Stochastic Optimization and Discretization

A Change in the Point of View

During the first part of the course, we have studied open-loop stochastic optimization problems, that is, problems in which the decisions correspond to deterministic variables which minimize a cost function defined as an expectation.

$$\min_{u \in \mathbb{U}^{\mathrm{ad}}} \mathbb{E}\big(j(u, \boldsymbol{W})\big) \ .$$

We now enter the realm of closed-loop stochastic optimization, that is, the case where on-line information is available to the decision maker. The decisions are thus functions of information and correspond to random variables.

$$\min_{\boldsymbol{U} \in \mathcal{U}^{\mathrm{ad}}} \mathbb{E}\big(j(\boldsymbol{U}, \boldsymbol{W})\big) \ .$$

A Change in the Point of View

During the first part of the course, we have studied open-loop stochastic optimization problems, that is, problems in which the decisions correspond to deterministic variables which minimize a cost function defined as an expectation.

$$\min_{u \in \mathbb{U}^{\mathrm{ad}}} \mathbb{E}\big(j(u, \boldsymbol{W})\big) .$$

We now enter the realm of closed-loop stochastic optimization, that is, the case where on-line information is available to the decision maker. The decisions are thus functions of information and correspond to random variables.

$$\min_{\boldsymbol{U} \in \mathcal{U}^{\mathrm{ad}}} \mathbb{E}\big(j(\boldsymbol{U}, \boldsymbol{W})\big) .$$

Variables and Constraints

The decision variable $\boldsymbol{U}$ is now a random variable and belongs to a functional space $\mathcal{U}$. A canonical example is: $\mathcal{U} = L^2(\Omega, \mathcal{A}, \mathbb{P}; \mathbb{U})$.

The contraints $\boldsymbol{U} \in \mathcal{U}^{\mathrm{ad}}$ on the r.v. $\boldsymbol{U}$ may be of different kinds:

We will mainly concentrate on measurability constraints.

Variables and Constraints

The decision variable $\boldsymbol{U}$ is now a random variable and belongs to a functional space $\mathcal{U}$. A canonical example is: $\mathcal{U} = L^2(\Omega, \mathcal{A}, \mathbb{P}; \mathbb{U})$.

The contraints $\boldsymbol{U} \in \mathcal{U}^{\mathrm{ad}}$ on the r.v. $\boldsymbol{U}$ may be of different kinds:

- point-wise constraints dealing with the possible values of $\boldsymbol{U}$:

$$\boldsymbol{U} \in \mathcal{U}^{\mathrm{ad}} = \left\{ \boldsymbol{U} \in \mathcal{U},\ \boldsymbol{U}(\omega) \in U^{\mathrm{ad}}\ \mathbb{P}\text{-a.s.} \right\} ,$$

- risk constraints, such as expectation or probability constraints:

$$\boldsymbol{U} \in \mathcal{U}^{\mathrm{ad}} = \left\{ \boldsymbol{U} \in \mathcal{U},\ \mathbb{P}\big(\Theta(\boldsymbol{U}) \leq \theta\big) \geq \pi \right\} ,$$

- measurability constraints which express the fact that a given amount of information $\boldsymbol{Y}$ is available to the decision maker:

$$\boldsymbol{U} \in \mathcal{U}^{\mathrm{ad}} = \left\{ \boldsymbol{U} \in \mathcal{U},\ \boldsymbol{U} \text{ measurable w.r.t. } \boldsymbol{Y} \right\} .$$

We will mainly concentrate on measurability constraints.

Variables and Constraints

The decision variable $\boldsymbol{U}$ is now a random variable and belongs to a functional space $\mathcal{U}$. A canonical example is: $\mathcal{U} = L^2(\Omega, \mathcal{A}, \mathbb{P}; \mathbb{U})$.

The contraints $\boldsymbol{U} \in \mathcal{U}^{\mathrm{ad}}$ on the r.v. $\boldsymbol{U}$ may be of different kinds:

- point-wise constraints dealing with the possible values of $\boldsymbol{U}$:

$$\boldsymbol{U} \in \mathcal{U}^{\mathrm{ad}} = \big\{ \boldsymbol{U} \in \mathcal{U},\ \boldsymbol{U}(\omega) \in U^{\mathrm{ad}}\ \mathbb{P}\text{-a.s.} \big\} \ ,$$

- risk constraints, such as expectation or probability constraints:

$$\boldsymbol{U} \in \mathcal{U}^{\mathrm{ad}} = \Big\{ \boldsymbol{U} \in \mathcal{U},\ \mathbb{P}\big(\Theta(\boldsymbol{U}) \leq \theta\big) \geq \pi \Big\} \ ,$$

- measurability constraints which express the fact that a given amount of information $\boldsymbol{Y}$ is available to the decision maker:

$$\boldsymbol{U} \in \mathcal{U}^{\mathrm{ad}} = \big\{ \boldsymbol{U} \in \mathcal{U},\ \boldsymbol{U} \text{ measurable w.r.t. } \boldsymbol{Y} \big\} \ .$$

We will mainly concentrate on measurability constraints.

Variables and Constraints

The decision variable $\boldsymbol{U}$ is now a random variable and belongs to a functional space $\mathcal{U}$. A canonical example is: $\mathcal{U} = L^2(\Omega, \mathcal{A}, \mathbb{P}; \mathbb{U})$.

The contraints $\boldsymbol{U} \in \mathcal{U}^{\mathrm{ad}}$ on the r.v. $\boldsymbol{U}$ may be of different kinds:

- point-wise constraints dealing with the possible values of $\boldsymbol{U}$:

$$\boldsymbol{U} \in \mathcal{U}^{\mathrm{ad}} = \big\{ \boldsymbol{U} \in \mathcal{U},\ \boldsymbol{U}(\omega) \in U^{\mathrm{ad}}\ \mathbb{P}\text{-a.s.} \big\} ,$$

- risk constraints, such as expectation or probability constraints:

$$\boldsymbol{U} \in \mathcal{U}^{\mathrm{ad}} = \big\{ \boldsymbol{U} \in \mathcal{U},\ \mathbb{P}\big(\Theta(\boldsymbol{U}) \leq \theta\big) \geq \pi \big\} ,$$

- measurability constraints which express the fact that a given amount of information $\boldsymbol{Y}$ is available to the decision maker:

$$\boldsymbol{U} \in \mathcal{U}^{\mathrm{ad}} = \big\{ \boldsymbol{U} \in \mathcal{U},\ \boldsymbol{U} \text{ measurable w.r.t. } \boldsymbol{Y} \big\} .$$

We will mainly concentrate on measurability constraints.

Variables and Constraints

The decision variable $\boldsymbol{U}$ is now a random variable and belongs to a functional space $\mathcal{U}$. A canonical example is: $\mathcal{U} = L^2(\Omega, \mathcal{A}, \mathbb{P}; \mathbb{U})$.

The contraints $\boldsymbol{U} \in \mathcal{U}^{\mathrm{ad}}$ on the r.v. $\boldsymbol{U}$ may be of different kinds:

- point-wise constraints dealing with the possible values of $\boldsymbol{U}$:

$$\boldsymbol{U} \in \mathcal{U}^{\mathrm{ad}} = \left\{ \boldsymbol{U} \in \mathcal{U},\ \boldsymbol{U}(\omega) \in U^{\mathrm{ad}}\ \mathbb{P}\text{-a.s.} \right\} ,$$

- risk constraints, such as expectation or probability constraints:

$$\boldsymbol{U} \in \mathcal{U}^{\mathrm{ad}} = \left\{ \boldsymbol{U} \in \mathcal{U},\ \mathbb{P}\big(\Theta(\boldsymbol{U}) \leq \theta\big) \geq \pi \right\} ,$$

- measurability constraints which express the fact that a given amount of information $\boldsymbol{Y}$ is available to the decision maker:

$$\boldsymbol{U} \in \mathcal{U}^{\mathrm{ad}} = \left\{ \boldsymbol{U} \in \mathcal{U},\ \boldsymbol{U} \text{ measurable w.r.t. } \boldsymbol{Y} \right\} .$$

We will mainly concentrate on measurability constraints.

Variables and Constraints

The decision variable $\boldsymbol{U}$ is now a random variable and belongs to a functional space $\mathcal{U}$. A canonical example is: $\mathcal{U} = L^2(\Omega, \mathcal{A}, \mathbb{P}; \mathbb{U})$.

The contraints $\boldsymbol{U} \in \mathcal{U}^{\mathrm{ad}}$ on the r.v. $\boldsymbol{U}$ may be of different kinds:

- point-wise constraints dealing with the possible values of $\boldsymbol{U}$:

$$\boldsymbol{U} \in \mathcal{U}^{\mathrm{ad}} = \big\{ \boldsymbol{U} \in \mathcal{U},\ \boldsymbol{U}(\omega) \in U^{\mathrm{ad}}\ \mathbb{P}\text{-a.s.} \big\} \ ,$$

- risk constraints, such as expectation or probability constraints:

$$\boldsymbol{U} \in \mathcal{U}^{\mathrm{ad}} = \big\{ \boldsymbol{U} \in \mathcal{U},\ \mathbb{P}\big(\Theta(\boldsymbol{U}) \leq \theta\big) \geq \pi \big\} \ ,$$

- measurability constraints which express the fact that a given amount of information $\boldsymbol{Y}$ is available to the decision maker:

$$\boldsymbol{U} \in \mathcal{U}^{\mathrm{ad}} = \big\{ \boldsymbol{U} \in \mathcal{U},\ \boldsymbol{U} \text{ measurable w.r.t. } \boldsymbol{Y} \big\} \ .$$

We will mainly concentrate on measurability constraints.

Compact Formulation of a Closed-Loop Problem

Given a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, the essential ingredients of a stochastic optimization problem are

- noise $\boldsymbol{W}$: r.v. with values in a measurable space $(\mathbb{W}, \mathcal{W})$,
- decision $\boldsymbol{U}$: r.v. with values in a measurable space $(\mathbb{U}, \mathcal{U})$,
- information $\boldsymbol{Y}$: r.v. with values in a measurable space $(\mathbb{Y}, \mathcal{Y})$,
- cost function: measurable mapping $j : \mathbb{U} \times \mathbb{W} \to \mathbb{R}$.

The σ-field generated by $\boldsymbol{Y}$ is denoted by $\mathcal{B} \subset \mathcal{A}$.

With all these elements at hand, the problem is written as follows:

$$\min_{\boldsymbol{U} \preceq \boldsymbol{Y}} \mathbb{E}\big(j(\boldsymbol{U}, \boldsymbol{W})\big) .$$

The notation $\boldsymbol{U} \preceq \boldsymbol{Y}$ (or equivalently $\boldsymbol{U} \preceq \mathcal{B}$) is used to express that the r.v. $\boldsymbol{U}$ is measurable w.r.t. to the σ-field generated by $\boldsymbol{Y}$.

Compact Formulation of a Closed-Loop Problem

Given a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, the essential ingredients of a stochastic optimization problem are

- noise $\boldsymbol{W}$: r.v. with values in a measurable space $(\mathbb{W}, \mathcal{W})$,
- decision $\boldsymbol{U}$: r.v. with values in a measurable space $(\mathbb{U}, \mathcal{U})$,
- information $\boldsymbol{Y}$: r.v. with values in a measurable space $(\mathbb{Y}, \mathcal{Y})$,
- cost function: measurable mapping $j : \mathbb{U} \times \mathbb{W} \to \mathbb{R}$.

The σ-field generated by $\boldsymbol{Y}$ is denoted by $\mathcal{B} \subset \mathcal{A}$.

With all these elements at hand, the problem is written as follows:

$$\min_{\boldsymbol{U} \preceq \boldsymbol{Y}} \ \mathbb{E}\big(j(\boldsymbol{U}, \boldsymbol{W})\big) .$$

The notation $\boldsymbol{U} \preceq \boldsymbol{Y}$ (or equivalently $\boldsymbol{U} \preceq \mathcal{B}$) is used to express that the r.v. $\boldsymbol{U}$ is measurable w.r.t. to the σ-field generated by $\boldsymbol{Y}$.

Representation of Measurability Constraints

Consider the information structure of the stochastic optimization problem in a compact form, that is, the measurability constraints

$$\boldsymbol{U} \preceq \boldsymbol{Y} \ .$$

This information structure may be interpreted in different ways.

Representation of Measurability Constraints

Consider the information structure of the stochastic optimization problem in a compact form, that is, the measurability constraints

$$\boldsymbol{U} \preceq \boldsymbol{Y} \; .$$

This information structure may be interpreted in different ways.

- From the functional point of view, using Doob's Theorem, the decision $\boldsymbol{U}$ is expressed as a measurable function of $\boldsymbol{Y}$:

  $$\boldsymbol{U} = \varphi(\boldsymbol{Y}) \; .$$

  In this setting, the decision variable becomes the function φ.
- From the algebraic point of view, the constraints are expressed in terms of σ-field, that is,

  $$\sigma(\boldsymbol{U}) \subset \sigma(\boldsymbol{Y}) \; .$$

Representation of Measurability Constraints

Consider the information structure of the stochastic optimization problem in a compact form, that is, the measurability constraints

$$\boldsymbol{U} \preceq \boldsymbol{Y} .$$

This information structure may be interpreted in different ways.

- From the functional point of view, using Doob's Theorem, the decision $\boldsymbol{U}$ is expressed as a measurable function of $\boldsymbol{Y}$:

  $$\boldsymbol{U} = \varphi(\boldsymbol{Y}) .$$

  In this setting, the decision variable becomes the function φ.
- From the algebraic point of view, the constraints are expressed in terms of σ-field, that is,

  $$\sigma\big(\boldsymbol{U}\big) \subset \sigma\big(\boldsymbol{Y}\big) .$$

Question: *how to take this last representation into account?*

Representation of Measurability Constraints

Consider the information structure of the stochastic optimization problem in a compact form, that is, the measurability constraints

$$\boldsymbol{U} \preceq \boldsymbol{Y} \ .$$

This information structure may be interpreted in different ways.

- From the functional point of view, using Doob's Theorem, the decision $\boldsymbol{U}$ is expressed as a measurable function of $\boldsymbol{Y}$:

$$\boldsymbol{U} = \varphi(\boldsymbol{Y}) \ .$$

  In this setting, the decision variable becomes the function φ.
- From the algebraic point of view, the constraints are expressed in terms of σ-field, that is,

$$\sigma(\boldsymbol{U}) \subset \sigma(\boldsymbol{Y}) \ .$$

Question: *how to take this last representation into account?*

Representation of Measurability Constraints

Consider the information structure of the stochastic optimization problem in a compact form, that is, the measurability constraints

$$\boldsymbol{U} \preceq \boldsymbol{Y} .$$

This information structure may be interpreted in different ways.

- From the functional point of view, using Doob's Theorem, the decision $\boldsymbol{U}$ is expressed as a measurable function of $\boldsymbol{Y}$:

  $$\boldsymbol{U} = \varphi(\boldsymbol{Y}) .$$

  In this setting, the decision variable becomes the function φ.
- From the algebraic point of view, the constraints are expressed in terms of σ-field, that is,

  $$\sigma(\boldsymbol{U}) \subset \sigma(\boldsymbol{Y}) .$$

Question: *how to take this last representation into account?*

Dynamic Information Structure (DIS)

This is the situation when $\mathcal{B} = \sigma(\boldsymbol{Y})$ depends on $\boldsymbol{U}$. For example, in the case where $\boldsymbol{Y} = h(\boldsymbol{U}, \boldsymbol{W})$, the constraint expression is

$$\boldsymbol{U} \preceq h(\boldsymbol{U}, \boldsymbol{W}) ,$$

which yields a (seemingly) implicit measurability constraint.

This is a source of huge complexity for stochastic optimization problems, known under the name of the dual effect of control. Indeed, the decision maker has to take care of the following double effect:

Dynamic Information Structure (DIS)

This is the situation when $\mathcal{B} = \sigma(\boldsymbol{Y})$ depends on $\boldsymbol{U}$. For example, in the case where $\boldsymbol{Y} = h(\boldsymbol{U}, \boldsymbol{W})$, the constraint expression is

$$\boldsymbol{U} \preceq h(\boldsymbol{U}, \boldsymbol{W}) ,$$

which yields a (seemingly) implicit measurability constraint.

This is a source of huge complexity for stochastic optimization problems, known under the name of the dual effect of control. Indeed, the decision maker has to take care of the following double effect:

Dynamic Information Structure (DIS)

This is the situation when $\mathcal{B} = \sigma(\boldsymbol{Y})$ depends on $\boldsymbol{U}$. For example, in the case where $\boldsymbol{Y} = h(\boldsymbol{U}, \boldsymbol{W})$, the constraint expression is

$$\boldsymbol{U} \preceq h(\boldsymbol{U}, \boldsymbol{W}) \ ,$$

which yields a (seemingly) implicit measurability constraint.

