universität wien

DISSERTATION

Titel der Dissertation

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

verfasst von

Anna Timonina

angestrebter akademischer Grad

Doctor of Philosophy (PhD)

Wien, 2014

Studienkennzahl lt. Studienblatt:	A 794 370 136
Dissertationsgebiet lt. Studienblatt:	Statistik und Operations Research
Betreuerin / Betreuer:	o.Univ.-Prof. Mag. Dr. Georg Pflug

Contents

Acknowledgements

In 2010 I was very lucky to get the PraeDoc position at the Institute for Statistics and Operations Research of the University of Vienna and to start working on my PhD thesis. The decision to pursue scientific research after completion of my M.Sc. degree was very right for me and I am happy that it was made! Conducting research at the University of Vienna is a very fascinating, challenging and an unforgettable task, and I am very grateful to my supervisor, o.Univ.Prof. Dr. Georg Ch. Pflug, for his continuous guidance, inspirational advices and irreplaceable support all the time!

In 2011 Prof. Pflug gave me his recommendation for the Young Scientist Summer Program at the International Institute for Applied Systems Analysis (IIASA), where I met the incredible RPV team, who I would like to thank for their beneficial assessment during the development of applications in the field of catastrophic risk-management and for the unique experience working with them.

At the same time, words can not express my gratefulness to Prof. Dr. Boris Polyak, who was my supervisor at the Moscow Institute for Physics and Technology (State University) during my B.Sc. and M.Sc. studies and who initially brough my interests to scietific research and, especially, to the areas of applied mathematics, stochastic and robust optimization, statistics, computer sciences.

Yours Sincerely,
Anna Timonina

Abstract

Approximation techniques are challenging, important and very often irreplaceable solution methods for multi-stage stochastic optimization programs. Applications for scenario process approximation include financial and investment planning, inventory control, energy production and trading, electricity generation planning, pension fund management, supply chain management and similar fields. In multi-stage stochastic optimization problems the amount of stage-wise available information is crucial. While some authors deal with filtration distances, in this paper we consider the concepts of nested distributions and their distances which allows to keep the setup purely distributional but at the same time to introduce information and information constraints. Also we introduce the distance between stochastic process and a tree and we generalize the concept of nested distance for the case of infinite trees, i.e. for the case of two stochastic processes given by their continuous distributions. We are making a step towards to a new method for distribution quantization that is the most suitable for multi-stage stochastic optimization programs as it takes into account both the stochastic process and the stage-wise information.
The main problems that are considered in this thesis are:

Scenario generation The stage-wise minimization of the Kantorovich distance gives us a well-known result for the optimal quantizers of the distributions at each stage. However, this result does not mean that the nested distance between initial and approximate problems is minimized. The aim of this thesis is to show that the quantizers received by the minimization of the nested distance are better (in the sense of minimal distance) than the quantizers received by the stage-wise minimization of the Kantorovich distance.

Computational efficiency When dealing with numerical calculation of the nested distance the question of the computational efficiency is of interest. It is necessary to understand how many values should the scenario process have at each stage in order to make the computational time small and at the same time to make the approximation good (i.e. to make the approximation error small). A compromise for this should be found.

Applications Stochastic programming offers a huge variety of applications. In the dissertation we would like to focus on the applications in the field of natural hazards risk-management.

Zusammenfassung

Approximationstechniken sind herausfordernde, wichtige und oft unersetzliche Loesungsmethoden fuer mehrstufige Stochastische Optimierungsprobleme. Anwendungen der Approximierung durch Szenario-Prozesse sind unter anderem Finanz- und Investitionsplanung, Energieproduktion und -handel, Supply-Chain-Management sowie aehnliche Gebiete. Ein wesentlicher Bestandteil von mehrstufigen stochastischen Optimierungsproblemen, ist der Umfang der je Stufe zu verfuegbaren Information. Waehrend einige Autoren diesen Aspekt ueber Filtrations Distanzen beruecksichtigen, verwenden wir in dieser Arbeit das Konzept von der nested-distributions und deren Distanzen. Diese Herangehensweise erlaubt uns einen rein verteilungsbasierten Zugang zu waehlen und dabei gleichzeitig die stufenweise Informationsenthuellung und Nebenbedingungen an diese Information einzufuehren.
Wir fuehren die Distanz zwischen einem stochastischen Prozess und einem Baum ein und verallgemeinern das Konzept der nested-distance fuer den Fall von unendlichen Baeumen, d.h. fuer den Fall von zwei stochastischen Prozessen gegeben durch ihre stetigen Verteilungen. Eine neue Methode zur Verteilungs-Quantisierung wird eingefuehrt. Diese beruecksichtigt sowohl den stochastischen Prozess, als auch die Information je Stufe. Die zentralen Probleme, die in dieser Arbeit behandelt werden, sind:

Szenariogenerierung Die stufenweise Minimierung der Kantorovich Distanz liefert uns ein wohlbekanntes Ergebnis fuer die optimale Quantisierung der Verteilungen in jeder Stufe. Allerdings bedeutet dieses Resultat nicht dass die nested-distance zwischen den urspruenglichen und den approximierten Problemen minimiert wird. Das Ziel dieser Arbeit ist zu zeigen, dass die Quantizer, die aus der Minimierung der nested-distance stammen, besser sind (im Sinn der minimalen Distanz) als die Quantizer aus der stufenweisen Minimierung der Kantorovich Distanz.

Berechnungseffizienz Bei der numerischen Ermittlung der nested-distance ist die Frage der Effizienz fuer die Berechnung von Interesse. Der Grad der Knoten (bushiness) des Szenarienbaumes ist entscheidend fuer die Qualitaet der Approximation und die Effizienz der Berechnung. Ein Kompromiss sollte dafuer gefunden werden.

Anwendungen Stochastische Optimierung hat eine grosse Zahl von Anwendungen. In dieser Dissertation konzentrieren wir uns auf die Anwendungen in den Bereichen des Risikomanagements von Naturgewalten.

Chapter 1

Introduction

1.1 Problem description

Suppose that a multi-stage stochastic optimization program (1.1) is given in the form with one loss/profit function $H(\cdot)$ [26, 24, 27, 28]:

$$\max_x \{\mathcal{A}[H(x,\xi);\mathcal{F}] : x \triangleleft \mathcal{F}, x \in \mathbb{X}\}, \tag{1.1}$$

where

- $\xi = (\xi_1, ..., \xi_T)$ is a stochastic process defined on the probability space $(\Omega, \mathcal{F}, P)$ and $\xi^t = (\xi_1, ..., \xi_t)$ is its history up to time t;
- $\mathcal{F} = (\mathcal{F}_1, ..., \mathcal{F}_T)$ is a filtration on the $(\Omega, \mathcal{F}, P)$ to which the process ξ is adapted (i.e. ξ_t is measurable with respect to σ-algebra $\mathcal{F}_t$, $\forall t = 1, ..., T$): we denote it as $\xi \triangleleft \mathcal{F}$, i.e. $\xi_t \triangleleft \mathcal{F}_t$, $\forall t = 1, ..., T$. The fact that filtration can be defined only on a particular probability space leads to the question of how to introduce information and information constraints, keeping the setup purely distributional, i.e. without any reference to a probability space;
- We add the trivial σ-algebra $\mathcal{F}_0 = \{\emptyset, \Omega\}$ as the first element of the filtration $\mathcal{F}$. A sequence of decisions $x = (x_0, ..., x_{T-1})$ with history $x^t = (x_0, ..., x_{t-1})$ must be adapted to $\mathcal{F}$: i.e. it must fulfill the non-anticipativity conditions $x \triangleleft \mathcal{F}$ [24, 27, 28], that means only those decisions are feasible, which are based on the information available at the particular time;
- $H(x, \xi)$ is the loss/profit functions;
- $\mathbb{X}$ is the set of constraints other than the non-anticipativity constraints;
- $\mathcal{A}(\cdot, \cdot)$ is a version-independent (also called law-invariant) multi-period acceptability functional (see [24, 27, 28]), i.e. it is defined on the probability space $(\Omega, \mathcal{F}, P)$ and depends only on the distribution of the outcome variable $H(x, \xi)$.

Our goal is to solve this problem numerically by the approximation though in general it may be difficult to solve it explicitly due to its functional form. So the necessity of approximation of the initial problem with a solvable one (1.2) arises immediately. The approximated problem can be written correspondingly in the form with the profit/loss function $H(\cdot)$:

$$\max_x \{\mathcal{A}[H(\widetilde{x}, \widetilde{\xi}); \widetilde{\mathcal{F}}] : \widetilde{x} \triangleleft \widetilde{\mathcal{F}}, \widetilde{x} \in \mathbb{X}\}, . \tag{1.2}$$

where the stochastic process ξ is replaced by a scenario process with finite number of values $\widetilde{\xi} = (\widetilde{\xi}_1, ..., \widetilde{\xi}_T)$ defined on a probability space $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{P})$: it is not natural to construct approximations on the same probability space, because concrete decision problems of type (1.1) are given without reference to a probability space, they are given in distributional setups [19, 18, 24, 26, 27, 28].
The distance between problems (1.1) and (1.2) determines approximation error. Previously, distances between the initial problem (1.1) and its approximation (1.2) have been defined only if both processes ξ and $\widetilde{\xi}$ and both filtrations $\mathcal{F}$ and $\widetilde{\mathcal{F}}$ are defined on the same probability space $(\Omega, \mathcal{F}, P)$. The introduction of the concept of *nested distribution* [24, 27, 28], which contains in one mathematical object the scenario values as well as the information structure under which decisions have to be made, brings the problem to the

purely distributional setup. The nested distance between these distributions is introduced in [27, 28] and turns out to be a generalization of the well-known Kantorovich distance for single-stage problems (see [15, 27, 28, 37] for more details).
The main problems that are considered in this thesis are:

Scenario generation When dealing with scenario generation for the multi-stage stochastic optimization problem, the natural question that arises is how to approximate (to quantize) the continuous distribution of the stochastic process in such a way, that the nested distance between the initial distribution and its finite multi-stage approximation is minimized. The stage-wise minimization of the Kantorovich distance gives us a well-known result for the optimal quantizers of the distributions at each stage [15, 25, 28, 37, 38]. However, this result does not mean that the nested distance between initial and approximate problems is minimized. The aim of this thesis is to show that the quantizers received by the minimization of the nested distance are better (in the sense of minimal distance) than the quantizers received by the stage-wise minimization of the Kantorovich distance.

Computational efficiency When dealing with numerical calculation of the nested distance the question of the computational efficiency is of interest. It is necessary to understand how many values should the scenario process have at each stage in order to make the computational time small and at the same time to make the approximation good (i.e. to make the approximation error small). A compromise for this should be found.

Applications Stochastic programming offers a huge variaty of applications. In the dissertation we would like to focus on the applications in the field of natural hazards risk-management.

1.2 Preliminaries

Consider a stochastic process $\xi = (\xi_1, ..., \xi_T)$ on some probability space $(\Omega, \mathcal{F}, P)$. The σ-algebra generated by the random variable ξ_t is denoted by $\sigma(\xi_t)$, the filtration generated by the process $(\xi_1, ..., \xi_T)$ is denoted by $\sigma(\xi) = (\sigma(\xi_1), \sigma(\xi_1, \xi_2), ..., \sigma(\xi_1, \xi_2, ..., \xi_T))$.
If random variable ξ_t is measurable w.r.t. the σ-algebra $\mathcal{F}_t$, we write $\xi_t \triangleleft \mathcal{F}_t$. The similar notation $\xi \triangleleft \mathcal{F}$ we use to indicate that the process $\xi = (\xi_1, ..., \xi_T)$ is adapted to the filtration $\mathcal{F} = (\mathcal{F}_1, ..., \mathcal{F}_T)$.
In our distributional setup, there is no predefined probability space and therefore there is no room for introducing other filtration than the one which is generated by the random process itself (see [26, 27, 28]). Hence, we can make the following assumption:
At time t, no other information is available to the decision-maker rather than the values of the scenario process $\xi_s,\ s \leq t,\ \forall t = 1, ..., T$. In different words we assume that the decision x_{t-1} is measurable with respect to the σ-algebra $\mathcal{F}_t$, which is the one generated by $(\xi_1, ..., \xi_T)$.

Definition 1.1 (**Tree process**). *A stochastic process $\nu = (\nu_1, ..., \nu_T)$ is called a tree process (see [24, 27, 32]), if $\sigma(\nu_1), \sigma(\nu_2), ..., \sigma(\nu_T)$ is a filtration.*

Notice, that considering the history process of any stochastic process $\xi = (\xi_1, \xi_2, ..., \xi_T)$, it is possible to transform it into a tree process:

$$\begin{aligned} \nu_1 &= \xi_1 \\ \nu_2 &= (\xi_1, \xi_2) \\ &\vdots \\ \nu_T &= (\xi_1, \xi_2, ..., \xi_T). \end{aligned}$$

Assuming now that all the state spaces of all tree process are Polish, we can factorize the joint distribution P of $\nu = (\nu_1, ..., \nu_T)$ into the chain of conditional distributions [24, 27]:

$$\begin{aligned} P_1(A) &= P(\nu_1 \in A) \\ P_2(A|u_1) &= P(\nu_2 \in A|\nu_1 = u_1) \\ &\vdots \\ P_t(A|u_1, ..., u_{t-1}) &= P(\nu_t \in A|\nu_1 = u_1, ..., \nu_{t-1} = u_{t-1}) \\ &\vdots \\ P_T(A|u_1, ..., u_{T-1}) &= P(\nu_T \in A|\nu_1 = u_1, ..., \nu_{T-1} = u_{T-1}), \end{aligned}$$

where A is a measurable set, in which values of the stochastic process $\nu = (\nu_1, ..., \nu_T)$ are lying with some particular probability; $u = (u_1, u_2, ..., u_T)$ is the realization of the stochastic process $\nu = (\nu_1, ..., \nu_T)$.
In this case, we write the chain of conditional probabilities in the following form:

$$P = P_1 \circ P_2 \circ ... \circ P_T.$$

Definition 1.2 (**Equivalence of tree processes**). *Let ν and $\widetilde{\nu}$ be two tree processes defined on possibly different probability spaces $(\Omega, \mathcal{F}, P)$ and $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{P})$ correspondingly. Let $P = P_1 \circ P_2 \circ ... \circ P_T$ and $\widetilde{P} = \widetilde{P}_1 \circ \widetilde{P}_2 \circ ... \circ \widetilde{P}_T$ be the pertaining chains of conditional probabilities. The tree processes ν and $\widetilde{\nu}$ are called equivalent (see [25]) if the following properties hold:*

1. *There are bijective functions* $k_1, ..., k_T$ *that map the state spaces of* ν *into the state spaces of* $\widetilde{\nu}$*;*

2. *The distributions of two tree processes* $\nu = (\nu_1, ..., \nu_T)$ *and* $k^{-1}(\widetilde{\nu}) = (k_1^{-1}(\widetilde{\nu_1}), ..., k_T^{-1}(\widetilde{\nu_T}))$ *coincide.*

Notice, that for tree processes defined on the same probability space, one could define equivalence simply as $\nu_t = k_t(\widetilde{\nu_t})\ \forall t = 1, ..., T$.
The random component $\xi = (\xi_1, ..., \xi_T)$ of the decision problem (1.1) is described by the continuous distribution function, while the random component $\widetilde{\xi} = (\widetilde{\xi_1}, ..., \widetilde{\xi_T})$ of the approximate decision problem (1.2) is discrete, i.e. it takes finite (and in fact few) number of values. Since every finitely valued stochastic process $(\widetilde{\xi_1}, ..., \widetilde{\xi_T})$ is representable as a tree, we deal with tree approximation of stochastic processes [24, 25, 27]. The tree that represents the finitely valued stochastic process $(\widetilde{\xi_1}, ..., \widetilde{\xi_T})$ we call *finitely valued tree.*
To solve the approximate problem (1.2) numerically we should represent the stochastic process $\widetilde{\xi}$ as a finitely valued tree. The tree can be determined by its graph (tree) structure, the scenario probabilities and by the values sitting on the nodes.
The **structure** of the tree could be given by the vector of predecessors *pred* or in the simplest case by the vector *bush* that describes *bushiness* of the tree at each stage.

Definition 1.3 (Bushiness). *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 (see [35]). Values* $b_1, b_2, ..., b_T$ *are called bushiness factors.*

Fig. 1.1 shows two trees with different bushiness factors (the bushiness vector of the left-hand side tree is [5 5 5], the bushiness vector of the right-hand side tree is [2 2 2]; hence, bushiness factor of the left-hand side tree is 5, while bushiness factor of the right-hand side tree is 2).

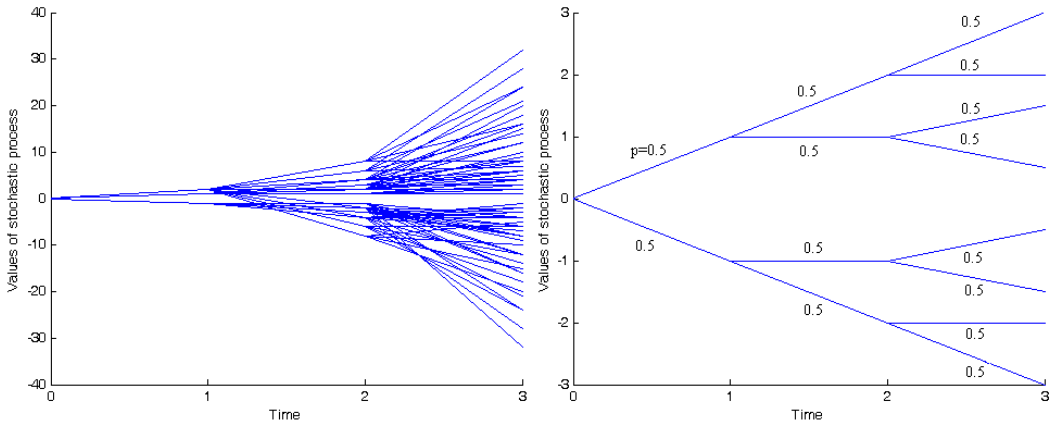


Figure 1.1: Comparison of the trees with different bushiness (left: bush=[5 5 5]; right: bush=[2 2 2]).

1.2.1 Encoding in MatLab

All numerical experiments of the thesis are perfomed in MatLab. The following variables and coding notations are used for the calculations:

The **scenario probabilities** are given by the vector *pScen* of probabilities or they can be obtained from the probabilities of each node generated according to the given distribution. Length of the vector of scenario probabilities equals the numbers of terminal nodes, i.e. the number of scenarios.
The **values** could be given by the matrix *val* or could be generated randomly or optimally according to some known distribution (the methods for scenario generation are described in [7, 8, 11, 17, 27, 29, 32] etc.). The number of columns equals to the number of nodes including the root; the number of rows is the dimension of the node information.
To summarize all input arguments:

pred	[1 by $N_{\mathbb{T}}$]	the predecessor vector
pScen	[1 by N_T]	the terminal probabilities (sum up to 1)
val	[D by $N_{\mathbb{T}}$]	the value matrix

where $N_{\mathbb{T}}$ is the number of nodes of the tree $\mathbb{T}$ including the root; T is the height of the tree; N_T is the number of scenarios (i.e. number of nodes at the stage T); D is the dimension of the node information.

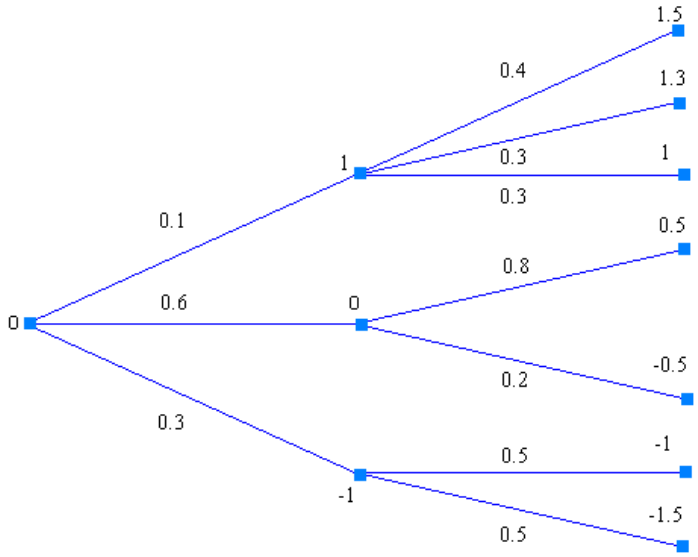


Figure 1.2: Sample tree (Tree 1).

For example, the tree in Fig. 1.2 can be described by the following vectors:
pred=[0 1 1 1 2 2 2 3 3 4 4],
pScen=[0.04 0.03 0.03 0.48 0.12 0.15 0.15],
val=[0 1 0 -1 1.5 1.3 1 0.5 -0.5 -1 -1.5].
Clearly, the unconditional probability of each node of the tree can be calculated backwards from the scenario probabilities: starting with the stage T, one should find the set of nodes $\mathcal{N} \in \mathcal{N}_T$ with the same predecessor $n \in \mathcal{N}_{T-1}$ and sum up the unconditional probabilities of each node in $\mathcal{N}$ ($\mathcal{N}_t$ is the set of nodes at stage t of the tree $\forall t = 1, ..., T$).

For the tree in Fig. 1.2 the uncoditional probabilities of the nodes are
p=[0.1 0.6 0.3 0.04 0.03 0.03 0.48 0.12 0.15 0.15].
At the same time, conditional probabilities $pNodes$ can be recursively obtained from the chain rule (Fig. 1.2):
$pNodes$=[0.1 0.6 0.3 0.4 0.3 0.3 0.8 0.2 0.5 0.5].

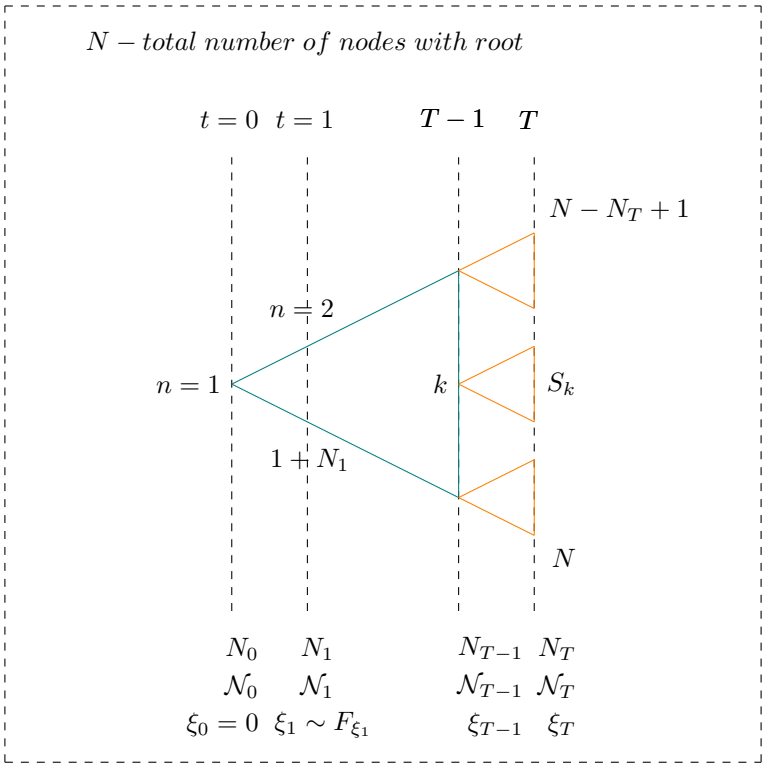


Figure 1.3: Notations sitting on a tree of height T.

In this thesis we distinguish between two types of approximations:

Approximation Type 1 is focused on the approximation of the continuous-state stochastic process by the tree (**Tree 1**) with a fixed tree structure (that is given either by the vector of predecessors *pred* or by the bushiness vector *bush*). If Tree 1 is given, we are not going to change its graph (tree) structure, it is constant for a particular problem (even we may already assume that the probabilities p and values z are assigned to the nodes).
We use the following notations refering to the tree of Type 1:

$\mathbb{T}_z$ is a tree structure $\mathbb{T}$ together with values z sitting on it;

$\mathbb{T}_p$ is a tree structure $\mathbb{T}$ together with probabilities p sitting on it;

$\mathbb{T}_{(z,p)}$ is a tree structure $\mathbb{T}$ together with values z and probabilities p sitting on it;

This type of the approximation is good for the development of new algorithms of scenario generation. In this case the stages of the tree are likely to be considered separately and the error of approximation assumed to be the sum of errors at each stage.
If the tree structure is not optimal, the disadvantage of the approximation of this type is that it hardly takes into account all the properties of stochastic processes.

Approximation Type 2 is focused on the approximation of the continuous-state stochastic process by the tree (**Tree 2**) with changeable tree structure. This approximation type is going to be used for a better approximation of the stochastic process $\xi = (\xi_1, ..., \xi_T)$ by the tree. The structure of Tree 2 is described by the bushiness vector $bush$ that may change. We use the following notations refering to the tree of Type 2:

$\mathbb{T}$ is a changeable tree structure (vector of predecessors or bushiness);

$\mathbb{T}^*$ is a tree structure of a very high bushiness, s.t. it well represents continuous distribution functions;

Combining these two approximation types, one can develop new methods and algorithms for the approximate solution of the multi-stage stochastic optimization problems.
Fig. 1.3 shows main notations used for the description of a tree of height T.

$N_{\mathbb{T}}$ is a number of nodes of the tree with root;

N_t is a number of nodes at the stage t of the tree $\forall t = 1, ..., T$;

$\mathcal{N}_t$ is a set of all node indices n at the stage t of the tree $\forall t = 1, ..., T$;

$\xi = (\xi_1, ..., \xi_T)$ is the stochastic process sitting on the tree;

F_{ξ_t} is a continuous probability distribution of the individual variable ξ_t.

Chapter 2

Distances between stochastic scenario processes

When are two stochastic scenario processes close to each other? To answer this question we should follow the track of historical distance development: from the distance between two points to the distance between two trees.
First of all, one should be able to calculate the distance between two points in $\mathbb{R}^D$-Euclidean space. Suppose that $x \in \mathbb{R}^D$ and $y \in \mathbb{R}^D$. The distance between this two points can be defined in the following ways:

1-norm distance $d_1 = \sum_{i=1}^{D} |x_i - y_i|$;

r-norm distance $d_r = \left(\sum_{i=1}^{D} |x_i - y_i|^r \right)^{\frac{1}{r}}$ with $1 \leq r \leq \infty$;

∞-norm distance $d_\infty = \max_i |x_i - y_i|$.

Hence, the distance between realizations of two $\mathbb{R}^D$-valued random variables X and Y, defined on some common probability space $(\Omega, \mathcal{F}, P)$ is given by $d_r(X(\omega), Y(\omega))$.
If we do not know the realizations of two random variables but measurable functions X and Y, mapping Ω into $\mathbb{R}^D$, then the candidate for measuring the L_r distance between these two random variables would be:

$$d_r(X, Y) = \int_\Omega d_r(X(\omega), Y(\omega)) P(d\omega).$$

This distance holds only for those two random variables X and Y that are defined on the same probability space $(\Omega, \mathcal{F}, P)$. Moreover, it is usual that the only information we have about random variables is their distribution functions and in this case it is not clear from the first sight how to define the distance. Hence, two natural questions arise: How to generalize the given distance to the case of random variables on different probability spaces $(\Omega_1, \mathcal{F}_1, P_1)$ and $(\Omega_2, \mathcal{F}_2, P_2)$? And how to find the distance between random variables given by their distribution functions only?

2.1 Distance between probability measures

To get an intuitive understanding of the distance between probability measures let us start with a description of the *Monge Transportation Problem* (see [20]).
Initially, the problem was formulated in the following way [20, 29, 31, 37, 38]:
Split two equally large volumes into small particles and then associate them with each other so that the sum of products of the paths along which the particles are transported (i.e. the distances between associated particles) to the volume of the transported particle is least. Along what paths must the particles be transported and what is the smallest transportation cost?
Denote the initial and final volumes by Ω and $\widetilde{\Omega}$ respectively and notice that in general it is not necessary to assume them to be of equal volumes rather they are bodies with equal masses though with not necessarily uniform densities. Let $\pi(\cdot,\cdot)$ be the probability measure on $\Omega \times \widetilde{\Omega}$ describing the shipping plan. Then the fraction of volume of Ω that is transported from w into $\widetilde{w}$ is defined by $\pi(w,\widetilde{w})$ for any $w \subset \Omega$ and $\widetilde{w} \subset \widetilde{\Omega}$. According to the constrains of the problem, $\pi(w,\widetilde{\Omega}) = \frac{Volume(w)}{Volume(\Omega)}$ and $\pi(\Omega,\widetilde{w}) = \frac{Volume(\widetilde{w})}{Volume(\widetilde{\Omega})}$. The projections of $\pi(\cdot,\cdot)$ on the first and the second coordinates, denoted by $P(\cdot)$ and $\widetilde{P}(\cdot)$, describe masses of Ω and $\widetilde{\Omega}$ respectively:

$$\begin{aligned} \pi(\Omega,\cdot) &= \widetilde{P}(\cdot), \\ \pi(\cdot,\widetilde{\Omega}) &= P(\cdot). \end{aligned}$$

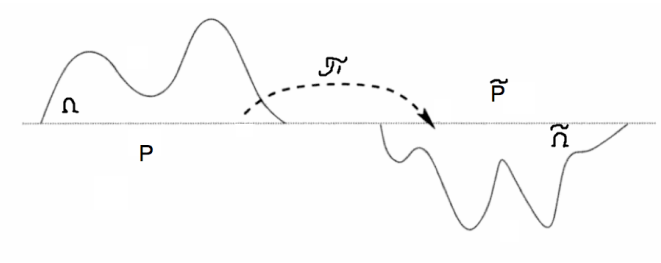


Figure 2.1: Monge transportation problem.

If we denote the cost of the transportation from w to $\widetilde{w}$ by $d(w,\widetilde{w})$ then the total cost is

$$\int_{\Omega\times\widetilde{\Omega}} d(w,\widetilde{w})\pi(dw,d\widetilde{w}),$$

where $\pi(dw,d\widetilde{w})$ is the mass shipped from the neighborhood of w to the neighborhood of $\widetilde{w}$.
Hence, we obtain the so-called *Kantorovich formulation of the Monge problem* [37, 38]:
Suppose that P and $\widetilde{P}$ are two Borel probability measures given on separable metric spaces (Ω, d_Ω) and $(\widetilde{\Omega}, d_{\widetilde{\Omega}})$ respectively and $\mathcal{P}^{(P,\widetilde{P})}$ is the space of all Borel probability measures π on $\Omega \times \widetilde{\Omega}$ with fixed marginals $P(\cdot) = \pi(\cdot,\widetilde{\Omega})$ and $\widetilde{P}(\cdot) = \pi(\Omega,\cdot)$. Evaluate the functional

$$\mathcal{D}(P,\widetilde{P}) = \inf_{\pi}\{\int_{\Omega\times\widetilde{\Omega}} d(w,\widetilde{w})\pi(dw,d\widetilde{w}) : \pi \in \mathcal{P}^{(P,\widetilde{P})}\},$$

where $d(w,\widetilde{w})$ is the cost function for the transportation of $w \in \Omega$ to $\widetilde{w} \in \widetilde{\Omega}$.

Definition 2.1 (**Kantorovich (multivariate) Distance**). *The Kantorovich distance [15, 37, 38] between measures can be defined in the following way:*

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

where $d(w,\widetilde{w})$ is a slight generalization of the distance between w and $\widetilde{w}$ to the case of different state spaces Ω and $\widetilde{\Omega}$.

The Kantorovich distance satisfies the following properties:

- **Nonnegativity:** $d_{KA}(P,\widetilde{P}) \geq 0$;
- **Triangle inequality:** $d_{KA}(\widetilde{P}_1,\widetilde{P}_2) \leq d_{KA}(P,\widetilde{P}_1) + d_{KA}(P,\widetilde{P}_2)$;
- **Strictness:** $d_{KA}(\widetilde{P}_1,\widetilde{P}_2) = 0 \Longleftrightarrow \widetilde{P}_1 = \widetilde{P}_2$.

Definition 2.2 (**Wasserstein Distance**). *The Wasserstein distance [37, 38] between measures is a generalization of the Kantorovich distance for order $r \geq 1$:*

$$\begin{aligned} d_{WA_r}(P,\widetilde{P}) &= \inf_{\pi}\{\int_{\Omega\times\widetilde{\Omega}} d(w,\widetilde{w})^r\pi[dw,d\widetilde{w}]\}^{\frac{1}{r}},\\ \pi[\cdot\times\widetilde{\Omega}] &= P(\cdot),\\ \pi[\Omega\times\cdot] &= \widetilde{P}(\cdot). \end{aligned}$$

Example 2.1. *Consider two discrete measures $\widetilde{P}_1 = \sum_{n_1=1}^{N_1} p_{n_1}\delta_{z_{n_1}}$ and $\widetilde{P}_2 = \sum_{n_2=1}^{N_2} p_{n_2}\delta_{z'_{n_2}}$ sitting on N_1 and N_2 points respectively. The distance between these measures can be found by solving the following linear problem:*

$$\begin{aligned} \min_{\pi} \sum_{n_1,n_2} \pi_{n_1,n_2} \quad & d_{n_1,n_2},\\ s.t. \sum_{n_2} \pi_{n_1,n_2} \quad &= \quad p_{n_1},\\ \sum_{n_1} \pi_{n_1,n_2} \quad &= \quad p_{n_2},\\ \pi_{n_1,n_2} \quad &\geq \quad 0, \end{aligned}$$

where $d_{n_1,n_2} = d(z_{n_1}, z'_{n_2})$ is a matrix that represents costs related to the transportation of masses from z_{n_1} to z'_{n_2}.

Example 2.2. *Consider two random variables: $\xi \sim P$ and $\widetilde{\xi}$ which takes only finitely many values $\widetilde{\xi} \in \{z_1,...,z_N\}$ (Fig. 2.2). Given the supporting points z_n, $n = 1,...,N$, we calculate probabilities p_n, $n = 1,...,N$ by the minimization of the Wasserstein distance $d_{WA_r}(P, \sum_{n=1}^N p_n\delta_{z_n})$, $r \geq 1$. The distance $d_{WA_r}(P, \sum_{n=1}^N p_n\delta_{z_n})$ between measures P and $\sum_{n=1}^N p_n\delta_{z_n}$ can be found using the following formula:*

$$d_{WA_r}(P, \sum_{n=1}^N p_n\delta_{z_n}) = \inf_{p_n}\Big(\inf\{\mathbb{E}[d(\xi,\widetilde{\xi})] :\ \xi\sim P,\ \widetilde{\xi}\in\{z_1,...,z_N\}\}\Big) =$$

$$= \int \min_n d(u,z_n)^r dP(u).$$

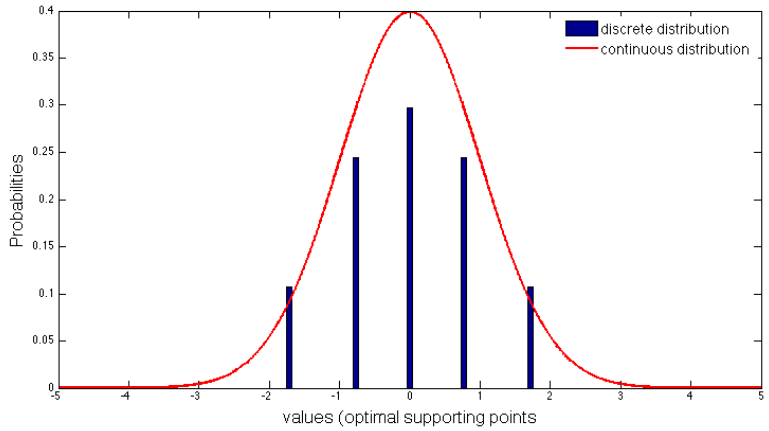


Figure 2.2: Continuous v.s. discrete distribution with minimal distance between each other.

Lemma 2.1. *Consider probability measure P sitting on N points $\xi \in Z = \{z_1, z_2, ..., z_N\}$, i.e*

$$P = \sum_{n=1}^{N} p_n \delta_{z_n}.$$

Split the probabilities p_n, $\forall n = 1, ..., N$, so that

$$\widetilde{P} = \sum_{n=1}^{N} \left(\sum_{m=M_{n-1}+1}^{M_n} \widetilde{p}_m \right) \delta_{z_n},$$

where $M_0 = 0$, $M_N = M \leq N$, $M_n - M_{n-1} \geq 1$, $\forall n = 1, ..., N$ and

$$\sum_{m=M_{n-1}+1}^{M_n} \widetilde{p}_m = p_n, \ \forall n = 1, ..., N.$$

Then the optimal solutions of problems (2.1) and (2.2) are equal to zero:

$$d_{WA_r}(P, \widetilde{P}) = d_{KA}(P, \widetilde{P}) = 0.$$

Remark 2.1. *Notice, that*

1. *the split $\widetilde{P}$, that satisfies conditions of the Lemma, is not unique for the measure P, and that probability measures P and $\widetilde{P}$ are equal;*

2. *$\widetilde{P}$ is the probability measure as well as P, because the set of points Z is not changed and $\sum_{m=M_{n-1}+1}^{M_n} \widetilde{p}_m = p_n$.*

Proof. The proof obviously follows from the definition of the Wasserstein and Kantorovich distances for the same probability measures.
Let us show the proof for the Kantorovich distance. The multivariate distance between measures $P = \sum_{n=1}^{N} p_n \delta_{z_n}$ and $\widetilde{P} = \sum_{m=1}^{M} \widetilde{p}_n \delta_{\widetilde{z}_n}$ can be found by solving the following

linear problem:

$$\begin{aligned} \min_{\pi} \sum_{n,n} \pi_{n,m} \quad & d_{n,m}, \\ s.t. \sum_{m} \pi_{n,m} \quad &= \quad p_n, \\ \sum_{n} \pi_{n,m} \quad &= \quad \widetilde{p}_m, \\ \pi_{n,m} \quad &\geq \quad 0, \end{aligned}$$

where $d_{n,m} = d(z_n, \widetilde{z}_m)$ is the matrix that represents the weighted costs related to the transportation of masses from z_n to $\widetilde{z}_m$.
In our case, according to the conditions of the Lemma, the matrix $d_{n,m},\ n, m = 1, ..., N$ is square and its diagonal contains only zero elements. Hence, the cheapest transportation plan is the one, that transports i to i, as soon as $d_{n,m}(i, i) = 0,\ i = 1, ..., N$. The constraints of the problem with such a transportation plan are clearly satisfied due to the conditions of the Lemma. □

Assume now that we search for the approximation of the continuous distribution function P by a discrete one, that sits only on N points: we call this discrete distribution $\widetilde{P}^N$. The following theorem answers the question how the approximation depends on the number of points N.

