Akademia Górniczo-Hutnicza
im. Stanisława Staszica w Krakowie

AGH University of Krakow

Deep learning-based estimation of time-dependent parameters in Markov models with application to SDEs - theoretical foundations

Andrzej Kałuża, Paweł M. Morkisz, Bartłomiej Mulewicz, Paweł Przybyłowicz, Martyna Wiącek

AGH University of Krakow, Faculty of Applied Mathematics

14 XI 2023

Outline

- **PROBLEM:**
  approximation of time-dependent parameters (from real data)
- **IDEA:**
  approximation task → optimization problem (maximum likelihood approach)
- We have good tools, because it fits into the **deep learning framework**. We can find appropriate neural network, train it to fit our data and obtain the closest possible values.

[1] O. Dürr, B. Sick, E. Murina, *Probabilistic Deep Learning*, Manning, 2020

Chapter 4 → Building loss functions with the likelihood approach

- **CASE 1:** linear relationship between input and output
- **CASE 2:** non-linear relationship between input and output, const. variance

$$loss = MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

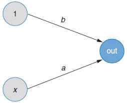


Figure 4.11 Simple linear regression as an fcNN without a hidden layer. This model computes the output directly from the input as out = $a \cdot x + b$.

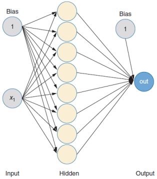


Figure 4.15 Extended linear regression. The eight neurons in the hidden layer give the features from which the output "out" is computed.

Figure: Neural network architectures for case 1 and case 2. Source: [1].

I. Introduction - inspiration

[1] O. Dürr, B. Sick, E. Murina, *Probabilistic Deep Learing*, Manning, 2020
Chapter 4 → Building loss functions with the likelihood approach

- **CASE 3:** non-linear relationship between input and output, nonconstant variance

$$loss = NLL = \sum_{i=1}^{n} -\log\left(\frac{1}{\sqrt{2\pi\sigma_{x_i}^2}}\right) + \frac{(\mu_{x_i} - y_i)^2}{2\sigma_{x_i}{}^2}$$

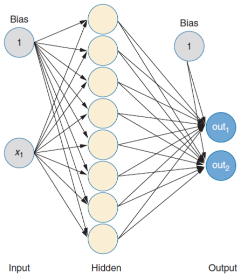


Figure 4.16 You can use an NN with two output nodes to control the parameters μ_x and σ_x of the conditional outcome distribution $N(\mu_x, \sigma_x)$ for regression tasks with nonconstant variance.

Figure: Neural network architectures for case 3. Source: [1].

AGH

- We have $x_1, x_2, \ldots, x_n$ - observed realization of the sequence of random variables $X_1, X_2, \ldots, X_n$,
- We assume (discrete) Markov property, i.e.:

$$\mathbb{P}(X_{k+1} \in A \mid X_1, \ldots, X_k) = \mathbb{P}(X_{k+1} \in A \mid X_k),$$

for all Borel sets A and $k = 0, 1, \ldots, n-1$, where $\mathbb{P}(X_1 \in A \mid X_0) := \mathbb{P}(X_1 \in A)$

- We assume that for all $k = 0, 1, \ldots, n-1$ there exists conditional density $f_{X_{k+1}|X_k}(\cdot\,|x_k)$, with $f_{X_1|X_0}(\cdot\,|x_0) := f_{X_1}(\cdot)$ and $f_{X_{k+1}|X_k} = f_{X_{k+1}|X_k}(\cdot\,|x_k, \Theta)$, $\quad f_{X_{k+1}|X_k} : \mathbb{R}^d \times \mathbb{R}^d \times B([0,T]) \to \mathbb{R}$

- Transition densities might depend on the parameter function $\Theta : [0,T] \to \mathbb{R}^s$ in the functional way, and this dependence might be nonlinear.

- By the Markov property:

$$f_{(X_1,X_2,\ldots,X_n)}(x_1, x_2, \ldots, x_n, \Theta) = \prod_{k=0}^{n-1} f_{X_{k+1}|X_k}(x_{k+1}|x_k, \Theta).$$

AGH

MaxLike approach:

$$\mathcal{L}(\Theta) = -\ln f_{(X_1, X_2, \ldots, X_n)}(x_1, x_2, \ldots, x_n, \Theta)$$

$$= -\sum_{k=0}^{n-1} \ln\left[f_{X_{k+1}|X_k}(x_{k+1}|x_k, \Theta)\right]$$

Aim

$$\Theta^* = \mathrm{argmin}_{\Theta \in B([0,T])} \mathcal{L}(\Theta)$$

Problem: number of optimization parameters - in this setting the problem is infinite dimensional.

