

Neural Fractional SDE-Netによる金融時系列生成

Financial Time Series Generation with Neural Fractional SDE-Net

林 晃平^{1*} 中川 慧²
Kohei Hayashi, Kei Nakagawa

¹ 東京大学大学院数理科学研究科

¹ Graduate School of Mathematical Science, The University of Tokyo

² 野村アセットマネジメント株式会社

² Nomura Asset Management Co, Ltd.

Abstract: 本研究では、ニューラルネットワークを使用した時系列データの生成手法を提案する。株価のような現実の金融市場における時系列データは不規則にサンプリングされる場合が多く、更にそのノイズ構造は独立同分布やBrown運動（正規分布）よりも複雑であると言われている。このような特性を持つ時系列データを生成するために、Brown運動をベースとしたNeural Stochastic Differential Equation (SDE)モデルを拡張・一般化し、長期記憶特性を示すHurst指数が半分より大きい非整数階Brown運動に基づくNeural Fractional SDE-Net (fSDE-Net)を提案する。また、理論的にはfSDE-Netの数値解析手法を確立し、fSDE-Netの解の存在と一意性を示す。更に、人工データと実データを用いた実証分析を行い、fSDE-Netモデルが時系列データのHurst指数などの分布特性をうまく複製できることを示す。

1 はじめに

時系列データは様々な分野で現れ、そのモデリングは、時間発展の背後にある現象をより深く理解したり将来のシナリオを予測するための基本的な道具である。近年、深層学習(DNN)を用いた時系列データに特化した生成モデルが注目されている。特に、画像生成分野での[1]によるGenerative adversarial network (GAN)の成功([2])によって時系列データの生成においてもGANが利用されている([3, 4, 5, 6])。一般に、時系列データは、時刻でインデックスが付けられ、時間的に等間隔に連続した時点でサンプリングされた一連のデータである。しかし、時系列データが不規則にサンプリングされることも考えられる。実際例えば、金融データ(典型的には株価)は通常は営業日のみサンプリングされる。このように入力データのサンプリングが不規則な場合、時系列モデルがDNNのように十分な表現力を持っていても、ダイナミクスの学習は困難であると考えられる。このような課題を克服するアプローチとして、DNNと微分方程式(ordinary differential equation, ODE)の数値解法を組み合わせたNeural ODEモデル([7])が時系列データ生成の新しい方向性を示す手法として提案された。[8]は、ResNet([7])の時間的極限操作によって、ODEとして解釈できることを

指摘した。その後[9]は、ODE-Netの応用として不規則にサンプリングされた時系列データを生成した。一般に、ResNetの代わりにNeural ODEを用いることでパラメータの数が大幅に削減され、メモリを大幅に節約することができる。また、損失関数の勾配を計算する際に双対法を利用することができ、微分方程式を効率的に計算できる。Neural ODEが提案された後、いくつかの論文で、Brown運動によって駆動される確率微分方程式(stochastic differential equation, SDE)を利用して攪乱項をODEに導入されており、これらはNeural SDEと総称される。観測データのノイズ項をモデル化するとき、ノイズが独立かつ同一分布に従っていれば、Brown運動は最も自然な選択である。しかし、実際の時系列データには、金融データのようにノイズの増分が時間的に相関を持つ場合など、Brown運動よりも複雑なノイズ構造をもつ場合がある。また、Neural ODEのようにResNetから時間極限をとってNeural SDEを導出する場合にも、ノイズ項は必ずしもBrown運動に限定されない。実際、例えばドロップアウトやランダム初期化によってノイズ項を独立同分布として与えても、学習の進行やバッチ由来のランダム性を考慮することで層を跨いだ相関を持つことがある。

これらの観点から、より現実的な方法でノイズをモデル化する候補として、非整数階Brown運動([10])が挙げられる。非整数階Brown運動(fBm)は、0から1の間の値をとるHurst指数と呼ばれる実数 H でパラメー

*連絡先: 〒153-8914 東京都目黒区駒場 3-8-1
E-mail: kohei@ms.u-tokyo.ac.jp

タ付けられた連続値確率過程である。fBm は $H = 1/2$ のとき標準 Brown 運動に一致することが知られており、標準 Brown 運動の一般化になっている。また、Hurst 指数が大きくなると、fBm は時間的に長期の相関を示し、サンプルパスの正則性が改善し、解析が容易になる。

fBm は気象、通信、経済、ファイナンスなど多くの応用分野で重要なモデルである [11]。特に数理ファイナンスの分野では、fBm を用いたモデルが複数提案され、現実の市場を適切に記述するためのアプローチとして多くの成果を上げている [12]。例えば [13] は、一般的な株式リターンに長期記憶性があることを指摘した。一方で、市場のボラティリティは短期的に観察するとパスの正則性が低いことが報告されている。前述の fBm の性質により、この2つの視点はそれぞれ Hurst 指数 $H > 1/2$ 、 $H < 1/2$ の fBm によって再現できる。一方、fBm を用いたモデルでは、裁定機会の存在がしばしば証明されるが、同時に取引にコストがかかると裁定取引が不可能になることも示されており、より自然な現実のマーケットの記述として注目されている。すなわち、このような実際の金融市場における時系列データ性質は、fBm をノイズとして用いる強い動機付けとなる。そこで、本論文では既存の Neural SDE を一般化し、特に $H > 1/2$ の場合の非整数階 Brown 運動を用いて時系列の長期記憶性を反映させる新たな時系列生成モデルを提案する。