Theorem 2.1 (see [10, 28, 40])**.** *Suppose that the D-dimensional distribution P has a density f with*
$\int |u|^{1+\delta} f(u)du < \infty$ *for some $\delta > 0$ and suppose that*

$$[\int f(x_t)^{\frac{D}{D+1}} dx_t]^{\frac{D+1}{D}} \leq c,\ \forall t = 0, ..., T-1$$

where c is some constant.
Then $\exists \widetilde{P}^N$, such that

$$d(P, \widetilde{P}^N) \leq cN^{-\frac{1}{D}},$$

where we denote by $\widetilde{P}^N$ a discrete approximation of P that sits on N points,
and for $N \to \infty$

$$d(P, \widetilde{P}^N) \to 0.$$

Proof. The result of the theorem follows from Zador-Gersho formula (see [10, 28, 40]). □

2.2 Distance between conditional probability measures

The notion of the Kantorovich distance between probability measures can be easily extended to the case of conditional probabilities (also called transition probabilities) [15, 18, 19, 25, 27].
Let P be the probability measure on $\mathbb{R}^{DT}$. It can be dissected into the chain of conditional (transition) probabilities:

$$P = P_1 \circ P_2 \circ ... \circ P_T, \tag{2.1}$$

where P_1 $(t = 1)$ is the unconditional probability on $\mathbb{R}^D$ and P_t, $\forall t = 2, ..., T$ is the conditional probability of time t given the history up to time $t-1$.
Now suppose that $u = (u_1, ..., u_T) \in \mathbb{R}^{DT}$ and denote by $u^t = (u_1, ..., u_t)$ its history up to time t. The probability measure P can be dissected into the chain (2.1) if and only if the following statement holds for the sequence $A_1, ...A_T$ of Borel sets:

$$\begin{aligned} &P(A_1 \times A_2 \times ... \times A_T) = \\ &\int_{A_1}\int_{A_2}...\int_{A_T} P_T(du_T|u^{T-1})...P_t(du_t|u^{t-1})...P_2(du_2|u_1)P_1(du_1). \end{aligned} \tag{2.2}$$

Here, $P_t(A_t|u^{t-1})$ $\forall t = 2, ..., T$ is the conditional probability of $\xi_t \in A_t$ $\forall t = 2, ..., T$ given the history $\xi^{t-1} = u^{t-1}$ $\forall t = 2, ..., T$ and P_1 is the unconditional probability. As soon as we work on the Polish spaces, we can guarantee the existence of conditional probabilities (see [2] (Chapter 4, theorem 1.6), [19, 18]).

Definition 2.3 (**Lipschitz property**). *Let P be a probability on $\mathbb{R}^{DT}$, dissected into the chain of transition probabilities*

$$P = P_1 \circ P_2 \circ ... \circ P_T.$$

We say that the multivariate distribution P has the $(K_2, ..., K_T)$-Lipschitz property, if

$$d_{KA}(P_t(\cdot|u), P_t(\cdot|v)) \leq K_t d(u, v),\ \forall t = 2, ..., T,$$

where $d(u, v)$ is the Euclidean distance between points u and v.

Remark 2.2 (**Zero Lipschitz constant**). *If $K_t = 0$, then $d_{KA}(P_t(\cdot|u), P_t(\cdot|v)) = 0$, i.e. P_t is independent on u and v.*

If P and $\widetilde{P}$ are two transition probabilities on $\mathbb{R}_1 \times \mathcal{F}_2$ and d is a probability distance, we can define

$$\begin{aligned} d_{KA}(P, \widetilde{P}) &= \sup_u d_{KA}(P(\cdot|u), \widetilde{P}(\cdot|u)), \\ d_{WA_r}(P, \widetilde{P}) &= \sup_u d_{WA_r}(P(\cdot|u), \widetilde{P}(\cdot|u)). \end{aligned}$$

2.3 Nested distance between scenario processes

The Kantorovich distance does not use any information about the structure of the stochastic scenario process, that may exist. Suppose that there is a tree structure, on which the stochastic scenario process is sitting and let us consider the stochastic scenario process together with all available information that is introduced by the tree process [24, 27, 28].

Definition 2.4 (**Scenario Process-and-Information Structure**). *If ν is a tree process and $\mathcal{F} = \sigma(\nu)$, then any process ξ adapted to $\mathcal{F}$ can be written as $\xi_t = f_t(\nu_t)$ for some measurable function f_t. We call a pair consisting of a scenario process ξ and a tree process, such that $\xi \lhd \sigma(\nu)$, a scenario process-and-information structure. We denote it as $(\Omega, \mathcal{F}, P, \xi)$.*

The distribution of such a process-and-information structure is called *a nested distribution* $\mathbb{P}$. This concept is explained in [24, 26, 27]. The nested distributions are defined in a pure distributional setup. The relation between the nested distribution $\mathbb{P}$ and the process-and-information-structure $(\Omega, \mathcal{F}, P, \xi)$ is comparable to the relation between a probability measure P on $\mathbb{R}^D$ and a $\mathbb{R}^D$-valued random variable ξ with distribution P.
We may alternatively consider either the nested distribution $\mathbb{P}$ or its realization with Ω being the state space, $\mathcal{F}$ being an information and ξ being a process. We symbolize this fact by $\mathbb{P} \sim (\Omega, \mathcal{F}, P, \xi)$.
To generalize the Kantorovich distance for the multi-stage case let us consider two *process-and-information structures*

$$\mathbb{P} \sim (\Omega, \mathcal{F}, P, \xi)$$

and

$$\widetilde{\mathbb{P}} \sim (\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{P}, \widetilde{\xi}).$$

We intend to minimize the effort or costs that have to be taken into account when passing from one process-and-information structure to another.
To be able to calculate these costs we need [27, 28]:

1. a transportation cost function between Ω and $\widetilde{\Omega}$, that is

$$d(w, \widetilde{w}) := \sum_{t=1}^{T} d_t(\xi_t(w_t), \widetilde{\xi}_t(\widetilde{w}_t)),$$

   where d_t is the distance available in the state space of the processes ξ_t and $\widetilde{\xi}_t$; w_t is a predecessor of w that belongs to the stage t of the tree $\forall t = 1, ..., T$ (in symbol we denote it $w_t = pred_t(w)$, where $w = w_T$). For finitely valued trees with finite number of nodes $N_{\mathbb{T}}$ we will denote by $n_t \in \mathcal{N}_t$ the t-stage predecessor of a node $n_T \in \mathcal{N}_T$ from the last stage of the tree (leave-node) $\forall t = 1, ..., T$.
   We do not consider the term $d_0(\xi_0, \widetilde{\xi}_0)$ at $t = 0$, as soon as we assume the values ξ_0 and $\widetilde{\xi}_0$ are equal at the root of the trees.

2. to take into account the gradually increasing information, provided by the filtrations. As the complete information is available only at the final stage T (with $\mathcal{F}_T$ and $\widetilde{\mathcal{F}}_T$) and the measure π in general is not adapted to the situation of lacking information, we should use filtrations to create a multi-stage distance.

Definition 2.5 (**Nested Distance**). *The multi-stage distance (see [27, 28]) of order* $r \geq 1$ *of two process-and-information structures is*

$$dl_r(\mathbb{P}, \widetilde{\mathbb{P}}) = \inf_{\pi}\Big(\int d(w, \widetilde{w})^r \pi(dw, d\widetilde{w})\Big)^{\frac{1}{r}}, \tag{2.3}$$

subject to

$$\mathbb{P} \sim (\Omega, \mathcal{F}, P, \xi),\ \widetilde{\mathbb{P}} \sim (\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{P}, \widetilde{\xi})$$

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

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

We denote by $dl(\mathbb{P}, \widetilde{\mathbb{P}})$ *the nested distance of order* $r = 1$, *i.e.* $dl_1(\mathbb{P}, \widetilde{\mathbb{P}}) = dl(\mathbb{P}, \widetilde{\mathbb{P}})$.

Consider two nested distributions $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ with corresponding multivariate distributions P and $\widetilde{P}$. Suppose, that these nested distributions are represented as trees of the height T. In line with the work of G.Ch. Pflug and A.Pichler [28], nested distributions $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ possess the recursive structure and, hence, the nested distance $dl(\mathbb{P}, \widetilde{\mathbb{P}})$ can be calculated by the recursive procedure.
Let $\mathcal{N}_t$ and $\widetilde{\mathcal{N}}_t$ be the sets of all nodes at stage t of trees $\mathbb{P}$ and $\widetilde{\mathbb{P}}$, $\forall t = 1, ..., T$ and suppose that $n_t \in \mathcal{N}_t$ and $\widetilde{n}_t \in \widetilde{\mathcal{N}}_t$. Then, the recursive procedure for the nested distance $dl(\mathbb{P}, \widetilde{\mathbb{P}})$ calculation can be described in three steps:

Distance between leaves of the tree ($t = T$)**:** Let $dl(\mathbb{P}^{t:T|n_{t-1}}, \widetilde{\mathbb{P}}^{t:T|\widetilde{n}_{t-1}})$ be the nested distance between subtrees $\mathbb{P}^{t:T|n_{t-1}}$ and $\widetilde{\mathbb{P}}^{t:T|\widetilde{n}_{t-1}}$ starting from nodes n_{t-1} and $\widetilde{n}_{t-1}$ at stage $t-1$ of tree models $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ respectively and let $d_{KA}(P_T(\cdot|n_{T-1}), \widetilde{P}_T(\cdot|\widetilde{n}_{T-1}))$ be the Kantorovich distance between probability distributions P_T and $\widetilde{P}_T$ conditional on the nodes n_{T-1} and $\widetilde{n}_{T-1}$.
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 (2.3) with the distance matrix

$$d_{n_t, \widetilde{n}_t} = d(n_t, \widetilde{n}_t) + dl(\mathbb{P}^{t+1:T|n_t}, \widetilde{\mathbb{P}}^{t+1:T|\widetilde{n}_t}), \tag{2.4}$$

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 and given the underlying probabilities sitting on the nodes of the trees.

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

It may be difficult to calculate $dl_r(\mathbb{P}, \widetilde{\mathbb{P}})$ numerically, especially if $\mathbb{P}$ is continuous. In this case it is easier to estimate the Wasserstein distance $d_{WA_r}(P, \widetilde{P})$, but then one should know the relationship between dl_r and d_{WA_r}.

Theorem 2.2 (**Lower Bound, see [28]**). *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.

Proof. Notice, that the trivial sigma-algebra on $\Omega \times \widetilde{\Omega}$ is $\mathcal{F}_0 \otimes \widetilde{\mathcal{F}}_0 = \{0, \Omega \times \widetilde{\Omega}\}$. As soon as the following conditions hold

$$\begin{aligned} P(\cdot) = P(\cdot|\mathcal{F}_0) = \pi(\cdot \times \Omega|\mathcal{F}_0 \otimes \widetilde{\mathcal{F}_0}) = \pi(\cdot \times \Omega); \\ \widetilde{P}(\cdot) = \widetilde{P}(\cdot|\mathcal{F}_0) = \pi(\widetilde{\Omega} \times \cdot|\mathcal{F}_0 \otimes \widetilde{\mathcal{F}_0}) = \pi(\widetilde{\Omega} \times \cdot), \end{aligned}$$

the marginal conditions for trivial sigma-algebra are just the marginal conditions of the Wasserstein distance, that is why it obviously follows that

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

The same lower bound holds for order $r = 1$ and Kantorovich distance:

$$d_{KA}(P, \widetilde{P}) \leq dl(\mathbb{P}, \widetilde{\mathbb{P}}),$$

where $dl(\mathbb{P}, \widetilde{\mathbb{P}})$ denotes the nested distance of order $r = 1$. □

Calculating $d_{WA_r}(P, \widetilde{P})$ between joint distributions P and $\widetilde{P}$ we do not consider gradually increasing information at all, that is why d_{WA_r} is the lower bound for dl_r and it does not help us in estimation of the nested distance and in its minimization (minimization of the Wasserstein/Kantorovich distance, obviously, does not necessarily lead to the minimization of the nested distance). Nevertheless, by taking into account the gradually increasing information, we can produce several upper bounds for $dl(\mathbb{P}, \widetilde{\mathbb{P}})$, that can be used for the numerical minimization of $dl(\mathbb{P}, \widetilde{\mathbb{P}})$.
The following notations are used for the probability measures sitting on a treestructure:

P_t is an unconditional probability distribution sitting at the stage t;

$P_t(|u^{t-1})$ are conditional (on the past) probability distributions sitting at the stage t. At the stage t there are so many conditional distributions as many nodes are at the stage $t-1$;

$P^{t:T}$ is an unconditional joint probability distribution between stages t to T;

$P^{t:T}(|u^{t-1})$ are conditional multivariate probability distributions between stages t to T. At the stage t there are so many conditional distributions as many nodes are at the stage $t-1$.

Fig. 2.3 shows how conditional and unconditional continuous probability distributions are located on the tree.

Theorem 2.3 (**Upper Bound, see [28]**). *Let $\mathbb{P}$ be a nested distribution, P its multivariate distribution, which is dissected into the chain $P = P_1 \circ P_2 \circ ... \circ P_T$ of conditional*

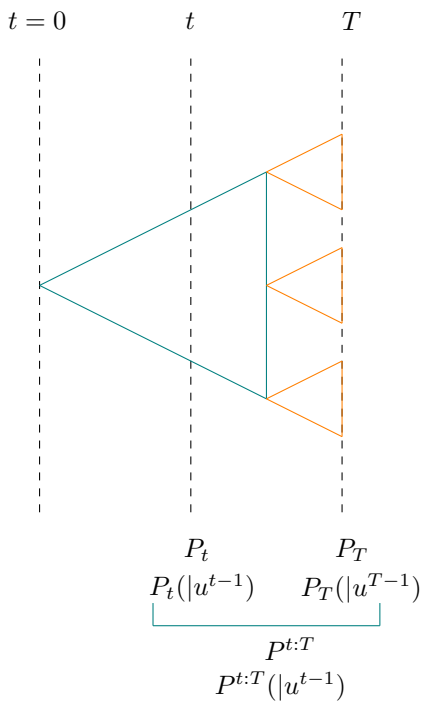


Figure 2.3: Probability measures on the treestructure.

distributions.
Then

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

Moreover, if $\forall t = 1, ..., T$ *the Lipschitz property (2.3) holds, i.e. if*

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

then

$$dl(\mathbb{P}, \widetilde{\mathbb{P}}) \leq \sum_{t=1}^{T} d_{KA}(P_t, \widetilde{P}_t) \prod_{s=t+1}^{T} (K_s + 1). \tag{2.5}$$

Proof. According to the definition of the nested distance we can write

$$dl(\mathbb{P}, \widetilde{\mathbb{P}}) \leq \int d(u, v)\pi(du, dv), \ \forall \pi \in \Pi,$$

where $d(u, v) = \sum_{t=1}^{T} d(u_t, v_t) = d(u^{T-1}, v^{T-1}) + d(u_T, v_T)$ and Π is the feasible set.
Let $\pi^*(u^{T-1}, v^{T-1})$ be the transportation plan optimal for the $dl_{T-1}(\mathbb{P}, \widetilde{\mathbb{P}})$ and
$\pi^*(u_T, v_T|u^{T-1}, v^{T-1})$ be the transportation plan optimal for the
$d_{KA}(P^T(\cdot|u^{T-1}), \widetilde{P}^T(\cdot|v^{T-1}))$.
Using the following transportation plan

$$\pi(du, dv) = \pi^*(du_T, dv_T|u^{T-1}, v^{T-1})\pi^*(du^{T-1}, dv^{T-1}) \in \Pi$$

we rewrite the nested distance

$$\begin{aligned} & dl_T(\mathbb{P},\widetilde{\mathbb{P}}) \leq \\ \leq & \int [d(u^{T-1},v^{T-1})+d(u_T,v_T)]\pi^*(du_T,dv_T|u^{T-1},v^{T-1})\pi^*(du^{T-1},dv^{T-1}) = \\ = & \int d(u^{T-1},v^{T-1})\pi^*(du^{T-1},dv^{T-1}) + \\ + & \int d(u_T,v_T)\pi^*(du_T,dv_T|u^{T-1},v^{T-1})\pi^*(du^{T-1}dv^{T-1}) = \\ = & dl_{T-1}(\mathbb{P},\widetilde{\mathbb{P}}) + \int d_{KA}(P_T(\cdot|u^{T-1}),\widetilde{P}_T(\cdot|v^{T-1}))\pi^*(du^{T-1},dv^{T-1}), \end{aligned} \tag{2.6}$$

where $d_{KA}(P_T(\cdot|u^{T-1}),\widetilde{P}_T(\cdot|v^{T-1})) \leq \sup_{u^{T-1},v^{T-1}} d_{KA}(P_T(\cdot|u^{T-1}),\widetilde{P}_T(\cdot|v^{T-1}))$.
Hence,

$$dl_t(\mathbb{P},\widetilde{\mathbb{P}}) \leq dl_{t-1}(\mathbb{P},\widetilde{\mathbb{P}}) + \sup_{u^{t-1},v^{t-1}} d_{KA}(P_t(\cdot|u^{t-1}),\widetilde{P}_t(\cdot|v^{t-1})),\ \forall t=1,...,T \tag{2.7}$$

Applying this recursion for all t we receive the first statement of the theorem.
Now, according to the equation (2.6) and in line with the triangle inequality and the Lipschitz property of the probability measure P_t we obtain the following bound

$$\begin{aligned} & d_{KA}(P_T(\cdot|u^{T-1}),\widetilde{P}_T(\cdot|v^{T-1})) \leq \\ \leq & d_{KA}(P_T(\cdot|u^{T-1}),P_T(\cdot|v^{T-1})) + d_{KA}(P_T(\cdot|v^{T-1}),\widetilde{P}_T(\cdot|v^{T-1})) \leq \\ \leq & K_T d(u^{T-1},v^{T-1}) + \sup_v d_{KA}(P_T(\cdot|v^{T-1}),\widetilde{P}_T(\cdot|v^{T-1}))) = \\ = & K_T d(u^{T-1},v^{T-1}) + d_{KA}(P_T,\widetilde{P}_T) \end{aligned}$$

and, hence,

$$dl_t(\mathbb{P},\widetilde{\mathbb{P}}) \leq (1+K_t)dl_{t-1}(\mathbb{P},\widetilde{\mathbb{P}}) + d_{KA}(P_t,\widetilde{P}_t),\ \forall t=1,...,T.$$

Applying this recursion for all t we receive the statement of the theorem.
Notice, that in case $\xi_t,\ t=1,...,T$ are i.i.d. random variables $K_t=0\ \forall t=1,...,T$ and, hence,

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

□

Now, as soon as every tree can be represented as a structure of its subtrees (Fig. 2.4) and the change in a subtree of tree $\mathbb{P}$ leads to the change in the nested distance $dl(\mathbb{P},\widetilde{\mathbb{P}})$, we would like to study how some particular changes in the subtree structure influence the total nested distance.

Definition 2.6 (Clairvoyant tree (new)). *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. 2.5).*

Remark 2.3. *Values at the stage $t=2$ of the clairvoyant tree in the Fig. 2.5 may coincide with each other, as soon as they appear from the same node of the initial tree model.*

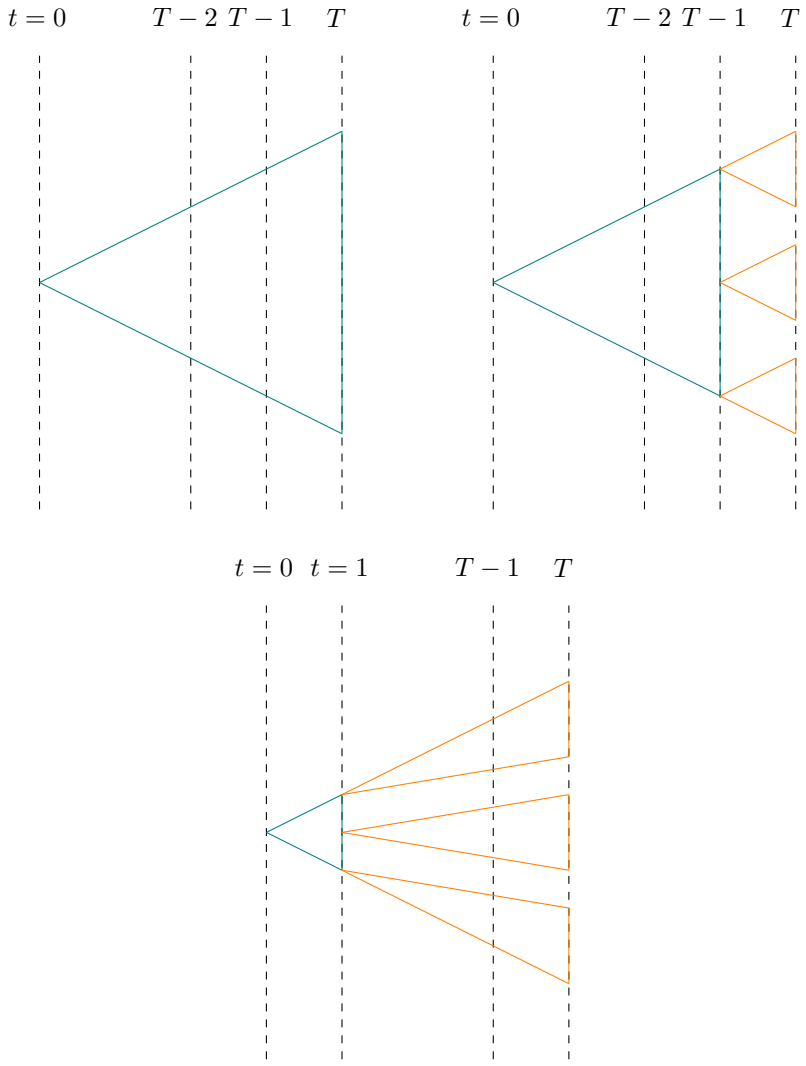


Figure 2.4: Tree representation in terms of subtrees.

Consider a nested distribution $\mathbb{P}$ with corresponding multivariate distribution P, which is dissected into the chain $P = P_1 \circ P_2 \circ ... \circ P_T$ of conditional distributions and let us represent this nested distribution as a tree.
Now, 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. In this case, the values sitting on the nodes of the clairvoyant tree stay unchanged with respect to the tree $\mathbb{P}$ and the corresponding multivariate distribution $P_{c(t)}$ can be dissected into the chain $P_{c(t)} = P_{c(t),1} \circ P_{c(t),2} \circ ... \circ P_{c(t),T}$, where

$$P_{c(t)}(\xi_t, ..., \xi_T | \xi_{t-1}, ..., \xi_1) = P(\xi_t | \xi_1, ..., \xi_{t-1}) \cdot ... \cdot P(\xi_T | \xi_{T-1}, ..., \xi_1), \tag{2.8}$$

that clearly follows from the chain probability of the initial tree $\mathbb{P}$ and the clairvoyant tree $\mathbb{P}_{c(t)}$:

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

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

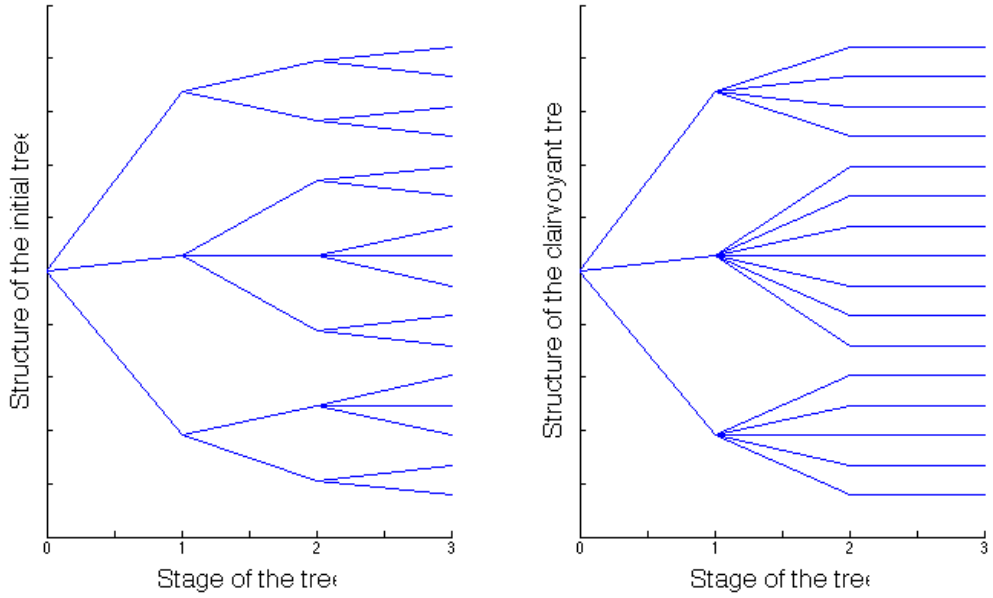


Figure 2.5: Example: clairvoyant tree at stage t=1 (v.s. initial one).

to time $t-1$.
Let us use the following notation for probability distributions sitting at the stage t of the clairvoyant tree $\mathbb{P}_{c(t)}$:

$P^{t|t+1:T}(|u^t)$ are conditional multivariate distributions sitting at the stage t of the clairvoyant tree. According to the equation (2.8), these distributions use the probability information of the future stages of the tree.

Theorem 2.4 (Improved 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:

$$d_{KA}(P, \widetilde{P}) = dl(\mathbb{P}_{c(1)}, \widetilde{\mathbb{P}}_{c(1)}) \leq dl(\mathbb{P}_{c(2)}, \widetilde{\mathbb{P}}_{c(2)}) \leq ... \leq dl(\mathbb{P}_{c(t)}, \widetilde{\mathbb{P}}_{c(t)}) \leq ...$$
$$... \leq dl(\mathbb{P}_{c(T-1)}, \widetilde{\mathbb{P}}_{c(T-1)}) \leq dl(\mathbb{P}_{c(T)}, \widetilde{\mathbb{P}}_{c(T)}) = dl(\mathbb{P}, \widetilde{\mathbb{P}}). \tag{2.9}$$

Proof. Relaxing the whole filtration structure of trees $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ (i.e. receiving $\mathbb{P}_{c(1)}$ and $\widetilde{\mathbb{P}}_{c(1)}$), we can claim, that

$$d_{KA}(P, \widetilde{P}) = dl(\mathbb{P}_{c(1)}, \widetilde{\mathbb{P}}_{c(1)}), \tag{2.10}$$

as soon as the nested distance (2.5) and the Kantorovich distance (2.1) coincide in case filtration conditions are relaxed, according to their definitions.
Now, consider subtrees of $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ at stage t and notice, that relaxing the filtration structure of each subtree at stage t (i.e. receiving $\mathbb{P}_{c(t)}$ and $\widetilde{\mathbb{P}}_{c(t)}$), we decrease the nested distance between subtees to the Kantorovich distance between conditional multivariate distributions $P^{t:T}(|u^t)$ and $\widetilde{P}^{t:T}(|u^t)$ in line with equality (2.10). Therefore, the resulting nested distance $dl(\mathbb{P}, \widetilde{\mathbb{P}})$ also decreases due to the recursive structure (2.4).
Hence,

$$dl(\mathbb{P}_{c(t)}, \widetilde{\mathbb{P}}_{c(t)}) \leq dl(\mathbb{P}, \widetilde{\mathbb{P}}).$$

The same can be written for the stage $t-1$ of the trees:

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

At the same time, we can write

$$dl(\mathbb{P}_{c(t-1)}, \widetilde{\mathbb{P}}_{c(t-1)}) \leq dl(\mathbb{P}_{c(t)}, \widetilde{\mathbb{P}}_{c(t)}),$$

taking into account that the filtration relaxation at the stage $t-1$ automatically implies the filtration relaxation at the stage t.
Combining all the inequalities for stages $t = 1, ..., T$, we receive

$$d_{KA}(P, \widetilde{P}) = dl(\mathbb{P}_{c(1)}, \widetilde{\mathbb{P}}_{c(1)}) \leq dl(\mathbb{P}_{c(2)}, \widetilde{\mathbb{P}}_{c(2)}) \leq ... \leq dl(\mathbb{P}_{c(t)}, \widetilde{\mathbb{P}}_{c(t)}) \leq ...$$
$$... \leq dl(\mathbb{P}_{c(T-1)}, \widetilde{\mathbb{P}}_{c(T-1)}) \leq dl(\mathbb{P}_{c(T)}, \widetilde{\mathbb{P}}_{c(T)}),$$

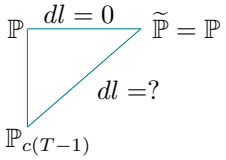
where $dl(\mathbb{P}_{c(T)}, \widetilde{\mathbb{P}}_{c(T)}) = dl(\mathbb{P}, \widetilde{\mathbb{P}})$ in line with the tree structure. □

Notice, that Theorem 2.4 holds only if both nested distributions $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ are changed to the clairvoyant ones at stage $t = 1, ..., T$. Counterexample 2.3 shows, that making clairvoyant only one of two trees $\mathbb{P}$ or $\widetilde{\mathbb{P}}$ does not lead to decrease or increase in the nested distance.

Example 2.3 (Counterexample). *Consider nested distributions $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ and suppose that one of them becomes clairvoyant at stage $T-1$, i.e. we get $\mathbb{P}_{c(T-1)}$. Now, we would like to check if the following inequality holds:*

$$dl(\mathbb{P}_{c(T-1)}, \widetilde{\mathbb{P}}) \leq dl(\mathbb{P}, \widetilde{\mathbb{P}}), \tag{2.11}$$

that is similar to the lower bound (2.9), but with just one clairvoyant tree.
To check it, we consider a simple example: clearly, if the nested distributions $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ are equal, i.e. $\mathbb{P} = \widetilde{\mathbb{P}}$, then the nested distance between them is equal to zero: $dl(\mathbb{P}, \mathbb{P}) = 0$. Hence, if we now make one of these nested distributions clairvoyant at stage $T-1$ and calculate the distance $dl(\mathbb{P}_{c(T-1)}, \mathbb{P})$ it, obviously, becomes greater or equal than zero ($dl(\mathbb{P}_{c(T-1)}, \mathbb{P}) \geq 0$) and it should be equal to zero in order to satisfy (2.11):



So, taking some tree as an example, we can check the inequality (2.11).
Consider the following tree:

Tree

$$\begin{aligned}
pred &= [0\ 1\ 1\ 1\ 2\ 2\ 3\ 3\ 3\ 4\ 4\ 5\ 5\ 6\ 6\ 7\ 7\ 8\ 8\ 8\ 9\ 9\ 10\ 10\ 10\ 11\ 11] \\
pNodes &= [1\ 0.2\ 0.3\ 0.5\ 0.5\ 0.5\ 0.6\ 0.2\ 0.2\ 0.7\ 0.3\ 0.5\ 0.5\ 0.1\ 0.9 \\
&\quad 0.2\ 0.8\ 0.1\ 0.2\ 0.7\ 0.5\ 0.5\ 0.2\ 0.3\ 0.5\ 0.9\ 0.1] \\
val &= [0\ 40.68\ 93.83\ 25.54\ 53.32\ 95.48\ 26.77\ 25.01\ 92.77\ 6.86\ 29.94 \\
&\quad 59.16\ 20.33\ 63.59\ 79.84\ 50.17\ 65.08\ 79.6\ 23.34\ 60.08\ 11.25 \\
&\quad 51.58\ 83.78\ 92.08\ 49.82\ 27.76\ 65.25]
\end{aligned}$$

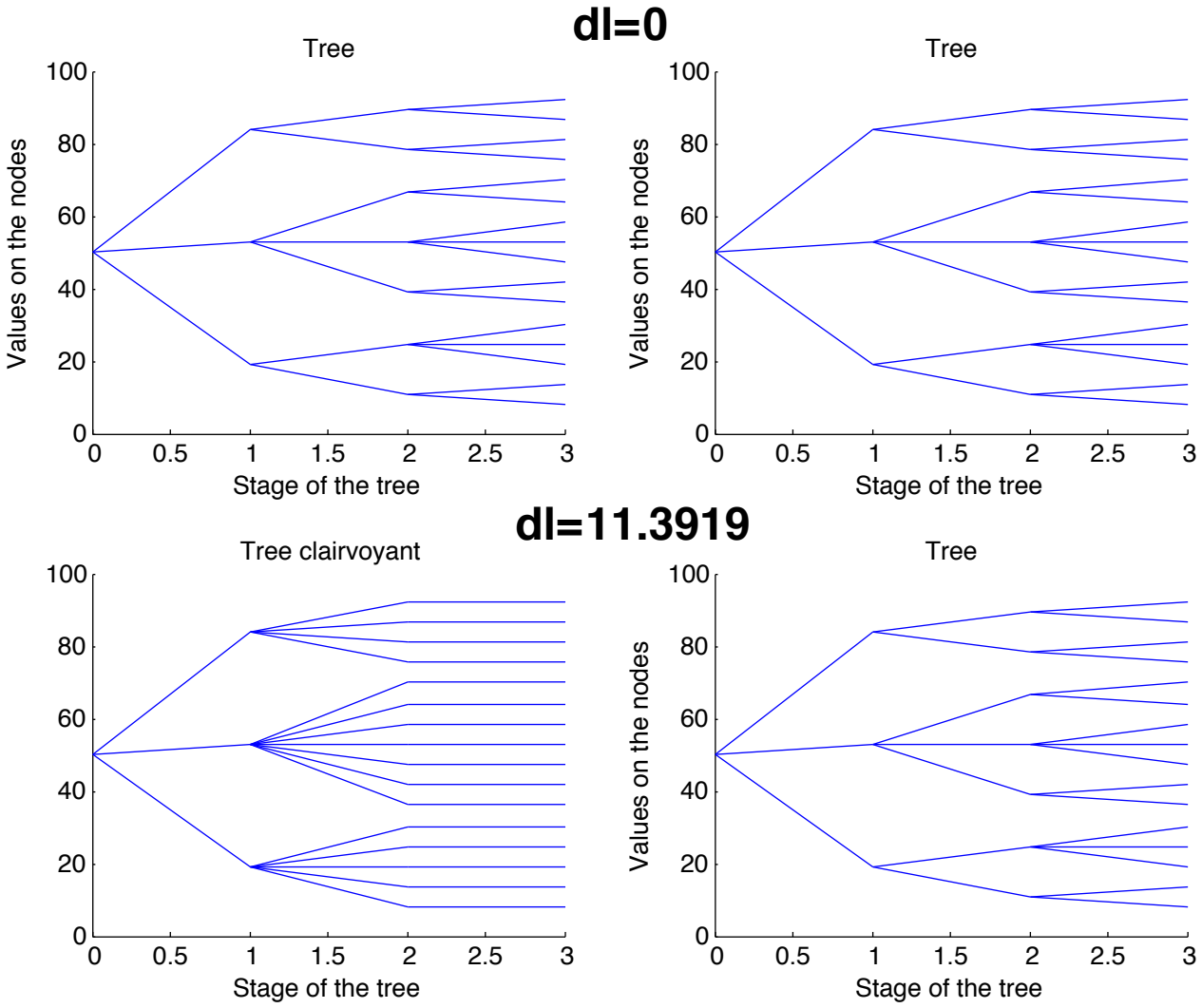


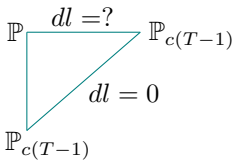
Figure 2.6: Increase in the nested distance due to the claivoyance of one tree.

Changing the structure of the tree so that it becomes clairvoyant at stage $T-1$ *and calculating necessary nested distances, we obtain the counterexample for the inequality (2.11), i.e.* $11.3919 = dl(\mathbb{P}_{c(T-1)}, \mathbb{P}) > dl(\mathbb{P}, \mathbb{P}) = 0$ *for the given tree (Fig. 2.6).*
Notice, that the values and probabilities on the nodes up to time $T-2$ *stay unchanged for the clairvoyant tree, while the values and the probabillities of the stages* $T-1$ *and* T *change in line with the change in the predecessor structure (Fig. 2.6):*

Tree clairvoyant

$$\begin{aligned} pred \;=\; & [0\ 1\ 1\ 1\ 2\ 2\ 2\ 2\ 3\ 3\ 3\ 3\ 3\ 3\ 4\ 4\ 4\ 4\ 4\ 5\ 6 \\ & 7\ 8\ 9\ 10\ 11\ 12\ 13\ 14\ 15\ 16\ 17\ 18\ 19\ 20] \end{aligned}$$

Hence, the nested distance $dl(\mathbb{P}_{c(T-1)}, \widetilde{\mathbb{P}})$ *is not a lower bound for the distance* $dl(\mathbb{P}, \widetilde{\mathbb{P}})$.
Clearly, the distance $dl(\mathbb{P}_{c(T-1)}, \widetilde{\mathbb{P}})$ *is also not an upper bound for the nested distance* $dl(\mathbb{P}, \widetilde{\mathbb{P}})$. *To see this, consider the same example from the opposite point: let nested distribution* $\widetilde{\mathbb{P}}$ *be equal to the clairvoyant distribution of* $\mathbb{P}$, *i.e.* $\widetilde{\mathbb{P}} = \mathbb{P}_{c(T-1)}$:



Clearly, the nested distance $dl(\mathbb{P}, \mathbb{P}_{c(T-1)}) \geq 0$*, and, particularly for the chosen tree it is equal to* $dl(\mathbb{P}, \mathbb{P}_{c(T-1)}) = 11.3919$*. Nevertheless, the nested distance* $dl(\mathbb{P}_{c(T-1)}, \mathbb{P}_{c(T-1)})$ *is obviously equal to zero, that means the following statement holds in this case:*

$$0 = dl(\mathbb{P}_{c(T-1)}, \mathbb{P}_{c(T-1)}) < dl(\mathbb{P}, \mathbb{P}_{c(T-1)}),$$

where we again changed just one tree structure to the clairvoyant one (Fig. 2.7).

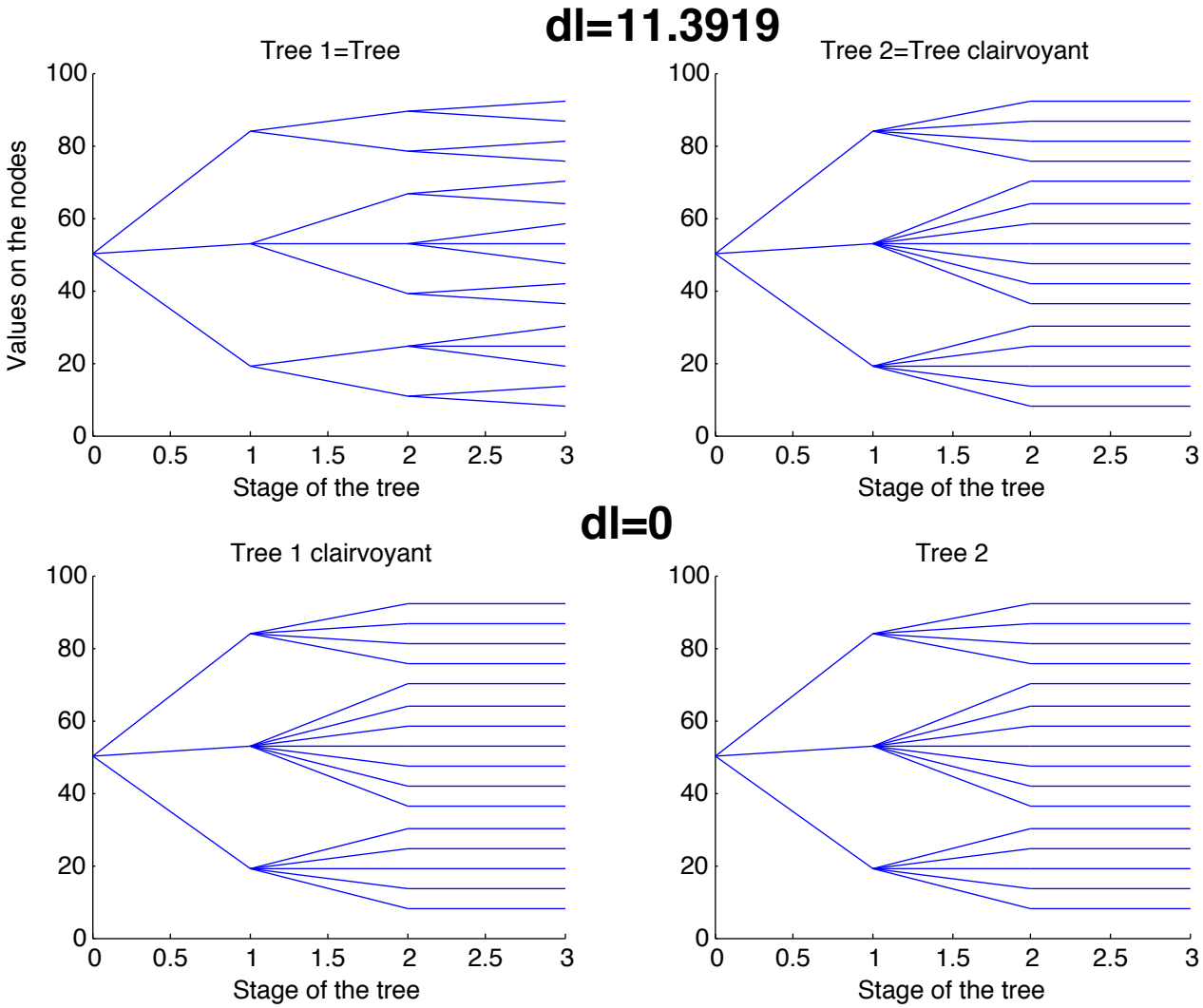


Figure 2.7: Decrease in the nested distance due to the claivoyance of one tree.

