Fractional Brownian motion

Fractional Brownian motion (fBM) is a natural generalization of standard Brownian motion.

A **zero-mean Gaussian process** is uniquely characterized by its **covariance function** $R(s,t) := \mathbb{E}(X_s X_t)$. $R(s,t)$ uniquely defines the process, since it determines the **covariance matrix** of $(X_{t_1}, ..., X_{t_n})$ for any ordered pair of time values $0 < t_1 < ... < t_n$, with (i,j)th element $R_{ij} := R(t_i, t_j)$, and for a zero-mean Gaussian vector, we only need its covariance to describe its **joint pdf**. Specifically, the joint density of a n-dimensional Gaussian random vector $\mathbf{X} = (X_1, ..., X_n)$ (with zero mean vector) is given by

$$p(\mathbf{x}) \quad = \quad p(x_1, .., x_n) \quad = \quad (2\pi)^{-n/2} \det(\Sigma)^{-\frac{1}{2}} \, e^{-\frac{1}{2}\mathbf{x}^T \mathbf{\Sigma^{-1}} \mathbf{x}} \tag{1}$$

where $\mathbf{\Sigma}$ is the matrix with i,jth element $\Sigma_{ij} = \mathbb{E}(X_i X_j)$, and det denotes the **determinant** of a matrix (if $n = 1$ this reduces to the density of a one-dimensional $N(0, \sigma^2)$ random variable, i.e. $p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{x^2}{\sigma^2}}$).

We can **sample** a Gaussian process X at $(t_1, ..., t_n)$ as $\mathbf{X} = \mathbf{CZ}$, where Z is a **column vector** $(Z_1, .., Z_n)$ of standard $N(0,1)$ random variables and C is the **unique lower triangular** $n \times n$ matrix such that $\mathbf{CC}^\top = \mathbf{\Sigma}$ (C is known as the **Cholesky decomposition** of Σ, see also FM06). Lower triangular means that $C_{ij} = 0$ if $i < j$, so the matrix looks like

$$C \quad = \quad \begin{bmatrix} C_{11} & 0 & .. & 0 \\ C_{21} & C_{22} & ... & 0 \\ ... & & & \\ C_{n1} & C_{n2} & ... & C_{nn} \end{bmatrix}$$

Note for any (non-zero) vector $\mathbf{x} = (x_1, ..., x_n)$:

$$\mathbb{E}((\sum_{i=1}^n x_i X_{t_i})^2) \quad = \quad \sum_{i=1}^n \sum_{j=1}^n R(t_i, t_j) x_i x_j \quad = \quad \mathbf{x}^\top \mathbf{\Sigma} \, \mathbf{x} \quad \geq \quad \mathbf{0}\,.$$

If we make the natural non-degeneracy assumption that $X_{t_1}, ..., X_{t_n}$ are **linearly independent** i.e. that $\sum_{i=1}^n x_i X_{t_i} \neq 0$ a.s. when at least one of the x_i's are non-zero, then clearly the square of this quantity $(\sum_{i=1}^n x_i X_{t_i})^2 > 0$ a.s. and hence its expectation $\mathbb{E}((\sum_{i=1}^n x_i X_{t_i})^2) = \sum_{i=1}^n \sum_{j=1}^n R(t_i, t_j) x_i x_j > \quad 0$. Hence Σ is **positive definite**, which (from standard results in linear algebra) implies that Σ has **positive determinant** and is **invertible** (which is needed for (1)).

The Cholesky method gives the correct covariance for X because

$$\mathbb{E}(\mathbf{X}\mathbf{X}^\top) \quad = \quad \mathbf{C}\mathbb{E}(\mathbf{Z}\mathbf{Z}^\top)\mathbf{C}^\top \quad = \quad \mathbf{C}\mathbf{C}^\top \quad = \quad \mathbf{\Sigma}$$

(where here we interpret $\mathbf{X}$ as a column vector), and we have used that

$$\mathbf{Z}\mathbf{Z}^\top \quad = \quad \begin{bmatrix} Z_1^2 & Z_1 Z_2 & .. & Z_1 Z_n \\ Z_2 Z_1 & Z_2^2 & ... & Z_2 Z_n \\ ... & & & \\ Z_n Z_1 & ... & ... & Z_n^2 \end{bmatrix}$$

so $\mathbb{E}(\mathbf{Z}\mathbf{Z}^\top) = \mathbf{I}$, i.e. the **identity matrix**.

