Backward Stochastic Differential Equation : Numerical Results

Majdi Rabia

January 25, 2017

Sommaire

1. Introduction
2. Random Forest : a machine learning method
3. Max leaf nodes analysis (on European call example : expected 7.15)
4. Numerical Analysis
5. Plan next month

Exmple of a Non-linear driver : bid-ask model

$$\begin{cases} f(t, Y_t, Z_t) = -Z_t\theta - rY + (R-r)(Y - \frac{Z_t}{\sigma})^- \\ \theta = \frac{\mu - r}{\sigma} \\ \xi(X_T) = (X_T - K_1) - 2(X_T - K_2) \end{cases}$$

Sommaire

1 Introduction

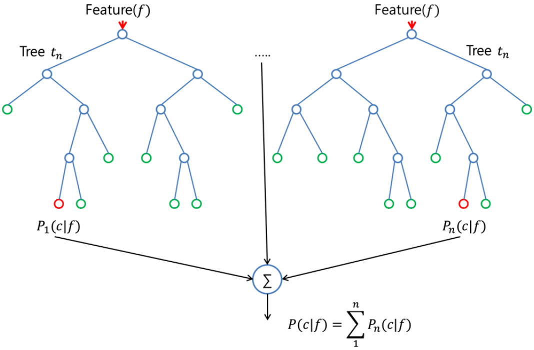
2 Random Forest : a machine learning method
- Overview
- Most sensitive parameters
- Other parameters
- Remarks

3 Max leaf nodes analysis (on European call example : expected 7.15)

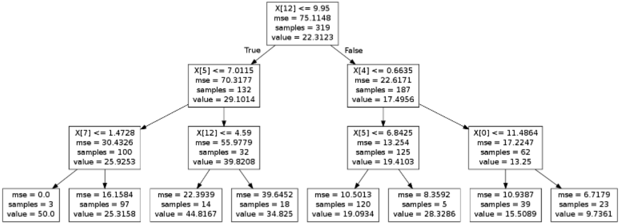
4 Numerical Analysis

5 Plan next month

Overview


Feature(f)
Tree t_n
.....
Feature(f)
Tree t_n
$P_1(c|f)$
$P_n(c|f)$
Σ
$P(c|f) = \sum_1^n P_n(c|f)$

Max leaf nodes



This parameter helps avoid an overfitting by taking bigger values, but reaching a point, this will lead to some inaccuracies(cf following analysis on a specific example)

Accuracy

- **n_estimators** : Represents the number of trees randomly generated. We try to take the biggest value possible in our simulations, as the MSE decreases with this parameter.
- **max_features** : The number of features to consider when looking for the best split:
- **max_depth** : The maximum depth of the tree
- **min_samples_split** : The minimum number of samples required to split an internal node
- **min_samples_leaf** : The minimum number of samples required to be at a leaf node

Performance

- **n_jobs** : The number of jobs to run in parallel for both fit and predict
- **warm_start** : When set to True, reuse the solution of the previous call to fit and add more estimators to the ensemble, otherwise, just fit a whole new forest.

- **Accuracy** : We focused especially on the maximum number of leafs allowed in the Random Forest simulation, which seemed to be the most sensitive parameter in one dimension. However, in higher dimension, results are not bad, but could be improved by constrain on other parameters.
- **Performance** Warm_start parameter could be helpful. From what I understand, it seems to keep in memory the last generated trees, and adding useful other estimators. This leads to better time performance, but I am going to analyse this on simple data to understand exactly how it works. Indeed, it could be used for $\mathbb{E}[Y_{t+dt}|\mathcal{F}_t]$, as it is stable, but might be dangerous for $\mathbb{E}[Y_{t+dt}\Delta B_t|\mathcal{F}_t]$

Sommaire

1. Introduction
2. Random Forest : a machine learning method
3. Max leaf nodes analysis (on European call example : expected 7.15)
4. Numerical Analysis
5. Plan next month