Theorem 2.5 (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

$$P = P_1 \circ P_2 \circ ... \circ P_T;$$
$$\widetilde{P} = \widetilde{P}_1 \circ \widetilde{P}_2 \circ ... \circ \widetilde{P}_T.$$

Then, the clairvoyant nested distributions $\mathbb{P}_{c(t)}$ *and* $\widetilde{\mathbb{P}}_{c(t)}$ *are defined* $\forall t = 1, ..., T$ *and*

$$dl(\mathbb{P}, \widetilde{\mathbb{P}}) \leq \sum_{t=1}^{T} \sup_{u^{t-1}, v^{t-1}} d_{KA}(P^{t|t+1:T}(\cdot|u^{t-1}), \widetilde{P}^{t|t+1:T}(\cdot|v^{t-1})), \tag{2.12}$$

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* $\widetilde{\mathbb{P}}_{c(t)}$ *correspondingly.*
Moreover, if $\forall t = 1, ..., T$ *the Lipschitz property holds for conditional joint distribution* $P^{t|t+1:T}$ *with constants* $K^{t|t+1:T}$*, i.e. if*

$$d_{KA}(P^{t|t+1:T}(\cdot|u^{t-1}), P^{t|t+1:T}(\cdot|v^{t-1})) \leq K^{t|t+1:T} d(u^{t-1}, v^{t-1}),$$

then (according to [28])

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

Proof. Consider two nested distributions $\mathbb{P}$ and $\widetilde{\mathbb{P}}$. According to the Theorem 2.3 we can write

$$dl_T(\mathbb{P}, \widetilde{\mathbb{P}}) \leq dl_{T-1}(\mathbb{P}, \widetilde{\mathbb{P}}) + \sup_{u^{T-1}, v^{T-1}} d_{KA}(P_T(\cdot|u^{T-1}), \widetilde{P}_T(\cdot|v^{T-1})).$$

According to the definition of the clairvoyant tree (2.6), we can derive clairvoyant nested distributions $\mathbb{P}_{c(t)}$ and $\widetilde{\mathbb{P}}_{c(t)}$ $\forall t = 1, ..., T$.
Consider the clairvoyant distributions $\mathbb{P}_{c(T-1)}$ and $\widetilde{\mathbb{P}}_{c(T-1)}$ (Fig. 2.8).

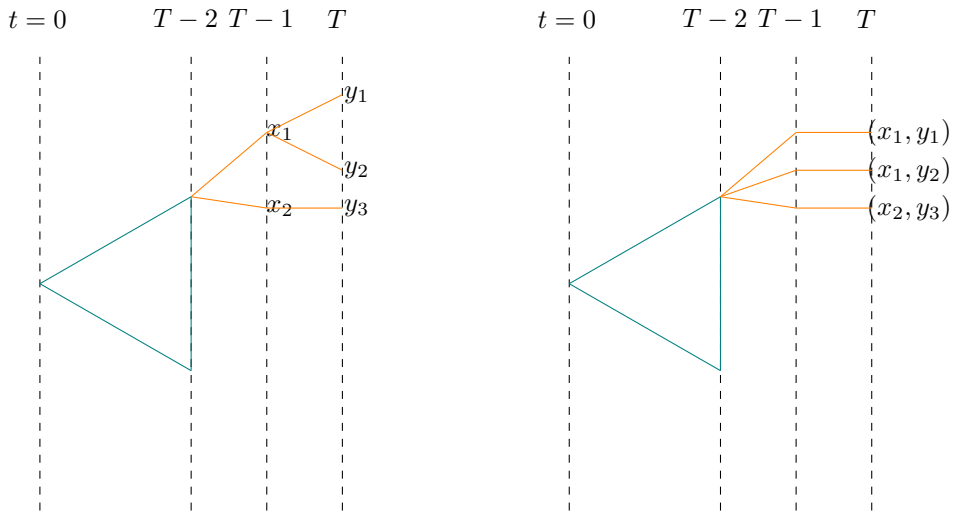


Figure 2.8: Decrease of the nested distance between subtrees.

Notice that the nested distributions of the initial and the clairvoyant trees are equal up to the stage $T-1$ according to the Lemma 2.1 and, hence,

$$\begin{aligned}\mathbb{P}^{1:T-1}_{c(T-1)} &\equiv \mathbb{P}^{1:T-1}\\ \widetilde{\mathbb{P}}^{1:T-1}_{c(T-1)} &\equiv \widetilde{\mathbb{P}}^{1:T-1},\end{aligned}$$

where we denote by $\mathbb{P}^{1:t}$ the nested distribution up to time t.
Therefore,

$$dl_{T-1}(\mathbb{P},\widetilde{\mathbb{P}}) = dl_{T-1}(\mathbb{P}_{c(T-1)},\widetilde{\mathbb{P}}_{c(T-1)}).$$

Recall, that the multivariate distribution sitting at the stage $T-1$ of the clairvoyant tree is $P^{T-1|T}(\cdot|u^{T-2})$ and, from here we can conclude the following:

$$\begin{aligned}dl_T(\mathbb{P},\widetilde{\mathbb{P}}) &\leq dl_{T-1}(\mathbb{P},\widetilde{\mathbb{P}}) + \sup_{u^{T-1},v^{T-1}} d_{KA}(P_T(\cdot|u^{T-1}),\widetilde{P}_T(\cdot|v^{T-1})),\\ dl_{T-1}(\mathbb{P}_{c(T-1)},\widetilde{\mathbb{P}}_{c(T-1)}) &\leq dl_{T-2}(\mathbb{P}_{c(T-1)},\widetilde{\mathbb{P}}_{c(T-1)}) +\\ &+ \sup_{u^{T-2},v^{T-2}} d_{KA}(P^{T-1|T}(\cdot|u^{T-2}),\widetilde{P}^{T-1|T}(\cdot|v^{T-2})),\end{aligned}$$

and, hence,

$$dl_{T-1}(\mathbb{P},\widetilde{\mathbb{P}}) \leq dl_{T-2}(\mathbb{P},\widetilde{\mathbb{P}}) + \sup_{u^{T-2},v^{T-2}} d_{KA}(P^{T-1|T}(\cdot|u^{T-2}),\widetilde{P}^{T-1|T}(\cdot|v^{T-2})),$$

where we used the fact that

$$\begin{aligned}dl_{T-1}(\mathbb{P},\widetilde{\mathbb{P}}) &= dl_{T-1}(\mathbb{P}_{c(T-1)},\widetilde{\mathbb{P}}_{c(T-1)})\\ dl_{T-2}(\mathbb{P},\widetilde{\mathbb{P}}) &= dl_{T-2}(\mathbb{P}_{c(T-1)},\widetilde{\mathbb{P}}_{c(T-1)}).\end{aligned}$$

Analogically, we consider the clairvoyant distributions $\mathbb{P}_{c(t)}$ and $\widetilde{\mathbb{P}}_{c(t)}$ and get the following bound:

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

Iterating the same procedure $\forall t = 1,...,T$ we get the recursion

$$\begin{aligned}dl_T \leq dl_{T-1} + &\sup_{u^{T-1},v^{T-1}} d_{KA}(P_T(\cdot|u^{T-1}),\widetilde{P}_T(\cdot|v^{T-1})),\\ dl_{T-1} \leq &\; dl_{T-2} + \sup_{u^{T-2},v^{T-2}} d_{KA}(P^{T-1|T}(\cdot|u^{T-2}),\widetilde{P}^{T-1|T}(\cdot|v^{T-2})),\\ &\vdots\\ &dl_t \leq dl_{t-1} + \sup_{u^{t-1},v^{t-1}} d_{KA}(P^{t|t+1:T}(\cdot|u^{t-1}),\widetilde{P}^{t|t+1:T}(\cdot|v^{t-1})).\end{aligned}$$

Applying this recursion for all t we receive the first statement of the theorem.
Now, according to the definition of the nested distance and following the same steps as in the Theorem 2.3 we can write

$$dl(\mathbb{P},\widetilde{\mathbb{P}}) \leq \int d(u,v)\pi(du,dv),\ \forall \pi \in \Pi,$$

where $d(u^t, v^t) = \sum_{s=1}^{t} d(u_s, v_s) = d(u^{t-1}, v^{t-1}) + d(u_t, v_t)$ and Π is the feasible set.
Let $\pi^*(u^{t-1}, v^{t-1})$ be the transportation plan optimal for the $dl_{t-1}(\mathbb{P}, \widetilde{\mathbb{P}})$ and
$\pi^*(u_t, v_t|u^{t-1}, v^{t-1})$ be the transportation plan optimal for the
$d_{KA}(P^{t|t+1:T}(\cdot|u^{t-1}), \widetilde{P}^{t|t+1:T}(\cdot|v^{t-1}))$.
Using the following transportation plan

$$\pi(du^t, dv^t) = \pi^*(du_t, dv_t|u^{t-1}, v^{t-1})\pi^*(du^{t-1}, dv^{t-1}) \in \Pi$$

we rewrite the nested distance

$$\begin{aligned}
& dl_t(\mathbb{P}, \widetilde{\mathbb{P}}) \leq \\
\leq & \int [d(u^{t-1}, v^{t-1}) + d(u_t, v_t)]\pi^*(du_t, dv_t|u^{t-1}, v^{t-1})\pi^*(du^{t-1}, dv^{t-1}) = \\
= & \int d(u^{t-1}, v^{t-1})\pi^*(du^{t-1}, dv^{t-1}) + \\
+ & \int d(u_t, v_t)\pi^*(du_t, dv_t|u^{t-1}, v^{t-1})\pi^*(du^{t-1}dv^{t-1}) = \\
= & dl_{t-1}(\mathbb{P}, \widetilde{\mathbb{P}}) + \int d_{KA}(P^{t|t+1:T}(\cdot|u^{t-1}), \widetilde{P}^{t|t+1:T}(\cdot|v^{t-1}))\pi^*(du^{t-1}, dv^{t-1}),
\end{aligned}$$

where

$$\begin{aligned}
& d_{KA}(P^{t|t+1:T}(\cdot|u^{t-1}), \widetilde{P}^{t|t+1:T}(\cdot|v^{t-1})) \leq \\
\leq & d_{KA}(P^{t|t+1:T}(\cdot|u^{t-1}), P^{t|t+1:T}(\cdot|v^{t-1})) + \\
+ & d_{KA}(P^{t|t+1:T}(\cdot|v^{t-1}), \widetilde{P}^{t|t+1:T}(\cdot|v^{t-1})) \leq \\
\leq & K^{t:T} d(u^{t-1}, v^{t-1}) + \sup_v d_{KA}(P^{t|t+1:T}(\cdot|v^{t-1}), \widetilde{P}^{t|t+1:T}(\cdot|v^{t-1}))) = \\
= & K^{t|t+1:T} d(u^{t-1}, v^{t-1}) + d_{KA}(P^{t|t+1:T}, \widetilde{P}^{t|t+1:T})
\end{aligned}$$

according to the triangle inequality, assumptions of the theorem and the notation for conditional probability measures

$$d_{KA}(P^{t|t+1:T}, \widetilde{P}^{t|t+1:T}) = \sup_v d_{KA}(P^{t|t+1:T}(\cdot|v^{t-1}), \widetilde{P}^{t|t+1:T}(\cdot|v^{t-1}))).$$

Hence,

$$dl_t(\mathbb{P}, \widetilde{\mathbb{P}}) \leq (1 + K^{t|t+1:T}) dl_{t-1}(\mathbb{P}, \widetilde{\mathbb{P}}) + d_{KA}(P^{t|t+1:T}, \widetilde{P}^{t|t+1:T}),\ \forall t = 1, ..., T.$$

Applying this recursion for all t we receive the second statement of the theorem.

□

Corollary 2.1. *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*

$$P = P_1 \circ P_2 \circ ... \circ P_T;$$
$$\widetilde{P} = \widetilde{P}_1 \circ \widetilde{P}_2 \circ ... \circ \widetilde{P}_T.$$

Then,

$$dl(\mathbb{P}, \widetilde{\mathbb{P}}) \leq \sum_{t=1}^{T} \sup_{u^{t-1}, v^{t-1}} d_{KA}(P^{t:T}(\cdot|u^{t-1}), \widetilde{P}^{t:T}(\cdot|v^{t-1})), \tag{2.13}$$

where $P^{t:T}(|u^{t-1})$ *and* $\widetilde{P}^{t:T}(|v^{t-1})$ *are conditional multivariate distribution sitting at the stages* $t, ..., T$ *of the corresponding trees.*
Moreover, if $\forall t = 1, ..., T$ *the Lipschitz property holds for joint distributions of stages* t *to* T *with constants* $K^{t:T}$*, i.e. if*

$$d_{KA}(P^{t:T}(\cdot|u^{t-1}), P^{t:T}(\cdot|v^{t-1})) \leq K^{t:T} d(u^{t-1}, v^{t-1}),$$

then

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

Proof. Immediately follows from the Theorem 2.5 and from the probability chain (2.8) of the corresponding clairvoyant trees. □

Notice, that if the components $\xi_t,\ \forall t = 1, ..., T$ of the stochastic process $\xi = (\xi_1, ..., \xi_T)$ are i.i.d. random variables, it is enough to use the upper bound 2.3, as soon as it cannot be improved due to the independence between the tree stages. However, if the components $\xi_t,\ \forall t = 1, ..., T$ of the stochastic process $\xi = (\xi_1, ..., \xi_T)$ are not i.i.d. random variables, the upper bound 2.5 should be considered.

Example 2.4 (**Gaussian Distribution**). *Let the stochastic process* $\xi = (\xi_1, \ldots, \xi_T)$ *sitting on the tree* $\mathbb{T}$ *be distributed as a multivariate normal variable with mean vector* $(\mu_1, \ldots, \mu_T)$ *and non-singular covariance matrix* $C = (c_{s,t})_{t=1,...,T;s=1,...,T}$.
Consider the main submatrix

$$C_t = \begin{pmatrix} c_{1,1} & \ldots & c_{1,t} \\ \vdots & \ddots & \vdots \\ c_{t,1} & \ldots & c_{t,t} \end{pmatrix}$$

of C *and dissect it into*

$$C_t = \left(\begin{array}{c|c} C_{t-1} & c^t \\ \hline (c^t)^\top & c_{t,t} \end{array}\right)$$

Let $\mu^t = (\mu_1, \ldots, \mu_t)$*. Then every conditional distribution of* ξ_t *given the history* ξ^{t-1} *is also a normal distribution of the form (see [16] for the details)*

$$\mathcal{N}(\mu_t + (\xi^{t-1} - \mu^{t-1}) C_{t-1}^{-1} c^t, c_{t,t} - (c^t)^\top C_{t-1}^{-1} c^t). \tag{2.14}$$

The similar result can be written for every joint distribution of $\xi^{t:T} = (\xi_t, ..., \xi_T)$.
Consider

$$C = \begin{pmatrix} c_{1,1} & \ldots & c_{1,T} \\ \vdots & \ddots & \vdots \\ c_{T,1} & \ldots & c_{T,T} \end{pmatrix}$$

and dissect it into

$$C_T = \left(\begin{array}{c|c} C_{t-1} & C^{t:T} \\ \hline (C^{t:T})^\top & C^{t:T,t:T} \end{array}\right).$$

Then the form of the conditional on the stages $1, .., t-1$ *distribution for the stages* $t, ..., T$ *is*

$$\mathcal{N}(\mu^{t:T} + (\xi^{t-1} - \mu^{t-1}) C_{t-1}^{-1} C^{t:T}, C^{t:T,t:T} - (C^{t:T})^\top C_{t-1}^{-1} C^{t:T}), \tag{2.15}$$

where $\mu^{t:T} = (\mu_t, ..., \mu_T)$.
We claim that

$$\begin{aligned} d_{KA}(P_t(\cdot|\xi_1^{t-1}), P_t(\cdot|\xi_2^{t-1})) &\leq ||(\xi_1^{t-1} - \xi_2^{t-1})C_{t-1}^{-1}c^t||, \quad (2.16) \\ d_{KA}(P^{t:T}(\cdot|\xi_1^{t-1}), P^{t:T}(\cdot|\xi_2^{t-1})) &\leq ||(\xi_1^{t-1} - \xi_2^{t-1})C_{t-1}^{-1}C^{t:T}||, \quad (2.17) \end{aligned}$$

where ξ_1^{t-1} *and* ξ_2^{t-1} *are two different process histories up to time* $t-1$.
And, hence,

$$\begin{aligned} K_t &= ||C_{t-1}^{-1}c^t||, \quad (2.18) \\ K^{t:T} &= ||C_{t-1}^{-1}C^{t:T}||. \quad (2.19) \end{aligned}$$

Indeed, if we let $\xi \sim \mathcal{N}(0, \sigma_t)$ *and* $\xi_1 = \nu_1 + \xi$ *and* $\xi_2 = \nu_2 + \xi$, *then* $\xi_1 \sim \mathcal{N}(\nu_1, \sigma_t)$, $\xi_2 \sim \mathcal{N}(\nu_2, \sigma_t)$ *and therefore*

$$d_{KA}(\mathcal{N}(\nu_1, \sigma_t), \mathcal{N}(\nu_2, \sigma_t)) \leq \mathbb{E}(||\xi_1 - \xi_2||) = ||\nu_1 - \nu_2||.$$

Setting $\nu_1 = \mu_t + (\xi_1^{t-1} - \mu^{t-1})C_{t-1}^{-1}c^t$, $\nu_2 = \mu_t + (\xi_2^{t-1} - \mu^{t-1})C_{t-1}^{-1}c^t$ *and* $\sigma_t = c_{t,t} - (c^t)^\top C_{t-1}^{-1}c^t$ *implies (2.16) and (2.18).*
Analogically, one can prove statements (2.17) and (2.19).

Example 2.5 (Independent increments). *Consider stochastic process* $\eta = (\eta_1, \ldots, \eta_T)$ *and suppose that the mean* μ_t *and the variance* σ_t *are given for each* η_t $\forall t = 1, ..., T$, *i.e. we know that* $\eta_t \sim \mathcal{N}(\mu_t, \sigma_t)$. *We would like to calculate the covariance matrix* $C = (c_{s,t})_{t=1,...,T;s=1,...,T}$ *of the multivariate joint distribution of the stochastic process* $\xi = (\xi_1, \ldots, \xi_T)$, *knowing that the process is*

$$\begin{aligned} \xi_1 &= \eta_1, \\ \xi_2 &= \eta_1 + \eta_2, \\ \xi_3 &= \eta_1 + \eta_2 + \eta_3, \\ &\vdots \\ \xi_T &= \eta_1 + \eta_2 + \eta_3 + ... + \eta_T. \end{aligned}$$

From here we easily receive, that

$$\begin{aligned} \eta_1 &= \xi_1, \\ \eta_2 &= \xi_2 - \xi_1, \\ \eta_3 &= \xi_3 - \xi_2, \\ &\vdots \\ \eta_T &= \xi_T - \xi_{T-1}. \end{aligned}$$

As soon as $\xi_t - \xi_{t-1}$ *are independent random variables, we write the probability of* x_t *being in the given bounds in the following way:*

$$\begin{aligned} &P(\xi_1 \in [x_1, y_1], \xi_2 \in [x_2, y_2], ..., \xi_T \in [x_T, y_T]) = \\ &= \int_{x_1}^{y_1} \int_{x_2}^{y_2} ... \int_{x_T}^{y_T} f_1(\xi_1) f_2(\xi_2 - \xi_1) ... f_T(\xi_T - \xi_{T-1}) d\xi_1 ... \xi_T, \end{aligned}$$

where f_t is the probability density function for the stage t.
As soon as we know the mean and the variance for each of the density functions, we can calculate the resulting covariance matrix

$$P(\xi_1 \in [x_1, y_1], \xi_2 \in [x_2, y_2], ..., \xi_T \in [x_T, y_T]) \sim$$
$$\sim \int e^{-\frac{1}{2\sigma_1^2}t_1^2 - \frac{1}{2\sigma_2^2}(t_2-t_1)^2 - ... - \frac{1}{2\sigma_T^2}(t_T-t_{T-1})^2} dt_1...t_T,$$

where we used the following change of variables:

$$\begin{aligned} t_1 &= \xi_1 - \mu_1, \\ t_2 &= \xi_2 - \mu_1 - \mu_2, \\ &\vdots \\ t_T &= \xi_T - \sum_{t=1}^{T} \mu_t. \end{aligned}$$

After opening the brackets under the exponential function we receive

$$P(\xi_1 \in [x_1, y_1], \xi_2 \in [x_2, y_2], ..., \xi_T \in [x_T, y_T]) \sim$$
$$\sim \int e^{-\frac{1}{2}(\xi-\mu)^T C^{-1}(\xi-\mu)} d\xi_1...\xi_T,$$

where

$$C^{-1} = \begin{pmatrix} \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} & -\frac{1}{\sigma_2^2} & 0 & \cdots & 0 \\ -\frac{1}{\sigma_2^2} & \frac{1}{\sigma_2^2} + \frac{1}{\sigma_3^2} & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \frac{1}{\sigma_{T-1}^2} + \frac{1}{\sigma_T^2} & -\frac{1}{\sigma_T^2} \\ 0 & \cdots & 0 & -\frac{1}{\sigma_T^2} & \frac{1}{\sigma_T^2} \end{pmatrix},$$

is tridiagonal and

$$C = \begin{pmatrix} \sigma_1^2 & \cdots & & \cdots & \sigma_1^2 \\ \vdots & \sigma_1^2 + \sigma_2^2 & & \cdots & \sigma_1^2 + \sigma_2^2 \\ \vdots & \vdots & \sigma_1^2 + \sigma_2^2 + \sigma_3^2 & \cdots & \sigma_1^2 + \sigma_2^2 + \sigma_3^2 \\ \vdots & \vdots & \vdots & & \vdots \\ \sigma_1^2 & \sigma_1^2 + \sigma_2^2 & \sigma_1^2 + \sigma_2^2 + \sigma_3^2 & \cdots & \sum_{t=1}^{T} \sigma_t^2 \end{pmatrix}.$$

Based on the inequality (2.16) and using the Cauchy-Schwarz inequality, we can see that

$$d_{KA}(P_t(\cdot|\xi_1^{t-1}), P_t(\cdot|\xi_2^{t-1})) \leq ||(\xi_1^{t-1} - \xi_2^{t-1})C_{t-1}^{-1}c^t|| \leq ||C_{t-1}^{-1}c^t||||\xi_1^{t-1} - \xi_2^{t-1}||,$$

where the Lipschitz constant K_t can be calculated by the finding of $C_{t-1}^{-1}c^t$.
As soon as each of C_{t-1}^{-1} is tridiagonal and c^t is the following vector

$$c^t = \begin{pmatrix} \sigma_1^2 \\ \sigma_1^2 + \sigma_2^2 \\ \sigma_1^2 + \sigma_2^2 + \sigma_3^2 \\ \vdots \\ \sum_{t=1}^{T-1} \sigma_t^2 \end{pmatrix},$$

the multiplication of the matrix C_{t-1}^{-1} *and the vector* c^t *will give us such a vector* $\forall t = 2, ..., T$*:*

$$C_{t-1}^{-1} c^t = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}, \tag{2.20}$$

with only one non-zero element.
As it is expected, in this case $K^{t:T} = K_t = 1$ *due to the tridiagonal structure of the matrix* C^{-1}*.*

2.3.1 Nested distance between stochastic process and scenario tree

- Using the Kantorovich distance $d_{KA}(P, \widetilde{P})$ (see definition 2.1) we are able to measure stage-wise the difference between continuous distribution function P and its discrete approximation $\widetilde{P}$. We are not able, though, to measure the distance between stochastic process given by the continuous distribution function and the whole tree that contains all available information (incl. stage-wise information).

- The multi-stage distance (see definition 2.5) gives us an opportunity to measure the difference between two finite trees ($\widetilde{\mathbb{P}}^{(1)}$ and $\widetilde{\mathbb{P}}^{(2)}$). But, in case one of the trees is infinite (i.e. in case of stochastic process given by the continuous distribution function) there is a numerical difficulty to find the shipping plan $\pi(dw, d\widetilde{w})$.

In this section we are going to introduce an approximation of the multi-stage distance between stochastic process given by its continuous distribution function and a tree with given structure, probabilities and values (Tree 1, Fig. 2.9).

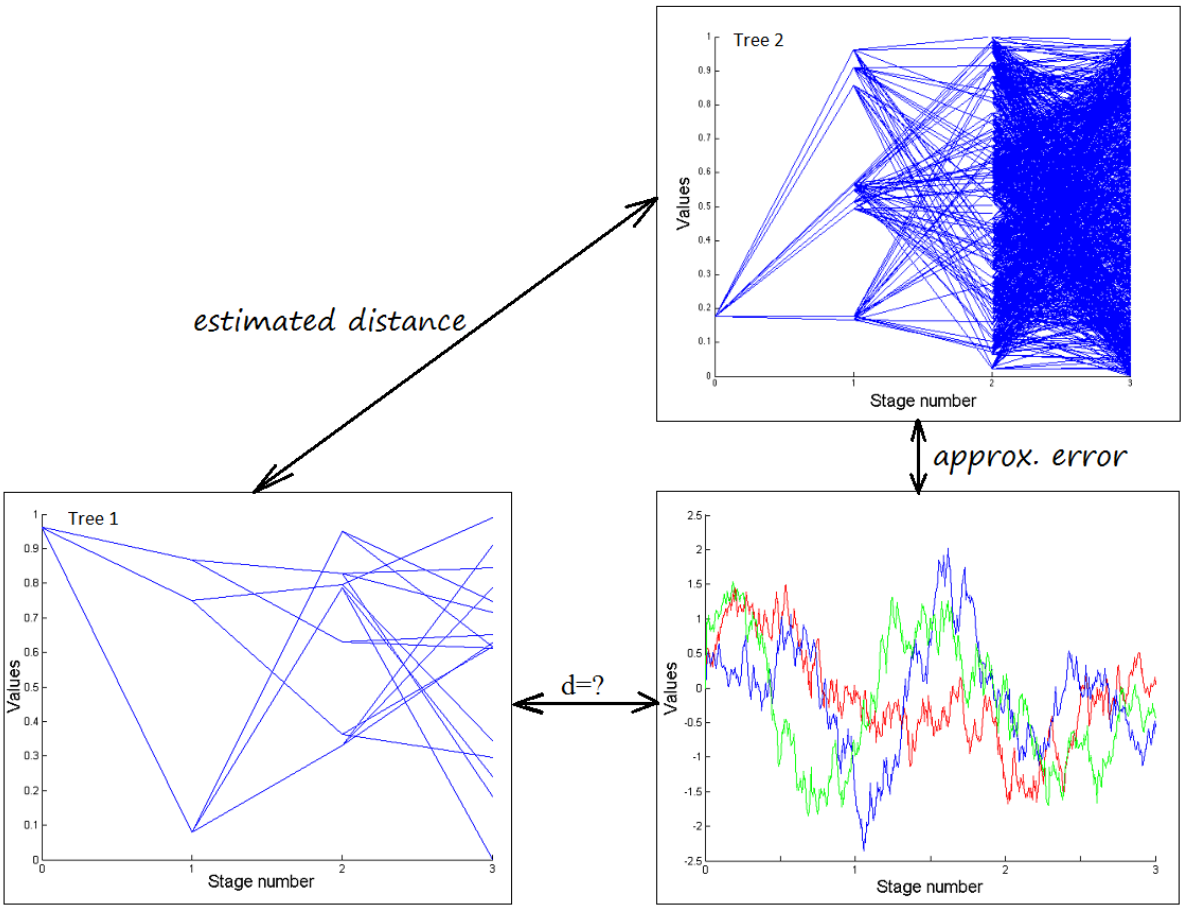


Figure 2.9: Distance approximation for the case of one infinite tree.

Suppose that Tree 1 is given: we know the tree structure, the probabilities of the nodes $p_{n,i}$ and the values sitting on the nodes $z_{n,i}$, where every node of the tree is identified by two numbers (n, i), where n is the identification number of its predecessor and $1 \leq i \leq S_n$,

where S_n is a total number of successors of n. The stochastic process $\xi = (\xi_1, ..., \xi_T)$ is given by the joint distribution function P. We would like to find the distance between this stochastic process and the given tree.
Suppose that $\mathbb{P}$ is a nested distribution of the stochastic process ξ, $\widetilde{\mathbb{P}}$ is a nested distribution of the given tree (Tree 1) and $\hat{\mathbb{P}}$ is a nested distribution of the approximate stochatic process $\hat{\xi}$ (Tree 2), lying in the ε-neighborhood of the initial nested distribution $\mathbb{P}$. We would like to calculate the distance $dl(\mathbb{P}, \widetilde{\mathbb{P}})$ between stochastic process ξ and the given tree (Tree 1) by the approximation of this stochastic process by a tree with high bushiness (Tree 2).
According to the triangle inequality that holds for the nested distance, we can write (see Fig. 2.9)

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

In order to approximate the distance $dl(\mathbb{P}, \widetilde{\mathbb{P}})$ between stochastic process and Tree 1 by the distance $dl(\widetilde{\mathbb{P}}, \hat{\mathbb{P}})$ between Tree 1 and the process approximation (Tree 2) we should guarantee that the distance between stochastic process ξ and its approximation (Tree 2) is small enough[1]:

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

Let P be a probability on $\mathbb{R}^{DT}$, dissected into transition probabilities $P_1, ..., P_T$.
Let us use the following notations:

$$k_t = \prod_{s=t+1}^{T} (K_s + 1),$$

$$k^{t|t+1:T} = \prod_{s=t+1}^{T} (K^{s|s+1:T} + 1),$$

where K_t and $K^{t|t+1:T}$ are Lipschitz constants introduced in Theorem 2.3 and Theorem 2.5. Notice, that $k_T = k^T = 1$ as soon as $K_{T+1} = K^{T+1} = 0$.
According to the Definition 2.3 and to the Theorems 2.2, 2.3, 2.5, we can find the bounds of the distance between stochastic process and a tree for the multivariate distributions with Lipschitz property (for example, Gaussian distribution):

Upper bounds:

$$dl(\mathbb{P}, \widetilde{\mathbb{P}}) \leq \sum_{t=1}^{T} k_t d_{KA}(P_t, \widetilde{P}_t) = U_1 \tag{2.21}$$

The upper bound U_1 bounds the nested distance by the sum of Kantorovich distances between conditional distributions and their approximations at each stage of the tree. This upper bound does not pay attention on the distances between subtrees, i.e. it does not take into account the future values of the stochastic scenario process $(t+1, ..., T)$ at the stage t.

$$dl(\mathbb{P}, \widetilde{\mathbb{P}}) \leq \sum_{t=1}^{T} k^{t|t+1:T} d_{KA}(P^{t|t+1:T}, \widetilde{P}^{t|t+1:T}) = U_2 \tag{2.22}$$

[1]We do not say minimized as the minimum may be not obtained due to the computational error.

The upper bound U_2 bounds the nested distance by the sum of Kantorovich distances between conditional joint distributions and their approximations at each stage of the tree. This upper bound pays attention on the distances between subtrees, i.e. it takes into account the future values of the stochastic scenario process $(t+1, ..., T)$ at the stage t, that makes it more precise than U_1.

$$dl(\mathbb{P}, \widetilde{\mathbb{P}}) \leq dl(\widetilde{\mathbb{P}}, \hat{\mathbb{P}}) + \sum_{t=1}^{T} k_t d_{KA}(P_t, \hat{P}_t) = U_3, \tag{2.23}$$

$$dl(\mathbb{P}, \widetilde{\mathbb{P}}) \leq dl(\widetilde{\mathbb{P}}, \hat{\mathbb{P}}) + \sum_{t=1}^{T} k^{t|t+1:T} d_{KA}(P^{t|t+1:T}, \hat{P}^{t|t+1:T}) = U_4, \tag{2.24}$$

where $\hat{P}_t$ is a corresponding to $\hat{\mathbb{P}}$ probability distribution at the stage t; $\hat{P}^{t|t+1:T}$ is a corresponding to $\hat{\mathbb{P}}$ multivariate probability distribution sitting at the stages $t, ..., T$. Upper bounds U_3 and U_4 are based on the triangle inequality and on the bounds U_1 and U_2.

Lower bounds:

$$L_1 = d_{KA}(P, \widetilde{P}) \leq dl(\mathbb{P}, \widetilde{\mathbb{P}}). \tag{2.25}$$

The lower bound (2.25) does not take into account the tree structure at all, bounding the nested distance by the Kantorovich distance between joint distribution and its approximation.

$$\begin{aligned} &d_{KA}(P, \widetilde{P}) = dl(\mathbb{P}_{c(1)}, \widetilde{\mathbb{P}}_{c(1)}) \leq dl(\mathbb{P}_{c(2)}, \widetilde{\mathbb{P}}_{c(2)}) \leq ... \leq dl(\mathbb{P}_{c(t)}, \widetilde{\mathbb{P}}_{c(t)}) \leq ... \\ &... \leq dl(\mathbb{P}_{c(T-1)}, \widetilde{\mathbb{P}}_{c(T-1)}) \leq dl(\mathbb{P}_{c(T)}, \widetilde{\mathbb{P}}_{c(T)}) = dl(\mathbb{P}, \widetilde{\mathbb{P}}). \end{aligned} \tag{2.26}$$

The chain of lower bounds (2.26) for the nested distance $dl(\mathbb{P}, \widetilde{\mathbb{P}})$ is received from the Theorem 2.4 and is based on the clairvoyant nested distributions (2.6) $\mathbb{P}_{c(t)}$ and $\widetilde{\mathbb{P}}_{c(t)}$ at stage t. This chain of lower bounds is, first of all, clearly better than the lower bound (2.25) and, moreover, it takes into account the tree structures of initial nested distribution $\mathbb{P}$ and $\widetilde{\mathbb{P}}$.

Now, we can prove the sufficient condition for the convergence of the nested distance to zero (see [35]):

Theorem 2.6. *(Sufficient condition for the convergence) Suppose that the stochastic process $\xi = (\xi_1, ..., \xi_T)$ is approximated by the tree with bushiness vector $bush = [b_1, ..., b_T]$ ($\mathbb{P}$ and $\widetilde{\mathbb{P}}$ are corresponding to the process and to the tree nested distributions). Suppose that the $(K_2, ..., K_T)$-Lipschitz property holds for the corresponding multivariate distribution P, that is dissected into the chain $P = P_1 \circ P_2 ... \circ P_T$ of conditional distributions. If the bushiness factors $b_1, ..., b_T$ are increasing and the conditions of the Theorem 2.1 are satisfied for each P_t, then the multi-stage distance $dl(\mathbb{P}, \widetilde{\mathbb{P}})$ is converging to zero.*

Proof. As the conditions of the Theorem 2.1 are satisfied for all P_t, $t = 1, ..., T$, then $d_{KA}(P_t, \widetilde{P}_t^{b_t}) \to 0 \ \forall t = 1, ..., T$ if the bushiness factor $b_t \to \infty \ \forall t = 1, ..., T$, where by $\widetilde{P}_t^{b_t}$ we denote the approximation $\widetilde{P}_t$ of P_t sitting on b_t points.
As soon as the $(K_2, ..., K_T)$-Lipschitz property holds for P, that is dissected into the chain

$P = P_1 \circ P_2 ... \circ P_T$ of conditional distributions, then according to the Theorem 2.3, we can write

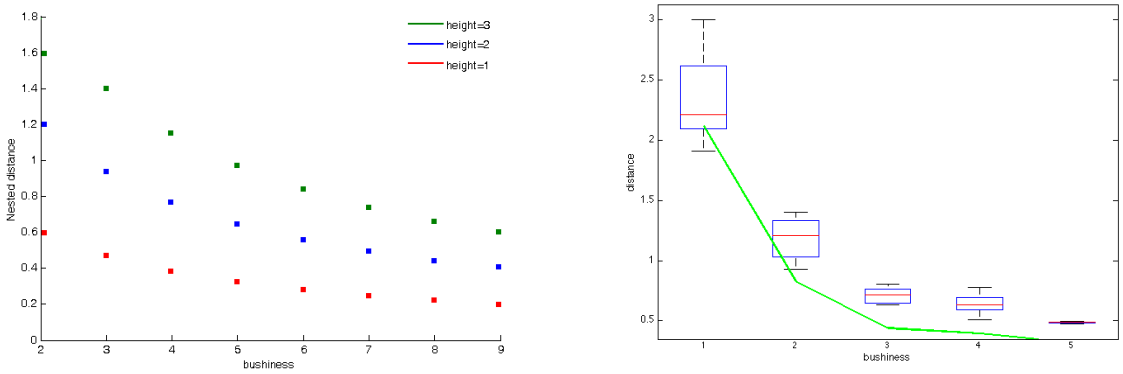
$$dl(\mathbb{P}, \widetilde{\mathbb{P}}) \leq \sum_{t=1}^{T} k_t d_{KA}(P_t, \widetilde{P}_t^{b_t}).$$

This implies that the upper bound for the nested distance is converging to zero when the bushiness of the tree is increasing and, hence,

$$dl(\mathbb{P}, \widetilde{\mathbb{P}}) \to 0.$$

□

The convergence in bushiness holds for the upper bound U_2 that depends on the approximation $\hat{\mathbb{P}}$ and $\hat{P}_t$ of the stochastic process. The rate of the convergence depends on the method of scenario generation [11, 28, 32]. Fig. 2.10 shows the convergence of the nested distance between normally distributed stochastic process and a tree with *optimally generated* scenarios (Fig. 2.10 (a), see [8, 17, 37] for the details) as well as with a tree with *randomly generated* scenarios (boxplot in the Fig. 2.10 (b), see [32]). As one can see in the Fig. 2.10 (b) the convergence of the nested distance between the stochastic process and the tree with randomly generated scenarios holds in probability. The reason is that the conditions of the Theorem 2.1 are satisfied almost surely for a tree with randomly generated values sitting on it and, hence, the conditions of the Theorem 2.6 that imply the convergence are satisfied in probability for this case.



a. Optimal quantization.

b. Random generation.

Figure 2.10: Convergence of the nested distance to zero.

