Tree Generation through Clustering

Convergence and a test case

Jean-Sébastien Roy

EDF Research & Development

17 August 2006

Introduction

For multistage problems, stochastic programming is usually performed on scenario trees.

An essential problem to solve to apply this method is: How to generate the tree when only historical scenarios are available ?

A classical method is to use a clustering heuristic. The goal of this short talk is to present the convergence of this heuristic and illustrate it on an example.

Plan

Overview of stochastic programming

Tree generation through clustering

Illustration on an example

Plan

Overview of stochastic programming

Tree generation through clustering

Illustration on an example

Plan

Overview of stochastic programming

Tree generation through clustering

Illustration on an example

Plan

Overview of stochastic programming

Tree generation through clustering

Illustration on an example

Stochastic optimization is difficult problem, even more so in the multistage case [Shapiro, 2006], but appears nevertheless quite frequently in a large number of domains and applications (finance, production, power, logistic, scheduling, ...).

Multistage problems differ from two stages ones by their information structure. In the multistage case, not only an optimal decision must be obtained for the first stage, but also decision rules for the following stages, which are functions of the information acquired at each stage.

This progressive acquisition of information and the associated measurability constraints make this problem much harder.

Therefore, contrarily to the two stages case, it is not possible to use a simple Monte Carlo sampling approach.

I'll put aside dynamic programming based methods which are a separate class with their strengths and weaknesses and I will concentrate on stochastic programming based methods. Among these methods, we can find:

1. Closed loop stochastic gradient methods [Barty et al., 2005] which adapt stochastic gradient techniques to the functional nature of the multistage problems ;
2. Methods of progressive refinement of a discrete approximation of the problem [Casey and Sen, 2006];
3. And the most common methods which first discretize the problem completely before solving it using classical means.

Stochastic Optimization
A priori discretization

Among those methods we can find:

1. Linear decision rules based methods [Thénié and Vial, 2004, Holt et al., 1955] or quantization [Barty, 2004], which discretize the multistage problem into a two stages one ;
2. And the most widely used technique usually associated with "Stochastic Programing" which consists in discretizing the problem into a deterministic finite dimension problem, solve by large scale optimization methods, such as linear programming. This discretization is usually performed with the help of a tree, a structure which intrinsically takes into account measurability constraints on the decisions.

While my approach will be slightly different, I'll first concentrate on that last class of methods.

Two widely different cases can occur:

1. Either we know how to perform conditional sampling on the uncertainties, typically when the processes involved are known, or when the uncertainties are independent. In that case, a large amount of tree building techniques are available, either Monte-Carlo based [Shapiro, 2003, Shapiro, 2006] or Quasi Monte Carlo based [Pennanen and Koivu, 2005].
2. Or we cannot perform such sampling and only scenarios are available, typically when only historical data are available, or the scenario generation procedure is not adapted to conditional sampling (e.g., some physical simulations). In this last case, if the uncertainties are continuous, the scenarios do no lend themselves directly to a tree structure.

In this last case, a few techniques are available (see [Dupačová et al., 2000] for a survey). Mainly:

1. Moment matching techniques [Høyland and Wallace, 2001, Høyland et al., 2003];
2. Probability metric methods [Römisch and Heitsch, 2006, Pflug, 2001, Hochreiter and Pflug, 2006], and among them, the classical tree generation by clustering method, typically used for its simplicity.

Clustering is somewhat simpler that the SCENRED tree reduction method [Römisch and Heitsch, 2006] and optimal tree generation [Pflug, 2001]. Moreover, clustering does not assume the problem is already formulated on a tree, which is almost never the case.

Plan

Problem setting

let $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ two random variables with support $\Xi_1 \subset \mathbb{R}^{p_1}$ and $\Xi_2 \subset \mathbb{R}^{p_2}$, with $\Xi_1 \times \Xi_2$ equipped with the euclidean norm denoted $\|\cdot\|$. We look for the solution of the following three-stage stochastic problem:

$$J = \min_{u_0, u_1(\cdot)} \mathbb{E}\left[f\left(u_0, u_1\left(\boldsymbol{\xi}_1\right), \boldsymbol{\xi}_1, \boldsymbol{\xi}_2\right)\right] \qquad (1)$$

