, $\mathcal{U}_2(\cdot|\mathcal{F}_1)$ is called *acceptance consistent*, if for all $Y \in \mathcal{Y}$ the implication

$$\mathrm{ess}\inf \mathcal{U}_2(Y|\mathcal{F}_1) \leq \mathcal{U}_1(Y)$$

holds. It is called *rejection consistent*, if

$$\mathrm{ess}\sup \mathcal{U}_2(Y|\mathcal{F}_1) \geq \mathcal{U}_1(Y).$$

(see e.g. Weber, 2006).
Proposition. If $\mathcal{U}_1(0) = 0$ and $\mathcal{U}_2(0|\mathcal{F}_1) = 0$ *a.s.* and $\mathcal{U}_1(\cdot)$, $\mathcal{U}_2(\cdot|\mathcal{F}_1)$ are translation equivariant then time consistency implies acceptance and rejection consistency.

The relation between time consistency and recursivity

Theorem. (Artzner et. al., 2007) A pair $\mathcal{U}_1(\cdot)$, $\mathcal{U}_2(\cdot|\mathcal{F}_1)$ with translation equivariant $\mathcal{U}(\cdot|\mathcal{F}_1)$, the property $\mathcal{U}(0|\mathcal{F}_1) = 0$ and monotonic $\mathcal{U}(\cdot)$ is time consistent if and only if it is recursive.
Proof. Let the pair be recursive and let $\mathcal{U}_2(Y|\mathcal{F}_1) \leq \mathcal{U}_2(\tilde{Y}|\mathcal{F}_1)$. Then, by monotonicity, $\mathcal{U}_1(Y) = \mathcal{U}_1(\mathcal{U}_2(Y|\mathcal{F}_1)) \leq \mathcal{U}_1(\mathcal{U}_2(\tilde{Y}|\mathcal{F}_1)) = \mathcal{U}_1(\tilde{Y})$. Conversely, let the pair be time consistent. By assumption,

$$\mathcal{U}_2(\mathcal{U}_2(Y|\mathcal{F}_1)|\mathcal{F}_1) = \mathcal{U}_2(\mathcal{U}_2(Y|\mathcal{F}_1) + 0|\mathcal{F}_1) = \mathcal{U}_2(Y|\mathcal{F}_1) + 0.$$

Setting $\tilde{Y} = \mathcal{U}_2(Y|\mathcal{F}_1)$ and using the time consistency, leads to

$$\mathcal{U}_1(\tilde{Y}) = \mathcal{U}_1(\mathcal{U}_2(Y|\mathcal{F}_1)) = \mathcal{U}_1(Y),$$

which is the equation of recursivity.

Enforcing time consistency by composition (nesting)

Let a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a filtration $\mathfrak{F} = (\mathcal{F}_0, \ldots, \mathcal{F}_T)$ of σ-fields $\mathcal{F}_t$, $t = 0, ..., T$, with $\mathcal{F}_T = \mathcal{F}$ be given. Let $\mathcal{Y}_t := L_p(\mathcal{F}_t)$ for $t = 1, \ldots, T$ and some $p \in [1, +\infty)$.
Let, for each $t = 1, \ldots, T$, conditional acceptability mappings $\mathcal{U}_{t-1} := \mathcal{U}(\cdot\,|\mathcal{F}_{t-1})$ from $\mathcal{Y}_T$ to $\mathcal{Y}_{t-1}$ be given. Introduce a multi-period probability functional $\mathcal{U}$ on $\mathcal{Y} := \times_{t=1}^{T} \mathcal{Y}_t$ by compositions of the conditional acceptability mappings $\mathcal{U}_{t-1}$, $t = 1, \ldots, T$, namely,

$$\begin{aligned} \mathcal{U}(Y; \mathfrak{F}) &:= \mathcal{U}_0[Y_1 + \cdots + \mathcal{U}_{T-2}[Y_{T-1} + \mathcal{U}_{T-1}(Y_T)]\cdot] \\ &= \mathcal{U}_0 \circ \mathcal{U}_1 \circ \cdots \circ \mathcal{U}_{T-1}(\sum_{t=1}^{T} Y_t) \end{aligned}$$

for every $Y_t \in \mathcal{Y}_t$. (Ruszczynski and Shapiro, 2006). Notice that these functionals are recursive in a trivial way.