At the same time, the upper bound U_1 does not depend on the stochastic process approximation tree with $\hat{\mathbb{P}}$ and $\hat{P}_t$, that is why it does not change when the bushiness of the approximate tree increases. It is natural to guess that U_1 is better than U_2, i.e. $U_1 \leq U_2$, if the approximation tree of the stochastic process has low bushiness and U_2 is better than U_1 if the bushiness is high.
Hence, the way to receive a better approximation for the nested distance $dl(\mathbb{P}, \widetilde{\mathbb{P}})$ between stochastic process and a tree is to increase the bushiness of the tree that approximates the stochastic process until the neccesary level ε of the gap between the lower L and the upper bound U_2 is obtained.

2.3.2 Nested distance between stochastic scenario processes

As soon as we are able to find the bounds for the distance between continuous-state stochastic process and a tree, we can generalize these bounds for the case of two stochastic processes, while the distance between two continuous-state stochastic processes is the nested distance between two infinitely large trees (i.e. the trees with infinitely many nodes, Fig. 2.11) [27].

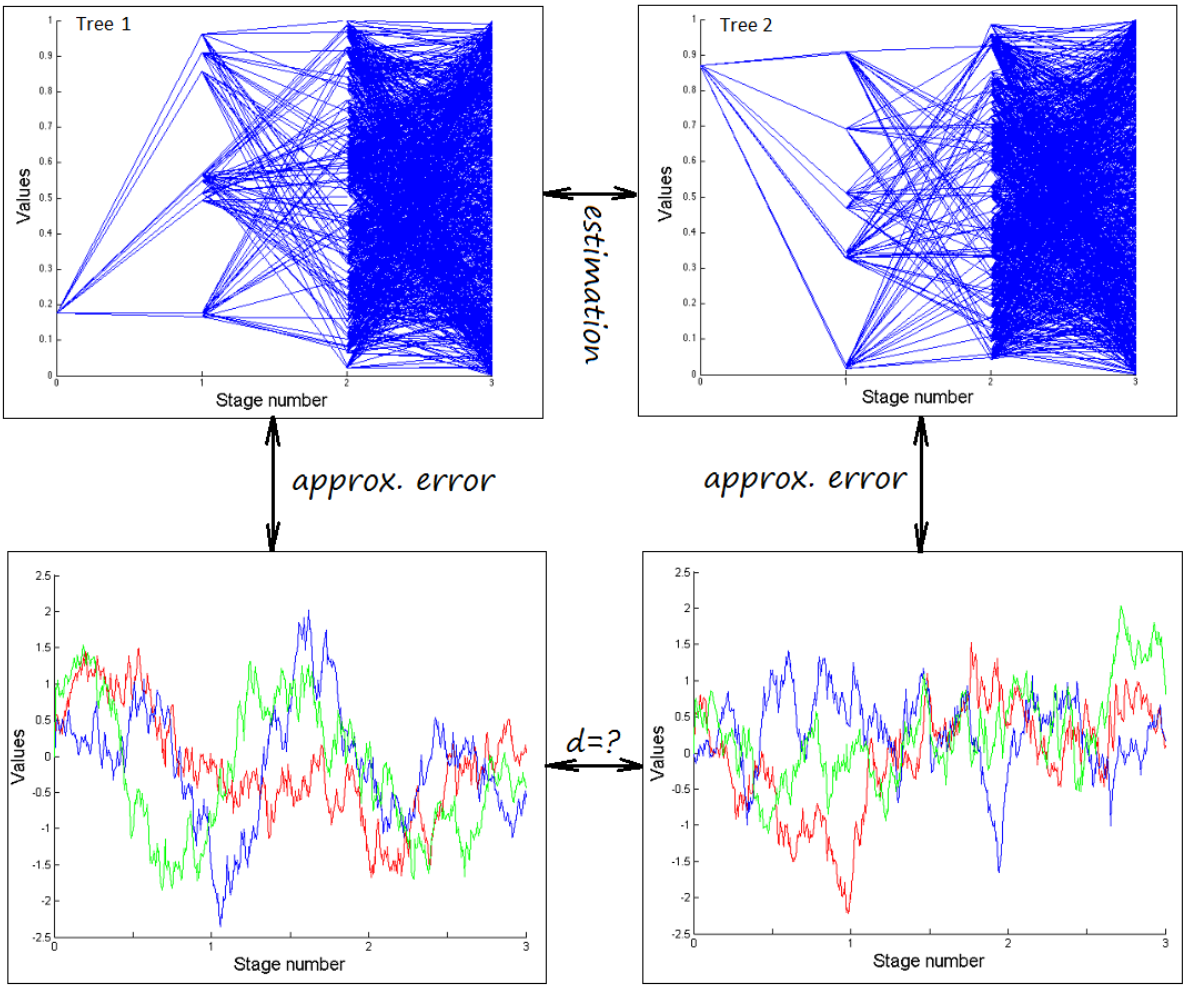


Figure 2.11: Distance approximation for the case of two infinite trees.

Suppose that $\mathbb{P}_1$ is a nested distribution of the stochastic process ξ_1, $\mathbb{P}_2$ is a nested distribution of the stochastic process ξ_2, $\widetilde{\mathbb{P}_1}$ is a nested distribution of the approximate stochatic process $\widetilde{\xi_1}$ and $\widetilde{\mathbb{P}_2}$ is a nested distribution of the approximate stochatic process $\widetilde{\xi_2}$. Corresponding distributions at each tree stage are $P_{1,t}$ and $P_{2,t}$ $\forall t = 1, ..., T$ and the joint distributions between stages $t, ..., T$ are $P_1^{t:T}$ and $P_2^{t:T}$. The similar notations hold for the nested distribution approximations $\widetilde{\mathbb{P}_1}$ and $\widetilde{\mathbb{P}_2}$.
We would like to calculate the distance $dl(\mathbb{P}_1, \mathbb{P}_2)$ between two stochastic processes (i.e. infinitely large trees).
According to the triangle inequality that holds for the nested distance, we can claim (see Fig. 2.11) $dl(\mathbb{P}_1, \mathbb{P}_2) \leq dl(\mathbb{P}_1, \widetilde{\mathbb{P}}_1) + dl(\mathbb{P}_2, \widetilde{\mathbb{P}}_1)$, and $dl(\mathbb{P}_2, \widetilde{\mathbb{P}}_1) \leq dl(\widetilde{\mathbb{P}}_1, \widetilde{\mathbb{P}}_2) + dl(\mathbb{P}_2, \widetilde{\mathbb{P}}_2)$.

Then,

$$dl(\mathbb{P}_1, \mathbb{P}_2) \leq dl(\mathbb{P}_1, \widetilde{\mathbb{P}}_1) + dl(\mathbb{P}_2, \widetilde{\mathbb{P}}_2) + dl(\widetilde{\mathbb{P}}_1, \widetilde{\mathbb{P}}_2).$$

In order to calculate the distance $dl(\mathbb{P}_1, \mathbb{P}_2)$ between two stochastic processes as a distance $dl(\widetilde{\mathbb{P}}_1, \widetilde{\mathbb{P}}_2)$ between two trees that approximate these stochastic processes, we should guarantee that the distance between stochastic process ξ_1 and its approximation is small enough (as well as the distance between stochastic process ξ_2 and its approximation): $dl(\mathbb{P}_1, \widetilde{\mathbb{P}}_1) \leq \varepsilon_1$, and $dl(\mathbb{P}_2, \widetilde{\mathbb{P}}_2) \leq \varepsilon_2$.
Let $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ be two nested distributions and let P and $\widetilde{P}$ be the pertaining multivariate distributions respectively.
For the multivariate distributions with $(K_2, ..., K_T)$-Lipschitz property the following bounds hold

Upper bounds:

$$dl(\mathbb{P}_1, \mathbb{P}_2) \leq \sum_{t=1}^{T} k_t d_{KA}(P_{1,t}, P_{2,t}) = U_1 \tag{2.27}$$

The upper bound U_1 bounds the nested distance by the sum of Kantorovich distances between conditional distributions at each stage of the tree. This upper bound does not pay attention on the distances between subtrees, i.e. it does not take into account the future values of the stochastic scenario processes $(t+1, ..., T)$ at the stage t.

$$dl(\mathbb{P}_1, \mathbb{P}_2) \leq \sum_{t=1}^{T} k^{t|t+1:T} d_{KA}(P_1^{t|t+1:T}, P_2^{t|t+1:T}) = U_2 \tag{2.28}$$

The upper bound U_2 bounds the nested distance by the sum of Kantorovich distances between conditional joint distributions at each stage of the tree. This upper bound pays attention on the distances between subtrees, i.e. it takes into account the future values of the stochastic scenario processes $(t+1, ..., T)$ at the stage t, that makes it more precise than U_1.

$$\begin{aligned} dl(\mathbb{P}_1, \mathbb{P}_2) \leq dl(\widetilde{\mathbb{P}}_1, \widetilde{\mathbb{P}}_2) \quad &+ \quad \sum_{t=1}^{T} k_{1,t} d_{KA}(P_{1,t}, \widetilde{P}_{1,t}) + \\ &+ \quad \sum_{t=1}^{T} k_{2,t} d_{KA}(P_{2,t}, \widetilde{P}_{2,t}) = U_3, \end{aligned} \tag{2.29}$$

$$\begin{aligned} dl(\mathbb{P}_1, \mathbb{P}_2) \leq dl(\widetilde{\mathbb{P}}_1, \widetilde{\mathbb{P}}_2) \quad &+ \quad \sum_{t=1}^{T} k_1^{t|t+1:T} d_{KA}(P_1^{t|t+1:T}, \widetilde{P}_1^{t|t+1:T}) + \\ &+ \quad \sum_{t=1}^{T} k_2^{t|t+1:T} d_{KA}(P_2^{t|t+1:T}, \widetilde{P}_2^{t|t+1:T}) = U_4, \end{aligned} \tag{2.30}$$

where one could choose $k^{t|t+1:T} = \max\{k_1^{t|t+1:T}, k_2^{t|t+1:T}\}$ and
$k_t = \max\{k_{1,t}, k_{2,t}\}$ if necessary for numerical calculation and use

$$dl(\mathbb{P}_1, \mathbb{P}_2) \leq dl(\widetilde{\mathbb{P}}_1, \widetilde{\mathbb{P}}_2) + \sum_{t=1}^{T} k_t \left(d_{KA}(P_{1,t}, \widetilde{P}_{1,t}) + d_{KA}(P_{2,t}, \widetilde{P}_{2,t}) \right)$$

or

$$dl(\mathbb{P}_1, \mathbb{P}_2) \leq dl(\widetilde{\mathbb{P}}_1, \widetilde{\mathbb{P}}_2) + \sum_{t=1}^{T} k_1^{t|t+1:T} \left(d_{KA}(P_1^{t|t+1:T}, \widetilde{P}_1^{t|t+1:T}) + d_{KA}(P_2^{t|t+1:T}, \widetilde{P}_2^{t|t+1:T}) \right).$$

Notice, that upper bounds U_3 and U_4 are based on the triangle inequality and on the bounds U_1 and U_2.

Lower bounds:

$$L = d_{KA}(P_1, P_2) \leq dl(\mathbb{P}_1, \mathbb{P}_2) \tag{2.31}$$

The lower bound does not take into account the tree structure at all, bounding the nested distance below by the Kantorovich distance between two joint distributions.

$$\begin{aligned} &d_{KA}(P_1, P_2) = dl(\mathbb{P}_{1,c(1)}, \mathbb{P}_{2,c(1)}) \leq ... \leq dl(\mathbb{P}_{1,c(t)}, \mathbb{P}_{2,c(t)}) \leq ... \\ &... \leq dl(\mathbb{P}_{1,c(T-1)}, \mathbb{P}_{2,c(T-1)}) \leq dl(\mathbb{P}_{1,c(T)}, \mathbb{P}_{2,c(T)}) = dl(\mathbb{P}_1, \mathbb{P}_2). \end{aligned} \tag{2.32}$$

This chain of lower bounds is received from the Theorem 2.4 and is based on the clairvoyant nested distributions $\mathbb{P}_{1,c(t)}$ and $\mathbb{P}_{2,c(t)}$ at stage t, derived from nested distributions $\mathbb{P}_1$ and $\mathbb{P}_2$. It, clearly, takes into account the tree structures of initial nested distribution $\mathbb{P}_1$ and $\mathbb{P}_2$.

Which upper bound is better ((2.27) or (2.29); (2.28) or (2.30)) depends on the bushiness of the approximate tree (Tree 2). Notice, that upper bounds (2.27) and (2.28) for the distance do not depend on the bushiness of the tree, because they do not contain approximations of the stochastic process but just the continuous distribution function P_t. The same as for the case of the distance between stochastic process and a tree, it is natural to guess that (2.27) is better than (2.29) ((2.28) is better than (2.30)), i.e. $U_1 \leq U_3$ ($U_2 \leq U_4$), if the approximation tree of the stochastic process has low bushiness and the upper bound (2.29) is better than the upper bound (2.27) ((2.30) is better than (2.28)) if the bushiness is high.
This means that the gap between the best upper and lower bounds U and L for the distance between stochastic process and a tree does not change while $U_1 \leq U_3$ ($U_2 \leq U_4$) and decreases as long as the upper bound U_3 (U_4) decreases, i.e. as long as the bushiness of the tree increases.
Hence, several methods for the nested distance approximation arise:

Increasing the bushiness: One way to receive an approximation for the nested distance $dl(\mathbb{P}_1, \mathbb{P}_2)$ between stochastic processes is to increase the bushiness of trees that approximate these stochastic processes until the necessary level ε of the gap between the lower L and the upper bound U is obtained. In this case we are talking either about upper bound (2.29) or (2.30), assuming, of course, that $U_3 \leq U_1$ and $U_4 \leq U_2$:

$$dl(\mathbb{P}_1, \mathbb{P}_2) = \lim_{bush \to \infty} dl(\widetilde{\mathbb{P}}_1, \widetilde{\mathbb{P}}_2).$$

Bounding from above: The other way to receive an approximation for the nested distance between continuous-state stochastic scenario processes is to calculate the upper bounds (2.27) or (2.28), that are not dependent on the bushiness vector of the approximation tree. Though this approximation is not so precise as the approximation received by the increase in the bushiness, but numerically it requires much less power.

Chapter 3

Optimal quantization of stochastic scenario processes

3.1 Kantorovich distance minimization

Consider a random variable ξ_t given by its continuous D-dimensional distribution function F_{ξ_t} and suppose that we should approximate this distribution by a discrete one with at most N values. Approximation of a random vector ξ_t defined on $(\Omega, \mathcal{F}, P)$ by a measurable function with at most N values can be done by:

Monte-Carlo (random) generation that chooses N random points from the distribution F_{ξ_t} and assigns equal probabilities to them (see [7, 32, 26]);

Quasi Monte-Carlo generation which basic idea is to replace random samples in Monte-Carlo methods by deterministic points that are uniformly distributed in $[0,1]^D$, where D is the number of dimensions (see [32, 26]);

Optimal quantization that minimizes the Kantorovich distance between the continuous distribution and its discrete approximation.
This type of quantization

1. finds optimal supporting points z_i, $i = 1, ..., N$ by the minimization of

$$\mathcal{D}(z_1, ..., z_N) = \int \min_i d(x, z_i) P(dx)$$

over $z_1, ..., z_N$;

2. then, given the supporting points z_i, we calculate their probabilities p_i by the minimization of the Kantorovich distance $d_{KA}(P, \sum_{i=1}^N p_i \delta_{z_i})$ (2.1) (see [8, 17, 28, 29, 35, 37, 38]).

Now suppose that $\xi = (\xi_1, ..., \xi_T)$ is the stochastic process with known joint probability distribution $F_{\xi_1,...,\xi_T}(u_1, ..., u_T)$ (with the corresponding joint density $f_{\xi_1,...,\xi_T}(u_1, ..., u_T)$). And suppose that this stochastic process is sitting on the tree with given treestructure $\mathbb{T}$. The joint density for the stochastic process $\xi = (\xi_1, ..., \xi_T)$ can be written in the form

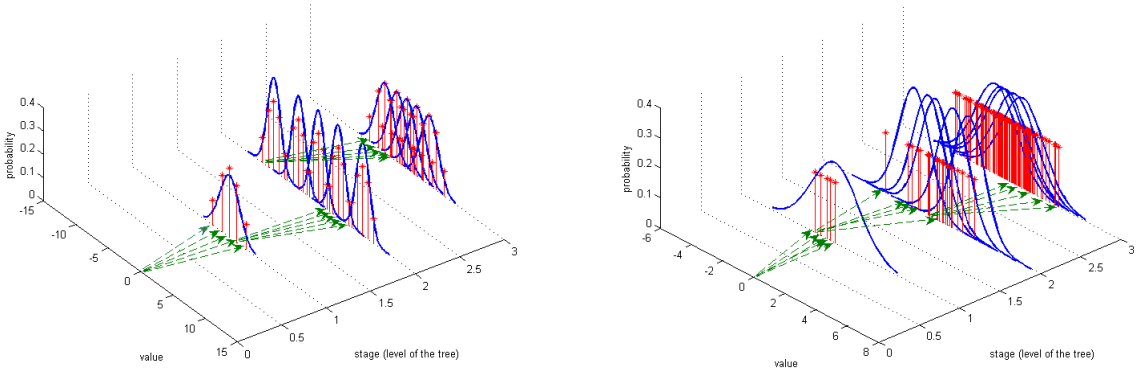
$$f_{(\xi_1,...,\xi_T)}(u_1, ..., u_T) = f_{\xi_1}(u_1) f_{\xi_2|\xi_1}(u_2|u_1) ... f_{\xi_T|\xi_1...\xi_{T-1}}(u_T|u_1...u_{T-1}),$$

where

$$f_{\xi_t|\xi_1...\xi_{t-1}}(u_t|u_1...u_{t-1}) = \frac{\int ... \int f_{\xi_1...\xi_T}(u_1, ..., u_t, v_{t+1}...v_T) dv_{t+1}...dv_T}{\int ... \int f_{\xi_1...\xi_T}(u_1, ..., u_{t-1}, v_t...v_T) dv_t dv_{t+1}...dv_T}.$$

Hence, in order to be able to generate points from the distribution at each stage of the tree, we should know the conditional distribution at each stage given the information from the previous stages of the tree. That means, we should, first of all, find the form of each $f_{\xi_t|\xi_1...\xi_{t-1}}(u_t|u_1...u_{t-1})$, $t = 1, ..., T$ in order to generate $\xi = (\xi_1, ..., \xi_T)$ with known joint probability density function $f_{\xi_1,...,\xi_T}(u_1, ..., u_T)$.
Let us assume now that all the conditional distribution functions $F_n = F_{\xi_t|\xi^{t-1}}$ of each random variable ξ_t $\forall t = 1, ..., T$ are known (index n refers to the predecessor number at the previous stage - at the stage t there are so many conditional distribution functions F_n, as many nodes N_{t-1} are at the stage $t-1$). We would like to approximate the stochastic process $\xi = (\xi_1, ..., \xi_T)$ by the tree with T stages and with known treestructure $\mathbb{T}$. As soon as the treestructure is given, the number S_n of successors of each node n of the tree is known too. Hence, to approximate the stochastic process $\xi = (\xi_1, ..., \xi_T)$ by

this tree we need to approximate each continuous F_n by S_n discrete points - that can be done either by Monte-Carlo generation or by optimal quantization. If we use optimal quantization to approximate the distribution functions, the Kantorovich distance between the distribution and its approximation at each stage of the tree is minimized [8, 17, 37]. Fig. 3.1 compares stage-wise optimal quantization (Fig. 3.1 (a)) with stage-wise random generation (Fig. 3.1 (b)) for one-dimensional Gaussian distribution sitting on the tree with given treestructure (bush=[5 5 5]) (see [17]).



a. Optimal quantization. b. Random generation.

Figure 3.1: Optimal quantization and random generation for the normal distribution.

Such a stage-wise tree approximation of the stochastic process $\xi = (\xi_1, ..., \xi_T)$ can be described in the following algorithmic form:

1. At the root (zero-stage) of the tree there is only one node (the root) which probability is one. We assume that the value ξ_0 (we denote it ξ_0 for convenience) at the root is known explicitely as soon as it refers to the present moment in time. This is true for both P and $\widetilde{P}$, that is why we do not talk abut the distance at zero-stage: $d_0(P_0, \widetilde{P}_0) = 0$;

2. First stage of the tree has N_1 nodes (this number depends on the structure of the tree). All of them are dependent only on the root and given by the unconditional distribution function F_{ξ_1}. Hence, one should generate N_1 values of ξ_1 according to the unconditional distribution F_{ξ_1} of ξ_1;

3. For each of the following stages $t = 2, ..., T$ we should generate ξ_t according to the conditional distribution $F_{\xi_t|\xi^{t-1}}$ of ξ_t conditional on the already chosen of the random variables ξ^{t-1}.

3.1.1 One-dimensional case

Till this point we have not specified the number of dimensions of the random variable ξ_t, which in general could have several dimensions. Now, consider stochastic process $\xi = (\xi_1, ..., \xi_T)$ for which each random variable ξ_t, $t = 1, ..., T$ is one-dimensional ($D = 1$) and suppose that the joint probability distribution function $F_{\xi_1,...,\xi_T}(u_1, ..., u_T)$ of this stochastic process is known. We would like to show how to approximate this stochastic process by the tree with known probabilities $p_{n,i}$ of the nodes (n is the predecessor number and $1 \leq i \leq S_n$, where S_n is a total number of successors of n).
Suppose that we have found the form of each of the conditional distributions $F_{\xi_t|\xi^{t-1}}(u_t|u_1...u_{t-1})$, $t = 2, ..., T$ and of the unconditional distribution $F_{\xi_1}(u_1)$. Fig. 3.2 shows how to construct and to approximate the conditional distributions at each stage of the tree based on the joint distribution function of $\xi = (\xi_1, ..., \xi_T)$ for the case of Gaussian distribution.

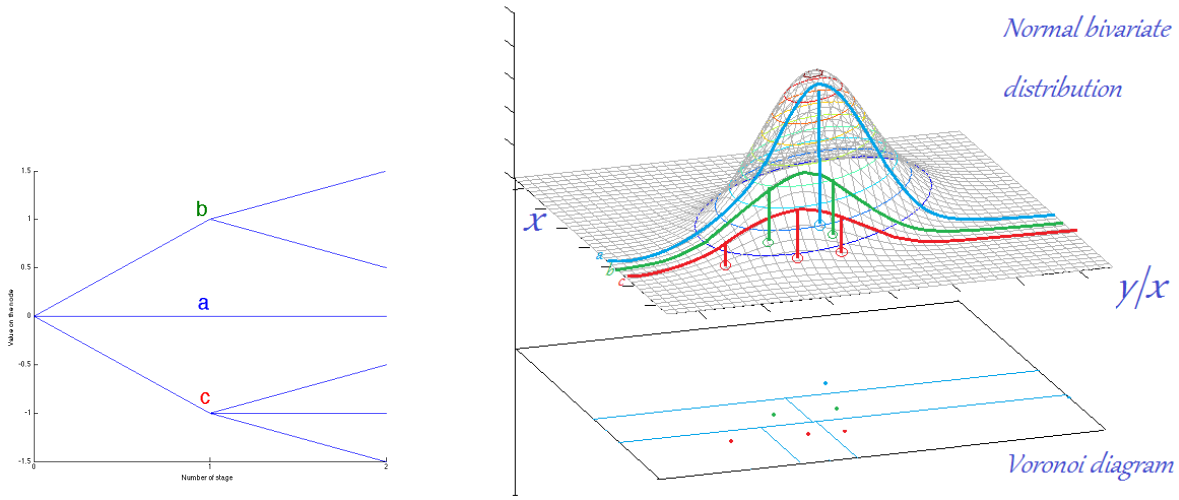


Figure 3.2: Stage-wise conditional distributions sitting on the tree (Gaussian case).

Remark 3.1. *Notice, that the joint distribution $F_{\xi_1,...,\xi_T}(u_1, ..., u_T)$ sitting on the tree in the Fig. 3.2 is two-dimensional Gaussian. However, the conditional distributions $F_{\xi_t|\xi^{t-1}}(u_t|u_1...u_{t-1})$, $t = 2, ..., T$ at each stage of the tree are one-dimensional Gaussian. Instead of approximation of the joint distribution, we approximate the one-dimensional distributions at each stage of the tree (blue, green and red cuts of the joint distribution).*

Given the probabilities $p_{n,i}$ we are able to calculate the intervals in which the realization of our stochastic process is lying for each of the nodes using the quantile function:

$$F^{-1}_{\xi_t|\xi^{t-1}}(p_{n,i}) = \inf\{u \in \mathbb{R} : \ p_{n,i} \leq F_{\xi_t|\xi^{t-1}}(u)\}.$$

Denote the bounds for the realization of the stochastic process by $b_{n,i-1}$ and $b_{n,i}$ for each node (n, i) of the tree. Calculation of these bounds can be done in two steps:

1. *Calculation at the root:*
   The index of the root is $n = 1$ and it is the predecessor for all the tree nodes at

stage $t = 1$.
Let $\overline{p}_{1,i} = \sum_{j=1}^{i} p_{1,j}$ then

$$b_{1,0} = -\infty,$$
$$b_{1,i} = F_{\xi_1}^{-1}(\overline{p}_{1,i}),\ \forall i = 1, ..., S_n$$
$$b_{1,S_1} = \infty.$$

Notice, that $b_{n,0}$ refers to the lower bound of the first successor of n and is always equal to $-\infty$.

2. *Calculation at other nodes:*
Now we calculate the bounds recursively: suppose all bounds at stage $t-1$ are calculated. For each node at the stage t consider all its predecessors at stages $1, ..., t-1$, denoting by $(1, n_2, ..., n_{t-1})$ respectively, and the corresponding intervals $[b_{1,i_1-1}, b_{1,i_1}]$, $[b_{n_2,i_2-1}, b_{n_2,i_1}]$,...,$[b_{n_{t-1},i_{t-1}-1}, b_{n_{t-1},i_{t-1}}]$, in which the historical values of the stochastic process are lying. The index i_t in the formula refers to the successor of n_t that realized in the history ($\forall t = 1, ..., T$).
Let $\overline{p}_{n,i} = \sum_{j=1}^{i} p_{n,j}$. Now we can find

$$b_{n,0} = -\infty,$$
$$b_{n,i} = F^{-1}_{\xi_t|\xi^{t-1}}(b_{1,i_1-1}, b_{1,i_1}, ..., b_{n_{t-1},i_{t-1}-1}, b_{n_{t-1},i_{t-1}}; \overline{p}_{n,i}),$$
$$b_{n,S_n} = \infty,$$

where the conditional distributions are

$$F_{\xi_t|\xi^{t-1}}(u_1, v_1, ..., u_{t-1}, v_{t-1}; \xi) = \tag{3.1}$$
$$= \frac{P(u_1 \le \xi_1 \le v_1, ..., u_{t-1} \le \xi_{t-1} \le v_{t-1}; \xi_t \le \xi)}{P(u_1 \le \xi_1 \le v_1, ..., u_{t-1} \le \xi_{t-1} \le v_{t-1})}$$

and the inverse of $F_{\xi_t|\xi^{t-1}}$

$$F^{-1}_{\xi_t|\xi^{t-1}}(u_1, v_1, ..., u_{t-1}, v_{t-1}; p)$$

is such that

$$F^{-1}_{\xi_t|\xi^{t-1}}(u_1, v_1, ..., u_{t-1}, v_{t-1}; F_{\xi_t|\xi^{t-1}}(u_1, v_1, ..., u_{t-1}, v_{t-1}; \xi)) = \xi.$$

As soon as we have calculated the bounds for the stochastic process we can approximate the process by the midpoints of the obtained intervals.

Example 3.1. *Consider the treestructure with probabilities in the Fig. 3.3 (a). Suppose that the stochastic process $(\xi_1, \ldots, \xi_T)$ is distributed as multivariate normal variable with mean vector $(\mu_1, \ldots, \mu_T)$ and non-singular covariance matrix $C = (c_{s,t})_{s,t=1,...,T}$. The value at the root is known and suppose that it is equal to 0. From [16] we know that the conditional distributions*[1] *at each stage are of the form*

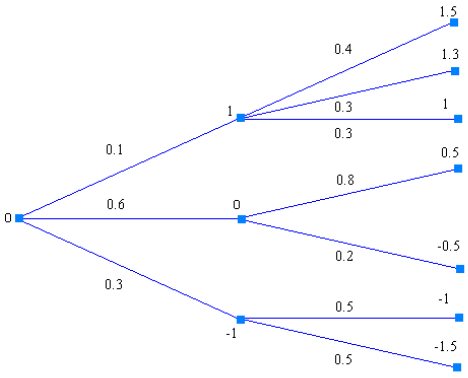
$$\mathcal{N}(\mu_t + (\xi^{t-1} - \mu^{t-1})C_{t-1}^{-1}c^t, c_{t,t} - (c^t)^\top C_{t-1}^{-1}c^t), \tag{3.2}$$

[1]Here we consider the conditional Gaussian distribution that depends on the historical values ξ^{t-1} but not on the bounds in which these historical values are lying. To improve the approximation one may construct a distribution function conditional on the bounds $[b_{1,i_1-1}, b_{1,i_1}]$, $[b_{n_2,i_2-1}, b_{n_2,i_1}]$,...,$[b_{n_{t-1},i_{t-1}-1}, b_{n_{t-1},i_{t-1}}]$ of the stochastic process.

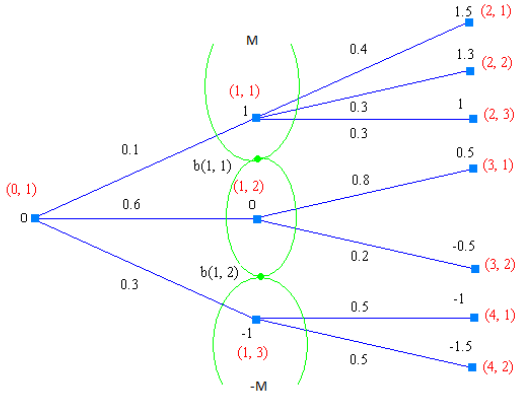
where the covariance matrix is dissected into

$$C_t = \begin{pmatrix} c_{1,1} & \dots & c_{1,t} \\ \vdots & \ddots & \vdots \\ c_{t,1} & \dots & c_{t,t} \end{pmatrix} = \left(\begin{array}{c|c} C_{t-1} & c^t \\ \hline (c^t)^\top & c_{t,t} \end{array}\right), \tag{3.3}$$

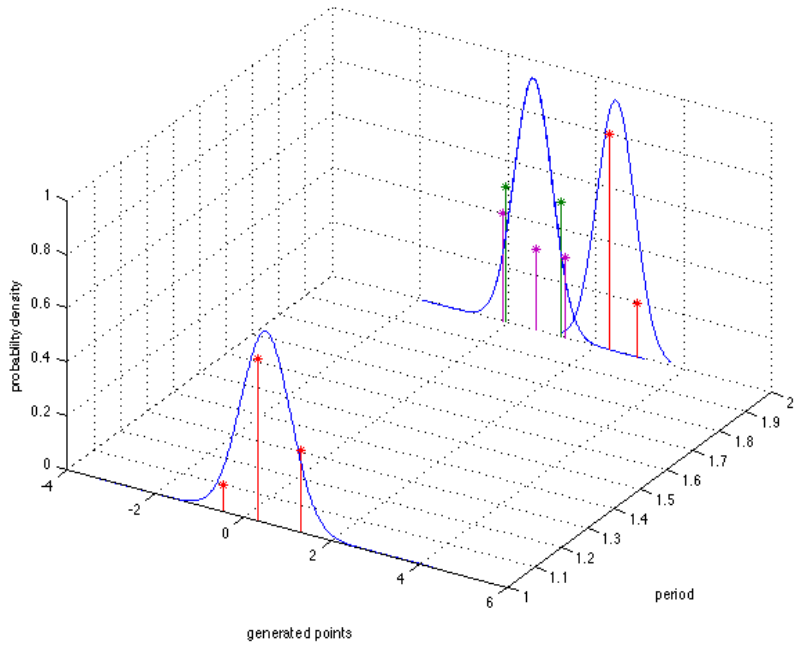
and μ^t is the history of the mean values up to time t.



(a) Given treestructure with probabilities



(b) Bounds for the Gaussian stochastic process



(c) Approximation of a normal distribution

Figure 3.3: Approximation of a normal distribution using the bounds for the stochastic process sitting on given treestructure.

Using the quantile function for the conditional normal distribution (3.2) at each stage of the tree (Fig. 3.3 (a)) we find lower and upper bounds at each node for the Gaussian distribution (for example, $b(1,1)$ and $b(1,2)$ for the stage $t = 1$ as it is shown in the Fig.

3.3 (b)). Then we approximate the stochastic process by the midpoints of the obtained bounds (Fig. 3.3 (c)).
For simplicity in this example, we use the notation $F_n(\cdot)$ to stress, that the probability distribution is conditional on the node n of the tree. Notice, that at one stage t there might be several probability distributions F_n.
At $t = 0$:

$$P(\xi_0 = 0) = 1$$

At $t = 1$ the predecessor node is $n = 1$:

$$P(-\infty < \xi_1 \leq F_1^{-1}(0.1)) = 0.1$$

$$P(F_1^{-1}(0.1) < \xi_1 \leq F_1^{-1}(0.1 + 0.6)) = 0.6$$

$$P(F_1^{-1}(0.1 + 0.6) < \xi_1 < F_1^{-1}(0.1 + 0.6 + 0.3) = \infty) = 0.3$$

The bissection is

$$(-\infty, F_1^{-1}(0.1)] \cup (F_1^{-1}(0.1), F_1^{-1}(0.7)] \cup (F_1^{-1}(0.7), \infty),$$

$$b(1, 1) = F_1^{-1}(0.1),$$

$$b(1, 2) = F_1^{-1}(0.7).$$

At $t = 2$, the first predecessor node is $n = 2$:

$$P(-\infty < \xi_2 \leq F_2^{-1}(0.4) | -\infty < \xi_1 \leq F_1^{-1}(0.1)) = 0.4$$

$$P(F_2^{-1}(0.4) < \xi_2 \leq F_2^{-1}(0.4 + 0.3) | -\infty < \xi_1 \leq F_1^{-1}(0.1)) = 0.3$$

$$P(F_2^{-1}(0.4 + 0.3) < \xi_2 < \infty | -\infty < \xi_1 \leq F_1^{-1}(0.1)) = 0.3$$

The bissection is

$$(-\infty, F_2^{-1}(0.4)] \cup (F_2^{-1}(0.4), F_2^{-1}(0.7)] \cup (F_2^{-1}(0.7), \infty).$$

At $t = 2$, the second predecessor node is $n = 3$:

$$P(-\infty < \xi_2 \leq F_3^{-1}(0.8) | F_1^{-1}(0.1) < \xi_1 \leq F_1^{-1}(0.1 + 0.6)) = 0.8$$

$$P(F_3^{-1}(0.8) < \xi_2 < \infty | F_1^{-1}(0.1) < \xi_1 \leq F_1^{-1}(0.1 + 0.6)) = 0.2$$

The bissection is

$$(-\infty, F_3^{-1}(0.8)] \cup (F_3^{-1}(0.8), \infty).$$

At $t = 2$, the third predecessor node is $n = 4$:

$$P(-\infty < \xi_2 \leq F_4^{-1}(0.5) | F_1^{-1}(0.1 + 0.6) < \xi_1 < \infty) = 0.5$$

$$P(F_4^{-1}(0.5) < \xi_2 < \infty | F_1^{-1}(0.1 + 0.6) < \xi_1 < \infty) = 0.5$$

The bissection is

$$(-\infty, F_4^{-1}(0.5)] \cup (F_4^{-1}(0.5), \infty).$$

Such an approximation method helps us to calculate the upper bound (2.21) of the nested distance between stochastic process $\xi = (\xi_1, ..., \xi_T)$ with Lipschitz property and a tree $\mathbb{T}_z$ with values $z_{t,n}$ sitting on the nodes $n = 1, ..., N$ of the tree (with nested distributions $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ respectively):

$$d(\mathbb{P}, \widetilde{\mathbb{P}}) \leq \sum_{t=1}^{T} \sum_{n \in \mathcal{N}_t} \mathbb{E}(|\xi_t - z_{t,n}| \, |\xi^{t-1} \in \mathbb{B}^{t-1}) P(\xi^t \in \mathbb{B}^t) \prod_{s=t+1}^{T} (K_s + 1),$$

where $\mathbb{B}_t$ is the bounds in the form $\mathbb{B}_t = \{\xi_t : b_{n_t, i_t - 1} \leq \xi_t \leq b_{n_t, i_t}\}$ ($\mathbb{B}^t$ is a history of all the bounds up to time t); $\mathcal{N}_t$ is a set of all node indexes n at the stage t of the tree[2].

[2]For the numerical calculation of the upper bound it is necessary to compute the multivariate probabilities $P(\xi^t \in \mathbb{B}^t)$. One can read more about the methods for it in [9, 34].

3.1.2 Multi-dimensional case

Suppose that the tree $\mathbb{T}$ is given and that at each stage of this tree a multi-dimensional random variable $\xi_t,\ t = 1, ..., T$ with known continuous multivariate distribution function F_{ξ_t} is sitting. The same as for the one-dimensional random variable, we would like to approximate the stochastic process $\xi = (\xi_1, ..., \xi_T)$ by this given tree.
In case of one-dimensional random variable the problem is relaxed to the approximation based on the quantiles of the one-dimensional distribution. In case of multi-dimensional random variable the problem is more involved, as there is no quantile function for the multi-dimensional distribution F_{ξ_t}. The way to approximate the multi-dimensional distribution is to use *the Voronoi diagram* (read [39] for the details), that allows to divide the space into given number of regions and to approximate the distribution by the midpoints of these regions (Fig. 3.4).

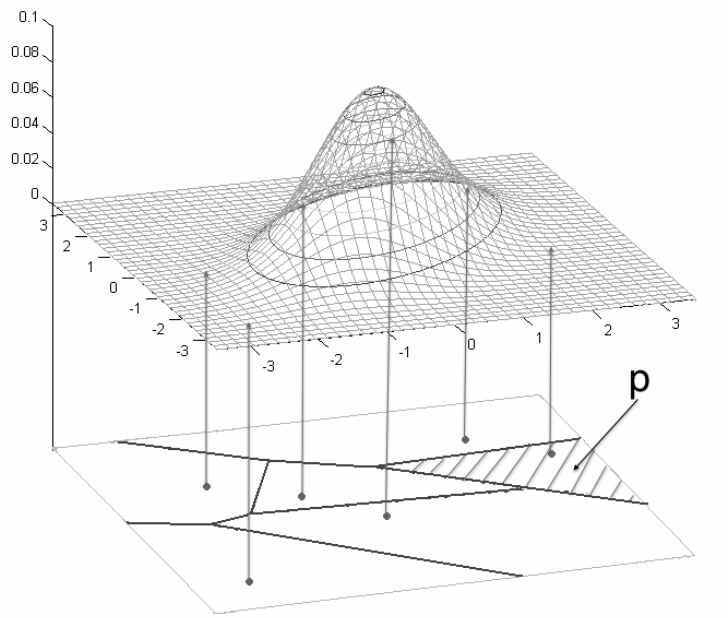


Figure 3.4: Bivariate Gaussian distribution sitting on the Voronoi diagram.