Subject to

$$\begin{cases} u_0 \in U_0 \subset \mathbb{R}^{n_0} \\ \forall \xi_1 \in \Xi_1,\ u_1\left(\xi_1\right) \in U_1 \subset \mathbb{R}^{n_1} \end{cases}$$

Suppose that for all $u_0 \in U_0$ $u_1\left(\xi_1\right) \in U_1$, ξ_1 and ξ_2, $f\left(u_0, u_1\left(\boldsymbol{\xi}_1\right), \xi_1, \xi_2\right)$ is finite and that $f\left(u_0, u_1\left(\boldsymbol{\xi}_1\right), \boldsymbol{\xi}_1, \boldsymbol{\xi}_2\right)$ is measurable and integrable for all u_0 and $u_1\left(\cdot\right)$ considered.

Hypothesis on the dynamic

We suppose that there exists a function h_1 such that $\boldsymbol{\xi}_2 = h_1\left(\boldsymbol{\xi}_1, \boldsymbol{\varepsilon}_2\right)$ with $\boldsymbol{\varepsilon}_2$ an $\mathbb{R}^{p_2'}$ valued r.v. independent of $\boldsymbol{\xi}_1$.

Solving from historical scenarios

Let $\left(\xi_1^j, \xi_2^j, \varepsilon_2^j\right)_{j=1,\ldots,N}$ a sample of N independent drawings of the rand variables triplet $(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \boldsymbol{\varepsilon}_2)$.

We look for the solution of problem (1), using only the values $\left(\xi_1^j, \xi_2^j\right)_{j=1,\ldots,N}$ which correspond to the historical scenarios.

The transfer function h_1 is supposed unknown.

Tree generation method

Scenario clustering

We perform a clustering of the N drawings into M_N clusters $(P_i)_{i=1,\dots,M_N}$ (a partition of the indices $j \in \{1,\dots,N\}$) with $1 \le M_N \le N$.

Let:

- $\|P_i\| = \max_{j,j' \in P_i} \left\| \xi_1^j - \xi_1^{j'} \right\|$ the diameter of cluster P_i;
- $\theta_N = \max_{i=1,\dots,N} \|P_i\|$ the maximum diameter;
- $\forall i = 1,\dots,M_N$, $N_i = \mathrm{card} P_i$, the number of drawings in each cluster;
- $m_N = \min_{i=1,\dots,M_N} N_i$ the smallest number of drawings in a cluster;
- $\forall i = 1,\dots,M_N$, $\forall j \in P_i$, $p_j = i$, the cluster to which each drawing belongs to.

Clustering Hypothesis

Assume that when $N \to \infty$, the clustering procedure leads to $\boldsymbol{\theta}_N \to 0$, p.s., and $\exists \alpha > 0$ t.q. $\sum_{N=1}^{\infty} \mathbb{P}\left[\mathbf{m}_N < N^{\alpha}\right] < \infty$.

Example

If $\boldsymbol{\xi}_1$ follows an uniform law over $\Xi_1 = [0, 1]$, we can take $M_N = \left\lceil \sqrt{N} \right\rceil$, $\theta_N = \frac{1}{M_N}$, the M_N clusters being defined by the contiguous intervals of width θ_N.

For all non-empty cluster P_i, we arbitrarily choose a representative element denoted $\hat{\xi}_1^i$ s.t. $\forall j \in P_i$, $\left\|\hat{\xi}_1^i - \xi_1^j\right\| \leq \theta_N$ and $\hat{\xi}_1^i \in \Xi_1$.

Example

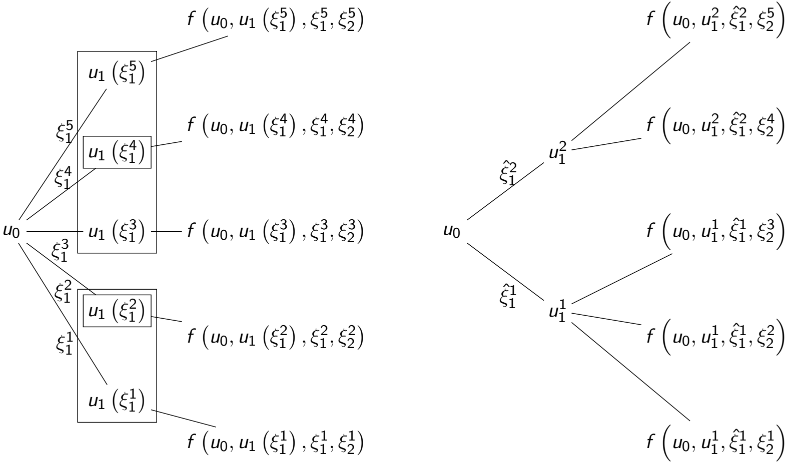
If Ξ_1 is convex, the barycenter of each cluster, $\hat{\xi}_1^i = \frac{1}{N_i} \sum_{j \in P_i} \xi_1^j$, is an adequate choice. Otherwise, we can always choose $\hat{\xi}_1^i$ among the elements of the cluster, i.e., $\left(\xi_1^j\right)_{j \in P_i}$.