NN-MaxLike approach

$$\mathcal{L}(w) = -\sum_{k=0}^{n-1} \ln\left[f_{X_{k+1}|X_k}(x_{k+1}|x_k, \Theta(w))\right],$$

- Θ is modeled by artificial neural network
- customized loss function $\mathcal{L}$ is the crucial part
- weights $w \in \mathbb{R}^N$ and we assume that N is strictly smaller than n

(Better!) Aim

$$w^* = \operatorname{argmin}_{w \in \mathbb{R}^N} \mathcal{L}(w)$$

Number of optimization parameters is $O(N)$, which depends on architecture of NN but is independent of n.

- $x_1, x_2, \ldots, x_n$ - realizations of the $\mathbb{R}^d$-valued random vectors

$$X(t_k) = \mu(t_k) + \varepsilon(t_k), \quad k = 1, 2, \ldots, n \tag{1}$$

- $(\varepsilon(t_k))_k$ – independent random variables $\sim N(0, \Sigma(t_k))$ with $\Sigma(t_k) \in \mathbb{R}^{d \times d}$, $\mu(t_k) \in \mathbb{R}^d$.
- The symmetric covariance matrix $\Sigma(t_k)$ – strictly positive definite for all $k = 1, 2, \ldots, n$.

$$\Theta(t) = [\mu(t), \Sigma(t)], \qquad \Theta(t, w) = [\mu(t, w)), \Sigma(t, w)]$$

$(X(t_k))_k$ – independent and normally distributed, hence for $x \in \mathbb{R}^d$.

$$f_{X_k|X_{k-1}}(x|x_{k-1}, \Theta) = f_{X_k}(x|\Theta) = (2\pi)^{-d/2}(\det(\Sigma(t_k)))^{-1/2} \times \exp\Big(-\frac{1}{2}(x - \mu(t_k))^T \Sigma(t_k)^{-1}(x - \mu(t_k))\Big). \tag{2}$$

AGH

NN-MaxLike approach

- We approximate the parameter function $\Theta = [\mu, \Sigma]$ via the neural network $\Theta(w) = [\mu(w), \Sigma(w)]$.

- The loss function has the following form

$$\mathcal{L}(w) = \frac{1}{2}\sum_{k=1}^{n}\Bigg[\ln\Big[(2\pi)^d \cdot \det(\Sigma(t_k, w))\Big] + (x_k - \mu(t_k, w))^T \cdot \Sigma^{-1}(t_k, w) \cdot (x_k - \mu(t_k, w))\Bigg],$$

www.agh.edu.pl

Step 1. Synthetic data generation - all parameters time-dependent.

- We generate data according to the distribution:

$$X(t_k) = \mu(t_k) + \varepsilon(t_k), \quad k = 0, 1, 2, \dots, n \tag{3}$$

  where $\mu(t_k) = [\mu_1(t_k), \mu_2(t_k)]$, $\varepsilon(t_k)$ are independent random variables with the distribution $\mathcal{N}(0, \Sigma(t_k))$, and $t_k = k\frac{2\pi}{n}$, $n = 3000$.
- $X_1 = 0.5 + \sin(t)$, $X_2 = \cos(t)$, $\sigma_1 = 0.1$, $\sigma_2 = 0.15$, $\varrho = 0.5$
- The covariance matrix is as follows

$$\Sigma(t) = \begin{bmatrix} \sigma_1^2|\mu_1(t)| & \varrho\sigma_1\sigma_2|\mu_1(t)|\cdot|\mu_2(t)| \\ \varrho\sigma_1\sigma_2|\mu_1(t)|\cdot|\mu_2(t)| & \sigma_2^2|\mu_2(t)| \end{bmatrix}$$

- $\sigma_1(t) = \sigma_1\sqrt{\mu_1(t)}$, $\sigma_2(t) = \sigma_2\sqrt{\mu_2(t)}$, $\varrho(t) = \varrho\sqrt{|\mu_1(t)|}\sqrt{|\mu_2(t)|}$ with $\varrho = 0.5$.
- So the modelled parameter

$$\Theta(t) = [\mu_1(t), \mu_2(t), \sigma_1(t), \sigma_2(t), \varrho(t)]\,.$$

Step 2. Creating and training the neural network model.