To assign probabilities to the approximate points we need to calculate the area of the regions in the Voronoi diagram: the way to do this is to generate N random points from the distribution function F_{ξ_t} and to calculate how many of them are in the specific region (suppose, k). The ration $\frac{k}{N}$ is the necessary probability p (Fig. 3.4). Alternatively, one can use other algorithms for the multidimensional integration (for example, [9, 34]).

3.1.3 Computational algorithms

Numerical algorithms for distributional quantization should solve the problem of finding N points z_1,...,z_N from a given distribution F_ξ, such that the Kantorovich distance between the given continuous distribution and the discrete distribution sitting on the obtained points is minimized [8, 11, 28, 29, 32].
At first, consider L_1-distance $d(z_1, \ldots, z_N) = \int \min_i(|x - z_i|)\, dF_\xi(x)$ to be minimized, where F_ξ is a distribution function on $\mathbb{R}^D$.
We distinguish between one-dimensional and multi-dimensional cases:

One-dimensional case as soon as the solution of the quantization problem has been found, one can define *the breakpoints* $q_i = \frac{1}{2}(z_i + z_{i+1})$ for $i = 1, \ldots, N-1$, $q_0 = -\infty$ and $q_N = \infty$, the intervals $I_i = [q_{i-1}, q_i]$ for $i = 1, \ldots N$ and the probabilities $p_i = F_\xi(q_i) - F_\xi(q_{i-1})$. Then the following equations must hold: $z_i = F_\xi^{-1}(p_1 + \cdots + p_{j-1} + \frac{p_i}{2}) = F_\xi^{-1}(\frac{F_\xi(q_{i-1}) + F_\xi(q_i)}{2})$, i.e. z_i is the conditional median within I_i;

Multi-dimensional case in case random variable is multi-dimensional, the quantile function cannot be used any more as there is no analogue for it in the multi-dimensional space. It is possible, though, to use random number generation technique in order to calculate probabilities of the Voronoi cells (or regions) [39].

Secondly, consider L_2-distance $d(z_1, \ldots, z_N) = \int \min_i(x - z_i)^2\, dF_\xi(x)$ to be minimized, where F_ξ is the distribution function on $\mathbb{R}^D$.
Iterative Algorithms 1, 2, 3, 4, 5 solve the problem for one-dimensional and multi-dimensional cases (converges at least to local solution). Moreover,

1. for a one-dimensional distribution it is possible to start with probabilities but not with points in the stage-wise quantization algorithm (see Algorithm 3);

2. in multi-dimensional spaces the start with probabilities is much more involved as there is no analogue for the quantile function in this case. However, numerical Algorithms 4, 5 work for the minimization of $d(z_1, \ldots, z_N)$ for both one-dimensional and multi-dimensional cases.

Algorithm 1 (see [25]) One-dimensional algorithm; L_1-distance

---

1. Start with $z_1, \ldots, z_N$ (starting points);

2. Find the breakpoints q_i, the intervals I_i and the probabilities p_i;

3. Define the new points as the conditional medians

$$z_i^{new} = F_\xi^{-1}\left(\frac{F_\xi(q_{i-1}) + F_\xi(q_i)}{2}\right)$$

and iterate until convergence.

Algorithm 2 (see [25]) Multi-dimensional algorithm; L_1-distance

1. Start with some starting values $z_1, \ldots, z_N$ and set $k_1 = k_2 = k_N = 1$, where k_i is the number of points in the i-th Voronoi cell $\forall i = 1, ..., N$. Set $k = N$ (the total number of points);
2. Sample ξ from distribution F_ξ. Find the index j such that
$$\|\xi - z_j\|_1 = \min_i \|\xi - z_i\|_1;$$
3. Set $z_j^{new} = z_j + \frac{1}{k^{0.75}} sign(\xi - z_j)$, $k_j^{new} = k_j + 1$ and $k^{new} = k + 1$. Leave all other indices untouched;
4. the estimates for the probabilities are $p_i = k_i/k$, $\forall i = 1, ..., N$;
5. Iterate until convergence.

Algorithm 3 (new variant) One-dimensional algorithm; L_2-distance (start with probabilities)

1. Start with equal probabilities $p_1 = p_2 = p_N = \frac{1}{N}$;
2. Find the bounds $b_i = F_\xi^{-1}(p_1 + \cdots + p_i)$ for $i = 1, \ldots N-1$, set $b_0 = -\infty$, $b_N = \infty$ and define the intervals $I_i = [b_{i-1}, b_i]$ for $i = 1, \ldots N$;
3. Define the new points as
$$z_i = \frac{\int_{I_i} x f_\xi(x)\, dx}{p_i};$$
4. Find the breakpoints $q_i = \frac{1}{2}(z_i + z_{i+1})$ for $i = 1, \ldots, N-1$ and $q_0 = -\infty$ and $q_N = \infty$;
5. Find the new probabilities as $p_i^{new} = F_\xi(q_i) - F_\xi(q_{i-1})$ and go to 2 until convergence.

Algorithm 4 (see [25]) One-dimensional algorithm; L_2-distance

1. Start with $z_1, \ldots, z_N$;
2. Find the breakpoints q_i, the intervals I_i and the probabilities p_i;
3. Define the new points as the conditional means
$$z_i^{new} = \left(\int_{I_i} x\, dF_\xi(x) \right) / p_i;$$
and iterate until convergence.

Algorithm 5 (see [25]) Multi-dimensional algorithm; L_2-distance

1. Start with some starting values $z_1, \ldots, z_N$ and set $k_1 = k_2 = k_N = 1$, Set $k = N$;
2. Sample ξ from distribution F_ξ. Find the index j such that
$$\|\xi - z_j\|_2 = \min_i \|\xi - z_i\|_2;$$
3. Set $z_j^{new} = z_j + \frac{1}{k^{0.75}}(\xi - z_j)$, $k_j^{new} = k_j + 1$ and $k^{new} = k + 1$. Leave all other indices untouched;
4. the estimates for the probabilities are $p_i = k_i/k$;
5. Iterate until convergence.

Fig. 3.5 shows how the multi-dimensional approximation algorithm works for a simple tree with Gaussian distribution sitting on it: light-grey points on the Voronoi diagrams are the random starting values for the algorithm; final discrete approximations are shown in circles.

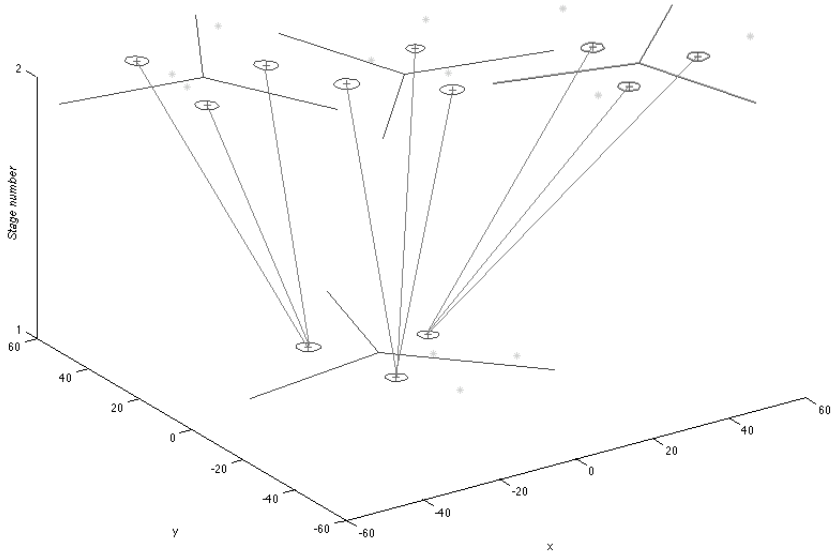


Figure 3.5: Voronoi diagrams for a normal distribution sitting on the tree with $bush = [3\ 3]$.

3.2 Nested distance minimization

Let $\mathbb{P}$ be a nested distribution, P its continuous multivariate distribution, which is dissected into the chain $P = P_1 \circ P_2 \circ ... \circ P_T$ of conditional distributions. We would like to construct another nested distribution $\widetilde{\mathbb{P}}$ that has finite tree structure $\widetilde{\mathbb{T}}$ and such that the nested distance $dl(\mathbb{P}, \widetilde{\mathbb{P}})$ is minimized.
Using the optimal quantization method on the stage-wise basis we minimize Kantorovich distance at each stage of the tree with $n_t \in \mathcal{N}_t$ being a node of the tree, i.e.

$$\begin{aligned} \widetilde{P}_t^+(\cdot|n_{t-1}) &= \mathcal{A}^+(P_t) \in \arg\min_{\widetilde{P}_t} d_{KA}(P_t(\cdot|n_{t-1}), \widetilde{P}_t(\cdot|n_{t-1})),\ \forall t = 1, ..., T \qquad (3.4) \\ \widetilde{P}^+ &= \widetilde{P}_1 \circ \widetilde{P}_2 \circ ... \circ \widetilde{P}_T, \end{aligned}$$

that gives us a nested distribution $\widetilde{\mathbb{P}}^+ = \mathcal{A}^+(\mathbb{P})$ and where we denote by $\mathcal{A}^+(\cdot)$ the operator that applies the stage-wise optimal quantization method to the nested distribution $\mathbb{P}$ at the tree structure $\widetilde{\mathbb{T}}$.
Clearly, the stage-wise minimization of the Kantorovich distance (3.4) does not necessarily lead to the minimization of the nested distance between stochastic process and the tree, i.e. it may happen that

$$\widetilde{\mathbb{P}}^+ \neq \mathcal{A}^*(\mathbb{P}) \in \arg\min_{\widetilde{\mathbb{P}}} dl(\mathbb{P}, \widetilde{\mathbb{P}}),$$

where we denote by $\mathcal{A}^*(\cdot)$ the operator that receives the approximation of $\mathbb{P}$ on the tree structure $\widetilde{\mathbb{T}}$ that minimizes the nested distance $dl(\mathbb{P}, \widetilde{\mathbb{P}})$.
Nevertheless, the stage-wise distance minimization leads to the minimization of the upper bound (2.3) for the nested distance, i.e.

$$\begin{aligned} U(\mathbb{P}) &= \min_{\widetilde{\mathbb{P}}} \sum_{t=1}^{T} d_{KA}(P_t, \widetilde{P}_t) \prod_{s=t+1}^{T} (K_s + 1). \\ \widetilde{\mathbb{P}}^+ &= \mathcal{A}^+(\mathbb{P}) \in \arg\min_{\widetilde{\mathbb{P}}} \sum_{t=1}^{T} d_{KA}(P_t, \widetilde{P}_t) \prod_{s=t+1}^{T} (K_s + 1), \end{aligned}$$

where $d_{KA}(P_t, \widetilde{P}_t) = \sup_{n_t} d_{KA}(P_t(\cdot|n_{t-1}), \widetilde{P}_t(\cdot|n_{t-1}))$.
Hence, it is important to study what the stage-wise distribution approximation gives for the nested distribution approximation, i.e. for the calculation of the minimal nested distance:

$$\widetilde{\mathbb{P}}^* = \mathcal{A}^*(\mathbb{P}) \in \arg\min_{\widetilde{\mathbb{P}}} dl(\mathbb{P}, \widetilde{\mathbb{P}}), \qquad (3.5)$$

where

- $\mathcal{A}^+(\cdot)$ is the random operator, that applies the stage-wise optimal quantization method to the preliminary nested distribution $\mathbb{P}$ and receives nested distribution $\mathbb{P}^+$ as the result (i.e. it maps $\mathbb{P}$ into $\mathbb{P}^+$);
- $\hat{\mathcal{A}}(\cdot)$ is the random operator, that produces nested distribution $\hat{\mathbb{P}}$, that is better than $\mathbb{P}^+$ in the sense of a smaller nested distance $dl(\mathbb{P}, \hat{\mathbb{P}})$;
- $\mathcal{A}^*(\cdot)$ is the random operator, that receives optimal nested distribution $\mathbb{P}^*$ based on some initial nested distribution $\mathbb{P}$.

For this, we need to understand, what the stage-wise minimization of the Kantorovich distance (3.4) means for the minimization of the nested distance (3.5), and to check if there is a better quantization of the distribution sitting on the tree than the quantization that minimizes the Kantorovich distance on the stage-wise basis, i.e. we would like to study the relationship between $\widetilde{\mathbb{P}}^+ = \mathcal{A}^+(\mathbb{P})$ and $\widetilde{\mathbb{P}}^* = \mathcal{A}^*(\mathbb{P})$. The main interest is to find out how to deal with the case when $dl(\mathbb{P}, \mathcal{A}^*(\mathbb{P}))$ is strictly smaller than $dl(\mathbb{P}, \mathcal{A}^+(\mathbb{P}))$ and how to estimate the nested distribution $\widetilde{\mathbb{P}}^* = \mathcal{A}^*(\mathbb{P})$ in this case:

- First of all, we demonstrate on the example, that

$$\exists \mathbb{P} : \ dl(\mathbb{P}, \mathcal{A}^+(\mathbb{P})) < U(\mathbb{P}), \tag{3.6}$$

  where $U(\mathbb{P})$ is the minimal upper bound (2.3).
  This statement automatically implies, that it is not enough to minimize the Kantorovich distance stage-wise to receive the minimal nested distance;

- Secondly, we would like to show that

$$\exists \hat{\mathbb{P}} : \ dl(\mathbb{P}, \hat{\mathbb{P}}) < dl(\mathbb{P}, \mathcal{A}^+(\mathbb{P})). \tag{3.7}$$

- Then, we develop an algorithm that allows to calculate the nested distribution $\hat{\mathbb{P}}$, that is a better approximation for the continuous nested distribution $\mathbb{P}$;

- Consequently, we find an algorithm to receive a better approximation for the optimal nested distribution $\widetilde{\mathbb{P}}^* = \mathcal{A}^*(\mathbb{P})$, that is the nested distribution with minimal distance to the initial distribution $\mathbb{P}$.

To prove the statement (3.6) it is enough to find such $\mathbb{P}$ that satisfies the condition. However, it is not easy to calculate the distance between continuous nested distribution $\mathbb{P}$ and its approximation $\widetilde{\mathbb{P}}^+ = \mathcal{A}^+(\mathbb{P})$ explicitely due to the infinitely large transportation plan in this case (see definition 2.5), that is why it is necessary to find a suitable algorithm for this first.

Lemma 3.1 (**Nested Distance Convergence**). *Let $\mathbb{P}$ be a nested distribution with continuous multivariate distribution P, which is dissected into the chain $P = P_1 \circ P_2 \circ \ldots \circ P_T$ of conditional distributions and suppose that there is another nested distribution $\widetilde{\mathbb{P}}$ that sits on a given finite tree structure $\widetilde{\mathbb{T}}$.*
If conditions of the Theorem 2.6 are satisfied, then there exists such finite nested distribution $\mathbb{Q}_b$ sitting on the a tree $\mathbb{T}_b$ with bushiness factor b, for which the following convergence result holds:

$$\exists \mathbb{Q}_b : \ \lim_{b\to\infty} dl(\mathbb{P}, \mathbb{Q}_b) = 0, \ \lim_{b\to\infty} dl(\mathbb{Q}_b, \widetilde{\mathbb{P}}) = dl(\mathbb{P}, \widetilde{\mathbb{P}}).$$

Proof. The result directly follows from the triangle inequality

$$dl(\mathbb{P}, \widetilde{\mathbb{P}}) \leq dl(\mathbb{P}, \mathbb{Q}_b) + dl(\mathbb{Q}_b, \widetilde{\mathbb{P}}),$$

and the conditions of the Theorem 2.6, under which the nested distance $dl(\mathbb{P}, \mathbb{Q}_b)$ is converging to zero when the bushiness factor b goes to infinity:

$$b \to \infty \Rightarrow dl(\mathbb{P}, \mathbb{Q}_b) \to 0 \Rightarrow dl(\mathbb{Q}_b, \widetilde{\mathbb{P}}) \to dl(\mathbb{P}, \widetilde{\mathbb{P}}).$$

□

Notice, that nested distributions $\hat{\mathbb{P}} = \hat{\mathcal{A}}(\mathbb{P})$, $\widetilde{\mathbb{P}}^+ = \mathcal{A}^+(\mathbb{P})$ and $\widetilde{\mathbb{P}}^* = \mathcal{A}^*(\mathbb{P})$ are sitting on the same tree structure $\widetilde{\mathbb{T}}$ and the difference between them is implied only by the values $\hat{z}$, z^+, z^* and the probabilities $\hat{p}$, p^+ and p^* correspondingly. All of them, clearly, depend on the continuous-state nested distribution $\mathbb{P}$. Approximating the continuous nested distribution $\mathbb{P}$ by the finite tree $\mathbb{Q}_b$ with high bushiness factor b, that satisfies conditions of Lemma 3.1, we can calculate the nested distance between continuous stochastic process and a tree, i.e. $dl(\mathbb{P}, \hat{\mathcal{A}}(\mathbb{P}))$ and $dl(\mathbb{P}, \mathcal{A}^+(\mathbb{P}))$, by calculating the distances between finite trees, i.e. $dl(\mathbb{Q}_b, \hat{\mathbb{P}})$ and $dl(\mathbb{Q}_b, \widetilde{\mathbb{P}}^+)$.

Example 3.2 (Gaussian distribution). *Let the stochastic process $(\xi_1, \dots, \xi_T)$ be distributed as a multivariate normal variable with mean vector $\mu = (\mu_1, \dots, \mu_T)$ and non-singular covariance matrix $C = (c_{s,t})_{s,t=1,\dots,T}$. Denote the conditional expectations and variances of ξ_t given ξ_{t-1} by $\mu_t(\xi_{t-1})$ and $\sigma_t^2(\xi_{t-1})$ correspondingly. Let $\mathbb{P}$ be a nested distribution of this stochastic process.*
According to the equation (2.16), Gaussian distribution has Lipschitz property with $K_t = ||C_{t-1}^{-1} c^t||$, hence, the upper bound (2.3) holds for it.
Now, we would like to show that the minimization of the nested distance is in general not the same as the stage-wise minimization of the Kantorovich distance and that the minimization of the nested distance may lead to a better quantization result.

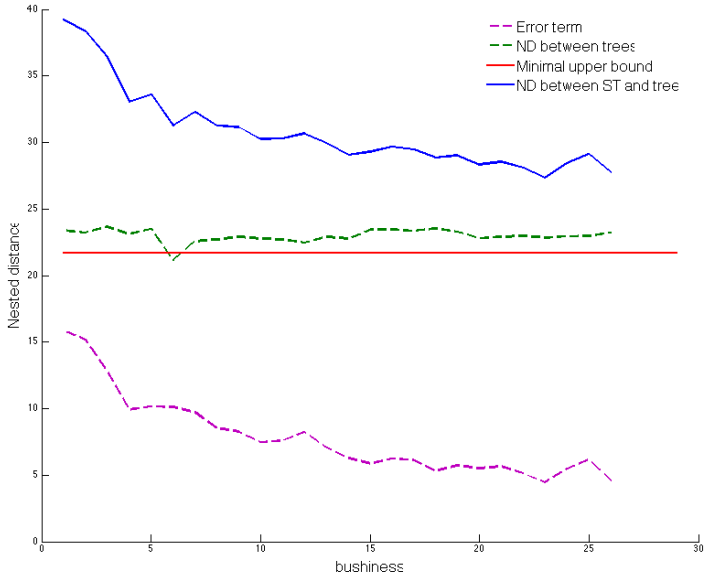


Figure 3.6: Behavior of the nested distance between Gaussian stochastic process and optimally quantized tree w.r.t. the minimal upper bound.

First of all, we fix a tree structure $\widetilde{\mathbb{T}}$.
We place values and probabilities on the tree structure $\widetilde{\mathbb{T}}$ by the stage-wise minimization of the Kantorovich distance (3.4), i.e. the nested distribution sitting on this tree is $\widetilde{\mathbb{P}}^+ = \mathcal{A}^+(\mathbb{P})$. For the given mean vector μ and the covariance matrix C we are now able to calculate the minimal upper bound (2.3) (Fig. 3.6, Fig. 3.7 (red constant line)):

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

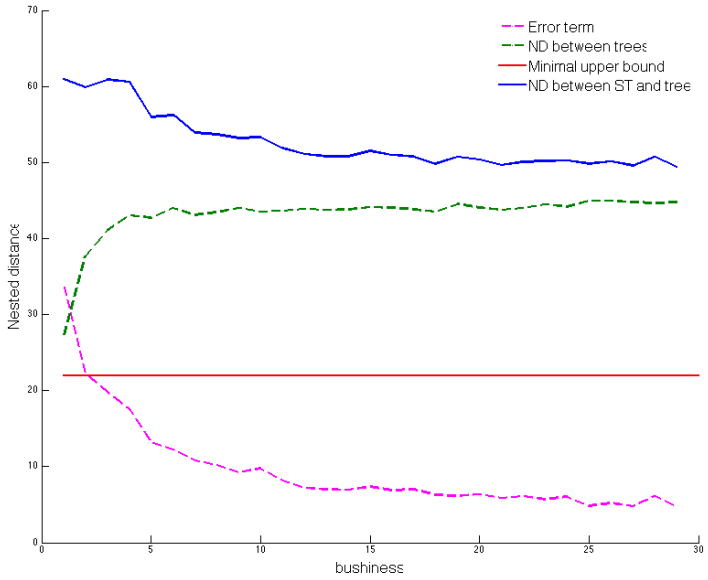


Figure 3.7: Behavior of the nested distance between Gaussian stochastic process and sampled tree w.r.t. the minimal upper bound.

Now, we sample other tree structures $\mathbb{T}_b$ *that depend on the bushiness factor* b *changing from 1 to 30, i.e.* $b = 1, 2, ..., 30$. *The height of the tree is fixed and equal to* $T = 3$.
Placing the values and the probabilities on each tree structure $\mathbb{T}_b$ *by the stage-wise minimization of the Kantorovich distance (3.4), we calculate the nested distance* $dl(\mathbb{Q}_b, \widetilde{\mathbb{P}}^+)$, *where* $\mathbb{Q}_b$ *is the nested distribution sitting on the tree* $\mathbb{T}_b$.
As soon as the nested distance $dl(\mathbb{Q}_b, \widetilde{\mathbb{P}}^+)$ *is converging to the distance* $dl(\mathbb{P}, \widetilde{\mathbb{P}}^+)$, *when bushiness* b *is going to infinity, i.e.* $b \to \infty$, *we need to check if* $\exists b : dl(\mathbb{Q}_b, \widetilde{\mathbb{P}}^+) < U(\mathbb{P})$.
Fig. 3.6 shows the behavior of the nested distance $dl(\mathbb{Q}_b, \widetilde{\mathbb{P}}^+)$ *and the error term* $dl(\mathbb{P}, \mathbb{Q}_b)$ *with respect to the minimal upper bound (2.3): it is clear that the resulting nested distance (i.e. at* $b = 30$*) is very close to the minimal upper bound* $U(\mathbb{P})$, *but does not appear to be lower than* $U(\mathbb{P})$ *(Fig. 3.6).*
From here, we can conclude that the stage-wise minimization of the Kantorovich distance in case of Gaussian stochastic process gives a fine approximation for the nested distance $dl(\mathbb{P}, \mathcal{A}^+(\mathbb{P}))$.
Nevertheless, notice, that the nested distance $dl(\mathbb{P}, \mathcal{A}^+(\mathbb{P})$ *is not necessarily the minimal nested distance, i.e.*

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

Therefore, we would like to check the statement (3.7).
For this, we sample values and probabilities on the tree structure $\widetilde{\mathbb{T}}$ *so that they are not optimal on the stage-wise basis any more, i.e. the nested distribution sitting on this tree is* $\hat{\mathbb{P}} = \hat{\mathcal{A}}(\mathbb{P})$ *now. However, we continue sampling tree structures* $\mathbb{T}_b$ *depending on the bushiness factor* b *and we place values and probabilities on each tree structure* $\mathbb{T}_b$ *by the stage-wise minimization of the Kantorovich distance (3.4) with nested distribution* $\mathbb{Q}_b$ *sitting on the tree* $\mathbb{T}_b$.
We calculate the nested distance $dl(\mathbb{Q}_b, \hat{\mathbb{P}})$, $\forall b = 1, ..., 30$ *and monitor if it falls below*

minimal upper bound (2.3) for any $\hat{\mathbb{P}}$*, i.e. we check if*

$$\exists \hat{\mathbb{P}}, b_0 : \ dl(\mathbb{Q}_b, \hat{\mathbb{P}}) < U(\mathbb{P}), \ \forall b > b_0,$$

where we use the fact that $\mathbb{Q}_b \to \mathbb{P}$*, if* $b \to \infty$*, and the simulation result of Fig. 3.6, that says* $U(\mathbb{P}) \simeq dl(\mathbb{P}, \mathcal{A}^+(\mathbb{P}))$ *for the Gaussian stochastic process.*
Fig. 3.7 shows the typical result of our tree sampling: it did not find any nested distribution $\hat{\mathbb{P}}$ *that satisfies the following inequality:*

$$dl(\mathbb{P}, \hat{\mathbb{P}}) < dl(\mathbb{P}, \mathcal{A}^+(\mathbb{P})).$$

Though these simulations show, that for the Gaussian distribution with covariance matrix independent of the historical values of the stochastic process ξ*, the following holds:*

$$dl(\mathbb{P}, \mathcal{A}^*(\mathbb{P})) \simeq dl(\mathbb{P}, \mathcal{A}^+(\mathbb{P})) \simeq \min_{\widetilde{\mathbb{P}}} \sum_{t=1}^{T} d_{KA}(P_t, \widetilde{P}_t) \prod_{s=t+1}^{T} (K_s + 1),$$

it cannot guarantee that this result is always true, especially for other types of stochastic processes. Hence, the relationship between $dl(\mathbb{P}, \mathcal{A}^*(\mathbb{P}))$*,* $dl(\mathbb{P}, \mathcal{A}^+(\mathbb{P}))$ *and* $U(\mathbb{P})$ *is still of interest here.*

Example 3.3 (Covariance Dependent on the History of Stochastic Process).

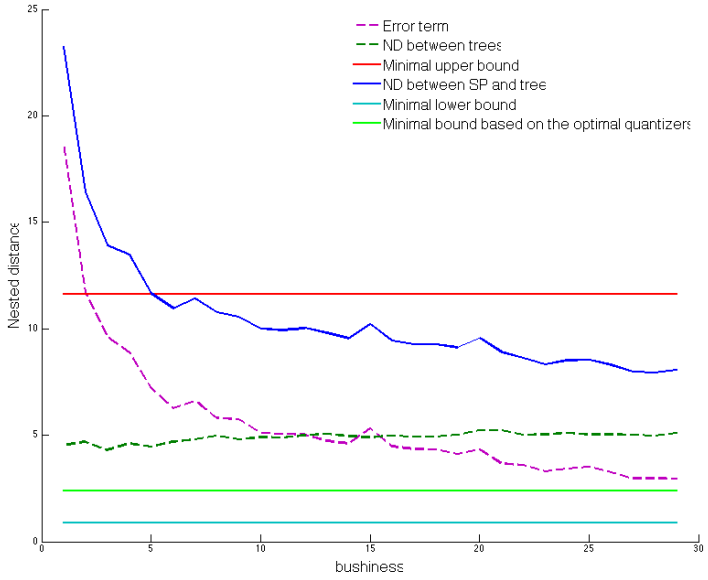


Figure 3.8: Minimal nested distance between stochastic process and the tree w.r.t. minimal upper bound.

Though the Kantorovich distance minimization still may result in a good approximation for the minimal nested distance in the Gaussian case, it is not clear if this property holds for other models. Let us, for example, check if the situation changes if we include the dependence of the covariance on the previous values to the formula (3.2) of the stage-wise normal distribution $\mathcal{N}(\mu, \sigma)$*:* $\sigma = c_{t,t} - (c^t)^\top C_{t-1}^{-1} c^t + \frac{1}{2}(\xi^{t-1} - \mu^{t-1})(\xi^{t-1} - \mu^{t-1})^T$*.*
Fig. 3.8 shows that the nested distance $dl(\mathbb{P}, \mathcal{A}^+(\mathbb{P}))$ *between stochastic process and the*

fixed tree becomes sufficiently lower than the minimal upper bound (2.3) in this case and, hence, it is not enough to find the minimal upper bound (2.3) in order to approximate the stochastic process by the tree with given treestructure. Therefore, the other possible quantization methods based on the minimization of the nested distance between stochastic process and a tree rather than on the stage-wise minimization of the Kantorovich distance (2.21) should be found and studied.

To conclude, examples (3.2) and (3.3) demonstrate, that the stage-wise minimization of the Kantorovich distance (3.4) may result in a good or bad approximation for the minimal nested distance. At the same time, these examples show that the stage-wise method is more stable and, clearly, much more time-consuming than the method, based on the iterative increase in bushiness, that can produce lower nested distance for some cases, but on the different side, it may end up in a very rough approximation.
This proves the necessity for the development of a new stable algorithm, that can produce better bounds for the nested distance, rather than upper bound (2.3) or triangle inequality.

3.2.1 Backtracking Optimal Quantization

Consider nested distribution $\widetilde{\mathbb{P}}^* = \mathcal{A}^*(\mathbb{P})$, that is one of the solutions of the nested distance minimization problem

$$\widetilde{\mathbb{P}}^* = \mathcal{A}^*(\mathbb{P}) \in \arg\min_{\widetilde{\mathbb{P}}} dl(\mathbb{P}, \widetilde{\mathbb{P}}).$$

As it is shown in the previous section, minimization of the upper bound (2.3) between continuous-state stochastic process and scenario tree in general does not result in the minimal nested distance between them, i.e.

$$dl(\mathbb{P}, \mathcal{A}^*(\mathbb{P})) \leq \min_{\widetilde{\mathbb{P}}} \sum_{t=1}^{T} d_{KA}(P_t, \widetilde{P}_t) \prod_{s=t+1}^{T} (K_s + 1). \tag{3.8}$$

Hence, the question that arises is how to obtain the minimal nested distance or, at least, a better approximation for it.
Consider the upper bound (2.5) for the nested distance between stochastic process and scenario tree. As well as the minimization of the upper bound (2.3), minimization of the upper bound (2.5) gives us an approximation for the minimal nested distance, i.e.

$$dl(\mathbb{P}, \mathcal{A}^*(\mathbb{P})) \leq \min_{\widetilde{P}^{t|t+1:T}} \sum_{t=1}^{T} d_{KA}(P^{t|t+1:T}, \widetilde{P}^{t|t+1:T}) \prod_{s=t+1}^{T} (K^{s|s+1:T} + 1). \tag{3.9}$$

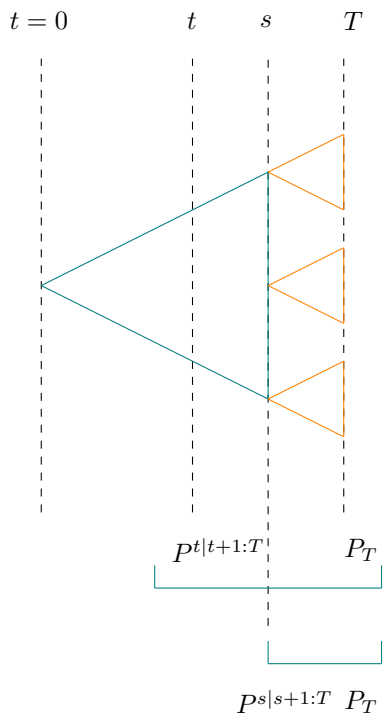


Figure 3.9: Representation of the upper bound for the nested distance on the tree.

In contrast to the upper bound (3.8), where the stage-wise optimal quantizers of P_t can be received easily for the given continuous distribution function conditional on the historical values of the stochastic process, the upper bound (3.9) is terminated over the joint

distribution functions $P^{t|t+1:T}$ of the stages $t, ..., T$ and contains stage-wise additional information about the future values of the scenario process.
From the first sight this information seem to be unreachable for the solution of the minimization problem (3.9), as soon as the future values are unknown because of the non-anticipativity condition. However, with a specific numerical procedure it is possible to solve this problem with all necessary constraints.
First of all, notice that the sum in (3.9) can be splited into two parts:

$$\begin{aligned} & \sum_{t=1}^{T} d_{KA}(P^{t|t+1:T}, \widetilde{P}^{t|t+1:T}) \prod_{s=t+1}^{T} (K^{s|s+1:T} + 1) = \\ = & \sum_{t=1}^{T-1} d_{KA}(P^{t|t+1:T}, \widetilde{P}^{t|t+1:T}) \prod_{s=t+1}^{T} (K^{s|s+1:T} + 1) + d_{KA}(P_T, \widetilde{P}_T), \end{aligned}$$

where the term $d_{KA}(P_T, \widetilde{P}_T)$ coincides with the last term of the sum in the upper bound (3.8):

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

Hence, starting from the leaves and finding the optimal quantizers for the leaves first (i.e. at stage T), we obtain the optimal quantizers for the stage $T-1$, minimizing the distance between the continuous joint distribution $P^{T-1|T}$ and its approximation $\widetilde{P}^{T-1|T}$ conditional on the quantizers for the stages T. Running this procedure backwards to the root, we obtain the optimal quantizers for the stage t, minimizing the distance between continuous joint distribution $P^{t|t+1:T}$ for the stages $t, ..., T$ and its approximation given the quantizers for the stages $t+1, ..., T$.

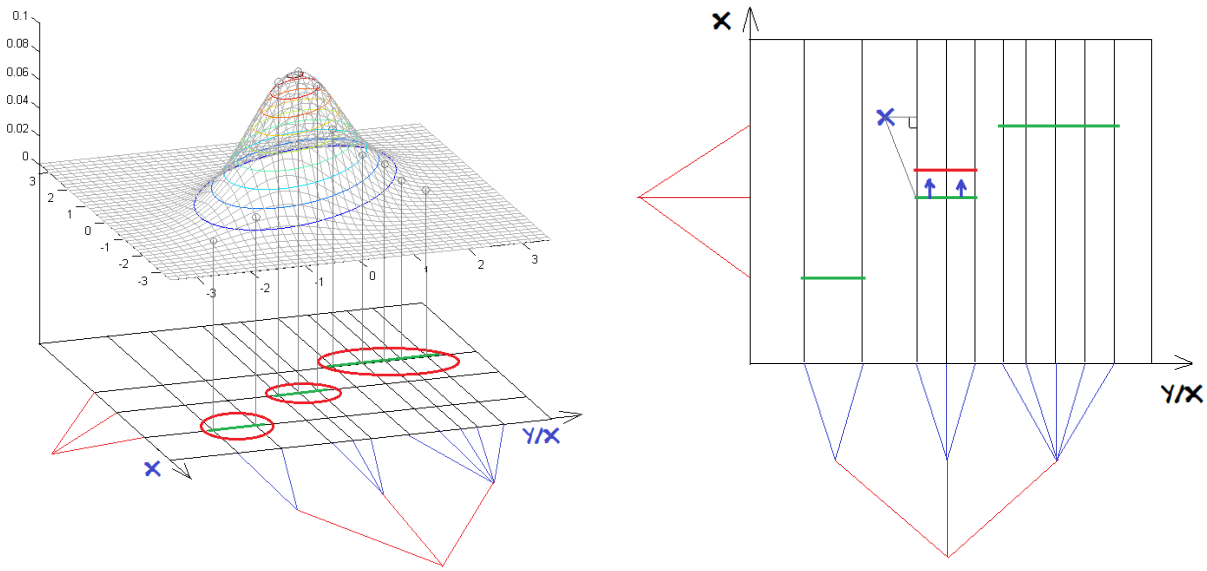


Figure 3.10: Algorithm for the scenario adaptation

Notice, that doing this we do not relax the non-anticipativity condition, we just take into account the possible values of $\xi_{t+1}, , , , \xi_T$, that follow the particular value of ξ_{t-1}; at the

same time we assign the values to ξ_t, so that the Kantorovich distance between the joint distribution for the stages $t, ..., T$ and its approximation conditional on probable future scenarios is minimized. Fig. 3.9 shows how the upper bound for the nested distance is constructed.
The numerical procedure can be discribed by two main steps (see [35]):

1. FORWARD procedure: using the stage-wise optimal quantization algorithm for the calculation of the upper bound (3.8) we receive the first approximation of the quantizers for the stages $t = 1, ..., T$;

2. BACKWARD procedure: 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 $P^{t|t+1:T}$ and its discrete approximation $\widetilde{P}^{t|t+1:T}$ is minimized.

We are going to call this procedure is either *the Backtracking Optimal Quantization* or *the QOND quantization (Quantization based On the Nested Distance minimization).*

Algorithm 6 Backtracking Optimal Quantization

1. Start with some starting values $z_1, \dots, z_N$ sitting on the nodes of the tree. The best choice for the start would be the optimal quantizers received from the stage-wise minimization of the Kantorovich distances $d_{KA}(P_t, \widetilde{P}_t)$, $\forall t = 1, ..., T$. Set $s = T$;

2. **while** $s \neq 1$ **do**
   Fix future scenarios $z_{future} = \{z_n,\ n \in \mathcal{N}_s, z_n,\ n \in \mathcal{N}_{s+1}, ..., z_n,\ n \in \mathcal{N}_T\}$ (clearly, the number of scenarios is equal to the number of leaves of the tree);
   Find all multivariate distributions $P^{s-1|s:T}$ (there are so many of them, as many nodes are at the stage $s - 1$);
   Adapt the quantizers $z_n,\ n \in \mathcal{N}_{s-1}$ based on the the fixed future scenarios z_{future}. For this, sample N random points ξ from $P^{s-1|s:T}$ and for each of them move the set of closest scenarios starting at the node $n,\ n \in \mathcal{N}_{s-1}$ towards it, keeping the future scenarios z_{future} unchanged (Fig. 3.10):

$$\begin{pmatrix} z_n^{new} \\ z_{future} \end{pmatrix} = \begin{pmatrix} z_n \\ z_{future} \end{pmatrix} + \frac{1}{k^{0.75}} \left(\xi - \begin{pmatrix} z_n \\ z_{future} \end{pmatrix} \right), \ \forall n \in \mathcal{N}_{s-1},$$

   where k is the number of sample ($k \leq N$);
   Adapt the probabilities of the quantizers $z_n,\ n \in \mathcal{N}_{s-1}$. For this, calculate the number of points that fall in the neighborhood of $\{z_n,\ n \in \mathcal{N}_{s-1}, ..., z_n,\ n \in \mathcal{N}_T\}$ (suppose, it equals to k_n) and, then, calculate new probabilities as

$$p_n = \frac{k_n}{N};$$

   Set $s = s - 1$.
   end while

Remark 3.2. *Notice, that in this algorithm we adapt several scenarios simultaneously, as soon as they start from the same node n, $n \in \mathcal{N}_{s-1}$ (Fig. 3.10);*

Backtracking Optimal Quantization always allows to improve the existing quantizers or, at least, to stay the same on the sense of minimal nested distance.

Fig. 3.11 shows the work of the Backtracking Optimal Quantization algorithm for the bivariate Gaussian distribution of the stochastic process ξ_{T-1}, ξ_T: the values of ξ_T are fixed (future scenarios) and the change is allowed and performed only in the coordinate ξ_{T-1}. As soon as the algorithm depends on the random samples, the convergence is in probability.