本研究の貢献は以下の通りである。

- まず SDE-Net を fBm を用いて拡張した fSDE-Net を提案した。理論的解析として、fSDE-Net の解の存在と一意性および Euler 法による数値解の収束性を証明した。
- 次に fSDE-Net の応用として、株式市場の時系列データに対する生成モデルを構築した。数値実験により、複数の生成モデルと比較することで元の時系列の分布特性、特に長期記憶性を再現することができることを示した。

2 先行研究

Neural SDE: Neural ODE にノイズ項を加え拡張し、SDE を考えるのには次の2つの動機がある。第一の動機は、入力データ自体のランダム性を考慮することである。[14] と [15] では、[16] で導入した離散版である深層潜在 Gauss モデルの連続的な拡張として SDE が考えられている。一方、[17] は DNN の変換を力学系の確率的な状態変化と見なしすことで不確実性をモデル化するために SDE を検討し、特に分布外検出タスクに適用している。また、[18] は時系列の生成モデルとして Neural SDE を用いた。SDE を考慮する第二の動機は、ニュー

ラルネットワークのパラメータのランダム化である。この点に関して、[19] はノイズ項を例えばドロップアウト正則化効果として与え SDE を考えることにより汎化性能と敵対的頑健性が改善されることを示した。一方、[20] は正則化効果に限定せず、重みとバイアスがランダムに与えられると仮定し、より一般的な設定で Neural SDE を導出した。

他の Neural Differential Equation モデルの拡張: Neural ODE の拡張は SDE 以外にもいくつか存在する。例えば、[21, 22] では観測データにジャンプがある場合を考え、どちらのモデルもランダムな時間にジャンプがある Neural ODE で記述されている。また [21] は、時間的な発展は連続的で、確率的な事象の発生により中断される可能性があるシステムを考察している。これに対して、[22] は2つの観測値間の条件付き期待値を Neural ODE を用いてモデル化し、新しい観測が行われるたびにジャンプするモデル化を行っている。一方、[23] は Controlled ODE (CDE) によって Neural ODE モデルを拡張できることを示し、RNN の連続表現として Neural CED を提案した。

3 準備

3.1 非整数階 Brown 運動

ここでは、非整数階ブラウン運動の定義と基本的な性質について概説する。 $H \in (0, 1)$ を定数とする。平均0の実数値 Gauss 過程 $B^H = \{B_t^H\}_{t \geq 0}$ が Hurst 指数 H の非整数階 Brown 運動 (fractional Brownian motion, fBm) であるとは、ほとんど確実に $B_0^H = 0$ かつ、任意の $s, t \geq 0$ に対して次を満たすときをいう。

$$\text{Cov}(B_s^H, B_t^H) = \frac{1}{2}(|t|^{2H} + |s|^{2H} - |t-s|^{2H}). \quad (1)$$

fBm は $H = 1/2$ のとき、標準 Brown 運動と一致する。また、fBm から $\{B_{t+1}^H - B_t^H : t = 0, 1, \dots\}$ という新たな離散確率過程を考えることができ、これを非整数階 Gauss ノイズ (fractional Gaussian noise, fGn) という。fBm の増分は $H > 1/2$ ($H < 1/2$) のとき正 (負) の相関をもつことは定義から明らかである。また、 $B_t^H - B_s^H \sim \mathcal{N}(0, |t-s|^{2H})$ なので、fBm は定常増分である。特に、Kolmogorov-Chentsov 連続性定理により、Hurst 指数 H の fBm のほとんど全てのパスは、任意の $\varepsilon > 0$ に対して $(H - \varepsilon)$ -Hölder 正則性を持つ。すなわち、任意の $s, t \geq 0$ に対して、 $|B_t^H - B_s^H| \leq C|t-s|^{H-\varepsilon}$ を満たす $C > 0$ が存在する。換言すれば、fBm のサンプルパスは Hurst 指数が大きくなるほどパスの正則性が向上する。(図 1 参照。)

次に、確率過程の長期記憶性を定義する。 $X = \{X_t\}_{t \geq 0}$ を確率過程とし、その増分を $X_{s,t} = X_t - X_s$ と書く。

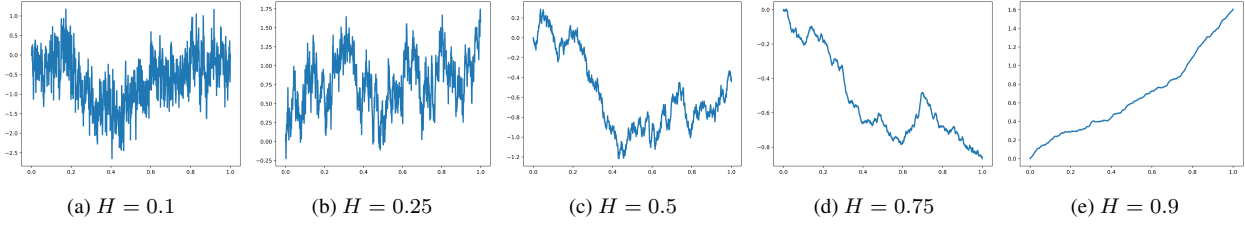


図 1: 種々の Hurst 指数に対する非整数階 Brown 運動のサンプルパス.

