Approximation of Continuous-State Scenario Processes in Multi-Stage Stochastic Optimization and its Applications

Anna Timonina

University of Vienna,
Abraham Wald PhD Program in Statistics and Operations Research.
Supervisor Prof. Georg Pflug

International Institute for Applied Systems Analysis (IIASA),
Risk, Policy and Vulnerability Program

Initial Problem Description

$$\max_x \{\mathcal{A}[H(x,\xi);\mathcal{F}] : x \lhd \mathcal{F}, x \in \mathbb{X}\},$$

$\mathcal{A}(\cdot,\cdot)$ is a multi-period acceptability functional ($\mathbb{E}, \mathbb{AV@R}$...);
$H(x,\xi)$ is the (intermediate) loss/profit function;
$\xi = (\xi_1, ..., \xi_T)$ is a stochastic process on $(\Omega, \mathcal{F} = (\mathcal{F}_1, ..., \mathcal{F}_T), P)$, where $\mathcal{F}$ is a filtration.
$x = (x_0, ..., x_{T-1})$ are the decision functions;
$x \lhd \mathcal{F}$ is a non-anticipativity condition (means that x_t is measurable w.r.t. $\mathcal{F}_t$ for all t).
That is a variational problem that could be solved only in very special cases.

Problem Approximation

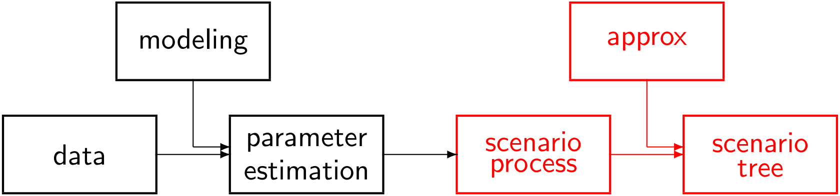
$$\max_x \{\mathcal{A}[H(x,\xi);\mathcal{F}] : x \lhd \mathcal{F}, x \in \mathbb{X}\},$$

How to solve this problem though in general it is unsolvable?
We approximate this problem by a simpler one:

$$\max_x \{\mathcal{A}[H(\widetilde{x},\widetilde{\xi});\widetilde{\mathcal{F}}] : \widetilde{x} \lhd \widetilde{\mathcal{F}}, \widetilde{x} \in \mathbb{X}\},$$

that is a tree structured problem of huge but finite dimension, where
$\widetilde{\Omega}$ is a finite probability space;
$\widetilde{\mathcal{F}}$ is a finite filtration;
$\widetilde{x}$ is a high dimensional vector.

From data to tree models



Example

Energy demand, that can be described by, for example, by SARIMA or GARCH models, can be represented in terms of the scenario tree as soon as the parameters of the model are estimated and the approximation is done. For example, GARCH model with $\epsilon_t = \sigma_t z_t$

$$\xi_t = \mu + \beta_1 \xi_{t-1} + \epsilon_t,$$

where $\sigma_t^2 = \kappa + \gamma_1 \sigma_{t-1}^2 + \ldots + \gamma_n \sigma_{t-n}^2 + \alpha_1 \epsilon_{t-1}^2 + \ldots + \alpha_m \epsilon_{t-m}^2.$

Filtration as a tree

$\xi = (\xi_1, ..., \xi_T)$ a stochastic process defined on $(\Omega, \mathcal{F}, P)$;
$\widetilde{\xi} = (\widetilde{\xi}_1, ..., \widetilde{\xi}_T)$ a *finite valued* scenario process defined on $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{P})$.

It is not natural to construct approximations on the same probability space, because concrete decision problems are given without reference on a probability space, they are given in distributional setups.
As the stochastic process ξ is given by its continuous probability distribution only, we need to approximate this distribution by discrete one, i.e. to generate points from this distribution.
Hence, our aim is

1. To generate points from the given distribution (using different quantization algorithms);
2. To solve the multi-stage stochastic optimization program using the generated points.

