Rough Paths in Machine Learning
Assessed Coursework

Benedikt Petko

Department of Mathematics, Imperial College London

Source code used to compute the tasks can be accessed from the repository `https://github.com/benediktpetko/RPML-MATH97229`.

1 Setting

We solve the stochastic differential equation

$$\begin{cases} dY_t &= (1 - Y_t)\, dt + 2Y_t^2 \circ dW_t \\ Y_0 &= 0 \end{cases} \tag{1}$$

numerically using the Milstein scheme [1] in time horizon $[0, T]$.

For time-step $\Delta t := \frac{T}{K}$ define $t_i = i\Delta$ for $i = 0, \ldots, K$ and the approximated solution according to the Milstein scheme modified for Stratonovich integration:

$$\hat{Y}(t_{i+1}) := \hat{Y}(t_i) + (1 - \hat{Y}(t_i))\Delta t + 2\hat{Y}(i)^2 N_i \Delta t + 2\hat{Y}(i)^2 (N_i \Delta t)^2$$

where $\{N_i : i = 0, \ldots, K\}$ are independent standard normal random variables.

Concretely, we choose time horizon $T = 0.25$, $K = 250$ number of steps and $N = 1600$ number of replications. See Figures 1and 2 for sample paths and empirical distribution at terminal time.

2 Comparison: increment features and signature

Here, we implement and compare linear regression of the target variable Y_T with respect to two distinct sets of features based on the driving signal $X_t := (t, W_t)$, or more precisely its increments

$$\left(X_{(i+1)T/K} - X_{iT/K} \right)_{i=0}^{K-1}$$

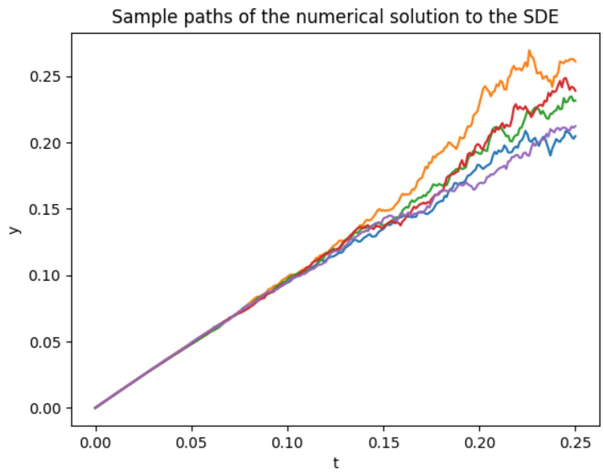

Figure 1: 5 sample paths of the simulated solution to Eq. 1 up to $T = 0.25$ with step size 0.001

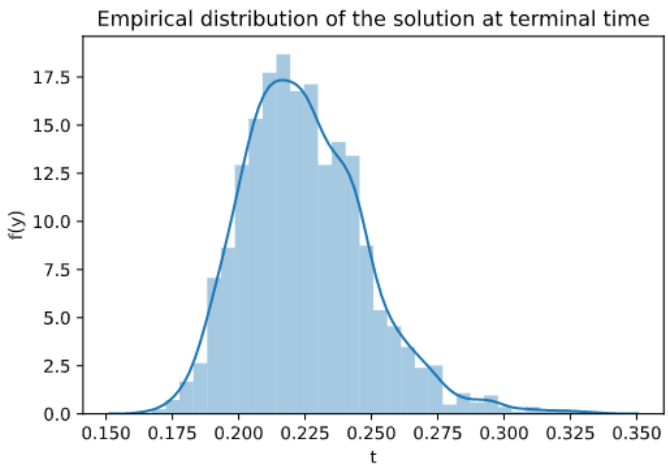

Figure 2: Empirical density of $Y_{0.25}$ based on 1600 replications with step size 0.001

Sampling $N = 1600$ replications, we obtain sample paths of the driving signal,

$$\left(X^{(j)}_{(i+1)T/K} - X^{(j)}_{(iT/K)}\right)_{i=0}^{k} \text{ for } j = 1, \ldots, 1600$$

Splitting these 1600 sample paths into the training set of paths indexed by $j = 1, \ldots, 800$ and test set of paths indexed by $j = 801, ldots, 1600$, we can use the former to fit a linear regression model and the latter to measure out-of-sample goodness of fit criteria. The linear regression can be applied with the following two distinct kinds of features.