A zero-mean Gaussian process B_t^H is called standard **fractional Brownian motion** (fBM) with **Hurst exponent** $H \in (0,1)$ if it has covariance function

$$R_H(s,t) \quad = \quad \mathbb{E}(B_t^H B_s^H) - \mathbb{E}(B_t^H)\mathbb{E}(B_s^H) \quad = \quad \mathbb{E}(B_t^H B_s^H) \quad = \quad \frac{1}{2}(t^{2H} + s^{2H} - (t-s)^{2H}) \tag{2}$$

$0 \leq s \leq t$ (note B^H can be defined for all $t \in \mathbb{R}$ or just $t \in [0, \infty)$).

- For $H = \frac{1}{2}$ and $0 \leq s \leq t$, we see that $R_H(s,t) = \frac{1}{2}(t + s - (t-s)) = s$, so we see that for $H = \frac{1}{2}$, $R_H(s,t) = \min(s,t)$, i.e. when $H = \frac{1}{2}$, fBM is just a **standard Brownian motion**.
- When $H \in (0, \frac{1}{2})$, B^H is **rougher** than standard BM, and when $H \in (\frac{1}{2}, 1)$, B^H is smoother than standard BM (see simulations in Figure 1 below); more specifically B^H is $H - \varepsilon$ **Hölder continuous** which means that $|B_t^H - B_s^H| \leq c_1(\omega)|t-s|^{H-\varepsilon}$ a.s. for any $\varepsilon \in [0, H)$ where $c_1(\omega)$ is a (in general random) constant depending on B^H itself (this comes partly from the **Kolmogorov Continuity Theorem**, see below for full statement and application to fBM.

We now prove some basic fundamental properties of fBM:

- $R(as, at) = a^{2H} R(s,t)$, so

$$X_{a(.)} \sim a^H X_{(.)}$$

(i.e. both processes on the left and right here have the same joint distribution at $(t_1, ..., t_n)$), so the process X is said to be **self-similar**, and in particular for a single fixed t-value we have $B^H_{at} \sim a^H B^H_t$. Note for $H = \frac{1}{2}$ this reduces to well known property of BM that $B_{at} \sim \sqrt{a} B_t$.

- From (2), for $0 \le s \le t$, we see that

$$\begin{aligned} \mathbb{E}((B^H_t - B^H_s)^2) &= \mathbb{E}((B^H_t)^2) + \mathbb{E}((B^H_s)^2) - 2\mathbb{E}(B^H_s B^H_t) &= t^{2H} + s^{2H} - (t^{2H} + s^{2H} - (t-s)^{2H}) \\ &&= (t-s)^{2H} \end{aligned}$$

so $B^H_t - B^H_s \sim N(0, |t-s|^{2H})$; since the answer only depends on the difference $t - s$, we say that B^H has **stationary increments**.

- Setting $s = n$ and $t = n+1$ in the previous bullet point, we see that $X_n = B^H_n - B^H_{n-1} \sim N(0,1)$ since $1^{2H} = 1$, so $(X_n)_{n=1}^{\infty}$ is a **discrete-time** Gaussian process with $X_n \sim N(0,1)$ for all k (but the X_n's are not independent of each other), and X is known as **fractional Gaussian noise** (**fGn**)); then $\rho_n = \mathbb{E}(X_{k+n} X_k)$ depends only on n (not k) and has **autocovariance** function:

$$\begin{aligned} \rho(n) := \mathbb{E}(X_{k+n} X_k) &= \mathbb{E}((B^H_{k+n} - B^H_{k+n-1})(B^H_k - B^H_{k-1})) \\ &= R_H(k+n, k) + R_H(k+n-1, k-1) - R_H(k+n, k-1) - R_H(k+n-1, k) \\ &= \frac{1}{2}[(n+1)^{2H} - n^{2H} - (n^{2H} - (n-1)^{2H})] \sim const. \times n^{2H-2} \qquad (n \to \infty) \end{aligned}$$

so we see X is stationary since $\rho(n)$ only depend on the time difference n and not k.