The nested $\mathbb{A}$V@R

Example. Consider the conditional Average Value-at-Risk (of level $\alpha \in (0,1]$) as conditional acceptability mapping

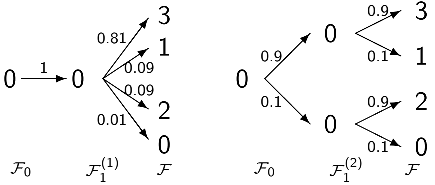
$$\mathcal{U}_{t-1}(Y_t) := \mathbb{A}\mathsf{V@R}_\alpha(\cdot\,|\mathcal{F}_{t-1})$$

for every $t = 1, \dots, T$. Then the multi-period probability functional

$$n\mathbb{A}\mathsf{V@R}_\alpha(Y;\mathfrak{F}) = \mathbb{A}\mathsf{V@R}_\alpha(\cdot\,|\mathcal{F}_0) \circ \cdots \circ \mathbb{A}\mathsf{V@R}_\alpha(\cdot\,|\mathcal{F}_{T-1})\left(\sum\nolimits_{t=1}^{T} Y_t\right)$$

satisfies is called the *nested Average Value-at-Risk*.

Time consistency contradicts information monotonicity.



In both examples, the final income Y is the same, but in the right example, the filtration is finer. One calculates

$$\mathbb{A}\mathsf{V@R}_{0.1}[\mathbb{A}\mathsf{V@R}_{0.1}(Y|\mathcal{F}_1^{(1)})] = 0.9 > 0 = \mathbb{A}\mathsf{V@R}_{0.1}[\mathbb{A}\mathsf{V@R}_{0.1}(Y|\mathcal{F}_1^{(2)})].$$

Notice that

$$\mathbb{E}[\mathbb{A}\mathsf{V@R}_{0.1}(Y|\mathcal{F}_1^{(1)})] = \mathbb{E}[\mathbb{A}\mathsf{V@R}_{0.1}(Y|\mathcal{F}_1^{(2)})] = 0.9.$$

Information monotonicity

- The expectation is information monotone.
- The essential infimum (or essential supremum) is information monotone

⚡ **Theorem.**(R. Kovacevic, G.P.) If a $\mathcal{U}_t(\cdot|\cdot)$ are distortion functionals, but neither the conditional expectation nor the essential infimum, then information monotonicity of the nested functional $\mathcal{U}$ does not hold.

Decomposing the final $\mathbb{A}$V@R: Random level $\mathbb{A}$V@R's

Let $\alpha \lhd \mathcal{F}_t$ be a random variable with values in [0,1]. Define the $\mathbb{A}$V@R with random level α as

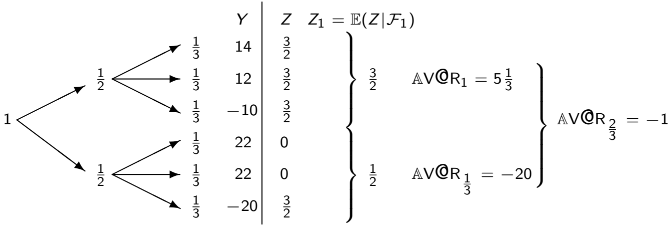
$$\mathbb{A}\mathsf{V@R}_\alpha(Y|\mathcal{F}_t) = \mathbf{inf}\{\mathbb{E}(YZ|\mathcal{F}_t) : \mathbb{E}(Y|\mathcal{F}_t) = 1, 0 \leq Z; \alpha Z \leq 1\}.$$

It has an alternate characterization for $\alpha > 0$ by

$$\mathbb{A}\mathsf{V@R}_\alpha(Y|\mathcal{F}_t) = \mathbf{sup}\{Q - \frac{1}{\alpha}\mathbb{E}([Q - Y]_+|\mathcal{F}_t) : Q \lhd \mathcal{F}_t\}.$$

The $\mathbb{A}$V@R with random level obeys all properties like the usual $\mathbb{A}$V@R, i.e. translation-equivariance, concavity, monotonicity, and positive homogeneity. Moreover, $\alpha \mapsto \mathbb{A}\mathsf{V@R}_\alpha$ is convex.

Illustration: Artzner's Example



The total $\mathbb{A}\mathrm{V@R}_{\frac{2}{3}}$ is -1, while $\mathbb{A}\mathrm{V@R}_{\frac{2}{3}}(Y|\mathcal{F}_1) \equiv 1$.

Theorem. Nested decomposition of the $\mathbb{A}$V@R
Let $Y \in L^1(\mathcal{F}_T)$, $\mathcal{F}_t \subset \mathcal{F}_\tau \subset \mathcal{F}_T$.

1. For $\alpha \in [0,1]$ the Average Value-at-Risk obeys the decomposition

$$\mathbb{A}\mathsf{V@R}_\alpha(Y) = \inf \mathbb{E}\left[Z_t \cdot \mathbb{A}\mathsf{V@R}_{\alpha \cdot Z_t}(Y|\mathcal{F}_t)\right], \tag{3}$$

   where the infimum is among all densities $Z_t \lhd \mathcal{F}_t$ with $0 \le Z_t$, $\alpha Z_t \le \mathbb{1}$ and $\mathbb{E}Z_t = 1$. For $\alpha > 0$ the infimum in (3) is attained.

