APPROXIMATE LIKELIHOOD CONSTRUCTION FOR ROUGH DIFFERENTIAL EQUATIONS

ANASTASIA PAPAVASILIOU AND KASIA TAYLOR

Abstract. The paper is split in two parts: first, we construct the exact likelihood for a discretely observed rough differential equation, driven by a piecewise linear path. In the second part, we use this likelihood in order to construct an approximation to the likelihood for a discretely observed rough differential equation. Finally, We show that the approximation error disappears as the sampling frequency goes to zero.

1. Setting and Main Ideas

In the first part of the paper, we consider the following type of differential equations

$$dY_t^{\mathcal{D}} = a(Y_t^{\mathcal{D}};\theta)dt + b(Y_t^{\mathcal{D}};\theta)dX_t^{\mathcal{D}}, \quad Y_0 = y_0,\ t \le T, \tag{1}$$

where $X^{\mathcal{D}}$ is a realisation of a random piecewise linear path in $\mathbb{R}^m$ corresponding to partition $\mathcal{D}$ of $[0,T]$. We also assume that $\theta \in \Theta$, where Θ is the parameter space. Moreover, we request that $a(\cdot,\theta) : \mathbb{R}^d \to \mathbb{R}^d$ and $b(\cdot,\theta) : \mathbb{R}^d \to L(\mathbb{R}^m,\mathbb{R}^d)$ are Lip(1), which are sufficient conditions for the existence and uniqueness of the solution $Y^{\mathcal{D}}$, which is a bounded variation path on $\mathbb{R}^d$.

We will use I_θ to denote the Itô map defined by (1). That is, I_θ maps the path $X^{\mathcal{D}}$ to the path $Y^{\mathcal{D}}$ and we write

$$Y^{\mathcal{D}} = I_\theta(X^{\mathcal{D}}).$$

First, we develop a framework for performing statistical inference for differential equation (1), assuming that we know the distribution of $X^{\mathcal{D}}$. More precisely, we will aim to construct the likelihood of discrete observations of $Y^{\mathcal{D}}$ on the grid $\mathcal{D}$, which we will denote by $y_{\mathcal{D}}$. The main idea is to use the observations to explicitly construct the Itô map that maps a finite parametrization of $Y^{\mathcal{D}}$ to a finite parametrization of $X^{\mathcal{D}}$. Typically, $Y^{\mathcal{D}}$ will be parametrized by the observations $y_{\mathcal{D}} := \{y_{t_i};\ t_i \in \mathcal{D}\}$ and $X^{\mathcal{D}}$ will be parametrized by the corresponding normalised increments $(\Delta x)_{\mathcal{D}} := \{\frac{x_{t_{i+1}} - x_{t_i}}{t_{i+1}-t_i};\ t_i, t_{i+1} \in \mathcal{D}\}$.

In section 2, we study the existence and uniqueness for the pair $(X^{\mathcal{D}}, Y^{\mathcal{D}})$ for $Y^{\mathcal{D}}$ parametrised by the given dataset $y_{\mathcal{D}} = \{y_{t_i};\ t_i \in \mathcal{D}\}$. We give conditions for

Date: July 21, 2016.

Paper No. 16-02, www.warwick.ac.uk/go/crism

existence, which is necessary for the methodology to work. Then, we show that for a and b in Lip(2) and b non singular, the solution will be unique for the case $m = d$ and it will have $m - d$ degrees of freedom for the case $m > d$. Since existence will not in generally be true for the case $d > m$, this case will not be considered.

In section 3, we explicitly construct the likelihood, treating separately the cases where we have uniqueness and where we have one or more degrees of freedom. Finally, in section 4, we demonstrate how the method works in the simple case of a discretely observed Ornstein-Uhlenbeck model driven by a picewise linear approximation to fractional Brownian motion.

In the second part of the paper, we consider equation

$$dY_t = a(Y_t;\theta)dt + b(Y_t;\theta)dX_t, \quad Y_0 = y_0,\ t \leq T, \tag{2}$$

where $X \in G\Omega_p(\mathbb{R}^m)$ is the realisation of a random geometric p-rough path, defined as the p-variation limit of a random sequence of nested piecewise linear paths. Let us denote by $\mathcal{D}(n,T)$ the sequence of nested partitions of $[0,T]$ and by $\pi_n(X)$ the corresponding sequence of piecewise linear paths, such that $d_p(\pi_n(X), X) \to 0$ as $n \to \infty$. We now assume that for each $\theta \in \Theta$, $a(\cdot,\theta)$ and $b(\cdot,\theta)$ are Lip$(\gamma+1)$, for some $\gamma > p$, which are sufficient conditions for the existence and uniqueness of the solution $Y = I_\theta(X) \in G\Omega_p(\mathbb{R}^d)$. Moreover, as before, for b non-singular, the pair (X,Y) is unique. If we denote by $Y(n)$ the response to the piecewise linear path $\pi_n(X)$, i.e. $Y(n) = I_\theta(\pi_n(X))$, then the continuity of the Itô map in the p-variation topology implies that $d_p(Y(n), Y) \to 0$ as $n \to \infty$.

To simplify notation, we will assume that the partitions $\mathcal{D}(n,T)$ are the dyadic partitions of $[0,T]$, i.e. they are homogeneous with interval size $\delta = 2^{-n}$. We write $\mathcal{D}(n) = \{k2^{-n};\ k = 0,\ldots,N\}$, where $N = 2^nT$.

In section 5, we use the likelihood constructed before to construct an approximate likelihood of observing a realisation of (2) on grid $\mathcal{D}(n)$ for some fixed n – denoted by $y_{\mathcal{D}(n)}$. The main idea behind the construction is to replace the model (2) that produces the data by (1), which is tractable and converges to (2) for $n \to \infty$. However, one also needs to normalise the likelihood appropriately, so that the limit still depends on the parameter that we want to estimate.

In section 6, we make precise in what sense the likelihood constructed in the previous section is approximate. Replacing the complicated model by a simpler one approximating the actual model, when we can construct the likelihood corresponding to the simpler model exactly, is not an uncommon approach for performing statistical inference for otherwise intractable models. For example, this is done in [**?**] where the authors replace the actual multiscale model by its limiting diffusion and use that to construct the likelihood. They show that the approximation error due to the mismatch between data (coming from the multiscale model) and model (the limiting equation) disappears in the limit. Following a similar approach, we show

Paper No. 16-02, www.warwick.ac.uk/go/crism

that, under suitable conditions, an appropriate distance between the likelihood for discrete observations on a grid $\mathcal{D}(n)$ of the corresponding process $Y(n)$ and of the limiting process Y respectively disappears, as $n \to \infty$.

Finally, in section **??**, we construct the limiting likelihood for a discretely observed OU process and construct the MLE as well as posterior distributions for the drift and diffusion parameter of the OU model.

2. Existence and Uniqueness

We are given a set of points $y_{\mathcal{D}}$ in $\mathbb{R}^d$, where $\mathcal{D}$ is the fixed partition of $[0, T]$. In this section, we study the existence and uniqueness of piecewise linear path $X^{\mathcal{D}}$, whose response $Y^{\mathcal{D}}$ through (1) goes through points $y_{\mathcal{D}}$, i.e. $Y^{\mathcal{D}}_{t_i} = y_{t_i}$ for each $t_i \in \mathcal{D}$.

