Studying Self-similar Processes Using the Crossing Tree

Nazarov Ivan, 101мНОД(ИССА)
Supervised by: Decrouez Geoffrey Gerard

August 14, 2016

Abstract

The Crossing Tree is a recent tool, created to study scale invariance properties exhibited by network traffic, financial time-series, and natural phenomena in physics and biology, such as turbulence, earthquakes and heart beat. It provides an ad-hoc representation of the data across a wide range of resolutions and independent of its time scale. In this paper the theoretical and practical construction of the Crossing Tree is presented. This tool is applied to analyse self-similar signals of processes known as Hurst self-similar processes with stationary increments to gather evidence of the common features of the constructed crossing trees in this class of processes.

Contents

1 Introduction **2**

2 The crossing tree **3**
- 2.1 Crossing Times . . . 3
- 2.2 Tree Structure . . . 4
- 2.3 Crossing tree on time series data . . . 5
- 2.4 Analysis using the crossing tree . . . 6
- 2.5 The crossing tree of Brownian motion . . . 8

3 H-sssi processes **10**
- 3.1 Definition . . . 10
- 3.2 Fractional Brownian motion . . . 11
- 3.3 Hermite processes . . . 11
- 3.4 Weierstrass function . . . 12

4 the Crossing tree and H-sssi processes **13**
- 4.1 Generation of fBm, Hermite and Weierstrass processes . . . 13
- 4.2 Simulation results . . . 14

5 Conclusion and further work **25**

1. Introduction

Scale-invariance and self-similarity has been widely used lately to analyse a wide range of time series, from medicine, to seismology, and to financial data. Essentially, scale-invariance of some process implies that its dynamics is not characterized by well defined dominant time-scales. Instead their dynamics involve all time-scales. Analysing this kind of data should rely on identifying the mechanism relating the scales with one another. The crossing tree, described in section 2 (p. 3), is a natural tool for analysing data at all scales simultaneously.

The structure of this paper is as follows. The second chapter (p. 3) of this study describes the crossing tree, its theoretical construction as well as practical computation. In the third chapter we review the definition and classical examples of self-similar processes with stationary increments (p. 10). The last chapter (p. 13) discusses the results of a Monte-Carlo simulation of crossing trees constricted for samples paths of various H-sssi processes, and attempts verify whether these processes share common crossing tree statistical properties.

2. The crossing tree

The crossing tree as a tool for analysing self-similarity goes back to the study of diffusion on fractal sets (see [2]). It was later reintroduced by Jones and Shen in [10] in the context of studying self-similarity and regularity patterns in network traffic data, where where it was successfully used in detecting structural changes, manifested in changes in local regularity of the time series. In this section I describe theoretical construction of the crossing tree and how it is recovered from the discrete time series data.

2.1 Crossing Times

Consider a path of real-valued continuous process $\{X(t)\}$ with time running through $[0, +\infty)$. Continuity is understood in the sense of the path of the process being almost surely continuous, i.e. the set of discontinuous paths of $\{X(t)\}$ is a subset of a set of probability zero. At the heart of the crossing tree lie crossing times of a uniformly spaced grid on $\mathbb{R}$ given by $X(0) + \delta 2^n \mathbb{Z}$ for some $n \geq 0$ and base grid spacing $\delta > 0$.[1] Each $n \geq 0$ represents the "resolution" through which sample paths of $\{X(t)\}$ are studied.

In particular, the crossing times T^n_k of process $\{X(t)\}$ at resolution $n \geq 0$ are defined as follows: let $T^n_0 = 0$ and for any $k \geq 0$ put

$$T^n_{k+1} = \inf\Big\{ t \geq T^n_k : \big|X(t) - X(T^n_k)\big| \geq \delta 2^n \Big\}$$

In other words T^n_{k+1}, if it is finite, is the first passage time of a path of $X(t)$ across a level of $\delta 2^n \mathbb{Z}$ that is different from the line of the same grid crossed previously. Fixing the zero-th crossing at 0 automatically aligns the grid with the process, so in effect, without loss of generality one may consider processes X_t, which start at the origin $X(0) = 0$.