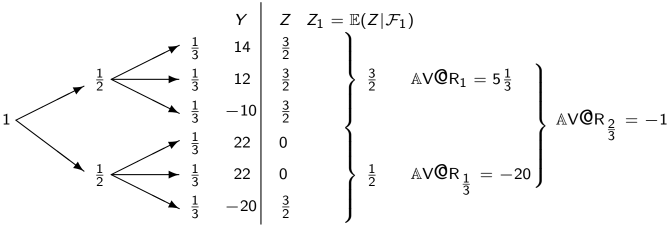
2. Moreover if Z is the optimal dual density for the $\mathbb{A}$V@R, that is $\mathbb{A}\mathsf{V@R}_\alpha(Y) = \mathbb{E}YZ$ with $Z \ge 0$, $\alpha Z \le \mathbb{1}$ and $\mathbb{E}Z = 1$, then $Z_t = \mathbb{E}[Z|\mathcal{F}_t]$ is the best choice in (3).

3. The *conditional* Average Value-at-Risk at random level $\alpha \lhd \mathcal{F}_t$ $(0 \le \alpha \le \mathbb{1})$ has the recursive (nested) representation

$$\mathbb{A}\mathsf{V@R}_\alpha(Y|\mathcal{F}_t) = \mathbf{inf}\mathbb{E}\left[Z_\tau \cdot \mathbb{A}\mathsf{V@R}_{\alpha \cdot Z_\tau}(Y|\mathcal{F}_\tau)\middle| \mathcal{F}_t\right], \tag{4}$$

   where the infimum is among all densities $Z_\tau \lhd \mathcal{F}_\tau$ with $0 \le Z_\tau$, $\alpha Z_\tau \le \mathbb{1}$ and $\mathbb{E}[Z_\tau|\mathcal{F}_t] = \mathbb{1}$.

Illustration



		Y	Z	$Z_1 = \mathbb{E}(Z\vert\mathcal{F}_1)$
$\frac{1}{2}$	$\frac{1}{3}$	14	$\frac{3}{2}$	$\frac{3}{2}$
	$\frac{1}{3}$	12	$\frac{3}{2}$	
	$\frac{1}{3}$	-10	$\frac{3}{2}$	
$\frac{1}{2}$	$\frac{1}{3}$	22	0	$\frac{1}{2}$
	$\frac{1}{3}$	22	0	
	$\frac{1}{3}$	-20	$\frac{3}{2}$	

$\mathbb{A}\mathrm{V@R}_1 = 5\frac{1}{3}$

$\mathbb{A}\mathrm{V@R}_{\frac{1}{3}} = -20$

$\mathbb{A}\mathrm{V@R}_{\frac{2}{3}} = -1$

The total $\mathbb{A}\mathrm{V@R}$ is $\mathbb{A}\mathrm{V@R}_\alpha(Y) = \mathbb{E}[Z_1 \mathbb{A}\mathrm{V@R}_{\alpha Z_1}(Y|\mathcal{F}_1)] = -1$, while $\mathbb{A}\mathrm{V@R}_{\frac{2}{3}}(Y|\mathcal{F}_1) \equiv 1$.

Notice that for $t < \tau$

$$\mathbb{A}\mathrm{V@R}_\alpha(Y|\mathcal{F}_t) \leq \mathbb{E}[\mathbb{A}\mathrm{V@R}_\alpha(Y|\mathcal{F}_\tau)|\mathcal{F}_t)] \leq \mathbb{E}(Y|\mathcal{F}_t)$$

A typical multistage decision problem

Let $H(x_0, \xi_1, \ldots, x_{T-1}, \xi_T)$ be some profit function depending on the random scenario process $\xi = (\xi_1, \ldots, \xi_T)$ and the decisions
$x = (x_0, \ldots, x_{T-1})$
The multistage decision problem is

$$\begin{array}{rl} \text{maximize} & \mathbb{E}H\left(x, \xi\right) + \gamma \cdot \mathbb{A}\mathrm{V@R}\left[H\left(x, \xi\right)\right] \\ \text{s.t.} & x \lhd \mathcal{F} \\ & x \in \mathcal{X}, \end{array} \tag{5}$$

where $H(x, \xi)$ is a short notation for $H(x_0, \xi_1, \ldots, x_{T-1}, \xi_T)$.

We require the real-valued function H to be concave in x, for x in a convex set, such that (ξ any fixed state)

$$H\left((1-\lambda)\,x' + \lambda x'', \xi\right) \geq (1-\lambda)\,H\left(x', \xi\right) + \lambda H\left(x'', \xi\right).$$

By the monotonicity property and concavity of the utility functional $\mathbb{A}\mathsf{V@R}$, the mapping $x \mapsto \mathbb{A}\mathsf{V@R}\left[H\left(x, \xi\right)\right]$ is concave as well.
With $x_{t_1:t_2}$ we denote the subvector $x_{t_1}, x_{t_1+1}, \ldots, x_{T_2}$.

