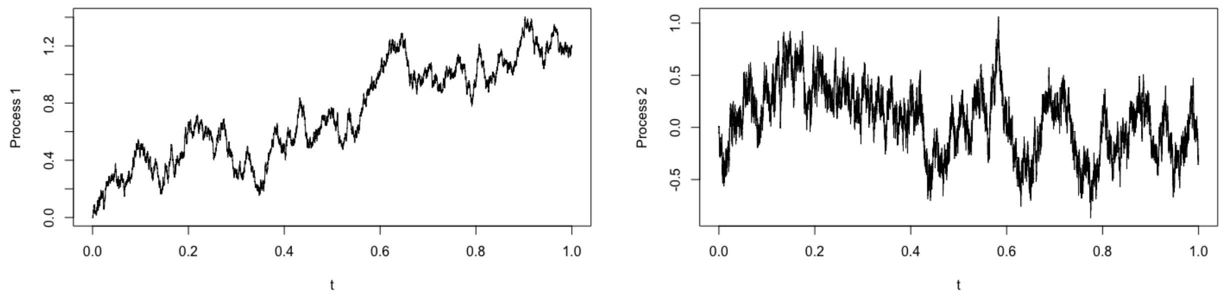
Hurst index estimation under measurement errors

Introduction

What is the difference between the following two processes?



Clearly, the second process is rougher (i.e., less regular) than the first process. In this project, we want to study a class of processes X_t^H called *fractional Brownian motions (fBM)*. If you are familiar with *Brownian motion (BM)*, fBMs are generalizations of BM indexed by a parameter $H \in (0,1)$ called the *Hurst index*: if $H = \frac{1}{2}$, fBM is simply BM; if $H \in (\frac{1}{2}, 1)$, fBM is *smoother* than BM; and if $H \in (0, \frac{1}{2})$, fBM is *rougher* than BM. For example, in the simulation above, the first process is a BM, while the second is an fBM with $H = 0.3$. It is known that, with probability 1, the paths of fBM with Hurst index H are Hölder continuous of order α if and only if $\alpha > H$. Hence, H can be interpreted as a parameter of *regularity* (*roughness/smoothness*).

Estimating H as a parameter of regularity is often important in applications. Suppose that we can observe an fBM X^H on a fine grid $I_n = \{\frac{i}{n} : i = 1, \ldots, n\}$ of $[0,1]$, where n is large. Can we figure out H from these observations? Put in proper statistical language, this question becomes: Can we *consistently estimate* H based on the data, that is, can we construct an estimator $\widehat{H}_n$ from the data $\{X^H_{i/n} : i = 1, \ldots, n\}$ such that $\widehat{H}_n$ converges in probability to H as $n \to \infty$? And if so, can we determine the *rate* (i.e., the speed of convergence) of $\widehat{H}_n$? The answer is "yes" to both questions, and the idea behind is simple: by the properties of fBM, the variance of an increment $\Delta_i^n X^H := X^H_{i/n} - X^H_{(i-1)/n}$ is $\mathbb{E}[(\Delta_i^n X^H)^2] = Cn^{2H}$ for some constant C. From this, one can show that

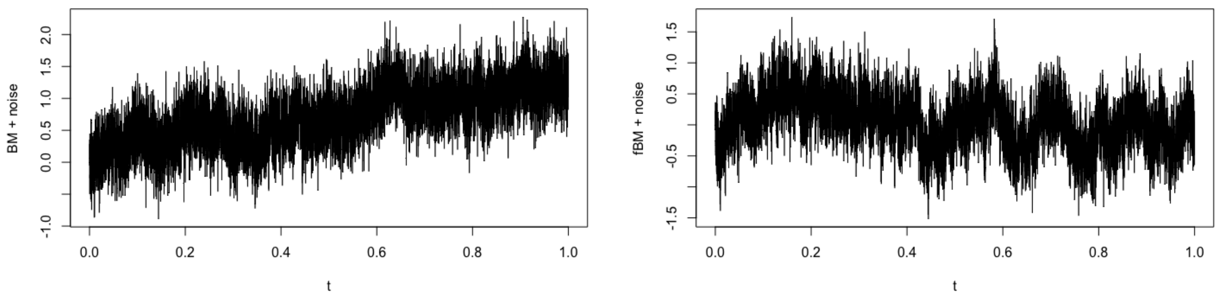
$$V^n(X^H) := \frac{1}{n} \sum_{i=1}^{n} (\Delta_i^n X^H)^2 \sim Cn^{2H},$$

in the sense that $n^{-2H} V^n(X^H) \xrightarrow{\mathbb{P}} C$. The functional $V^n(X^H)$ is called the *(normalized) quadratic variation* of X^H and the fact that H appears in the scaling of this quantity allows us to construct a consistent estimator for H. It can further be shown that this estimator has a rate of $n^{-1/2}$, which is best possible.

But in practice, often one cannot observe $X^H_{i/n}$ directly: the instruments used to measure $X^H_{i/n}$ incur measurement errors ϵ_i^n, so that, in fact, one can only observe

$$Y^H_{\frac{i}{n}} = X^H_{\frac{i}{n}} + \epsilon_i^n,$$

where $(\epsilon_i^n)_{i=1,\ldots,n}$ is a sequence of iid normal random variables with mean 0 and variance v. If such a noise with $v = 0.1$ is added to the paths simulated above, what we get is this:



Suddenly, it is much less obvious, maybe even unclear, which one is BM plus noise and which one is fBM plus noise. Using some modifications of $V^n(X^H)$, one can prove that H can still be consistently estimated, but the optimal rate now goes down to $n^{-1/(4H+2)}$.

Project description

Apart from measurement noise, there is another phenomenon that complicates matters: in many situations encountered in practice, the stochastic fluctuations (i.e., the local variance of the increments) of X_t^H are not constant over time but time-varying and *random*. In particular, the resulting processes are no longer Gaussian. Depending on the field of applications, people refer to this as *stochastic volatility* or *intermittency*.

In the world of statistics, this means that we have to move from a *parametric setting* and to a *non-parametric / semi-parametric setting*. Fortunately, the quadratic variation based estimators of H still do their job: they continue to be consistent with rate $n^{-1/2}$, but the proof of this fact becomes much more involved. This is well known in the literature. What is open—and to fill this gap is precisely the goal of this project—is how to estimate H when both stochastic volatility / intermittency *and* measurement errors are present, which is by far the most realistic scenario in practice. We will mostly study this estimation problem from a theoretical perspective. Given time and interest, applications to data may be considered.

To be more precise, the goals of this project will be to:

1. Learn some basic results about BM and fBM. Here you will encounter some classic concepts in probability such as convergence, martingales and stochastic integration.

2. Study existing results about $V^n(X^H)$ and understand existing estimators for H when (a) measurement noise is present *or* (b) stochastic volatility / intermittency is present.

3. Find and analyze estimators for H when *both* measurement noise *and* stochastic volatility / intermittency are present. This is the main research problem of the project.

4. (Optional:) Apply the results to turbulence data.

Prerequisites

You must be well trained in analysis and have had at least one probability theory class—MATH GU4155 or equivalent. I assume that you are familiar with things like Taylor's theorem, law of large numbers, central limit theorems and convergence in probability / in distribution. If you already know what Brownian motion or stochastic integrals are, even better. Given that this project is quite ambitious, you need to have a strong affinity for proofs and be willing to learn fast. In exchange, you have the opportunity to delve into an interesting area of applied probability / mathematical statistics that bridges theory and applications. I am happy to provide reading material if you feel the need to catch up before the official start.