For these purposes we represent stochastic process $\widetilde{\xi} = (\widetilde{\xi}_1, ..., \widetilde{\xi}_T)$ *as a finitely valued tree.*

Sample tree

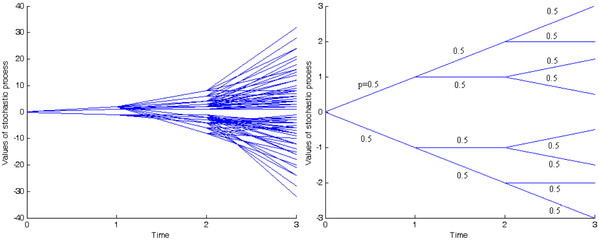


Figure: High bushiness v.s. Low bushiness

Definition

Consider a finitely valued stochastic process $\widetilde{\xi} = (\widetilde{\xi}_1, ..., \widetilde{\xi}_T)$ *that is represented by the tree with the same number of successors* b_t *for each node at the stage* t, $\forall t = 1, ..., T$. *The vector bush* $= (b_1, ..., b_T)$ *is a bushiness vector of the tree.*

Kantorovich distance between measures

Definition

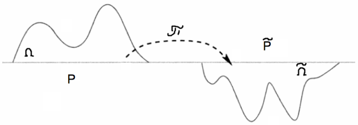
The **Kantorovich distance** *between measures is defined as*

$$d_{KA}(P, \widetilde{P}) = \inf_{\pi}\{\int_{\Omega\times\widetilde{\Omega}} d(w, \widetilde{w})\pi[dw, d\widetilde{w}]\},$$

$$\pi[\cdot \times \widetilde{\Omega}] = P(\cdot),$$

$$\pi[\Omega \times \cdot] = \widetilde{P}(\cdot).$$

Wasserstein distance: $d_{WA_r}(P, \widetilde{P}) = \inf_{\pi}\{\int_{\Omega\times\widetilde{\Omega}} d(w, \widetilde{w})^r \pi[dw, d\widetilde{w}]\}^{\frac{1}{r}}$ *under the same constraints.*



Nested distance between trees

Definition

The multistage distance (see [1, 2]) of two process-and-information structures is

$$dl_r(\mathbb{P}, \widetilde{\mathbb{P}}) = \inf_{\pi} \left(\int d(w, \widetilde{w})^r \pi(dw, d\widetilde{w}) \right)^{\frac{1}{r}},$$

$$dl_1(\mathbb{P}, \widetilde{\mathbb{P}}) := dl(\mathbb{P}, \widetilde{\mathbb{P}}),$$

where $\mathbb{P}$, $\widetilde{\mathbb{P}}$ *are nested distributions (containing information about filtration structure, values and probabilities)*

$$\pi[A \times \widetilde{\Omega} | \mathcal{F}_t \otimes \widetilde{\mathcal{F}_t}](w, \widetilde{w}) = P(A|\mathcal{F}_t)(w), \; (A \in \mathcal{F}_T, 1 \leq t \leq T),$$

$$\pi[\Omega \times B | \mathcal{F}_t \otimes \widetilde{\mathcal{F}}_t](w, \widetilde{w}) = \widetilde{P}(B|\widetilde{\mathcal{F}}_t)(\widetilde{w}), \; (B \in \widetilde{\mathcal{F}}_T, 1 \leq t \leq T).$$

Recursive Structure of the Nested Distance

Distance between leaves of the tree ($t = T$): The nested distance between last stages of tree models $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ conditional on nodes n_{T-1} and $\widetilde{n}_{T-1}$ is equal to the Kantorovich distance, i.e.

$$dl(\mathbb{P}^{T:T|n_{T-1}}, \widetilde{\mathbb{P}}^{T:T|\widetilde{n}_{T-1}}) = d_{KA}(P_T(\cdot|n_{T-1}), \widetilde{P}_T(\cdot|\widetilde{n}_{T-1})).$$

Nested distance between subtrees ($\forall t = 2, ..., T-1$): For each stage $t = 2, ..., T-1$, the nested distance $dl(\mathbb{P}^{t:T|n_{t-1}}, \widetilde{\mathbb{P}}^{t:T|\widetilde{n}_{t-1}})$ can be received as the solution of the linear program (0.3) with the distance matrix

$$d_{n_t,\widetilde{n}_t} = d(n_t, \widetilde{n}_t) + dl(\mathbb{P}^{t+1:T|n_t}, \widetilde{\mathbb{P}}^{t+1:T|\widetilde{n}_t}),$$

given that the distance $dl(\mathbb{P}^{t+1:T|n_t}, \widetilde{\mathbb{P}}^{t+1:T|\widetilde{n}_t})$ is calculated at the previous step.