This is a source of huge complexity for stochastic optimization problems, known under the name of the dual effect of control. Indeed, the decision maker has to take care of the following double effect:

- on the one hand, his decision affects the cost $\mathbb{E}\big(j(\boldsymbol{U}, \boldsymbol{W})\big)$,
- on the other hand, she makes the information more or less constrained, that is, a less or more large admissible set for $\boldsymbol{U}$.

Dynamic Information Structure (DIS)

This is the situation when $\mathcal{B} = \sigma(\boldsymbol{Y})$ depends on $\boldsymbol{U}$. For example, in the case where $\boldsymbol{Y} = h(\boldsymbol{U}, \boldsymbol{W})$, the constraint expression is

$$\boldsymbol{U} \preceq h(\boldsymbol{U}, \boldsymbol{W}) ,$$

which yields a (seemingly) implicit measurability constraint.

This is a source of huge complexity for stochastic optimization problems, known under the name of the dual effect of control. Indeed, the decision maker has to take care of the following double effect:

- on the one hand, his decision affects the cost $\mathbb{E}\big(j(\boldsymbol{U}, \boldsymbol{W})\big)$,
- on the other hand, she makes the information more or less constrained, that is, a less or more large admissible set for $\boldsymbol{U}$.

Static Information Structure (SIS)

This is the case when $\mathcal{B} = \sigma(\boldsymbol{Y})$ is fixed, defined independently of $\boldsymbol{U}$. Therefore, the terminology "static" expresses that the information σ-field $\mathcal{B}$ constraining the decision $\boldsymbol{U}$ cannot be modified by the decision maker. *It does not imply that no dynamics is present in the problem formulation.*[12]

[12] If time is involved in the problem, at each time t, a decision $\boldsymbol{U}_t$ is taken based on the available information $\boldsymbol{Y}_t$, inducing a measurability constraint $\boldsymbol{U}_t \preceq \boldsymbol{Y}_t$. But the issue of dynamic information depends on the dependency of $\boldsymbol{Y}_t$ w.r.t. the controls, and not on the presence of time t in the problem.

Static Information Structure (SIS)

This is the case when $\mathcal{B} = \sigma(\boldsymbol{Y})$ is fixed, defined independently of $\boldsymbol{U}$. Therefore, the terminology "static" expresses that the information σ-field $\mathcal{B}$ constraining the decision $\boldsymbol{U}$ cannot be modified by the decision maker. *It does not imply that no dynamics is present in the problem formulation.*[12]

- The situation where the information $\boldsymbol{Y}$ is a function of a exogenous noise $\boldsymbol{W}$, that is, $\boldsymbol{Y} = h(\boldsymbol{W})$, always induces a static information structure.
- Note that it may happen that $\boldsymbol{Y}$ functionally depends on $\boldsymbol{U}$ whereas the σ-field $\mathcal{B}$ generated by $\boldsymbol{Y}$ remains fixed.

[12] If time is involved in the problem, at each time t, a decision $\boldsymbol{U}_t$ is taken based on the available information $\boldsymbol{Y}_t$, inducing a measurability constraint $\boldsymbol{U}_t \preceq \boldsymbol{Y}_t$. But the issue of dynamic information depends on the dependency of $\boldsymbol{Y}_t$ w.r.t. the controls, and not on the presence of time t in the problem.

Static Information Structure (SIS)

This is the case when $\mathcal{B} = \sigma(\boldsymbol{Y})$ is fixed, defined independently of $\boldsymbol{U}$. Therefore, the terminology "static" expresses that the information σ-field $\mathcal{B}$ constraining the decision $\boldsymbol{U}$ cannot be modified by the decision maker. *It does not imply that no dynamics is present in the problem formulation.*[12]

- The situation where the information $\boldsymbol{Y}$ is a function of a exogenous noise $\boldsymbol{W}$, that is, $\boldsymbol{Y} = h(\boldsymbol{W})$, always induces a static information structure.
- Note that it may happen that $\boldsymbol{Y}$ functionally depends on $\boldsymbol{U}$ whereas the σ-field $\mathcal{B}$ generated by $\boldsymbol{Y}$ remains fixed.

[12] If time is involved in the problem, at each time t, a decision $\boldsymbol{U}_t$ is taken based on the available information $\boldsymbol{Y}_t$, inducing a measurability constraint $\boldsymbol{U}_t \preceq \boldsymbol{Y}_t$. But the issue of dynamic information depends on the dependency of $\boldsymbol{Y}_t$ w.r.t. the controls, and not on the presence of time t in the problem.

Position of the Problem...

We want to solve a closed-loop stochastic optimization problem, that is, a problem such that the decision variable $\boldsymbol{U}$ is a random variable which satisfies measurability conditions imposed by the information structure defined by the random variable $\boldsymbol{Y}$.

We assume that the problem is dual effect free, that is, we assume that the σ-field generated by the information variable $\boldsymbol{Y}$ does not depend on the control variable $\boldsymbol{U}$ (static information structure).

We manipulate the measurability conditions from the algebraic point of view, that is, $\sigma(\boldsymbol{U}) \subset \sigma(\boldsymbol{Y}) = \mathcal{B}$.

In order to numerically solve the optimization problem, we need to approximate the problem by using a finite representation of it.

Position of the Problem. . .

We want to solve a closed-loop stochastic optimization problem, that is, a problem such that the decision variable $\boldsymbol{U}$ is a random variable which satisfies measurability conditions imposed by the information structure defined by the random variable $\boldsymbol{Y}$.

We assume that the problem is dual effect free, that is, we assume that the σ-field generated by the information variable $\boldsymbol{Y}$ does not depend on the control variable $\boldsymbol{U}$ (static information structure).

We manipulate the measurability conditions from the algebraic point of view, that is, $\sigma(\boldsymbol{U}) \subset \sigma(\boldsymbol{Y}) = \mathcal{B}$.

In order to numerically solve the optimization problem, we need to approximate the problem by using a finite representation of it.

Position of the Problem...

We want to solve a closed-loop stochastic optimization problem, that is, a problem such that the decision variable $\boldsymbol{U}$ is a random variable which satisfies measurability conditions imposed by the information structure defined by the random variable $\boldsymbol{Y}$.

We assume that the problem is dual effect free, that is, we assume that the σ-field generated by the information variable $\boldsymbol{Y}$ does not depend on the control variable $\boldsymbol{U}$ (static information structure).

We manipulate the measurability conditions from the algebraic point of view, that is, $\sigma(\boldsymbol{U}) \subset \sigma(\boldsymbol{Y}) = \mathcal{B}$.

In order to numerically solve the optimization problem, we need to approximate the problem by using a finite representation of it.

Position of the Problem. . .

We want to solve a closed-loop stochastic optimization problem, that is, a problem such that the decision variable $\boldsymbol{U}$ is a random variable which satisfies measurability conditions imposed by the information structure defined by the random variable $\boldsymbol{Y}$.

We assume that the problem is dual effect free, that is, we assume that the σ-field generated by the information variable $\boldsymbol{Y}$ does not depend on the control variable $\boldsymbol{U}$ (static information structure).

We manipulate the measurability conditions from the algebraic point of view, that is, $\sigma(\boldsymbol{U}) \subset \sigma(\boldsymbol{Y}) = \mathcal{B}$.

In order to numerically solve the optimization problem, we need to approximate the problem by using a **finite representation** of it.

and Problem under Consideration

The standard form of the problem we are interested in is

$$\min_{\boldsymbol{U} \in \mathcal{U}} \mathbb{E}\big(j(\boldsymbol{U}, \boldsymbol{W})\big) ,$$

subject to

$$\boldsymbol{U} \text{ is } \mathcal{B}\text{-measurable} ,$$

where $\mathcal{B} = \sigma(\boldsymbol{Y})$ is a fixed σ-field.

In order to obtain a numerically tractable approximation of this problem, we have to approximate

- the noise $\boldsymbol{W}$ by a "finite" noise $\boldsymbol{W}_n$ (Monte Carlo, ...),
- the σ-field $\mathcal{B}$ by a "finite" σ-field $\mathcal{B}_n$ (partition, ...).

Question: $\mathcal{V}(\boldsymbol{W}_n, \mathcal{B}_n) \longrightarrow \mathcal{V}(\boldsymbol{W}, \mathcal{B})$?

and Problem under Consideration

The standard form of the problem we are interested in is

$$\mathcal{V}(\boldsymbol{W}, \mathcal{B}) = \min_{\boldsymbol{U} \in \mathcal{U}} \mathbb{E}\big(j(\boldsymbol{U}, \boldsymbol{W})\big) ,$$

subject to

$$\boldsymbol{U} \text{ is } \mathcal{B}\text{-measurable} ,$$

where $\mathcal{B} = \sigma(\boldsymbol{Y})$ is a fixed σ-field.

In order to obtain a numerically tractable approximation of this problem, we have to approximate

- the noise $\boldsymbol{W}$ by a "finite" noise $\boldsymbol{W}_n$ (Monte Carlo,...),
- the σ-field $\mathcal{B}$ by a "finite" σ-field $\mathcal{B}_n$ (partition,...).

Question: $\mathcal{V}(\boldsymbol{W}_n, \mathcal{B}_n) \longrightarrow \mathcal{V}(\boldsymbol{W}, \mathcal{B})$?

and Problem under Consideration

The standard form of the problem we are interested in is

$$\mathcal{V}(\boldsymbol{W}, \mathcal{B}) = \min_{\boldsymbol{U} \in \mathcal{U}} \mathbb{E}\big(j(\boldsymbol{U}, \boldsymbol{W})\big) \ ,$$

subject to

$$\boldsymbol{U} \text{ is } \mathcal{B}\text{-measurable} \ ,$$

where $\mathcal{B} = \sigma(\boldsymbol{Y})$ is a fixed σ-field.

In order to obtain a numerically tractable approximation of this problem, we have to approximate

- the noise $\boldsymbol{W}$ by a "finite" noise $\boldsymbol{W}_n$ (Monte Carlo,...),
- the σ-field $\mathcal{B}$ by a "finite" σ-field $\mathcal{B}_n$ (partition,...).

Question: $\mathcal{V}(\boldsymbol{W}_n, \mathcal{B}_n) \longrightarrow \mathcal{V}(\boldsymbol{W}, \mathcal{B})$?

and Problem under Consideration

The standard form of the problem we are interested in is

$$\mathcal{V}(\boldsymbol{W}, \mathcal{B}) = \min_{\boldsymbol{U} \in \mathcal{U}} \mathbb{E}\big(j(\boldsymbol{U}, \boldsymbol{W})\big) \ ,$$

subject to

$$\boldsymbol{U} \text{ is } \mathcal{B}\text{-measurable} \ ,$$

where $\mathcal{B} = \sigma(\boldsymbol{Y})$ is a fixed σ-field.

In order to obtain a numerically tractable approximation of this problem, we have to approximate

- the noise $\boldsymbol{W}$ by a “finite” noise $\boldsymbol{W}_n$ (Monte Carlo,. . .),
- the σ-field $\mathcal{B}$ by a “finite” σ-field $\mathcal{B}_n$ (partition,. . .).

Question: $\mathcal{V}(\boldsymbol{W}_n, \mathcal{B}_n) \longrightarrow \mathcal{V}(\boldsymbol{W}, \mathcal{B})$?

and Problem under Consideration

The standard form of the problem we are interested in is

$$\mathcal{V}(\boldsymbol{W}, \mathcal{B}) = \min_{\boldsymbol{U} \in \mathcal{U}} \mathbb{E}\big(j(\boldsymbol{U}, \boldsymbol{W})\big) ,$$

subject to

$$\boldsymbol{U} \text{ is } \mathcal{B}\text{-measurable} ,$$

where $\mathcal{B} = \sigma(\boldsymbol{Y})$ is a fixed σ-field.

In order to obtain a numerically tractable approximation of this problem, we have to approximate

- the noise $\boldsymbol{W}$ by a "finite" noise $\boldsymbol{W}_n$ (Monte Carlo,. . .),
- the σ-field $\mathcal{B}$ by a "finite" σ-field $\mathcal{B}_n$ (partition,. . .).

Question: $\mathcal{V}(\boldsymbol{W}_n, \mathcal{B}_n) \longrightarrow \mathcal{V}(\boldsymbol{W}, \mathcal{B})$?

and Problem under Consideration

The standard form of the problem we are interested in is

$$\mathcal{V}(\boldsymbol{W}, \mathcal{B}) = \min_{\boldsymbol{U} \in \mathcal{U}} \mathbb{E}\big(j(\boldsymbol{U}, \boldsymbol{W})\big) ,$$

subject to

$$\boldsymbol{U} \text{ is } \mathcal{B}\text{-measurable} ,$$

where $\mathcal{B} = \sigma(\boldsymbol{Y})$ is a fixed σ-field.

In order to obtain a numerically tractable approximation of this problem, we have to approximate

- the noise $\boldsymbol{W}$ by a "finite" noise $\boldsymbol{W}_n$ (Monte Carlo,...),
- the σ-field $\mathcal{B}$ by a "finite" σ-field $\mathcal{B}_n$ (partition,...).

Question: $\mathcal{V}(\boldsymbol{W}_n, \mathcal{B}_n) \longrightarrow \mathcal{V}(\boldsymbol{W}, \mathcal{B})$?

A Specific Instance of the Problem

A specific instance of the problem is the one which incorporates dynamical systems, that is, the stochastic optimal control problem:

$$\min_{(\boldsymbol{U}_0,\dots,\boldsymbol{U}_{T-1},\boldsymbol{X}_0,\dots,\boldsymbol{X}_T)} \mathbb{E}\bigg(\sum_{t=0}^{T-1} L_t(\boldsymbol{X}_t, \boldsymbol{U}_t, \boldsymbol{W}_{t+1}) + K(\boldsymbol{X}_T)\bigg)$$

subject to

$$\boldsymbol{X}_0 = f_{-1}(\boldsymbol{W}_0) \; ,$$

$$\boldsymbol{X}_{t+1} = f_t(\boldsymbol{X}_t, \boldsymbol{U}_t, \boldsymbol{W}_{t+1}) \; , \quad t = 0, \dots, T-1 \; ,$$

$$\boldsymbol{U}_t \preceq \boldsymbol{Y}_t \; , \quad t = 0, \dots, T-1 \; .$$

Assuming that $\sigma(\boldsymbol{Y}_t)$ are fixed σ-fields, a widely used approach to discretize this optimization problem is the so-called scenario tree method. We present it before considering the general case.

A Specific Instance of the Problem

A specific instance of the problem is the one which incorporates dynamical systems, that is, the stochastic optimal control problem:

$$\min_{(\boldsymbol{U}_0,\dots,\boldsymbol{U}_{T-1},\boldsymbol{X}_0,\dots,\boldsymbol{X}_T)} \mathbb{E}\left(\sum_{t=0}^{T-1} L_t(\boldsymbol{X}_t, \boldsymbol{U}_t, \boldsymbol{W}_{t+1}) + K(\boldsymbol{X}_T)\right)$$

subject to

$$\begin{aligned} \boldsymbol{X}_0 &= f_{-1}(\boldsymbol{W}_0) , \\ \boldsymbol{X}_{t+1} &= f_t(\boldsymbol{X}_t, \boldsymbol{U}_t, \boldsymbol{W}_{t+1}) , \quad t = 0, \dots, T-1 , \\ \\ \boldsymbol{U}_t &\preceq \boldsymbol{Y}_t , \quad t = 0, \dots, T-1 . \end{aligned}$$

Assuming that $\sigma(\boldsymbol{Y}_t)$ are fixed σ-fields, a widely used approach to discretize this optimization problem is the so-called scenario tree method. We present it before considering the general case.

Lecture Outline

1. Stochastic Programming: the Scenario Tree Method
   - Scenario Tree Method Overview
   - Some Details about the Method
2. Stochastic Optimal Control and Discretization Puzzles
   - Working out an Example
   - Naive Monte Carlo-Based Discretization
   - Scenario Tree-Based Discretization
   - A Constructive Proposal
3. A General Convergence Result
   - Convergence of Random Variables
   - Convergence of σ-Fields
   - The Long-Awaited Convergence Theorem

A Standard Stochastic Optimal Control Problem

Consider the following stochastic optimal control problem with a static (non-anticipative) information structure.

$$\min_{(\boldsymbol{U}_0,\dots,\boldsymbol{U}_{T-1},\boldsymbol{X}_0,\dots,\boldsymbol{X}_T)} \mathbb{E}\left(\sum_{t=0}^{T-1} L_t(\boldsymbol{X}_t, \boldsymbol{U}_t, \boldsymbol{W}_{t+1}) + K(\boldsymbol{X}_T) \right)$$

subject to

$$\begin{aligned} \boldsymbol{X}_0 &= f_{-1}(\boldsymbol{W}_0) \ , \\ \boldsymbol{X}_{t+1} &= f_t(\boldsymbol{X}_t, \boldsymbol{U}_t, \boldsymbol{W}_{t+1}) \ , \quad t = 0, \dots, T-1 \ , \end{aligned}$$

$$\boldsymbol{U}_t \preceq h_t(\boldsymbol{W}_0, \dots, \boldsymbol{W}_t) \ , \qquad t = 0, \dots, T-1 \ .$$

Almost sure constraints (e.g. bound constraints on $\boldsymbol{X}_t$ and $\boldsymbol{U}_t$) may also be present in the formulation.