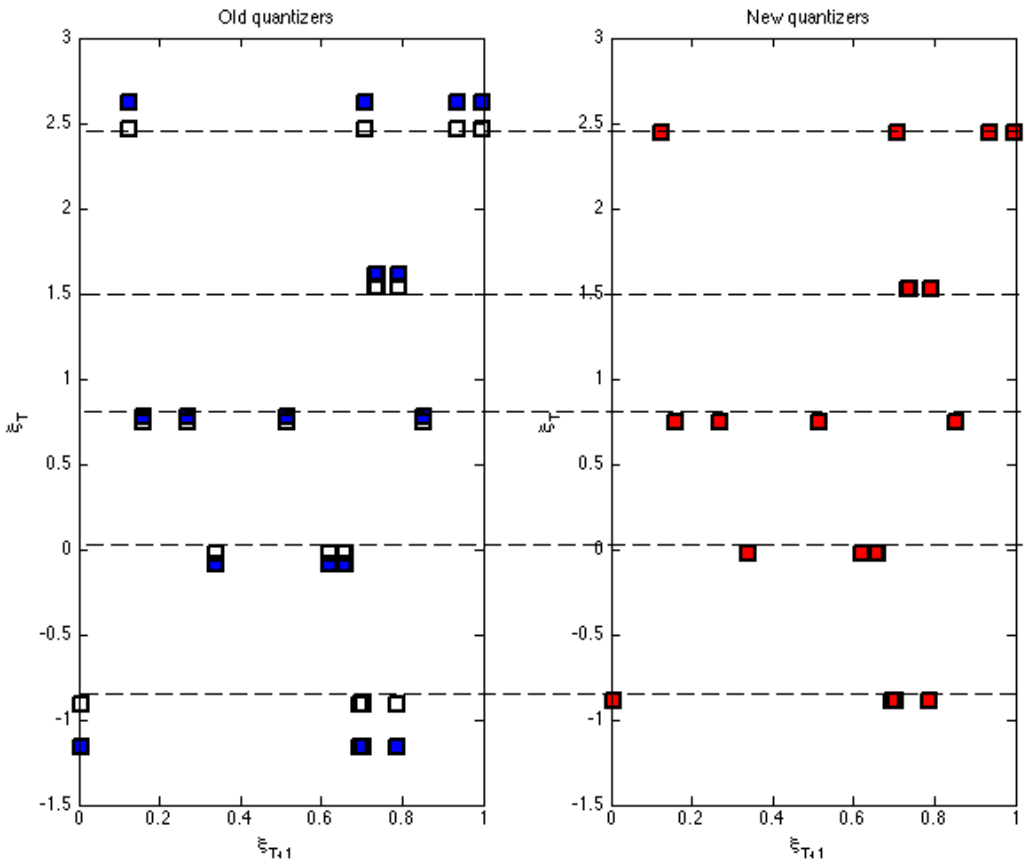


Figure 3.11: Backtracking Optimal Quantization algorithm for the bivariate Gaussian distribution.

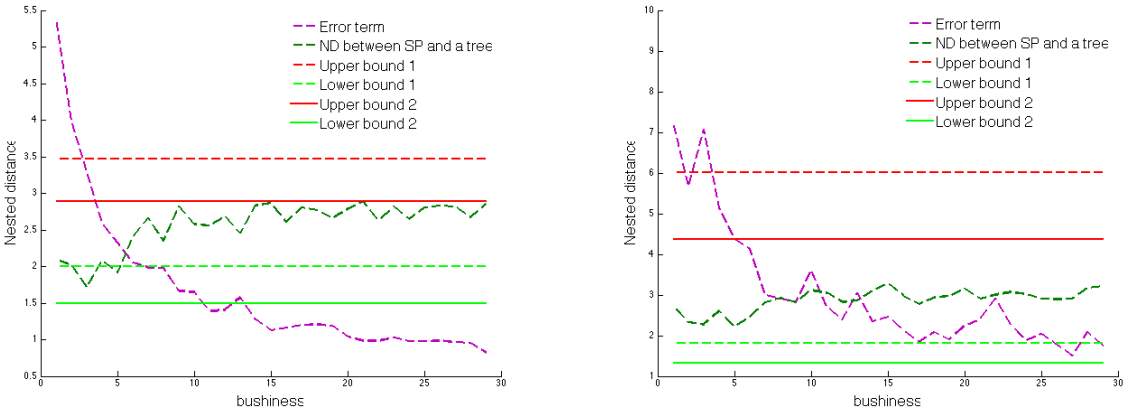
Example 3.4 (**Gaussian distribution**). *Let the stochastic process* $(\xi_1, \ldots, \xi_T)$ *be distributed as a multivariate normal variable with mean vector* $\mu = (\mu_1, \ldots, \mu_T)$ *and non-singular covariance matrix* $C = (c_{s,t})_{s,t=1,\ldots,T}$. *Denote the conditional expectations and variances of* ξ_t *given* ξ_{t-1} *by* $\mu_t(\xi_{t-1})$ *and* $\sigma_t^2(\xi_{t-1})$ *correspondingly. Let* $\mathbb{P}$ *be a nested distribution of this stochastic process.*
According to the equation (2.16), Gaussian distribution has Lipschitz property with $K_t =$

$||C_{t-1}^{-1}c^t||$, hence, upper bounds (3.8) and (3.9) hold for it.

$$U_1(\mathbb{P}) = \min_{\widetilde{\mathbb{P}}} \sum_{t=1}^{T} d_{KA}(P_t, \widetilde{P}_t) \prod_{s=t+1}^{T} (K_s + 1);$$

$$U_2(\mathbb{P}) = \min_{\widetilde{\mathbb{P}}} \sum_{t=1}^{T} d_{KA}(P^{t:t+1:T}, \widetilde{P}^{t|t+1:T}) \prod_{s=t+1}^{T} (K^{t|t+1:T} + 1).$$

Now, we would like to show that the upper bound (3.9) of the nested distance gives better approximation for the minimal nested distance $dl(\mathbb{P}, \mathcal{A}^*(\mathbb{P}))$ than the stage-wise minimization of the Kantorovich distance (3.8).
The same as in the Examples 3.2 and 3.3, we fix a tree structure $\widetilde{\mathbb{T}}$ and calculate the minimal upper bound (3.8). After this, we place values and probabilities on the tree structure $\widetilde{\mathbb{T}}$ by the minimization of the upper bound (3.9), i.e. using the Backtracking Optimal Quantization algorithm. Therefore, we get some nested distribution $\hat{\mathbb{P}} = \hat{\mathcal{A}}(\mathbb{P})$. For the given mean vector μ and the covariance matrix C, we now know the minimal upper bounds (3.8) and (3.9) (Fig. 3.12 (red dashed and constant lines correspondingly)).



a. b.

Figure 3.12: Minimal nested distance between stochastic process and the tree w.r.t. the minimal upper bound (2.22).

First of all, notice, that the upper bound (3.9) appears to be lower than the upper bound (3.8) and, hence, it better represents the minimal nested distance $dl(\mathbb{P}, \mathcal{A}^*(\mathbb{P}))$:

$$U_2(\mathbb{P}) \leq U_1(\mathbb{P}).$$

Now, we sample tree structures $\mathbb{T}_b$ depending on the bushiness factor b with stage-wise optimal values and probabilities sitting on it (i.e. nested distribution $\mathbb{Q}_b$), we calculate the nested distance $dl(\mathbb{Q}_b, \hat{\mathbb{P}})$, $\forall b = 1, ..., 30$ and monitor if it converges to the minimal upper bound (3.9), where we use the fact that $\mathbb{Q}_b \to \mathbb{P}$, if $b \to \infty$.
Fig. 3.12 shows the behavior of the nested distance $dl(\mathbb{P}, \hat{\mathbb{P}}) = dl(\mathbb{P}, \hat{\mathcal{A}}(\mathbb{P}))$ with respect to the minimal upper bounds (3.8) and (3.9). It comes out that the upper bound (3.9) is better than the upper bound (3.8) both for the case of the covariance independent of the previous values of the process (Fig. 3.12 (a)) and for the covariance with the dependency (Fig. 3.12 (b)), as well as the nested distance $dl(\mathbb{P}, \hat{\mathcal{A}}(\mathbb{P}))$ is much closer to the minimal upper bound (3.9).

Fig. 3.13 demonstrates the dependency of the minimal upper bounds on the bushiness factor for optimal quantization and for the QOND quantization of the probability distribution on the tree. QOND quantization (Backtracking Optimal Quantization) heuristically always shows a better result in terms of nested distance minimization, than the Optimal quantization.

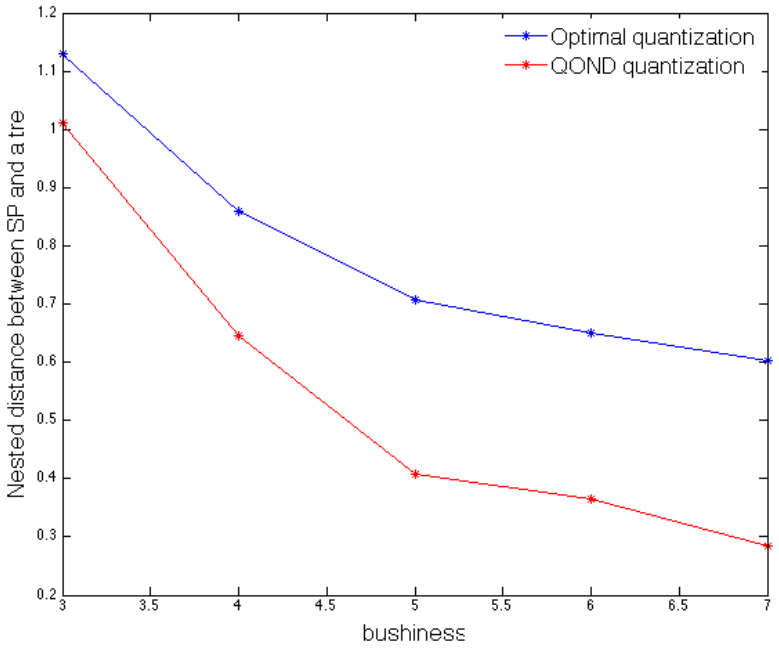


Figure 3.13: Nested distance convergence for optimal stage-wise quantization v.s. QOND quantization.

Example 3.5 (Decrease in the nested distance). *Suppose, that the nested distribution $\mathbb{P}$ corresponding to the continuous-state stochastic process ξ is approximated by the finite nested distribution $\widetilde{\mathbb{P}} = \mathcal{A}(\mathbb{P})$ sitting on a tree structure $\widetilde{\mathbb{T}}$, where $\mathcal{A}$ is the random operator, that maps $\mathbb{P}$ into $\widetilde{\mathbb{P}}$. Using the Backtracking Optimal Quantization, we can receive another approximation for $\mathbb{P}$ out of $\widetilde{\mathbb{P}}$, i.e. we can improve $\widetilde{\mathbb{P}}$:*

$$\widetilde{\mathbb{P}}_{improved} = \mathcal{B}(\widetilde{\mathbb{P}}) = \mathcal{B}(\mathcal{A}(\mathbb{P})),$$

where $\mathcal{B}$ is the random operator, that maps $\widetilde{\mathbb{P}}$ into $\widetilde{\mathbb{P}}_{improved}$.
In the Example 3.4 we have shown, that the upper bound (3.9) is lower than the upper bound (3.8) in the Gaussian case. However, it still may be possible that the nested distance $dl(\mathbb{P}, \widetilde{\mathbb{P}}_{improved}) = dl(\mathbb{P}, \mathcal{B}(\mathcal{A}(\mathbb{P})))$ is higher than the nested distance $dl(\mathbb{P}, \mathcal{A}(\mathbb{P}))$, that would mean the Backtracking Optimal Quantization leads to the deterioration in the nested distance and that the improvement in the upper bound does not lead to the improvement in the nested distance, though.
In this example we would like to show that this is not the case and, that the improvement in the upper bound leads to the improvement in the nested distance, i.e.

$$dl(\mathbb{P}, \mathcal{B}(\mathcal{A}(\mathbb{P}))) \leq dl(\mathbb{P}, \mathcal{A}(\mathbb{P})).$$

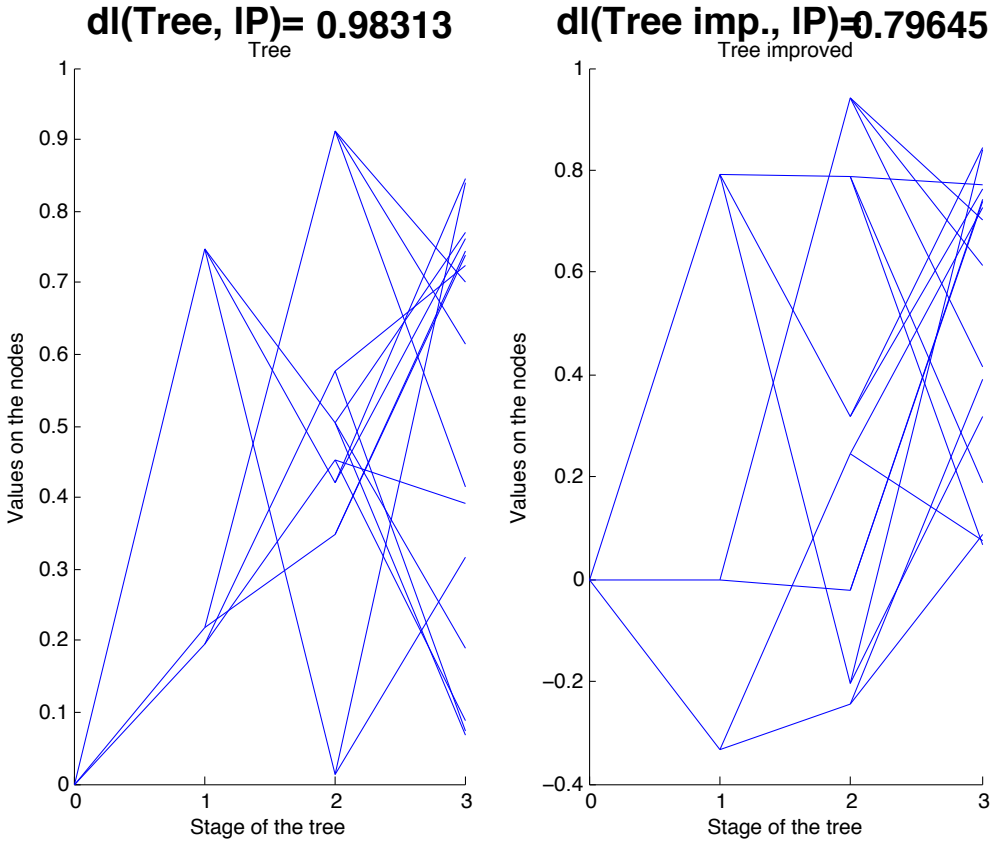


Figure 3.14: Decrease in the nested distance due to the improvement in random quantizers.

In order to estimate the true nested distances $dl(\mathbb{P}, \mathcal{B}(\mathcal{A}(\mathbb{P})))$ *and* $dl(\mathbb{P}, \mathcal{A}(\mathbb{P}))$*, we use triangle inequalities*

$$\begin{aligned} dl(\mathbb{P}, \mathcal{A}(\mathbb{P})) &\leq dl(\mathbb{Q}_b, \mathcal{A}(\mathbb{P})) + dl(\mathbb{Q}_b, \mathbb{P}); \\ dl(\mathbb{P}, \mathcal{B}(\mathcal{A}(\mathbb{P}))) &\leq dl(\mathbb{Q}_b, \mathcal{B}(\mathcal{A}(\mathbb{P}))) + dl(\mathbb{Q}_b, \mathbb{P}), \end{aligned}$$

where $\mathbb{Q}_b$ *is the nested distribution sitting on the tree* $\mathbb{T}_b$*, depending on the bushiness factor* b *and*

$$dl(\mathbb{Q}_b, \mathbb{P}) \leq \sum_{t=1}^{T} d_{KA}(P_t, Q_t) \prod_{s=t+1}^{T} (1 + K_s); \tag{3.10}$$

$$dl(\mathbb{Q}_b, \mathbb{P}) \leq \sum_{t=1}^{T} d_{KA}(P^{t|t+1:T}, Q^{t|t+1:T}) \prod_{s=t+1}^{T} (1 + K^{s|s+1:T}), \tag{3.11}$$

where Q *is the multivariate distribution corresponding to the nested distribution* $\mathbb{Q}_b$*, dissected into the chain of conditional probabilities* $Q = Q_1 \circ Q_2 ... \circ Q_T$*.*
Clearly, if the upper bounds (3.10) and (3.11) are converging to zero, i.e.

$$dl(\mathbb{Q}_b, \mathbb{P}) \to 0, \ b \to \infty,$$

then

$$\begin{aligned} dl(\mathbb{Q}_b, \mathcal{A}(\mathbb{P})) &\to dl(\mathbb{P}, \mathcal{A}(\mathbb{P})); \\ dl(\mathbb{Q}_b, \mathcal{B}(\mathcal{A}(\mathbb{P}))) &\to dl(\mathbb{P}, \mathcal{B}(\mathcal{A}(\mathbb{P}))). \end{aligned}$$

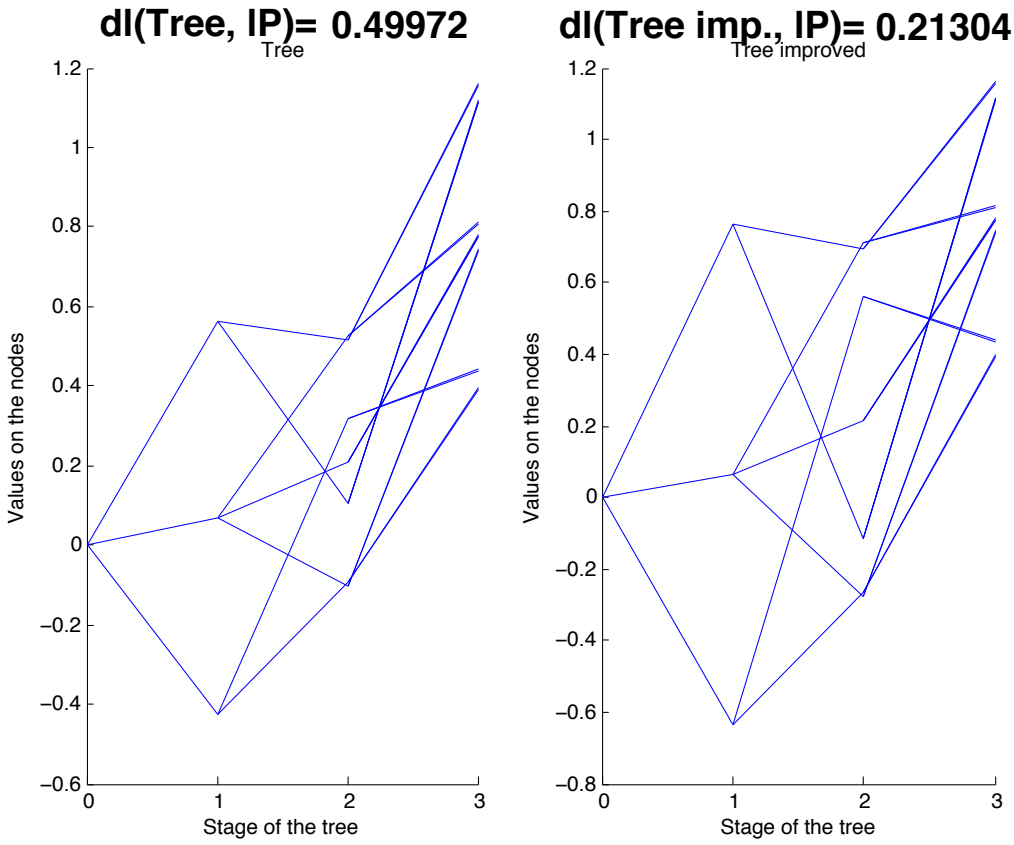


Figure 3.15: Decrease in the nested distance due to the improvement in stage-wise optimal quantizers.

Hence, in order to check if the Backtracking Optimal Quantization improves the nested distance, we need to show that

$$dl(\mathbb{Q}_b, \mathcal{B}(\mathcal{A}(\mathbb{P}))) \leq dl(\mathbb{Q}_b, \mathcal{A}(\mathbb{P})),\ b \to \infty.$$

First of all, we sample random values and probabilities on the tree structures $\widetilde{\mathbb{T}}$*, i.e. we receive* $\widetilde{\mathbb{P}} = \mathcal{A}(\mathbb{P})$ *and, then, we change the quantizers in line with the Backtracking Optimal Quantization, i.e. we calculate* $\mathcal{B}(\widetilde{\mathbb{P}})$*.*
Placing values and probabilities on the tree structure $\mathbb{T}_b$ *by the stage-wise minimization of the Kantorovich distance (3.4), we receive* $\mathcal{A}^+(\mathbb{Q}_b)$*. Now, we can calculate nested distances* $dl(\widetilde{\mathbb{P}}, \mathcal{A}^+(\mathbb{Q}_b))$ *and* $dl(\mathcal{B}(\widetilde{\mathbb{P}}), \mathcal{A}^+(\mathbb{Q}_b))$ *and we show, that*

$$dl(\mathcal{A}^+(\mathbb{Q}_b), \mathcal{B}(\widetilde{\mathbb{P}})) \leq dl(\mathcal{A}^+(\mathbb{Q}_b), \widetilde{\mathbb{P}}).$$

Secondly, we sample values and probabilities on the tree structures $\widetilde{\mathbb{T}}$ *by the stage-wise minimization of the Kantorovich distance (3.4), i.e. we receive* $\widetilde{\mathbb{P}}^+ = \mathcal{A}^+(\mathbb{P})$*. We also imply the Backtracking Optimal Quantization to the nested distributions* $\widetilde{\mathbb{P}}^+$ *and* $\mathcal{A}^+(\mathbb{Q}_b)$ *and we demonstrate that*

$$dl(\mathcal{B}(\widetilde{\mathbb{P}}^+), \mathcal{B}\{\mathcal{A}^+(\mathbb{Q}_b)\}) \leq dl(\widetilde{\mathbb{P}}^+, \mathcal{A}^+(\mathbb{Q}_b)),$$

left-hand side of which is converging to the distance $dl(\mathcal{B}(\widetilde{\mathbb{P}}^+), \mathbb{P})$ *if* $b \to \infty$*, while the right-hand side is approaching* $dl(\widetilde{\mathbb{P}}^+, \mathbb{P})$*.*

Fig. 3.14 and Fig. 3.15 show, that the change in the quantizers implied by the Backtracking Optimal Quantization algorithm leads to the decrease in the nested distance of the finite tree $\widetilde{\mathbb{T}}$ to the continuous-stage stochastic process ξ. Fig. 3.14 demonstrates that the improvement can be received, if the quantizers initially sitting on the tree structure $\widetilde{\mathbb{T}}$ are random. Moreover, Fig. 3.15 proves that the decrease in the nested distance is obtained even if the quantizers initially sitting on the tree structure $\widetilde{\mathbb{T}}$ are already stage-wise optimal.

Chapter 4

Applications

4.1 Inventory control problem

Let us consider a multi-stage inventory control problem as an example for the multi-stage stochastic optimization program which has been deeply studied in [26] and for which the explicit solution is known. That gives us an opportunity to compare approximate solutions obtained by the developed algorithms with the true solution and, hence, to see the quality of the approximation.
The grocery shop has to place regular orders one period ahead.

- The random process $\xi_1, ..., \xi_T$ defines the demand for goods of the shop at times $t = 1, ..., T$;
- x_{t-1}, $t = 1, ..., T$ is an order of goods one period ahead; the cost for ordering one piece is one. Notice, that the decision should be made one period before the random variable is observed;
- Unsold goods may be stored in the inventory with a storage loss $1 - l_t$;
- If the demand exceeds the inventory plus the newly arriving order, the demand has to be fulfilled by rapid orders (which are immediately delivered), for a price of $u_t > 1$ per piece (the shop pays $u_t > 1$ per piece);
- The selling price is s_t ($s_t > 1$) and the final inventory K_T has a value $l_T K_T$.

Let K_t be the inventory volume right after all sales have been effectuated at time t with $K_0 = 0$

$$K_t = [l_{t-1}K_{t-1} + x_{t-1} - \xi_t]^+ \geq 0,\ t = 1, ..., T.$$

Shortage at time t is

$$M_t = [l_{t-1}K_{t-1} + x_{t-1} - \xi_t]^- \geq 0,\ t = 1, ..., T.$$

These two equations can be merged into

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

The profit of the whole operation is

$$H(x_0, \xi_1, x_1, \xi_2, ..., \xi_{T-1}, \xi_T) = \sum_{t=1}^{T} s_t\xi_t - \sum_{t=0}^{T-1} x_t - \sum_{t=1}^{T} u_t M_t + l_T K_T.$$

The problem is to maximize the expected profit

$$\begin{aligned} \underset{x}{\text{maximize}} \quad & \mathbb{E}\left[\sum_{t=1}^{T}(s_t\xi_t - x_{t-1} - u_t M_t) + l_T K_T\right] \\ \text{subject to} \quad & x_t \triangleleft \mathcal{F}_t,\ t = 1, ..., T, \\ & l_{t-1}K_{t-1} + x_{t-1} - \xi_t = K_t - M_t. \end{aligned}$$

Notice, that $\mathbb{E}[\sum_{t=1}^{T} s_t\xi_t]$ does not depend on the decisions and can be removed from the optimization problem, hence, the problem is

$$\begin{aligned} \underset{x}{\text{maximize}} \quad & \mathbb{E}\left[-\sum_{t=1}^{T}(x_{t-1} + u_t M_t) + l_T K_T\right] \\ \text{subject to} \quad & x_t \triangleleft \mathcal{F}_t,\ t = 1, ..., T, \\ & l_{t-1}K_{t-1} + x_{t-1} - \xi_t = K_t - M_t. \end{aligned} \tag{4.1}$$

To write the solution of this problem in the convenient form let us introduce the notions of value-at-risk and average-value-at-risk (see [26]).

Definition 4.1 (**Value-at-Risk**). *The α-quantile of the profit distribution is called value-at-risk of level α:*

$$\mathbb{V}@\mathbb{R}_\alpha(\xi) = \mathbb{V}@\mathbb{R}_\alpha\{F\} = F^{-1}(\alpha),$$

where F is a distribution function of the random variable ξ.

Definition 4.2 (**Average Value-at-Risk**). *The average value-at-risk of level $0 < \alpha \leq 1$ of ξ is defined as:*

$$\mathbb{AV}@\mathbb{R}_\alpha(\xi) = \mathbb{AV}@\mathbb{R}_\alpha\{F\} = \frac{1}{\alpha}\int_0^\alpha F^{-1}(u)du,$$

where F is a distribution function of ξ.

Fig. 4.1 shows the value-at-risk and the average-value-at-risk graphically for the Gaussian distribution function.

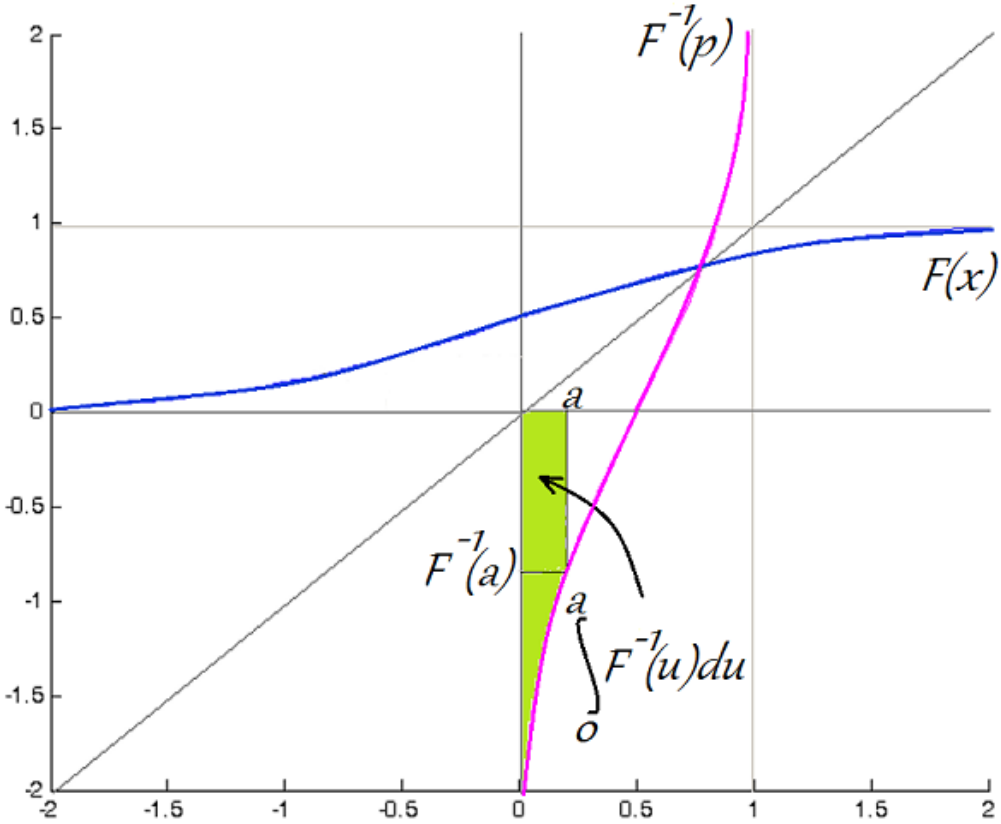


Figure 4.1: Inverse CDF of the Gaussian distribution.

According to [26], the explicit solution (i.e. optimal value m^* and optimal decision x_t, $\forall t = 1, ..., T$) of the problem (4.1) can be written in terms of the average value-at-risk:

$$\begin{aligned} m^* &= \sum_{t=1}^{T} l_t \mathbb{E}(-\xi_t) + \sum_{t=1}^{T}(1 - l_t)\mathbb{E}(\mathbb{AV}@\mathbb{R}_{\alpha_t}(-\xi_t|\mathcal{F}_{t-1})), \\ x_t &= \mathbb{V}@\mathbb{R}_{\beta_{t+1}}(\xi_{t+1}|\mathcal{F}_t) - l_t K_t, \end{aligned}$$

where $\beta_t = 1 - \alpha_t = \frac{u_t - 1}{u_t - l_t}$, $K_t = [\mathbb{V}@\mathbb{R}_{\beta_t}(\xi_t|\mathcal{F}_{t=1}) - \xi_t]^+$.
And applying the following identity with $0 < \alpha < 1$

$$\mathbb{E}(Y) = \alpha \mathbb{AV}@\mathbb{R}_\alpha(Y) - (1-\alpha)\mathbb{AV}@\mathbb{R}_{1-\alpha}(-Y)$$

to the optimal value m^*, we obtain

$$m^* = -\sum_{t=1}^{T} u_t \mathbb{E}(\xi_t) + \sum_{t=1}^{T}(u_t - 1)\mathbb{E}(\mathbb{AV}@\mathbb{R}_{\beta_t}(\xi_t|\mathcal{F}_{t-1})). \tag{4.2}$$

Example 4.1 (**Normally distributed demand**). *Suppose that the random variable ξ is distributed according to the Gaussian distribution, i.e. $\xi \sim \mathcal{N}(0,1)$.*
The distribution function of ξ is known and equal to

$$F(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-\frac{t^2}{2}} dt.$$

If $0 < \alpha \leq 0.5$, then

$$\begin{aligned} \mathbb{AV}@\mathbb{R}_\alpha(\xi) &= \frac{1}{\alpha}\int_0^\alpha F^{-1}(u)du = -\frac{1}{\alpha}\int_{-\infty}^{F^{-1}(\alpha)} F(u)du + F^{-1}(\alpha), \\ \mathbb{AV}@\mathbb{R}_\alpha(\xi) &= -\frac{1}{\alpha\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left[F^{-1}(\min(\alpha, 1-\alpha))\right]^2\right), \end{aligned}$$

for $0 < \alpha < 1$.
In general case, when $\xi \sim \mathcal{N}(\mu, \sigma^2)$

$$\mathbb{AV}@\mathbb{R}_\alpha(\xi) = \mu - \frac{\sigma}{\alpha\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left[F^{-1}(\min(\alpha, 1-\alpha))\right]^2\right),$$

for $0 < \alpha < 1$.
Now let the demand $(\xi_1, \ldots, \xi_T)$ in multi-inventory control problem be distributed as multivariate normal variable with mean vector $(\mu_1, \ldots, \mu_T)$ and non-singular covariance matrix $C = (c_{s,t})_{t=1,\ldots,T;s=1,\ldots,T}$, then

$$\begin{aligned} \mathbb{AV}@\mathbb{R}_{\alpha_t}(\xi_t|\mathcal{F}_{t-1}) &= \mathbb{AV}@\mathbb{R}_{\alpha_t}(\xi_t|\xi_{t-1}), \\ \mathbb{AV}@\mathbb{R}_{\alpha_t}(\xi_t|\mathcal{F}_{t-1}) &= \mu_t(\xi_{t-1}) - \frac{\sigma_t(\xi_{t-1})}{\alpha_t\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left[F^{-1}(\min(\alpha_t, 1-\alpha_t))\right]^2\right), \end{aligned}$$

where $0 < \alpha_t < 1$; $\mu_t(\xi_{t-1})$ and $\sigma_t^2(\xi_{t-1})$ are conditional expectations and variances of ξ_t given ξ_{t-1}.

Definition 4.3 (**Multi-period Average Value-at-Risk**). *Let $(\xi_1, ..., \xi_T)$ be an integrable stochastic process. For a given sequence of constants $c = (c_1, ..., c_T)$, probabilities $\alpha = (\alpha_1, ..., \alpha_T)$, and a filtration $\mathcal{F} = (\mathcal{F}_1, ..., \mathcal{F}_T)$, the multi-period average value-at-risk is defined as*

$$\mathbb{AV}@\mathbb{R}_{\alpha,c}(\xi; \mathcal{F}) = \sum_{t=1}^{T} c_t \mathbb{E}(\mathbb{AV}@\mathbb{R}_{\alpha_t}(\xi_t|\mathcal{F}_{t-1})).$$

Therefore, as soon as $0 < \alpha < 1$ *and* $0 < \beta = 1 - \alpha < 1$, *we can conclude*

$$\mathbb{AV@R}_{\beta,c}(\xi;\mathcal{F}) = \sum_{t=1}^{T} c_t \left[\mathbb{E}(\mu_t(\xi_{t-1})) - \frac{\mathbb{E}(\sigma_t(\xi_{t-1}))}{\beta_t\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left[F^{-1}(\min(\beta_t, 1-\beta_t))\right]^2\right)\right].$$

Using the expressions (3.2) and (3.3) for $\mu_t(\xi_{t-1})$ *and* $\sigma_t^2(\xi_{t-1})$, *that are conditional expectations and variances of* ξ_t *given* ξ_{t-1}, *i.e.*

$$\begin{aligned} \mu_t(\xi_{t-1}) &= \mu_t + (\xi^{t-1} - \mu^{t-1})C_{t-1}^{-1}c^t, \\ \sigma_t^2(\xi_{t-1}) &= c_{t,t} - (c^t)^\top C_{t-1}^{-1}c^t, \end{aligned}$$

we receive

$$\mathbb{AV@R}_{\beta_t}(\xi_t|\mathcal{F}_{t-1}) = \mu_t(\xi_{t-1}) - \frac{1}{\beta_t}\sqrt{\frac{\sigma_t^2(\xi_{t-1})}{2\pi}} \exp\left(-\frac{1}{2}\left[F^{-1}(\min(\beta_t, 1-\beta_t))\right]^2\right),$$

and, applying this equation to the optimal value (4.2), we get

$$m^* = -\sum_{t=1}^{T} u_t\mu_t + \sum_{t=1}^{T}(u_t - 1)\left[\mu_t - \frac{1}{\beta_t}\sqrt{\frac{\sigma_t^2(\xi_{t-1})}{2\pi}} \exp\left(-\frac{1}{2}\left[F^{-1}(\min(\beta_t, 1-\beta_t))\right]^2\right)\right].$$

So, the optimal value of the multi-stage inventory control problem depends only on the mean vector and the covariance matrix of the random process $\xi = (\xi_1, ..., \xi_T)$.

Example 4.2 (**Log-normally distributed demand**). *Suppose now that the stochastic process* $(\xi_1, ..., \xi_T)$ *behaves according to the log-normal distribution, i.e.*

$$\ln \xi_t = \varepsilon_t,$$

where $\varepsilon_t \sim \mathcal{N}(\mu_t, c_{t,t})$ *and* $\varepsilon_t|\varepsilon^{t-1} \sim \mathcal{N}(\mu_t + (\xi^{t-1} - \mu^{t-1})C_{t-1}^{-1}c^t, c_{t,t} - (c^t)^\top C_{t-1}^{-1}c^t)$.
Then, according to [26] and to the conditional mean and variance, i.e. to

$$\begin{aligned} \mu_t(\xi_{t-1}) &= \mu_t + (\xi^{t-1} - \mu^{t-1})C_{t-1}^{-1}c^t, \\ \sigma_t^2(\xi_{t-1}) &= c_{t,t} - (c^t)^\top C_{t-1}^{-1}c^t, \end{aligned}$$

we get

$$\mathbb{AV@R}_{\alpha_t}(\xi_t|\xi_{t-1}) = \frac{1}{\alpha_t}\exp\left(\mu_t + \frac{\sigma_t^2(\xi_{t-1})}{2}\right) \cdot F\left(F^{-1}(\alpha_t) - \sqrt{\sigma_t^2(\xi_{t-1})}\right).$$

So, the solution of the multi-inventory control problem for the log-normal demand is

$$m^* = \sum_{t=1}^{T} \exp\left(\mu_t + \frac{\sigma_t^2(\xi_{t-1})}{2}\right)\left[u_t - (u_t - 1)\frac{1}{\beta_t}\Upsilon\right],$$

where $\Upsilon = F\left(F^{-1}(\beta_t) - \sqrt{c_{t,t} - (c^t)^\top C_{t-1}^{-1}c^t}\right)$ *and* $\beta_t = 1 - \alpha_t = \frac{u_t - 1}{u_t - l_t}$.
The same as for the Gaussian stochastic process, the optimal value of the multi-stage inventory control problem with log-normally distributed demand depends only on the mean vector and the covariance matrix of the random process $\xi = (\xi_1, ..., \xi_T)$.

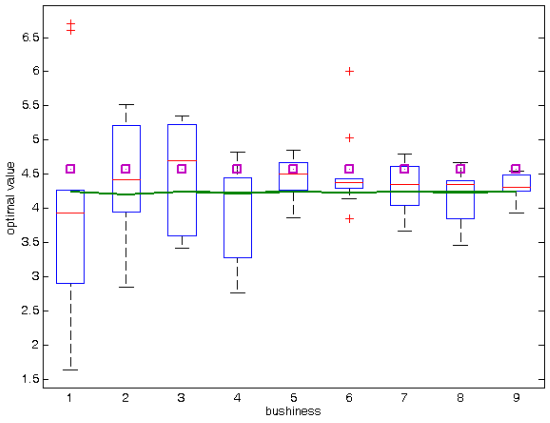


Figure 4.2: Convergence of the approximate solution to the true one for the normaly distributed demand.

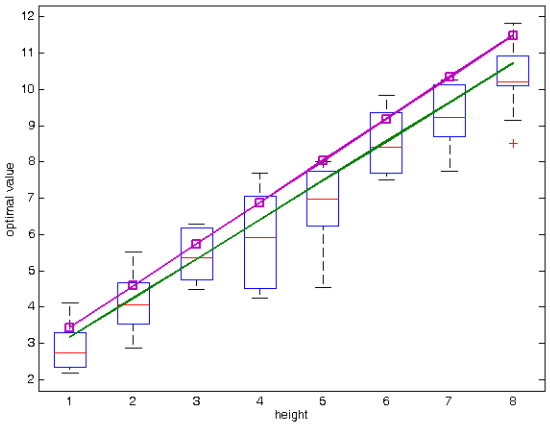


Figure 4.3: Linear behavior of the solution w.r.t. the number of periods for the normaly distributed demand.