このとき、確率過程 X の増分が長期記憶性を持つとは、任意の $h > 0$ に対して

$$\sum_{n=1}^{\infty} |\text{Cov}(X_{0,h}, X_{(n-1)h,nh})| = \infty$$

が成立することをいう。(1)式から、fBm の増分に対しては $H > 1/2$ のときのみ長期依存性を持つことがわかる。

3.2 Hurst 指数の推定

ここでは [24] に倣い、時系列データに長距離依存性があるかを検定するための R/S 統計量を定義する。 $X = \{X_t : t = 0, 1, \dots, T\}$ を離散時系列とし、 X の時刻 T における R_T を次で定義する。

$$R_T = \max_{0 \leq k \leq T} \sum_{t=0}^k (X_t - \bar{X}_T) - \min_{0 \leq k \leq T} \sum_{t=0}^k (X_t - \bar{X}_T)$$

ここで、 $\bar{X}_T = (1/T) \sum_{0 \leq t \leq T} X_t$ は標本平均である。更に、標本標準偏差 S_T を次で定義する。

$$S_T = \left(\frac{1}{T} \sum_{t=0}^T (X_t - \bar{X}_T)^2 \right)^{1/2}.$$

このとき、 R_T/S_T を R/S 統計量という。[24]において、fGn の $T^{-H}(R_T/S_T)$ が $T \rightarrow \infty$ のとき定数に確率収束することが示されている。このとき特に、 R/S 統計量は漸近的に次の関係を満たす。

$$\log(R_T/S_T) = H \log T + \text{const.} + o_p(1) \quad (2)$$

ここで、 $o_p(1)$ は $T \rightarrow \infty$ のとき 0 に収束する項である。この等式を用いて、Hurst 指数は (2) 式の線形回帰式によって求めることができる。

3.3 非整数階 Brown 運動で駆動される確率微分方程式

提案手法である f SDE-Net を実装するため、Brown 運動を fBm に置き換えた SDE を考察する。以下、簡

単のためにドリフトとボラティリティが時間的に一様であると仮定し、以下のような fBm によって駆動される SDE を考える。

$$X_t = X_0 + \int_0^t b(X_s) ds + \int_0^t \sigma(X_s) dB_s^H. \quad (3)$$

SDE(3) 式の解を定義するためには、fBm に関する確率積分を定義する必要がある。しかし、Hurst 指数 H を持つ fBm は $H \neq 1/2$ のときセミマルチンゲールでないことが知られているため、通常の Itô 積分の意味では定義できない。しかし、前述の通り非整数階 Brown 運動のパスは Hurst 指数が大きいほど高い Hölder 正則性をもつ。 $H > 1/2$ の場合には、このようなパスの正則性を利用して fBm に関する積分を以下のように定義することができる。以下、 $[0, T]$ 上のすべての実数値 α -Hölder 連続関数の集合を $C^\alpha([0, T])$ と表記する。 $f \in C^\alpha([0, T])$ 及び $g \in C^\beta([0, T])$ を実数値関数とする。このとき、 $\alpha + \beta > 1$ ならば次の積分は収束することが [25] で知られている。

$$\int_0^T f(t) dg(t) = \lim_{|\Delta| \rightarrow 0} \sum_{i=1}^{n-1} f(t_i)(g(t_{i+1}) - g(t_i)).$$

ここで、 $\Delta = \{0 = t_0 \leq t_1 \leq \dots \leq t_{n-1} \leq t_n = T\}$ は $[0, T]$ の分割であり、 $|\Delta| = \max_{0 \leq i \leq n-1} |t_{i+1} - t_i|$ と定義した。この積分は Young 積分と呼ばれる。以上の結果をもとに、 $X = \{X_t : t \in [0, T]\}$ が (3) 式の SDE の解であるとは、fBm に対する積分が Young 積分で定義される X のほとんどすべてのパスが (3) 式を満たすことと定義する。

次に、(3) 式の SDE の well-posedness について述べる。輸送項 b と拡散項 σ が十分滑らかな関数であれば、(3) 式の解は fBm と同じ正則性を持ち、 $H > 1/2$ のとき fBm に関する積分が well-defined となると期待される。実際、次のように (3) 式の解の一意解の存在が保証されている。 $C_b^m(\mathbb{R})$ を、次数 m までの導関数がある有界であるような m 階連続微分可能な実数値関数の集合とする。

Proposition 3.1. $H > 1/2$ かつ $b, \sigma \in C_b^2(\mathbb{R})$ を仮定する、このとき、(3) 式は唯一の解をもつ。

4 提案手法:fSDE-Net

ここでは提案手法である fSDE-Net と先行研究である Neural ODE (SDE) との対応を述べる. 残差マップ $f(h_t, \theta_t)$ をもつ多層 ResNet のブロックを次のように表現する.

$$h_{t+1} = h_t + f(h_t, \theta_t). \quad (4)$$

ここで, $t \in \{0, \dots, T\}$ かつ h_t は i 番目の ResNet 層の入力であり, h_{t+1} は次の層への出力である. ResNet のパラメータを θ_t と書く. (4) 式において, 層数を追加し時間無限小の極限をとると, 隠れ層の連続表現は次の微分方程式 (ODE) として書ける.