A Standard Stochastic Optimal Control Problem

Consider the following stochastic optimal control problem with a static (non-anticipative) information structure.

$$\min_{(\boldsymbol{U}_0,\dots,\boldsymbol{U}_{T-1},\boldsymbol{X}_0,\dots,\boldsymbol{X}_T)} \mathbb{E}\left(\sum_{t=0}^{T-1} L_t(\boldsymbol{X}_t, \boldsymbol{U}_t, \boldsymbol{W}_{t+1}) + K(\boldsymbol{X}_T)\right)$$

subject to

$$\begin{aligned} \boldsymbol{X}_0 &= f_{-1}(\boldsymbol{W}_0) \ , \\ \boldsymbol{X}_{t+1} &= f_t(\boldsymbol{X}_t, \boldsymbol{U}_t, \boldsymbol{W}_{t+1}) \ , \quad t = 0, \dots, T-1 \ , \end{aligned}$$

$$\boldsymbol{U}_t \preceq h_t(\boldsymbol{W}_0, \dots, \boldsymbol{W}_t) \ , \qquad t = 0, \dots, T-1 \ .$$

Almost sure constraints (e.g. bound constraints on $\boldsymbol{X}_t$ and $\boldsymbol{U}_t$) may also be present in the formulation.

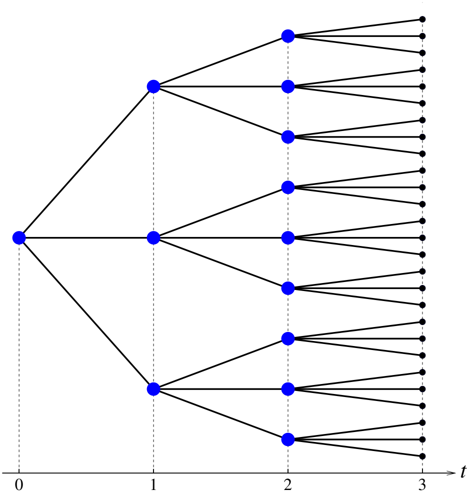
Scenario Tree Methodology

Aim: obtain a finite dimensional approximation of the problem.

Scenario Tree Methodology

Aim: obtain a finite dimensional approximation of the problem.

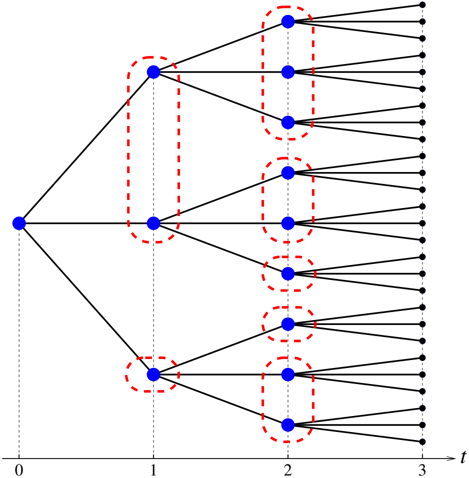
1. Discretize the noise process $\{\boldsymbol{W}_t\}$ using a scenario tree.
2. Copy out the measurability constraints on this structure:
$$\boldsymbol{U}_t \preceq h_t(\boldsymbol{W}_0, \ldots, \boldsymbol{W}_t).$$
3. Write the dynamics and cost functions at the tree nodes:
$$\boldsymbol{X}_{t+1} = f_t(\boldsymbol{X}_t, \boldsymbol{U}_t, \boldsymbol{W}_{t+1}).$$
4. Solve the problem using adequate mathematical programming techniques.



Scenario Tree Methodology

Aim: obtain a finite dimensional approximation of the problem.

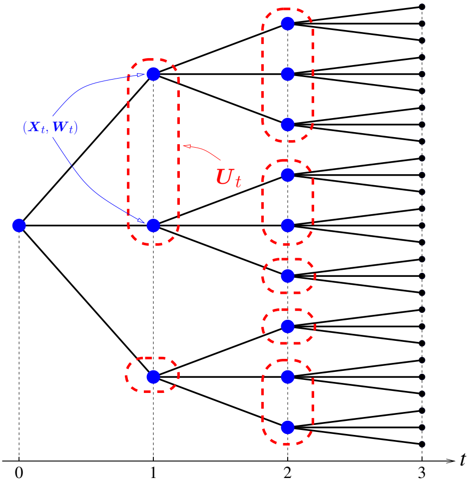
1. Discretize the noise process $\{\boldsymbol{W}_t\}$ using a scenario tree.
2. Copy out the measurability constraints on this structure:
   $$\boldsymbol{U}_t \preceq h_t(\boldsymbol{W}_0, \ldots, \boldsymbol{W}_t).$$
3. Write the dynamics and cost functions at the tree nodes:
   $$\boldsymbol{X}_{t+1} = f_t(\boldsymbol{X}_t, \boldsymbol{U}_t, \boldsymbol{W}_{t+1}).$$
4. Solve the problem using adequate mathematical programming techniques.



Scenario Tree Methodology

Aim: obtain a finite dimensional approximation of the problem.

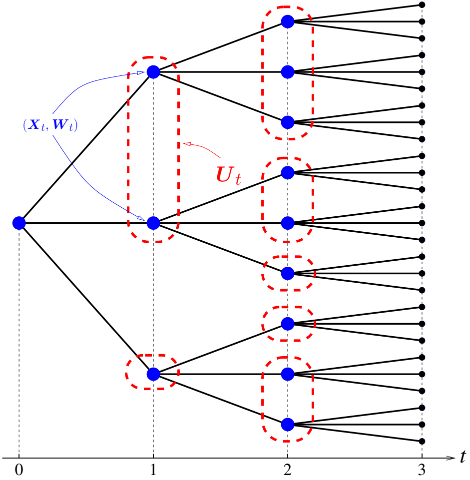
1. Discretize the noise process $\{\boldsymbol{W}_t\}$ using a scenario tree.
2. Copy out the measurability constraints on this structure:
$$\boldsymbol{U}_t \preceq h_t(\boldsymbol{W}_0, \ldots, \boldsymbol{W}_t).$$
3. Write the dynamics and cost functions at the tree nodes:
$$\boldsymbol{X}_{t+1} = f_t(\boldsymbol{X}_t, \boldsymbol{U}_t, \boldsymbol{W}_{t+1}).$$
4. Solve the problem using adequate mathematical programming techniques.



Scenario Tree Methodology

Aim: obtain a finite dimensional approximation of the problem.

1. Discretize the noise process $\{\boldsymbol{W}_t\}$ using a scenario tree.
2. Copy out the measurability constraints on this structure:
   $$\boldsymbol{U}_t \preceq h_t(\boldsymbol{W}_0, \dots, \boldsymbol{W}_t).$$
3. Write the dynamics and cost functions at the tree nodes:
   $$\boldsymbol{X}_{t+1} = f_t(\boldsymbol{X}_t, \boldsymbol{U}_t, \boldsymbol{W}_{t+1}).$$
4. Solve the problem using adequate mathematical programming techniques.



1. Discretize the Random Inputs

The tree architecture is characterized by the fact that each node of the tree corresponds to a unique past noise history but is generally followed by several possible future histories.

The tree is obtained by repeatedly using a finite approximation of the conditional probability laws $\mathbb{P}(W_t \mid W_0, \ldots, W_{t-1})$:

$$\mathbb{P}(W_0) \approx \{w_0^1, \ldots, w_0^{n_0}\} \rightsquigarrow \mathbb{P}(W_1 \mid W_0 = w_0^i) \approx \{w_1^{i,1}, \ldots, w_0^{i,n_1}\} \ \ldots$$

Note that this discretization scheme is much more sophisticated than the standard Monte Carlo sampling of $(W_0, \ldots, W_T)$.

The starting point may be a given collection of scenarios from which one constructs a tree by grouping the scenarios according to their (approximate) common past.

1. Discretize the Random Inputs

The tree architecture is characterized by the fact that each node of the tree corresponds to a unique past noise history but is generally followed by several possible future histories.

The tree is obtained by repeatedly using a finite approximation of the conditional probability laws $\mathbb{P}(\boldsymbol{W}_t \mid \boldsymbol{W}_0, \dots, \boldsymbol{W}_{t-1})$:

$$\mathbb{P}(\boldsymbol{W}_0) \approx \{w_0^1, \dots, w_0^{n_0}\} \;\rightsquigarrow\; \mathbb{P}(\boldsymbol{W}_1 \mid \boldsymbol{W}_0 = w_0^i) \approx \{w_1^{i,1}, \dots, w_0^{i,n_1}\} \;\dots$$

Note that this discretization scheme is much more sophisticated than the standard Monte Carlo sampling of $(\boldsymbol{W}_0, \dots, \boldsymbol{W}_T)$.

The starting point may be a given collection of scenarios from which one constructs a tree by grouping the scenarios according to their (approximate) common past.

1. Discretize the Random Inputs

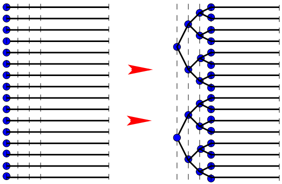
The tree architecture is characterized by the fact that each node of the tree corresponds to a unique past noise history but is generally followed by several possible future histories.

The tree is obtained by repeatedly using a finite approximation of the conditional probability laws $\mathbb{P}(\boldsymbol{W}_t \mid \boldsymbol{W}_0, \ldots, \boldsymbol{W}_{t-1})$:

$$\mathbb{P}(\boldsymbol{W}_0) \approx \{w_0^1, \ldots, w_0^{n_0}\} \quad \rightsquigarrow \quad \mathbb{P}(\boldsymbol{W}_1 \mid \boldsymbol{W}_0 = w_0^i) \approx \{w_1^{i,1}, \ldots, w_0^{i,n_1}\} \quad \ldots$$

Note that this discretization scheme is much more sophisticated than the standard Monte Carlo sampling of $(\boldsymbol{W}_0, \ldots, \boldsymbol{W}_T)$.

The starting point may be a given collection of scenarios from which one constructs a tree by grouping the scenarios according to their (approximate) common past.



2. Copy out the Measurability Constraints

Assume that the information consists of the exact observation of all past noises: $\boldsymbol{Y}_t = (\boldsymbol{W}_0, \ldots, \boldsymbol{W}_t)$. Then, a different decision has to be attached at each node of the scenario tree.

But the method can face more general situations by grouping nodes of the scenario tree in order to represent the information structure induced by the $h_t(\boldsymbol{W}_0, \ldots, \boldsymbol{W}_t)$'s.

In all cases, the information structure is entirely coded within the scenario tree by means of those groups of nodes (one decision for each group of nodes).

For example, the so-called perfect memory information structure $h_t(\boldsymbol{W}_0, \ldots, \boldsymbol{W}_t) = (\hbar_0(\boldsymbol{W}_0), \ldots, \hbar_t(\boldsymbol{W}_t))$ leads to a grouping of scenario tree nodes at each time step t, and ultimately produces a tree structure called the decision tree.

2. Copy out the Measurability Constraints

Assume that the information consists of the exact observation of all past noises: $\boldsymbol{Y}_t = (\boldsymbol{W}_0, \ldots, \boldsymbol{W}_t)$. Then, a different decision has to be attached at each node of the scenario tree.

But the method can face more general situations by grouping nodes of the scenario tree in order to represent the information structure induced by the $h_t(\boldsymbol{W}_0, \ldots, \boldsymbol{W}_t)$'s.

In all cases, the information structure is entirely coded within the scenario tree by means of those groups of nodes (one decision for each group of nodes).

For example, the so-called perfect memory information structure $h_t(\boldsymbol{W}_0, \ldots, \boldsymbol{W}_t) = \big(\hbar_0(\boldsymbol{W}_0), \ldots, \hbar_t(\boldsymbol{W}_t)\big)$ leads to a grouping of scenario tree nodes at each time step t, and ultimately produces a tree structure called the decision tree.

2. Copy out the Measurability Constraints

Assume that the information consists of the exact observation of all past noises: $\boldsymbol{Y}_t = (\boldsymbol{W}_0, \ldots, \boldsymbol{W}_t)$. Then, a different decision has to be attached at each node of the scenario tree.

But the method can face more general situations by grouping nodes of the scenario tree in order to represent the information structure induced by the $h_t(\boldsymbol{W}_0, \ldots, \boldsymbol{W}_t)$'s.

In all cases, the information structure is entirely coded within the scenario tree by means of those groups of nodes (one decision for each group of nodes).

For example, the so-called perfect memory information structure $h_t(\boldsymbol{W}_0, \ldots, \boldsymbol{W}_t) = (\hbar_0(\boldsymbol{W}_0), \ldots, \hbar_t(\boldsymbol{W}_t))$ leads to a grouping of scenario tree nodes at each time step t, and ultimately produces a tree structure called the decision tree.

2. Copy out the Measurability Constraints

Assume that the information consists of the exact observation of all past noises: $\boldsymbol{Y}_t = (\boldsymbol{W}_0, \ldots, \boldsymbol{W}_t)$. Then, a different decision has to be attached at each node of the scenario tree.

But the method can face more general situations by grouping nodes of the scenario tree in order to represent the information structure induced by the $h_t(\boldsymbol{W}_0, \ldots, \boldsymbol{W}_t)$'s.

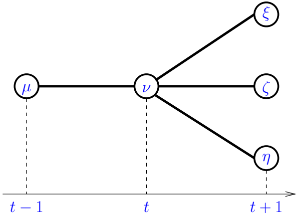
In all cases, the information structure is entirely coded within the scenario tree by means of those groups of nodes (one decision for each group of nodes).

For example, the so-called perfect memory information structure $h_t(\boldsymbol{W}_0, \ldots, \boldsymbol{W}_t) = \big(\hbar_0(\boldsymbol{W}_0), \ldots, \hbar_t(\boldsymbol{W}_t)\big)$ leads to a grouping of scenario tree nodes at each time step t, and ultimately produces a tree structure called the decision tree.

3. Write the Dynamics and Cost Functions

Consider a node $\nu \in \mathcal{N}$ of the scenario tree at time t, and denote:

- $\mathfrak{f}(\nu)$ the predecessor of node ν $(= \mu)$,
- $\pi(\nu)$ the probability of node ν,
- $\theta(\nu)$ the time index of node ν $(= t)$,
- $\gamma(\nu)$ the control index of node ν.



Note that the probability function π satisfies the following conditions:

$$\pi(\nu) = \sum_{\xi \in \mathfrak{f}^{-1}(\nu)} \pi(\xi) \;, \qquad \sum_{\nu \in \theta^{-1}(t)} \pi(\nu) = 1 \;.$$

Then, the dynamic equation from node μ to node ν writes

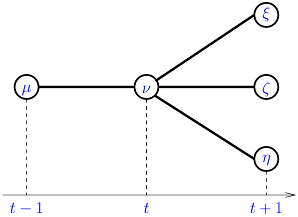
$$x_\nu = f_{\theta(\mathfrak{f}(\nu))}\big(x_{\mathfrak{f}(\nu)}, u_{\gamma(\mathfrak{f}(\nu))}, w_\nu\big) \;.$$

The cost induced by the transition is: $L_{\theta(\mathfrak{f}(\nu))}\big(x_{\mathfrak{f}(\nu)}, u_{\gamma(\mathfrak{f}(\nu))}, w_\nu\big)$.

3. Write the Dynamics and Cost Functions

Consider a node $\nu \in \mathcal{N}$ of the scenario tree at time t, and denote:

- $\mathfrak{f}(\nu)$ the predecessor of node ν $(= \mu)$,
- $\pi(\nu)$ the probability of node ν,
- $\theta(\nu)$ the time index of node ν $(= t)$,
- $\gamma(\nu)$ the control index of node ν.



Note that the probability function π satisfies the following conditions:

$$\pi(\nu) = \sum_{\xi \in \mathfrak{f}^{-1}(\nu)} \pi(\xi) \quad , \quad \sum_{\nu \in \theta^{-1}(t)} \pi(\nu) = 1 \; .$$

Then, the dynamic equation from node μ to node ν writes

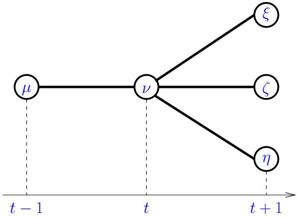
$$x_\nu = f_{\theta(\mathfrak{f}(\nu))}\big(x_{\mathfrak{f}(\nu)}, u_{\gamma(\mathfrak{f}(\nu))}, w_\nu\big) \; .$$

The cost induced by the transition is: $L_{\theta(\mathfrak{f}(\nu))}\big(x_{\mathfrak{f}(\nu)}, u_{\gamma(\mathfrak{f}(\nu))}, w_\nu\big)$.

