Convergence of approximations of stochastic optimization problems subject to measurability constraints.

Pierre Carpentier and SOWG

ENSTA — ENPC — EDF

April 13, 2006
Outline of the presentation

1. Introduction
2. Counterexample
3. Convergence theorem
4. Conclusions
1 Introduction
   - Problem statement
   - Strong convergence topology of $\sigma$-fields
   - A two-stage convergence result

2 Counterexample

3 Convergence theorem

4 Conclusions
Prototype Problem

Stochastic optimization problem under consideration:

\[ V(\xi, \mathcal{F}) = \min_{\mathbf{u} \in L^2(\Omega, \mathcal{A}, \mathbb{P}; \mathcal{U})} \mathbb{E}\left[ j(\mathbf{u}, \xi) \right] , \quad (1a) \]

subject to \( \mathbf{u} \) \( \mathcal{F} \)-measurable. \( \quad (1b) \)

- \((\Omega, \mathcal{A}, \mathbb{P})\) a probability space,
- \(\xi\) a random variable on \(\Xi\) (noise),
- \(\mathbf{u}\) a random variable on \(U\) (control),
- \(\mathcal{F}\) a sub \(\sigma\)-field of \(\mathcal{A}\) (observation, usually generated by \(y\)).
Prototype Problem

Stochastic optimization problem under consideration:

\[ V(\xi, \mathcal{F}) = \min_{u \in L^2(\Omega, \mathcal{A}, \mathbb{P}; U)} \mathbb{E}[j(u, \xi)] , \]  

subject to \( u \ \mathcal{F}-\text{measurable} \) . \hspace{1cm} (1a)

\[ \min \mathbb{E} \left[ \sum_{t=0}^{T-1} L_{t+1}(x_t, u_t, \xi_{t+1}) + K(x_T) \right] \] ,

subject to \( \begin{cases} x_0 &= f_0(\xi_0) \\ x_{t+1} &= f_{t+1}(x_t, u_t, \xi_{t+1}) \\ u_t &= \sigma(\xi_0, \ldots, \xi_t) - \text{measurable} . \end{cases} \) 

Pierre Carpentier and SOWG
CIMODE’06
Prototype Problem

Stochastic optimization problem under consideration:

\[ V(\xi, \mathcal{F}) = \min_{u \in L^2(\Omega, \mathcal{A}, \mathbb{P}; \mathcal{U})} \mathbb{E}[j(u, \xi)], \]  

subject to \[ u \text{ } \mathcal{F}-\text{measurable}. \] \hspace{1cm} (1a)

In order to obtain a tractable approximation of problem (1),

1. the expectation in (1a) must be discretized,
2. and the \( \sigma \)-field \( \mathcal{F} \) in (1b) must be discretized.

These two discretizations are a priori independent.

The first discretization is somewhat traditional (Monte Carlo), whereas the last one is not so well-known...
Strong convergence topology of $\sigma$-fields (Neveu)

Coarest topology such that conditional expectation is continuous w.r.t. the $\sigma$-field:

$$\lim_{n \to +\infty} \mathcal{F}_n = \mathcal{F} \iff \lim_{n \to +\infty} \left\| \mathbb{E}[f \mid \mathcal{F}_n] - \mathbb{E}[f \mid \mathcal{F}] \right\|_{L^1} = 0 \quad \forall f \in L^1.$$ 

Main properties of the strong topology (Cotter)

- The strong convergence topology is metrizable.
- The set of $\sigma$-fields generated by a finite partition is dense.
- If $y_n \xrightarrow{P} y$ and $\sigma(y_n) \subset \sigma(y)$, then $\sigma(y_n) \to \sigma(y)$.