$$\frac{dh_t}{dt} = f(t, h_t, \theta).$$

この関係式を利用したネットワークが Neural ODE である. この式に Brown 運動 $B = \{B_t : t \in [0, T]\}$ で駆動されるノイズ項を追加すると,

$$dh_t = f(t, h_t, \theta)dt + g(t, h_t, \theta)dB_t.$$

となり, これが Neural SDE に対応する. Neural SDE では, 時系列のトレンド部分は輸送項ネットワーク f で学習し, ノイズの影響は拡散項 g で学習する. 我々の提案する fSDE-Net は, Neural SDE において Brown 運動でモデル化されたノイズを fBm B^H に次のように一般化するものである.

$$dh_t = f(t, h_t, \theta)dt + g(t, h_t, \theta)dB_s^H.$$

4.1 数値計算スキーム

提案手法である fSDE-Net を実装するためには, SDE(3) を数値的に解く必要がある. Hurst 指数が $H > 1/2$ より大きい場合は古典的な Euler 法が有効である. ここでは離散化解の収束を保証する結果を述べる. $X^n = \{X_{i/n}^n : i = 0, \dots, [nT]\}$ を $X_0^n = X_0$ 及び次の漸化式により定まる離散系列とする.

$$X_{(i+1)/n}^n = X_{i/n}^n + b(X_{i/n}^n) \frac{1}{n} + \sigma(X_{i/n}^n)(B_{(i+1)/n}^H - B_{i/n}^H).$$

X^n を単関数を用いて $[0, T]$ 上の連続時間過程に拡張し, 同じ記号を用いて $X_t^n = X_{[nt]/n}^n$ と表す. このとき, 離散化した解 X^n の連続解への収束について, 次の命題が成立する. (証明は [26] を参照.)

Proposition 4.1. $H > 1/2$ とし, 輸送項および拡散項が $b \in C_b^2(\mathbb{R})$ と $\sigma \in C_b^3(\mathbb{R})$ を満たすとする. $X = \{X_t : t \in [0, T]\}$ を (3) 式の唯一解とし, $X^n = \{X_t^n : t \in [0, T]\}$ を上のように定義する. このとき, $n \rightarrow \infty$ で $n^{2H-1} \|X^n - X\|_{L^\infty([0, T])}$ はほとんど確実にある確率変数に収束する.

上の命題により, 離散化された解は精度 $O(n^{1-2H})$ で (3) の解に収束することが保証される.

4.2 誤差逆伝播法

ODE-Net に実装された高速な数値解法である双対法を fSDE-Net に対しても実現するために, ここでは fSDE の解のパラメータ微分について述べる. ある空間 Θ にパラメータを持ち, 各 $\theta \in \Theta$ に対して次の Hurst 指数 $H > 1/2$ の次の SDE に従う系列を考える.

$$dX_t = b(\theta, X_t)dt + \sigma(\theta, X_t)dB_t^H. \quad (5)$$

ここでは簡単のため, パラメータ空間 Θ はユークリッド空間の開集合であると仮定する. $\mathcal{L} : C([0, T]) \rightarrow \mathbb{R}$ を作用素とする. fSDE-Net においては, 作用素 \mathcal{L} は損失汎関数で, ニューラルネットのパラメータ θ に対する勾配を計算する必要がある. このとき, 損失汎関数がある良い性質を満たしていれば, 合成関数の微分に関する連鎖率により, $\partial_\theta \mathcal{L}(X)(\theta) = D\mathcal{L}(X) \circ \partial_\theta X(\theta)$ が各 $\theta \in \Theta$ に対して成り立つ. ここで, $D\mathcal{L}$ は $C([0, T])$ の各要素を $C([0, T])$ 上の線形汎関数の空間へ移す汎関数 \mathcal{L} の Fréchet 微分である. 上記のように fSDE の解をパスごとに与えれば, (5) 式は ODE として解釈でき, 次の結果が容易に証明できる.

Theorem 4.2. 各 $\theta \in \Theta$ に対して, $X_t = X_t^\theta$ を (5) 式の唯一の解とする. $b, \sigma : \Theta \times \mathbb{R} \rightarrow \mathbb{R}$ は $\theta \in \Theta$ および $x \in \mathbb{R}$ で微分可能であり, 各 θ に対して, $b_\theta(\theta, \cdot), b_x(\theta, \cdot), \sigma_\theta(\theta, \cdot), \sigma_x(\theta, \cdot) \in C_b^2(\mathbb{R})$ を仮定する. このとき, 各 $\theta \in \Theta$ に対して, $Y_t = \partial_\theta X_t$ は次の確率微分方程式を満たす.

$$dY_t = (b_\theta(\theta, X_t) + b_x(\theta, X_t)Y_t)dt + (\sigma_\theta(\theta, X_t) + \sigma_x(\theta, X_t)Y_t)dB_t^H. \quad (6)$$

Proof. (5) 式を積分表記すれば, 仮定より積分記号下での微分を適用することで (6) を導出できる. またこのとき, (6) 式の well-posedness は仮定より保証される. \square

5 fSDE-Net による金融時系列生成モデル

5.1 fBm による金融時系列モデル

数理ファイナンスでは, 時系列は通常, 標準 Brown 運動で駆動される SDE を用いてモデル化される. 一方, より現実的な市場を表現するために, fBm を用いたモデルも複数提案されている. 例えば, 株価過程 $S = \{S_t : t \geq 0\}$ に対しては, 非整数階 Black-Scholes モデ

Algorithm 1 Optimization of fSDE-Net generator

Input: $\{S_0, \dots, S_T\}$ - the stock price process until time horizon T , sample size M , learning rate η , number of optimization steps k

Output: $(\theta_b, \theta_\sigma)$ - the optimized parameter for fSDE-Net generator with some Hurst index H

while not converged **do**

for k steps **do**

 Let $\{\hat{S}_0^{(i)}, \dots, \hat{S}_T^{(i)}\}_{i=1}^M$ be the realization of generated paths of size M and let $p_\theta(t, \cdot)$ be the pdf of \hat{r}_t .