3. Write the Dynamics and Cost Functions

Consider a node $\nu \in \mathcal{N}$ of the scenario tree at time t, and denote:

- $\mathfrak{f}(\nu)$ the predecessor of node ν $(= \mu)$,
- $\pi(\nu)$ the probability of node ν,
- $\theta(\nu)$ the time index of node ν $(= t)$,
- $\gamma(\nu)$ the control index of node ν.



Note that the probability function π satisfies the following conditions:

$$\pi(\nu) = \sum_{\xi \in \mathfrak{f}^{-1}(\nu)} \pi(\xi) \quad , \quad \sum_{\nu \in \theta^{-1}(t)} \pi(\nu) = 1 \ .$$

Then, the dynamic equation from node μ to node ν writes

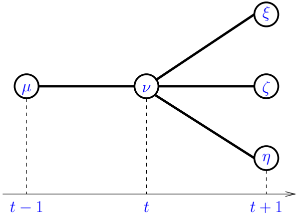
$$x_\nu = f_{\theta(\mathfrak{f}(\nu))}\big(x_{\mathfrak{f}(\nu)}, u_{\gamma(\mathfrak{f}(\nu))}, w_\nu\big) \ .$$

The cost induced by the transition is: $L_{\theta(\mathfrak{f}(\nu))}\big(x_{\mathfrak{f}(\nu)}, u_{\gamma(\mathfrak{f}(\nu))}, w_\nu\big)$.

3. Write the Dynamics and Cost Functions

Consider a node $\nu \in \mathcal{N}$ of the scenario tree at time t, and denote:

- $\mathfrak{f}(\nu)$ the predecessor of node ν ($= \mu$),
- $\pi(\nu)$ the probability of node ν,
- $\theta(\nu)$ the time index of node ν ($= t$),
- $\gamma(\nu)$ the control index of node ν.



Note that the probability function π satisfies the following conditions:

$$\pi(\nu) = \sum_{\xi \in \mathfrak{f}^{-1}(\nu)} \pi(\xi) \quad , \quad \sum_{\nu \in \theta^{-1}(t)} \pi(\nu) = 1 \ .$$

Then, the dynamic equation from node μ to node ν writes

$$x_\nu = f_{\theta(\mathfrak{f}(\nu))}\big(x_{\mathfrak{f}(\nu)}, u_{\gamma(\mathfrak{f}(\nu))}, w_\nu\big) \ .$$

The cost induced by the transition is: $L_{\theta(\mathfrak{f}(\nu))}\big(x_{\mathfrak{f}(\nu)}, u_{\gamma(\mathfrak{f}(\nu))}, w_\nu\big)$.

4. Solve the Approximated Problem

The initial stochastic optimization problem boils down to

$$\min \left(\sum_{\nu \in \mathcal{N} \setminus \theta^{-1}(0)} \pi(\nu) L_{\theta(\mathfrak{f}(\nu))}\big(x_{\mathfrak{f}(\nu)}, u_{\gamma(\mathfrak{f}(\nu))}, w_\nu\big) + \sum_{\nu \in \theta^{-1}(T)} \pi(\nu) K(x_\nu) \right) ,$$

subject only to the dynamics constraints

$$x_\nu = f_{-1}(w_\nu) \qquad \forall \nu \in \theta^{-1}(0) ,$$

$$x_\nu = f_{\theta(\mathfrak{f}(\nu))}\big(x_{\mathfrak{f}(\nu)}, u_{\gamma(\mathfrak{f}(\nu))}, w_\nu\big) \qquad \forall \nu \in \mathcal{N} \setminus \theta^{-1}(0) .$$

The initial infinite dimensional stochastic optimization problem is approximated by a finite dimensional deterministic problem, that can be solved using relevant mathematical programming tools.

Note that this approximation corresponds to an optimal control problem with an arborescent (rather than linear) time structure.

4. Solve the Approximated Problem

The initial stochastic optimization problem boils down to

$$\min \left(\sum_{\nu \in \mathcal{N} \setminus \theta^{-1}(0)} \pi(\nu) L_{\theta(\mathfrak{f}(\nu))}\big(x_{\mathfrak{f}(\nu)}, u_{\gamma(\mathfrak{f}(\nu))}, w_\nu\big) + \sum_{\nu \in \theta^{-1}(T)} \pi(\nu) K(x_\nu) \right) ,$$

subject only to the dynamics constraints

$$x_\nu = f_{-1}(w_\nu) \qquad \forall \nu \in \theta^{-1}(0) ,$$

$$x_\nu = f_{\theta(\mathfrak{f}(\nu))}\big(x_{\mathfrak{f}(\nu)}, u_{\gamma(\mathfrak{f}(\nu))}, w_\nu\big) \qquad \forall \nu \in \mathcal{N} \setminus \theta^{-1}(0) .$$

The initial infinite dimensional stochastic optimization problem is approximated by a finite dimensional deterministic problem, that can be solved using relevant mathematical programming tools.

Note that this approximation corresponds to an optimal control problem with an arborescent (rather than linear) time structure.

4. Solve the Approximated Problem

The initial stochastic optimization problem boils down to

$$\min \left(\sum_{\nu \in \mathcal{N} \setminus \theta^{-1}(0)} \pi(\nu) L_{\theta(\mathfrak{f}(\nu))}\big(x_{\mathfrak{f}(\nu)}, u_{\gamma(\mathfrak{f}(\nu))}, w_\nu\big) + \sum_{\nu \in \theta^{-1}(T)} \pi(\nu) K(x_\nu) \right) ,$$

subject only to the dynamics constraints

$$x_\nu = f_{-1}(w_\nu) \qquad \forall \nu \in \theta^{-1}(0) ,$$

$$x_\nu = f_{\theta(\mathfrak{f}(\nu))}\big(x_{\mathfrak{f}(\nu)}, u_{\gamma(\mathfrak{f}(\nu))}, w_\nu\big) \qquad \forall \nu \in \mathcal{N} \setminus \theta^{-1}(0) .$$

The initial infinite dimensional stochastic optimization problem is approximated by a finite dimensional deterministic problem, that can be solved using relevant mathematical programming tools.

Note that this approximation corresponds to an optimal control problem with an arborescent (rather than linear) time structure.

Facts and Questions about the Scenario Tree Method

Dual Effect: it is mandatory that no dual effect holds true.

White noise: the noise process $(W_0, \ldots, W_T)$ may be correlated.

Perfect memory : this property is not required although useful.

Complexity: the amount of scenarios needed to achieve a given accuracy grows exponentially w.r.t. the number of time steps T of the problem (see [Shapiro, 2006]).

Tree structure: how to build a tree which is at the same time representative of the problem and numerically tractable?

Extrapolation: how to obtain feedback laws once the optimal decisions on the nodes of the scenario tree have been computed?

A huge literature is available on the scenario tree method...

Facts and Questions about the Scenario Tree Method

Dual Effect: it is mandatory that no dual effect holds true.

White noise: the noise process $(W_0, \ldots, W_T)$ may be correlated.

Perfect memory : this property is not required although useful.

Complexity: the amount of scenarios needed to achieve a given accuracy grows exponentially w.r.t. the number of time steps T of the problem (see [Shapiro, 2006]).

Tree structure: how to build a tree which is at the same time representative of the problem and numerically tractable?

Extrapolation: how to obtain feedback laws once the optimal decisions on the nodes of the scenario tree have been computed?

A huge literature is available on the scenario tree method. . .

Facts and Questions about the Scenario Tree Method

Dual Effect: it is mandatory that no dual effect holds true.

White noise: the noise process $(\boldsymbol{W}_0, \ldots, \boldsymbol{W}_T)$ may be correlated.

Perfect memory : this property is not required although useful.

Complexity: the amount of scenarios needed to achieve a given accuracy grows exponentially w.r.t. the number of time steps T of the problem (see [Shapiro, 2006]).

Tree structure: how to build a tree which is at the same time representative of the problem and numerically tractable?

Extrapolation: how to obtain feedback laws once the optimal decisions on the nodes of the scenario tree have been computed?

A huge literature is available on the scenario tree method. . .

Facts and Questions about the Scenario Tree Method

Dual Effect: it is mandatory that no dual effect holds true.

White noise: the noise process $(\boldsymbol{W}_0, \ldots, \boldsymbol{W}_T)$ may be correlated.

Perfect memory : this property is not required although useful.

Complexity: the amount of scenarios needed to achieve a given accuracy grows exponentially w.r.t. the number of time steps T of the problem (see [Shapiro, 2006]).

Tree structure: how to build a tree which is at the same time representative of the problem and numerically tractable?

Extrapolation: how to obtain feedback laws once the optimal decisions on the nodes of the scenario tree have been computed?

A huge literature is available on the scenario tree method. . .

Facts and Questions about the Scenario Tree Method

Dual Effect: it is mandatory that no dual effect holds true.

White noise: the noise process $(\boldsymbol{W}_0, \ldots, \boldsymbol{W}_T)$ may be correlated.

Perfect memory : this property is not required although useful.

Complexity: the amount of scenarios needed to achieve a given accuracy grows exponentially w.r.t. the number of time steps T of the problem (see [Shapiro, 2006]).

Tree structure: how to build a tree which is at the same time representative of the problem and numerically tractable?

Extrapolation: how to obtain feedback laws once the optimal decisions on the nodes of the scenario tree have been computed?

A huge literature is available on the scenario tree method...

Facts and Questions about the Scenario Tree Method

Dual Effect: it is mandatory that no dual effect holds true.

White noise: the noise process $(\boldsymbol{W}_0, \dots, \boldsymbol{W}_T)$ may be correlated.

Perfect memory : this property is not required although useful.

Complexity: the amount of scenarios needed to achieve a given accuracy grows exponentially w.r.t. the number of time steps T of the problem (see [Shapiro, 2006]).

Tree structure: how to build a tree which is at the same time representative of the problem and numerically tractable?

Extrapolation: how to obtain feedback laws once the optimal decisions on the nodes of the scenario tree have been computed?

A huge literature is available on the scenario tree method...

Facts and Questions about the Scenario Tree Method

Dual Effect: it is mandatory that no dual effect holds true.

White noise: the noise process $(\boldsymbol{W}_0, \ldots, \boldsymbol{W}_T)$ may be correlated.

Perfect memory : this property is not required although useful.

Complexity: the amount of scenarios needed to achieve a given accuracy grows exponentially w.r.t. the number of time steps T of the problem (see [Shapiro, 2006]).

Tree structure: how to build a tree which is at the same time representative of the problem and numerically tractable?

Extrapolation: how to obtain feedback laws once the optimal decisions on the nodes of the scenario tree have been computed?

A huge literature is available on the scenario tree method. . .

Facts and Questions about the Scenario Tree Method

Dual Effect: it is mandatory that no dual effect holds true.

White noise: the noise process $(\boldsymbol{W}_0, \ldots, \boldsymbol{W}_T)$ may be correlated.

Perfect memory : this property is not required although useful.

Complexity: the amount of scenarios needed to achieve a given accuracy grows exponentially w.r.t. the number of time steps T of the problem (see [Shapiro, 2006]).

Tree structure: how to build a tree which is at the same time representative of the problem and numerically tractable?

Extrapolation: how to obtain feedback laws once the optimal decisions on the nodes of the scenario tree have been computed?

A huge literature is available on the scenario tree method...

(Very) Compact View of the Scenario Tree Approach

The stochastic optimal control problem under consideration depends on both a noise process $\boldsymbol{W}$ and a sequence of σ-fields $\mathcal{B}$. It can thus be represented under the compact form:

$$\mathcal{V}(\boldsymbol{W}, \mathcal{B}) = \min_{\boldsymbol{U} \preceq \mathcal{B}} \mathbb{E}\big(j(\boldsymbol{U}, \boldsymbol{W})\big) , \quad \text{with} \quad \mathcal{B} = \sigma\big(h(\boldsymbol{W})\big) .$$

The aim of the scenario tree method is to

- approximate the noise $\boldsymbol{W}$ by a "finite" noise $\boldsymbol{W}_n$,
- and deduce the approximated information $\mathcal{B}_n = \sigma\big(h(\boldsymbol{W}_n)\big)$.

In this framework, an unique approximation is performed to obtain the approximated solution $\mathcal{V}(\boldsymbol{W}_n, \mathcal{B}_n)$, and it is possible to prove that $\mathcal{V}(\boldsymbol{W}_n, \mathcal{B}_n) \to \mathcal{V}(\boldsymbol{W}, \mathcal{B})$ (see [Pennanen, 2005]).

But the noise has been discretized in a very specific way...

(Very) Compact View of the Scenario Tree Approach

The stochastic optimal control problem under consideration depends on both a noise process $\boldsymbol{W}$ and a sequence of σ-fields $\mathcal{B}$. It can thus be represented under the compact form:

$$\mathcal{V}(\boldsymbol{W}, \mathcal{B}) = \min_{\boldsymbol{U} \preceq \mathcal{B}} \mathbb{E}\big(j(\boldsymbol{U}, \boldsymbol{W})\big) , \quad \text{with} \quad \mathcal{B} = \sigma\big(h(\boldsymbol{W})\big) .$$

The aim of the scenario tree method is to

- approximate the noise $\boldsymbol{W}$ by a “finite” noise $\boldsymbol{W}_n$,
- and deduce the approximated information $\mathcal{B}_n = \sigma\big(h(\boldsymbol{W}_n)\big)$.

In this framework, an unique approximation is performed to obtain the approximated solution $\mathcal{V}(\boldsymbol{W}_n, \mathcal{B}_n)$, and it is possible to prove that $\mathcal{V}(\boldsymbol{W}_n, \mathcal{B}_n) \to \mathcal{V}(\boldsymbol{W}, \mathcal{B})$ (see [Pennanen, 2005]).

But the noise has been discretized in a very specific way. . .

(Very) Compact View of the Scenario Tree Approach

The stochastic optimal control problem under consideration depends on both a noise process $\boldsymbol{W}$ and a sequence of σ-fields $\mathcal{B}$. It can thus be represented under the compact form:

$$\mathcal{V}(\boldsymbol{W}, \mathcal{B}) = \min_{\boldsymbol{U} \preceq \mathcal{B}} \mathbb{E}\Big(j(\boldsymbol{U}, \boldsymbol{W})\Big) \;, \quad \text{with} \quad \mathcal{B} = \sigma\Big(h(\boldsymbol{W})\Big) \;.$$

The aim of the scenario tree method is to

- approximate the noise $\boldsymbol{W}$ by a “finite” noise $\boldsymbol{W}_n$,
- and deduce the approximated information $\mathcal{B}_n = \sigma\Big(h(\boldsymbol{W}_n)\Big)$.

In this framework, an unique approximation is performed to obtain the approximated solution $\mathcal{V}\big(\boldsymbol{W}_n, \mathcal{B}_n\big)$, and it is possible to prove that $\mathcal{V}\big(\boldsymbol{W}_n, \mathcal{B}_n\big) \to \mathcal{V}\big(\boldsymbol{W}, \mathcal{B}\big)$ (see [Pennanen, 2005]).

But the noise has been discretized in a very specific way…

(Very) Compact View of the Scenario Tree Approach

The stochastic optimal control problem under consideration depends on both a noise process $\boldsymbol{W}$ and a sequence of σ-fields $\mathcal{B}$. It can thus be represented under the compact form:

$$\mathcal{V}(\boldsymbol{W}, \mathcal{B}) = \min_{\boldsymbol{U} \preceq \mathcal{B}} \mathbb{E}\big(j(\boldsymbol{U}, \boldsymbol{W})\big) , \quad \text{with} \quad \mathcal{B} = \sigma\big(h(\boldsymbol{W})\big) .$$

The aim of the scenario tree method is to

- approximate the noise $\boldsymbol{W}$ by a "finite" noise $\boldsymbol{W}_n$,
- and deduce the approximated information $\mathcal{B}_n = \sigma\big(h(\boldsymbol{W}_n)\big)$.

In this framework, an unique approximation is performed to obtain the approximated solution $\mathcal{V}\big(\boldsymbol{W}_n, \mathcal{B}_n\big)$, and it is possible to prove that $\mathcal{V}\big(\boldsymbol{W}_n, \mathcal{B}_n\big) \rightarrow \mathcal{V}\big(\boldsymbol{W}, \mathcal{B}\big)$ (see [Pennanen, 2005]).

But the noise has been discretized in a very specific way...

A simple SOC problem

$$\min_{\boldsymbol{U} \preceq \boldsymbol{W}_0} \mathbb{E}\left(\varepsilon \boldsymbol{U}^2 + \left(\boldsymbol{W}_0 + \boldsymbol{U} + \boldsymbol{W}_1\right)^2\right)$$

- The noises $\boldsymbol{W}_0$ and $\boldsymbol{W}_1$ are independent random variables, each with a uniform probability distribution over $[-1, 1]$.
- The initial state is $\boldsymbol{X}_0 = \boldsymbol{W}_0$.
- The decision variable $\boldsymbol{U}$ is measurable w.r.t. $\boldsymbol{W}_0$: $\boldsymbol{U} \preceq \boldsymbol{W}_0$.
- The final state is $\boldsymbol{X}_1 = \boldsymbol{X}_0 + \boldsymbol{U} + \boldsymbol{W}_1$.