Resulting nested distance ($t = 1$): The resulting nested distance is implied by the previous step for the case $t = 1$.

Nested distance approximation

The fundamental result of [2] says that

$$|v(\mathbb{P}) - v(\widetilde{\mathbb{P}})| \leq L_\beta dl(\mathbb{P}, \widetilde{\mathbb{P}}).$$

According to the triangle inequality that holds for the nested distance, we can write

$$dl(\mathbb{P}, \widetilde{\mathbb{P}}) \leq dl(\mathbb{P}^*, \widetilde{\mathbb{P}}) + dl(\mathbb{P}, \mathbb{P}^*).$$

In order to approximate the distance $d(\mathbb{P}, \widetilde{\mathbb{P}})$ between stochastic process and tree $\widetilde{\mathbb{P}}$ by the distance $d(\widetilde{\mathbb{P}}, \mathbb{P}^*)$ between tree $\widetilde{\mathbb{P}}$ and the process approximation (tree $\mathbb{P}^*$) we should guarantee that the distance between stochastic process ξ and its approximation (tree $\mathbb{P}^*$) is small enough:

$$dl(\mathbb{P}, \mathbb{P}^*) \leq \varepsilon.$$

We can guarantee it if we are increasing the bushiness of the tree $\mathbb{P}^*$, because in this case $dl(\mathbb{P}, \mathbb{P}^*) \to 0$ and $dl(\mathbb{P}^*, \widetilde{\mathbb{P}}) \to dl(\mathbb{P}, \widetilde{\mathbb{P}})$.

Initial Lower Bound

Theorem (**Lower Bound (initial), see [2]**)

Let $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ be two nested distributions and let P and $\widetilde{P}$ be the pertaining multivariate distributions respectively. Then

$$d_{WA_r}(P, \widetilde{P}) \leq dl_r(\mathbb{P}, \widetilde{\mathbb{P}}),$$

where $d_{WA_r}(P, \widetilde{P})$ is the Wasserstein distance of order r.

Initial Upper Bound

Theorem (**Upper Bound (initial), see [2]**)

Let $\mathbb{P}$ be a nested distribution, P its multivariate distribution, which is dissected into the chain $P = P_1 \circ P_2 \circ \ldots \circ P_T$ of conditional distributions. If $\forall t = 1, \ldots, T$ the Lipschitz property holds, i.e.

$$d_{KA}(P_t(\cdot|u), P_t(\cdot|v)) \leq K_t d(u, v),$$

then

$$d_{KA}(P, \widetilde{P}) \leq dl(\mathbb{P}, \widetilde{\mathbb{P}}) \leq \sum_{t=1}^{T} d_{KA}(P_t, \widetilde{P}_t) \prod_{s=t+1}^{T} (K_s + 1).$$

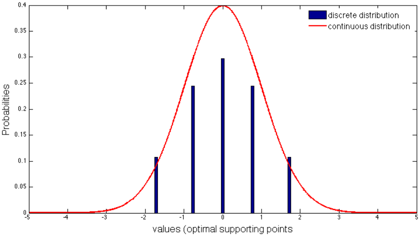
Stage-wise optimal quantization

Optimal quantization means:

1. to find optimal supporting points z_i, $i = 1, ..., N$ $(z_1 \leq z_2 \leq ... \leq z_N)$:

$$\min_{z=(z_1,...,z_N)} \int \min_s d(x, z_s)^r dP(x)$$

2. given the supporting points z_i, to find the probabilities p_i, such that



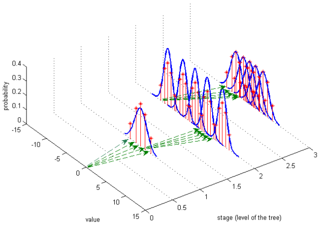
$$d_{KA}(P, \widetilde{P}) \to min.$$

Stage-wise optimal quantization on a treestructure

The stage-wise optimal tree approximation of the stochastic process $\xi = (\xi_1, ..., \xi_T)$ solves the minimization problem

$$dl(\mathbb{P}, \widetilde{\mathbb{P}}) \leq \sum_{t=1}^{T} d_{KA}(P_t, \widetilde{P}_t) \prod_{s=t+1}^{T} (K_s + 1) \rightarrow \min$$