$N = 1000$

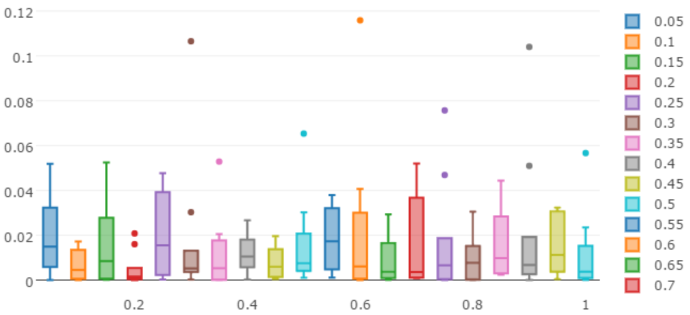


Figure: MSE against $\frac{RF\_max\_leaf\_nodes}{N}$ with N $= 1000$

$N = 10000$

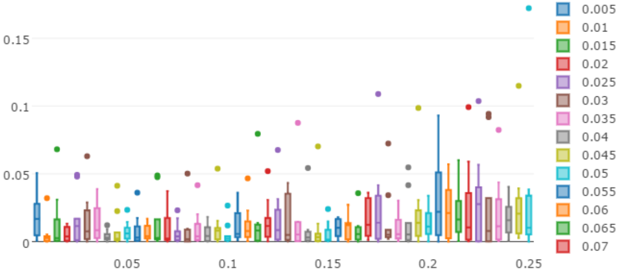


Figure: MSE against $\frac{RF\_max\_leaf\_nodes}{N}$ with N $= 10000$

$N = 10000$

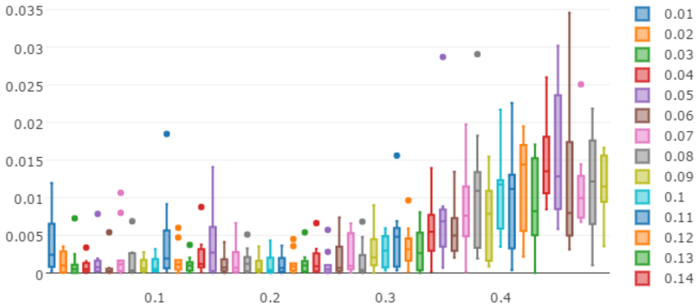


Figure: MSE against $\frac{RF\_max\_leaf\_nodes}{N}$ with N $= 100000$

Remarks

- Getting a clear analysis from the previous three figures is not easy. It seems like there exists a stable region (with high variance) around 10 to 20%.
- We will take in the following simulations 10 and 15%

Sommaire

1 Introduction

2 Random Forest : a machine learning method

3 Max leaf nodes analysis (on European call example : expected 7.15)

4 Numerical Analysis

5 Plan next month

European Call One Dimension : expected = 7.15

	stat parameter	values
RandomForest	mean	7.1891
RandomForest	std	0.0993
RandomForest	95% confidence interval	[7.1794, 7.1988]
RandomForest	min	6.9923
RandomForest	max	7.392
LSM	mean	7.1553
LSM	std	0.0375
LSM	95% confidence interval	[7.1516, 7.159]
LSM	min	7.0907
LSM	max	7.2242
Derivative	mean	7.1779
Derivative	std	0.3088
Derivative	95% confidence interval	[7.1476, 7.2082]
Derivative	min	6.7418
Derivative	max	7.7752
Mesh	mean	7.1875
Mesh	std	0.1387
Mesh	95% confidence interval	[7.1739, 7.2011]
Mesh	min	6.9085
Mesh	max	7.4492

Spread Call One Dimension : expected = 2.95 (Gobet)

	stat parameter	values
RandomForest	mean	2.9694
RandomForest	std	0.0436
RandomForest	95% confidence interval	[2.9651, 2.9737]
RandomForest	min	2.8896
RandomForest	max	3.0422
Derivative	mean	2.9403
Derivative	std	0.1441
Derivative	95% confidence interval	[2.9121, 2.9685]
Derivative	min	2.7534
Derivative	max	3.2832
LSM	mean	2.9381
LSM	std	0.0154
LSM	95% confidence interval	[2.9366, 2.9396]
LSM	min	2.9159
LSM	max	2.9787
Mesh	mean	2.841
Mesh	std	0.0711
Mesh	95% confidence interval	[2.834, 2.848]
Mesh	min	2.6986
Mesh	max	2.9682
GradBoosting	mean	2.9687
GradBoosting	std	0.0465
GradBoosting	95% confidence interval	[2.9641, 2.9733]
GradBoosting	min	2.886
GradBoosting	max	3.0544

European bid-ask Basket Option on 7 assets : expected = 3.30 (Black-Scholes)

	stat parameter	values	time
GradBoost	mean	3.6126	107.8s
GradBoost	std	0.0403	
GradBoost	95% confidence interval	[3.6087, 3.6165]	
GradBoost	min	3.5353	
GradBoost	max	3.6843	
RF	mean	3.4706	313.8s
RF	std	0.0416	
RF	95% confidence interval	[3.4665, 3.4747]	
RF	min	3.4054	
RF	max	3.5668	
derivative	mean	3.3306	43.024s
derivative	std	0.1796	
derivative	95% confidence interval	[3.2602, 3.401]	
derivative	min	3.0466	
derivative	max	3.5348	
mesh	mean	3.4356	125.33
mesh	std	0.0752	
mesh	95% confidence interval	[3.4282, 3.443]	
mesh	min	3.2461	
mesh	max	3.5529	

Remarks

- The one dimensional examples look good after the calibration of Max_leaf_nodes parameter for the spread option
- I am going to check why it does not fit for the simple call example as the difference R - r is smaller ...
- The high dimensional example is not bad only for the Derivative method.
- Moreover, the time spent in Random Forest is way too high. I think it would be good to analyse the tuning of the other parameter for Random Forest ...

Sommaire

1 Introduction

2 Random Forest : a machine learning method

3 Max leaf nodes analysis (on European call example : expected 7.15)

4 Numerical Analysis

5 Plan next month

	finished
	started but not finished
	not started

	Simulations BSDE
1D	Bid-ask european call
	bid-ask call combination
	CVA (does not involve Z)
high dim	bid-ask geometric call
	Simulations 2 -BSDE
	Portfolio optimisation Touzi, warin, fahim

	Code improvement
performance	parallelization (checked how to implement with Deborshee)
	try the warm_start parameter in RF
accuracy	adapt analysis made for max_leaf_nodes to other params

MSc thesis
1) BSDE theory
2) Random Forest Regression
3) RF parameters tuning
4) Mesh Method
5) Z as derivative
6) Numerical Simulations
7) 2-BSDE and non linear PDE
8) Numerical Simulations
Article