By definition, the crossing times of some fixed level $n \geq 0$ are non-decreasing: $T^n_k \leq T^n_{k+1}$ for all $k \geq 0$. However, since the paths of the process are almost surely continuous, we show that $T^n_k < T^n_{k+1}$. We proceed by contradiction. Indeed, by right-continuity of $X(t)$ (a fixed path of $\{X(t)\}$) for $\delta 2^{n-1}$ there is $\eta > 0$ with $|X(s) - X(T^n_k)| < \delta 2^{n-1}$ for all $s \in [T^n_k, T^n_k + \eta)$. However, by definition of T^n_{k+1}, in the η vicinity of T^n_{k+1}, there exists $\hat{s}$ such that $|X(\hat{s}) - X(T^n_k)| \geq \delta 2^n$. Thus, if $T^n_k = T^n_{k+1}$ then one has $\delta 2^n \leq \delta 2^{n-1}$ – a contradiction.

Another important property of crossing times is that if $T^n_{k+1} < +\infty$, then $|X(T^n_{k+1}) - X(T^n_k)| = \delta 2^n$. Indeed, the definition of T^n_{k+1} implies that there exits a sequence $s_j \downarrow T^n_{k+1}$ with $|X(T^n_{k+1}) - X(T^n_k)| \geq \delta 2^n$. Since the path $X(t)$ is continuous at T^n_{k+1} it must be true that

$$|X(T^n_{k+1}) - X(T^n_k)| = \lim_{j\to\infty} |X(s_j) - X(T^n_k)| \geq \delta 2^n \,.$$

Now, since the function $t \mapsto |X(t) - X(T^n_k)| - \delta 2^n$ is continuous at T^n_{k+1}, $|X(T^n_{k+1}) - X(T^n_k)| > \delta 2^n$ would entail the existence of $s \in [T^n_k, T^n_{k+1})$ such that $|X(s) - X(T^n_k)| - \delta 2^n > 0$. This would contradict the fact that T^n_{k+1} is a lower bound of the set of all $s \geq T^n_k$ with $|X(s) - X(T^n_k)| \geq \delta 2^n$.

So far the crossing times of an almost surely continuous process $\{X(t)\}$, have the following properties (almost surely):

1. $T^n_k < T^n_{k+1}$ for all $k \geq 0$;

2. $|X(T^n_{k+1}) - X(T^n_k)| = \delta 2^n$ for all $k \geq 0$ with $T^n_{k+1} < +\infty$.

In order to see the natural tree structure of the crossings, one has to investigate the properties of the crossing times at different resolutions. We suppose, that for some integers $k, m \geq 0$, $T^n_k \in [T^{n+1}_m, T^{n+1}_{m+1})$ and $X(T^{n+1}_m) = X(T^n_k)$.

By definition of T^n_{k+1} for all $s \in [T^n_k, T^n_{k+1})$ it must be true that

$$|X(s) - X(T^n_k)| < \delta 2^n \,,$$

[1] The addition $x_0 + A$, for A in an affine space such as $\mathbb{R}$ represents shifting of every element of the set A by a common value x_0.

otherwise T^n_{k+1} would not have been the least lower bound. Since the process takes the same value at both times T^{n+1}_m and T^n_k, it must be true that $|X(s) - X(T^{n+1}_m)| < \delta 2^{n+1}$ for such s, whence $s \leq T^{n+1}_{m+1}$. Therefore $T^n_{k+1} \leq T^{n+1}_{m+1}$.

However, if one further assumes that $T^{n+1}_{m+1} < +\infty$, i.e. that the $(n+1)$-st crossing takes place, then $T^n_{k+1} \neq T^{n+1}_{m+1}$. Indeed,

- $|X(T^n_{k+1}) - X(T^n_k)| = \delta 2^n$;
- $|X(T^{n+1}_{m+1}) - X(T^{n+1}_m)| = \delta 2^{n+1}$.

together with $X(T^{n+1}_m) = X(T^n_k)$ entail the contradiction $\delta 2^n = \delta 2^{n+1}$.

The assumption that $(n+1)$-level crossing took place (grid $\delta 2^{n+1}\mathbb{Z}$) allows further refinement of crossing times' behaviour: under the assumptions above, $T^n_{k+2} \leq T^{n+1}_{m+1}$ holds.