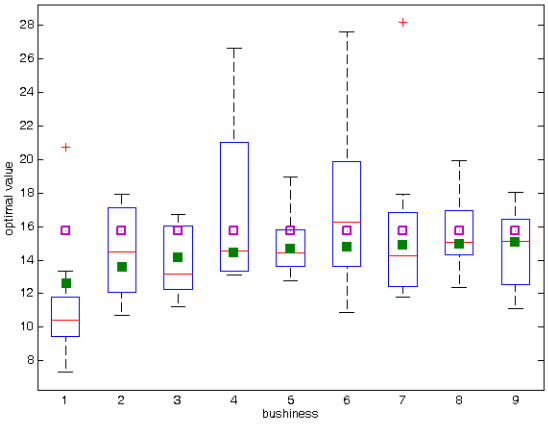


Figure 4.4: Convergence of the approximate solution to the true one.

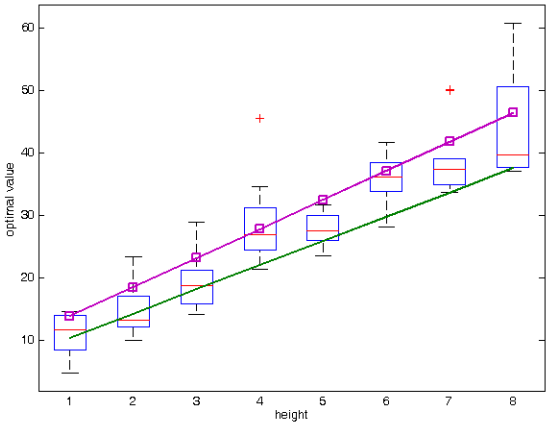


Figure 4.5: Linear behavior of the solution w.r.t. the number of periods.

As soon as the multi-inventory control problem can be solved explicitely (that we have shown for normal and log-normal demand in the Examples 4.1 and 4.2), it plays an important role for the conparison of different distribution approximation algorithms: as the distribution of the random variable ξ_t is known, we can approximate it by the simpler (discrete) distribution and find the approximate solution of the inventory control problem.
At first, consider the algorithm for generation of scenario vector $\xi = (\xi_1, ..., \xi_T)$ with known distribution F (with ξ_0 sitting at the root of the tree):

1. Zero stage of the tree has only one node that has probability 1. We assume that the value on the root is known (suppose for convenience, it is equal to $\xi_0 = 0$);

2. First stage of the tree has N_1 nodes (the number depends on the structure *pred* of the tree). We should generate N_1 values of ξ_1 according to the unconditional distribution of ξ_1 (as soon as it depends only on the root);

3. For each of the following stages t we should generate N_t points of ξ_t according to the conditional distribution of ξ_t given the distribution of the history ξ^{t-1}.

Approximating the stochastic process ξ defined on $(\Omega, \mathcal{F}, P)$ by the tree $\mathbb{T}_{b_t}$ (depending on the bushiness factor b_t at each stage of the tree), one can numerically calculate the solution of the inventory control problem using random scenario generation or optimal quantization techniques, as well as the Backtracking Optimal Quantization, that is the most crucial in the dissertation.
Fig. 4.2 and Fig. 4.4 show the explicit solution m^* (pink) for the normal and log-normal cases correspondingly and compare random generation algorithm (boxplot) with optimal quantization (green line) for both cases for the trees with different bushiness factors b_t. One can see that optimal quantization algorithm converges to the explicit solution in bushiness, as it is proved theoretically in the Theorem 2.6. Because of the stationarity of the demand processes the optimal values are linear in T (Fig. 4.3,4.5) in both cases.

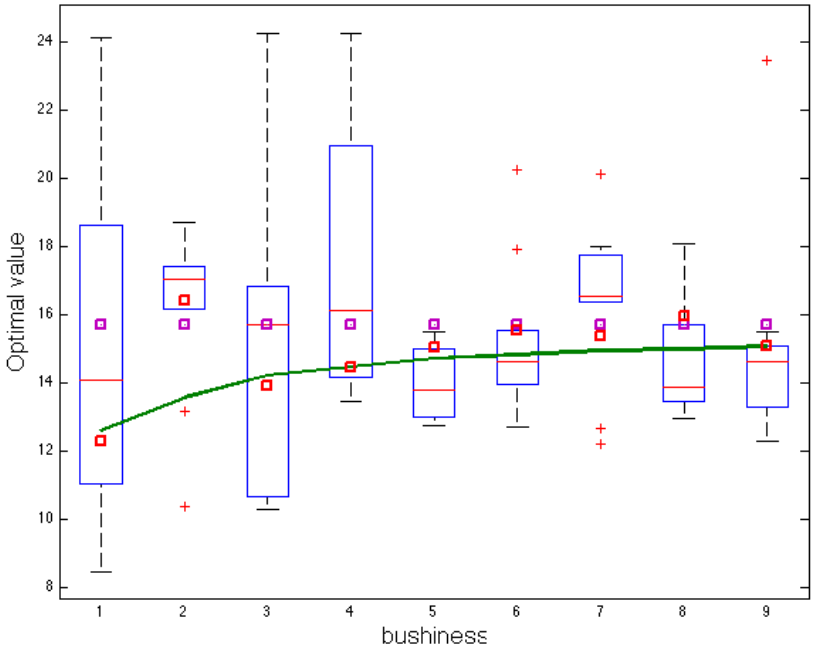


Figure 4.6: Inventory control problem solution incl. QOND quantization.

Fig. 4.6 includes the Backtracking Optimal Quantization approach (Algorithm 6) (red squares on the picture) for the solution of the inventory control problem. One can see that it estimates the solution much more explicit than random or optimal stage-wise methods for the trees with high bushiness.

4.2 Risk-management of extreme events

An essential part of stresses and risks in societies and their environments is imposed by catastrophic events (CAT-events) (read [4, 12, 13, 14, 36]). These stresses and risks can be reduced by the development of a strategy, that promotes the adaptation, resilience and resistance of societies to catastrophes and contributes to a decrease of risk and vulnerability. That is why the research, devoted to finding the optimal strategies for risk management of catastrophic events, is motivated by different needs of people on international, national and local policy levels. The goal of this chapter is to study the problem of catastrophe risk management from the multi-period and multi-hazard points of view by considering it as a multi-stage stochastic optimization program with random variables, describing the catastrophic event by their probability distributions, and to find optimal pre-event and post-event strategies [36].
The approach is to start from the description of decision model in terms of the systems dynamics functions, the constraint sets and the objective function, to specify distributions of the random variables, but not to specify a probability space on which these random variables are defined. Then these distributions should be approximated in an appropriate setting by simpler, discrete distributions and the decision functions and the optimal values in both cases are to be compared.
As there might be the dependence of one catastrophic event on another, the multi-hazard approach should be considered. For the multi-hazard case the multi-stage stochastic optimization program depends on the random variable that describes the compound loss after the sequence of natural hazard events, that can be estimated from the marginal distributions of the losses after each type of CAT-events [5, 6, 35, 36].

Research presented in this chapter was conducted at the International Institute for Applied Systems Analysis (IIASA) in line with the Young Scientist Summer Program 2011. Annual Mikhalevich Award 2011 was granted for the quality of the results.

4.2.1 Multi-period approach

Consider a government, which may lose a part of its stock S_t at time $t = 1, ..., T$ because of the random natural hazard event of loss size ξ_t. At each period it has budget B_t to spend. The government is going to take some budget allocation decision for investmets x_t, insurance z_t, consumption C_t and a new debt D_t at each period (Fig. 4.7).

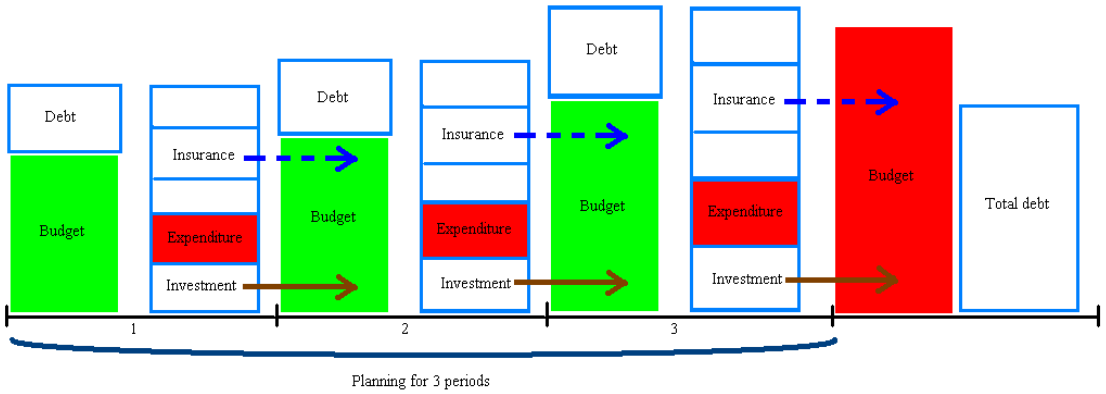
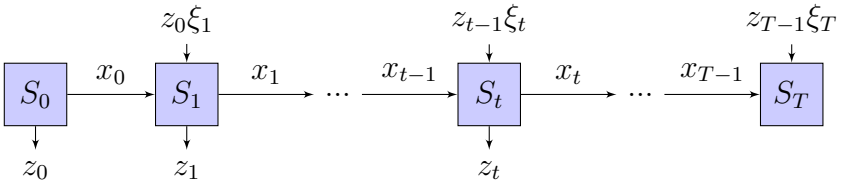


Figure 4.7: Multi-period budget allocation procedure.

As the stock is depreciating in time, the policy-maker should invest some amount in the stock of this period to rise it to the new level. Hence, the stock is summed up of the previous depreciated stock and the investment. The income from the stock is to be split into investment, insurance, consumption and savings. If the policy-maker takes debts, then he pays debt service as well. The decision about the amount of investment, insurance, consumption and debt should be the best possible for the maximization of the objective of the policy-maker, i.e. optimization problem arises.
We are going to consider three possible examples of the decision making problem for the government.
At first, suppose that the decision-maker should decide how much to spend on investment x_t and how much on insurance z_t.
In the simplest case the decision-maker would like to maximize his expected stock in the last period without considering each period consumption and without possibility of taking a debt. In this situation the stock S_{t+1} in the period $t + 1$ is summed up of the part of stock S_t and investments x_t from the previous period that are not damaged by the natural hazard event and of the insurance that is obtained in case of damage event. Budget at each period is greater or equal to the investment plus insurance with risk premium.
Components of the model: S_t stock quantity at period t; B_t available budget at period t; r return on the stock; δ depreciation rate; ξ_t random relative loss because of the natural hazard; V insurance load; E_t loss expectation $E_t = \mathbb{E}(\xi_t)$.
Decision Variables: z_t insured value of the stock; x_t amount of investment.



1. At first stage policy-maker invests and insures into the next period;

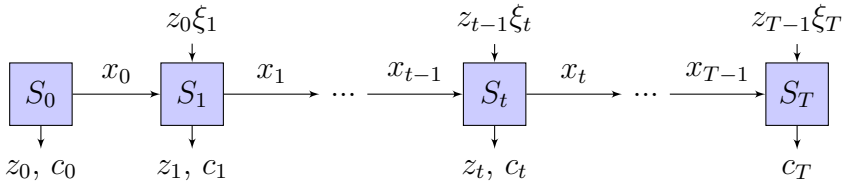
2. At all stages except the first and the last he invests, insures, loses some part of his stock in case of natural hazard event and obtains insurance from the previous period in case of natural hazard event;

3. At the last stage he only loses some part of his stock in case of natural hazard event and obtains insurance from the previous period in case of natural hazard event.

The multi-stage stochastic optimization program that describes this decision problem is:

$$\begin{aligned} \underset{z_t, x_t}{\text{maximize}} \quad & \mathbb{E}(S_T) \\ \text{subject to} \quad & x_t \triangleleft \mathcal{F}_t,\ z_t \triangleleft \mathcal{F}_t,\ t = 1, ..., T-1; \\ & B_t = rS_t \geq x_t + E_t(1+V)z_t S_t, t = 1, ..., T-1; \\ & S_{t+1} = [(1-\delta)S_t + x_t](1-\xi_t) + z_t \xi_t S_t,\ t = 1, ..., T-1. \end{aligned}$$

Here $z_t \triangleleft \mathcal{F}_t$ is a shorthand for the non-anticipativity condition, saying that all decisions must be based on the available information.
Now, suppose that the policy-maker should also decide how much to spend on consumption in each period. Moreover a minimal level for consumption per period is given.
In this case we are going to add the decision about consumption to the objective function and to the budget constraint. The constraint for the next period stock is not going to change.
Components of the model: S_t stock quantity at period t; B_t available budget at period t; r return on the stock; δ depreciation rate; ξ_t random relative loss because of the natural hazard V insurance load; E_t loss expectation $E_t = \mathbb{E}(\xi_t)$.
Decision Variables: c_t consumption at period t; z_t insured value of the stock; x_t amount of investment.



This can be written in the form of multi-stage stochastic optimization program:

$$\begin{aligned} \underset{c_t, z_t, x_t}{\text{maximize}} \quad & (1-\alpha)\sum_{t=1}^{T} \rho^{-t}\mathbb{E}(c_t) + \alpha\mathbb{E}(S_T) \\ \text{subject to} \quad & x_t \triangleleft \mathcal{F}_t,\ z_t \triangleleft \mathcal{F}_t,\ t = 1, ..., T-1,\ c_t \geq c,\ t = 1, ..., T; \\ & B_t = rS_t \geq c_t + x_t + E_t(1+V)z_t S_t, t = 1, ..., T-1; \\ & rS_T \geq c_T; \\ & S_{t+1} = [(1-\delta)S_t + x_t](1-\xi_t) + z_t \xi_t S_t,\ t = 1, ..., T-1, \end{aligned}$$

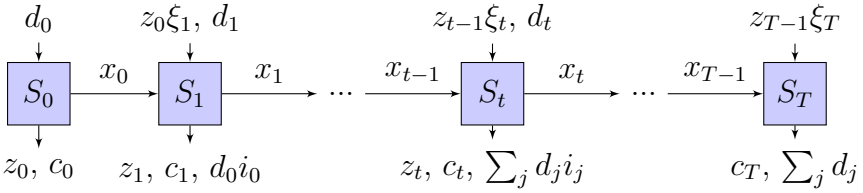
where

1. at first stage policy-maker consumes, insures and invests into the next period;

2. at all stages except the first and the last he invests, insures, consumes, loses some part of his stock in case of natural hazard event and obtains insurance from the previous period in case of natural hazard event;

3. at the last stage he consumes, loses some part of his stock in case of natural hazard event and obtains insurance from the previous period in case of natural hazard event.

Finally, let's add the possibility of taking loans at each period except for the last one. At each period except the first one the interest on the debt must be paid. During the last period all the principals should be returned.
The multi-stage stochastic optimization problem changes in the following manner in this case.
Components of the model: S_t stock quantity at period t; B_t available budget at period t; r return on the stock; δ depreciation rate; ξ_t random relative loss because of the natural hazard; i_t interest rate of the debt; V insurance load; E_t loss expectation $E_t = \mathbb{E}(\xi_t)$.
Decision Variables: d_t debt that might be taken at each period t; c_t consumption at period t; z_t insured value of the stock; x_t amount of investment.



1. At first stage policy-maker consumes, insures and invests into the next period and has a opportunity to take debt;

2. At all stages except the first and the last he invests, insures, consumes, loses some part of his stock in case of natural hazard event and obtains insurance from the previous period in case of natural hazard event; he has the opportunity to take debt and should return interests on previously taken debts;

3. At the last stage he consumes, loses some part of his stock in case of natural hazard event and obtains insurance from the previous period in case of natural hazard event; he should return all taken debts.

This can be written in the form of multi-stage stochastic optimization program:

$$
\begin{aligned}
\underset{d_t,c_t,z_t,x_t}{\text{maximize}} \quad & (1-\alpha)\sum_{t=1}^{T}\rho^{-t}\mathbb{E}(c_t) + \alpha\mathbb{E}(S_T) \\
\text{subject to} \quad & x_t \triangleleft \mathcal{F}_t,\ z_t \triangleleft \mathcal{F}_t,\ t=\overline{1,T}\ c_t \geq c,\ , d_t \leq d,\ t=\overline{1,T}; \\
& rS_t - \sum_{j=1}^{t-1} d_j i_j \geq c_t + x_t + E_t(1+V)z_t S_t, t=\overline{1,T-1}; \\
& rS_T - \sum_{t=1}^{T} d_t \geq c_T; \\
& S_{t+1} = [(1-\delta)S_t + x_t](1-\xi_t) + z_t\xi_t S_t + d_{t+1},\ t=\overline{1,T-2}; \\
& S_T = [(1-\delta)S_{T-1} + x_{T-1}](1-\xi_T) + z_T\xi_T S_T.
\end{aligned}
\tag{4.3}
$$

In all these decision problems the attitude towards the risk can be introduced by the utility function $u(c) = \frac{c^\gamma - 1}{\gamma}$ with a constant relative risk aversion $R = 1-\gamma$. Changing γ we can

introduce risk-averse ($\gamma < 1$), risk-neutral ($\gamma = 1$) or risk-loving ($\gamma > 1$) policy-maker. In this case objective function will be

$$\underset{d_t, c_t, z_t, x_t}{\text{maximize}} \quad (1-\alpha)\sum_{t=1}^{T} \rho^{-t}\mathbb{E}(u(c_t)) + \alpha\mathbb{E}(u(S_T))$$

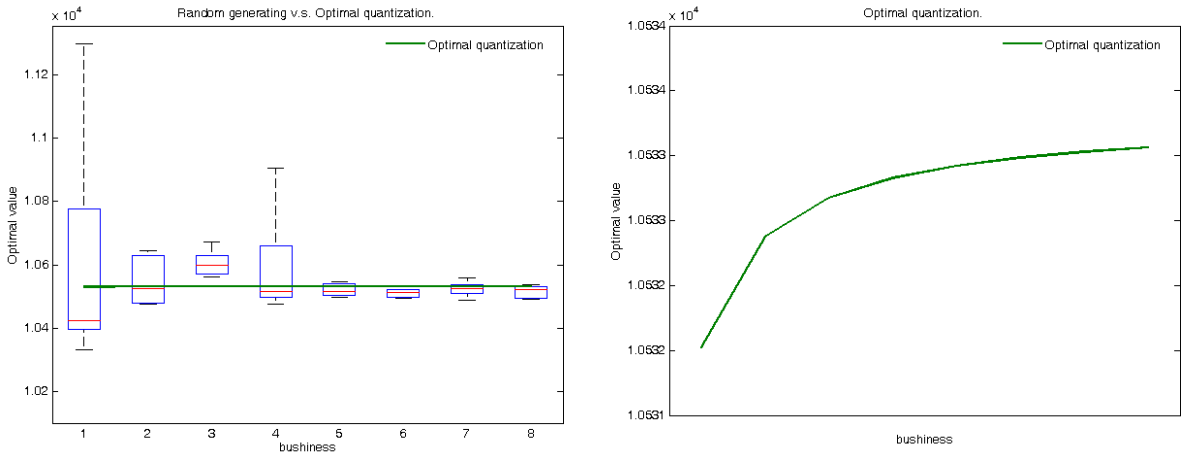
As large deviations are possible in the random process ξ that describes losses after natural hazard event the loss distribution of the natural hazard event in the specific region could be choosen of the Pareto type that is heavy-tailed, which probability density function is

$$f_X(x) = k\frac{x_m^k}{x^{k+1}}, \ \forall x \geq x_m$$

and

$$f_X(x) = 0, \ \forall x < x_m.$$

Suppose that based on the historical observations in this region the policy-maker derives scale and shape parameters x_m and k of the Pareto type loss distribution and solves the budget allocation decision problem (4.3) with a use of the multi-stage stochastic approximation methods.
Suppose that the structure of the tree is described by the bushiness vector *bush*. Increasing the bushiness of the tree and generating Pareto losses at each stage independently of the previous one and solving the problem for each considered number of leaves and for generated losses we should obtain convergence to the solution of the initial problem (Fig. 4.8: a, b).



a. b.

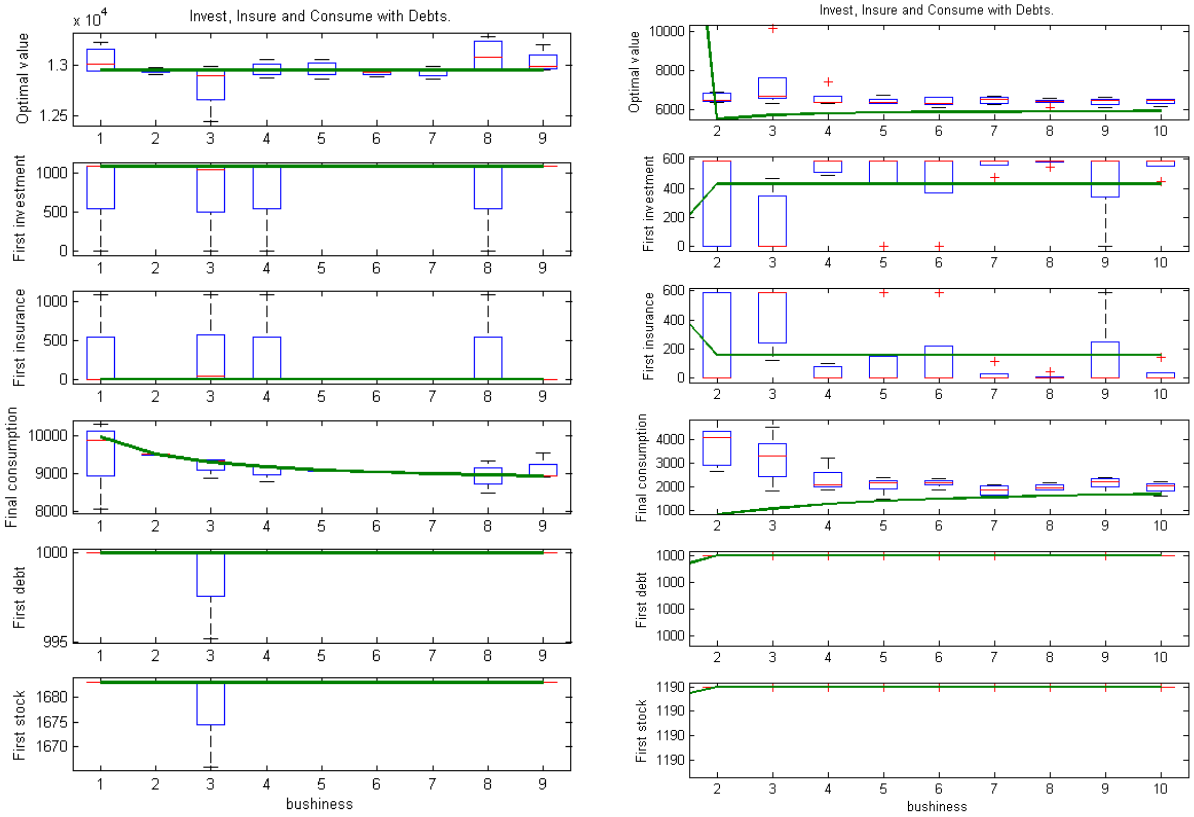
Figure 4.8: Random generating v.s. Optimal quantization

In the Fig. 4.8 one can see the convergence of the optimal value for the problem with investment, insurance, consumption and debts in case of optimal quantization and random generating. The policy-maker is supposed to be *risk-neutral* that means $\gamma = 1$.
From Fig. 4.8 we can conclude that:

- Increasing the bushiness of the tree we obtain convergence both for the random generating and for optimal quantization technique;

- Convergence in probability is for the case of random generating;
- Convergence of the optimal value is for the case of optimal quantization;
- Optimal quantization technique implies faster convergence than random generating technique.

If policy-maker is risk-neutral ($\gamma = 1$) the solution for the decision problem is of "bang-bang"-type (the "bang-bang" solution means, that the optimal decision is either to allocate all the budget to insurance or to investment). Moreover the decision for this case is to invest much and not to insure at all. At the initial stage the policy-maker is going to take the highest possible debts and to make people consume the least possible amount in order to maximize the final consumption. If policy-maker is risk-averse ($\gamma < 1$) the decision is both to invest and to insure. The proportion of the insurance depends on the given parameters, i.e. depreciation rate, interest rate, initial stock and so on.



a. Risk-neutral case.

b. Risk-averse case.

Figure 4.9: Risk-neutral vs. Risk-averse (first decision)

In the Fig. 4.9 the risk-neutral case is compared with the risk averse. The conclusion is that:

- Risk-neutral policy maker does not insure but only invest;
- The debt for the risk-neutral policy-maker is maximal possible;

- Risk-averse policy maker goes both for insurance and investment;
- At the initial stage the consumption is minimal and the debt is maximal possible.

In order to understand the quality of the decision it is necessary to consider how this decision changes in time while the expected loss in all the periods remains the same.

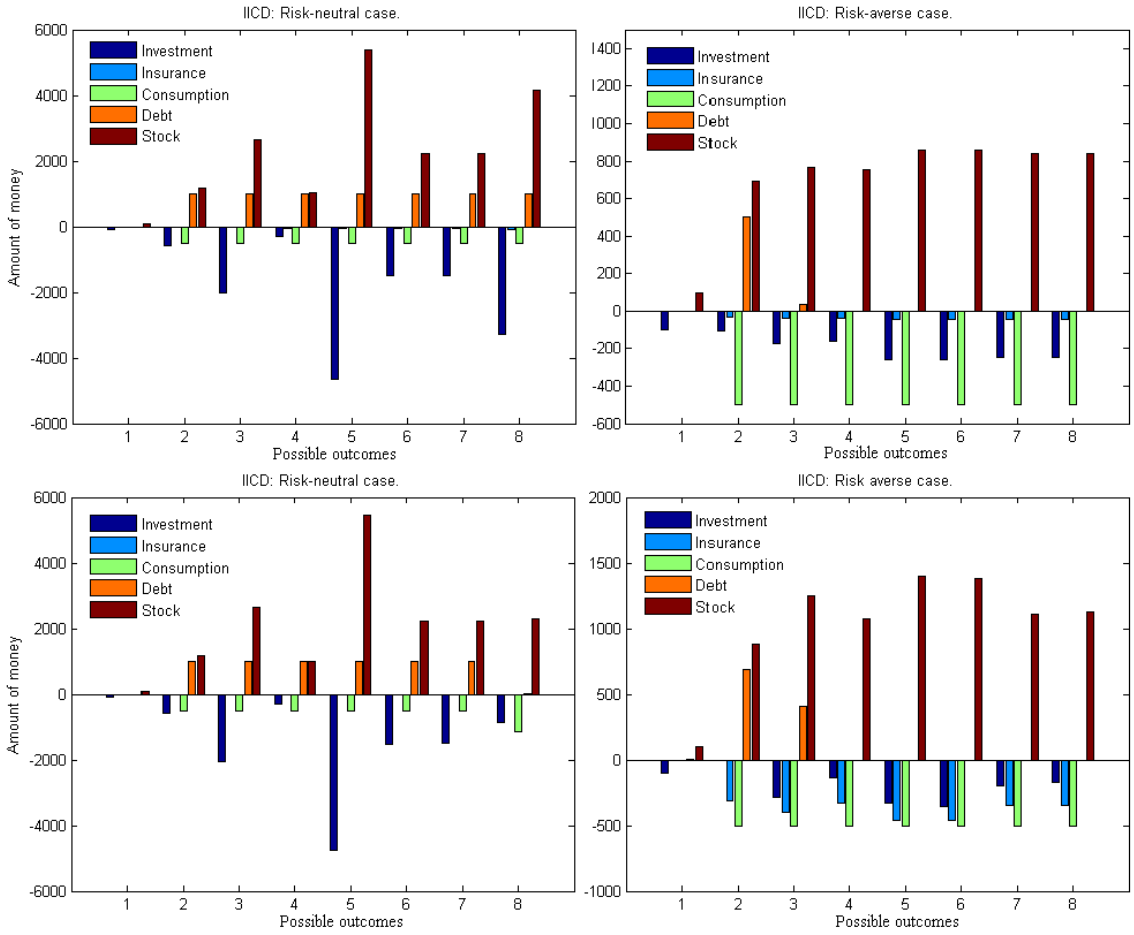


Figure 4.10: Risk-neutral vs. Risk-averse for all possible outcomes

In the Fig. 4.10 from the histogram one can see how the decision about investment, insurance, consumption and debts depends on the *possible outcome*. We use the term possible outcome meaning that at the next period there are several possible losses with different probabilities given by the loss distribution. Using the optimal quantization technique it is possible to obtain these possible losses for each stage. If we fix the tree then we fix the number of possible losses at each particular stage. The more the number is the closer the result to the desired solution.
On the left-hand side of the Fig. 4.10 one can see two pictures that correspond to the case of risk-neutral policy-maker. The lower one shows the optimal decision for losses with higher possible deviation than the upper one.
Hence, from these two pictures we can derive that

- Risk-neutral policy-maker does not insure, but only invest;

- In case of lower possible deviations of losses the optimal debts tend to be maximal possible while the consumption tends to be on the lower bound of the life standard in all the periods;
- In case of higher possible deviations of losses the optimal debts tend to decrease while the consumption tends to increase over periods. Outcome 8 from the low left picture of Fig. 4.10 confirms it.

On the right-hand side of the Fig. 4.10 one can see two pictures that correspond to the case of risk-averse policy-maker. The lower one shows the optimal decision for losses with higher possible deviation than the upper one. From these pictures we can conclude that

- Risk-averse policy-maker goes both for investment and insurance;
- The policy-maker tries to avoid debts and takes only the amount that is necessary for the recovery after CAT-events;
- The higher the deviation of losses the more policy-maker should insure.

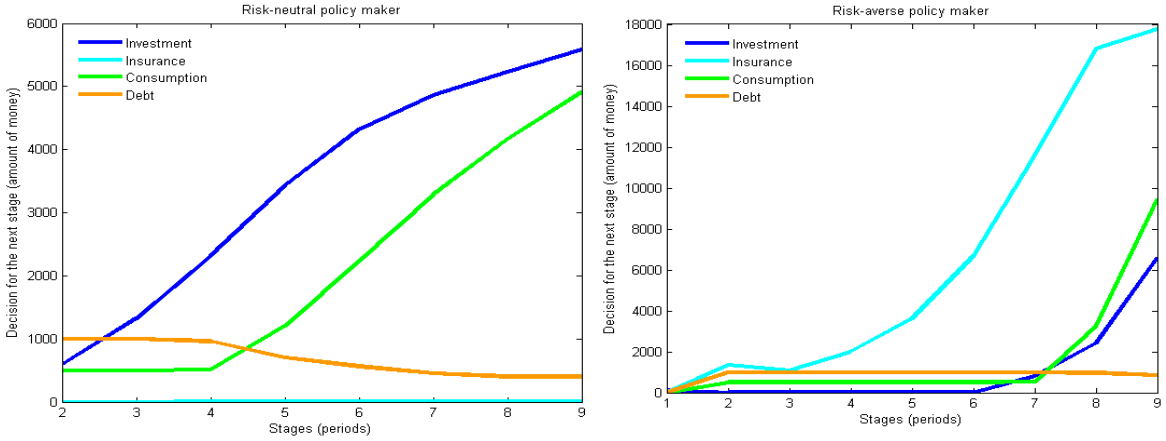


Figure 4.11: Risk-neutral vs. Risk-averse for all periods of interest.

The dependence of the optimal decision on the number of period is shown in the Fig. 4.11 for risk-neutral and risk-averse policy makers. For both of them the optimal debt decreases beginning with some stage while optimal consumption increases.
In Fig. 4.12, 4.13 we can see that the government is able to make the stock and the consumption increase for both high and low losses. In Fig. 4.12 one can see that if the losses are small then the government is able to increase his stock with an acceleration. If the losses are higher then at some level of the losses the growing of the stock starts to decelerate. In Fig. 4.13 one can see that the policy-maker tries to keep the life standard on the lower bound in order to maximize final consumption. The lower losses the greater the final consumption.

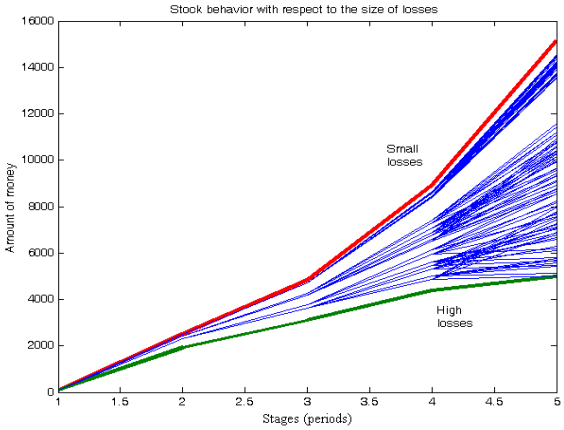


Figure 4.12: Stock behavior.

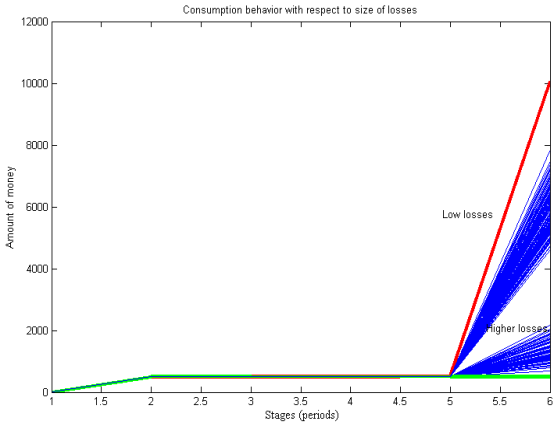


Figure 4.13: Consumption behavior.

4.2.2 Multi-hazard approach

Suppose now that in some country there is the high probability of occurance of a group of natural hazards. In this case we have two possibilities: first is that the sequence of natural hazards has low deviation in size, i.e. the policy-maker is somehow sure about the possible loss range of this group of natural hazard events (the resulting loss is close to the expected one), and second that the deviation in damage is large and, hence, the range is high (the resulting loss is possibly much higher than the expectation). The best example for the first case is the countries that are located in some stable region (for example, Europe). The example for the second case is Japan where tsunami may follow an earthquake and the size of loss may be absolutely unexpected.
Let ξ be the random variable that describes absolute loss after the group of natural hazard events (or multiple natural hazard events) in one region and suppose that large deviations are not possible. So we deal with the countries from the first group. The loss distribution that corresponds nicely to the situation when the variable represents the compound loss from a sequence of many natural hazard events is the log-normal distribution, which probability density function is

$$f_X(x;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}.$$

Now, if we deal with a risk-averse policy-maker that would like to maximize his stock in a given number of periods and to improve the standard of life in the country over all the periods where that is possible then we should consider the same multi-stage stochastic

optimization problem that we considered before:

$$\begin{aligned}
\underset{d_t,c_t,z_t,x_t}{\text{maximize}} \quad & (1-\alpha)\sum_{t=1}^{T}\rho^{-t}\mathbb{E}(c_t)+\alpha\mathbb{E}(S_T)\\
\text{subject to} \quad & x_t \triangleleft \mathcal{F}_t,\ z_t \triangleleft \mathcal{F}_t,\ t=\overline{1,T}\ c_t \geq c,\ , d_t \leq d,\ t=\overline{1,T};\\
& rS_t - \sum_{j=1}^{t-1} d_j i_j \geq c_t + x_t + E_t(1+V)z_t S_t, t=\overline{1,T-1};\\
& rS_T - \sum_{t=1}^{T} d_t \geq c_T;\\
& S_{t+1} = [(1-\delta)S_t + x_t](1-\xi_t) + z_t\xi_t S_t + d_{t+1},\ t=\overline{1,T-2};\\
& S_T = [(1-\delta)S_{T-1} + x_{T-1}](1-\xi_T) + z_T\xi_T S_T.
\end{aligned}$$

The only one difference is that for the case of the unique natural hazard we use the loss distribution of the Pareto type that allows high deviations while for the stable countries that would like to make a decision how to deal with the groups of natural hazards of low damage we use log-normal distribution that does not allow high deviations in size with respect to the expected value.
Comparing the decision about the investment, the insurance, the consumption and the debt for the case of unique natural hazard event with the case of multiple natural hazards with resulting low loss (Fig. 4.14) we derive that for the case of multiple natural hazards the government is able to grow his stock very rapidly dealing with investment at most. The amount of the insurance is close to zero and start to rise slightly only after some periods. For the case of the unique natural hazard the grow of the capital is not so rapid because of the possibility of high deviations in size of the damage after CAT-event.

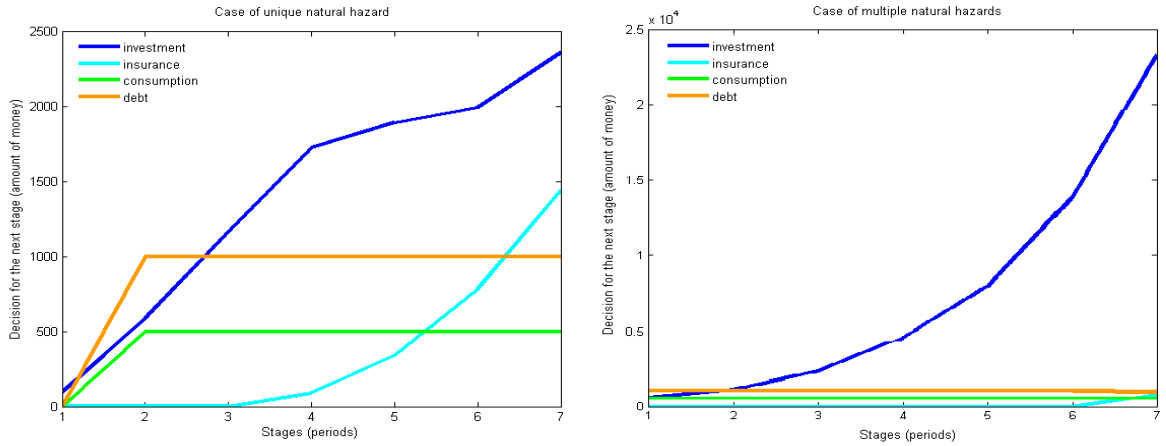


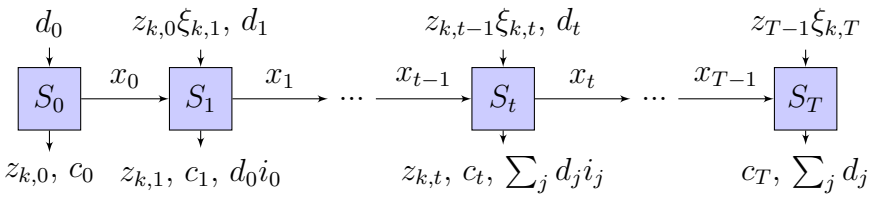
Figure 4.14: Comparison of the unique natural hazard with the case of multiple natural hazards.