Custom loss function - the accurate formula or $d = 2$:

$$\mathcal{L}(w) = \sum_{i=1}^{n} \Bigg[\ln\left(2\pi\sigma_1(t_i,w)\sigma_1(t_i,w)\sqrt{1-\varrho^2(t_i,w)} \right) + \frac{1}{2(1-\varrho^2(t_i,w))}\Bigg(\left(\frac{x_i^{(1)}-\mu_1(t_i,w)}{\sigma_1(t_i,w)}\right)^2 + \left(\frac{x_i^{(2)}-\mu_2(t_i,w)}{\sigma_2(t_i,w)}\right)^2 - 2\varrho(t_i,w)\frac{(x_i^{(1)}-\mu_1(t_i,w))(x_i^{(2)}-\mu_2(t_i,w))}{\sigma_1(t_i,w)\sigma_2(t_i,w)} \Bigg) \Bigg].$$

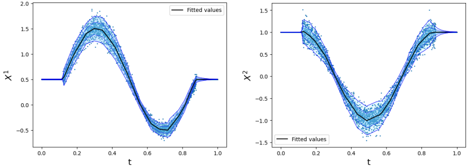


Figure: Fitted values of μ_1 and μ_2 for X_1 and X_2,respectively, together with 68% and 95% prediction intervals

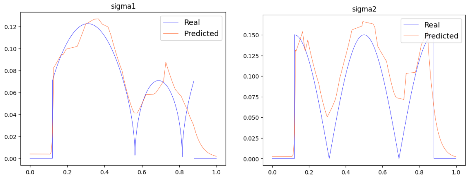


Figure: Parameters reproduced by neural network

IV. Application 2 - SDEs parameters estimation

- SDEs are used for multiple modeling applications - finances, energy consumption, etc.
- We typically don't have access to the SDEs coefficients but just a single observed trajectory
- The typical modeling of real phenomena has two crucial parts - one is choosing the right model, and the second is estimating its parameters (calibration).
- Proposed approach enables us to approximate the underlying SDE based on the single trajectory by using deep learning framework. We only need to assume the type of the SDE, e.g. Black - Scholes.

Calibration of SDEs

$(\Omega, \Sigma, (\Sigma_t)_{t\geq 0}, \mathbb{P}),\ T > 0,\ (W(t))_{t\geq 0},$

$$\begin{cases} dX(t) = a(t, X(t), \Theta(t))dt + b(t, X(t), \Theta(t))dW(t),\ t \in [0, T], \\ X(0) = x_0 \end{cases} \tag{4}$$

Aim

Having discrete observations $x_0, x_1, \ldots, x_n$ of X at $t_i = iT/n$, $i = 0, 1, \ldots, n$, estimate the parameter function $\Theta : [0, T] \to \mathbb{R}^s$.

Euler scheme: $h = T/n$, $t_k = kh$, $\Delta W_k = W(t_{k+1}) - W(t_k)$

$$
\begin{aligned}
&X^E(t_0) = x_0 \\
&X^E(t_{k+1}) = X^E(t_k) + a(t_k, X^E(t_k), \theta(t_k))h + b(t_k, X^E(t_k), \theta(t_k))\Delta W_k, \\
&k = 0, 1, \ldots, n-1
\end{aligned}
$$

$(X(t_k))_{k=0,1,\ldots,n}$, $(X^E(t_k))_{k=0,1,\ldots,n}$ - discrete Markov processes

AGH

- For $k = 0, 1, \ldots, n-1$ we have the transition density for Euler scheme:

$$f_{X^E(t_{k+1})|X^E(t_k)}(x|x_k, \Theta) = (2\pi)^{-d/2} \det\Big(h(b \cdot b^T)(t_k, x_k, \Theta(t_k))\Big)^{-\frac{1}{2}}$$

$$\times \exp\Big[-\frac{1}{2h}(x - x_k - a(t_k, x_k, \Theta(t_k))h)^T (b \cdot b^T)^{-1}(t_k, x_k, \Theta(t_k))(x - x_k - a(t_k, x_k, \Theta(t_k))h)\Big]$$

for $x = [x^1, \ldots, x^d] \in \mathbb{R}^d$, $x_k = [x_k^1, \ldots, x_k^d] \in \mathbb{R}^d$.

- Quasi-MaxLike:

$$\mathcal{L}(\Theta) = -\sum_{k=0}^{n-1} \ln\left[f_{X^E(t_{k+1})|X^E(t_k)}(x_{k+1}|x_k, \Theta)\right] =$$

$$\frac{1}{2}\sum_{k=0}^{n-1}\left[\ln\Big[(2\pi)^d \det\Big(h(b \cdot b^T)(t_k, x_k, \Theta(t_k))\Big)\Big]\right.$$

$$\left. + \frac{1}{h}(\Delta x_k - a(t_k, x_k, \Theta(t_k))h)^T (b \cdot b^T)^{-1}(t_k, x_k, \Theta(t_k))(\Delta x_k - a(t_k, x_k, \Theta(t_k))h)\right].$$

Number of unknown parameters is $O(n)$, so use NN!