The goal is to minimize the expectation of $\left(\varepsilon \boldsymbol{U}^2 + \boldsymbol{X}_1^2\right)$, where ε is a “small” positive number (cheap control).

Note that this example matches the Markovian setting.

A simple SOC problem

$$\min_{\boldsymbol{U} \preceq \boldsymbol{W}_0} \mathbb{E}\left(\varepsilon \boldsymbol{U}^2 + (\boldsymbol{W}_0 + \boldsymbol{U} + \boldsymbol{W}_1)^2\right)$$

- The noises $\boldsymbol{W}_0$ and $\boldsymbol{W}_1$ are independent random variables, each with a uniform probability distribution over $[-1, 1]$.
- The initial state is $\boldsymbol{X}_0 = \boldsymbol{W}_0$.
- The decision variable $\boldsymbol{U}$ is measurable w.r.t. $\boldsymbol{W}_0$: $\boldsymbol{U} \preceq \boldsymbol{W}_0$.
- The final state is $\boldsymbol{X}_1 = \boldsymbol{X}_0 + \boldsymbol{U} + \boldsymbol{W}_1$.

The goal is to minimize the expectation of $\left(\varepsilon \boldsymbol{U}^2 + \boldsymbol{X}_1^2\right)$, where ε is a "small" positive number (cheap control).

Note that this example matches the Markovian setting.

Exact Solution of the Problem

$$\mathbb{E}\Big(\varepsilon \boldsymbol{U}^2 + (\boldsymbol{W}_0 + \boldsymbol{U} + \boldsymbol{W}_1)^2\Big) =$$

$$\mathbb{E}\Big(\underbrace{\boldsymbol{W}_0^2}_{1/3} + \underbrace{\boldsymbol{W}_1^2}_{1/3} + (1+\varepsilon)\boldsymbol{U}^2 + 2\boldsymbol{U}\boldsymbol{W}_0 + 2\underbrace{\boldsymbol{U}\boldsymbol{W}_1}_{0} + 2\underbrace{\boldsymbol{W}_0\boldsymbol{W}_1}_{0}\Big)$$

Exact Solution of the Problem

$$\mathbb{E}\Big(\varepsilon \boldsymbol{U}^2 + (\boldsymbol{W}_0 + \boldsymbol{U} + \boldsymbol{W}_1)^2\Big) =$$

$$\mathbb{E}\Big(\underbrace{\boldsymbol{W}_0^2}_{1/3} + \underbrace{\boldsymbol{W}_1^2}_{1/3} + (1+\varepsilon)\boldsymbol{U}^2 + 2\boldsymbol{U}\boldsymbol{W}_0 + 2\underbrace{\boldsymbol{U}\boldsymbol{W}_1}_{0} + 2\underbrace{\boldsymbol{W}_0\boldsymbol{W}_1}_{0}\Big)$$

- The problem is thus equivalent to

$$\min_{\boldsymbol{U} \preceq \boldsymbol{W}_0} \frac{2}{3} + \mathbb{E}\Big((1+\varepsilon)\boldsymbol{U}^2 + 2\boldsymbol{U}\boldsymbol{W}_0\Big) ,$$

- By the first order optimality condition, the optimal solution is

$$\boldsymbol{U}^\sharp = -\frac{\boldsymbol{W}_0}{1+\varepsilon} .$$

- The associated optimal cost is readily calculated to be

$$J^\sharp = \frac{1}{3} \times \frac{1+2\varepsilon}{1+\varepsilon} = \frac{1}{3} + O(\varepsilon) .$$

Exact Solution of the Problem

$$\mathbb{E}\Big(\varepsilon \boldsymbol{U}^2 + (\boldsymbol{W}_0 + \boldsymbol{U} + \boldsymbol{W}_1)^2\Big) =$$

$$\mathbb{E}\Big(\underbrace{\boldsymbol{W}_0^2}_{1/3} + \underbrace{\boldsymbol{W}_1^2}_{1/3} + (1+\varepsilon)\boldsymbol{U}^2 + 2\boldsymbol{U}\boldsymbol{W}_0 + 2\underbrace{\boldsymbol{U}\boldsymbol{W}_1}_{0} + 2\underbrace{\boldsymbol{W}_0\boldsymbol{W}_1}_{0}\Big)$$

- The problem is thus equivalent to

$$\min_{\boldsymbol{U} \preceq \boldsymbol{W}_0} \frac{2}{3} + \mathbb{E}\Big((1+\varepsilon)\boldsymbol{U}^2 + 2\boldsymbol{U}\boldsymbol{W}_0\Big) ,$$

- By the first order optimality condition, the optimal solution is

$$\boldsymbol{U}^\sharp = -\frac{\boldsymbol{W}_0}{1+\varepsilon} .$$

- The associated optimal cost is readily calculated to be

$$J^\sharp = \frac{1}{3} \times \frac{1+2\varepsilon}{1+\varepsilon} = \frac{1}{3} + O(\varepsilon) .$$

Exact Solution of the Problem

$$\mathbb{E}\Big(\varepsilon \boldsymbol{U}^2 + (\boldsymbol{W}_0 + \boldsymbol{U} + \boldsymbol{W}_1)^2\Big) =$$

$$\mathbb{E}\Big(\underbrace{\boldsymbol{W}_0^2}_{1/3} + \underbrace{\boldsymbol{W}_1^2}_{1/3} + (1+\varepsilon)\boldsymbol{U}^2 + 2\boldsymbol{U}\boldsymbol{W}_0 + 2\underbrace{\boldsymbol{U}\boldsymbol{W}_1}_{0} + 2\underbrace{\boldsymbol{W}_0\boldsymbol{W}_1}_{0}\Big)$$

- The problem is thus equivalent to

$$\min_{\boldsymbol{U} \preceq \boldsymbol{W}_0} \frac{2}{3} + \mathbb{E}\Big((1+\varepsilon)\boldsymbol{U}^2 + 2\boldsymbol{U}\boldsymbol{W}_0\Big) ,$$

- By the first order optimality condition, the optimal solution is

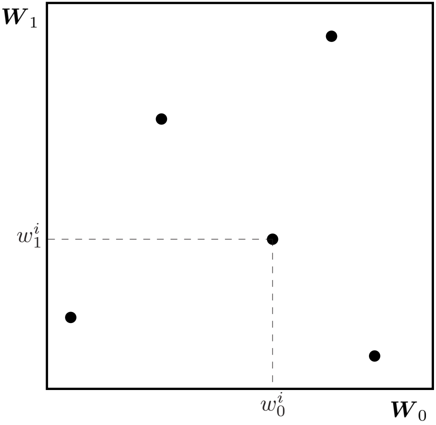
$$\boldsymbol{U}^\sharp = -\frac{\boldsymbol{W}_0}{1+\varepsilon} .$$

- The associated optimal cost is readily calculated to be

$$J^\sharp = \frac{1}{3} \times \frac{1+2\varepsilon}{1+\varepsilon} = \frac{1}{3} + \mathrm{O}(\varepsilon) .$$

Noise Discretization “a la Monte Carlo”

We crudely sample the optimization problem.



- To that purpose, we first consider a realization of a N-sample of the couple $(\boldsymbol{W}_0, \boldsymbol{W}_1)$, that is, points in the square $\Omega = [-1, 1]^2$:

$$\left\{(w_0^i, w_1^i)\right\}_{i=1,\ldots,N}$$

- This sample will be used to approximate the expectation by the Monte Carlo method.

Discretized Information Structure

- We consider the N realizations $\{u^i\}_{i=1,\ldots,N}$ of the decision variable $\boldsymbol{U}$, corresponding to the discretization of the noise,
- and we have to keep in mind that $\boldsymbol{U}$ should be measurable w.r.t. the first component $\boldsymbol{W}_0$ of the noise $(\boldsymbol{W}_0, \boldsymbol{W}_1)$:

$$\boldsymbol{U} \preceq \boldsymbol{W}_0 \ .$$

- To translate this condition in our discrete framework issued from a Monte Carlo sample, we impose the constraint

$$\forall (i,j) \in \{1, \ldots, N\}^2 \ , \ \ w_0^i = w_0^j \implies u^i = u^j \ ,$$

which prevents $\boldsymbol{U}$ from taking different values whenever two samples of the noise display the same value on the first component (corresponding to $\boldsymbol{W}_0$).

Discretized Information Structure

- We consider the N realizations $\{u^i\}_{i=1,\ldots,N}$ of the decision variable $\boldsymbol{U}$, corresponding to the discretization of the noise,
- and we have to keep in mind that $\boldsymbol{U}$ should be measurable w.r.t. the first component $\boldsymbol{W}_0$ of the noise $(\boldsymbol{W}_0, \boldsymbol{W}_1)$:

$$\boldsymbol{U} \preceq \boldsymbol{W}_0 \ .$$

- To translate this condition in our discrete framework issued from a Monte Carlo sample, we impose the constraint

$$\forall (i,j) \in \{1,\ldots,N\}^2 \ , \ \ w_0^i = w_0^j \implies u^i = u^j \ ,$$

which prevents $\boldsymbol{U}$ from taking different values whenever two samples of the noise display the same value on the first component (corresponding to $\boldsymbol{W}_0$).

Discretized Information Structure

- We consider the N realizations $\{u^i\}_{i=1,\dots,N}$ of the decision variable $\boldsymbol{U}$, corresponding to the discretization of the noise,
- and we have to keep in mind that $\boldsymbol{U}$ should be measurable w.r.t. the first component $\boldsymbol{W}_0$ of the noise $(\boldsymbol{W}_0, \boldsymbol{W}_1)$:

$$\boldsymbol{U} \preceq \boldsymbol{W}_0 \ .$$

- To translate this condition in our discrete framework issued from a Monte Carlo sample, we impose the constraint

$$\forall (i,j) \in \{1,\dots,N\}^2 \ , \ \ w_0^i = w_0^j \implies u^i = u^j \ ,$$

which prevents $\boldsymbol{U}$ from taking different values whenever two samples of the noise display the same value on the first component (corresponding to $\boldsymbol{W}_0$).

The Measurability Constraint is Not Effective!

The expression of the cost after discretization is

$$\frac{1}{N}\left(\sum_{i=1}^{N} \varepsilon\big(u^i\big)^2 + \big(w_0^i + u^i + w_1^i\big)^2\right) ,$$

and it is minimized w.r.t. $(u^1, \ldots, u^N)$ under the constraints

$$u^i = u^j \quad \text{whenever} \quad w_0^i = w_0^j .$$

Since the N sample trajectories (w_0^i, w_1^i) of $(\mathbf{W}_0, \mathbf{W}_1)$ are produced by a Monte Carlo sampling over $[-1, 1]^2$, then, with probability 1,

$$w_0^i \neq w_0^j \qquad \forall (i, j) \quad \text{such that} \quad i \neq j .$$

The constraints are in fact never effective, so that the discretized cost can be minimized independently for each individual sample i.

The Measurability Constraint is Not Effective!

The expression of the cost after discretization is

$$\frac{1}{N}\left(\sum_{i=1}^{N} \varepsilon\big(u^i\big)^2 + \big(w_0^i + u^i + w_1^i\big)^2\right) ,$$

and it is minimized w.r.t. $(u^1, \ldots, u^N)$ under the constraints

$$u^i = u^j \quad \text{whenever} \quad w_0^i = w_0^j \; .$$

Since the N sample trajectories (w_0^i, w_1^i) of $(\boldsymbol{W}_0, \boldsymbol{W}_1)$ are produced by a Monte Carlo sampling over $[-1, 1]^2$, then, with probability 1,

$$w_0^i \neq w_0^j \qquad \forall (i, j) \quad \text{such that} \quad i \neq j \; .$$

The constraints are in fact never effective, so that the discretized cost can be minimized independently for each individual sample i.

The Measurability Constraint is Not Effective!

The expression of the cost after discretization is

$$\frac{1}{N}\left(\sum_{i=1}^{N} \varepsilon\left(u^i\right)^2 + \left(w_0^i + u^i + w_1^i\right)^2\right) ,$$

and it is minimized w.r.t. $(u^1, \dots, u^N)$ under the constraints

$$u^i = u^j \quad \text{whenever} \quad w_0^i = w_0^j \; .$$

Since the N sample trajectories (w_0^i, w_1^i) of $(\boldsymbol{W}_0, \boldsymbol{W}_1)$ are produced by a Monte Carlo sampling over $[-1, 1]^2$, then, with probability 1,

$$w_0^i \neq w_0^j \qquad \forall (i, j) \quad \text{such that} \quad i \neq j \; .$$

The constraints are in fact never effective, so that the discretized cost can be minimized independently for each individual sample i.

Something is Wrong...

The optimization problem associated to the i-th sample is

$$\min_{u_i \in \mathbb{R}} \varepsilon \big(u^i\big)^2 + \big(w_0^i + u^i + w_1^i\big)^2 ,$$

which yields the optimal value and the optimal cost

$$u_\flat^i = -\frac{w_0^i + w_1^i}{1+\varepsilon} \quad , \quad j_\flat^i = \varepsilon \frac{(w_0^i + w_1^i)^2}{1+\varepsilon} .$$

The averaged cost over the N samples is equal to

$$\frac{1}{N}\sum_{i=1}^{N} \frac{\varepsilon\big(w_0^i + w_1^i\big)^2}{1+\varepsilon} \underset{N\to+\infty}{\longrightarrow} \frac{2\,\varepsilon}{3\,(1+\varepsilon)} = 0 + O(\varepsilon) .$$

This cost is far below the true optimal cost $J^\sharp = 1/3 + O(\varepsilon)$!
However, any admissible solution (any $\mathbf{U}$ such that $\mathbf{U} \preceq \mathbf{W}_0$) cannot achieve a cost better than the optimal cost $J^\sharp$...

Something is Wrong. . .

The optimization problem associated to the i-th sample is

$$\min_{u_i \in \mathbb{R}} \varepsilon \big(u^i \big)^2 + \big(w_0^i + u^i + w_1^i \big)^2 ,$$

which yields the optimal value and the optimal cost

$$u_\flat^i = -\frac{w_0^i + w_1^i}{1+\varepsilon} \quad , \quad j_\flat^i = \varepsilon \frac{(w_0^i + w_1^i)^2}{1+\varepsilon} .$$

The averaged cost over the N samples is equal to

$$\frac{1}{N} \sum_{i=1}^{N} \frac{\varepsilon (w_0^i + w_1^i)^2}{1+\varepsilon} \underset{N \to +\infty}{\longrightarrow} \frac{2\,\varepsilon}{3\,(1+\varepsilon)} = 0 + \mathrm{O}(\varepsilon) .$$

This cost is far below the true optimal cost $J^\sharp = 1/3 + \mathrm{O}(\varepsilon)$!
However, any admissible solution (any $\boldsymbol{U}$ such that $\boldsymbol{U} \preceq \boldsymbol{W}_0$) cannot achieve a cost better than the optimal cost $J^\sharp$ . . .

Something is Wrong...

The optimization problem associated to the i-th sample is

$$\min_{u_i \in \mathbb{R}} \varepsilon \big(u^i\big)^2 + \big(w_0^i + u^i + w_1^i\big)^2 ,$$

which yields the optimal value and the optimal cost

$$u_\flat^i = -\frac{w_0^i + w_1^i}{1+\varepsilon} \quad , \quad j_\flat^i = \varepsilon \frac{(w_0^i + w_1^i)^2}{1+\varepsilon} .$$

The averaged cost over the N samples is equal to

$$\frac{1}{N} \sum_{i=1}^{N} \frac{\varepsilon (w_0^i + w_1^i)^2}{1+\varepsilon} \underset{N \to +\infty}{\longrightarrow} \frac{2\,\varepsilon}{3\,(1+\varepsilon)} = 0 + \mathrm{O}(\varepsilon) .$$

This cost is far below the true optimal cost $J^\sharp = 1/3 + \mathrm{O}(\varepsilon)$!
However, any admissible solution (any $\boldsymbol{U}$ such that $\boldsymbol{U} \preceq \boldsymbol{W}_0$) cannot achieve a cost better than the optimal cost $J^\sharp$...

Real Value of the Discretized Problem Solution

- The resolution of the discretized problem derived from the Monte Carlo procedure yields N optimal values

$$u_\flat^i = -\frac{w_0^i + w_1^i}{1+\varepsilon} ,$$

  but not a random variable. The associated cost value of order ε is just a fake cost estimation, because we have not produced an admissible control for the initial problem, namely a random variable $\boldsymbol{U}_\flat$ measurable with respect to $\boldsymbol{W}_0$.

- To evaluate the true cost of this naive approach, we must first derive an admissible control for the initial problem, that is, a random variable $\boldsymbol{U}_\flat$ over $[-1,1]^2$ with constant value along every vertical line of this square (since the horizontal axis corresponds to the first component $\boldsymbol{W}_0$ of the noise).