As typical for Markov decision processes, we define the *value function*

$$\mathcal{V}_t\left(x_{0:t-1},\alpha,\gamma\right) := \operatorname{esssup}_{x_{t:T}} \mathbb{E}\left[\left. H\left(x_{0:T}\right)\right| \mathcal{F}_t\right] + \gamma \cdot \mathbb{A}\mathrm{V@R}_\alpha\left(H\left(x_{0:T}\right)\left|\mathcal{F}_t\right.\right).$$

The value function depends on

- the decisions up to time $t-1$, $x_{0:t-1}$, where $x_{t:T}$ is chosen such that $\left(x_{0:T}\right)=\left(x_{0:t-1},\,x_{t,T}\right)\in\mathcal{X}$,
- the random model parameters $\alpha\lhd\mathcal{F}_t$ and $\gamma\lhd\mathcal{F}_t$ and
- the current status of the system due to the filtration $\mathcal{F}_t$.

Evaluated at initial time $t=0$ and assuming the sigma-algebra $\mathcal{F}_0$ trivial the value function relates to the initial problem as

$$\begin{aligned}\sup_{x_{0:T}} \mathbb{E}H\left(x_{0:T}\right)+\gamma\cdot\mathbb{A}\mathrm{V@R}_\alpha\left(H\left(x_{0:T}\right)\right) = & \\ &= \operatorname{esssup}_{x_{0:T}} \mathbb{E}\left[H\left(x_{0:T}\right)\left|\mathcal{F}_0\right.\right]+\gamma\cdot\mathbb{A}\mathrm{V@R}_\alpha\left(H\left(x_{0:T}\right)\left|\mathcal{F}_0\right.\right)\\ &= \mathcal{V}_0\left([],\alpha,\gamma\right).\end{aligned}$$

Theorem. Dynamic Programming Principle. Assume that H is random upper semi-continuous with respect to x and ξ valued in some convex, compact subset of $\mathbb{R}^n$.

1. The value function evaluates to

$$\mathcal{V}_T\left(x_{0:T-1}, \alpha, \gamma\right) = (1+\gamma)\operatorname{esssup}_{x_T} H\left(x_{0:T}\right)$$

   at terminal time T.

2. For any $t < \tau$, $(t, \tau \in \mathbf{T})$ the recursive relation

$$\begin{aligned} & \mathcal{V}_t\left(x_{0:t-1}, \alpha, \gamma\right) \\ = \quad & \operatorname{esssup}_{x_{t:\tau-1}} \operatorname{essinf}_{Z_{t:\tau}} \mathbb{E}\left[\mathcal{V}_\tau\left(x_{0:\tau-1}, \alpha \cdot Z_{t:\tau}, \gamma \cdot Z_{t:\tau}\right) \middle| \mathcal{F}_t\right], \end{aligned}$$

   where $Z_{t:\tau} \lhd \mathcal{F}_\tau$, $0 \leq Z_{t:\tau}$, $\alpha Z_{t:\tau} \leq \mathbb{1}$ and $\mathbb{E}\left[Z_{t:\tau} | F_t\right] = \mathbb{1}$, holds true.

The Algorithm

Step 0 Let $x^0_{0:T}$ be any feasible, initial solution of the problem (5). Set $k \leftarrow 0$. Set

$$\mathcal{Y}(x^0_{0:T}) = \mathbb{E}H(x^0_{0:T}) + \gamma \mathbb{A}\mathsf{V@R}_\alpha(H(x^0_{0:T}))$$

Step 1 Find Z^k, such that $0 \leq Z^k \leq \frac{1}{\alpha}$, $\mathbb{E}Z^k = 1$ and define

$$Z^k_t := \mathbb{E}\left(Z^k | \mathcal{F}_t\right). \tag{6}$$

A good initial choice is often Z^k satisfying

$$\mathbb{E}Z^k H\left(x^k_{0:T}\right) = \mathbb{A}\mathsf{V@R}_\alpha\left(H\left(x^k_{0:T}\right)\right). \tag{7}$$

Step 2 (check for local improvement). Choose

$$\begin{aligned} x^{k+1}_t \in & \operatorname{argmax}_{x_t \lhd \mathcal{F}_t} \mathbb{E}\left[H\left(x^k_{0:T}\right)\middle| \mathcal{F}_t\right] & (8) \\ & + \gamma Z^k_t \mathbb{A}\mathsf{V@R}_{\alpha Z^k_t}\left(H\left(x^k_{0:T}\right)\middle| \mathcal{F}_t\right) & (9) \end{aligned}$$

at any arbitrary stage t and a node specified by $\mathcal{F}_t$.