Based on these notions, Barty has proposed a discretization scheme in order to approximate problem (1).
Discretization scheme

1. Approximate $\mathcal{F}$ by $\mathcal{F}_k$ generated by a finite partition of $\Omega$:

$$V(\xi, \mathcal{F}_k) = \min_{u \mathcal{F}_k-\text{measurable}} \mathbb{E}[j(u, \xi)].$$

2. Approximate $\xi$ by a finitely valued random variable $\xi_n$:

$$V(\xi_n, \mathcal{F}_k) = \min_{u \mathcal{F}_k-\text{measurable}} \mathbb{E}[j(u, \xi_n)].$$

Convergence theorem (Barty)

1. Information structure discretization error:

$$|V(\xi, \mathcal{F}) - V(\xi, \mathcal{F}_k)| \to 0 \text{ as } \mathcal{F}_k \to \mathcal{F} \text{ strongly.}$$

2. Mean computation discretization error:

$$|V(\xi, \mathcal{F}_k) - V(\xi_n, \mathcal{F}_k)| \to 0 \text{ as } \xi_n \to \xi \text{ in distribution.}$$

Discretization scheme independent in $\xi$ and $\mathcal{F}$?
1. Introduction

2. Counterexample
   - Formulation and exact solution
   - Discretization scheme
   - Approximated solution
   - What is wrong?

3. Convergence theorem

4. Conclusions
Formulation

- \( x \) and \( w \): independent uniformly distributed random variables on \([-1, 1]\) (initial state and noise),
- \( u \): random variable based on the observation of \( x \) (control),
- \( z = x + u + w \) (final state).
- The problem is formulated on \(([-1, 1]^2, \mathcal{B}_{[-1,1]^2}, \mu)\):
  \[
  \min_{u \in \sigma(x)-\text{measurable}} \mathbb{E} \left[ \epsilon u^2 + z^2 \right].
  \]  

Exact resolution using dynamic programming

\[
 u^\#(x) = -\frac{x}{1 + \epsilon} \quad \text{and} \quad J^\# = V((x, w), \mathcal{F}) = \frac{1}{3} \left( 1 + \frac{\epsilon}{1 + \epsilon} \right),
\]

with \( \mathcal{F} = \mathcal{B}_{[-1,1]} \otimes \{ \emptyset, [-1, 1] \} \): sub \( \sigma \)-field generated by \( x \).
Figure: Partition of $[-1, 1]^2$ and associated sample.
Let \( n \in \mathbb{N}^* \). Let \( (F_n^{(1)}, \ldots, F_n^{(n)}) \) be a partition of \([-1, 1]^2\), with
\[
F_n^{(k)} = \left( \frac{2(k-1)}{n} - 1, \frac{2k}{n} - 1 \right] \times [-1, 1].
\]

Let \( \mathcal{F}_n \) be the sub \( \sigma \)-field generated by \( (F_n^{(1)}, \ldots, F_n^{(n)}) \).

- \( (\mathcal{F}_n)_{n \in \mathbb{N}} \) strongly converges to \( \mathcal{F} \),
- \( u \) is \( \mathcal{F}_n \) - measurable \( \iff \) \( u \) is constant over each \( F_n^{(k)} \)
\[
\iff u(x, w) = \sum_{k=1}^{n} u_n^{(k)} \mathbb{I}_{F_n^{(k)}}(x, w).
\]
Let $(\zeta_n)_{n \in \mathbb{N}}$ be a deterministic sequence of elements in $[-1, 1]^2$ such that the associated sequence of empirical probability laws narrowly converges to $\mu$. For $n \in \mathbb{N}^*$ and $k \in \{1, \ldots, n\}$, let

$$(x_n^{(k)}, w_n^{(k)}) = \left(\frac{2k - 1}{n} - 1 + \frac{\zeta_k,1}{n}, \zeta_k,2\right),$$

and define the approximation $(x_n, w_n)$ of $(x, w)$ by

$$(x_n, w_n) = \sum_{k=1}^{n} \left(x_n^{(k)}, w_n^{(k)}\right) \mathbb{1}_{F_n^{(k)}}(x, w).$$