 Compute the gradient of loss function

$$\mathcal{L}(\theta) = -\frac{1}{T} \sum_{t=0}^T \log p_\theta(t, \log S_t).$$

 Descent parameters: $\theta \leftarrow \theta - \eta \nabla_\theta \mathcal{L}(\theta)$.

end for

end while

ル ([27]) や長期記憶確率ボラティリティモデル ([28]) などが知られている。

- $H \neq 1/2$ である非整数階 Black-Scholes モデルは次のように記述される。

$$\begin{cases} dr_t = rdt \\ dS_t = r_t S_t dt + \sigma_t S_t dB_t^H \end{cases}$$

- $H > 1/2$ であり $W = \{W_t : t \geq 0\}$ を Wiener 過程として長期記憶確率ボラティリティモデルは次のように記述される。

$$\begin{cases} dS_t = rS_t dt + \sigma_t S_t dW_t \\ d\sigma_t = \alpha \sigma_t dt + \beta dB_t^H \end{cases}$$

株価の時間発展が上記のモデルで記述されるとき、裁定取引の機会が存在することが知られている。一方、取引コストが存在する場合には裁定取引の機会が実現されないことも同時に証明されている。このことから、金融時系列を fBm でモデル化することは合理的であり、同時にパスの長期記憶性を考慮することが可能になる。以下では、対数価格過程の時間発展 $\log S_t$ が以下の SDE で記述されると仮定する。

$$d(\log S_t) = b(\theta, \log S_t)dt + \sigma(\theta, \log S_t)dB_t^H \quad (7)$$

輸送項 b と拡散項 σ をニューラルネットワーク (NN) により定め、それらを NN パラメータ θ について最適化する。

5.2 ニューラルネットワークの構成

ニューラルネットワークによって (7) 式の輸送項と拡散項を近似するために、以下の L -層の多層パーセプトロン (MLP) をニューラルネットワークとして使用する。

$$h_i^\ell = \sum_{j=1}^{N_{\ell-1}} w_{ij}^{\ell-1} x_j^{\ell-1} + b_i^{\ell-1}, \quad x_i^\ell = \varphi(h_i^\ell)$$

$\ell = 1, \dots, L$ かつ $i = 1, \dots, N_\ell$ である。このとき、 x^0 と x^L をそれぞれ入力データと出力データとし、NN 関数は $x^L = f_\theta(x^0)$ を満たし、 $\theta = \{w_{ij}^\ell, b_i^\ell\}_{i,j,\ell}$ とする。以降、 Θ を θ が値をとる NN パラメータの空間とする。このとき、fSDE-Net の解の存在と一意性が成り立ち、更に数値解法の精度保証が成立する。

Theorem 5.1 (Informal). $H > 1/2$ とする。ニューラルネットワーク構造が MLP で与えられ、活性化関数が \tanh の (7) 式の fSDE-Net の生成器は一意解を持ち、更に陽的 Euler 法を用いて数値的に解くことができる。

Proof. 活性化関数 \tanh は滑らかであり、任意の次数の導関数が有界であることに注意する。従って、関数の合成が正則性と有界性を遺伝することに注目すると、ニューラルネットワーク関数 f_θ は、任意の $\theta \in \Theta$ に対して命題 3.1 および 4.1 の仮定を満たすことがわかり、証明が完了する。□

5.3 アルゴリズム

以下では、 S_0, \dots, S_T を過去の株価とする。数値計算の都合上、元データを対数リターン過程 $\{r_0, \dots, r_T\}$ に変換し標準化することで、 $r_t = \log(S_{t+1}/S_t)$ が平均 0、分散 1 になるようにする。この前処理の後、輸送項と拡散項が MLP で与えられる fSDE(7) 式を、パラメータ $\theta \in \Theta$ について解き、その連続解を入力データの離散観測点に射影することで系列 $\hat{S}_0, \dots, \hat{S}_T$ を生成することができる。このように、fSDE-Net は不規則にサンプリングされた入力系列から元の経路と同じタイムスタンプを持つ合成パスを生成することが可能である。

この方法により M 個のサンプルパスを作成し、 i 番目の実現を $\{\hat{S}_0^{(i)}, \dots, \hat{S}_T^{(i)}\}$ で表す。次に、最尤推定法を適用して NN のパラメータを最適化し、より現実的なパスを生成することが可能になる。各タイムスタンプ $t \in 0, 1, \dots, T$ に対して、 $p_t(\theta, \cdot)$ を対数リターン $r_t = \log(\hat{S}_{t+1}^{(i)}/\hat{S}_t^{(i)})$ の確率密度関数とし、次の対数尤度を最大化する。

$$L(\theta) = \frac{1}{T} \sum_{t=0}^T \log p_\theta(t, \log S_t).$$

表 1: 非整数階 OU 過程 (fOU) と SP500 (SPX) の各パフォーマンス指標. 太字は最も良いものを表す.