Step 3 (Verification). Accept $x_{0:t}^{k+1}$ if

$$\mathcal{Y}\left(x_{0:T}^{k}\right) \leq \mathbb{E}H\left(x_{0:T}^{k+1}\right) + \gamma \mathbb{A}\mathsf{V@R}_{\alpha}\left(H\left(x_{0:T}^{k+1}\right)\right),$$

else try another feasible Z^k (for example $Z^k \leftarrow \frac{1}{2}\left(\mathbb{1} + Z^k\right)$, $Z^k \leftarrow (1+\alpha)\mathbb{1} - \alpha Z^k$ or $Z^k = \mathbb{1}_B$ $(P(B) \geq \alpha)$) and repeat Step 2. If no direction Z^k can be found providing an improvement, then $x_{0:T}$ is already optimal. Set

$$\mathcal{Y}\left(x_{0:T}^{k+1}\right) := \mathbb{E}H\left(x_{0:T}^{k+1}\right) + \gamma \mathbb{A}\mathsf{V@R}_{\alpha}\left(H\left(x_{0:T}^{k+1}\right)\right), \qquad (10)$$

increase $k \leftarrow k+1$ and continue with Step 1 unless

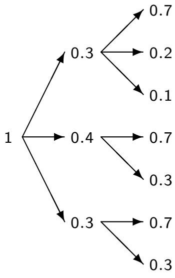
$$\mathcal{Y}\left(x_{0:T}^{k+1}\right) - \mathcal{Y}\left(x_{0:T}^{k}\right) < \varepsilon,$$

where $\varepsilon > 0$ is the desired improvement in each cycle k.

Extension for general distortion functionals

Decomposition Theorem. Let $\mathcal{U}$ be a positively homogeneous, version independent acceptability functional.

1. $\mathcal{U}_h$ obeys the decomposition
$$\mathcal{U}_h\left(Y\right)=\inf\mathbb{E}\left[Z\cdot\mathcal{U}_Z\left(Y|\mathcal{F}_t\right)\right],\tag{11}$$
   where the infimum is among all feasible, positive random variables $Z\lhd\mathcal{F}_t$ satisfying $\mathbb{E}Z=1$ and $h(U)\prec_{SSD}Z$ for $U\sim\mathit{Uniform}[0,1]$.
2. Let $\mathcal{F}_t\subset\mathcal{F}_\tau$. The utility functional obeys the nested decomposition
$$\mathcal{U}\left(Y|\mathcal{F}_t\right)=\operatorname{essinf}\mathbb{E}\left[Z_\tau\cdot\mathcal{U}_{Z_\tau}\left(Y|\mathcal{F}_\tau\right)\middle|\mathcal{F}_t\right],$$
   the essential infimum being among all feasible random variables $Z_\tau\lhd\mathcal{F}_\tau$.



Y	Z	$\mathbb{E}(Z\|\mathcal{F})$	$Z'_{0.7}\cdot Z_{0.7}$		$Z'_{0.4}\cdot Z_{0.4}$		$\mathcal{R}_Z(Y\|\mathcal{F})$
1	$\frac{2}{7}$	1	0	1	$\frac{5}{7}$	1	$6.07 = \frac{3}{5}\mathbb{A}V@R_{0.7} + \frac{2}{5}\mathbb{A}V@R_{0.4}$
7	2.66		3.33		1.67		
8	2.66		3.33		1.67		
0	0	0.2	0	0	0	0.5	$3 = 0\cdot\mathbb{A}V@R_1 + 1\cdot\mathbb{A}V@R_{0.7}$
3	$\frac{2}{3}$		0		1.67		
4	$\frac{38}{21}$	2.06	$\frac{40}{21}$	2.33	1.67	1.67	$5.94 = \frac{21}{31}\mathbb{A}V@R_{0.3} + \frac{10}{31}\mathbb{A}V@R_0$
9	2.66		3.33		1.67		

Nested decomposition of $\mathcal{R} = \frac{3}{5}\mathrm{U}\mathbb{A}\mathrm{V@R}_{0.7}(Y) + \frac{2}{5}\mathrm{U}\mathbb{A}\mathrm{V@R}_{0.4}(Y)$. We get

$$\mathcal{R}(Y) = \mathbb{E}[Z|\mathcal{R}_Z(Y|\mathcal{F}_t)] = 6.07\cdot1\cdot0.3+3\cdot0.2\cdot0.4+5.94\cdot2.06\cdot0.3 = 5.74$$

Utility functionals are typically not time consistent

Theorem. Suppose that the positively homogeneous functional $\mathcal{U}$ has a Kusuoka representation

$$\mathcal{U}(Y) = \inf\{\int_0^1 \mathbb{A}\mathsf{V@R}_\alpha(Y)\, d\mu(\alpha) : \mu \in \mathcal{M}\}.$$