Finally, we approximate the initial stochastic problem (1) by the following deterministic tree structured problem:

$$\tilde{J}_N = \min_{u_0, \left(u_1^i\right)_{i=1,\ldots,M_N}} \frac{1}{N} \sum_{j=1,\ldots,N} f\left(u_0, u_1^{p_j}, \hat{\xi}_1^{p_j}, \xi_2^j\right)$$

subject to $u_0 \in U_0$ and $\forall i = 1, \ldots, M_N$, $u_1^i \in U_1$.

Tree generation method
Example



Initial scenario fan

Tree generated by clustering

Figure: Tree generation from 5 scenarios with $P_1 = \{1, 2, 3\}$, $P_2 = \{4, 5\}$, $\hat{\xi}_1^1 = \xi_1^2$ and $\hat{\xi}_1^2 = \xi_1^4$.

Convergence

For all $\eta > 0$, we study the probability $\mathbb{P}\left[\left|\tilde{\mathbf{J}}_N - J\right| \geq \eta\right]$ that the approximate solution of the tree is not η-optimal.

On Quasi Monte Carlo

Please note that the following study does no take into account the low discrepancy properties of the representative elements obtained after clustering and is therefore only useful to prove convergence and does not provide any reasonable clue about the convergence speed.

Notations

Let $v(u_0, \xi_1)$ denote the second stage value:

$$v(u_0, \xi_1) = \min_{u_1 \in U_1} \mathbb{E}\left[f(u_0, u_1, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2) \middle| \boldsymbol{\xi}_1 = \xi_1 \right]$$

$$J = \min_{u_0 \in U_0} \mathbb{E}\left[v(u_0, \boldsymbol{\xi}_1) \right]$$

Let $\tilde{v}_i(u_0, \xi_1)$ denote the discrete approximation of $v(u_0, \xi_1)$ through the clusters:

$$\tilde{v}_i(u_0, \xi_1) = \min_{u_1 \in U_1} \frac{1}{N_i} \sum_{j \in P_i} f\left(u_0, u_1, \xi_1, \xi_2^j\right)$$

$$\tilde{J}_N = \min_{u_0 \in U_0} \frac{1}{N} \sum_{i=1,\ldots,M_N} N_i \tilde{v}_i\left(u_0, \hat{\xi}_1^i\right) = \min_{u_0 \in U_0} \frac{1}{N} \sum_{j=1,\ldots,N} \tilde{v}_{p_j}\left(u_0, \hat{\xi}_1^{p_j}\right)$$

Notations

Let $v(u_0, \xi_1)$ denote the second stage value:

$$v(u_0, \xi_1) = \min_{u_1 \in U_1} \mathbb{E}\left[f(u_0, u_1, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2) \middle| \boldsymbol{\xi}_1 = \xi_1 \right]$$

$$J = \min_{u_0 \in U_0} \mathbb{E}\left[v(u_0, \boldsymbol{\xi}_1) \right]$$

Let $\tilde{v}_i(u_0, \xi_1)$ denote the discrete approximation of $v(u_0, \xi_1)$ through the clusters:

$$\tilde{v}_i(u_0, \xi_1) = \min_{u_1 \in U_1} \frac{1}{N_i} \sum_{j \in P_i} f\left(u_0, u_1, \xi_1, \xi_2^j\right)$$

$$\tilde{J}_N = \min_{u_0 \in U_0} \frac{1}{N} \sum_{i=1,\ldots,M_N} N_i \tilde{v}_i\left(u_0, \hat{\xi}_1^i\right) = \min_{u_0 \in U_0} \frac{1}{N} \sum_{j=1,\ldots,N} \tilde{v}_{p_j}\left(u_0, \hat{\xi}_1^{p_j}\right)$$