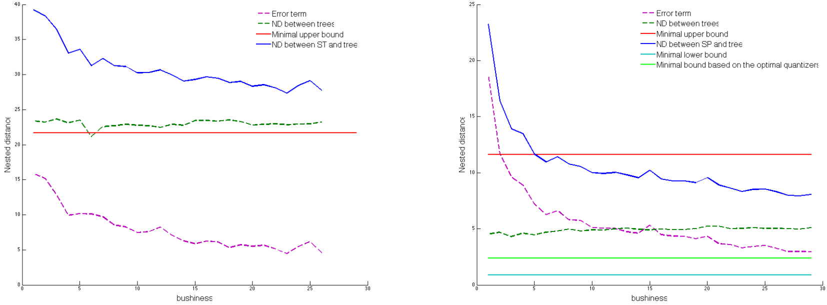
1. First stage of the tree has N_1 nodes. We generate N_1 values of ξ_1 according to the unconditional probability distribution of ξ_1;
2. For each of the following stages $t = 2, ..., T$ we generate ξ_t according to the conditional distribution of ξ_t given the historical values of the random variables ξ^{t-1}.



New results

Is there any chance to improve?

Yes, it can and the minimal upper bound DOES NOT mean that the nested distance is minimal:



a. b.

Figure: Behavior of the nested distance between stochastic process and a tree.

Clairvoyant tree structure

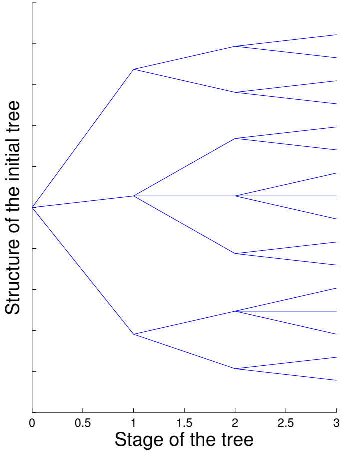


Figure: From primal to the clairvoyant tree.

From primal to the clairvoyant tree

Definition (**Clairvoyant tree)**

Tree model $\mathbb{P}_c$ is called clairvoyant at/from stage t and denoted by $\mathbb{P}_{c(t)}$ if every node $n \in \mathcal{N}_s$, $\forall s = t+1, ..., T-1$ has unique successor with probability 1 ($\mathcal{N}_s$ is the set of all node indexes at the stage s) (see Fig. 3).

Out of the tree $\mathbb{P}$, it is possible to receive another tree $\mathbb{P}_{c(t)}$ by changing the tree structure of $\mathbb{P}$ so, that it becomes clairvoyant starting from the stage t.

$$P(\xi_1, \xi_2, ..., \xi_T) = P(\xi_1) \cdot P(\xi_2|\xi_1) \cdot ... \cdot P(\xi_T|\xi_1, ..., \xi_{T-1}),$$
$$P(\xi_1, \xi_2, ..., \xi_T) = P(\xi_1) \cdot P(\xi_2|\xi_1) \cdot ... \cdot P_{c(t)}(\xi_t, ..., \xi_T|\xi_{t-1}, ..., \xi_1).$$

Notice, that here we used the fact that the multivariate probability $P(\xi_1, \xi_2, ..., \xi_T)$ is the same for the initial and the clairvoyant trees, as well as the conditional probabilities up to time $t-1$.