- Note it is quicker to simulate an fBM B^H by simulating a fGN X and then (using self-similarity) we define $B^H_{\frac{j}{n}} = \frac{1}{n^H} X_j$ for $j = 1..n$; this is quicker because the covariance matrix of fGN is constant along diagonals (since $\rho(n)$ above only depends on n), so computing the covariance matrix of X only requires $O(n)$ (not $O(n^2)$) distinct computations. We call this type of matrix a **Toeplitz matrix**.)

(By convexity of the function $g(n) := n^{2H}$), we see that $\mathbb{E}(X_{k+n} X_k) > 0$ if $H \in (\frac{1}{2}, 1)$ (which we call **persistent**) and $\mathbb{E}(X_{k+n} X_k) < 0$ for $H \in (0, \frac{1}{2})$ (which we call **anti-persistent**). Loosely speaking, for $H > \frac{1}{2}$, if B^H was increasing in the past, it is more likely to increase in the future, and vice versa. Similarly for $H < \frac{1}{2}$, if B^H was increasing in the past, it is more likely to decrease in the future, and vice versa.

- There exists a function $k(s,t)$ such that B^H can be realized as $B^H_t = \int_0^t k(s,t) dB_s$ where B is standard Brownian motion, and $k(s,t) \sim const.(t-s)^{H-\frac{1}{2}}$ as $s \nearrow t$, so k blows up as $s \nearrow t$ when $H \in (0, \frac{1}{2})$.

The Cholesky matrix above approximates the function k such that $B^H_t = \int_0^t k(s,t) dB_s$ where B is a standard Brownian motion **if you use the same Z vector to generate B and B^H, which you should do for this Task in Part 2 of the rough vol project**. In particular, for $0 < t < u$ we have the **conditional decomposition**:

$$B^H_u = \int_0^t k(s,u) dB_s + \int_t^u k(s,u) dB_s \,.$$

The two expressions on the right hand side are **independent**, and conditioned on B up to time t, B^H has conditional distribution which is $N(\int_0^t k(s,u) dB_s, \int_t^u k(s,u)^2 ds)$. In this sense we see that the process B^H has **memory**. Since $\int_0^t k(s,u) dB_s \neq B_t$ when $H \neq \frac{1}{2}$ and not just a simple function of B_t, we see that B^H is not a martingale, nor is Markov.

- A commonly used simpler version of this process is the **Riemann-Liouville** process $Z^H_t = \sqrt{2H} \int_0^t (t-s)^{H-\frac{1}{2}} dB_s$, for which $\mathbb{E}((Z^H_t)^2) = \mathbb{E}((B^H_t)^2)$ and is also self-similar, but no longer has stationary increments. Note that $Z^H_t \sim B^H_t$, but B^H and Z do not have the same covariance function and Z does not have stationary increments.

We now recall the **Kolmogorov continuity theorem**:

Theorem 0.1 *Let $\alpha, \varepsilon, c > 0$ and X be a random process which satisfies*

$$\mathbb{E}(|X_t - X_s|^\alpha) \quad \leq \quad c|t-s|^{1+\varepsilon}\,.$$

*Then X is γ-**Hölder continuous** for all $\gamma \in [0, \frac{\varepsilon}{\alpha})$.*

Application to fBM: From above, we know that

$$B_t^H - B_s^H \;\sim\; N(0,(t-s)^{2H}) \;\sim\; (t-s)^H Z$$

where $Z \sim N(0,1)$, so $\mathbb{E}(|B_t^H - B_s^H|^q) = \mathbb{E}(|Z|^q)(t-s)^{qH}$. Then applying the Kolmogorov continuity theorem to fBM with $\alpha = q$ and $1+\varepsilon = qH$, we see that B^H is γ-Hölder continuous for all $0 < \gamma < \frac{\varepsilon}{\alpha} = \frac{qH-1}{q}$ for any $q > 1/H$ which ensures that $qH - 1 > 0$.