Notations (continued)

Let $\hat{J}$ denote the SAA approximation using drawings $\left(\xi_1^j\right)$ and the exact second stage:

$$\hat{J}_N = \min_{u_0} \frac{1}{N} \sum_{j=1,\ldots,N} v\left(u_0, \xi_1^j\right)$$

Finally, though hypothesis (11) on the dynamic, $\forall j = 1, \ldots, N$, $\xi_2^j = h_1\left(\xi_1^j, \varepsilon_2^j\right)$, let $\hat{v}_i\left(u_0, \xi_1\right)$ denote the value of the second stage approximated using the values$\boldsymbol{\xi}_2 | \boldsymbol{\xi}_1 = \xi_1$ generated from $h_1\left(\xi_1, \varepsilon_2^j\right)$ for $j \in P_i$:

$$\hat{v}_i\left(u_0, \xi_1\right) = \min_{u_1 \in U_1} \frac{1}{N_i} \sum_{j \in P_i} f\left(u_0, u_1, \xi_1, h_1\left(\xi_1, \varepsilon_2^j\right)\right)$$

Notations (continued)

Let $\hat{J}$ denote the SAA approximation using drawings $\left(\xi_1^j\right)$ and the exact second stage:

$$\hat{J}_N = \min_{u_0} \frac{1}{N} \sum_{j=1,\dots,N} v\left(u_0, \xi_1^j\right)$$

Finally, though hypothesis (11) on the dynamic, $\forall j = 1, \dots, N$, $\xi_2^j = h_1\left(\xi_1^j, \varepsilon_2^j\right)$, let $\hat{v}_i\left(u_0, \xi_1\right)$ denote the value of the second stage approximated using the values $\boldsymbol{\xi}_2 | \, \boldsymbol{\xi}_1 = \xi_1$ generated from $h_1\left(\xi_1, \varepsilon_2^j\right)$ for $j \in P_i$:

$$\hat{v}_i\left(u_0, \xi_1\right) = \min_{u_1 \in U_1} \frac{1}{N_i} \sum_{j \in P_i} f\left(u_0, u_1, \xi_1, h_1\left(\xi_1, \varepsilon_2^j\right)\right)$$

Proof Outline

The proof is composed of three steps:

1. First we prove the convergence of the second stage towards its exact value;
2. Then we perform the same proof for the first stage;
3. Finally we conclude on the overall convergence.

Convergence of the second stage
Splitting into parts

We compare, $\forall j = 1, \ldots, N$, $v\left(u_0, \xi_1^j\right)$ and $\tilde{v}\left(u_0, \hat{\xi}_1^{p_j}\right)$ using the following upper bounding:

$$\begin{aligned}
\left|v\left(u_0, \xi_1^j\right) - \tilde{v}\left(u_0, \hat{\xi}_1^{p_j}\right)\right| \quad \leq \quad & \left|v\left(u_0, \xi_1^j\right) - v\left(u_0, \hat{\xi}_1^{p_j}\right)\right| \\
& + \left|v\left(u_0, \hat{\xi}_1^{p_j}\right) - \hat{v}_{p_j}\left(u_0, \hat{\xi}_1^{p_j}\right)\right| \\
& + \left|\hat{v}_{p_j}\left(u_0, \hat{\xi}_1^{p_j}\right) - \tilde{v}_{p_j}\left(u_0, \hat{\xi}_1^{p_j}\right)\right|
\end{aligned}$$

The difference $\left| v\left(u_0, \xi_1^j\right) - v\left(u_0, \hat{\xi}_1^{p_j}\right) \right|$ an be upper bounded using regularity properties of h_1 and f.

Hypothesis on the regularity

Assume that:

1. h_1 is uniformly C_{h_1}-Lipschitz in ξ_1;
2. And f is uniformly C_f-Lipschitz in (ξ_1, ξ_2).

Therefore:

$$\left| v\left(u_0, \xi_1^j\right) - v\left(u_0, \hat{\xi}_1^{p_j}\right) \right| \leq \theta_N C_f \sqrt{1 + C_{h_1}^2}$$

Convergence of the second stage

First part : regularity

The difference $\left| v\left(u_0, \xi_1^j\right) - v\left(u_0, \hat{\xi}_1^{p_j}\right) \right|$ an be upper bounded using regularity properties of h_1 and f.