The log-normal loss distribution answers how to deal with the situation for hazards with small losses that are close to the expected value but does not give the strategy to survive for countries with a very high damage range provided by natural hazard events. To deal with the situation when one natural hazard with an unexpectedly high loss is followed by another one (the earthquake and the tsunami near Fukushima in Japan, for example) we should study the dependence of different types of natural hazards.
The dependence can be studied by finding the copula function [5, 14] for chosen natural hazard events. The copula C contains all information on the dependence structure between the components of $(\xi^{(1)}, \xi^{(2)}, ..., \xi^{(N)})$ whereas the marginal cumulative distribution functions F_i contain all information on the marginal distributions that are supposed to be Pareto as in the case for the unique natural hazard event:

$$C(u^{(1)}, ..., u^{(N)}) = P(\xi^{(1)} \leq F^{-1}(u^{(1)}), ..., \xi^{(N)} \leq F^{-1}(u^{(N)}))$$

Using the multidimensional random generating and optimal quantization techniques and given the loss distribution it is possible to model losses after the group of different natural hazard events of various size. For the purposes of the multidimensional optimal quantization the Voronoi iteration (Lloyd algorithm) [39] can be implemented and the multi-stage stochastic optimization problem is going to change in the following way in this case[1]:
Components of the model: S_t stock quantity at period t; B_t available budget at period t; r] return on the stock; δ depreciation rate; $\xi_{i,t}$ random relative loss because of the natural hazard i at time t; i_t interest rate of the debt; V insurance load; $E_{i,t}$ loss expectation $E_{i,t} = \mathbb{E}(\xi_{i,t})$.
Decision Variables: d_t debt that might be taken at each period t; c_t consumption at period t; $z_{i,t}$ insured value of the stock for each hazard i separately; x_t amount of investment.

[1]Notice, that quantization techniques used for joint losses between countries (copuled losses) can model multi-period/multi-hazard situations in multi-region environment: for example, flood losses in Europe that are very much dependent in different countries, as soon as river basins in Europe are connected to each other, could be modeled by applying quantization techniques on copuled flood losses.



The multi-stage stochastic optimization program in this case is:

$$
\begin{aligned}
\underset{d_t,c_t,z_t,x_t}{\text{maximize}} \quad & (1-\alpha)\sum_{t=1}^{T}\rho^{-t}\mathbb{E}(c_t) + \alpha\mathbb{E}(S_T) \\
\text{subject to} \quad & x_t \triangleleft \mathcal{F}_t,\ z_t \triangleleft \mathcal{F}_t,\ t=\overline{1,T}\ c_t \geq c,\ , d_t \leq d,\ t=\overline{1,T}; \\
& rS_t - \sum_{j=1}^{t-1} d_j i_j \geq c_t + x_t + \sum_i E_{i,t}(1+V_i) z_{i,t} S_t,\ t=\overline{1,T-1}; \\
& rS_T - \sum_{t=1}^{T} d_t \geq c_T; \\
& S_{t+1} = [(1-\delta)S_t + x_t](1 - \sum_i \xi_{i,t}) + \sum_i z_{i,t}\xi_{i,t} S_t + d_{t+1},\ t=\overline{1,T-2}; \\
& S_T = [(1-\delta)S_{T-1} + x_{T-1}](1 - \sum_i \xi_{i,T}) + \sum_i z_{i,T}\xi_{1,T} S_T,
\end{aligned}
$$

where

1. at first stage policy-maker consumes, insures against possible natural hazards, invests into the next period and has a opportunity to take debt;
2. at all stages except the first and the last one he invests, insures against possible natural hazards, consumes, loses some part of his stock in case of natural hazards events and obtains insurance from the previous period in case of any natural hazard event; he has the opportunity to take debt and should return interests on previously taken debts;
3. at the last stage he consumes, loses some part of his stock in case of natural hazard events and obtains insurance from the previous period in case of any natural hazard event; he should return all taken debts.

As one can see from this problem:

- the losses $\xi_{i,t}$ are different for different hazard events;
- the insurances $z_{i,t}$ are different for different types of natural hazards.

4.2.3 Empirical results

For the convenient development of the risk-management strategies, we created a tool CATOPT, that does multi-stage stochastic optimization for the purposes of hazard risk reduction. Using it, one can find the optimal budget allocation strategy, taking into account historical data on catastrhic events. Fig. 4.15 shows the program window on CATOPT with calculated optimal strategy for Bermuda Islands.

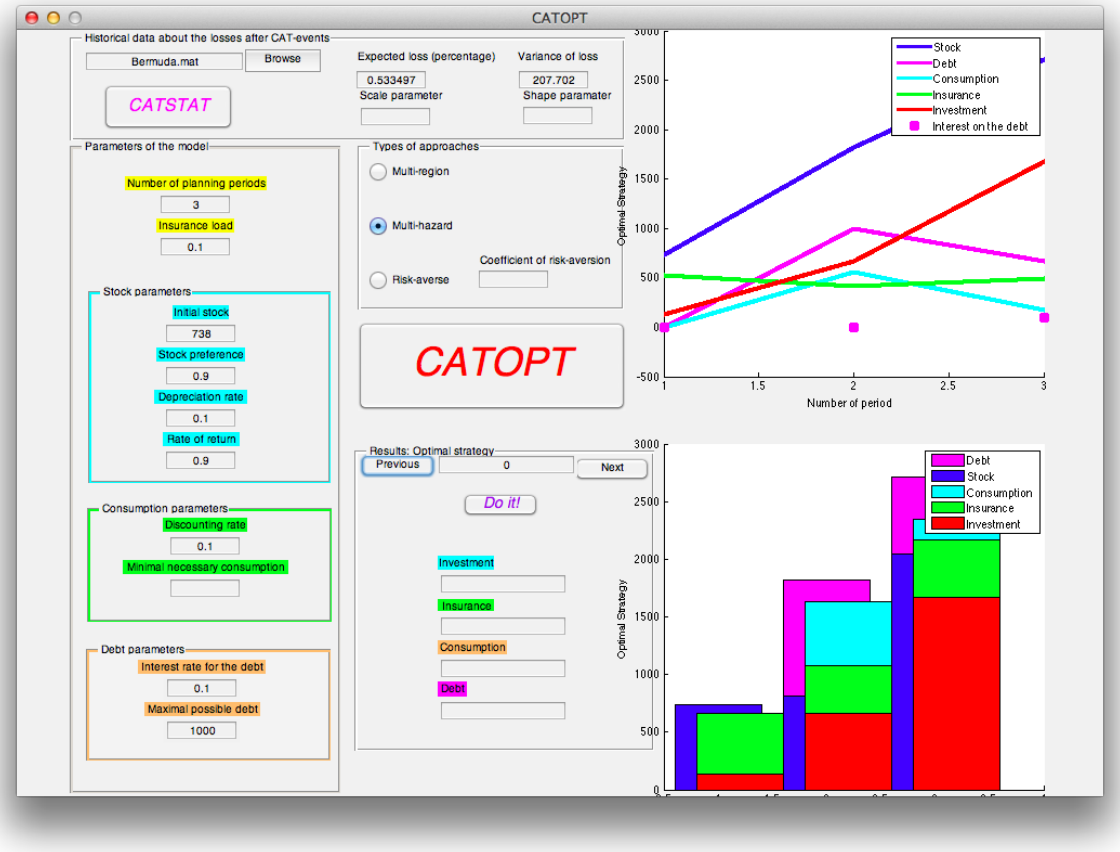


Figure 4.15: CATOPT tool.

Example 4.3 (**CATOPT application**). *Considering Austria as the rich country with low probability of CAT-event and, hence, with low risk, we can find the optimal strategy for the devision of the budget into investment, insurance, consumption and debt. The solution for this country (and for the most of Europe) would be to invest much and to avoid insurance and debts (Fig. 4.16).*

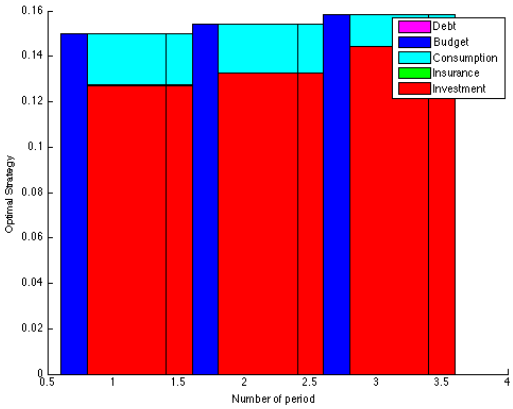


Figure 4.16: Austria optimal strategy for 3 years

A different result would appear for such countries as Madagascar or Bermuda as long as they can be considered as poor countries with high risk of CAT-event. Madagascar and Bermuda should insure against the catastrophic events as well as consider the possibility of taking debts (Fig. 4.17).

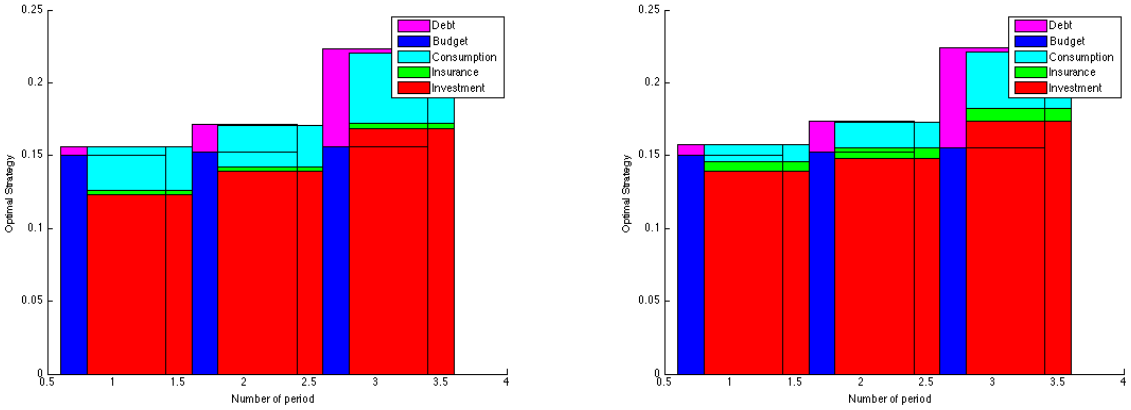


Figure 4.17: Bermuda and Madagascar optimal strategy for 3 years

Chapter 5

Appendix

5.1 OPTGEN

OPTGEN is the distributional quantization tool, that does quantization with the described in chapter 3 algorithms:

Monte-Carlo generation with a use of existing MATLAB tools for Monte Carlo generation;

Quasi Monte-Carlo generation the basic idea of Quasi Monte-Carlo generation is to replace random samples in Monte-Carlo methods by deterministic points that are uniformly distributed in $[0,1]^D$, where D is the number of dimensions;

Optimal quantization that finds the optimal supporting points z_i, $i = 1, ..., N$ by the minimization of $\int \min_i d(x, z_i)P(dx)$ over $z_1, ..., z_N$, and, then, given the supporting points z_i, it finds the probabilities p_i, which minimize $d_{KA}(P, \sum_{i=1}^{N} p_i \delta_{z_i})$ for the Kantorovich distance.

QOND Quantization a new method for quantization based on the minimization of the nested distance between continuous-state stochastic process and a tree. This method is very suitable for multi-stage stochastic optimization programs, as it takes into account all available tree information.

OPTGEN does the distributional quantization with the described algorithms. On the Fig. 5.1 one can see the program window of OPTGEN.

1. Create a file data.mat with the treestructure and parameters of the distribution at each stage of the tree and load it to the OPTGEN system.

2. Choose a type of the distribution (normal, log-normal, exponentional or Pareto) and a method of quantization. Notice that the parameters of Pareto and exponential distribution are not expectation and covariance matrix but scale and shape parameters.

3. After one presses the button OPTGEN the quantization is completed and one can see the results of the Monte-Carlo and Optimal quantization for log-normal distribution on the pictures (Fig. 5.1).

4. As soon as the optimization for quantization is termited the program creates a file with all the quantizers and probabilities of them as well as it reminds the treestructure.

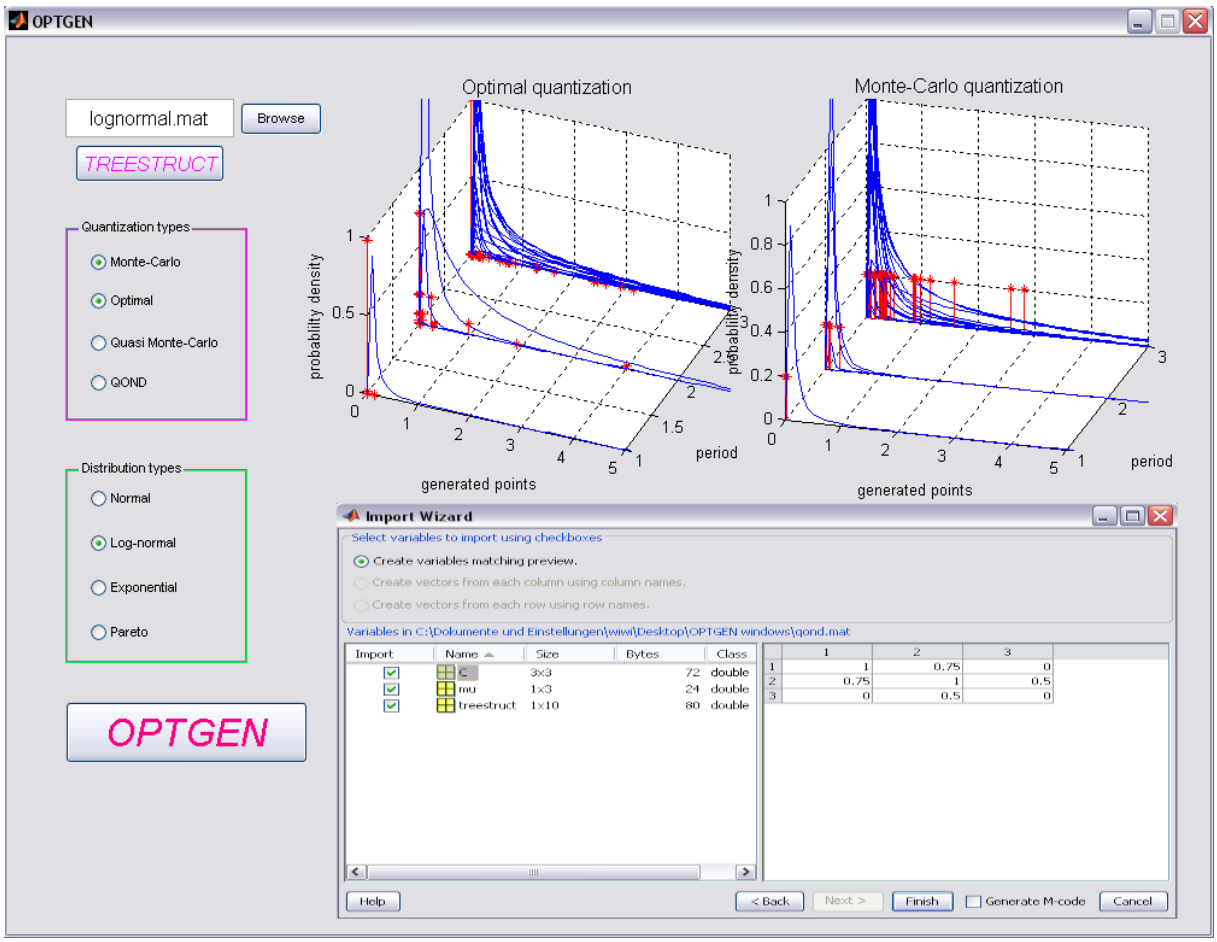


Figure 5.1: OPTGEN tool

5.2 Description of figures

Figure	Description and parameters used
Fig. 1.1	Height of both trees is $T = 3$. Bushiness of the left-hand side tree is $bush = [5\ 5\ 5]$ while bushiness of the rigth-hand side tree is $bush = [2\ 2\ 2]$.
Fig. 1.2	Tree with predecessor vector $pred = [0\ 1\ 1\ 1\ 2\ 2\ 2\ 3\ 3\ 4\ 4]$, values $val = [0\ 1\ 0\ -1\ 1.5\ 1.3\ 1\ 0.5\ -0.5\ -1\ -1.5]$ sitting on the nodes and scenario probabilities $pScen = [0.04\ 0.03\ 0.03\ 0.48\ 0.12\ 0.15\ 0.15]$. Height of the tree is $T = 2$.
Fig. 1.3	Notations for a tree structure.
Fig. 2.1	Illustration for the Monge Transportation Problem.
Fig. 2.2	$N = 5$ optimal supporting points and corresponding probabilities for the Gaussian distribution $\mathcal{N}(0, 1)$.
Fig. 2.3	Notations for the probability measures on a treestructure.
Fig. 2.4	Illustration of the subtree construction for a tree of the height T.
Fig. 2.5	Illustration for the clairvoyant tree (at stage $t = 1$) that is obtained from some initial tree with given structure with $N_{\mathbb{T}} = 27$ and $T = 3$. Predecessor vector of the initial tree is $pred = [0\ 1\ 1\ 1\ 2\ 2\ 3\ 3\ 3\ 4\ 4\ 5\ 5\ 6\ 6\ 7\ 7\ 8\ 8\ 8\ 9\ 9\ 10\ 10\ 10\ 11\ 11]$.
Fig. 2.6, 2.7	Counterexample for the specific behavior of nested distance between initial tree and its clairvoyant representation. Initial tree has the following structure: $pred = [0\ 1\ 1\ 1\ 2\ 2\ 3\ 3\ 3\ 4\ 4\ 5\ 5\ 6\ 6\ 7\ 7\ 8\ 8\ 8\ 9\ 9\ 10\ 10\ 10\ 11\ 11]$ $pNodes = [1\ 0.2\ 0.3\ 0.5\ 0.5\ 0.5\ 0.6\ 0.2\ 0.2\ 0.7\ 0.3\ 0.5\ 0.5\ 0.1\ 0.9$ $0.2\ 0.8\ 0.1\ 0.2\ 0.7\ 0.5\ 0.5\ 0.2\ 0.3\ 0.5\ 0.9\ 0.1]$ $val = [0\ 40.68\ 93.83\ 25.54\ 53.32\ 95.48\ 26.77\ 25.01\ 92.77\ 6.86\ 29.94$ $59.16\ 20.33\ 63.59\ 79.84\ 50.17\ 65.08\ 79.6\ 23.34\ 60.08\ 11.25$ $51.58\ 83.78\ 92.08\ 49.82\ 27.76\ 65.25]$
Fig. 2.8	Illustration for the decrease of nested distance between subtrees due to the clairvoyant structure.
Fig. 2.9	Illustration for the approximation of the nested distance between continuous stochastic process and finite tree.
Fig. 2.10	Inventory Control Problem with parameters $l = 0.9 * ones(1, T)$, $u = 1.1 * ones(1, T)$ for the Gaussian distribution $\mathcal{N}(ones(1, T), eye(T, T))$. The trees considered are of heights $T = 1$, $T = 2$ and $T = 3$ and bushiness changing from $bush = 2 * ones(1, T)$ to $bush = 9 * ones(1, T)$. The comparison is performed for random and optimal scenario generation techniques.
Fig. 2.11	Illustration for the approximation of the nested distance between two continuous stochastic processes.
Fig. 3.1	Stage-wise optimal quantization v.s. stage-wise random generation of multivariate Gaussian distribution for the tree with bushiness $bush = [5\ 5\ 5]$.
Fig. 3.2	Illustration for the stage-wise tree approximation of a stochastic process with bivariate Gaussian distribution sitting on a tree structure.

Fig. 3.4	Illustration for the direct approximation of the bivariate Gaussian distribution, i.e. Voronoi diagram.
Fig. 3.5	The tree is given by the bushiness $bush = [3\ 3]$ and at each stage of the tree a continuous bivariate Gaussian distribution is sitting. The figure illutrates an approximation of each of the bivariate distributions by the Voronoi diagram with 3 cells (i.e. the number of outgoing branches from a node).
Fig. 3.6, 3.8	Gaussian distribution $\mathcal{N}(\mu, C)$, where $C_{1,1}$=[0.4001 0.2320; 0.2320 0.3101], $C_{2,2}$=[0.3100 0.3500; 0.3500 1.2200], $C_{3,3}$=[0.5000 0.4400; 0.4400 1.1800], $C_{1,2}$=[0.3010 0.1630; 0.3240 0.4210], $C_{1,3}$=[0.2420 0.1650; 0.2030 0.4820], $C_{2,3}$=[0.2500 0.1700; 0.1300 0.3400]. $C = [C_{1,1},\ C_{1,2},\ C_{1,3};\ C_{1,2},\ C_{2,2},\ C_{2,3};\ C_{1,3},\ C_{2,3},\ C_{3,3}]$. μ^T=[0 0; 1 1; 2 2]. Bushiness vector of the approximate tree is $bush$=[3 4 3].
Fig. 3.9	Notations for the probability measures on a clairvoyant treestructure.
Fig. 3.10	Illustration, that describes the scenario adaptation algorithm.
Fig. 3.11	Joint Adaptation algorithm for the Gaussian distribution $\mathcal{N}(\mu, C)$, where μ=[0.7 0.5] and C=[0.6630 1.0311; 1.0311 1.7036]. Approximation tree structure is given by the predecessor vector $pred$=[0 1 1 1 1 1 2 2 2 2 3 3 3 3 4 4 4 4 5 5 5 5 6 6 6 6].
Fig. 3.13	Inventory Control Problem with parameters $l = 0.9 * ones(1, T)$, $u = 1.1 * ones(1, T)$ for the Gaussian distribution $\mathcal{N}(ones(1, T), eye(T, T))$. The tree height is $T = 3$ and bushiness is changing from $bush = 3 * ones(1, T)$ to $bush = 7 * ones(1, T)$. The comparison is performed for stage-wise optimal scenario generation and QOND quantization techniques.
Fig. 3.14	Gaussian distribution $\mathcal{N}(\mu, C)$ with parameters μ=[0.06856 0.2113 0.7773], C=[0.4054 0.0808 0.2945; 0.0808 0.2715 0.0869; 0.2945 0.0869 0.2200]; Treestructure is defined by $pred$=[0 1 1 1 2 2 3 3 3 4 4 5 5 6 6 7 7 8 8 8 9 9 10 10 10 11 11]; $vals$ and $pScen$ are chosen randomly
Fig. 3.15	Gaussian distribution $\mathcal{N}(\mu, C)$ with parameters μ=[0.06856 0.2113 0.7773], C=[0.4054 0.0808 0.2945; 0.0808 0.2715 0.0869; 0.2945 0.0869 0.2200]; Treestructure is defined by $pred$=[0 1 1 1 2 2 3 3 3 4 4 5 5 6 6 7 7 8 8 8 9 9 10 10 10 11 11]; $vals$ and $pScen$ are obtained by the stage-wise optimal algorithm.
Fig. 4.1	Illustration of the inverse CDF for the Gaussian distribution.
Fig. 4.2, 4.3	Inventory Control Problem with parameters l=[0.93 0.93 0.92], u=[1.1 1.1 1.1] for the Gaussian distribution $\mathcal{N}(\mu, C)$, where C=[1 0.75 0; 0.75 1 0.5; 0 0.5 1] and μ=[1 0.5 0.75]. The tree height is $T = 3$ and it is defined by $pred$=[0 1 1 1 2 2 3 4 4 4].
Fig. 4.4, 4.5, 4.6	Inventory Control Problem with parameters l=[0.93 0.93 0.92], u=[1.1 1.1 1.1] for the log-normal distribution $ln\mathcal{N}(\mu, C)$, where C=[1 0.75 0; 0.75 1 0.5; 0 0.5 1] and μ=[1 0.5 0.75]. The tree height is $T = 3$ and it is defined by $pred$=[0 1 1 1 2 2 3 4 4 4].

Chapter 6

Bibliography

Bibliography

[1] S. Boyd, L. Vandenberghe. *Convex Optimization.* Cambridge University Press, 2004.

[2] R. Durrett. *Probability: Theory and Examples.* Duxburry Press, Belmont, CA, 2nd edition, 2004. ISBN 0-534-42441-4.

[3] P. Embrechts, S. Resnick, G. Samorodnitsky. *Extreme Value Theory as a Risk Management Tool.* North American Actuarial Journal, Volume 3 (2), pp. 30-41, 1999.

[4] Y. Ermoliev, T. Ermolieva, G. Fischer, M. Makowski. *Extreme Events, Discounting and Stochastic Optimization.* Annals of Operations Research, Volume 177 (1), pp. 9-19, 2010.

[5] Y. Ermoliev, M. Makowski, K. Marti, G. Ch. Pflug eds. *Coping with Uncertainty. Lecture Notes in Economics and Mathematical Systems.* Springer Verlag, 2006. ISBN 3-540-35258-9.

[6] Y. Ermoliev, K. Marti, G. Ch. Pflug eds. *Dynamic Stochastic Optimization. Lecture Notes in Economics and Mathematical Systems.* Springer Verlag, 2004. ISBN 3-540-40506-2.

[7] G. S. Fishman. *Monte Carlo: Concepts, Algorithms, and Applications.* Springer, New York, 1995.

[8] J.-C. Fort, G. Pagés. *Asymptotics of Optimal Quantizers for Some Scalar Distributions.* Journal of Computational and Applied Mathematics, Volume 146 (2), Amsterdam, The Netherlands, Elsevier Science Publishers B. V., pp. 253-275, 2002.

[9] A. Genz. *Numerical Computation of Multivariate Normal Probabilities.* Journal of Computational and Graphical Statistics, Volume 1, pp. 141-149, 1992.

[10] S. Graf and H. Luschgy. *Foundations of Quantization for Probability Distributions.* Lecture Notes in Mathematics, Volume 1730, Springer, Berlin, 2000.

[11] H. Heitsch and W. Römisch. *Scenario Tree Modeling for Multi-stage Stochastic Programs.* Mathematical Programming, Volume 118, pp. 371–406, 2009.

[12] S. Hochrainer, G. Ch. Pflug. *Optimizing Financial Strategies for Governments Against the Economic Impacts of Natural Disasters in Developing Countries.* Paper presented at the IIASA-DPRI Conference, Beijing, China, 14-18 September 2005.

[13] S. Hochrainer-Stigler, G. Pflug, N. Lugeri. *Flood Risk in a Changing Climate: A Multilevel Approach for Risk Management.* Beitrag in Sammelband Integrated Catastrophe Risk Modeling - Supporting Policy Processes, Chapter 16, pp. 263-280, Springer Science+Business Media Dordrecht, 2013. ISBN: 978-94-007-2226-2.

[14] S. Hochrainer, A. Timonina, K. Williges, G. Pflug, R. Mechler. *Modelling the economic and fiscal risks from natural disasters. Insights based on the CatSim model.* Background Paper prepared for the Global Assessment Report on Disaster Risk Reduction, 2013.

[15] L. Kantorovich. *On the Translocation of Masses.* C.R. (Doklady) Acad. Sci. URSS (N.S.)., Volume 37, pp. 199201, 1942.

[16] R. S. Lipster and A. N. Shiryayev. *Statistics of Random Processes.* Springer-Verlag, Volume 2, New York, 1978. ISBN 0-387-90236-8.

[17] H. Luschgy, G. Pagés, and B. Wilbertz. *Asymptotically Optimal Quantization Schemes for Gaussian Processes.* ESAIM: Probability and Statistics, Volume 14, pp. 93-116, 2010.

[18] R. Mirkov. *Tree Approximations of Dynamic Stochastic Programs: Theory and Applications.* VDM Verlag, pp. 1-176, 2008. ISBN 978-363-906-131-4.

[19] R. Mirkov, G. Ch. Pflug. *Tree Approximations of Stochastic Dynamic Programs.* SIAM Journal on Optimization, Volume 18 (3), pp. 1082-1105, 2007.

[20] G. Monge. *Mémoire sue la théorie des déblais et de remblais.* Histoire de lAcadémie Royale des Sciences de Paris, avec les Mémoires de Mathématique et de Physique pour la même année, pp. 666-704, 1781.

[21] R. B. Nelsen. *An Introduction to Copulas.* Number 139 in Lecture Notes in Statistics, Springer, 1998.

[22] G. Ch. Pflug. *Optimization of Stochastic Models.* The Kluwer International Series in Engineering and Computer Science, Volume 373, Kluwer Academic Publishers, Boston, MA, 1996. ISBN 0-7923-9780-0.

[23] G. Ch. Pflug. *Some Remarks on the Value-at-Risk and the Conditional Value-at- Risk.* Probabilistic Constrained Optimization: Methodology and Applications, Kluwer Academic Publishers, pp. 272-281, Dordrecht, 2000.

[24] G. Ch. Pflug. *Version-independence and Nested distributions in Multi-stage Stochastic Optimization.* SIAM Journal on Optimization, Volume 20 (3), pp. 1406-1420, 2010.

[25] G. Ch. Pflug. *Scenario Tree Generation for Multiperiod Financial Optimization by Optimal Discretization.* Mathematical Programming, Series B, Volume 89 (2), pp. 251-257, 2001.

[26] G. Ch. Pflug, W. Römisch. *Modeling, Measuring and Managing Risk.* World Scientific Publishing, pp. 1-301, 2007. ISBN 978-981-270-740-6.

[27] G. Ch. Pflug, A. Pichler. *A Distance for Multi-stage Stochastic Optimization Models.* SIAM Journal on Optimization, Volume 22, pp. 1-23, 2012.

[28] G. Ch. Pflug, A. Pichler. *Approximations for Probability Distributions and Stochastic Optimization Problems.* Springer Handbook on Stochastic Optimization Methods in Finance and Energy (G. Consigli, M. Dempster, M. Bertocchi eds.), Int. Series in OR and Management Science, Volume 163, Chapter 15, pp. 343-387, 2011.

[29] S. T. Rachev. *Probability Metrics and the Stability of Stochastic Models.* Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley & Sons Ltd., Chichester, 1991. ISBN 0-471-92877-1.

[30] S. T. Rachev, W. Römisch. *Quantitative Stability in Stochastic Programming: the Method of Probability Metrics.* Math. Oper. Research, Volume 27 (4), pp. 792-818, 2002. ISSN 0364-765X.

[31] S. T. Rachev, W. Rüschendorf. *Mass Transportation Problems.* Probability and its Applications, Springer-Verlag, Volume 1 (Theory), New York, 1998. ISBN 0-387-98350-3.

[32] W. Römisch. *Scenario Generation.* Wiley Encyclopedia of Operations Research and Management Science (J.J. Cochran ed.), Wiley, 2010.

[33] R. Y. Rubinstein, D. R. Kroese. *Simulation and the Monte-Carlo Method.* Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., 2nd. ed., Hoboken, New Jersey, 2008. ISBN 978-0-470-17794-5.

[34] T. Szántai. *Improved Bounds and Simulation Procedures on the Value of the Multivariate Normal Probability Distribution Function.* Annals of Operations Research, Volume 100, pp. 85-101, 2000.

[35] 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.

[36] A.V. Timonina. *Optimal Strategies for Risk-Management of Catastrophic Events.* International Institute for Applied Systems Analysis, Research paper at Young Scientist Summer Program, 2011.

[37] C. Villani. *Topics in Optimal Transportation.* Graduate Studies in Mathematics, American Mathematical Society, Volume 58, Providence, RI, 2003. ISBN 0-8218-3312-X.

[38] C. Villani. *Optimal Transport, Old and New.* Springer, 2008.

[39] G. Voronoi. *Nouvelles Applications des Paramétres Continus á la théorie des Formes Quadratiques.* Journal für die Reine und Angewandte Mathematik, Volume 133, pp. 97-178, 1907.

[40] P. L. Zador. *Asymptotic quantization error of continuous signals and the quantization dimension.* IEEE Trans. Inform. Theory, Volume 28, pp. 139-149, 1982.

Chapter 7

Curriculum Vitae

Anna Timonina, M.Sc.

Work experience

October 2012–present — **Research assistant**, *International Institute for Applied System Analysis, Program on Risk, Policy and Vulnerability*, Vienna, Austria.

Sep. 2010–Dec. 2013 — **Project assistant**, *University of Vienna, Department of Statistics and Operations Research*, Vienna, Austria.

Summer 2011, 2012 — **Young Scientists Summer Program**, *International Institute for Applied System Analysis, Program on Risk, Policy and Vulnerability*, Vienna, Austria.
Research under the title "Optimal strategies for risk management of catastrophic events". Anna V. Timonina is the recipient of the Mikhalevich Award 2011 for the outstanding work in IIASA's Young Scientists Summer Program.

Education

2011–present — **Abraham Wald PhD Program in Statistics and Operations Research**, *University of Vienna*, Vienna, Austria.
Department of Statistics and Operations Research. Dissertation is devoted to the multi-stage stochastic approximation and optimization.

2009–2011 — **M.Sc. degree in Applied Mathematics and Physics**, *Moscow Institute of Physics and Technology (State University)*, Moscow, Russia.
Department of Radio Engineering and Cybernetics. Participation in the project "PageRank: new regularization and simulation models".

2005–2009 — **B.Sc. degree in Applied Mathematics and Physics**, *Moscow Institute of Physics and Technology (State University)*, Moscow, Russia.
Department of Radio Engineering and Cybernetics. Dissertation topic is "Iterative algorithms for ranking with application to the Scientometrics and to the Internet".

Extra education

2009–2010 — **Postgraduate program**, *Institute for Control Sciences of Russian Academy of Science*, Moscow, Russia.
Specialization is System analysis. Information Control and Processing.

2009–2010 — **International educational program in Financial Economics**, *International College of Economics and Finance, State University - Higher School of Economics in cooperation with London School of Economics*, Moscow, Russia.

2009–2009 — **Diploma of Higher Extra Education**, *Moscow Institute of Physics and Technology (State University)*, Moscow, Russia.
Specialization: English translator in the field of professional communication.

Vienna, Oskar-Morgenstern-Platz 1 – 1090
+436765623331 • ☎ *+43-1-4277-386 33* • FAX *+43-1-4277-838633*
• ✉ *anna.timonina@univie.ac.at*
• *http://homepage.univie.ac.at/anna.timonina/*

Awards

May 2013 **Best Student Paper Prize**, 10th International Conference on Computational Management Science.
The award is given for the research "Multi-Stage Stochastic Optimization: the Distance Between Stochastic Scenario Processes".

2011-2012 **Mikhalevich Award**, International Institute for Applied System Analysis.
The award is given for the research "Optimal Strategies for Risk Management of Catastrophic Events".

November 2009 **Winner Diploma**, 52nd Scientific Conference of Moscow Institute of Physics and Technology.
The Winner Diploma is given for the research of new regularizations and simulation methods for PageRank.

2008–2009 **Scholarship named after Abramov-Frolov**, Moscow Institute of Physics and Technology (State University).

2006–2008 **Scholarship named after A. Menshikov**, Moscow Institute of Physics and Technology (State University).

May 2005 **Third Diploma**, Moscow State University.
The diploma is awarded both at the Lomonosov Olympiad in mathematics and at the Lomonosov Olympiad in Russian and Literature

Publications

[1] A. V. Timonina "Multi-Stage Stochastic Optimization: the Distance Between Stochastic Scenario Processes", *Computational Management Science, Springer-Verlag Berlin Heidelberg, August 2013, DOI 10.1007/s10287-013-0185-3*.

[2] S. Hochrainer, A. Timonina, K. Williges, G. Pflug, R. Mechler "Modelling the economic and fiscal risks from natural disasters. Insights based on the CatSim model", *Global Assessment Report on Disaster Risk Reduction, 2013*.

[3] A. V. Timonina "Optimal Strategies for Risk-Management of CAT-events", *IIASA, YSSP 2011*.

[4] B.T. Polyak, A. V. Timonina "PageRank: New Regularizations and Simulation Models", *18th IFAC World Congress*, Volume No. 18, part No. 1, Milano, Italy, August, 2011, 11202–11207.

[5] A. V. Timonina "The Rank-Model and Its Investigations", *Stochastic Optimization in Informatics, No. 5*, (O.N. Granichin ed., ISSN 1992–2922), St.-Petersburg Univ., 2009, 139–156 (in Russian).

[6] A. V. Timonina "Iterative Algorithms for Ranking with Application to Scientometrics and to the Internet", *Collection of Studies of the 6th All-Russian Conference for Young Scientists"*, St.-Petersburg, Russia, 2009 (in Russian).

[7] A. V. Timonina "The Rank-Problem Modeling and Its Analysis", *Collection of Studies of the 52d Scientific Conference MIPT*, Moscow, Russia, 2009 (in Russian).

[8] A. V. Timonina "The Rank-Problem Modeling and Its Analysis", *Collection of Studies of the First Traditional School "Control, Information and Optimization"*, Pereslavl-Zalessky, Russia, 2009 (in Russian).

[9] A. V. Timonina "The Rank-Problem Modeling and Its Analysis", *Collection of Studies of the Second Traditional School "Control, Information and Optimization"*, Pereslavl-Zalessky, Russia, 2010 (in Russian).

Vienna, Oskar-Morgenstern-Platz 1 – 1090
+436765623331 • ☎ *+43-1-4277-386 33* • FAX *+43-1-4277-838633*
• ✉ *anna.timonina@univie.ac.at*
• *http://homepage.univie.ac.at/anna.timonina/*

Research projects

Approximation and convergence in multi-stage stochastic optimization. The project covers a topic between stochastics (statistics) and optimization. Decisions under uncertainty play an important role in finance, energy and commodity markets, production, logistics and the supply chain. This project focuses on multi-stage decisions. The goal is to study approximations of very large problems by simpler ones including estimates for the approximation error and ways of finding optimal approximations.

PageRank: new regularization and simulation models. PageRank problem is one of the most challenging problems in information technologies and numerical analysis due to its huge dimension and wide range of applications. It also attracts great attention of experts in control theory; it is closely related to consensus in multiagent systems. The original problem is replaced with finding eigenvector of modified matrix which can be effectively solved by power method. The aim is to demonstrate that the solution of the modified problem can be far enough from the original one, to propose an iterative regularization method which allows to find the desired solution, to induce an l^1-regularization which provides a solution of the page ranking problem which ignores low-ranking pages.

Insurance for Adaptation. Insurance for Adaptation project studies the efficiency and equity of available and alternative strategies and options for transferring disaster risk in Austria and Europe, as well as examining their link with risk reduction and adaptation. To date, there is little empirical evidence on the effectiveness of risk-transfer instruments for incentivizing the reduction of losses from extreme weather events and other hazards. This proposed research builds an evidence base on this link by examining experience at the (1) entrepreneurial scale, (2) national scale, and (3) European Union scale assessing EU wide risk sharing, and the performance of the European Union Solidarity Fund (EUSF)).

Optimal Strategies for Risk Management of Catastrophic Events. An essential part of stresses and risks on societies and their environments is imposed by worldwide catastrophic events. That is why the research, devoted to finding the optimal strategies for risk management of catastrophic events, is motivated by different needs of people on international, national and local policy levels. The goal of this project is to study the problem of catastrophe risk management from the multi-period, multi-hazard and multi-region points of view.

Languages

Russian	**Native speaker**	
English	**Fluent**	*Translator in the Field of Professional Communication; IELTS.*
German	**Advanced**	*Certificate of the Language Center on the University of Vienna*

Summary of qualifications

Educational and professional background Statistics, operations research, stochastic optimization, robust optimization, system analysis, information control and processing, data mining;

Software MATLAB, Stata, EViews; LaTex, Office.

Interests

1993–present	Music; Playing the violin (playing in the University Orchestra since 2010)
2000–present	Tennis

Vienna, Oskar-Morgenstern-Platz 1 – 1090
+436765623331 • ☎ *+43-1-4277-386 33* • FAX *+43-1-4277-838633*
• ✉ *anna.timonina@univie.ac.at*
• *http://homepage.univie.ac.at/anna.timonina/*