But $\frac{qH-1}{q} \nearrow H$ as $q \to \infty$ because the qH term dominates the 1, so we can make the stronger statement that B^H is γ-Hölder continuous for all $0 < \gamma < H$. Note the theorem does not tell us that B^H isn't smooth, but in FM14 last year we proved the more precise statement that fBM is not $H + \varepsilon$-Hölder continuous.

Recall from above that the joint density (i.e. pdf) of a n-dimensional Gaussian random vector $\mathbf{X} = (X_1, ..., X_n)$ (with zero mean vector) is given by

$$p(\mathbf{x}) \quad = \quad p(x_1, .., x_n) \quad = \quad (2\pi)^{-n/2}\det(\Sigma)^{-\frac{1}{2}}\, e^{-\frac{1}{2}\mathbf{x}^T\mathbf{\Sigma^{-1}}\mathbf{x}}$$

$\mathbf{x} = (x_1, .., x_n)^T$, where $\mathbf{\Sigma}$ is the matrix with i,jth element $\Sigma_{ij} = \mathbb{E}(X_i X_j)$, and det here denotes the determinant of a matrix (details not required).

It is easier to work with the log of the joint density:

$$\log p(x_1, .., x_n) \quad = \quad (...) - \frac{1}{2}(\mathbf{x}^T\mathbf{\Sigma^{-1}}\mathbf{x} + \log\det\mathbf{\Sigma})$$

where the (...) terms aren't important since they are constant, and recall that $\mathbf{\Sigma}$ is a matrix which depends on H. $X_t = \nu B_t^H$. Then for a given realization of $\mathbf{x} = (X_1, .., X_n)$, we then just find the H and ν value which maximizes $\log p(X_1, .., X_n)$.

Project advice

In Task 2, we consider a sample path of the process $X = \nu B^H$, so we can set $\Sigma_{ij} = \mathbb{E}(B_{i/n}^H B_{j/n}^H)$ and $\mathbb{E}(X_{i/n}X_{j/n}) = \nu^2\Sigma_{ij}$. Then the log of the joint density of $(X_{t_1}, ..., X_{t_n})$ is

$$\log p(x_1, .., x_n) \quad = \quad (...) - \frac{1}{2}(\frac{1}{\nu^2}\mathbf{x}^T\mathbf{\Sigma}^{-1}\mathbf{x} + \log\det(\nu^2\mathbf{\Sigma})\,.$$

You need to code up this log density, and then maximize it over (H, ν) using something like `minimize(loglikelihood,params,method='Nelder-Mead',tol=1e-3, options="maxiter":300)` in Python, where `params0` here is the initial guess for (H, ν). You can use chatGPT to help with this, but you must be able to explain the code in the oral exam.

For the first task in Part 2, need to use the self-similarity of fBM to derive equality of the form

$$\frac{1}{n}\mathbb{E}(\sum_{i=1}^{n}|B_{i\Delta}^H - B_{(i-1)\Delta}^H|^q) \quad = \quad \Delta^p\,\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(|(B_i^H - B_{i-1}^H)|^q) \quad = \quad f(T,n,q,H)$$

for some power p and some function f which you need to determine, expressed in terms $K_q = \mathbb{E}(|Z|^q)$ where $Z \sim N(0,1)$. An **estimator** $\hat{H}_n$ for H is said to be biased if $\mathbb{E}(\hat{H}_n) \neq H$ (or unbiased otherwise). Use **Jensen's inequality** to prove whether $\hat{H}_n$ is biased in the first task and think about how the size of Δ affects the answer.