If

$$\inf\{\mu([\epsilon, 1-\epsilon]) : \mu \in \mathcal{M}\} > 0$$

for some $\epsilon > 0$ and

$$\sup\{\mu([0,\gamma]) : \mu \in \mathcal{M}\} \to 0$$

for $\gamma \to 0$, then $\mathcal{U}$ is not time-consistent as such, but has to be randomly decomposed for ensuring time-consistency.
The only exceptions are

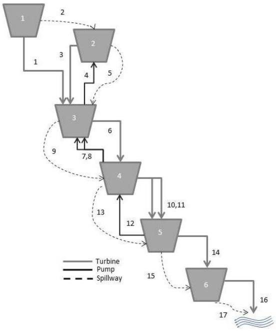
- the expectation
- the essential infimum
- the essential supremum

These are the same functionals, which are information monotone.

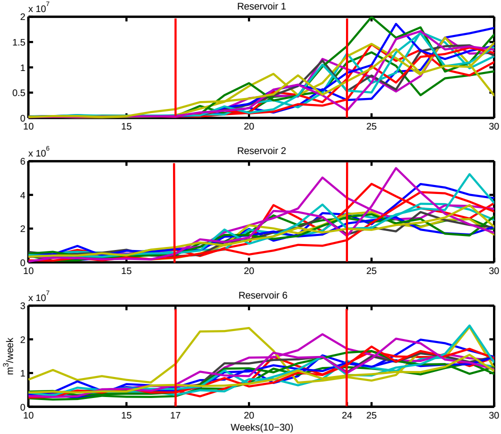
Conclusions

- Compositions of risk functionals are time consistent (but not interpretable) and information inconsistent
- Final risk functionals are typically information consistent but not time consistent
- Exceptions are only the expectation and the (essential) infimum resp. supremum
- When using time-inconsistent functionals one one has to decide:
  - either to accept time-inconsistent decisions in a rolling horizon setup
  - or to accept decision criteria which depend on the actual path the scenario process takes.

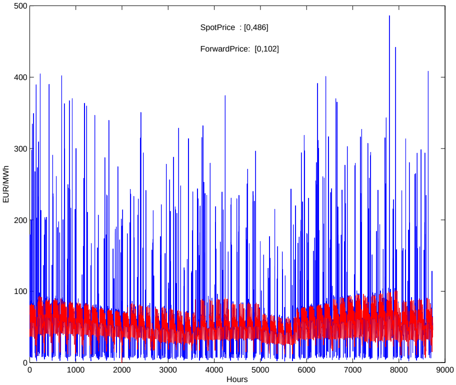
Case study: Management of a hydrosystem



The scenario process consist of 5 components: Spot prices, Pumping prices, Inflows for 3 reservoirs. Statistical model selection methods were used to find that the inflows can be represented by a 3-dimensional $SARMA(1,2),(2,2)_5 2$ process, while the spot and pumping prices can be modeled by an independent process, a superposition of an additive error model based on forward prices and a spike generating process.



Observations for Inflows



Observations for Spot/Forward prices

The decision model

maximize

$\lambda \mathbb{E}[x_T^c] - (1-\lambda)\mathbb{A}\mathsf{V@R}_{1-\alpha}[-x_T^c]$

subject to

$0 \leq x_{t,i}^f \leq \overline{x}_i^f,$

$\underline{x}_j^s \leq x_{t,j}^s \leq \overline{x}_j^s,$

$x_{end,j}^{',s} \leq x_{T,j}^s,$

$x_{t,j}^s = x_{t-1,j}^s + \xi_{t,j}^f + \sum_{\{i \in I | P_{max} > 0\}} A_{i,j} \cdot x_{t-1,i}^f + \sum_{\{i \in I | P_{max} = 0\}} A_{i,j} \cdot x_{t,i}^f,$

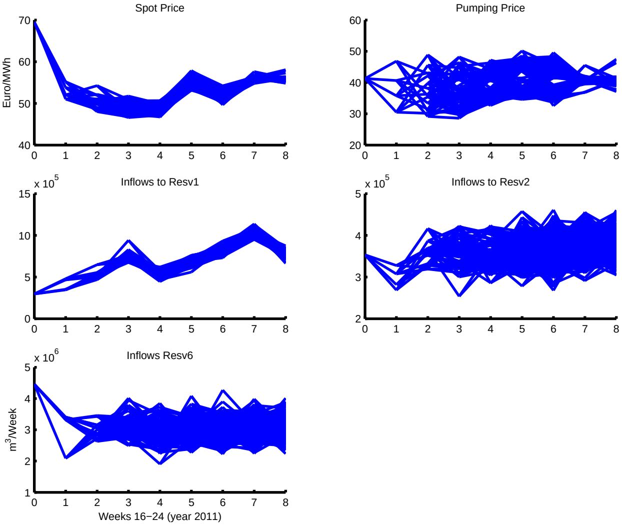
$x_{t,i}^e = x_{t-1,i}^f \cdot k^i \cdot \triangle t_{(t-1)},$