First, we discuss how to express $Y^{\mathcal{D}}$ in terms of $X^{\mathcal{D}}$. By construction, $X^{\mathcal{D}}$ is linear between grid points, i.e.

$$X^{\mathcal{D}}_t = X^{\mathcal{D}}_{t_i} + \Delta X^{\mathcal{D}}_{t_i} \left(t - t_i\right), \ \forall t \in [t_i, t_{i+1}), \ t_i, t_{i+1} \in \mathcal{D},$$

where $\Delta X^{\mathcal{D}}_{t_i} = \frac{X^{\mathcal{D}}_{t_{i+1}} - X^{\mathcal{D}}_{t_i}}{t_{i+1} - t_i}$. By definition, $Y^{\mathcal{D}} = I_\theta(X^{\mathcal{D}})$ which implies that for every $t \in [t_i, t_{i+1})$, $Y^{\mathcal{D}}_t$ satisfies

$$\begin{aligned} dY^{\mathcal{D}}_t &= a(Y^{\mathcal{D}}_t;\theta)dt + b(Y^{\mathcal{D}}_t;\theta)dX^{\mathcal{D}}_t = \\ &= \left(a(Y^{\mathcal{D}}_t;\theta)dt + b(Y^{\mathcal{D}}_t;\theta)\Delta X^{\mathcal{D}}_{t_i}\right) dt \end{aligned}$$

with initial conditions $Y^{\mathcal{D}}_{t_i} = y_{t_i}$. This is an ODE and we have already assumed sufficient regularity on a and b for existence and uniqueness of its solutions. The general form of the ODE is given by

$$d\tilde{Y}_t = \left(a(\tilde{Y}_t;\theta) + b(\tilde{Y}_t;\theta) \cdot c\right) dt, Y_0 = y_0 \tag{3}$$

and we will denote its solution by $F_t(y_0, c; \theta)$. Then,

$$Y^{\mathcal{D}}_t = F_{t-t_i}(y_{t_i}, \Delta X_{t_i}; \theta), \ \forall t \in [t_i, t_{i+1}). \tag{4}$$

In order to fit $Y^{\mathcal{D}}$ to the observed data $y_{\mathcal{D}}$, we need to solve for ΔX_{t_i}, using the terminal value, i.e. solve

$$F_{t_{i+1}-t_i}(y_{t_i}, \Delta X_{t_i}; \theta) = y_{t_{i+1}} \tag{5}$$

for $\Delta X_{t_i}(y_{t_i}, y_{t_{i+1}}; \theta)$. So, for every interval $[t_i, t_{i+1})$, we need to solve an independent system of d equations and m unknowns. That is, we need to study the existence and uniqueness of solutions with respect to c of the system

$$F_\delta(y_0, c; \theta) = y_1, \tag{6}$$

for every θ and for appropriate values of δ, y_0 and y_1. We are going to assume existence of solution, by requiring that $y_1 \in \cap_{\theta\in\Theta}\mathcal{M}_\delta(y_0;\theta)$, where

$$\mathcal{M}_\delta(y_0;\theta) = \{F_\delta(y_0, c;\theta); c \in \mathbb{R}^m\}. \tag{7}$$

Now, suppose that c_1 and c_2 are both solutions for a given $\theta \in \Theta$, i.e.

$$F_\delta(y_0, c_1, \theta) = y_1 = F_\delta(y_0, c_2, \theta).$$

We can write the difference as

$$F_\delta(y_0, c_2, \theta) - F_\delta(y_0, c_1, \theta) = \left(\int_0^1 D_c F_\delta(y_0, c_1 + s(c_2 - c_1);\theta)ds\right) \cdot (c_2 - c_1).$$

Thus, $F_\delta(y_0, c_1;\theta) = F_\delta(y_0, c_2;\theta)$ implies

$$\left(\int_0^1 D_c F_\delta(y_0, c_1 + s(c_2 - c_1);\theta)ds\right) \cdot (c_2 - c_1) = 0.$$

So, it is sufficient to show that $\forall\xi \in \mathbb{R}^m$, the rank of $d \times m$ matrix $D_c F_\delta(y_0, \xi;\theta)$ is d, which implies that the solution will have $m - d$ degrees of freedom, i.e. given $m - d$ coordinates of c, the other coordinates are uniquely defined. In particular, for $d = m$ we get uniqueness.

Since the vector field of (3) is linear with respect to c, we know that $F_t(y_0, c;\theta)$ will be continuously differentiable with respect to c for every y_0, θ and t in the appropriate bounded interval [1]. Thus, we define a new auxiliary process as $Z_t(c) = D_c F_t(y_0, c;\theta) \in \mathbb{R}^{d\times m}$, or,

$$Z_t^{i,\alpha}(c) = \frac{\partial}{\partial c_\alpha} F_t^i(y_0, c;\theta), \text{ for } i = 1, \dots, d, \ \alpha = 1, \dots, m. \tag{8}$$

Then, assuming one additional degree of regularity, $Z_t(c)$ satisfies

$$\begin{aligned}
\frac{d}{dt} Z_t^{i,\alpha}(c) &= \frac{d}{dt}\frac{\partial}{\partial c_\alpha} F_t^i(y_0, c;\theta) = \frac{\partial}{\partial c_\alpha}\frac{d}{dt} F_t^i(y_0, c;\theta) = \\
&= \frac{\partial}{\partial c_\alpha}\left(a_i(F_t(y_0, c;\theta)) + \sum_{\beta=1}^m c_\beta b_{i\beta}(F_t(y_0, c;\theta))\right) = \\
&= \sum_{j=1}^d \left(\partial_j a_i(F_t(y_0, c;\theta)) + \sum_{\beta=1}^m c_\beta \partial_j b_{i\beta}(F_t(y_0, c;\theta))\right) \bar{Z}_t^{j\alpha}(c) + b_{i\alpha}(F_t(y_0, c;\theta)),
\end{aligned}$$

where by $\bar{Z}_t^\alpha(c)$ we denote column $\alpha \in \{1, \dots, m\}$ of matrix $Z_t(c)$. More concisely, we write

$$\frac{d}{dt}\bar{Z}_t^\alpha(c) = \bigtriangledown(a + b \cdot c)(F_t) \cdot \bar{Z}_t^\alpha(c) + \bar{b}_\alpha(F_t), \tag{9}$$

Paper No. 16-02, www.warwick.ac.uk/go/crism

where $\bigtriangledown f$ of a function $f : \mathbb{R}^d \to \mathbb{R}^d$ we denote the $d \times d$ matrix defined as

$$(\bigtriangledown f(y))_{i,j} = \partial_j f_i(y).$$

Also, $\bar{b}_\alpha$ is column α of matrix b. Note that, for each fixed α, this is a linear equation of $\bar{Z}^\alpha(c)$ with non-homogeneous coefficients. Also note that the initial conditions will be

$$Z_0^{i,\alpha}(c) = \frac{\partial}{\partial c_\alpha} F_0^i(y_0, c; \theta) = \frac{\partial}{\partial c_\alpha} y_0 \equiv 0, \ \forall i = 1, \dots, d, \ \alpha = 1, \dots, m.$$

Thus, the solution to this equation will be

$$\bar{Z}_t^\alpha(c) = \int_0^t \exp\left(\mathbf{A}\right)_{s,t} \bar{b}_\alpha(F_s) ds, \tag{10}$$

where by $\exp\left(A\right)_{s,t}$ we denote the sum of iterated integrals

$$\exp\left(\mathbf{A}\right)_{s,t} = \sum_{k=0}^{\infty} \mathbf{A}_{s,t}^k$$

and

$$\mathbf{A}_{s,t}^k = \int \cdots \int_{s<u_1<\cdots<u_k<t} A(F_{u_1}) \cdots A(F_{u_k}) du_1 \dots du_k$$

for

$$A(y) = \bigtriangledown (a + b \cdot c)(y) \tag{11}$$

This is a $d \times d$ matrix and for $k = 0$ we get the identity matrix, i.e. $\mathbf{A}_{s,t}^0 = I_d$. Since each vector $\bar{Z}_\delta^\alpha$ is a column of the matrix $D_c F_\delta(y_0, c; \theta)$, the condition that the rank of this matrix is d is equivalent to d columns being linearly independent. Without loss of generality, let's consider the first d columns ($d \leq m$) and let us assume that

$$\lambda_1 \bar{Z}_\delta^1 + \cdots + \lambda_d \bar{Z}_\delta^d = \bar{0}, \tag{12}$$

for some $\lambda_1, \dots, \lambda_d \in \mathbb{R}$. We need to find conditions such that (12) is equivalent to $\lambda_1 = \cdots = \lambda_d = 0$. Using (10) we get that (12) is equivalent to

$$\int_0^\delta \exp\left(\mathbf{A}\right)_{s,\delta} \left(\lambda_1 \bar{b}_1(F_s) + \cdots + \lambda_d \bar{b}_d(F_s)\right) ds = \bar{0}.$$

Using the continuity of the integrated function with respect to s, we can deduce that there exists a $\delta' \in [0, \delta]$, such that we can write the above relationship as

$$\exp\left(\mathbf{A}\right)_{\delta',\delta} \left(\lambda_1 \bar{b}_1(F_{\delta'}) + \cdots + \lambda_d \bar{b}_d(F_{\delta'})\right) \cdot \delta = \bar{0}.$$

It is known that $\exp\left(\mathbf{A}\right)_{\delta',\delta}$ is invertible, with inverse equal to $\exp\left(\mathbf{A}\right)_{\delta,\delta'}$. Consequently, the above relationship can only be true if

$$\lambda_1 \bar{b}_1(F_{\delta'}) + \cdots + \lambda_d \bar{b}_d(F_{\delta'}) = \bar{0}.$$

Paper No. 16-02, www.warwick.ac.uk/go/crism

Assuming that the rank of $d \times m$ matrix $b(y)$ is d for every y, this implies that $\lambda_1 = \cdots = \lambda_d = 0$, which is what we required.