Hypothesis on the regularity

Assume that:

1. h_1 is uniformly C_{h_1}-Lipschitz in ξ_1;
2. And f is uniformly C_f-Lipschitz in (ξ_1, ξ_2).

Therefore:

$$\left| v\left(u_0, \xi_1^j\right) - v\left(u_0, \hat{\xi}_1^{p_j}\right) \right| \leq \theta_N C_f \sqrt{1 + C_{h_1}^2}$$

Convergence of the second stage
Second part: Monte Carlo

The difference $\left| v\left(u_0, \hat{\xi}_1^{p_j}\right) - \hat{v}_{p_j}\left(u_0, \hat{\xi}_1^{p_j}\right)\right|$ can be upper bounded using Shapiro's result on SAA approximation (cf. [Shapiro, 2006, Shapiro and Nemirovski, 2005]).

Define:

$$w\left(u_0, u_1, \xi_1\right) = \mathbb{E}\left[f\left(u_0, u_1, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2\right) \middle| \boldsymbol{\xi}_1 = \xi_1\right]$$

$$\hat{w}_i\left(u_0, u_1, \xi_1\right) = \frac{1}{N_i} \sum_{j \in P_i} f\left(u_0, u_1, \xi_1, h_1\left(\xi_1, \varepsilon_2^j\right)\right)$$

Assume that:

1. $U_0 \times U_1 \times \Xi_1$ is non-empty, closed and of finite diameter D_1;
2. There exists a constant σ_1 s.t.

   $$M_{(u_0,u_1,\xi_1)}(t) \leq e^{\frac{\sigma_1^2 t^2}{2}}, \forall t \in \mathbb{R}, \forall (u_0, u_1, \xi_1) \in U_0 \times U_1 \times \Xi_1$$

   where

   $$M_{(u_0,u_1,\xi_1)}(t) = \mathbb{E}\left[e^{f(u_0,u_1,\boldsymbol{\xi}_1,\boldsymbol{\xi}_2)-w(u_0,u_1,\boldsymbol{\xi}_1)} \middle| \boldsymbol{\xi}_1 = \xi_1 \right]$$

   is the moment generating function of the r.v. $f(u_0, u_1, \xi_1, \boldsymbol{\xi}_2) - w(u_0, u_1, \xi_1)$;
3. And that f is L_1-Lipschitz in (u_0, u_1, ξ_1) uniformly in ξ_2.

Convergence of the second stage
Application of the large deviations theorem

We then apply the theorem 1 from [Shapiro, 2006] which enables us to deduce that there exists $A_1 \geq 0$ s.t. $\forall \eta_1 > 0$, $\forall n \in \mathbb{N}^*$ and $\forall i = 1, \ldots, M_N$:

$$\mathbb{P}\left[\max_{(u_0,u_1,\xi_1)} \left|\hat{\mathbf{w}}_i\left(u_0, u_1, \xi_1\right) - w\left(u_0, u_1, \xi_1\right)\right| \geq \eta_1 \middle| \mathbf{N}_i = n\right] \leq A_1 \left(\frac{D_1 L_1}{\eta_1}\right)^{n_1} e^{-\frac{n\eta_1^2}{16\sigma_1^2}}$$

Which implies, $\forall \left(u_0, \xi_1\right) \in U_0 \times \Xi_1$, $\forall n \in \mathbb{N}^*$, and $\forall i = 1, \ldots, \mathbf{M}_N$:

$$\mathbb{P}\left[\left|\hat{\mathbf{v}}_i\left(u_0, \xi_1\right) - v\left(u_0, \xi_1\right)\right| \geq \eta_1 \middle| \mathbf{N}_i = n\right] \leq A_1 \left(\frac{D_1 L_1}{\eta_1}\right)^{n_1} e^{-\frac{n\eta_1^2}{16\sigma_1^2}}$$

Convergence of the second stage

Application of the large deviations theorem (cont.)