- $(x_n, w_n)_{n \in \mathbb{N}}$ converges in distribution to $(x, w)$,
- $(x_n, w_n)$ is constant over each subset $F_n^{(k)}$. 
Approximated problem

\[
\min_{(u_n^{(1)}, \ldots, u_n^{(n)}) \in \mathbb{R}^n} \sum_{k=1}^{n} \int_{F_n^{(k)}} \left( \epsilon (u_n^{(k)})^2 + (x_n^{(k)} + u_n^{(k)} + w_n^{(k)})^2 \right) \mu (dxdw) .
\]

Approximated solution

\[
\hat{u}_n^{(k)} = -\frac{x_n^{(k)} + w_n^{(k)}}{1 + \epsilon} \quad \text{and} \quad \hat{J}_n = V((x_n, w_n), F_n) \rightarrow \frac{2}{3} \left( \frac{\epsilon}{1 + \epsilon} \right).
\]

Approximated feedback in (2)

\[
\hat{u}_n(x, w) = -\sum_{k=1}^{n} \frac{x_n^{(k)} + w_n^{(k)}}{1 + \epsilon} \mathbb{I}_{F_n^{(k)}} (x, w) \sim \mathbb{E} \left[ \epsilon \hat{u}_n^2 + z^2 \right] \rightarrow \frac{2}{3}.
\]

Discretization fails to asymptotically give the optimal solution.
Same notions of convergence as in the two-stage procedure, but

- $\mathcal{F}$ and $\xi$ are \textit{independently} approximated:

  this “diagonal” discretization makes possible to solve each open-loop subproblem using a \textit{unique} sample of the random variable (a poor way to compute conditional expectations).

- The convergence notion used for $\xi$ is \textit{weak}:

  \[ \{(x_n, w_n)\}_{n \in \mathbb{N}} \] does not converge \textit{in probability} to $(x, w)$.

Note that with an stronger convergence notion (e.g. in $L^2$), each open-loop subproblem will be correctly approximated.

**Question:** can we expect a diagonal convergence when using a stronger convergence notion for the random variable?
1 Introduction

2 Counterexample

3 Convergence theorem
   • Notations
   • Theorem
   • Sketch of proof
   • Remarks

4 Conclusions
We go back to problem (1):

- \((\Omega, \mathcal{A}, \mathbb{P})\) is a probability space, and \(\mathcal{F}\) a sub \(\sigma\)-field of \(\mathcal{A}\),
- \(\mathcal{X}\) is the space of integrable \(\Xi\)-valued random variables with the topology of convergence in probability \(\mathbb{P}\),
- \(U\) is the space \(L^p(\Omega, \mathcal{A}, \mathbb{P}; U)\) with \(p \in [1, +\infty)\),
- \(\Delta(\mathcal{F})\) is the subset of \(\mathcal{F}\)-measurable random variables of \(U\).

Let \(j\) be a normal integrand on \(U \times \Xi\), and let \(J\) be the associated integral functional on \(U \times \mathcal{X}\):

\[ J(u, \xi) = \mathbb{E}[j(u, \xi)]. \]

\[ V(\xi, \mathcal{F}) = \min_{u \in \Delta(\mathcal{F})} J(u, \xi). \]
Theorem

Under the following assumptions:

**H1** \( J \) is a continuous function on \( \mathcal{U} \times \mathcal{X} \),

**H2** \( \{ \mathcal{F}_n \}_{n \in \mathbb{N}} \) strongly converges to \( \mathcal{F} \),

**H3** \( \forall n \in \mathbb{N}, \mathcal{F}_n \subset \mathcal{F} \),

**H4** \( \{ \xi_n \}_{n \in \mathbb{N}} \) converges to \( \xi \) in probability,

**H5** \( \{ J(\cdot, \xi_n) \}_{n \in \mathbb{N}} \) uniformly converges to \( J(\cdot, \xi) \) on \( \Delta(\mathcal{F}) \):

\[
\forall \epsilon > 0, \exists N \in \mathbb{N}, \forall n \geq N, \forall u \in \Delta(\mathcal{F}), |J(u, \xi_n) - J(u, \xi)| \leq \epsilon ,
\]

the convergence of the optimal costs holds true:

\[
\lim_{n \to +\infty} V(\xi_n, \mathcal{F}_n) = V(\xi, \mathcal{F}) . \quad (3)
\]
\[
\limsup_{n \to +\infty} V(\xi_n, \mathcal{F}_n) \leq V(\xi, \mathcal{F})
\]