データ	手法	Hurst 指数	周辺分布	ACF	加重 ACF	R^2
fOU	元系列	0.9 (真値)	-	-	-	-
	RNN	0.490 ± 0.105	1.070 ± 0.026	1.582	0.985	-17.470 ± 2.217
	SDE	0.495 ± 0.119	1.619 ± 0.065	1.709	1.123	-204.401 ± 56.750
	fSDE	0.866 ± 0.120	0.233 ± 0.143	3.487	1.805	-1.934 ± 1.855
	元系列	0.591	-	-	-	-
SPX	RNN	0.479 ± 0.094	0.473 ± 0.012	2.715	1.404	-3.185 ± 0.401
	SDE	0.513 ± 0.115	0.232 ± 0.067	3.052	1.720	-2.844 ± 1.282
	fSDE	0.529 ± 0.121	0.383 ± 0.062	2.785	1.469	-4.036 ± 1.132
	元系列	0.591	-	-	-	-

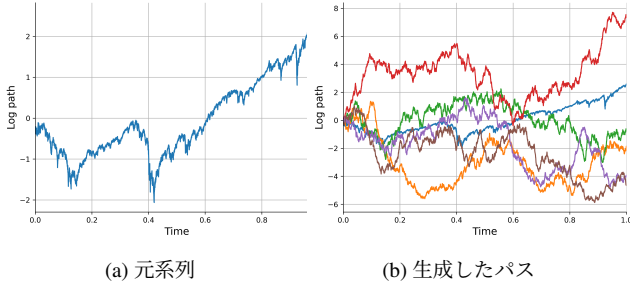


図 2: (a)S&P500 の対数価格と (b)fSDE-Net で生成した対数価格のパスの比較.

あるいは, 同様に NN のパラメータ $\theta \in \Theta$ について, $\mathcal{L}(\theta) = -L(\theta)$ という損失関数を最小化する.

ここで, 対数リターン分布は正規分布に従うと仮定し, 時刻 t における各尤度は平均と分散を $\{\hat{r}_t^{(1)}, \dots, \hat{r}_t^{(M)}\}$ として計算した正規分布の確率密度関数を用いて近似している. ここで, $\hat{r}_t^{(i)}$ は時間 t における i 番目の生成されたパスの対数リターンである. fSDE-Net の正確なステップはアルゴリズム 1 に記載されている. パラメータ θ に対する損失の勾配を計算するために, 効率的な計算と省メモリを実現するために, 定理 4.2 が役立つ.

6 実験

6.1 実験設定

本研究では次のデータ群を利用する.

人工データ: 人口データとして, 次の確率微分方程式で記述される非整数階 Ornstein-Uhlenbeck (OU) 過程を用いる.

$$dX_t = \alpha X_t dt + \beta dB_t^H.$$

ここで, $\alpha = -0.05$, $\beta = 0.1$, $H = 0.9$ とした. 特に $\alpha < 0$ の場合, 定常性が成り立つことが知られている. 非整数階 OU 過程の実現値は, 区間 $[0, 1]$ を 1000 個に分割することにより生成した. そのうち最初の 900 サンプルを訓練に使用し, 残り 100 個をテストに使用する.

マーケットデータ: 実データとしては, Bloomberg から取得した 2000 年 1 月から 2021 年 11 月までの S&P500 指数 (SPX) の終値を使用する. 訓練には 2020 年 12 月までのデータを使用し, 残りをテストに使用する. SPX の R/S 統計量による推定 Hurst 指数は先行研究 [29] で得られている値に近い 0.591 であり, この観測区間における長期記憶性を示している.

fSDE の数値解法を使用することで, fSDE-Net による時系列データを生成できる. 図 2 は, SPX の原系列と fSDE-Net によって生成された SPX の対数リターンのサンプルパスの比較を示している. 次に, いくつかの基準に基づいて fSDE-Net から生成された時系列データを定量的に評価する. 比較のため, 既存の時系列生成モデルとして再帰的ニューラルネットワーク (RNN) 及び SDE-Net を使用する. ネットワーク構造としては, SDE-Net 及び fSDE-Net は 20 個の隠れユニットをもつ 2 層 MLP を使用し, RNN に対しては 40 個のユニットを使用する.

6.2 評価指標

生成された時系列データが長期記憶性を持つかどうかを Hurst 指数を用いて確認するほか, 生成されたパスを元のパスと比較するために, ここでは周辺分布, 依存性, 予測スコアという基準を用いて評価する.

Hurst 指数: Hurst 指数 H を前述 [24] の方法で推定する. 具体的には, R/S 統計量が (2) 式を満たすことを利用し最初の 100 データを除去した後に線形回帰により Hurst 指数を推定する. Hurst 指数の値を比較することで, $H > 1/2$ かどうか, すなわち長期記憶性をもつかどうかを確認する.