We proceed by contradiction, and suppose otherwise: $T^{n+1}_{m+1} < T^n_{k+2}$. Then by definition of T^n_{k+2} and because $T^n_{k+1} \leq T^{n+1}_{m+1}$, it must be true that

$$|X(T^{n+1}_{m+1}) - X(T^n_{k+1})| < \delta 2^n \,.$$

However $|X(T^{n+1}_{m+1}) - X(T^{n+1}_m)| = \delta 2^{n+1}$ by the established properties of crossing times on the grid of $n+1$ resolution. Then by the triangle inequality and the assumption that $X(T^{n+1}_m) = X(T^n_k)$ one gets

$$\begin{aligned} |X(T^{n+1}_{m+1}) - X(T^{n+1}_m)| &\leq |X(T^{n+1}_{m+1}) - X(T^n_{k+1})| + |X(T^n_{k+1}) - X(T^{n+1}_m)| \\ &= |X(T^{n+1}_{m+1}) - X(T^n_{k+1})| + |X(T^n_{k+1}) - X(T^n_k)| \,, \end{aligned}$$

whence the following contradiction emerges

$$\delta 2^{n+1} \leq |X(T^{n+1}_{m+1}) - X(T^n_{k+1})| + \delta 2^n < \delta 2^{n+1} \,.$$

Therefore, under the stated conditions $T^n_{k+2} \leq T^{n+1}_{m+1}$.

2.2 Tree Structure

Let's define the *n-th level crossing* as the slice of the path of the process $\{X(t)\}$ over the time half-interval $[T^n_k, T^n_{k+1})$ for some integer $k \geq 0$, during which it makes a move of $\pm\delta 2^n$. The uncovered relationship between crossing times of consecutive levels imply that if $T^n_k \in [T^{n+1}_m, T^{n+1}_{m+1})$ with $X(T^{n+1}_m) = X(T^n_k)$ and $T^{n+1}_{m+1} < +\infty$ then

$$\big[T^n_k, T^n_{k+1}\big) \cup \big[T^n_{k+1}, T^n_{k+2}\big) \subseteq \big[T^{n+1}_m, T^{n+1}_{m+1}\big) \,.$$

The first condition, $X(T^{n+1}_m) = X(T^n_k)$ aligns the different resolution grid with one another. The last requirement that $[T^{n+1}_m, T^{n+1}_{m+1})$ be a finite interval is natural, since it actually means that the $(n+1)$-st level crossing actually occurred.

These observations actually state that within the $(n+1)$-st level crossing there musts be exactly an even number of crossing of the n-th level (subcrossings). Indeed, if $T^n_k = T^{n+1}_m$ and $T^{n+1}_{m+1} < +\infty$, then there are two possibilities for the crossing time T^n_{k+2}:

1. the process crossed a new level of $(n+1)$-st grid $\delta 2^{n+1}$ in which case $T^{n+1}_{m+1} \leq T^n_{k+2}$. Hence there are exactly *two* subcrossings.
2. the process moved back to the level $X(T^n_k)$, which did not incur a crossing of $(n+1)$-st grid, since $X(T^n_{k+2}) = X(T^{n+1}_m)$. Yet this fluctuation was registered by the grid of resolution n. In this case, the properties of crossing times entail

   $$T^n_{k+2} < T^n_{k+3} < T^n_{k+4} \leq T^{n+1}_{m+1} \,.$$

   Since T^{n+1}_{m+1} is finite, recurrent application of this argument implies that there must be exactly an even number of n-th level subcrossings in an $(n+1)$-st level crossing.