Improved Lower Bound

Theorem (**Lower Bound (new)**)

Let $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ be two nested distributions with corresponding multivariate distributions P and $\widetilde{P}$ and suppose that clairvoyant nested distributions $\mathbb{P}_{c(t)}$ and $\widetilde{\mathbb{P}}_{c(t)}$ are defined respectively $\forall t = 1, ..., T$.
Then the following chain of lower bounds for the nested distance between $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ holds:

$$d_{KA}(P, \widetilde{P}) = dl(\mathbb{P}_{c(1)}, \widetilde{\mathbb{P}}_{c(1)})) \leq ... \leq dl(\mathbb{P}_{c(t)}, \widetilde{\mathbb{P}}_{c(t)}) \leq ...$$
$$... \leq dl(\mathbb{P}_{c(T-1)}, \widetilde{\mathbb{P}}_{c(T-1)}) \leq dl(\mathbb{P}_{c(T)}, \widetilde{\mathbb{P}}_{c(T)}) = dl(\mathbb{P}, \widetilde{\mathbb{P}}).$$

Counterexample: what if to change only one of the trees to clairvoyant? (see MatLab example)

Improved Upper Bound

Theorem (**Improved Upper Bound (new)**)

Let $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ be two nested distributions with corresponding multivariate distributions P and $\widetilde{P}$ dissected into the chain of conditional probabilities $P = P_1 \circ P_2 \circ ... \circ P_T$ and $\widetilde{P} = \widetilde{P}_1 \circ \widetilde{P}_2 \circ ... \circ \widetilde{P}_T$. Then,

$$dl(\mathbb{P}, \widetilde{\mathbb{P}}) \leq \sum_{t=1}^{T} \sup_{u^{t-1}, v^{t-1}} d_{KA}(P^{t|t+1:T}(\cdot|u^{t-1}), \widetilde{P}^{t|t+1:T}(\cdot|v^{t-1})),$$

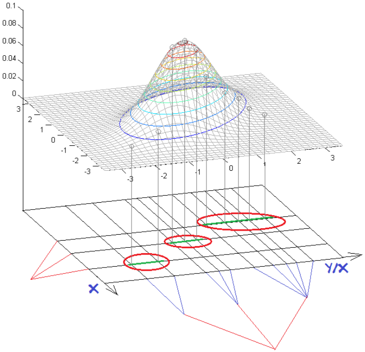
where $P^{t|t+1:T}(|u^{t-1})$ and $\widetilde{P}^{t|t+1:T}(|u^{t-1})$ are conditional multivariate distributions sitting at the stage t of clairvoyant trees $\mathbb{P}_{c(t)}$ and $\widetilde{\mathbb{P}}_{c(t)}$. Moreover, if $\forall t = 1, ..., T$ the Lipschitz property holds, then

$$dl(\mathbb{P}, \widetilde{\mathbb{P}}) \leq \sum_{t=1}^{T} d_{KA}(P^{t|t+1:T}, \widetilde{P}^{t|t+1:T}) \prod_{s=t+1}^{T} (K^{s|s+1:T} + 1).$$

Backtracking Optimal Quantization

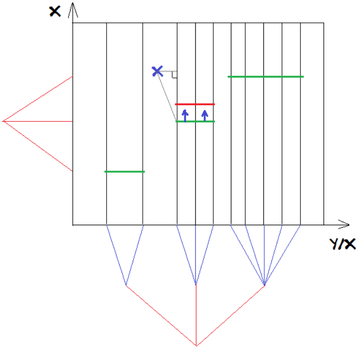
$$dl(\mathbb{P}, \widetilde{\mathbb{P}}) \leq \sum_{t=1}^{T} d_{KA}(P^{t|t+1:T}, \widetilde{P}^{t|t+1:T}) \prod_{s=t+1}^{T} (K^{s|s+1:T} + 1) \rightarrow min.$$

1. FORWARD: using the stage-wise optimal quantization algorithm we receive the first approximation of the quantizers for the stages $t = 1, ..., T$;
2. BACKWARD: starting with the stage $T - 1$ we adapt all the quantizers of the stages $t = 1, ..., T - 1$ so, that the Kantorovich distance between the joint probability distribution of $P^{t:T}$ and its discrete approximation is minimized.



Backtracking Optimal Quantization

In order to adapt the quantizers ξ_t of the stage t to the possible future scenarios $t+1:T$ we generate N random points $x_{t:T}$ and for each of them we move the closest quantizer towards it keeping the future quantizers $\xi_{t+1},...,\xi_T$ at the given level.