周辺分布: 元の時系列データと生成されたデータの経験分布を比較する. ρ と $\hat{\rho}$ を, それぞれ元のデータと生成データのパスの経験確率密度関数とする. このとき, $\mathcal{B} = \{B_1, \dots, B_K\}$ に対して,

$$\frac{1}{2} \sum_{B \in \mathcal{B}} |\rho(B) - \hat{\rho}(B)|$$

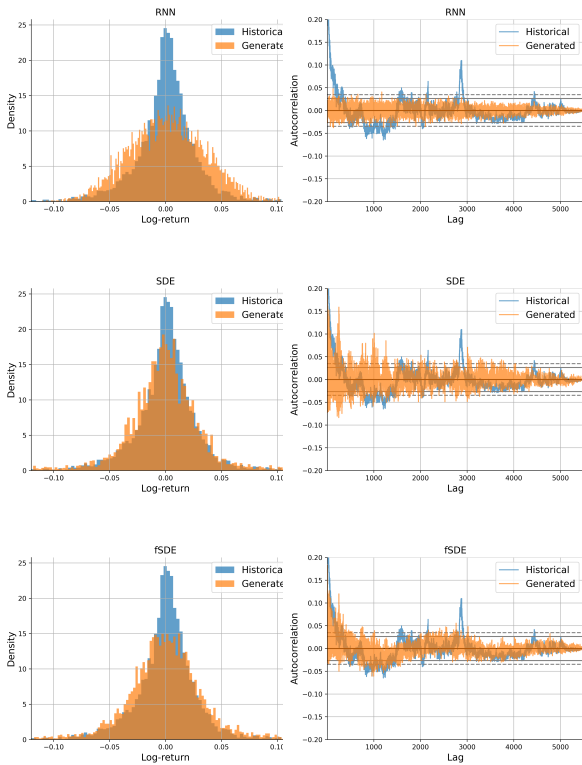


図 3: 対数リターン過程のヒストグラム (左列) とコロログラム (右列) について, 生成したパスと元データの S&P500 指数で比較したもの. RNN (上段), SDE-Net (中段), fSDE-Net (下段).

を元データと生成されたデータの分布間距離とする. ここで, $B \in \mathcal{B}$ に対して $\rho(B) = |B|^{-1} \int_B \rho(x) dx$ と定義した. $\hat{\rho}$ についても同様に定義される. また, 上記距離は $[0, 1]$ に値を取る.

依存性: 以下, $r_{0:T} = (r_0, \dots, r_T)$ と書く. また, $\gamma(\tau) = \text{Corr}(|r_t|, |r_{t+\tau}|)$ をラグ τ の絶対値対数リターンの相関係数とし, $C: \mathbb{R}^T \rightarrow [-1, 1]^S: r_{0:T} \mapsto (\gamma(1), \dots, \gamma(S))$ をラグ $S \leq T-1$ までの自己相関係数とする. このとき, [5] に基づき次を ACF スコアとして使用する.

$$\text{ACF} = \left\| C(r_{0:T}) - \frac{1}{M} \sum_{1 \leq i \leq M} C(\hat{r}_{0:T}^{(i)}) \right\|_2$$

ここで, $\|\cdot\|_2$ はユークリッド距離である. 更に, 次のような加重 ACF スコアを用いて, 自己相関がラグが大きいために減衰するかどうか (すなわち裾が重いか) を検証する.

$$\text{ACF}_w = \left\| C(r_{0:T}) \circ w - \frac{1}{M} \sum_{1 \leq i \leq M} C(\hat{r}_{0:T}^{(i)}) \circ w \right\|_2.$$

ここで, \circ は Hadamard 積であり, $w = (2i/(S+1))_{i=1, \dots, S}$ を重みとして使用する.

予測力: 最後に, 将来の時間発展を予測できるかどうかを評価するために, テストデータで計算された決定係数 R^2 を比較する.

6.3 結果

表 1 は, 学習率 0.04 の Adam を最適化ソルバーとして使用し, 100 ステップの計算を行った場合の性能指標を示したものである. 表中の Hurst 指数を見ると, fSDE-Net は他の手法に比べて高い値を示している. 特に, fSDE-Net で生成されたサンプルパスは長記憶性を示している. また, 非整数階 OU 過程においては, fSDE-Net はより良い周辺分布を生成することができる. 一方, SPX に対する値を見ると, fSDE-Net は他の手法と比較して優位性がないように見える. (図 3 を参照.) また, R^2 スコアは全て負の値を取っており, これは実世界の時系列に対するモデル化及び予測の難しさを示しているといえる.

7 結論

本論文では, Hurst 指数が $1/2$ 以上の非整数階 Brown 運動を用いた Neural ODE モデルの新しい枠組みである fSDE-Net を提案した. さらに, Euler 法による fSDE の数値解法を構築し, 長距離依存性を持ち, 更に不規則にサンプリングされた時系列データの生成に fSDE-Net を応用した. 数値シミュレーションの結果, fSDE-Net は既存の RNN や SDE-Net よりも長期記憶性を持つパスを生成する能力が高いことがわかった. 更なる研究の発展として, 時系列のパスのラフさを考慮するために, 正則性の低いノイズを含む SDE の解の理論とその数値解法を構築することで, Hurst 指数が $1/2$ 以下の場合への拡張がある.

今後の展望

本研究で提案した fSDE-Net のような金融時系列に特化した生成モデルの応用例として, オプション取引戦略やトレーディング戦略の評価がある. その裏付けとして, GAN を用いた DNN が古典的な DNN よりも現実的な時系列を生成することが実証されている. また, fBm によって駆動されるモデルは実データの当てはまりの良さを動機として提案されており, 我々の生成手法がヘッジ戦略の評価や構築にうまく適用できると期待される.