We have shown the following results:

Lemma 2.1. *Suppose that* $\mathrm{rank}\,(b(y,\theta)) = d$ *for every* y *and that* $a(\cdot,\theta)$ *and* $b(\cdot,\theta)$ *are* $\mathrm{Lip}(2)$. *Then*

$$\mathrm{rank}\,(Z_t(c)) = \mathrm{rank}\,(D_c F_t(y_0, c; \theta)) = d.$$

Note that the construction of the process Z can also be done for $X \in G\Omega_p(\mathbb{R}^m)$, provided that its piecewise linear approximations converge in p-variation and that the vector field functions a and b are now $\mathrm{Lip}(\gamma+1)$. Uniqueness of the pair (X,Y) for given Y follows by taking limits. We make this statement formal in the following

Corollary 2.2. *Suppose that* $\mathrm{rank}\,(b(y,\theta)) = d$ *for every* y *and that* $a(\cdot,\theta)$ *and* $b(\cdot,\theta)$ *are* $\mathrm{Lip}(\gamma+1)$. *Then, for a given* Y*, the solution* (X,Y) *of* (2) *is unique.*

3. Construction of the Likelihood

In this section, we construct the exact likelihood of observing the process $Y^{\mathcal{D}}$ on a fixed grid $\mathcal{D}$, denoted by $y_{\mathcal{D}} = Y^{\mathcal{D}}_{\mathcal{D}}$, where $Y^{\mathcal{D}}$ is the response to a piecewise linear path $X^{\mathcal{D}}$ on $\mathcal{D}$ through (1). The key realisation is that the values of $Y^{\mathcal{D}}$ on $\mathcal{D}$ actually completely describe the process $Y^{\mathcal{D}}$.

First, we need to impose a probability structure to the space. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $X^{\mathcal{D}}$ be a random variable, taking values in the space of piecewise linear paths on $\mathcal{D}$, equipped with the 1-variation topology. So, $X^{\mathcal{D}}$ is a random piecewise linear path on $\mathbb{R}^m$ corresponding to partition $\mathcal{D}$. Thus, it is fully described by the distribution of its values on the grid $\mathcal{D}$, or, equivalently, its increments. Let us denote that distribution by $\mathbb{P}_{\Delta X_{\mathcal{D}}}$.

The measure $\mathbb{P}_{\Delta X_{\mathcal{D}}}$ is a distribution on the finite dimensional space $\mathbb{R}^{m\times N}$, with $N = |\mathcal{D}|$ being the size of the partition. We will assume that this is absolutely continuous with respect to Lebesgue.

By the continuity of I_θ, $Y^{\mathcal{D}} = I_\theta(X^{\mathcal{D}})$ is also an implicitly finite dimensional random variable, whose distribution can be fully describe by the probability of its values on the grid. Below, we construct the likelihood of observing a realisation of $Y^{\mathcal{D}}$, corresponding to parametrisation $y_{\mathcal{D}}$.

3.1. Case I: Uniqueness.

Let us first consider the case where we have existence and uniqueness of solutions to system (6), so $m = d$. Then, for each dataset $y_{\mathcal{D}}$, the set $\{\Delta X_{t_i}(y_{t_i}, y_{t_{i+1}}; \theta), t_i \in \mathcal{D}\}$ will be uniquely defined as the collection of solutions of (6). This defines a map

$$I^{-1}_{\theta,\mathcal{D}}(y_{\mathcal{D}}) = \{\Delta X_{t_i}(y_{t_i}, y_{t_{i+1}}; \theta),\ t_i \in \mathcal{D}\}, \tag{13}$$

Paper No. 16-02, www.warwick.ac.uk/go/crism

which can be viewed as a transformation of the observed random variable in terms of the increments of the driving noise. Note that $y_{\mathcal{D}}$ and $\{\Delta X_{t_i}, t_i \in \mathcal{D}\}$ fully parametrize processes $Y^{\mathcal{D}}$ and $X^{\mathcal{D}}$. Thus, we can write the likelihood of observing $y_{\mathcal{D}}$ as

$$L_{Y^{\mathcal{D}}}(y_{\mathcal{D}}|\theta) = L_{\Delta X_{\mathcal{D}}}\left(I^{-1}_{\theta,\mathcal{D}}(y_{\mathcal{D}})\right)|DI^{-1}_{\theta,\mathcal{D}}(y_{\mathcal{D}})|,$$

where by $L_{\Delta X_{\mathcal{D}}}(\Delta x_{\mathcal{D}})$ we denote the Radon-Nikodym derivative of $\mathbb{P}_{\Delta X_{\mathcal{D}}}$ with respect to Lebesque. This will be explicitly known since we assumed that we know the distribution of $X^{\mathcal{D}}$. Finally, since ΔX_{t_i} only depends on y_{t_i} and $y_{t_{i+1}}$ and not the whole path, it is not hard to see that the Jacobian matrix will be block lower triangular and consequently, the determinant will be the product of the determinants of the blocks on the diagonal:

$$|DI^{-1}_{\theta,\mathcal{D}}(y_{\mathcal{D}})| = \prod_{t_i\in\mathcal{D}}\left|\nabla\Delta X_{t_i}(y_{t_i}, y; \theta)|_{y=y_{t_{i+1}}}\right|. \tag{14}$$

Note that, by definition,

$$F_{t_{i+1}-t_i}(y_{t_i}, \Delta X_{t_i}(y_{t_i}, y; \theta); \theta) \equiv y.$$

Thus,

$$D_c F_{t_{i+1}-t_i}(y_{t_i}, c; \theta)|_{c=\Delta X_{t_i}(y_{t_i}, y_{t_{i+1}};\theta)} \cdot \nabla\Delta X_{t_i}(y_{t_i}, y; \theta)|_{y=y_{t_{i+1}}} \equiv I_d$$

and, consequently,

$$\begin{aligned}\nabla\Delta X_{t_i}(y_{t_i}, y; \theta)|_{y=y_{t_{i+1}}} &= \left(D_c F_{t_{i+1}-t_i}(y_{t_i}, c; \theta)|_{c=\Delta X_{t_i}(y_{t_i}, y_{t_{i+1}};\theta)}\right)^{-1} = \\ &= \left(Z_{t_{i+1}-t_i}(\Delta X_{t_i}(y_{t_i}, y_{t_{i+1}}; \theta))\right)^{-1}.\end{aligned}$$

So, the likelihood can be written as

$$L_{Y^{\mathcal{D}}}(y_{\mathcal{D}}|\theta) = L_{\Delta X_{\mathcal{D}}}\left(I^{-1}_{\theta,\mathcal{D}}(y_{\mathcal{D}})\right)\left(\prod_{t_i\in\mathcal{D}}\left|Z_{t_{i+1}-t_i}\left(I^{-1}_{\theta,\mathcal{D}}(y_{\mathcal{D}})_{t_i}\right)\right|\right)^{-1}. \tag{15}$$