NN-Quasi-MaxLike approach:
The loss function for the neural network:

$$\tilde{\mathcal{L}}(w) = \frac{1}{2}\sum_{k=0}^{n-1}\Bigg[\ln\Big[(2\pi)^d \det\Big(h(b\cdot b^T)(t_k, x_k, \Theta(t_k, w))\Big)\Big]$$

$$+\frac{1}{h}\Big(\Delta x_k - a(t_k, x_k, \Theta(t_k, w))h\Big)^T (b\cdot b^T)^{-1}(t_k, x_k, \Theta(t_k, w))\Big(\Delta x_k - a(t_k, x_k, \Theta(t_k, w))h\Big)\Bigg],$$

where $\Delta x_k = x_{k+1} - x_k$.

Aim

$$w^* = \operatorname{argmin}_{w\in\mathbb{R}^N}\tilde{\mathcal{L}}(w)$$

EXAMPLE - Ornstein-Uhlenbeck proces with constant θ_1 and varying $\theta_2 = \theta_2(t)$

$$dX(t) = -\theta_1 X(t)\,\mathrm{d}t + \theta_2(t)\,\mathrm{d}W(t), \tag{5}$$

where $\theta_1 > 0$ and $\theta_2 : [0, T] \to [0, +\infty)$, $\Theta = [\theta_1, \theta_2]$.

Transition density:

$$f_{X(t_{k+1})|X(t_k)}(x|x_k, \Theta) = \left(2\pi \cdot \int_{t_k}^{t_{k+1}} \theta_2^2(u) e^{-2\theta_1 \cdot (t_{k+1}-u)}\,\mathrm{d}u\right)^{-1/2} \times \exp\left[-\frac{(x - e^{-\theta_1 \cdot h} \cdot x_k)^2}{2\int_{t_k}^{t_{k+1}} \theta_2^2(u) \cdot e^{-2\theta_1 \cdot (t_{k+1}-u)}\,\mathrm{d}u}\right]. \tag{6}$$

$$
\begin{cases} X(t,w) = a(t,X(t),\Theta(t))dt + b(t,X(t),\Theta(t))dW(t), \ \ t \in [0,T], \\ X(0,w) = x_0 \end{cases} \tag{7}
$$

$a : [0, T] \times \mathbb{R}^d \times \mathbb{R}^s \to \mathbb{R}^d$
$b : [0, T] \times \mathbb{R}^d \times \mathbb{R}^s \to \mathbb{R}^{d\times m}$
$\Theta : [0, T] \to \mathbb{R}^s$

- By $\|\cdot\|$ we mean the Euclidean norm in $\mathbb{R}^d$ or $\mathbb{R}^s$, or the Frobenius norm in $\mathbb{R}^{d\times m}$ (the meaning is clear from the context).
- For $\varrho \geq 0$ we denote by $B(0,\varrho) = \{x \in \mathbb{R}^d \mid \|x\| \leq \varrho\}$.
- For $H \in \mathbb{R}$ we take $H_+ = \max\{H, 0\}$.

We impose the following *Krylov-type* **conditions**.

Assumptions

(A0) $x_0 \in \mathbb{R}^d$,

(A1) a is Borel measurable and for all $(t,z) \in [0,T] \times \mathbb{R}^s$ the function $a(t,\cdot,z) : \mathbb{R}^d \to \mathbb{R}^d$ is continuous,

(A2) there exists $H \in \mathbb{R}$ such that for all $t \in [0,T], x \in \mathbb{R}^d, y \in \mathbb{R}^d, z \in \mathbb{R}^s$ it holds

$$< x - y, a(t,x,z) - a(t,y,z) > \leq H(1 + \|z\|)\|x - y\|^2,$$

(A3) for all $\varrho \in [0,+\infty)$ it holds

$$\sup_{(t,x)\in[0,T]\times B(0,\varrho)} \|a(t,x,0)\| < +\infty,$$

(A4) there exist $\alpha_1 \in (0,1], q \in [1,+\infty), L \in [0,+\infty)$ such that for all $t \in [0,T], x \in \mathbb{R}^d, z_1 \in \mathbb{R}^s, z_2 \in \mathbb{R}^s$ the following holds

$$\|a(t,x,z_1) - a(t,x,z_2)\| \leq L(1 + \|x\|^q)\|z_1 - z_2\|^{\alpha_1},$$

Assumptions

(B1) b is a Borel measurable function,

(B2) there exists $L \in [0, +\infty)$ such that for all $t \in [0, T], x \in \mathbb{R}^d, y \in \mathbb{R}^d, z \in \mathbb{R}^s$ it holds

$$\|b(t,x,z) - b(t,y,z)\| \leq L(1 + \|z\|) \cdot \|x - y\|,$$

(B3)

$$\sup_{0 \leq t \leq T} \|b(t,0,0)\| < +\infty,$$

(B4) there exist $\alpha_2 \in (0,1], L \in [0, +\infty)$ such that $t \in [0, T], x \in \mathbb{R}^d, z_1 \in \mathbb{R}^s, z_2 \in \mathbb{R}^s$

$$\|b(t,x,z_1) - b(t,x,z_2)\| \leq L(1 + \|x\|) \cdot \|z_1 - z_2\|^{\alpha_2},$$

and finally we assume that

(T1) Θ is Borel measurable and

$$\sup_{0 \leq t \leq T} \|\Theta(t)\| < +\infty.$$

AGH

Theorem 1.