Therefore, $\forall \left(u_0, \xi_1\right) \in U_0 \times \Xi_1$ and $\forall m \in \mathbb{N}^*$, since $M_N \leq N$:

$$\mathbb{P}\left[\max_{i=1,\ldots,M_N} \left|\hat{\mathbf{v}}_i\left(u_0, \xi_1\right) - v\left(u_0, \xi_1\right)\right| \geq \eta_1 \middle| \mathbf{m}_N \geq m\right] \leq A_1 N \left(\frac{D_1 L_1}{\eta_1}\right)^{n_1} e^{-\frac{m\eta_1^2}{16\sigma_1^2}}$$

which implies that $\forall \left(u_0, \xi_1\right) \in U_0 \times \Xi_1$ and $\forall m \in \mathbb{N}^*$:

$$\mathbb{P}\left[\max_{i=1,\ldots,M_N} \left|\hat{\mathbf{v}}_i\left(u_0, \xi_1\right) - v\left(u_0, \xi_1\right)\right| \geq \eta_1\right] \leq A_1 N \left(\frac{D_1 L_1}{\eta_1}\right)^{n_1} e^{-\frac{m\eta_1^2}{16\sigma_1^2}} + \mathbb{P}\left[\mathbf{m}_N < m\right]$$

Convergence of the second stage
Third part : regularity

Observe that:

$$\tilde{v}_i\left(u_0, \hat{\xi}_1^i\right) = \min_{u_1 \in U_1} \frac{1}{N_i} \sum_{j \in P_i} f\left(u_0, u_1, \hat{\xi}_1^i, h_1\left(\xi_1^j, \varepsilon_2^j\right)\right)$$

Therefore the difference $\left|\hat{v}_{p_j}\left(u_0, \hat{\xi}_1^{p_j}\right) - \tilde{v}_{p_j}\left(u_0, \hat{\xi}_1^{p_j}\right)\right|$ can be upper bounded using the regularity of the dynamic h_1 in ξ_1 since $\left\|\xi_1^j - \hat{\xi}_1^{p_j}\right\| \leq \theta_N$, then, using the regularity of f in ξ_2.

We obtain:

$$\left|\hat{v}_{p_j}\left(u_0, \hat{\xi}_1^{p_j}\right) - \tilde{v}_{p_j}\left(u_0, \hat{\xi}_1^{p_j}\right)\right| \leq \theta_N C_{h_1} C_f$$

Convergence of the second stage

Final upper bounding

Finally, let $K = C_f \left(C_{h_1} + \sqrt{1 + C_{h_1}^2} \right)$, we have obtained that $\forall m \in \mathbb{N}^*$:

$$\mathbb{P}\left[\max_{i=1,\ldots,M_N} \left| v\left(u_0, \boldsymbol{\xi}_1^j\right) - \tilde{\mathbf{v}}_i\left(u_0, \hat{\boldsymbol{\xi}}_1^i\right) \right| \geq \boldsymbol{\theta}_N K + \eta_1 \right] \leq A_1 N \left(\frac{D_1 L_1}{\eta_1} \right)^{n_1} e^{-\frac{m\eta_1^2}{16\sigma_1^2}} + \mathbb{P}\left[\mathbf{m}_N < m \right]$$

Which implies $\forall m \in \mathbb{N}^*$:

$$\mathbb{P}\left[\left| \hat{\mathbf{J}}_N - \tilde{\mathbf{J}}_N \right| \geq \boldsymbol{\theta}_N K + \eta_1 \right] \leq A_1 N \left(\frac{DL}{\eta_1} \right)^{n_1} e^{-\frac{m\eta_1^2}{16\sigma_1^2}} + \mathbb{P}\left[\mathbf{m}_N < m \right] \quad (2)$$

Therefore, if hypothesis (14) on the clustering procedure holds, the second stage value converge towards its exact value when the number of drawings grows.

Convergence of the first stage

Define:

$$\hat{z}(u_0) = \frac{1}{N} \sum_{j=1,\dots,N} v\left(u_0, \xi_1^j\right)$$

$$z(u_0) = \mathbb{E}\left[v\left(u_0, \boldsymbol{\xi}_1\right)\right]$$

As with the second stage approximation, the difference between J and $\hat{J}$ can be upper bounded using Shapiro's theorem.

Convergence of the first stage

Hypotheses for the large deviations theorem

Assume:

1. U_0 is of finite diameter D_0;
2. There exists σ_0 s.t.

   $$M_{u_0}(t) \leq e^{\frac{\sigma_0^2 t^2}{2}}, \forall t \in \mathbb{R}, \forall u_0 \in U_0$$

   where

   $$M_{u_0}(t) = \mathbb{E}\left[e^{v(u_0, \boldsymbol{\xi}_1) - z(u_0)}\right]$$

   is the moment generating function of the r.v. $v(u_0, \boldsymbol{\xi}_1) - z(u_0)$;
3. And that v is L_0-Lipschitz in u_0 uniformly in ξ_1.