3.2. **Case II: Degrees of Freedom.** Now suppose that $m > d$. Without loss of generality, let us assume that given coordinates $c_{d+1}, \ldots, c_m$, the remaining coordinates $c_1, \ldots, c_d$ are uniquely defined. Similar to previous case, we denote by $I^{-1}_{\theta,\mathcal{D},c_{d+1},\ldots,c_m}(y_{\mathcal{D}})$ the map from data points $y_{\mathcal{D}}$ to the first d increments, denoted by $\{\Delta X_{t_i}(y_{t_i}, y_{t_{i+1}}; \theta, c_{d+1}, \ldots, c_m)^i, t_i \in \mathcal{D}, i = 1, \ldots, d\}$, for fixed $c_{d+1}, \ldots, c_m$. As before, this can be viewed as a transformation of the observed random variable in terms of the first d increments of the driving noise and we get a similar formula for the likelihood:

$$\begin{aligned}&L_{Y^{\mathcal{D}}}(y_{\mathcal{D}}|\theta, c_{d+1}, \ldots, c_m) = \\ &L_{\Delta X_{\mathcal{D}}}\left(I^{-1}_{\theta,\mathcal{D},c_{d+1},\ldots,c_m}(y_{\mathcal{D}})\right) \cdot \left(\prod_{t_i\in\mathcal{D}}\left|Z_{t_{i+1}-t_i}\left(I^{-1}_{\theta,\mathcal{D},c_{d+1},\ldots,c_m}(y_{\mathcal{D}})_{t_i}\right)\right|\right)^{-1}.\end{aligned}$$

However, $c_{d+1}, \ldots, c_m$ will not be known in general, so we have to consider all possible values of them, leading to the formula

$$L_{Y^{\mathcal{D}}}(y_{\mathcal{D}}|\theta) = \int_{\mathbb{R}^{m-d}} L_{Y^{\mathcal{D}}}(y_{\mathcal{D}}|\theta, x_{d+1}, \ldots, x_m)\, \mathbb{P}_{c_{d+1},\ldots,c_m}(dx_{d+1}, \ldots, dx_m) =$$

$$\int_{\mathbb{R}^{m-d}} L_{\Delta X_{\mathcal{D}}}\left(I^{-1}_{\theta,\mathcal{D},x_{d+1},\ldots,x_m}(y_{\mathcal{D}})\right) \cdot \left(\prod_{t_i \in \mathcal{D}} \left| Z_{t_{i+1}-t_i}\left(I^{-1}_{\theta,\mathcal{D},x_{d+1},\ldots,x_m}(y_{\mathcal{D}})_{t_i}\right)\right|\right)^{-1} \cdot$$

$$\cdot \mathbb{P}_{c_{d+1},\ldots,c_m}(dx_{d+1}, \ldots, dx_m),$$

where $\mathbb{P}_{c_{d+1},\ldots,c_m}$ is the marginal distribution of $\mathbb{P}_{\Delta X_{\mathcal{D}}}$ on $\mathbb{R}^{(m-d)\times N}$.

4. Example: The 1d fractional O.U. process

To demonstrate the methodology, we will apply the ideas described in the previous section to a simple example. We consider the differential equation

$$dY^{\mathcal{D}}_t = -\lambda Y^{\mathcal{D}}_t dt + \sigma X^{\mathcal{D}}_t, \quad Y^{\mathcal{D}}_0 = 0, \tag{16}$$

where $X^{\mathcal{D}}_t$ is the piecewise linear interpolation to a fractional Brownian path with Hurst parameter h on a homogeneous grid $\mathcal{D} = \{k\delta; k = 0, \ldots, N\}$ where $N\delta = T$. Our goal will be to construct the likelihood of discretely observing a realisation of the solution $Y^{\mathcal{D}}(\omega)$ on the grid, for parameter values $\theta = (\lambda, \sigma) \in \mathbb{R}_+ \times \mathbb{R}_+$.

Our first task is to explicitly construct the parametrization of $Y^{\mathcal{D}}(\omega)$ in terms of its values on the grid $y_{\mathcal{D}}$, that completely determine the process. Let $X^{\mathcal{D}}(\omega)$ be the piecewise linear interpolation on $\mathcal{D}$ of the corresponding realisation of a fractional Brownian path driving (16). We will denote by x_{t_i} its values on the grid, i.e. $X^{\mathcal{D}}(\omega)_{t_i} = x_{t_i}$, $\forall t_i \in \mathcal{D}$. Since $X^{\mathcal{D}}(\omega)$ is the piecewise linear path defined on these points, $Y^{\mathcal{D}}(\omega)$ will be the solution to

$$dY^{\mathcal{D}}(\omega)_t = -\lambda Y^{\mathcal{D}}(\omega)_t dt + \sigma \frac{x_{(k+1)\delta} - x_{k\delta}}{\delta} dt,$$

which is given by

$$Y^{\mathcal{D}}(\omega)_t = Y^{\mathcal{D}}(\omega)_{k\delta} e^{-\lambda(t-k\delta)} + \frac{\sigma}{\lambda} \frac{x_{(k+1)\delta} - x_{k\delta}}{\delta}\left(1 - e^{-\lambda(t-k\delta)}\right), t \in [k\delta, (k+1)\delta).$$

We now need to solve for the unknown $\Delta x_{k+1} := x_{(k+1)\delta} - x_{k\delta}$: for $t = (k+1)\delta$. We get

$$y_{(k+1)\delta} = y_{k\delta} e^{-\lambda\delta} + \frac{\sigma \Delta x_{k+1}}{\lambda\delta}\left(1 - e^{-\lambda\delta}\right) \tag{17}$$

and, consequently,

$$I^{-1}_{\theta,\mathcal{D}}(y_{\mathcal{D}})_{k+1} := \Delta x_{k+1} = \frac{\lambda\delta\left(y_{(k+1)\delta} - y_{k\delta} e^{-\lambda\delta}\right)}{\sigma\left(1 - e^{-\lambda\delta}\right)}, \; k = 0, \ldots, N-1, \tag{18}$$

with $y_0 = 0$ and $\theta = (\lambda, \sigma)$. Thus, $Y^{\mathcal{D}}(\omega)$ is given by

$$Y^{\mathcal{D}}(\omega)_t = y_{k\delta} e^{-\lambda(t-k\delta)} + \frac{y_{(k+1)\delta} - y_{k\delta} e^{-\lambda\delta}}{1 - e^{-\lambda\delta}} \left(1 - e^{-\lambda(t-k\delta)}\right), \tag{19}$$

for $t \in [k\delta, (k+1)\delta)$ and $y_0 = 0$.

Clearly, in this case, the solution of system (6) always exists and is unique under the condition that $\sigma \neq 0$. Let us now compute the process Z defined in (8). In this case, since $d = 1$, this is a scalar process. It is easy to compute Z directly but we will use formula (10) instead, as a demonstration. First, we note that A defined in (11) will be $A(y) = \partial_y(-\lambda y + \sigma c) = \lambda$. Thus, (10) becomes

$$Z_t = \int_0^t \exp(-\lambda(t-s)) \frac{\sigma}{\delta} ds = \frac{\sigma}{\lambda\delta}(1 - e^{-\lambda t}).$$

We now have all the elements we need to write down the likelihood: from (15), we get

$$L_{Y^{\mathcal{D}}}\left(y_{\mathcal{D}} \middle|\, \theta\right) = L_{\Delta X_{\mathcal{D}}}\left(I^{-1}_{\theta,\mathcal{D}}(y_{\mathcal{D}})\right) \left(\frac{\lambda\delta}{\sigma(1 - e^{-\lambda\delta})}\right)^N.$$

Finally, we note that the likelihood of the increments $\Delta X_{\mathcal{D}}$ is a mean zero Gaussian distribution with covariance matrix given by

$$\left(\Sigma_h^{\mathcal{D}}\right)_{ij} = \frac{\delta^{2h}}{2}\left(|j-i+1|^{2h} + |j-i-1|^{2h} - 2|j-i|^{2h}\right),\ i,j = 1,\dots,N,$$

where h is the Hurst parameter of the fractional Brownian motion. Thus, the likelihood becomes

$$|2\pi\Sigma_h^{\mathcal{D}}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} I^{-1}_{\theta,\mathcal{D}}(y_{\mathcal{D}}) \left(\Sigma_h^{\mathcal{D}}\right)^{-1} I^{-1}_{\theta,\mathcal{D}}(y_{\mathcal{D}})^*\right) \left(\frac{\lambda\delta}{\sigma(1 - e^{-\lambda\delta})}\right)^N, \tag{20}$$

where we denote by z^* the transpose of a vector z. The corresponding log-likelihood is proportional to

$$\ell_Y\left(y_{\mathcal{D}} \middle|\, \theta\right) \propto -\frac{1}{2} I^{-1}_{\theta,\mathcal{D}}(y_{\mathcal{D}}) \left(\Sigma_h^{\mathcal{D}}\right)^{-1} I^{-1}_{\theta,\mathcal{D}}(y_{\mathcal{D}})^* + N \log\left(\frac{\lambda\delta}{\sigma(1 - e^{-\lambda\delta})}\right). \tag{21}$$

Finally, we can replace $I^{-1}_{\theta,\mathcal{D}}$ above with its exact expression, which gives

$$\ell_Y\left(y_{\mathcal{D}} \middle|\, \lambda, \sigma\right) \propto -\frac{\lambda^2\delta^2}{2\sigma^2(1 - e^{-\lambda\delta})^2} \left(\Delta^\lambda y\right)_{\mathcal{D}} \left(\Sigma_h^{\mathcal{D}}\right)^{-1} \left(\Delta^\lambda y\right)^*_{\mathcal{D}} + N \log\left(\frac{\lambda\delta}{\sigma(1 - e^{-\lambda\delta})}\right),$$

where by $\Delta^\lambda y_{k\delta} = y_{(k+1)\delta} - y_{k\delta} e^{\lambda\delta}$.