$x_t^c = x_{t-1}^c \cdot (1+r)^{\triangle t_{(t-1)}} + \sum_{\{i \in I | k^i > 0\}} x_{t-1,i}^e \cdot \xi_t^e + \sum_{\{i \in I | k^i < 0\}} x_{t-1,i}^i \cdot \xi_t^p.$

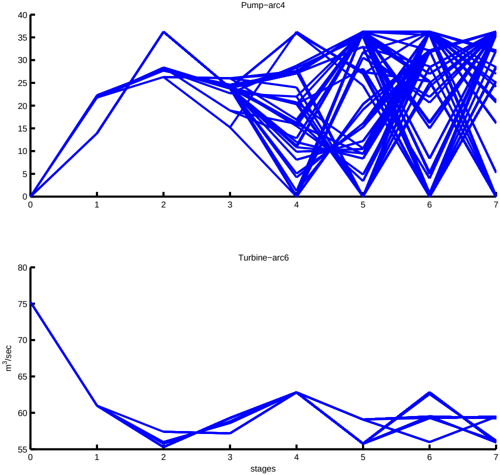
Generating a scenario tree

We generate a scenario tree in a way that the nested distance between the scenario process and the scenario tree is as small as possible.

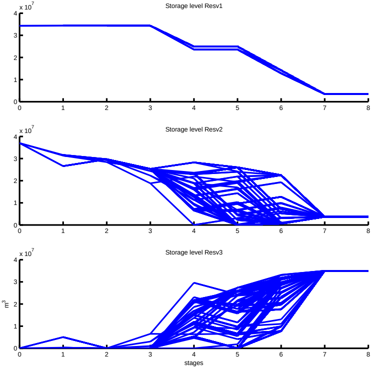
Number of stages	8
Minimal bushiness per stage	2,2,2,1,1,1,1,1
Maximal distance per stage	5,5,5,7,7,7,10,10
Number of scenarios (leaves)	392
Number of nodes 1532	



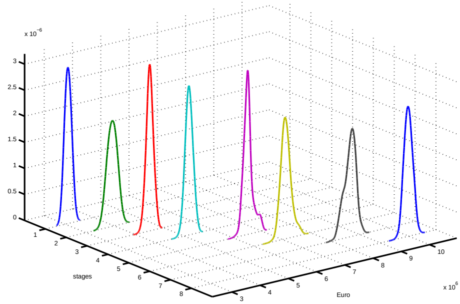
The generated five-dimensional tree



The pumping (top) and turbining (bottom) decisions



The storage levels



Accumulated Cash

The stochastic discretization algorithm

1. **Initialization.** Sample n random variates from distribution P, where n is much larger than s. Use a cluster algorithm to find s clusters. Let $Z(0) = \big(z^{(1)}(0), \ldots, z^{(s)}(0)\big)$ be the cluster medians. Set $k = 0$.
2. **Iteration.** Use a new independent sample $\xi(k)$ for the following stochastic optimization step: find the index $i \in \{1, \ldots s\}$ such that

$$\mathrm{d}\left(\xi(k),\, z^{(i)}(k)\right) = \min_{\ell} \mathrm{d}\left(\xi(k),\, z^{(\ell)}(k)\right).$$

Set

$$z^{(i)}(k+1) = z^{(i)}(k) - a_k \cdot r\,\mathrm{d}\left(\xi(k),\, z^{(i)}\right)^{r-1} \cdot \nabla_{z^{(i)}} \mathrm{d}\left(\xi(k),\, z^{(i)}\right),$$

and leave all other points unchanged to form the new point set $Z(k+1)$.

3. **Stopping criterion.** Set $k = k + 1$ and goto 56. Stop, if either the predetermined number of iterations are performed or if the relative change of the point set Z is below some threshold ϵ.
4. **Determination of the probabilities.** After having fixed the final point set Z, generate another sample $\xi(1), \ldots, \xi(n)$ and find the probabilities

   $$p_i = \frac{1}{n} \# \left\{ \ell : \mathrm{d}\left(\xi(\ell), z^{(i)}\right) = \min_k \mathrm{d}\left(\xi(\ell), z^{(k)}\right) \right\}$$

   and calculate an estimate for the distance $\mathrm{d}(P, \tilde{P})$.

The final approximate distribution is $\tilde{P} = \sum_{i=1}^{s} p_i \cdot \delta_{z^{(i)}}$.

The tree generation algorithm

Let a vector of minimal bushiness $b_1, \ldots, b_T$ and maximal distances $d_1, \ldots, d_T$ be given.