Example: Joint Adaptation

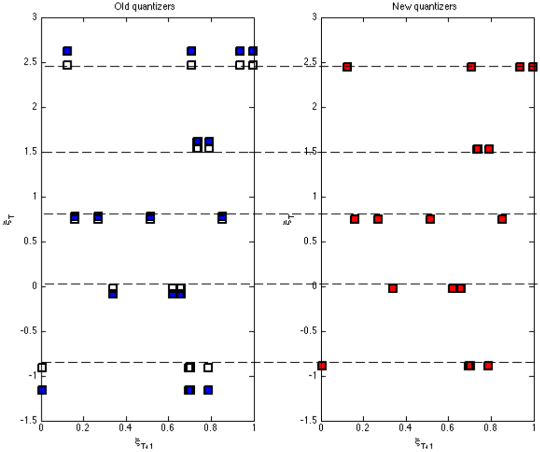


Figure: Joint Adaptation Algorithm.

Is there an improvement?

Yes, there is:

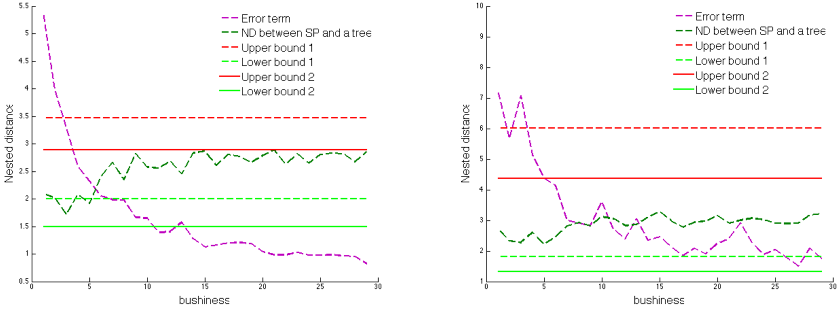


Figure: Improvement due to the Backtracking Optimal Quantization.

MatLab examples: does the tree improve?

Example: Backtracking Optimal Quantization (random)

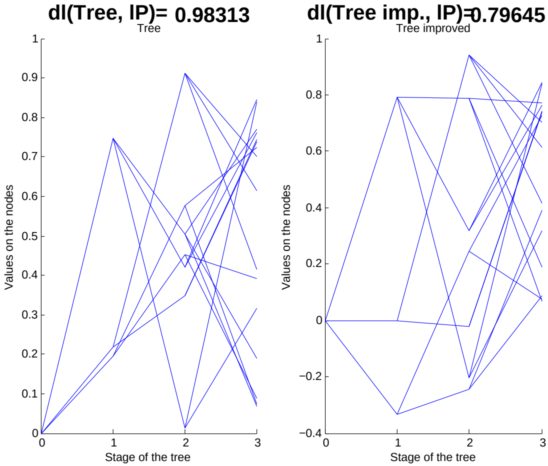


Figure: Decrease in the nested distance due to the improvement in random quantizers.

Example: Backtracking Optimal Quantization (optimal)

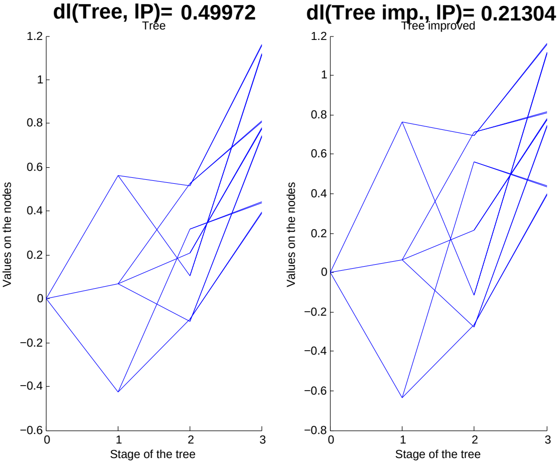


Figure: Decrease in the nested distance due to the improvement in stage-wise optimal quantizers.

Applications

Applications

Real life:

1. Energy, finance;
2. Risk-management of catastrophic events: scenario generation based on the historical probability loss disribution;

Virtual life:

1. Networks, Internet: PageRank problem of Google in the multi-stage environment.

Inventory Control Problem

Example: inventory control problem

The grocery shop has to place regular orders one period ahead.