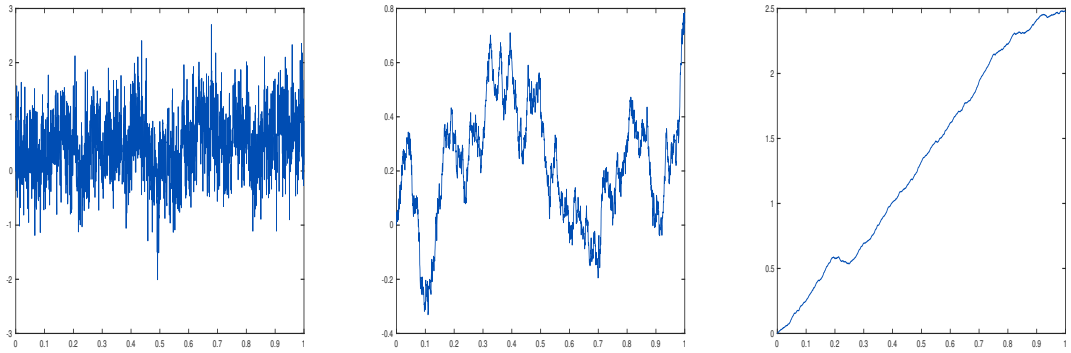

Figure 1: (i) Here we have plotted a Monte Carlo simulation of B^H using the Cholesky method for $H = .05$, $H = .5$ and $H = 0.9$

For the Monte Carlo task in Part 2, for extra marks I recommend pricing call options with Monte Carlo using the well known **Renault-Touzi formula**:

$$\begin{aligned} \text{CallPrice} &= \mathbb{E}(\max(S_T - K, 0)) = \mathbb{E}(C^{BS}(\tilde{S}_0, K, \sigma = 1, \tau = (1-\rho^2)\int_0^T \sigma_u^2 du, r = 0)) \\ \text{where} \quad \log \tilde{S}_0 &= \log S_0 - \frac{1}{2}\rho^2 \int_0^T \sigma_u^2 du + \rho \int_0^T \sigma_u dB_u \end{aligned} \tag{3}$$

where $\sigma_t = \sqrt{V_t}$ and $C^{BS}(S, K, \sigma, \tau, r)$ is the usual Black-Scholes (BS) formula from e.g. FM02, i.e. we use the BS formula with a **random** maturity τ and (when $\rho \neq 0$) a random initial stock price $\tilde{S}_0$ (proof of why this formula works is in Chap1 document from FM14). This formula means we only have to simulate the second Brownian motion B and not the first Brownian motion W to estimate the call price, which reduces the **sample variance** (one should also use antithetic variables as well, i.e. use the mirror image of B and then take the average of the two MC approximations for the call price).

We approximate (3) in the usual way with

$$\frac{1}{M}\sum_{i=1}^{M} C^{BS}(\tilde{S}_0^i, K, 1, (1-\rho^2)\int_0^T (\sigma_u^i)^2 du, 0)$$

where σ^i is the i'th Monte Carlo path, and M is the total number of MC paths (I am ignoring antithetic paths here). Then (using the MC paths are i.i.d), the variance of this expression is

$$\frac{1}{M}\text{Var}(C^{BS}(S_0, K, 1, (1-\rho^2)\int_0^T (\sigma_u^i)^2 du, 0))$$

so the **sample standard deviation** of the MC call price estimate scales like $\frac{1}{\sqrt{M}}$, which we can also use to condtruct e.g. confidence intervals for the call price.

It is important to check the convergence of a Monte Carlo estimate for a call price by varying the Number of time steps (with a fixed random seed) and the Number of paths (without fixing, since convergence is typically quite slow and bias large for rough models with realistic parameter values. I recommend e.g. at least 500 time steps and 1million paths for this task, if your computer doesn't crash (much better to use Colab than Cocalc).

Recall the definition of **implied volatility**: $\hat{\sigma}$ is the unique volatility value $\hat{\sigma}$ such that Call Price $= C^{BS}(S, K, \hat{\sigma}, T, r)$, where in our case here Call Price will be the Monte Carll price appoximation.

One important numerical trick to be aware of for the MLE computation: sometimes computing $\log(\det(\Sigma))$ will give NaN if the covar matrix is too big (i.e. number of data points N in the sample path is too large), but can get round this by multiplying the covariance matrix Σ by a constant $c_1 > 1$ (e.g 2, 4 or whatever makes it work), so $\det(c_1\Sigma) = c_1^N \det(\Sigma)$ (basic linear algebra result), hence

$$\log\det(\Sigma) = \log\det(c_1\Sigma) - N\log c_1$$

so the expression on the right hand side is what one should compute in Python to compute $\log\det(\Sigma)$.