- Iterate for $t = 0, \ldots, T-1$.
  For each node n of stage t
  - (i) Set $s = b_t$
  - (ii) Let $\tilde{\xi}_0, \ldots, \tilde{\xi}_{t-1}$ be the already found scenario values on the predecessors of this node. Let P be the conditional distribution of ξ_t given $\tilde{\xi}_0, \ldots, \tilde{\xi}_{t-1}$ form which one may sample. Use the stochastic discretization algorithm to generate the conditional distribution $\tilde{P}$ sitting on s points.
  - (iii) If the dstance $\mathrm{d}(P, \tilde{P})$ is smaller than ϵ_t, then set $s = s + 1$ and go to (ii).
- Stop, when all nodes of stage T are generated.

The tree reduction algorithm (Kovacevic and Pichler)

- **Step 1– Initialization**

  Set $k \leftarrow 0$, and let ξ^0 be process quantizers with related transport probabilities $\pi^0(i,j)$ between scenario i of the original $\mathbb{P}$-tree and scenario $\tilde{\xi}_j^0$ of the approximating $\mathbb{P}'$-tree; $\mathbb{P}^0 := \tilde{\mathbb{P}}$.

- **Step 2 – Improve the quantizers**

  Find improved quantizers $\tilde{\xi}_j^{k+1}$:

  - In case of the quadratic Wasserstein distance (Euclidean distance and Wasserstein of order $r = 2$) set

    $$\tilde{\xi}^{k+1}(n_t) := \sum_{m_t \in \mathcal{N}_t} \frac{\pi^k(m_t, n_t)}{\sum_{m_t \in \mathcal{N}_t} \pi^k(m_t, n_t)} \cdot \xi_t(m_t),$$

  - or find the barycenters by applying the steepest descent method, or the limited memory BFGS method.

- **Step 3 – Improve the probabilities**

  Setting $\pi \leftarrow \pi^k$ and $q \leftarrow q^{k+1}$ and calculate all conditional probabilities $\pi^{k+1}(\cdot,\cdot|m,n) = \pi^*(\cdot,\cdot|m,n)$, the unconditional transport probabilities $\pi^{k+1}(\cdot,\cdot)$ and the distance $\mathsf{dl}_r^{k+1} = \mathsf{dl}_r\left(\mathbb{P},\tilde{\mathbb{P}}\right)$.

- **Step 4**

  Set $k \leftarrow k+1$ and continue with Step 2 if

  $$\mathsf{dl}_r^{k+1} < \mathsf{dl}_r^{k} - \varepsilon,$$

  where $\varepsilon > 0$ is the desired improvement in each cycle k.

  Otherwise, set $\tilde{\xi}^* \leftarrow \tilde{\xi}^k$, define the measure

  $$\tilde{P}^{k+1} := \sum_j \delta_{\tilde{\xi}_j^{k+1}} \cdot \sum_i \pi^{k+1}(i,j),$$

  for which $\mathsf{dl}_r\left(\mathbb{P},\mathbb{P}^{k+1}\right) = \mathsf{dl}_r^{k+1}$ and stop.

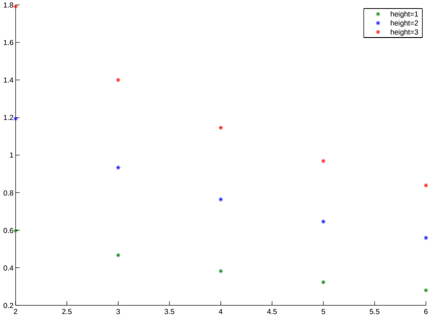
In case of the quadratic nested distance ($r = 2$) and the Euclidean distance the choice $\varepsilon = 0$ is possible.

Computational experience

Stages	4	5	5	6	7	7
Nodes of the initial tree	53	309	188	1,365	1,093	2,426
Nodes of the approx. tree	15	15	31	63	127	127
Time/ sec.	1	10	4	160	157	1,044

Approximation at work

Reducing the nested distance by making the tree bushier.



Lower bounds

$$Opt(\mathbb{P}): \quad v^*(\mathbb{P}) = \min\{\mathcal{R}_{\mathbb{P}}[H(x,\xi)] \,:\, x \lhd \mathfrak{F}; \mathbb{P} \sim (\Omega, \mathfrak{F}, P, \xi)\}$$

Lemma. Suppose that the functional $P \longmapsto \mathcal{R}_P(\cdot)$ is compound concave (i.e. the mapping $P \mapsto \mathcal{R}_P(Y)$ is concave for all random variables Y for which $\mathcal{R}$ is defined. Then the mapping $P \mapsto v^*(P)$ is also concave. Consequently, if one dissects the probability measure

$$P = \sum_{i=1}^{k} P(\omega_i)\delta_{\omega_i}. \tag{12}$$

then

$$\sum_{i=1}^{k} p_i v^*(P_i) \leq v^*(P).$$

Refinement Chains