- $\xi_1, \dots, \xi_T$ is the demand for goods of the shop at times $t = 1, \dots, T$;
- x_{t-1}, $t = 1, \dots, T$ is an order of goods one period ahead;
- $1 - l_t$ is a storage loss;
- rapid orders are possible for a price of $u_t > 1$ per piece;
- The selling price is s_t $(s_t > 1)$ and the final inventory K_T has a value $l_T K_T$.

$$\mathbb{E}[\sum_{t=1}^{T}(s_t\xi_t - x_{t-1} - u_t M_t) + l_T K_T] \to \max_x$$

$$\textit{subject to } x_t \lhd \mathcal{F}_t, \; t = 1, \dots, T,$$

$$l_{t-1}K_{t-1} + x_{t-1} - \xi_t = K_t - M_t.$$

$$K_t \geq 0, M_t \geq 0.$$

Inventory Control Problem

If we consider the demand $\xi \sim ln\mathcal{N}(\mu, C)$, we are able to calculate the exact analytical solution (violet color in the figure):

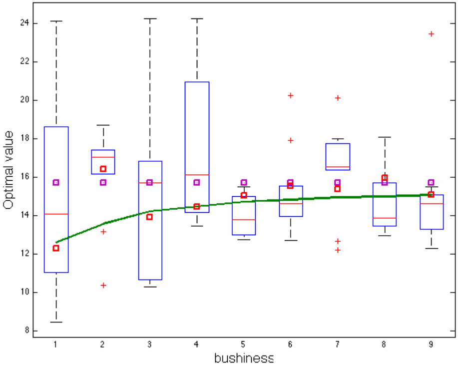


Figure: Inventory Control Problem.

Natural Disasters Risk-Management (IIASA)

Fundamentals

Governments of countries with high risk of natural hazard and low risk capital are to decide how much to spend for protection of the catastrophic event and how much for recovery after it.

Multi-period approach arises because of the uncertainty about the recurrence of the catastrophic event and the risk of zero risk capital at this point.

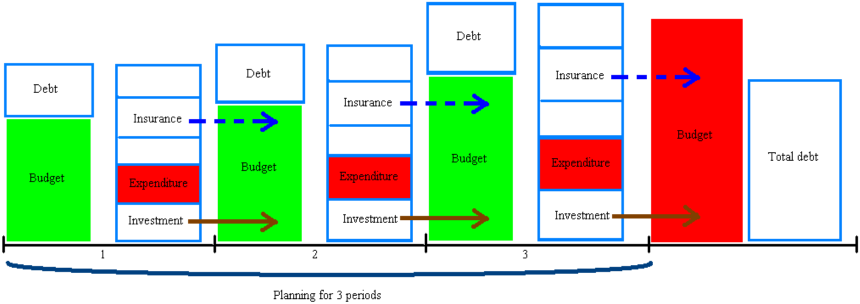
Multi-hazard approach appears as there might be the dependence of one catastrophic event on another, i.e. one catastrophic event occurs as a result of another with some probability.

Multi-period approach

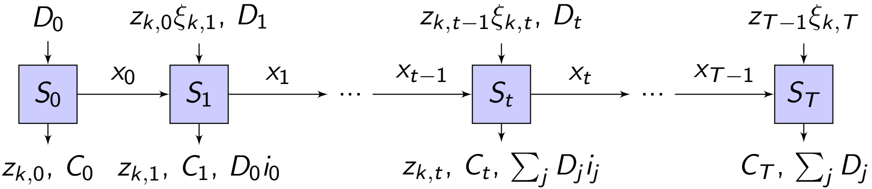
Multi-stage stochastic programs are the standard and well-established tool to support decision making under uncertainty.
The goal is

1. To construct the multi-stage decision model for adaptation and mitigation of natural hazard events;
2. To specify loss distributions for natural hazard events (type and parameters);
3. To approximate these distributions by discrete ones;
4. To construct a surrogate finite dimensional problem which is accessible to computer solution;
5. To solve the surrogate problem and to study the quality of approximation and to infer policy implications.