5. The Limiting Case

In the first part of the paper, we assumed that we observe the response to a differential equation driven by a piecewise linear path (1) and we constructed the exact likelihood of the observations. In this second part of the paper, we discretely observe the response to a differential equation (2) driven by a p-rough path X. We aim to construct an approximate likelihood for the observations, where we consider an approximation to be acceptable if it leads to asymptotically consistent estimators. A crucial assumption is that there exists a sequence of partitions $\mathcal{D}(n)$ (usually dyadic) such that the corresponding piecewise linear interpolations $\pi_n(X)$ of the path X converge in p-variation to the p-rough path X. This allows us to replace (2) by (1)

Let us denote by $y_{\mathcal{D}(n)}$ the sequence of observations of the limiting equation (2) on the grid $\mathcal{D}(n)$. We will use the likelihood $L_{Y^{\mathcal{D}(n)}}$ constructed in (15) to construct an approximate likelihood for the partially observed limiting equation – for simplicity, we will now denote it by $L_{Y(n)}$. Also, to simplify the exposition, we will focus on the case where we have uniqueness, i.e. $m = d$ and b is non-singular.

A first idea would be to define the approximate likelihood as $L_{Y(n)}\left(y_{\mathcal{D}(n)}|\theta\right)$. Then, we would hope to show that, for n large, this will be close to $L_{Y(n)}\left(y(n)_{\mathcal{D}(n)}|\theta\right)$ in a way that it allows the estimators constructed using this likelihood to inherit a lot of the properties of those constructed using exact likelihood $L_{Y(n)}\left(y(n)_{\mathcal{D}(n)}|\theta\right)$. Note that the difference between $y_{\mathcal{D}(n)}$ and $y(n)_{\mathcal{D}(n)}$ is that the first is the response to a realisation of the rough path x while the latter is the response to the piecewise linear approximation of x on the grid $\mathcal{D}(n)$, i.e. $y(n)_{\mathcal{D}(n)} = I_\theta(\pi_n(x))_{\mathcal{D}(n)}$, making the likelihood exact. Note that the two sequences converge in p-variation, for $n \to \infty$. So, we expect that the estimators constructed using the datasets $y_{\mathcal{D}(n)}$ and $y(n)_{\mathcal{D}(n)}$ will be close, provided that the estimator is continuous in the p-variation topology.

However, when the model involves more than one parameter, it is often the case that, in the limit, $L_{Y(n)}\left(y_{\mathcal{D}(n)}|\theta\right)$ as a function of θ scales differently for different coordinates of θ. In particular, this occurs because the drift component dt scales differently than the 'diffusion' component dX_t. Thus, we need to carefully normalise the likelihood appropriately, depending on which coordinate of θ we want to estimate at any time. Actually, it is equivalent and more convenient to work with the log-likelihood: normalising the log-likelihood involves adding functions to the log-likelihood that are independent of the parameters we want to estimate and thus do not alter the estimation, as both the maximum with respect to the parameter and the posterior on the parameter remain unaffected. So, we want to construct an expansion of the log-likelihood of the form

$$\ell_{Y(n)}\left(y_{\mathcal{D}(n)}|\theta\right) = \sum_{k=0}^{M} \ell^{(k)}_{Y(n)}\left(y_{\mathcal{D}(n)}|\theta\right) n^{-\alpha_k} + R_M(y_{\mathcal{D}(n)}, \theta) \tag{22}$$

for $M \in \mathbb{N}$ and $-\infty < \alpha_0 < \alpha_1 < \cdots < \alpha_M < \infty$, where $\ell^{(k)}_{Y(n)}\left(y_{\mathcal{D}(n)}|\theta\right)$ converges to a non-trivial limit (finite and non-zero) for every $k = 0, \ldots, M$ and the remainder $R_M(y_{\mathcal{D}(n)}, \theta)$ satisfies $\lim_{n\to\infty} n^{\alpha_M} R_M(y_{\mathcal{D}(n)}, \theta) = 0$. This will exist, assuming sufficient smoothness of the log-likelihood function of $\Delta x_{\mathcal{D}(n)}$. The construction is based on a Taylor expansion around the two components of the inverse that scale differently. More precisely, it is done as follows:

1. Using (15), we express the log-likelihood in terms of $I^{-1}_{\theta,\mathcal{D}(n)}(y_{\mathcal{D}(n)})$, which we will denote for simplicity by $I^{-1}_{\theta,n}(y_{\mathcal{D}(n)})$. That is,
$$\ell_{Y(n)}\left(y_{\mathcal{D}(n)}|\theta\right) = \ell_{\Delta X_{\mathcal{D}(n)}}\left(I^{-1}_{\theta,n}(y_{\mathcal{D}(n)})\right) - \sum_{i=0}^{N-1} \log\left|Z_{t_{i+1}-t_i}\left(I^{-1}_{\theta,n}(y_{\mathcal{D}(n)})\right)\right|.$$
2. Assuming sufficient regularity of $\ell_{\Delta X_{\mathcal{D}(n)}}$, we expand $\ell_{Y(n)}\left(y_{\mathcal{D}(n)}|\theta\right)$ in terms of the monomials of $I^{-1}_{\theta,n}(y_{\mathcal{D}(n)})$. Note that $Z_{t_{i+1}-t_i}\left(I^{-1}_{\theta,n}(y_{\mathcal{D}(n)})\right)$ given by (10) will always be smooth.
3. We have assumed uniqueness, which is equivalent to $b(y)$ being invertible. Thus, $I^{-1}_{\theta,n}(y_{\mathcal{D}(n)})$ can be expressed as
$$\begin{aligned} I^{-1}_{\theta,n}(y_{\mathcal{D}(n)})_{t_i,t_{i+1}} &= \int_{t_i}^{t_{i+1}} b^{-1}(Y(n, y_{\mathcal{D}(n)})_u) dY(n, y_{\mathcal{D}(n)})_u \\ &\quad - \int_{t_i}^{t_{i+1}} b^{-1}(Y(n, y_{\mathcal{D}(n)})_u) a(Y(n, y_{\mathcal{D}(n)})_u) du \\ &= I^{-1}_{\theta}\left(Y(n, y_{\mathcal{D}(n)})\right)_{t_i,t_{i+1}} \end{aligned} \tag{23}$$
where the process $Y(n, y_{\mathcal{D}(n)})$ is the response to a piecewise linear path, parametrised by its values on the grid $\mathcal{D}(n)$, given by $y_{\mathcal{D}(n)}$. Using this equivalence, we further expand the monomials of $I^{-1}_{\theta,n}(y_{\mathcal{D}(n)})$ in terms of monomials of the two vectors formed by these integrals, i.e.
$$\left\{\int_{t_i}^{t_{i+1}} b^{-1}(Y(n, y_{\mathcal{D}(n)})_u) dY(n, y_{\mathcal{D}(n)})_u\right\}_{i=0}^{N-1}$$
and
$$\left\{\int_{t_i}^{t_{i+1}} b^{-1}(Y(n, y_{\mathcal{D}(n)})_u) a(Y(n, y_{\mathcal{D}(n)})_u) du\right\}_{i=0}^{N-1},$$
where $N = 2^n T = \frac{T}{\delta}$, as before.
4. We classify each monomial to class k if the normalisation needed for it to converge to something "meaningful" is n^{α_k}, i.e. the monomial multiplied by n^{α_k} converges to a function that is not equivalent to 0. We denote by M the

number of such classes and by $\ell^{(k)}_{Y(n)}\left(y_{\mathcal{D}(n)}|\theta\right)$ the sum of all class k monomials multiplied by n^{α_k}.