Real Value of the Discretized Problem Solution

- The resolution of the discretized problem derived from the Monte Carlo procedure yields N optimal values

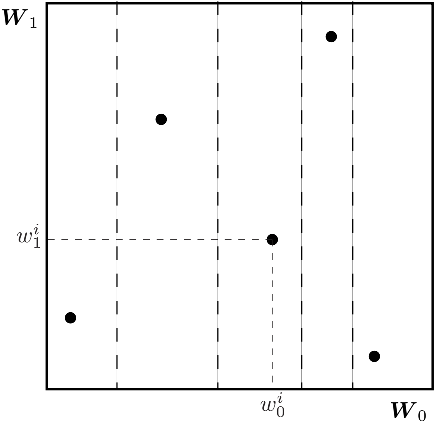
$$u_\flat^i = -\frac{w_0^i + w_1^i}{1+\varepsilon} ,$$

  but not a random variable. The associated cost value of order ε is just a fake cost estimation, because we have not produced an admissible control for the initial problem, namely a random variable $\boldsymbol{U}_\flat$ measurable with respect to $\boldsymbol{W}_0$.

- To evaluate the true cost of this naive approach, we must first derive an admissible control for the initial problem, that is, a random variable $\boldsymbol{U}_\flat$ over $[-1,1]^2$ with constant value along every vertical line of this square (since the horizontal axis corresponds to the first component $\boldsymbol{W}_0$ of the noise).

Construction of an Admissible Control (1)

We assume that the sample points have been renumbered so that the value of the sample first component w_0^i is increasing with i.



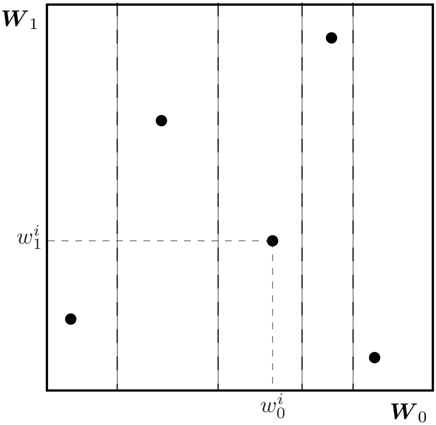
- Divide the square into N vertical strips by drawing vertical lines in the middle of segments $[w_0^i, w_0^{i+1}]$.
- The i-th strip is given by $[a^{i-1}, a^i] \times [-1, 1]$, with:

$$a^i = (w_0^i + w_0^{i+1})/2 ,$$

for $i = 2, \dots, N-1$, ($a^0 = -1$ and $a^N = 1$).

Construction of an Admissible Control (1)

We assume that the sample points have been renumbered so that the value of the sample first component w_0^i is increasing with i.



- Divide the square into N vertical strips by drawing vertical lines in the middle of segments $[w_0^i, w_0^{i+1}]$.
- The i-th strip is given by $[a^{i-1}, a^i] \times [-1, 1]$, with:

$$a^i = (w_0^i + w_0^{i+1})/2 \ ,$$

for $i = 2, \ldots, N-1$, ($a^0 = -1$ and $a^N = 1$).

Construction of an Admissible Control (2)

We construct a solution $\boldsymbol{U}_\flat$ as the function of (w_0, w_1) which is constant over each vertical strip defined on the square, the value of $\boldsymbol{U}_\flat$ in strip i being equal to the optimal value $u_\flat^i = -\frac{w_0^i + w_1^i}{1+\varepsilon}$:

$$\boldsymbol{U}_\flat(w_0, w_1) = \sum_{i=1}^{N} u_\flat^i \, \mathbf{1}_{[a^{i-1}, a^i] \times [-1,1]}(w_0, w_1) \ ,$$

where (w_0, w_1) ranges in the square $[-1, 1]^2$ and where $\mathbf{1}_A(\cdot)$ is the indicator function of the set A:

$$\mathbf{1}_A(x) = \begin{cases} 1 & \text{if } x \in A \ , \\ 0 & \text{otherwise} \ . \end{cases}$$

Note that the control $\boldsymbol{U}_\flat$ depends on the N samples (w_0^i, w_0^i) by means of the values of the mid-points a^i's and of the controls $u_\flat^i$'s.

Construction of an Admissible Control (2)

We construct a solution $\boldsymbol{U}_\flat$ as the function of (w_0, w_1) which is constant over each vertical strip defined on the square, the value of $\boldsymbol{U}_\flat$ in strip i being equal to the optimal value $u_\flat^i = -\frac{w_0^i + w_1^i}{1+\varepsilon}$:

$$\boldsymbol{U}_\flat(w_0, w_1) = \sum_{i=1}^{N} u_\flat^i \, \mathbf{1}_{[a^{i-1}, a^i] \times [-1,1]}(w_0, w_1) \ ,$$

where (w_0, w_1) ranges in the square $[-1, 1]^2$ and where $\mathbf{1}_A(\cdot)$ is the indicator function of the set A:

$$\mathbf{1}_A(x) = \begin{cases} 1 & \text{if } x \in A \ , \\ 0 & \text{otherwise} \ . \end{cases}$$

Note that the control $\boldsymbol{U}_\flat$ depends on the N samples (w_0^i, w_0^i) by means of the values of the mid-points a^i's and of the controls $u_\flat^i$'s.

Evaluation of the Expected Cost (1)

The corresponding cost value $\mathbb{E}\big(\varepsilon(\boldsymbol{U}_\flat)^2 + (\boldsymbol{W}_0 + \boldsymbol{U}_\flat + \boldsymbol{W}_1)^2\big)$ can be evaluated analytically (integration w.r.t. (w_0, w_1) over the square $[-1, 1]^2$), and is equal to

$$\frac{2}{3} + \sum_{i=1}^{N} \left((1+\varepsilon) \frac{a^i - a^{i-1}}{2} (u_\flat^i)^2 + \frac{(a^i)^2 - (a^{i-1})^2}{2} u_\flat^i \right) ,$$

where the values a^i and $u_\flat^i$ depend on the samples (w_0^i, w_1^i).

In order to assess the value of this estimate, we now compute its expectation when considering that the (w_0^i, w_1^i)'s are realizations of independent random variables $(\boldsymbol{W}_0^i, \boldsymbol{W}_1^i)$. This calculation is not straightforward because the w_0^i's have been reordered, so that we compute it numerically for different values of N.

Evaluation of the Expected Cost (1)

The corresponding cost value $\mathbb{E}\big(\varepsilon(\boldsymbol{U}_\flat)^2 + (\boldsymbol{W}_0 + \boldsymbol{U}_\flat + \boldsymbol{W}_1)^2\big)$ can be evaluated analytically (integration w.r.t. (w_0, w_1) over the square $[-1, 1]^2$), and is equal to

$$\frac{2}{3} + \sum_{i=1}^{N} \left((1+\varepsilon) \frac{a^i - a^{i-1}}{2} (u_\flat^i)^2 + \frac{(a^i)^2 - (a^{i-1})^2}{2} u_\flat^i \right) ,$$

where the values a^i and $u_\flat^i$ depend on the samples (w_0^i, w_1^i).

In order to assess the value of this estimate, we now compute its expectation when considering that the (w_0^i, w_1^i)'s are realizations of independent random variables $(\boldsymbol{W}_0^i, \boldsymbol{W}_1^i)$. This calculation is not straightforward because the w_0^i's have been reordered, so that we compute it numerically for different values of N.

Evaluation of the Expected Cost (2)

The cost provided by the admissible control $\boldsymbol{U}_\flat$ is estimated $2/3$.

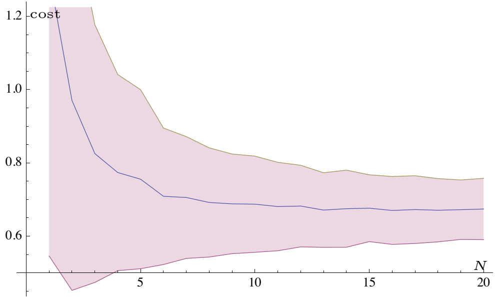


Figure: Estimated cost as a function of the number N of samples

This value **neither** corresponds to the true optimal cost $(1/3)$ **nor** to the cost of the discrete problem (0). **In fact, the value** $2/3$ **is equal to the one given by the best** open-loop control: $\boldsymbol{U}_\star \equiv 0$!

Evaluation of the Expected Cost (2)

The cost provided by the admissible control $\boldsymbol{U}_\flat$ is estimated $2/3$.

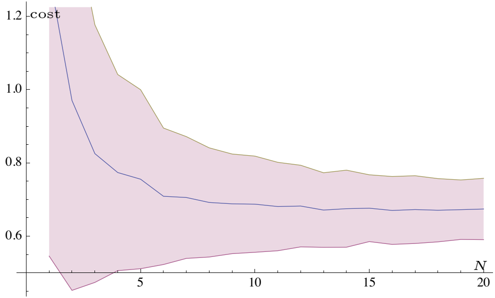
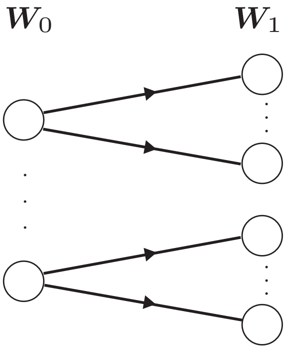


Figure: Estimated cost as a function of the number N of samples

This value **neither** corresponds to the true optimal cost $(1/3)$ **nor** to the cost of the discrete problem (0). In fact, the value $2/3$ is equal to the one given by the best open-loop control: $\boldsymbol{U}_\star \equiv 0$!

Scenario Tree Approach



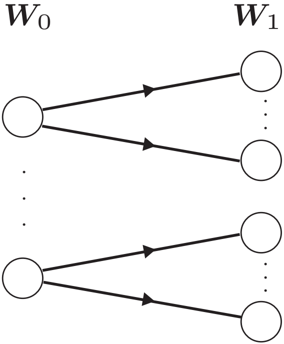
The scenario tree approach leads to $N_0 \times N_1$ scenarios:

$$\left\{\left(w_0^j, w_1^{jk}\right)\right\}_{j=1,\ldots,N_0}^{k=1,\ldots,N_1} .$$

- Notice that the discretization w_0^j of the first noise W_0 only depends on $j = 1, \ldots, N_0$,
- whereas the discretization w_1^{jk} of the noise W_1 "hangs" from a given j and depends on $k = 1, \ldots, N_1$.

From the measurability constraint, a different value u^j of the control U is associated to each value w_0^j of W_0.

Scenario Tree Approach



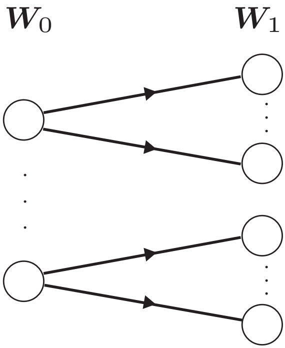
The scenario tree approach leads to $N_0 \times N_1$ scenarios:

$$\left\{\left(w_0^j, w_1^{jk}\right)\right\}_{j=1,\ldots,N_0}^{k=1,\ldots,N_1} .$$

- Notice that the discretization w_0^j of the first noise $\boldsymbol{W}_0$ only depends on $j = 1, \ldots, N_0$,
- whereas the discretization w_1^{jk} of the noise $\boldsymbol{W}_1$ "hangs" from a given j and depends on $k = 1, \ldots, N_1$.

From the measurability constraint, a different value u^j of the control $\boldsymbol{U}$ is associated to each value w_0^j of $\boldsymbol{W}_0$.

Scenario Tree Approach



The scenario tree approach leads to $N_0 \times N_1$ scenarios:

$$\left\{ \left(w_0^j, w_1^{jk} \right) \right\}_{j=1,\ldots,N_0}^{k=1,\ldots,N_1} .$$

- Notice that the discretization w_0^j of the first noise $\boldsymbol{W}_0$ only depends on $j = 1, \ldots, N_0$,
- whereas the discretization w_1^{jk} of the noise $\boldsymbol{W}_1$ "hangs" from a given j and depends on $k = 1, \ldots, N_1$.

From the measurability constraint, a different value u^j of the control $\boldsymbol{U}$ is associated to each value w_0^j of $\boldsymbol{W}_0$.

Scenario Tree Optimal Solution

On the scenario tree, the original cost $\mathbb{E}\big(\varepsilon \boldsymbol{U}^2 + (\boldsymbol{W}_0 + \boldsymbol{U} + \boldsymbol{W}_1)^2\big)$ is approximated by

$$\frac{1}{N_0}\sum_{j=1}^{N_0}\Big(\varepsilon (u^j)^2 + \frac{1}{N_1}\sum_{k=1}^{N_1}(u^j + w_0^j + w_1^{jk})^2\Big)$$

The solution of this approximated problem is

$$u_0^j = -\frac{w_0^j + \overline{w}_1^j}{1+\varepsilon}, \quad \text{where} \quad \overline{w}_1^j = \frac{1}{N_1}\sum_{k=1}^{N_1} w_1^{jk},$$

to be compared with the naive Monte Carlo solution $u_0^j = -\dfrac{w_0^j + w_1^j}{1+\varepsilon}$.

Note that $\overline{w}_1^j$ is an estimate of the expectation $\mathbb{E}(\boldsymbol{W}_1)$, since we assumed that $\boldsymbol{W}_0$ and $\boldsymbol{W}_1$ are independent random variables.

Scenario Tree Optimal Solution

On the scenario tree, the original cost $\mathbb{E}\big(\varepsilon \boldsymbol{U}^2 + (\boldsymbol{W}_0 + \boldsymbol{U} + \boldsymbol{W}_1)^2\big)$ is approximated by

$$\frac{1}{N_0}\sum_{j=1}^{N_0}\Big(\varepsilon(u^j)^2 + \frac{1}{N_1}\sum_{k=1}^{N_1}(u^j + w_0^j + w_1^{jk})^2\Big)$$

The solution of this approximated problem is

$$u_\natural^j = -\frac{w_0^j + \overline{w}_1^j}{1+\varepsilon} \; , \quad \text{where} \;\; \overline{w}_1^j = \frac{1}{N_1}\sum_{k=1}^{N_1} w_1^{jk} \; ,$$

to be compared with the naive Monte Carlo solution $u_\flat^j = -\dfrac{w_0^j + w_1^j}{1+\varepsilon}$.

Note that $\overline{w}_1^j$ is an estimate of the expectation $\mathbb{E}\big(\boldsymbol{W}_1\big)$, since we assumed that $\boldsymbol{W}_0$ and $\boldsymbol{W}_1$ are independent random variables.

Scenario Tree Optimal Solution

On the scenario tree, the original cost $\mathbb{E}\big(\varepsilon \boldsymbol{U}^2 + (\boldsymbol{W}_0 + \boldsymbol{U} + \boldsymbol{W}_1)^2\big)$ is approximated by

$$\frac{1}{N_0}\sum_{j=1}^{N_0}\Big(\varepsilon(u^j)^2 + \frac{1}{N_1}\sum_{k=1}^{N_1}(u^j + w_0^j + w_1^{jk})^2\Big)$$

The solution of this approximated problem is

$$u_\natural^j = -\frac{w_0^j + \overline{w}_1^j}{1+\varepsilon} \;, \quad \text{where } \; \overline{w}_1^j = \frac{1}{N_1}\sum_{k=1}^{N_1} w_1^{jk} \;,$$

to be compared with the naive Monte Carlo solution $u_\flat^j = -\dfrac{w_0^j + w_1^j}{1+\varepsilon}$.

Note that $\overline{w}_1^j$ is an estimate of the expectation $\mathbb{E}\big(\boldsymbol{W}_1\big)$, since we assumed that $\boldsymbol{W}_0$ and $\boldsymbol{W}_1$ are independent random variables.

Scenario Tree Optimal Cost

Let $(\overline{\sigma}_1^j)^2 = \frac{1}{N_1}\sum_{k=1}^{N_1}(w_1^{jk})^2$. The solution $u_\natural^j$ yields the cost

$$\frac{1}{N_0(1+\varepsilon)}\sum_{j=1}^{N_0}\Big(\varepsilon(w_0^j)^2 + 2\varepsilon w_0^j\overline{w}_1^j - (\overline{w}_1^j)^2 + (1+\varepsilon)(\overline{\sigma}_1^j)^2\Big) \; .$$

The two estimates $\overline{w}_1^j$ and $(\overline{\sigma}_1^j)^2$ converge as N_1 goes to infinity towards their asymptotic values, that is, 0 and 1/3, so that the scenario tree optimal cost is such that

$$\frac{1}{N_0(1+\varepsilon)}\sum_{j=1}^{N_0}\Big(\varepsilon(w_0^j)^2 + \frac{1+\varepsilon}{3}\Big) \underset{N_0\to+\infty}{\longrightarrow} \frac{1}{3} + O(\varepsilon) \; .$$

This cost is of the same order than the "true" optimal cost! However it does not correspond to an admissible solution...