Let us assume that $\Theta_i : [0, T] \to \mathbb{R}^s$, $i = 1, 2$, satisfy the assumption $(T1)$, $a : [0, T] \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ satisfies the assumptions (A1)-(A4), and $b : [0, T] \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^{d \times m}$ satisfies the assumptions (B1)-(B4). Then

$$\sup_{0 \le t \le T} \left(\mathbb{E} \| X_1(t) - X_2(t) \|^2 \right)^{1/2} \le C \Big(\| \Theta_1 - \Theta_2 \|_\infty^{\alpha_1} + \| \Theta_1 - \Theta_2 \|_\infty^{\alpha_2} \Big), \tag{8}$$

where $X_i = X(x_0, a, b, \Theta_i)$, $i = 1, 2$, is the unique strong solution of

$$\begin{cases} \mathrm{d}X_i(t) = a(t, X_i(t), \Theta_i(t))\,\mathrm{d}t + b(t, X_i(t), \Theta_i(t))\,\mathrm{d}W(t), & t \in [0, T], \\ X_i(0) = x_0, \end{cases} \tag{9}$$

and $C = \max\{C_1, C_2\}$.

We have the following result on existence and uniqueness of strong solution to (7).

Fact 1

Under the assumptions (A0)-(A4), (B1)-(B4), (T1) the SDE (7) has unique strong solution $X = X(x_0, a, b, \Theta)$*, such that for all* $p \in [2, +\infty)$

$$\mathbb{E}\Big(\sup_{0\leq t\leq T} \|X(t)\|^p\Big) < +\infty. \tag{10}$$

We stress that C in Theorem 1. does not depend on X_1. Therefore, if $\|\Theta_1 - \Theta_2\|_\infty \to 0$ (i.e. $\Theta_1 \to \Theta_2$ uniformly on $[0, T]$) then

$$\sup_{0 \leq t \leq T} \left(\mathbb{E} \| X(x_0, a, b, \Theta_1)(t) - X(x_0, a, b, \Theta_2)(t) \|^2 \right)^{1/2} \to 0, \tag{11}$$

and, in particular, $X(x_0, a, b, \Theta_1)(T) \to X(x_0, a, b, \Theta_2)(T)$ in law.

Lemma 2

(Gronwall's lemma with quadratic term, [5]) Let $\varphi : [0, T] \to [0, +\infty)$ *be a bounded Borel measurable function and let there exist* $a, \alpha, \beta \in [0, +\infty)$ *such that for all* $t \in [0, T]$

$$\varphi^2(t) \leq a + 2\alpha \int_0^t \varphi(s)\,\mathrm{d}s + 2\beta \int_0^t \varphi^2(s)\,\mathrm{d}s. \tag{12}$$

Then for all $t \in [0, T]$

$$\varphi(t) \leq \sqrt{a} \cdot e^{\beta t} + \alpha \cdot \frac{e^{\beta t} - 1}{\beta}. \tag{13}$$

PROOF

Let us denote by $\tilde{a}_i(t,x) = a(t,x,\Theta_i(t))$, $\tilde{b}_i(t,x) = b(t,x,\Theta_i(t))$, $i = 1,2$. Then for all $t \in [0,T]$

$$Z(t) = X_1(t) - X_2(t) = \int_0^t F(s)\,\mathrm{d}s + \int_0^t G(s)\,\mathrm{d}W(s), \tag{14}$$

where $F(s) = \tilde{a}_1(s, X_1(s)) - \tilde{a}_2(s, X_2(s))$, $G(s) = \tilde{b}_1(s, X_1(s)) - \tilde{b}_2(s, X_2(s))$. From the Itô formula (version from [5], Corollary 2.18) we get that

$$\|Z(t)\|^2 = \int_0^t \Big(2\langle Z(s), F(s)\rangle + \|G(s)\|^2\Big)\,\mathrm{d}s + 2\int_0^t \langle Z(s), G(s)\,\mathrm{d}W(s)\rangle, \tag{15}$$

where

$$\int_0^t \langle Z(s), G(s)\,\mathrm{d}W(s)\rangle = \sum_{k=1}^{d}\sum_{j=1}^{m}\int_0^t Z^k(s)\cdot G^{kj}(s)\,\mathrm{d}W^j(s). \tag{16}$$