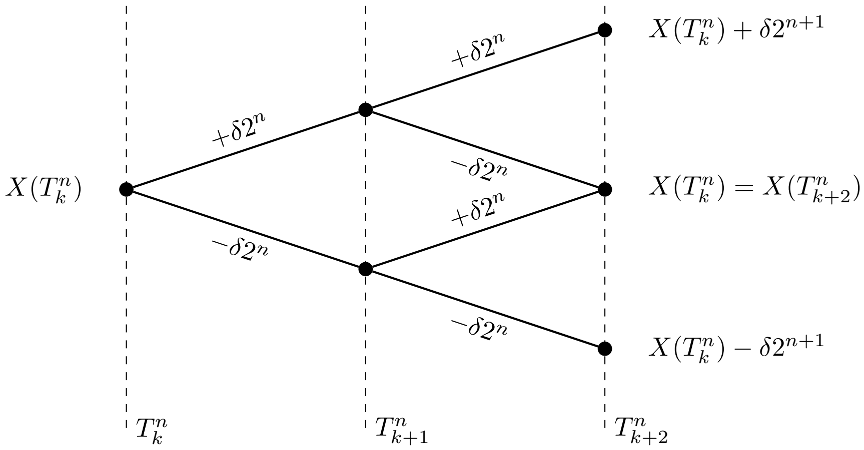

Figure 2.1: The structure of possible movements during the crossing $\left[T_k^n, T_{k+2}^n\right)$.

This described behaviour is depicted in figure 2.1.

Thus the properties of crossing times of a continuous process permit a natural tree structure of crossings. If tree levels are enumerated from bottom up then

- a node at the $(n+1)$-st tree level represents a crossing of the $\delta 2^{n+1}\mathbb{Z}$;
- n-th level children (offspring) of an $(n+1)$-st level node represent the multiple crossings of a finer gird $\delta 2^n\mathbb{Z}$ that took place during the larger crossing.

The crossing tree reveals patterns of subcrossings within a parent crossing, conditional on the direction of the latter. Denote the crossing orientations by

$$\alpha_k^n = \operatorname{sign}\bigl(X(T_{k+1}^n) - X(T_k^n)\bigr)\,.$$

Then, by construction of the crossing tree itself, subcrossing orientations of a complete crossing come in pairs by construction. The pairs of oppositely oriented crossings are *excursions*. Thus orientations (signs) $-+$ and $+-$ correspond to down-up and up-down excursions, respectively, whereas excursions $--$ or $++$ – to direct upward or downward crossings, respectively. Any complete parent crossing can be broken down into a series of excursions followed by a direct crossing, which actually concludes the parent crossing by passing over the next $\pm\delta 2^{n+1}$ grid line. Let V_k^n be 0 if the k-th n-level excursion (movements within a finer $(n-1)$-st gird) is up-down (if $\alpha_{2k}^{n-1}\|\alpha_{2k+1}^{n-1} = +-$) and 1 if it is down-up $(-+)$.

The above exposition considered the crossing tree for successively coarser resolutions $\delta 2^n\mathbb{Z}$ for $n \geq 0$. For example in [4] the crossing tree is defined for crossing times on this grid for any $n \in \mathbb{Z}$. The restriction to non-negative levels in this paper was motivated by the fact, that in practice, on real sample path of some supposedly continuous process it is never possible no meaningfully go beyond some basic finest resolution given by $\delta > 0$. This has to do with the fact that processes are sampled at finite frequency or over regularly spaced but finitely many points in time. However, it is important to note that the crossing tree for resolutions finer than the base scale give a one-to-one characterisation of the original continuous random process in terms of the base scale, and crossing times $(T_k^n)_{k\geq 0}$ and crossing orientations $(\alpha_k^n)_{k\geq 0}$, $n \in \mathbb{Z}$ (see [4] and [9]).

2.3 Crossing tree on time series data

In this section we briefly describe how the crossing tree is constructed for a sample path of some continuous process, given by time series data $(t_i, x_i)_{i=0}^N$.

In practice the crossing tree is built iteratively from the finest resolution specified by the base grid spacing $\delta > 0$. The crossing times of the leaves of the tree are computed for a fixed and given $\delta > 0$ in two passes: the first pass computes passage times on the passed $x_0 + \delta\mathbb{Z}$ grid levels, and the second pass adjusts the passage times and passed levels to eliminate re-crossings of the same level (recrossing events). The assumption that the time series data came from a continuous process is the main reason, that enables a sequential procedure for estimation of the crossing times of grid $\delta\mathbb{Z}$: recall, that $X(T_{k+1})^n = X(T_k^n) \pm \delta 2^n$ if $T_{k+1}^n < \infty$. We refer to the appendix for details of the procedure.

The choice of δ is extremely important as it affects the estimates of the crossing times and levels in the following way, reminiscent of the bias-variance trade-off:

- the grid with too small value of δ would make the resolution so fine as to regard the studied process as a continuous piecewise linear function. Each increment would be very likely to produce long consecutive unidirectional streaks of crossed levels, which would overestimate the number of crossings with only 2 subcrossings by introducing excess number of artificial crossings events.