Risk-management of CAT-events



Problem description



Risk-neutral:

$$(1-\alpha)\sum_{t=1}^{T}\rho^{-t}\mathbb{E}(c_t)+\alpha\mathbb{E}(S_T)\to\max_{d_t,c_t,z_t,x_t}$$

Risk-averse:

$$(1-\alpha)\sum_{t=1}^{T}\rho^{-t}\mathbb{E}(u(c_t))+\alpha\mathbb{E}(u(S_T))\to\max_{d_t,c_t,z_t,x_t},\ \textit{with}\ u(x)=\frac{x^{\gamma}-1}{\gamma}$$

Types of loss distributions

Pareto Heavy-tail distribution (that allows large deviation of losses) corresponds to the situation when natural hazard can be of different power.

Lognormal This type of loss distribution corresponds nicely to the situation when the variable represents the compound loss from a sequence of many natural hazard events.

Generating v.s. Quantization

Suppose that the distribution of the random variable ξ that describes losses in case of natural hazard is known. Then we can generate losses and their probabilities by

- Monte-Carlo random sampling;
- Choice of scenarios by optimal discretization;

and see how the optimal value and optimal decision change when we increase number of possible scenarios.

CAT optimal solution

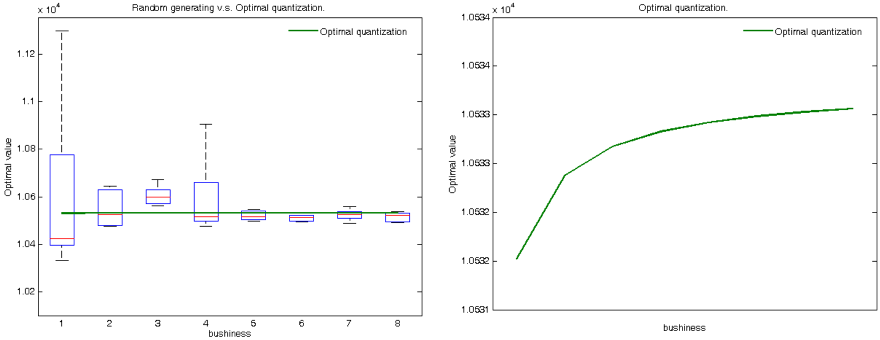


Figure: Stage-wise optimal quantization v.s. random generation

CAT-decisions

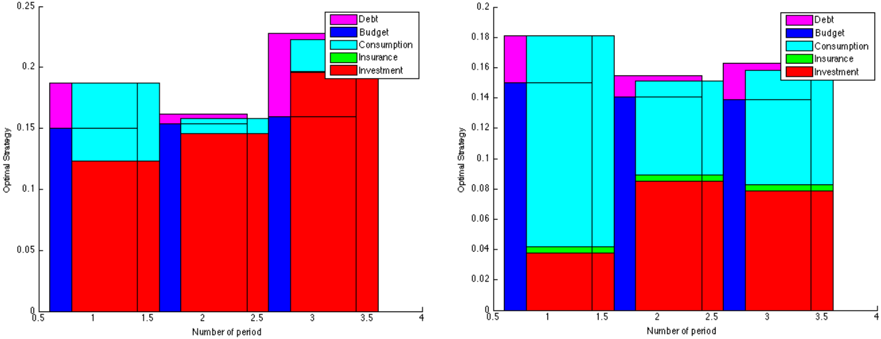


Figure: Mongolia and Mexico optimal strategy for 3 years

Thank you for your attention!

- G. Ch. Pflug, A. Pichler. *A distance for multi-stage stochastic optimization models.* SIAM Journal on Optimization 22, pp. 1-23, 2012.
- G. Ch. Pflug, A. Pichler. *Approximations for probability distributions and stochastic optimization problems.* Int. Series in OR and Management Science 163, Chapter 15, pp. 343-387, 2011.
- G. Ch. Pflug, W. Römisch. *Modeling, Measuring and Managing Risk.* 301 pages, World Scientific Publishing, 2007. ISBN 978-981-270-740-6.
- W. Römisch. *Scenario generation.* Wiley Encyclopedia of Operations Research and Management Science (J.J. Cochran ed.), Wiley, 2010.
- A.V. Timonina. *Multi-stage Stochastic Optimization: the Distance between Stochastic Scenario Processes.* Springer-Verlag Berlin Heidelberg, 2013. DOI 10.1007/s10287-013-0185-3.