We can show that

$$\int_0^T \mathbb{E}\Big[\|Z(t)\|^2 \cdot \|G(t)\|^2\Big]\,\mathrm{d}t \leq 2TL^2(1+\|\Theta_1\|_\infty^2)\cdot\mathbb{E}[\|Z\|_\infty^4]$$

$$+4TL^2(\|\Theta_1\|_\infty+\|\Theta_2\|_\infty)^{2\alpha_2}\cdot\Big(\mathbb{E}[\|Z\|_\infty^2]+\mathbb{E}[\|X_2\|_\infty^2\cdot\|Z\|_\infty^2]\Big) < +\infty,$$

using:

- Assumptions (B2), (B4)
- Auxiliary inequality (based on Fact 1 and Hölder's inequality)

$$\mathbb{E}[\|X_2\|_\infty^2\cdot\|Z\|_\infty^2] \leq \Big(\mathbb{E}[\|X_2\|_\infty^4]\Big)^{1/2}\cdot\Big(\mathbb{E}[\|Z\|_\infty^4]\Big)^{1/2} < +\infty. \qquad (17)$$

Hence, for all $t \in [0, T]$ $\mathbb{E}\int_0^t \langle Z(s), G(s)\,\mathrm{d}W(s)\rangle = 0$ and thus

$$\mathbb{E}\|Z(t)\|^2 = 2\mathbb{E}\int_0^t \langle Z(s), F(s)\rangle\,\mathrm{d}s + \int_0^t \mathbb{E}\|G(s)\|^2\,\mathrm{d}s. \tag{18}$$

For all $t \in [0, T]$

$$\|G(t)\| \leq L(1 + \|\Theta_1\|_\infty)\cdot\|Z(t)\| + L(1 + \|X_2\|_\infty)\cdot\|\Theta_1 - \Theta_2\|_\infty^{\alpha_2}, \tag{19}$$

and hence

$$\int_0^t \mathbb{E}\|G(s)\|^2\,\mathrm{d}s \leq 2L^2(1 + \|\Theta_1\|_\infty)^2\cdot\int_0^t \mathbb{E}\|Z(s)\|^2\,\mathrm{d}s$$
$$+4L^2(1 + \mathbb{E}[\|X_2\|_\infty^2])\cdot\|\Theta_1 - \Theta_2\|_\infty^{2\alpha_2}. \tag{20}$$

Moreover, for all $t \in [0, T]$, by the assumptions (A2), (A4) and Cauchy-Schwarz inequality

$$
\begin{aligned}
&\langle Z(t), F(t)\rangle = \langle X_1(t) - X_2(t), a(t, X_1(t), \Theta_1(t)) - a(t, X_2(t), \Theta_2(t))\rangle \\
&= \langle X_1(t) - X_2(t), a(t, X_1(t), \Theta_1(t)) - a(t, X_2(t), \Theta_1(t))\rangle \\
&\qquad + \langle X_1(t) - X_2(t), a(t, X_2(t), \Theta_1(t)) - a(t, X_2(t), \Theta_2(t))\rangle \\
&\leq H(1 + \|\Theta_2(t)\|) \cdot \|X_1(t) - X_2(t)\|^2 \\
&\qquad + |\langle X_1(t) - X_2(t), a(t, X_2(t), \Theta_1(t)) - a(t, X_2(t), \Theta_2(t))\rangle| \\
&\leq H_+(1 + \|\Theta_2\|_\infty) \cdot \|Z(t)\|^2 + \|Z(t)\| \cdot \|a(t, X_2(t), \Theta_1(t)) - a(t, X_2(t), \Theta_2(t))\| \\
&\leq H_+(1 + \|\Theta_2\|_\infty) \cdot \|Z(t)\|^2 + L(1 + \|X_2\|_\infty^q) \cdot \|\Theta_1 - \Theta_2\|_\infty^{\alpha_1} \cdot \|Z(t)\|. \qquad (21)
\end{aligned}
$$

Therefore, by the Hölder inequality we have for all $t \in [0, T]$

$$\mathbb{E}\int_0^t \langle Z(s), F(s)\rangle \, \mathrm{d}s \leq H_+(1 + \|\Theta_2\|_\infty) \cdot \int_0^t \mathbb{E}\|Z(s)\|^2 \, \mathrm{d}s$$

$$+L\|\Theta_1 - \Theta_2\|_\infty^{\alpha_1} \cdot \int_0^t \mathbb{E}\Big[\|Z(s)\| \cdot (1 + \|X_2\|_\infty^q)\Big] \, \mathrm{d}s$$

$$\leq H_+(1 + \|\Theta_2\|_\infty) \cdot \int_0^t \mathbb{E}\|Z(s)\|^2 \, \mathrm{d}s$$

$$+L\|\Theta_1 - \Theta_2\|_\infty^{\alpha_1} \cdot \Big(\mathbb{E}(1 + \|X_2\|_\infty^q)^2\Big)^{1/2} \cdot \int_0^t \Big(\mathbb{E}\|Z(s)\|^2\Big)^{1/2} \, \mathrm{d}s \quad (22)$$