Remark 5.1. *Note that while* (23) *is very useful for studying limiting behaviour, it is not useful for constructing the likelihood for fixed n, as constructing $Y(n, y_{\mathcal{D}(n)})$ corresponding to observations $y_{\mathcal{D}(n)}$ also requires the solution of* (5)*.*

Now, let θ_i be an arbitrary coordinate of the parameter θ and suppose that $\ell^{(k)}_{Y(n)}\left(y_{\mathcal{D}(n)}|\theta\right)$ are independent of θ_i for $k = 0, \dots, m-1$ while $\ell^{(m)}_{Y(n)}\left(y_{\mathcal{D}(n)}|\theta\right)$ depends on θ_i, in a way to be made precise later, which will depend on the way the constructed likelihood is used. Intuitively, we expect that the first $m-1$ components will be irrelevant to the estimation of the parameter and should be ignored, while the remaining log-likelihood should be normalised by n^{α_m}. Thus, we will say that coordinate θ_i of the parameter is of order m and, given observations of the limiting equation, we will use the dominating term $\ell^{(m)}_{Y(n)}\left(y_{\mathcal{D}(n)}|\theta\right)$ for its estimation. We will assume that all coordinates of the parameter are of finite order – otherwise, they cannot be estimated!

Below, we use this framework to build estimators for parameter θ using the constructed likelihoods. We discuss separately the two most common approaches, corresponding to the Frequentist or Bayesian paradigm.

5.1. **Frequentist Setting.** In the frequentist setting, we use the likelihood constructed above in order to construct the Maximum Likelihood Estimator (MLE) of the parameter $\theta \in \Theta$. We inductively define the MLEs of different co-ordinates of θ, as follows:

1. We start with the lower order α_0. We say that co-ordinates of the parameter θ are of order α_0 and we denote them by θ_0 if
$$\ell^{(0)}(y_{\mathcal{D}(n)}|\theta) = \ell^{(0)}(y_{\mathcal{D}(n)}|\theta_0), \forall \theta \in \Theta.$$
Then, we define their estimate as
$$\hat{\theta}_0(y_{\mathcal{D}(n)}) = \mathrm{argmax}_{\theta_0} \ell^{(0)}_{Y(n)}\left(y_{\mathcal{D}(n)}|\theta_0\right).$$
2. Suppose that we have defined parameters of order up to m and their MLEs, for some $m \geq 0$. Then, we say that the co-ordinates of θ that satisfy
$$\ell^{(m+1)}(y_{\mathcal{D}(n)}|\theta) = \ell^{(m+1)}(y_{\mathcal{D}(n)}|\hat{\theta}_0(y_{\mathcal{D}(n)}), \dots, \hat{\theta}_m(y_{\mathcal{D}(n)}), \theta_{m+1})$$
for all θ with coordinates of order less or equal to m equal to their MLE estimates are of order $m+1$, and we denote them by θ_{m+1}. We define their MLE as
$$\hat{\theta}_{m+1}(y_{\mathcal{D}(n)}) = \mathrm{argmax}_{\theta_{m+1}} \ell^{(m+1)}_{Y(n)}\left(y_{\mathcal{D}(n)}|\hat{\theta}_0(y_{\mathcal{D}(n)}), \dots, \hat{\theta}_m(y_{\mathcal{D}(n)}), \theta_{m+1}\right).$$

5.2. **Bayesian Setting.** In the Bayesian setting, we use the likelihood constructed above together with a prior distribution on the parameter space that we will denote by u, in order to construct the posterior distribution of the parameter $\theta \in \Theta$. We inductively define the posterior distributions of different co-ordinates of θ, as follows:

1. First, we start with the lower order α_0 and work our way up. We say that a co-ordinate θ_i of the parameter is of order α_k if the distance between the marginal posterior on θ_i and the marginal prior (for an appropriate choice of distance on the measure space) is non-zero for the first time when the posterior is computed using the scaling of the likelihood corresponding to $\ell^{(k)}_{Y(n)}\left(y_{\mathcal{D}(n)}|\theta_0\right)$. We will denote by $\theta(k)$ all the co-ordinates of the parameter that are of order k and by r the maximum order.
2. The posterior can be written as a product of the posteriors of parameters of different orders as follows:

$$\mathbb{P}(\theta|y_{\mathcal{D}(n)}) = \prod_{k=0}^{r} \mathbb{P}\left(\theta(k)|y_{\mathcal{D}(n)}, \theta(k-1), \ldots, \theta(0)\right),$$

   where for each k, $\mathbb{P}\left(\theta(k)|y_{\mathcal{D}(n)}, \theta(k-1), \ldots, \theta(0)\right)$ is computed at the relevant scale, i.e.

$$\frac{\exp\left(\ell^{(k)}_{Y(n)}\left(y_{\mathcal{D}(n)}|\theta(k), c\theta(k)\right) n^{-\alpha_k}\right) u(\theta(k), c\theta(k))}{\int_{\Theta_k} \exp\left(\ell^{(k)}_{Y(n)}\left(y_{\mathcal{D}(n)}|\tilde{\theta}(k), c\theta(k)\right) n^{-\alpha_k}\right) u(\tilde{\theta}(k), c\theta(k)) d\tilde{\theta}(k)},$$

   where Θ_k is the projection of the parameter space to the coordinates of order k and by $c\theta(k)$ we denote all the parameters that are not of order k.

6. Convergence of Approximate Likelihood

In this section, we study the behaviour of the approximate likelihoods constructed in section 5. The main result of this section is the following:

Theorem 6.1. *Let $\ell^{(k)}_{Y(n)}(\cdot|\theta)$ be the scaled likelihoods constructed in section 5, so that (22) holds. Let y be the response to a p-rough path x through (2) and $y(n)$ be the response to $\pi_n(x)$ through (1), where $\pi_n(x)$ is the piecewise linear interpolation of x on grid $\mathcal{D}(n) = \{k2^{-n}T,\ k = 0, \ldots, N\}$ for $N = 2^n T$. Then, assuming that the determinant of b is uniformly bounded from below*

$$\lim_{n\to\infty} \sup_{\theta} \left| \ell^{(k)}_{Y(n)}\left(y_{\mathcal{D}(n)}|\theta\right) - \ell^{(k)}_{Y(n)}\left(y(n)_{\mathcal{D}(n)}|\theta\right) \right| = 0, \tag{24}$$

for all k such that $\alpha_k \leq 0$.

First, we will show that the log-likelihood function, which is the logarithm of (15), is continuous with respect to the inverse Itô map of the data

$$I_{\theta,n}^{-1}(y_{\mathcal{D}(n)}) = I_{\theta}^{-1}(Y(n, y_{\mathcal{D}(n)})), \tag{25}$$

where, as before, $Y(n, y_{\mathcal{D}(n)})$ is the response to a piecewise linear path parametrised by its values on the grid, $y_{\mathcal{D}(n)}$. So, $I_{\theta}^{-1}(Y(n, y_{\mathcal{D}(n)}))$ will be exactly that piecewise linear path driving the process. More precisely, we will show that, under certain assumptions, the following holds:

$$\left|\ell_{Y(n)}\left(y_{\mathcal{D}(n)}|\theta\right) - \ell_{Y(n)}\left(\tilde{y}_{\mathcal{D}(n)}|\theta\right)\right| \leq \omega\left(d_p(I_{\theta,n}^{-1}(y_{\mathcal{D}(n)}), I_{\theta,n}^{-1}(\tilde{y}_{\mathcal{D}(n)})\right), \tag{26}$$

for some modulus of continuity function ω, independent of θ and n. This can be further split into two parts, one corresponding to the log-likelihood of the inverse and the other corresponding to the Jacobian correction. We start by deriving a bound for the part of the log-likelihood error corresponding to the Jacobian correction.