- A too large δ leads to a tree with fewer crossings and offspring, which is insufficient for reliable estimates of tree features.

One approach to selection of δ is to consider the increments of the time series and measure either their standard deviation, or inter-quartile range, or compute the median of their absolute values. In fact the effect of the base scale decreases with the level of the tree, provided the tree is tall enough.

2.4 Analysis using the crossing tree

Once a crossing tree has been constructed for a particular sample path of a continuous process $\{X(t)\}$ for a given base resolution $\delta > 0$, it can be used to gauge scale-invariance of the given realisation. Recall that the nodes of level n of the tree are crossings on the grid $\delta 2^n$, and the offspring of each level $n+1$ node correspond to the subcrossings of a finer grid $\delta 2^n$ that constitute the parent crossings. An example of a sample path of a process with superimposed crossing times and levels is given on figure 2.2. The corresponding crossing tree is depicted on fig. 2.3.

Let Z_k^n be the number of subcrossings of a finer grid $\delta 2^{n-1}$ that take place during the k-th crossing of grid $\delta 2^n$, and N_n be the total number of crossings of level n. In general N_n is defined as the last $k \geq 0$ such that $T_k^n < +\infty$. Since it is possible that the very last crossing of the n-th level is incomplete, i.e. the crossing time $T_{N_n}^n = +\infty$, we have the following inequality:

$$\sum_{k=1}^{N_{n+1}} Z_k^{n+1} \leq N_n \, .$$

The sequence of numbers of subcrossings, given by $(Z_k)_{k=1}^{N_n}$, can be used to estimate the offspring distribution. If the process is scale invariant and is ergodic, it is reasonable to expect that the offspring distributions of crossings of consecutive levels of the tree are identical, provided the levels are well sampled. To this end one can use a standard χ^2 squared test:

1. Take h bins $\{2m\}$ for $m = 1, h-1$ and $\{2h, 2h+2, \ldots\}$ and compute the empirical probabilities $\hat{p}_k^n$ of the number of offspring (Z_k^n) on each level $n = p, p+1, \ldots, q$ falling into each bin;

2. Pool all the offspring together into one pool Z, and compute the empirical probabilities $\bar{p}_i$ of the pooled offspring Z_k being in each of the bins;

3. Provided each bin has sufficiently many observations, the following statistic

$$t^{p,q} = \sum_{j=p}^{q} \sum_{i=1}^{h} N_i \frac{(\hat{p}_i^n - \bar{p}_i)^2}{\bar{p}_i}$$

   tests the hypothesis that the offspring distributions on all levels from p to q have the same distribution. The test statistic has asymptotic χ^2 distribution with $(q-p)(h-1)$ degrees of freedom.

Another important characteristic of the process, which is made apparent by the crossing tree, is the distribution of patterns of subcrossings within a parent crossing, conditional on its direction. Indeed, subcrossing orientations of a complete crossing come in pairs by construction of the crossing tree itself:

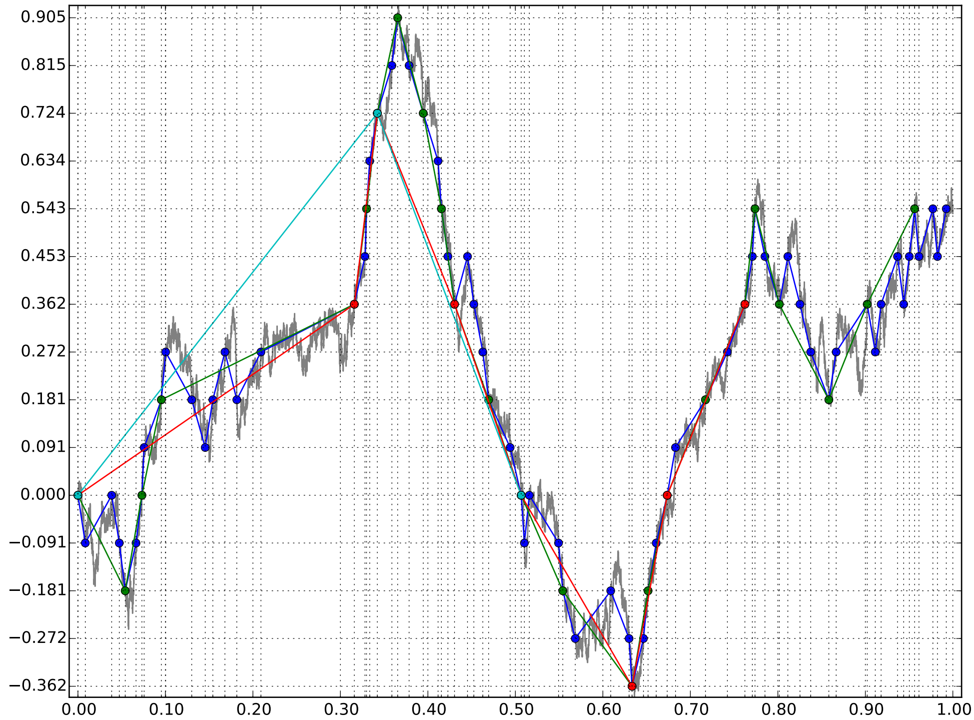

Figure 2.2: The crossing times and levels of a sample trajectory of a continuous process.

orientations $-+$ and $+-$ correspond to down-up and up-down excursions respectively, whereas $--$ or $++$ – to direct upward or downward crossings. In fact any complete crossing can be broken down into a series of excursions followed by a direct crossing, which actually end the current crossing, since they constitute a transitions to a next $\pm\delta 2^{n+1}$ grid line. Let V_k^n be 0 if the k-th n-level excursion (movements within a finer $(n-1)$-st gird) is up-down and 1 if it is down-up.

If the process is continuous, self-similar and has stationary and ergodic increments (see [11]), then the sequences of the number of subcrossings $Z^n = (Z_k^n)_{k\geq 0}$ are identically distributed, stationary and ergodic. The same result holds for the sequences of properly scaled crossing durations $2^{-n}W^n = (2^{-n}W_k^n)_{k\geq 0}$, where W_k^n is the duration of the k-th crossing of level n and is defined as $W_{k+1}^n = T_{k+1}^n - T_k^n$.

For H-sssi processes (see p.10) one can use the crossing tree to estimate the value of the Hurst index H by the following formula:

$$\hat{H} = \frac{\log 2}{\log \hat{\mu}}\,,$$

where $\hat{\mu} = \mathbb{E}Z_k^n$. Intuitively, if we one scales the time of an H-sssi process by μ, the expected number of finer subcrossings in a cruder crossing, then it is necessary to scale down the range of the process by $\frac{1}{2}$ – the multiplier by which the resolution is decreased when ascending the tree from the finest grid $\delta\mathbb{Z}$. In practice, in order to improve the accuracy of this estimate one attempts to pool sequences Z^n across several consecutive levels of the tree, over which the hypothesis of common distribution cannot be rejected.