Combining (20), (18), (22) we get for all $t \in [0, T]$ that

$$\varphi^2(t) \leq a + 2\alpha \int_0^t \varphi(s)\,\mathrm{d}s + 2\beta \int_0^t \varphi^2(s)\,\mathrm{d}s,$$

where $\varphi(t) = \left(\mathbb{E}\|Z(t)\|^2\right)^{1/2}$ is bounded Borel measuarable function while

$$a = 4TL^2(1 + \mathbb{E}[\|X_2\|_\infty^2]) \cdot \|\Theta_1 - \Theta_2\|_\infty^{2\alpha_2} \tag{23}$$

$$\alpha = L\|\Theta_1 - \Theta_2\|_\infty^{\alpha_1} \cdot \left(\mathbb{E}(1 + \|X_2\|_\infty^q)^2\right)^{1/2} \tag{24}$$

$$\beta = H_+(1 + \|\Theta_2\|_\infty) + L^2(1 + \|\Theta_1\|_\infty)^2. \tag{25}$$

Applying Lemma 2 (Gronwall with quadratic term) we get the thesis. □

Assumptions - discontinuous drift

In the following scalar case ($m = d = 1$) with additive noise

$$\begin{cases} \mathrm{d}X(t) = a(t, X(t))\,\mathrm{d}t + \Theta(t)\,\mathrm{d}W(t), \ t \in [0, T], \\ X(0) = x_0, \end{cases} \tag{26}$$

we consider the following *Zvonkin-type* for $a : [0, T] \times \mathbb{R} \to \mathbb{R}$ and $\Theta : [0, T] \to \mathbb{R}$:

(C0) $x_0 \in \mathbb{R}$,

(C1) a is Borel measurable,

(C2) there exists $H \in \mathbb{R}$ such that for all $t \in [0, T]$, $x, y \in \mathbb{R}$ it holds

$$(x - y)(a(t, x) - a(t, y)) \leq H|x - y|^2, \tag{27}$$

(C3) there exists $D \in [0, +\infty)$ such that for all $t \in [0, T], x \in \mathbb{R}$ it holds

$$|a(t, x)| \leq D, \tag{28}$$

(D1) Θ is Borel measurable,

(D2) there exist $d_0, d_1 \in (0, +\infty)$ such that for all $t \in [0, T]$

AGH

www.agh.edu.pl

Theorem 2 - discontinuous drift

Let us assume that $\Theta_i : [0, T] \to \mathbb{R}$, $i = 1, 2$, satisfy the assumptions $(D1), (D2)$, and $a : [0, T] \times \mathbb{R} \to \mathbb{R}$ satisfies the assumptions (C1)-(C3). Then

$$\sup_{0 \le t \le T} \left(\mathbb{E} \|X_1(t) - X_2(t)\|^2 \right)^{1/2} \le e^{TH_+} \|\Theta_1 - \Theta_2\|_{L^2[0,T]}, \tag{30}$$

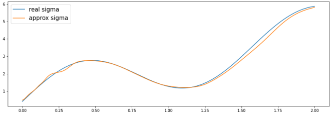
where $X_i = X(x_0, a, \Theta_i)$, $i = 1, 2$, is the unique strong solution of

$$\begin{cases} \mathrm{d}X_i(t) = a(t, X_i(t))\,\mathrm{d}t + \Theta_i(t)\,\mathrm{d}W(t), \ t \in [0, T], \\ X_i(0) = x_0. \end{cases} \tag{31}$$

Example 1 - Ornstein-Uhlenbeck process

$$dX(t) = \kappa(\mu - X(t))dt + \sigma(t)dW(t) \tag{32}$$

$\kappa = 2$, $\mu = 0.5$, $\sigma(t) = 2t + 0.4 + 1.5 \cdot \sin(4t)$.
Modelled parameter: $\theta(t) = \sigma(t)$.



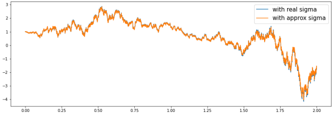


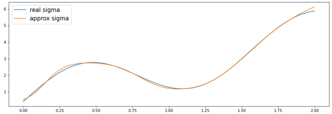
Figure: Parameter $\sigma(t)$ for $t \in [0, T]$ estimated with neural network and trajectories of the process with real and approximated values. Generated by the Euler-Maruyama scheme with the step-size $h = 0.0002$.