Lemma 6.2. *For $Z_{t_{i+1}-t_i}$ and $I_{\theta,n}^{-1}$ defined as in* (10) *and* (13) *respectively and under the additional assumption on b that*

$$\inf_y ||b(y)|| = \frac{1}{M_b} > 0,$$

for some $M_b > 0$, it holds that

$$\left|\sum_{t_i\in\mathcal{D}(n)} \log|Z_{t_{i+1}-t_i}\left(I_{\theta,n}^{-1}(y_{\mathcal{D}(n)})_{t_i}\right)| - \sum_{t_i\in\mathcal{D}(n)} \log|Z_{t_{i+1}-t_i}\left(I_{\theta,n}^{-1}(\tilde{y}_{\mathcal{D}(n)})_{t_i}\right)|\right| \leq M\cdot C\cdot\omega(d_p(I_{\theta,n}^{-1}(y_{\mathcal{D}(n)}), I_{\theta,n}^{-1}(\tilde{y}_{\mathcal{D}(n)})))$$

for some $M, C \in \mathbb{R}_+$ and modulus of continuity function ω.

Proof. We write

$$\sum_{t_i\in\mathcal{D}(n)} \log|Z_{t_{i+1}-t_i}\left(I_{\theta,n}^{-1}(y_{\mathcal{D}(n)})_{t_i}\right)| - \sum_{t_i\in\mathcal{D}(n)} \log|Z_{t_{i+1}-t_i}\left(I_{\theta,n}^{-1}(\tilde{y}^n_{\mathcal{D}(n)})_{t_i}\right)| = \sum_{t_i\in\mathcal{D}(n)} \log\frac{|Z_{t_{i+1}-t_i}\left(I_{\theta,n}^{-1}(y_{\mathcal{D}(n)})_{t_i}\right)|}{|Z_{t_{i+1}-t_i}\left(I_{\theta,n}^{-1}(\tilde{y}^n_{\mathcal{D}(n)})_{t_i}\right)|} \tag{27}$$

As before, using the continuity of integrated function within $Z_{t_{i+1}-t_i}$ with respect to the time variable, we write

$$Z_{t_{i+1}-t_i}\left(I_{\theta,n}^{-1}(y_{\mathcal{D}(n)})_{t_i}\right) = \exp\left(\mathbf{A}(y_{t_i}, I_{\theta,n}^{-1}(y_{\mathcal{D}(n)})_{t_i})\right)_{\zeta_i,t_{i+1}}\cdot b(F_{\zeta_i})\cdot(t_{i+1}-t_i)$$

and

$$Z_{t_{i+1}-t_i}\left(I_{\theta,n}^{-1}(\tilde{y}^n_{\mathcal{D}(n)})_{t_i}\right) = \exp\left(\mathbf{A}(\tilde{y}^n_{t_i}, I_{\theta,n}^{-1}(\tilde{y}^n_{\mathcal{D}(n)})_{t_i})\right)_{\eta_i,t_{i+1}}\cdot b(F_{\eta_i})\cdot(t_{i+1}-t_i)$$

Paper No. 16-02, www.warwick.ac.uk/go/crism

for some $\zeta_i, \eta_i \in [t_i, t_{i+1}]$. So, (27) simplifies to

$$\sum_{t_i \in \mathcal{D}(n)} \log \frac{|\exp\Big(\mathbf{A}(y_{t_i}, I^{-1}_{\theta,n}(y_{\mathcal{D}(n)})_{t_i})\Big)_{\zeta_i,t_{i+1}} \cdot b(F(y_{t_i}, I^{-1}_{\theta,n}(y_{\mathcal{D}(n)})_{t_i})_{\zeta_i})|}{|\exp\Big(\mathbf{A}(\tilde{y}^n_{t_i}, I^{-1}_{\theta,n}(\tilde{y}^n_{\mathcal{D}(n)})_{t_i})\Big)_{\eta_i,t_{i+1}} \cdot b(F(\tilde{y}^n_{t_i}, I^{-1}_{\theta,n}(\tilde{y}^n_{\mathcal{D}(n)})_{t_i})_{\eta_i})|} =$$

$$\sum_{t_i \in \mathcal{D}(n)} \log \frac{|\exp\Big(\mathbf{A}(y_{t_i}, I^{-1}_{\theta,n}(y_{\mathcal{D}(n)})_{t_i})\Big)_{\zeta_i,t_{i+1}}|}{|\exp\Big(\mathbf{A}(\tilde{y}^n_{t_i}, I^{-1}_{\theta,n}(\tilde{y}^n_{\mathcal{D}(n)})_{t_i})\Big)_{\eta_i,t_{i+1}}|} + \sum_{t_i \in \mathcal{D}(n)} \log \frac{|b(F(y_{t_i}, I^{-1}_{\theta,n}(y_{\mathcal{D}(n)})_{t_i})_{\zeta_i})|}{|b(F(\tilde{y}^n_{t_i}, I^{-1}_{\theta,n}(\tilde{y}^n_{\mathcal{D}(n)})_{t_i})_{\eta_i})|}.$$

Focusing on the second summand, we write

$$|\sum_{t_i \in \mathcal{D}(n)} \log \frac{|b(F(y_{t_i}, I^{-1}_{\theta,n}(y_{\mathcal{D}(n)})_{t_i})_{\zeta_i})|}{|b(F(\tilde{y}^n_{t_i}, I^{-1}_{\theta,n}(\tilde{y}^n_{\mathcal{D}(n)})_{t_i})_{\eta_i})|}| \leq$$

$$\sum_{t_i \in \mathcal{D}(n)} \log \left(1 + M \left| |b(F(y_{t_i}, I^{-1}_{\theta,n}(y_{\mathcal{D}(n)})_{t_i})_{\zeta_i})| - |b(F(\tilde{y}^n_{t_i}, I^{-1}_{\theta,n}(\tilde{y}^n_{\mathcal{D}(n)})_{t_i})_{\eta_i})| \right| \right) \leq$$

$$M \cdot \sum_{t_i \in \mathcal{D}(n)} \left| |b(F(y_{t_i}, I^{-1}_{\theta,n}(y_{\mathcal{D}(n)})_{t_i})_{\zeta_i})| - |b(F(\tilde{y}^n_{t_i}, I^{-1}_{\theta,n}(\tilde{y}^n_{\mathcal{D}(n)})_{t_i})_{\eta_i})| \right| \leq$$

$$M_b \cdot C_b \cdot \omega_1(d_p(I^{-1}_{\theta,n}(y_{\mathcal{D}(n)}), I^{-1}_{\theta,n}(\tilde{y}^n_{\mathcal{D}(n)}))),$$

where we used the inequality $\log(1+x) < x$, assumption that $\inf_y ||b(y)|| = \frac{1}{M_b} > 0$, the Lipschitz continuity of b and the universal limit theorem (see [4] for exact bound). Similarly, we get that

$$\sum_{t_i \in \mathcal{D}(n)} \log \frac{|\exp\left(\mathbf{A}(y_{t_i}, I^{-1}_{\theta,n}(y_{\mathcal{D}(n)})_{t_i})\right)_{\zeta_i,t_{i+1}}|}{|\exp\left(\mathbf{A}(\tilde{y}^n_{t_i}, I^{-1}_{\theta,n}(\tilde{y}^n_{\mathcal{D}(n)})_{t_i})\right)_{\eta_i,t_{i+1}}|} \leq M_A C_A \cdot \omega_2(d_p(I^{-1}_{\theta,n}(y_{\mathcal{D}(n)}), I^{-1}_{\theta,n}(\tilde{y}^n_{\mathcal{D}(n)}))).$$

under the additional assumption that $inf_y|exp(A)(y)_{t_i,t_{i+1}}| > 0$, which is trivially satisfied. Putting the two together, we get the result.

□

Since we have not specified the distribution of the driving process X, we will just assume that the log-likelihood of the inverse satisfies a similar bound. Then, (26) will hold under the additional assumption that

$$\left|\ell_{\Delta X_{\mathcal{D}(n)}}\left(\Delta x_{\mathcal{D}(n)}\right) - \ell_{\Delta X_{\mathcal{D}(n)}}\left(\Delta \tilde{x}_{\mathcal{D}(n)}\right)\right| \leq \omega\left(d_p(x, \tilde{x})\right), \tag{28}$$

where x and $\tilde{x}$ are p-rough paths with $\Delta x_{\mathcal{D}(n)}$ and $\Delta \tilde{x}_{\mathcal{D}(n)}$ denoting increments on $\mathcal{D}(n)$ respectively.