- \(\forall u \in \Delta(\mathcal{F}), \) define \(u_n = \mathbb{E}[u \mid \mathcal{F}_n].\) Then, \(\mathcal{F}_n \to \mathcal{F} \Longrightarrow u_n \to u.\)
  
  The set-valued mapping \(\Delta\) is thus lsc.
- \(J\) being usc, we conclude that the marginal function \(V\) is also usc.

\[
\liminf_{n \to +\infty} V(\xi_n, \mathcal{F}_n) \geq V(\xi, \mathcal{F})
\]

- From \(J(u, \xi_n) = J(u, \xi) + (J(u, \xi_n) - J(u, \xi)),\) we obtain:
  
  \[
  \min_{u \in \Delta(\mathcal{F}_n)} J(u, \xi_n) \geq \min_{u \in \Delta(\mathcal{F}_n)} J(u, \xi) + \min_{u \in \Delta(\mathcal{F}_n)} \left( J(u, \xi_n) - J(u, \xi) \right).
  \]
- Using \(\mathcal{F}_n \subset \mathcal{F} \implies \Delta(\mathcal{F}_n) \subset \Delta(\mathcal{F}),\) we deduce:
  
  \[
  V(\xi_n, \mathcal{F}_n) \geq V(\xi, \mathcal{F}) + \min_{u \in \Delta(\mathcal{F})} \left( J(u, \xi_n) - J(u, \xi) \right).
  \]
- The conclusion is then a consequence of \(H5.\)
1. Same result with:

\[ \Delta(\mathcal{F}) = \{ u \in U, \ u \ F - \text{measurable}, \ u(\omega) \in U^{ad} \ \mathbb{P} - \text{as} \} , \]

\[ U^{ad} \] being a closed convex subset of \( U \).

2. If \( \mathcal{F} \) is generated by a \( Y \)-valued random variable \( y \), \( \mathcal{F}_n \) may be constructed thanks to a *quantification operator* \( q_n \) on \( Y \):

\[ \mathcal{F}_n = \sigma(q_n \circ y) . \]

3. Assumption \( H5 \) is implied by:

\[ \exists \alpha > 0, \forall u \in U, \forall (\xi, \xi') \in \Xi \times \Xi, \ |j(u, \xi) - j(u, \xi')| \leq \alpha \| \xi - \xi' \|_\Xi . \]

4. Theorem does not make use of the right tool for convergence analysis, namely *epi-convergence*. 
Conclusions

1. Another sight on stochastic approximation.
2. Scenario trees are not built-in in stochastic programming.
3. Comparison with Pennanen work.
Conclusions

1. Another sight on stochastic approximation.
2. Scenario trees are not built-in in stochastic programming.
3. Comparison with Pennanen work.
Conclusions

1. Another sight on stochastic approximation.
2. Scenario trees are not built-in in stochastic programming.

3. Comparison with Pennanen work.
K. Barty

*Contributions à la discrétisation des contraintes de mesurabilité pour les problèmes d’optimisation stochastique.*


K.D. Cotter

*Similarity of information and behavior with a pointwise convergence topology.*


T. Pennanen

*Epi-Convergent Discretizations of Multistage Stochastic Programs.*


C. Strugarek and SOWG

*On the Fortet-Mourier metric for the stability of Stochastic Optimization Problems, an example.*

Figure: Partition of $[-1, 1]^2$ and Voronoi cells.