Example 2 - discontinuous Ornstein-Uhlenbeck process

We consider the following SDE (also known as threshold diffusion)

$$dX(t) = \Big(\mu - \kappa \cdot sign(X(t))\Big) dt + \sigma(t) dW(t), \qquad (33)$$

$\kappa = 2$, $\mu = 0.5$, $\sigma(t) = 2t + 0.4 + 1.5 \cdot \sin(4t)$.
Modelled parameter: $\theta(t) = \sigma(t)$.



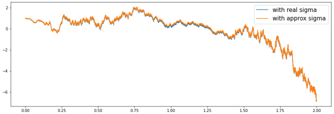


Figure: Parameter $\sigma(t)$ for $t \in [0, T]$ estimated with neural network and trajectories of the process with real and approximated values. Generated by the Euler-Maruyama scheme with the step-size $h = 0.0002$.

Example 3 - SDE with nonlinear coefficients

$$dX(t) = \kappa \cdot \cos(X(t))dt + \Big(\big(\sin(X(t)) + 1.5\big) \cdot \sigma(t) + 2\Big)dW(t), \quad (34)$$

$\kappa = 0.4$ and $\sigma(t) = 2 \cdot \sin(2\pi t) + t$.
Modelled parameter: $\theta(t) = \sigma(t)$, κ is known to the algorithm.

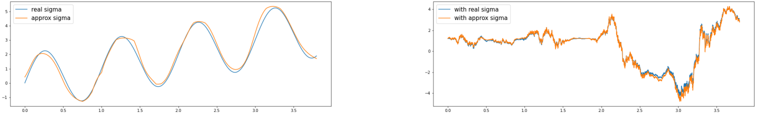


Figure: Parameter $\sigma(t)$ for $t \in [0, T]$ estimated with neural network and trajectories of the process with real and approximated values. Generated by the Euler-Maruyama scheme with the step-size $h = 0.00038$.

Evaluation of the results - histograms and qq-plots

To evaluate our methodology, we need to compare the trajectories generated with two sets of coefficients - real and approximated by the neural network. Proposed approach is to simulate multiple trajectories ($N = 1000$) for each option and compare both distributions of value $X(T)$. Our aim is to verify that the distributions are similar.

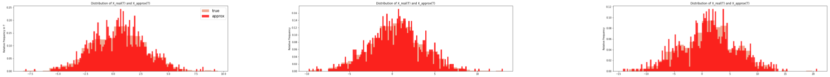


Figure: Histogram of values for real $X(T)$ and its approximation with computed $\sigma(t)$.

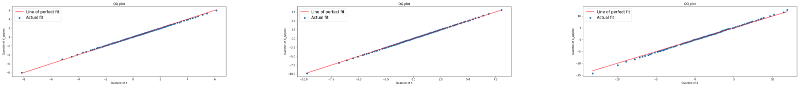


Figure: QQ plots - comparison of both distributions.

	Example 1		Example 2		Example 3	
	X real	X approx	X real	X approx	X real	X approx
empirical mean	0.515	0.515	0.495	0.497	1.243	1.256
empirical std dev	2.456	2.388	3.168	3.202	4.835	5.077

Table: Empirical distribution values for each example - mean and standard deviation

Kolmogorov-Smirnov test with the null hypothesis that both the distributions P, Q are identical based on statistcs $D_{n,m}$

$$D_{n,m} = max|F_P(x) - F_Q(x)| \tag{35}$$

$n = m = 1000$
F_P is CDF of P and F_Q CDF of Q

	Example 1	Example 2	Example 3
KS statistics	0.015	0.009	0.028
p-value of KS test	0.99987	0.99999	0.82821

Table: Comparison of distributions metrics for all the examples

$$D_{n,m} = max|F_P(x) - F_Q(x)| \tag{36}$$

No reason to reject null hypothesis for E1 and E2 ($\alpha = 0.05$).

Bibliography I

- A. Kałuża, P. M. Morkisz, B. Mulewicz, P. Przybyłowicz, M. Wiącek, *Deep learning-based estimation of time-dependent parameters in Markov models with application to nonlinear regression and SDEs*, IN PREPARATION
- O. Dürr, B. Sick, E. Murina, *Probabilistic Deep Learning with Python, Keras and TensorFlow Probability*, Manning, 2020.
- S.N. Cohen, R.J. Elliot, *Stochastic calculus and applications, 2nd ed.*, Probability and its applications, New York, Birkhäuser, 2015.
- R. Situ, *Theory of Stochastic Differential Equations with Jumps and Applications*, Springer Science+Business Media, 2005.
- E. Pardoux, E. Rascanu, *Stochastic Differential Equations, Backward SDEs, Partial Differential Equations*, Springer International Publishing Switzerland, 2014

Thank you for your attention!

More details on machine learning aspects tomorrow by
Pawel Morkisz – Machine Learning Seminar, 15:15
https://www.ida.liu.se/research/machinelearning/seminars/

martynawiacek@agh.edu.pl
morkiszp@agh.edu.pl