Scenario Tree Optimal Cost

Let $(\overline{\sigma}_1^j)^2 = \frac{1}{N_1}\sum_{k=1}^{N_1}(w_1^{jk})^2$. The solution $u_\natural^j$ yields the cost

$$\frac{1}{N_0(1+\varepsilon)}\sum_{j=1}^{N_0}\left(\varepsilon(w_0^j)^2 + 2\varepsilon w_0^j\overline{w}_1^j - (\overline{w}_1^j)^2 + (1+\varepsilon)(\overline{\sigma}_1^j)^2\right) .$$

The two estimates $\overline{w}_1^j$ and $(\overline{\sigma}_1^j)^2$ converge as N_1 goes to infinity towards their asymptotic values, that is, 0 and $1/3$, so that the scenario tree optimal cost is such that

$$\frac{1}{N_0(1+\varepsilon)}\sum_{j=1}^{N_0}\left(\varepsilon(w_0^j)^2 + \frac{1+\varepsilon}{3}\right) \underset{N_0\to+\infty}{\longrightarrow} \frac{1}{3} + \mathrm{O}(\varepsilon) .$$

This cost is of the same order than the "true" optimal cost!
However it does not correspond to an admissible solution. . .

Scenario Tree Optimal Cost

Let $(\overline{\sigma}_1^j)^2 = \frac{1}{N_1}\sum_{k=1}^{N_1}(w_1^{jk})^2$. The solution $u^j_\natural$ yields the cost

$$\frac{1}{N_0(1+\varepsilon)}\sum_{j=1}^{N_0}\left(\varepsilon(w_0^j)^2 + 2\varepsilon w_0^j\overline{w}_1^j - (\overline{w}_1^j)^2 + (1+\varepsilon)(\overline{\sigma}_1^j)^2\right) .$$

The two estimates $\overline{w}_1^j$ and $(\overline{\sigma}_1^j)^2$ converge as N_1 goes to infinity towards their asymptotic values, that is, 0 and $1/3$, so that the scenario tree optimal cost is such that

$$\frac{1}{N_0(1+\varepsilon)}\sum_{j=1}^{N_0}\left(\varepsilon(w_0^j)^2 + \frac{1+\varepsilon}{3}\right) \underset{N_0\to+\infty}{\longrightarrow} \frac{1}{3} + \mathrm{O}(\varepsilon) .$$

This cost is of the same order than the "true" optimal cost!

However it does not correspond to an admissible solution. . .

Scenario Tree Optimal Cost

Let $(\overline{\sigma}_1^j)^2 = \frac{1}{N_1}\sum_{k=1}^{N_1}(w_1^{jk})^2$. The solution $u_\natural^j$ yields the cost

$$\frac{1}{N_0(1+\varepsilon)}\sum_{j=1}^{N_0}\left(\varepsilon(w_0^j)^2 + 2\varepsilon w_0^j\overline{w}_1^j - (\overline{w}_1^j)^2 + (1+\varepsilon)(\overline{\sigma}_1^j)^2\right) .$$

The two estimates $\overline{w}_1^j$ and $(\overline{\sigma}_1^j)^2$ converge as N_1 goes to infinity towards their asymptotic values, that is, 0 and $1/3$, so that the scenario tree optimal cost is such that

$$\frac{1}{N_0(1+\varepsilon)}\sum_{j=1}^{N_0}\left(\varepsilon(w_0^j)^2 + \frac{1+\varepsilon}{3}\right) \underset{N_0\to+\infty}{\longrightarrow} \frac{1}{3} + \mathrm{O}(\varepsilon) .$$

This cost is of the same order than the “true” optimal cost!
However it does not correspond to an admissible solution. . .

Admissible Control and Associated Cost

As in the naive Monte Carlo method, we derive from the $u_\natural^j$'s an admissible solution $\boldsymbol{U}_\natural$ for the initial problem (piecewise constant fonction over N_0 strips of the square $[-1, 1]^2$). The cost provided by $\boldsymbol{U}_\natural$ is estimated $1/3$, corresponding to the true optimal cost.

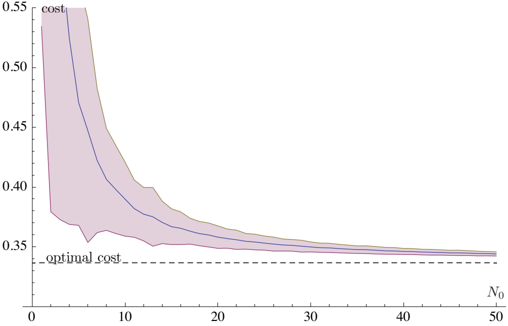


Figure: Estimated cost on a tree with N_0^2 scenarios

Where Do We Stand?

	True Solution	**Naive Monte Carlo**	**Scenario Tree**
Discrete Cost		$\mathrm{O}(\varepsilon)$	$1/3 + \mathrm{O}(\varepsilon)$
Optimal Control	$-\boldsymbol{W}_0/(1+\varepsilon)$	$-(w_0^i + w_1^i)/(1+\varepsilon)$	$-(w_0^i + \overline{w}_1^j)/(1+\varepsilon)$
Induced Cost	$1/3 + \mathrm{O}(\varepsilon)$	$2/3 + \mathrm{O}(\varepsilon)$	$1/3 + \mathrm{O}(\varepsilon)$

1. The naive Monte Carlo method
   - discretizes the noise process as a whole,
   - deduces the discretization of the measurability constraint,
   - yields a cost not better than the open-loop solution. . .
2. The scenario tree approach
   - discretizes the noise in a clever way (forward process),
   - deduces the discretization of the measurability constraint,
   - yields the optimal cost!

Clue: *the conditional probability laws are well estimated.*

Where Do We Stand?

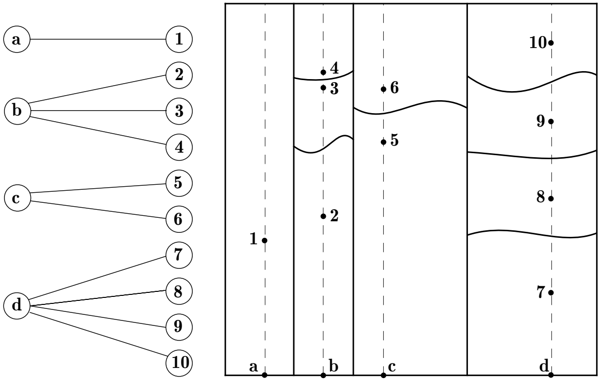
	True Solution	Naive Monte Carlo	Scenario Tree
Discrete Cost		$\mathrm{O}(\varepsilon)$	$1/3 + \mathrm{O}(\varepsilon)$
Optimal Control	$-\boldsymbol{W}_0/(1+\varepsilon)$	$-(w_0^i + w_1^i)/(1+\varepsilon)$	$-(w_0^i + \overline{w}_1^i)/(1+\varepsilon)$
Induced Cost	$1/3 + \mathrm{O}(\varepsilon)$	$2/3 + \mathrm{O}(\varepsilon)$	$1/3 + \mathrm{O}(\varepsilon)$

1. The naive Monte Carlo method
   - discretizes the noise process as a whole,
   - deduces the discretization of the measurability constraint,
   - yields a cost not better than the open-loop solution...
2. The scenario tree approach
   - discretizes the noise in a clever way (forward process),
   - deduces the discretization of the measurability constraint,
   - yields the optimal cost!

Clue: *the conditional probability laws are well estimated.*

Monte Carlo Interpretation of the Scenario Tree

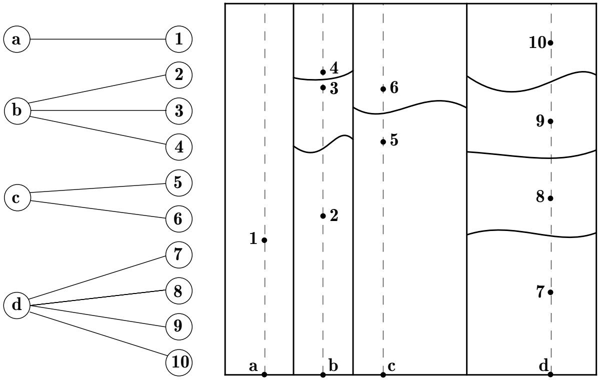
$$\boldsymbol{W}_0 \rightsquigarrow \{\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{d}\},\ \boldsymbol{W}_1 \rightsquigarrow \big\{\{\mathbf{1}\}, \{\mathbf{2}, \mathbf{3}, \mathbf{4}\}, \{\mathbf{5}, \mathbf{6}\}, \{\mathbf{7}, \mathbf{8}, \mathbf{9}, \mathbf{10}\}\big\}.$$



In a scenario tree, groups of samples are naturally aligned vertically!

Monte Carlo Interpretation of the Scenario Tree

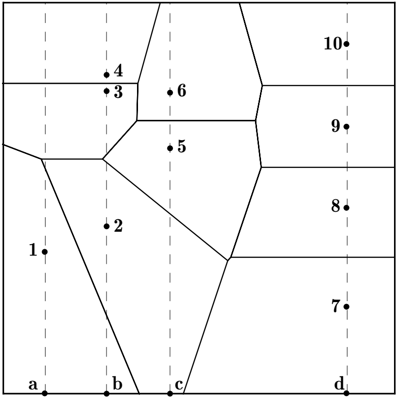
$$\boldsymbol{W}_0 \rightsquigarrow \{a, b, c, d\}, \ \boldsymbol{W}_1 \rightsquigarrow \{\{1\}, \{2, 3, 4\}, \{5, 6\}, \{7, 8, 9, 10\}\}.$$



In a scenario tree, groups of samples are naturally aligned vertically!

Voronoi Quantization

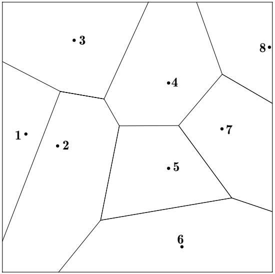
However, others quantizations of Ω are possible.



Given a set of points in the square $[-1, 1]^2$, the Voronoi tessellation minimizes the mean quadratic error among finite random variables taking given values. We in fact consider a discretized version of the random variable $(\boldsymbol{W}_0, \boldsymbol{W}_1)$, rather than a Monte Carlo sampling.

Independent Discretization of Noise and Information (1)

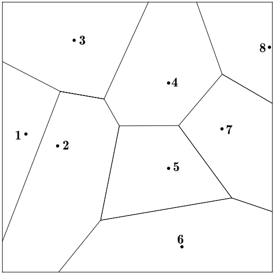
- Choose a discretization of the noise (8 cells).



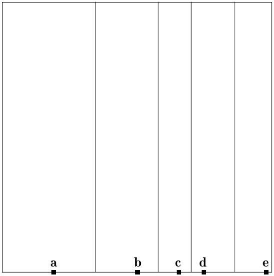
Noise

Independent Discretization of Noise and Information (2)

- Choose a discretization of the noise (8 cells).
- Choose a discretization of the information (5 cells).



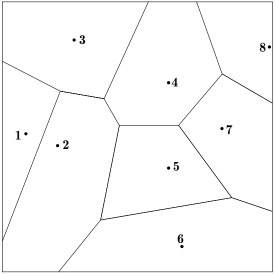
Noise



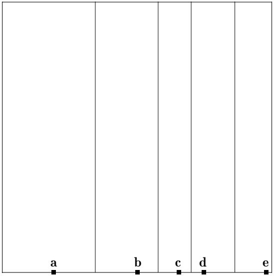
Information

Independent Discretization of Noise and Information (3)

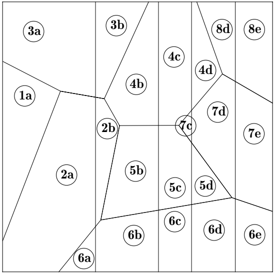
- Choose a discretization of the noise (8 cells).
- Choose a discretization of the information (5 cells).
- Combine both discretizations (21 non empty cells).



Noise



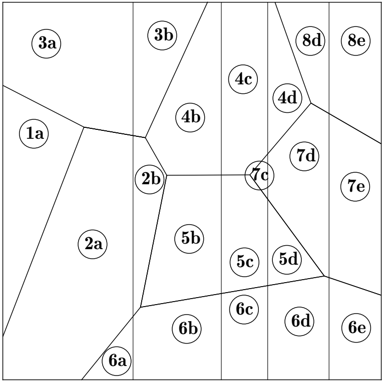
Information

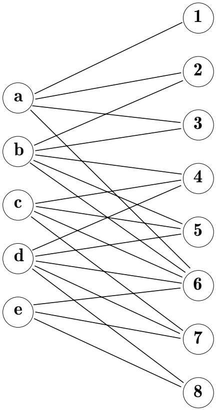


Mixing

Independent Discretization of Noise and Information (4)

$W_0 \rightsquigarrow \{a, b, c, d, e\}$, $W_1 \rightsquigarrow \{1, 2, 3, 4, 5, 6, 7, 8\}$.

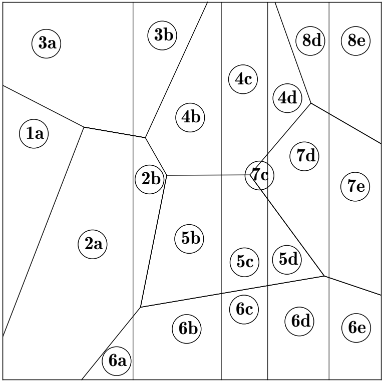


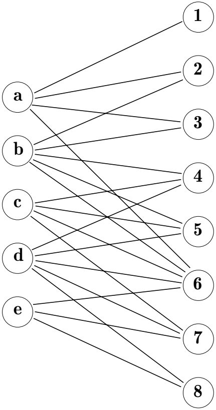


This approach does not necessarily produce a tree structure!

Independent Discretization of Noise and Information (4)

$\boldsymbol{W_0} \rightsquigarrow \{\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{d}, \boldsymbol{e}\}$, $\boldsymbol{W_1} \rightsquigarrow \{\boldsymbol{1}, \boldsymbol{2}, \boldsymbol{3}, \boldsymbol{4}, \boldsymbol{5}, \boldsymbol{6}, \boldsymbol{7}, \boldsymbol{8}\}$.





This approach does not necessarily produce a tree structure!

Discretized Optimization Problem

Using the notation $j(u, w_0, w_1) = \varepsilon u^2 + (w_0 + u + w_1)^2$, the discretized optimization problem is

$$\min_{\{u^k\}} \sum_{k \in \{\boldsymbol{a},\dots,\boldsymbol{e}\}} \sum_{i=1}^{8} \pi^{ik} \, j(u^k, w_0^i, w_1^i) \ ,$$

where π^{ik} is the probability weight of the cell ik, u^k is the control value on the cell k and w^i the noise value on the cell i. *Note that some of the weights π^{ik}'s are equal to zero.*

The solution of this discretized problem can be computed (finite dimensional optimization). We expect that the optimal cost of the discretized problem converges to the true optimal cost $J^\sharp$ as the numbers of points in the 2 discrete sets associated to information and noise ($\{a, \dots, e\}$ and $\{1, \dots, 8\}$ in our example) go to infinity.

Discretized Optimization Problem

Using the notation $j(u, w_0, w_1) = \varepsilon u^2 + (w_0 + u + w_1)^2$, the discretized optimization problem is

$$\min_{\{u^k\}} \sum_{k \in \{\boldsymbol{a},\ldots,\boldsymbol{e}\}} \sum_{i=1}^{8} \pi^{ik} j(u^k, w_0^i, w_1^i) ,$$

where π^{ik} is the probability weight of the cell ik, u^k is the control value on the cell k and w^i the noise value on the cell i. *Note that some of the weights π^{ik}'s are equal to zero.*

The solution of this discretized problem can be computed (finite dimensional optimization). We expect that the optimal cost of the discretized problem converges to the true optimal cost $J^\sharp$ as the numbers of points in the 2 discrete sets associated to information and noise ($\{\boldsymbol{a}, \ldots, \boldsymbol{e}\}$ and $\{1, \ldots, 8\}$ in our example) go to infinity.

Problem and its Approximation

We consider the general form of a stochastic optimisation problem:

$$\mathcal{V}(\boldsymbol{W}, \mathcal{B}) = \min_{\boldsymbol{U} \in \mathcal{U}} \mathbb{E}\big(j(\boldsymbol{U}, \boldsymbol{W})\big) ,$$

subject to

$$\boldsymbol{U} \text{ is } \mathcal{B}\text{-measurable} .$$

We consider a sequence of random noises $\{\boldsymbol{W}_n\}_{n\in\mathbb{N}}$ and another sequence of σ-fields $\{\mathcal{B}_n\}_{n\in\mathbb{N}}$ such that the $\boldsymbol{W}_n$'s and the $\mathcal{B}_n$'s have "finite" representations, e.g.