To prove Theorem 6.1, it remains to show that for y and $y(n)$ as described in the theorem, the piecewise linear paths $I^{-1}_{\theta,n}(y_{\mathcal{D}(n)})$ and $I^{-1}_{\theta,n}(y(n)_{\mathcal{D}(n)})$ converge in p-variation in such a way that the different scalings also converge. Let us first look into the different scalings and how these affect the requirements for the convergence of the inverse piecewise linear paths.

Paper No. 16-02, www.warwick.ac.uk/go/crism

Lemma 6.3. *Suppose that the log-likelihood has the form* (22) *and that it is uniformly continuous in the p-variation topology, with some modulus of continuity* ω *independent of* θ *or* n*, i.e.*

$$|\ell_{Y(n)}\left(y_{\mathcal{D}(n)}|\theta\right) - \ell_{Y(n)}\left(\tilde{y}_{\mathcal{D}(n)}|\theta\right)| \leq \omega\left(d_p(I^{-1}_{\theta,n}(y_{\mathcal{D}(n)}), I^{-1}_{\theta,n}(\tilde{y}_{\mathcal{D}(n)}))\right) \tag{29}$$

Then, if for every $k = 0, \ldots, M$

$$\lim_{n\to\infty} n^{\alpha_k}\omega\left(d_p(I^{-1}_{\theta,n}(y_{\mathcal{D}(n)}), I^{-1}_{\theta,n}(\tilde{y}_{\mathcal{D}(n)}))\right) = 0,$$

it follows that

$$\lim_{n\to\infty} |\ell^{(k)}_{Y(n)}\left(y_{\mathcal{D}(n)}|\theta\right) - \ell^{(k)}_{Y(n)}\left(\tilde{y}_{\mathcal{D}(n)}|\theta\right)| = 0. \tag{30}$$

Proof. Let's denote by $R_k(y_{\mathcal{D}(n)}, \theta)$ the remainder of the log-likelihood expansion truncated at α_k for some $k \geq 0$, i.e.

$$R_k(y_{\mathcal{D}(n)}, \theta) = \sum_{j=k+1}^{M} \ell^{(j)}_{Y(n)}\left(y_{\mathcal{D}(n)}|\theta\right) n^{-\alpha_j} + R_M(y_{\mathcal{D}(n)}, \theta).$$

Since functions $\ell^{(j)}_{Y(n)}$ are convergent, we get that

$$n^{\alpha_k} R_k(y_{\mathcal{D}(n)}, \theta) \to 0, \text{ as } n \to \infty.$$

We first prove (30) for $k = 0$. We re-write (29) as

$$\left|\left(\ell^{(0)}_{Y(n)}\left(y_{\mathcal{D}(n)}|\theta\right) - \ell^{(0)}_{Y(n)}\left(\tilde{y}_{\mathcal{D}(n)}|\theta\right)\right) + \left(R_0(y_{\mathcal{D}(n)}, \theta) - R_0(\tilde{y}_{\mathcal{D}(n)}, \theta)\right) n^{\alpha_0}\right| \leq \\ \leq n^{\alpha_0}\omega\left(d_p(I^{-1}_{\theta,n}(y_{\mathcal{D}(n)}), I^{-1}_{\theta,n}(\tilde{y}_{\mathcal{D}(n)}))\right)$$

which implies that

$$\left|\ell^{(0)}_{Y(n)}\left(y_{\mathcal{D}(n)}|\theta\right) - \ell^{(0)}_{Y(n)}\left(\tilde{y}_{\mathcal{D}(n)}|\theta\right)\right| \leq \\ n^{\alpha_0}\omega\left(d_p(I^{-1}_{\theta,n}(y_{\mathcal{D}(n)}), I^{-1}_{\theta,n}(\tilde{y}_{\mathcal{D}(n)}))\right) + n^{\alpha_0}\left|R_0(y_{\mathcal{D}(n)}, \theta) - R_0(\tilde{y}_{\mathcal{D}(n)}, \theta)\right|.$$

Given that the n^{α_0} times the remainder disappears in the limit, for every $\epsilon > 0$, we can find n_0 such that $\forall n \geq n_0$

$$\left|\ell^{(0)}_{Y(n)}\left(y_{\mathcal{D}(n)}|\theta\right) - \ell^{(0)}_{Y(n)}\left(\tilde{y}_{\mathcal{D}(n)}|\theta\right)\right| \leq n^{\alpha_0}\omega\left(d_p(I^{-1}_{\theta,n}(y_{\mathcal{D}(n)}), I^{-1}_{\theta,n}(\tilde{y}_{\mathcal{D}(n)}))\right) + \epsilon.$$

proving the statement for $k = 0$. The proof of (30) for $k > 0$ is similar, building inductively on k. □

Finally, we show the following

Lemma 6.4. *Let I_θ^{-1} be the inverse Itô map defined by* (1). *Moreover, let* $Y(n, I_{\theta_0}(x)_{\mathcal{D}(n)})$ *and* $Y(n, I_{\theta_0}(\pi_n(x))_{\mathcal{D}(n)})$ *be the responses to the piecewise linear map as in* (2)*, parametrised by its values on the grid $\mathcal{D}(n)$, given by $I_{\theta_0}(x)_{\mathcal{D}(n)}$ and $I_{\theta_0}(\pi_n(x))_{\mathcal{D}(n)}$ respectively, where x is a fixed rough path in $G\Omega_p(\mathbb{R}^d)$ and $\theta_0 \in \Theta$. Then,*

$$\lim_{n\to\infty} d_p\left(I_\theta^{-1}\left(Y(n, I_{\theta_0}(x)_{\mathcal{D}(n)})\right), I_\theta^{-1}\left(Y(n, I_{\theta_0}(\pi_n(x))_{\mathcal{D}(n)})\right)\right) = 0, \tag{31}$$

provided that $d_p(\pi_n(x), x) \to 0$ as $n \to \infty$.

Proof. Under the assumption that b is invertible, I_θ^{-1} is given by (23), which is an integral of the rough path $Y(n, \cdot)$. By construction, these are piecewise linear paths. Let us denote them by $x(n) = I_\theta^{-1}\left(Y(n, I_{\theta_0}(x)_{\mathcal{D}(n)})\right)$ and $\tilde{x}(n) = I_\theta^{-1}\left(Y(n, I_{\theta_0}(\pi_n(x))_{\mathcal{D}(n)})\right)$. In particular, $x(n)$ is that path that when driving the system will go through the data points $I_{\theta_0}(x)_{\mathcal{D}(n)}$, so its response converges point wise to Y. This implies that $x(n)$ will also converge point wise to x or that the distance between $x(n)$ and $\pi_n(x)$ disappears point wise as $n \to \infty$. Now, using the fact that if the distance between two piecewise linear paths disappears as $n \to \infty$ and if one of them converges in p-variation, then the other will also converge in p-variation to the same limit, we conclude that

$$d_p(x(n), \pi_n(x)) \to 0, \text{ as } n \to \infty.$$

We have assumed that

$$d_p(\pi_n(x), x) \to 0.$$

Using the universal limit theorem and the continuity of integration in p-variation, this implies that

$$d_p(\pi_n(x), \tilde{x}(x)) \to 0, \text{ as } n \to \infty.$$

The result follows. □

The three lemmas together prove the theorem.

References

[1] W. G. Kelley and A. C. Peterson. *The Theory of Differential Equations* (Second Edition), Springer, 2010

[2] P. K. Friz, N. B. Victoir. *Multidimensional Stochastic Processes as Rough Paths: Theory and Applications*. CUP, Cambridge, 2010.

[3] T.J. Lyons, M.J. Caruana and T. Lévy. *Differential Equations Driven by Rough Paths*. Ecole d' Été de Probabilités de Saint-Flour XXXIV -2004. Springer, 2007.

[4] F. Baudoin. *Rough Path Theory*. Online notes: `http://www.math.purdue.edu/ fbaudoin/Rough.pdf`.

Paper No. 16-02, www.warwick.ac.uk/go/crism