Convergence of the first stage
Application of the large deviation theorem

From theorem 1 of [Shapiro, 2006], $\exists A_0 \geq 0$, s.t. $\forall \eta_0 > 0$:

$$\mathbb{P}\left[\max_{u_0} \left| z\left(u_0\right) - \hat{z}\left(u_0\right) \right| \geq \eta_0 \right] \leq A_0 \left(\frac{D_0 L_0}{\eta_0} \right)^{n_0} e^{-\frac{N\eta_0^2}{16\sigma_0^2}}$$

Therefore

$$\mathbb{P}\left[\left| \hat{\mathbf{J}}_N - J \right| \geq \eta_0 \right] \leq A_0 \left(\frac{D_0 L_0}{\eta_0} \right)^{n_0} e^{-\frac{N\eta_0^2}{16\sigma_0^2}}$$

Global convergence

By combining this inequality with (2) we get, $\forall m \in \mathbb{N}^*$:

$$\mathbb{P}\left[\left|\tilde{\mathbf{J}}_N - J\right| \geq \boldsymbol{\theta}_N K + \eta_1 + \eta_0\right]$$

$$\leq A_0 \left(\frac{D_0 L_0}{\eta_0}\right)^{n_0} e^{-\frac{N\eta_0^2}{16\sigma_0^2}} + A_1 N \left(\frac{D_1 L_1}{\eta_1}\right)^{n_1} e^{-\frac{m\eta_1^2}{16\sigma_1^2}} + \mathbb{P}\left[\mathbf{m}_N < m\right] \tag{3}$$

Generalizations

Corollary

The generalization to any number of stages is straightforward.

Remark on having more scenario than manageable

It may be frequent that more scenario are available than can be possible to take into account as nodes in the tree. In this case, clustering can be applied to both the first and second stage, with similar constraints.

Generalizations

Corollary

The generalization to any number of stages is straightforward.

Remark on having more scenario than manageable

It may be frequent that more scenario are available than can be possible to take into account as nodes in the tree. In this case, clustering can be applied to both the first and second stage, with similar constraints.

Plan

Example chosen

We apply the previous methodology to the following example:

$$f\left(u_0, u_1\left(\boldsymbol{\xi}_1\right), \boldsymbol{\xi}_1, \boldsymbol{\xi}_2\right) = -\left[u_0\xi_1 + u_1\xi_2 - \left(u_0\xi_1 + u_1\xi_2\right)^2 - u_0^2 - u_1^2\right]$$

With $\boldsymbol{\xi}_1$ following an uniform law on $[1, 2]$ and $\boldsymbol{\xi}_2 = \boldsymbol{\xi}_1 + \varepsilon_2$ with ε_2 following and uniform law on $[-1, 1]$.

Methodologies compared

The trees are build using the same branching factor at each stage.

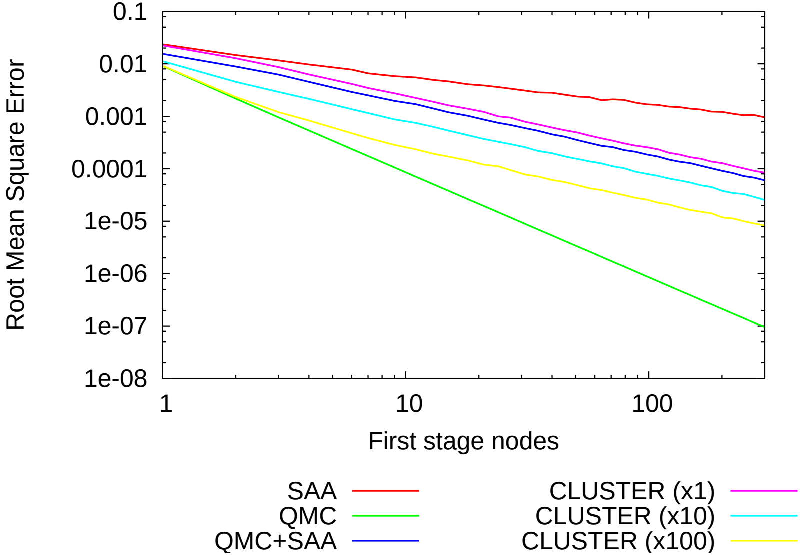
We compare the following methodologies as a function of the number of tree leaves:

1. SAA: drawing of a tree using conditional sampling;
2. QMC: idem SAA, but with low discrepancy drawings;
3. Cluster (x1): tree generation by clustering with as much scenarios as tree leaves. Scenarios are drawn at random;
4. Cluster (x10, x100): idem, but with 10 times more (resp. 100 times more) scenarios than tree leaves.