- $\boldsymbol{W}_n = \sum_{i=1}^{n} w^i \mathbf{1}_{\Omega_i}$, $(\Omega_1, \dots, \Omega_n)$ being a partition of Ω,
- $\mathcal{B}_n = \sigma(\Omega^1, \dots, \Omega^n)$, $(\Omega^1, \dots, \Omega^n)$ being a partition of Ω.

We are interested in the sequence of values $\{\mathcal{V}(\boldsymbol{W}_n, \mathcal{B}_n)\}_{n\in\mathbb{N}}$.

Problem and its Approximation

We consider the general form of a stochastic optimisation problem:

$$\mathcal{V}(\boldsymbol{W}, \mathcal{B}) = \min_{\boldsymbol{U} \in \mathcal{U}} \mathbb{E}\big(j(\boldsymbol{U}, \boldsymbol{W})\big) ,$$

subject to

$$\boldsymbol{U} \text{ is } \mathcal{B}\text{-measurable} .$$

We consider a sequence of random noises $\{\boldsymbol{W}_n\}_{n \in \mathbb{N}}$ and another sequence of σ-fields $\{\mathcal{B}_n\}_{n \in \mathbb{N}}$ such that the $\boldsymbol{W}_n$'s and the $\mathcal{B}_n$'s have "finite" representations, e.g.

- $\boldsymbol{W}_n = \sum_{i=1}^{n} w^i \mathbf{1}_{\Omega_i}$, $(\Omega_1, \dots, \Omega_n)$ being a partition of Ω,
- $\mathcal{B}_n = \sigma(\Omega^1, \dots, \Omega^n)$, $(\Omega^1, \dots, \Omega^n)$ being a partition of Ω.

We are interested in the sequence of values $\{\mathcal{V}(\boldsymbol{W}_n, \mathcal{B}_n)\}_{n \in \mathbb{N}}$.

Problem and its Approximation

We consider the general form of a stochastic optimisation problem:

$$\mathcal{V}(\boldsymbol{W}, \mathcal{B}) = \min_{\boldsymbol{U} \in \mathcal{U}} \mathbb{E}\big(j(\boldsymbol{U}, \boldsymbol{W})\big) ,$$

subject to

$$\boldsymbol{U} \text{ is } \mathcal{B}\text{-measurable} .$$

We consider a sequence of random noises $\big\{\boldsymbol{W}_n\big\}_{n\in\mathbb{N}}$ and another sequence of σ-fields $\big\{\mathcal{B}_n\big\}_{n\in\mathbb{N}}$ such that the $\boldsymbol{W}_n$'s and the $\mathcal{B}_n$'s have "finite" representations, e.g.

- $\boldsymbol{W}_n = \sum_{i=1}^n w^i \mathbf{1}_{\Omega_i}$, $(\Omega_1, \ldots, \Omega_n)$ being a partition of Ω,
- $\mathcal{B}_n = \sigma(\Omega^1, \ldots, \Omega^n)$, $(\Omega^1, \ldots, \Omega^n)$ being a partition of Ω.

We are interested in the sequence of values $\big\{\mathcal{V}(\boldsymbol{W}_n, \mathcal{B}_n)\big\}_{n\in\mathbb{N}}$.

Convergence Notions for $\boldsymbol{W}$

These are standard and well known notions.

Convergence Notions for $\boldsymbol{W}$

These are standard and well known notions.

- Convergence in distribution: $\boldsymbol{W}_n \xrightarrow{\mathcal{D}} \boldsymbol{W}$.

  $$\lim_{n \to +\infty} \mathbb{E}\Big(f(\boldsymbol{W}_n)\Big) = \mathbb{E}\Big(f(\boldsymbol{W})\Big) \text{ for all continuous bounded } f \ .$$

  This is the underlying concept in the Monte Carlo method: the empirical law defined by a N-sample $(\boldsymbol{W}^{(1)}, \ldots, \boldsymbol{W}^{(n)})$ of $\boldsymbol{W}$, that is, $\frac{1}{n}\sum_{i=1}^{n} \delta_{\boldsymbol{W}^{(i)}}$, weakly converges to $\mathbb{P}_{\boldsymbol{W}}$.

- Convergence in probability: $\boldsymbol{W}_n \xrightarrow{\mathbb{P}} \boldsymbol{W}$.

  $$\forall \epsilon > 0 \ , \quad \lim_{n \to +\infty} \mathbb{P}\Big(\big\|\boldsymbol{W}_n - \boldsymbol{W}\big\|_{\mathbb{W}} \geq \epsilon\Big) = 0 \ .$$

  This notion is much stronger than the previous one.

Convergence Notions for $\boldsymbol{W}$

These are standard and well known notions.

- Convergence in distribution: $\boldsymbol{W}_n \xrightarrow{\mathcal{D}} \boldsymbol{W}$.

  $$\lim_{n \to +\infty} \mathbb{E}\Big(f(\boldsymbol{W}_n) \Big) = \mathbb{E}\Big(f(\boldsymbol{W}) \Big) \text{ for all continuous bounded } f \; .$$

  This is the underlying concept in the Monte Carlo method: the empirical law defined by a N-sample $(\boldsymbol{W}^{(1)}, \dots, \boldsymbol{W}^{(n)})$ of $\boldsymbol{W}$, that is, $\frac{1}{n} \sum_{i=1}^{n} \delta_{\boldsymbol{W}^{(i)}}$, weakly converges to $\mathbb{P}_{\boldsymbol{W}}$.

- Convergence in probability: $\boldsymbol{W}_n \xrightarrow{\mathbb{P}} \boldsymbol{W}$.

  $$\forall \epsilon > 0 \; , \quad \lim_{n \to +\infty} \mathbb{P}\Big(\big\| \boldsymbol{W}_n - \boldsymbol{W} \big\|_{\mathbb{W}} \geq \epsilon \Big) = 0 \; .$$

  This notion is much stronger than the previous one.

Convergence Notions for $\mathcal{B}$

These notions and results are a little less well known...

Strong Convergence of σ-fields: $\mathcal{B}_n \to \mathcal{B}$.

$$\lim_{n \to +\infty} \mathbb{E}(f \mid \mathcal{B}_n) \xrightarrow{L^1} \mathbb{E}(f \mid \mathcal{B}) \ \text{ for all } \ f \in \mathrm{L}^1(\mathbb{R}) .$$

Main properties.

Convergence Notions for $\mathcal{B}$

These notions and results are a little less well known. . .

Strong Convergence of $\boldsymbol{\sigma}$-fields: $\mathcal{B}_n \to \mathcal{B}$.

$$\lim_{n \to +\infty} \mathbb{E}\big(\boldsymbol{f} \mid \mathcal{B}_n\big) \xrightarrow{L^1} \mathbb{E}\big(\boldsymbol{f} \mid \mathcal{B}\big) \text{ for all } \boldsymbol{f} \in \mathrm{L}^1(\mathbb{R}) \ .$$

Main properties.

Convergence Notions for $\mathcal{B}$

These notions and results are a little less well known...

Strong Convergence of $\boldsymbol{\sigma}$-fields: $\mathcal{B}_n \to \mathcal{B}$.

$$\lim_{n\to+\infty} \mathbb{E}\big(\boldsymbol{f} \mid \mathcal{B}_n\big) \xrightarrow{L^1} \mathbb{E}\big(\boldsymbol{f} \mid \mathcal{B}\big) \text{ for all } \boldsymbol{f} \in \mathrm{L}^1(\mathbb{R}) .$$

Main properties.

1. The topology of the strong convergence is metrizable, so that the space $\mathcal{A}^\star$ of sub-fields of $\mathcal{A}$ is a complete separable metric space.
2. The σ-fields generated by a finite partition of Ω are dense in $\mathcal{A}^\star$ equipped with the previous metric.
3. Let $\{\boldsymbol{Y}_n\}_{n\in\mathbb{N}}$ be a sequence of random variables such that

$$\boldsymbol{Y}_n \xrightarrow{\mathbb{P}} \boldsymbol{Y} \text{ and } \sigma(\boldsymbol{Y}_n) \subset \sigma(\boldsymbol{Y}) \ \forall n. \text{ Then, } \sigma(\boldsymbol{Y}_n) \to \sigma(\boldsymbol{Y}).$$

Convergence Notions for $\mathcal{B}$

These notions and results are a little less well known...

Strong Convergence of $\boldsymbol{\sigma}$-fields: $\mathcal{B}_n \to \mathcal{B}$.

$$\lim_{n\to+\infty} \mathbb{E}\big(\boldsymbol{f} \mid \mathcal{B}_n\big) \xrightarrow{L^1} \mathbb{E}\big(\boldsymbol{f} \mid \mathcal{B}\big) \text{ for all } \boldsymbol{f} \in \mathrm{L}^1(\mathbb{R}) .$$

Main properties.

1. The topology of the strong convergence is metrizable, so that the space $\mathcal{A}^\star$ of sub-fields of $\mathcal{A}$ is a complete separable metric space.
2. The σ-fields generated by a finite partition of Ω are dense in $\mathcal{A}^\star$ equipped with the previous metric.
3. Let $\{\boldsymbol{Y}_n\}_{n\in\mathbb{N}}$ be a sequence of random variables such that $\boldsymbol{Y}_n \xrightarrow{\mathbb{P}} \boldsymbol{Y}$ and $\sigma(\boldsymbol{Y}_n) \subset \sigma(\boldsymbol{Y})$ $\forall n$. Then, $\sigma(\boldsymbol{Y}_n) \to \sigma(\boldsymbol{Y})$.

Convergence Notions for $\mathcal{B}$

These notions and results are a little less well known...

Strong Convergence of $\boldsymbol{\sigma}$-fields: $\mathcal{B}_n \to \mathcal{B}$.

$$\lim_{n \to +\infty} \mathbb{E}(\boldsymbol{f} \mid \mathcal{B}_n) \xrightarrow{L^1} \mathbb{E}(\boldsymbol{f} \mid \mathcal{B}) \text{ for all } \boldsymbol{f} \in \mathrm{L}^1(\mathbb{R}) .$$

Main properties.

1. The topology of the strong convergence is metrizable, so that the space $\mathcal{A}^\star$ of sub-fields of $\mathcal{A}$ is a complete separable metric space.
2. The σ-fields generated by a finite partition of Ω are dense in $\mathcal{A}^\star$ equipped with the previous metric.
3. Let $\{\boldsymbol{Y}_n\}_{n\in\mathbb{N}}$ be a sequence of random variables such that $\boldsymbol{Y}_n \xrightarrow{\mathbb{P}} \boldsymbol{Y}$ and $\sigma(\boldsymbol{Y}_n) \subset \sigma(\boldsymbol{Y})$ $\forall n$. Then, $\sigma(\boldsymbol{Y}_n) \to \sigma(\boldsymbol{Y})$.

Convergence Theorem

Theorem

Let $\mathcal{W} = L^q(\Omega, \mathcal{A}, \mathbb{P}; \mathbb{W})$ *and* $\mathcal{U} = L^r(\Omega, \mathcal{A}, \mathbb{P}; \mathbb{U})$, *with* $1 \leq q < +\infty$ *and* $1 \leq r < +\infty$. *Under the assumptions*

- $\boldsymbol{H_1}$ *the sequence* $\{\mathcal{B}_n\}_{n \in \mathbb{N}}$ *strongly converges to* $\mathcal{B}$, *and* $\mathcal{B}_n \subset \mathcal{B}$,
- $\boldsymbol{H_2}$ *the sequence* $\{\boldsymbol{W}_n\}_{n \in \mathbb{N}}$ L^q *converges to* $\boldsymbol{W}$ *(in* L^q*-norm),*
- $\boldsymbol{H_3}$ *the normal integrand* j *is such that*

$$\forall (u, u') \in \mathbb{U}^2 \; , \; \forall (w, w') \in \mathbb{W}^2 \; ,$$

$$\big| j(u, w) - j(u', w') \big| \leq \alpha \, \|u - u'\|_{\mathbb{U}}^r + \beta \, \|w - w'\|_{\mathbb{W}}^q \; ,$$

the convergence of the approximated optimal costs holds true

$$\lim_{n \to +\infty} \mathcal{V}\big(\boldsymbol{W}_n, \mathcal{B}_n\big) = \mathcal{V}\big(\boldsymbol{W}, \mathcal{B}\big) \; .$$

Using epi-convergence, it is possible to obtain the same results under weaker assumptions and to ensure the convergence of the sequence of the solutions.

Convergence Theorem

Theorem

Let $\mathcal{W} = L^q(\Omega, \mathcal{A}, \mathbb{P}; \mathbb{W})$ and $\mathcal{U} = L^r(\Omega, \mathcal{A}, \mathbb{P}; \mathbb{U})$, with $1 \leq q < +\infty$ and $1 \leq r < +\infty$. Under the assumptions

- **H_1** *the sequence $\{\mathcal{B}_n\}_{n\in\mathbb{N}}$ strongly converges to $\mathcal{B}$, and $\mathcal{B}_n \subset \mathcal{B}$,*
- **H_2** *the sequence $\{\boldsymbol{W}_n\}_{n\in\mathbb{N}}$ L^q converges to $\boldsymbol{W}$ (in L^q-norm),*
- **H_3** *the normal integrand j is such that*

$$\forall (u, u') \in \mathbb{U}^2 , \ \forall (w, w') \in \mathbb{W}^2 ,$$

$$\left| j(u, w) - j(u', w') \right| \leq \alpha \left\| u - u' \right\|_{\mathbb{U}}^r + \beta \left\| w - w' \right\|_{\mathbb{W}}^q ,$$

the convergence of the approximated optimal costs holds true

$$\lim_{n\to+\infty} \mathcal{V}\big(\boldsymbol{W}_n, \mathcal{B}_n\big) = \mathcal{V}\big(\boldsymbol{W}, \mathcal{B}\big) .$$

Using epi-convergence, it is possible to obtain the same results under weaker assumptions and to ensure the convergence of the sequence of the solutions.

Conclusions

- In the discretization of a SOC problem, there are two issues:
  - noise discretization,
  - information discretization.
- The naive Monte Carlo discretization provides a too weak convergence notion (in distribution, not in probability).
- The scenario tree methodology provides an effective way to discretize stochastic optimal control problem, but the two discretizations of information and of noise are bundled.
- Independent discretizations of noise and information offer

Conclusions

- In the discretization of a SOC problem, there are two issues:
  - noise discretization,
  - information discretization.
- The naive Monte Carlo discretization provides a too weak convergence notion (in distribution, not in probability).
- The scenario tree methodology provides an effective way to discretize stochastic optimal control problem, but the two discretizations of information and of noise are bundled.
- Independent discretizations of noise and information offer

Conclusions

- In the discretization of a SOC problem, there are two issues:
  - noise discretization,
  - information discretization.
- The naive Monte Carlo discretization provides a too weak convergence notion (in distribution, not in probability).
- The scenario tree methodology provides an effective way to discretize stochastic optimal control problem, but the two discretizations of information and of noise are bundled.
- Independent discretizations of noise and information offer

Conclusions

- In the discretization of a SOC problem, there are two issues:
  - noise discretization,
  - information discretization.
- The naive Monte Carlo discretization provides a too weak convergence notion (in distribution, not in probability).
- The scenario tree methodology provides an effective way to discretize stochastic optimal control problem, but the two discretizations of information and of noise are bundled.
- Independent discretizations of noise and information offer

Conclusions

- In the discretization of a SOC problem, there are two issues:
  - noise discretization,
  - information discretization.
- The naive Monte Carlo discretization provides a too weak convergence notion (in distribution, not in probability).
- The scenario tree methodology provides an effective way to discretize stochastic optimal control problem, but the two discretizations of information and of noise are bundled.
- Independent discretizations of noise and information offer

Conclusions

- In the discretization of a SOC problem, there are two issues:
  - noise discretization,
  - information discretization.
- The naive Monte Carlo discretization provides a too weak convergence notion (in distribution, not in probability).
- The scenario tree methodology provides an effective way to discretize stochastic optimal control problem, but the two discretizations of information and of noise are bundled.
- Independent discretizations of noise and information offer
  - a greater latitude to select discretization schemes,
  - a way to obtain proper convergence results.

Conclusions

- In the discretization of a SOC problem, there are two issues:
  - noise discretization,
  - information discretization.

- The naive Monte Carlo discretization provides a too weak convergence notion (in distribution, not in probability).
- The scenario tree methodology provides an effective way to discretize stochastic optimal control problem, but the two discretizations of information and of noise are bundled.
- Independent discretizations of noise and information offer
  - a greater latitude to select discretization schemes,
  - a way to obtain proper convergence results.

Conclusions

- In the discretization of a SOC problem, there are two issues:
  - noise discretization,
  - information discretization.
- The naive Monte Carlo discretization provides a too weak convergence notion (in distribution, not in probability).
- The scenario tree methodology provides an effective way to discretize stochastic optimal control problem, but the two discretizations of information and of noise are bundled.
- Independent discretizations of noise and information offer
  - a greater latitude to select discretization schemes,
  - a way to obtain proper convergence results.