参考文献

- [1] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.
- [2] Jie Gui, Zhenan Sun, Yonggang Wen, Dacheng Tao, and Jieping Ye. A review on generative adversarial networks: Algorithms, theory, and applications. *IEEE Transactions on Knowledge and Data Engineering*, 2021.
- [3] Olof Mogren. C-rnn-gan: Continuous recurrent neural networks with adversarial training. *arXiv preprint arXiv:1611.09904*, 2016.
- [4] Jinsung Yoon, Daniel Jarrett, and Mihaela van der Schaar. Time-series generative adversarial networks. *Advances in Neural Information Processing Systems*, 32:5508–5518, 2019.
- [5] Magnus Wiese, Robert Knobloch, Ralf Korn, and Peter Kretschmer. Quant gans: Deep generation of financial time series. *Quantitative Finance*, 20(9):1419–1440, 2020.
- [6] Hao Ni, Lukasz Szpruch, Magnus Wiese, Shujian Liao, and Baoren Xiao. Conditional sig-wasserstein gans for time series generation. *arXiv preprint arXiv:2006.05421*, 2020.
- [7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [8] Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems*, pages 6572–6583, 2018.
- [9] Yulia Rubanova, Ricky TQ Chen, and David Duvenaud. Latent odes for irregularly-sampled time series. *arXiv preprint arXiv:1907.03907*, 2019.
- [10] Benoit B Mandelbrot and John W Van Ness. Fractional brownian motions, fractional noises and applications. *SIAM review*, 10(4):422–437, 1968.
- [11] Francesca Biagini, Yaozhong Hu, Bernt Øksendal, and Tusheng Zhang. *Stochastic calculus for fractional Brownian motion and applications*. Springer Science & Business Media, 2008.
- [12] S Rostek and R Schöbel. A note on the use of fractional brownian motion for financial modeling. *Economic Modelling*, 30:30–35, 2013.
- [13] Myron T Greene and Bruce D Fielitz. Long-term dependence in common stock returns. *Journal of Financial Economics*, 4(3):339–349, 1977.
- [14] Belinda Tzen and Maxim Raginsky. Theoretical guarantees for sampling and inference in generative models with latent diffusions. In *Conference on Learning Theory*, pages 3084–3114. PMLR, 2019.
- [15] Belinda Tzen and Maxim Raginsky. Neural stochastic differential equations: Deep latent gaussian models in the diffusion limit. *arXiv preprint arXiv:1905.09883*, 2019.
- [16] Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. Stochastic backpropagation and approximate inference in deep generative models. In *International conference on machine learning*, pages 1278–1286. PMLR, 2014.
- [17] Ling kai Kong, Jimeng Sun, and Chao Zhang. Sde-net: Equipping deep neural networks with uncertainty estimates. In *International Conference on Machine Learning*, pages 5405–5415. PMLR, 2020.
- [18] Patrick Kidger, James Foster, Xuechen Li, Harald Oberhauser, and Terry Lyons. Neural sdes as infinite-dimensional gans. *arXiv preprint arXiv:2102.03657*, 2021.
- [19] Xuanqing Liu, Si Si, Qin Cao, Sanjiv Kumar, and Cho-Jui Hsieh. Neural sde: Stabilizing neural ode networks with stochastic noise. *arXiv preprint arXiv:1906.02355*, 2019.
- [20] Stefano Peluchetti and Stefano Favaro. Infinitely deep neural networks as diffusion processes. In *International Conference on Artificial Intelligence and Statistics*, pages 1126–1136. PMLR, 2020.
- [21] Junteng Jia and Austin R Benson. Neural jump stochastic differential equations. *Advances in Neural Information Processing Systems*, 32:9847–9858, 2019.
- [22] Calypso Herrera, Florian Krach, and Josef Teichmann. Neural jump ordinary differential equations: Consistent continuous-time prediction and filtering. In *International Conference on Learning Representations*, 2021.
- [23] Patrick Kidger, James Morrill, James Foster, and Terry Lyons. Neural controlled differential equations for irregular time series. In *Advances in Neural Information Processing Systems*, volume 33, pages 6696–6707, 2020.
- [24] Benoit Mandelbrot. Statistical methodology for nonperiodic cycles: from the covariance to r/s analysis. In *Annals of Economic and Social Measurement, Volume 1, Number 3*, pages 259–290. NBER, 1972.
- [25] Laurence C Young. An inequality of the hölder type, connected with stieltjes integration. *Acta Mathematica*, 67(1):251, 1936.
- [26] Andreas Neuenkirch and Ivan Nourdin. Exact rate of convergence of some approximation schemes associated to sdes driven by a fractional brownian motion. *Journal of Theoretical Probability*, 20(4):871–899, 2007.
- [27] Walter Willinger, Murad S Taqqu, and Vadim Teverovsky. Stock market prices and long-range dependence. *Finance and stochastics*, 3(1):1–13, 1999.
- [28] Fabienne Comte and Eric Renault. Long memory in continuous-time stochastic volatility models. *Mathematical finance*, 8(4):291–323, 1998.
- [29] Erhan Bayraktar, H Vincent Poor, and K Ronnie Sircar. Estimating the fractal dimension of the s&p 500 index using wavelet analysis. *International Journal of Theoretical and Applied Finance*, 7(05):615–643, 2004.