Clusters are performed using a constant number of scenarios by cluster.

Convergence speed

RMSE as a function of the number of first stage nodes



Convergence speed
Observations

1. SAA convergence is very slow, even for such an easy problem;
2. The clustering procedure is significantly faster, without reaching Quasi Monte Carlo's speed. Remark: the trade-off between first stage and second stage node is not optimized.;
3. Using more scenarios get the clustering procedure nearer to Quasi Monte Carlo's performance, but does not improve the convergence speed.

Conclusion

On tree generation by clustering:

1. Simple and efficient methodology ;
2. Sound theoretical bases ;
3. Interesting properties in practice.

Two future improvements:

1. Take into account information on the cost function into the clustering and the representative element choice ;
2. Take into account the low discrepancy properties of the representative elements.

Conclusion

On tree generation by clustering:

1. Simple and efficient methodology ;
2. Sound theoretical bases ;
3. Interesting properties in practice.

Two future improvements:

1. Take into account information on the cost function into the clustering and the representative element choice ;
2. Take into account the low discrepancy properties of the representative elements.

Bibliography I

- Barty, K. (2004).
  Contributions à la discrétisation des contraintes de mesurabilité pour les problèmes d'optimisation stochastique.
  Thèse de doctorat, École Nationale des Ponts et Chaussées.

- Barty, K., Roy, J.-S., and Strugarek, C. (2005).
  A stochastic gradient type algorithm for closed loop problems.
  Stochastic Programming E-Print Series.

- Casey, M. and Sen, S. (2006).
  The scenario generation algorithm for multistage stochastic linear programming.
  Mathematics of Operations Research.
  to appear.

Bibliography II

Dupačová, J., Consigli, J., and Wallace, S. (2000).
Generating scenarios for multistage stochastic programs.
Annals of Operations Research, 100:25–53.

Hochreiter, R. and Pflug, G. C. (2006).
Financial scenario generation for stochastic multi-stage decision processes as facility location problem.
Annals of Operations Research.
to appear.

Holt, C., Modigliani, F., and Simon, H. (1955).
A linear decision rule for production and employment scheduling.
Management Science, 2(1):1–30.

Høyland, K., Kaut, M., and Wallace, S. (2003).
A heuristic for moment-matching scenario generation.
Computational Optimization and Applications, 24:169–185.

Bibliography III

Høyland, K. and Wallace, S. W. (2001).
Generating scenario trees for multi-stage decision problems.
Management Science, 47:295–307.

Pennanen, T. and Koivu, M. (2005).
Epi-convergent discretizations of stochastic programs via integration quadratures.
Numerische Mathematik, 100:141–163.

Pflug, G. C. (2001).
Scenario tree generation for multiperiod financial optimization by optimal discretization.
Mathematical Programming, 89:251–271.

Römisch, W. and Heitsch, H. (2006).
Scenrario tree modelling for multistage stochastic programs.
Technical Report Preprint 296, DFG Research Center MATHEON "Mathematics for key technologies".

Bibliography IV

Shapiro, A. (2003).
Inference of statistical bounds for multistage stochastic programming problems.
Math. Meth. Oper. Res., 58:57–68.

Shapiro, A. (2006).
On complexity of multistage stochastic programs.
Operations Research Letters, 34:1–8.

Shapiro, A. and Nemirovski, A. (2005).
On complexity of stochastic programming problems.
In Jeyakumar, V. and Rubinov, A., editors, *Continuous Optimization: Current Trends and Applications*, pages 111–144. Springer.

Bibliography V

Thénié, J. and Vial, J.-P. (2004).
Stochastic programming with linear decision rules.
In *Computational Management Science Conference*. University of Geneva - Logilab.