- Consider the increments of the driving signal themselves as features in the linear regression problem

$$\hat{Y}_T^{\text{inc}} = a_0 + \sum_{i=1}^{K-1} \sum_{j=1}^{2} a_{i,j} \left(X^j_{(i+1)T/K} - X^j_{iT/K}\right)$$

- Consider the signature of X_t up to level n as features to obtain the formulation

$$\hat{Y}_T^{\text{sig}} = L_0 + \sum_{j=1}^{n} \sum_{i_1 \ldots i_j = 1}^{2} S(X)_{0,T}^{j;i_1 \ldots i_j} L_{j;i_1 \ldots i_j}$$

  In the numerical experiment, the signature components $S(X)_{0,T}^{j;i_1 \ldots i_j}$ were computed by the iisignature package [2] for Python. Note that linear regression over signature features up to level n of the two-dimensional signal X_t entails fitting $2^{n+1} - 2$ parameters.

Table 1 below summarizes the performance of four sets of features, compared with respect to their mean squared error and R^2 scores, both measured on the test set.

Features	Mean squared error (MSE)	R^2 score
Increments	2.048×10^{-5}	0.96
Signature (level 2)	2.711×10^{-5}	0.95
Signature (level 4)	1.747×10^{-7}	0.9997
Signature (level 6)	1.803×10^{-9}	0.999996

Table 1: Comparison of results for 4 sets of features for fixed $K = 250, T = 0.25$ and $N = 1600$ replications.

.

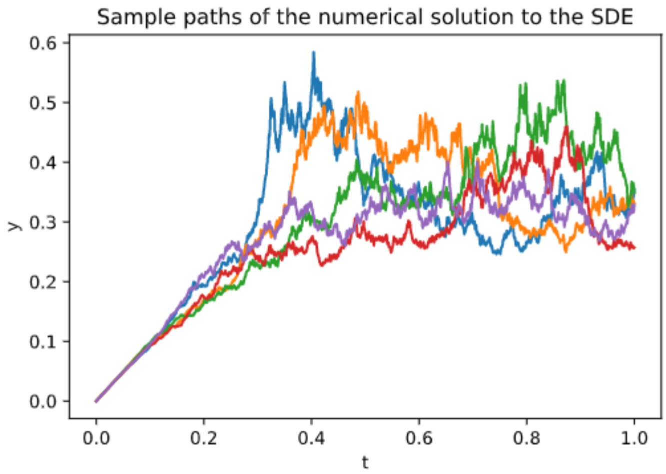

Figure 3: 5 sample paths of the simulated solution to Eq. 1 up to $T = 1$ with step size 0.001.

A remark on long time horizons

In our experiments, the Milstein scheme becomes unstable for time horizons T larger than 0.5. We propose to remedy this problem by discarding paths with mean higher than some cut-off value.

In the following, we chose terminal time $T = 1$ with $K = 1000$ time steps, $N = 20000$ replications and cut-off mean of 0.3. In our numerical experiment, this lead to keeping only $N' = 1738$ paths without singular behaviour and discarding the rest. See Figures 3 and 4 for observed sample paths and empirical distribution at terminal time, and Table 2 for a comparison of four feature sets in terms of the MSE and R^2 score they yield on the test set.

Features	Mean squared error (MSE)	R^2 score
Increments	0.02112	-0.86
Signature (level 2)	0.003331	0.73
Signature (level 4)	0.0001467	0.986
Signature (level 6)	9.146×10^{-5}	0.993

Table 2: Comparison of results for 4 sets of features for fixed $K = 1000, T = 1$ and $N = 1738$ non-singular replications obtained by discarding paths with mean higher than 0.3.

.

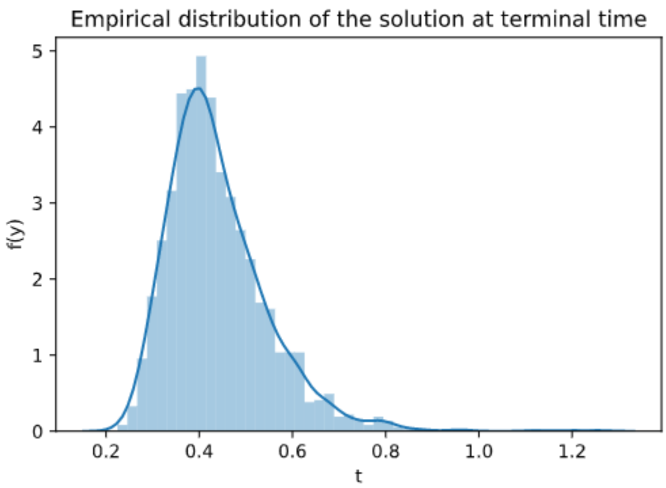

Figure 4: Empirical density of Y_1 based on 1738 replications with step size 0.001

References

[1] David García-Álvarez. *A comparison of a few numerical schemes for theintegration of stochastic differential equations in the Stratonovich interpretation*. Arxiv preprint: https://arxiv.org/abs/1102.4401. 2011.

[2] Jeremy Reizenstein and Benjamin Graham. *The iisignature library: efficient calculation of iterated-integral signatures and log signatures*. Arxiv preprint: https://arxiv.org/abs/1802.08252. 2018.