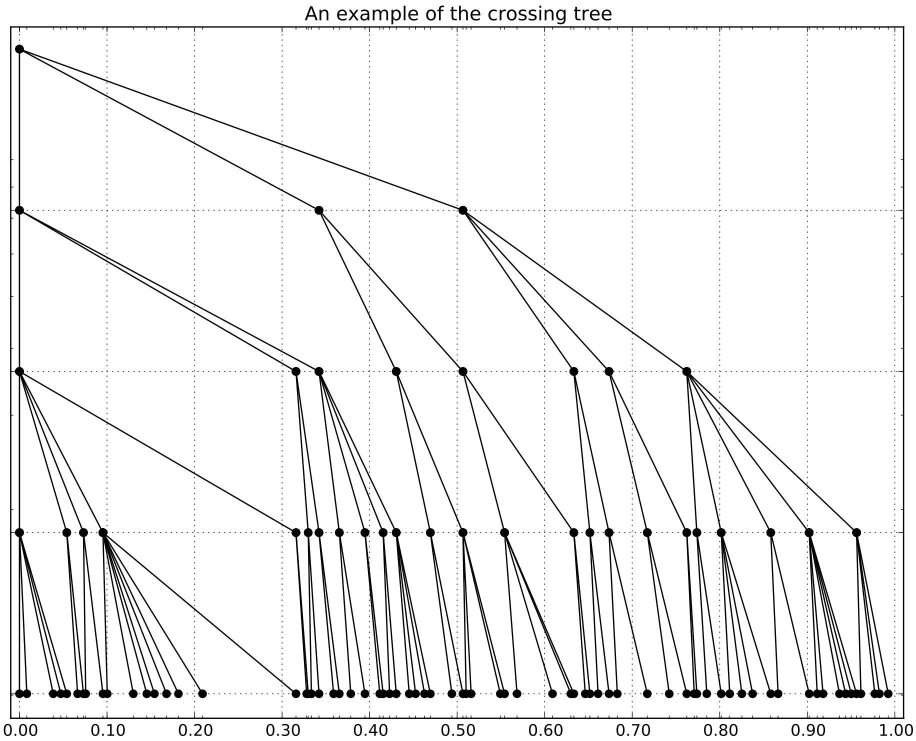

Figure 2.3: The tree structure of the crossing times for the path depicted in fig. 2.2.

2.5 The crossing tree of Brownian motion

One of most well studied processes is the Brownian motion. Usually it is axiomatically as a unique stochastic process $\big\{W(t)\big\}$ with the following four properties:

- $W(t)$ is almost surely 0 at $t = 0$;
- $\{W(t)\}$ has independent and stationary increments;
- for all $t \geq 0$ the random variable $W(t)$ has Gaussian distribution with parameters mean 0 and variance t;
- the sample paths of $\{W(t)\}$ are almost surely continuous functions.

This process exhibits scale-invariance properties, or self-similarity, since for any $a > 0$, the process $\big\{a^{-\frac{1}{2}}B(at)\big\}$ and $\{B(t)\}$ are stochastically equivalent in the sense of equality of all their finite-dimensional joint distributions.

The paper [9] gives a rigorous proof of the statistical properties of the offspring distribution, crossing lengths and the excursions for crossing trees built over paths of Brownian motion. Theorem 1 of [9], p. 640, states that the Brownian motion is the unique continuous process $\{B(t)\}$ for which the crossing tree has infinitely many offspring at each level ($N_n = +\infty$), and:

BM0 $B(0) = 0$;

BM1 the properly scaled crossing durations $\delta^{-2}4^{-n/2H}W_k^n$ for $H = \frac{1}{2}$ are identically distributed with mean 1 and finite variance, and are mutually independent within each level n;

BM2 the random variables Z_k^n are iid with probability distribution

$$\mathbb{P}\big(Z_k^n = 2m\big) = 2^{1-H^{-1}}\big(1 - 2^{1-(1/2)^{-1}}\big)^{m-1},$$

for all $m = 1, 2, \ldots$;

BM3 the excursions V_k^n are also iid random variables with $\mathbb{P}(V_k^n = 1) = 2^{-2H^{-1}}$, for $H = \frac{1}{2}$. Bern($\frac{1}{2}$).

This theorem crucially depends on the strong Markov property of Bm, which makes Z_k^n independent random variables. The characterisation of Brownian motion in terms of the properties of its crossing tree is based on ideas used in the construction of Brownian motion on a nested fractal (see [2]).

3. H-sssi processes

This section covers the basic definitions and properties of self-similar processes studied in this project. We consider a process $\{X(t)\}$ to be a continuous -time real-valued stochastic process defined for all $t \in [0, +\infty)$. The contents of this section are quite general and are based on the following papers and textbooks: [13], [1], [3] [6] and [7] to name but a few.

3.1 Definition

Before proceeding with the definitions and properties, we clarify what is meant by stochastic equivalence of random processes. Processes $\{X(t)\}$ and $\{Y(t)\}$ are equivalent in finite-dimensional distributions, or symbolically $\{X(t)\} \stackrel{\mathcal{D}}{=} \{Y(t)\}$, if for all $n \geq 1$ and all $(t_k)_{k=1}^n \in [0, +\infty)$ with $t_k < t_{k+1}$,

$$\big(X(t_k)\big)_{k=1}^n \stackrel{\mathcal{D}}{\sim} \big(Y(t_k)\big)_{k=1}^n \text{ holds},$$

where $A \stackrel{\mathcal{D}}{\sim} B$ denotes equality of distribution of random variables A and B.

A process $\{X(t)\}_{t\geq 0}$ is called **self-similar**, or **ss** for short, if for any $a > 0$ there exists $b > 0$ such that

$$\{X(at)\} \stackrel{\mathcal{D}}{=} \{bX(t)\},$$

that is, their finite-dimensional distributions coincide.

A process $\{X(t)\}$ is said to be stochastically continuous at $t \geq 0$ if $\lim_{h\to 0} \mathbb{P}\{|X(t+h) - X(t)| \geq \epsilon\} = 0$ for arbitrary $\epsilon > 0$.

It was shown by Lamperti in 1962 that whenever a self-similar process $\{X(t)\}$ has non-degenerate point distributions (non-trivial), and is stochastically continuous at $t = 0$, there (necessarily) exists a constant $H \geq 0$ such that any $a > 0$ the constant b in the definition of self-similarity is given by $b = a^H$, i.e

$$\{X(at)\} \stackrel{\mathcal{D}}{=} \{a^H X(t)\}.$$

Uniqueness of H follows from the fact that $b_1 X \stackrel{\mathcal{D}}{\sim} b_2 X$ implies $b_1 = b_2$ for any non-degenerate random variable X.

This theorem gives the definition of H-**s**elf-**s**imilarity: a process $\{X(t)\}$ is H-ss with Hurst exponent H if for all $a > 0$,

$$\{X(t)\} \stackrel{\mathcal{D}}{=} \{a^{-H} X(at)\}$$

A stochastic processes $\{X(t)\}$ is said to have **S**tationary **I**ncrements, or **si**, if all finite-dimensional distributions of the process $\{X(t+s) - X(t)\}_{s\geq 0}$ are independent of t, which equivalently means that

$$\{X(t+s) - X(t)\} \stackrel{\mathcal{D}}{=} \{X(s) - X(0)\}.$$

A particularly nice corollary to the definition of an H-self-similar process is that if $\{X(t)\}$ is H-ss then $X(t) \stackrel{\mathcal{D}}{\sim} t^H X(1)$. In turn, this allows, for example to show that all zero-mean stochastic processes with stationary increments share similar auto-correlation pattern for all $t \neq s$

$$\mathbb{E}X(s)X(t) = \frac{1}{2}\Big(t^{2H} + s^{2H} - |t-s|^{2H}\Big)\mathbb{E}|X(1)|^2.$$

Indeed, the identity $2ab = a^2 + b^2 - (a-b)^2$ implies that

$$\begin{aligned}
\mathbb{E}X(s)X(t) &= \frac{1}{2}\Big(\mathbb{E}X(s)^2 + \mathbb{E}X(t)^2 - \mathbb{E}\big(X(s) - X(t)\big)^2\Big) \\
&= \frac{1}{2}\Big(\mathbb{E}X(s)^2 + \mathbb{E}X(t)^2 - \mathbb{E}X(|s-t|)^2\Big) \\
&= \frac{1}{2}\Big(|s|^{2H}\mathbb{E}X(1)^2 + |t|^{2H}\mathbb{E}X(1)^2 - |s-t|^{2H}\mathbb{E}X(1)^2\Big),
\end{aligned}$$

where the second and the third lines follow from the stationarity of the increments and the mentioned corollary, respectively.

Lastly, recall that a process $\{X(t)\}$ is said to have **i**ndependent **i**ncrements if for any integer $n \geq 1$ and every $(t_k)_{k=0}^n \in [0, \infty)$ with $0 = t_0$ and $t_k < t_{k+1}$ the random variables $\big(X(t_k) - X(t_{k-1})\big)k = 1^n$ and $X(0)$ are jointly independent.

To summarize a process $\{X(t)\}$ is H-sssi, if it is H-self-similar with Hurst exponent H and has stationary increments.

3.2 Fractional Brownian motion

It was mentionend in section 2.5 that Brownian motion is an example of a self-similar process. In fact it is an $1/2$-sssi process. Indeed, let $a > 0$ and consider the process $V(t) = \frac{1}{\sqrt{a}} B(at)$. Obviously $V(0) = 0$ almost surely, since $V(0) = 1/\sqrt{a} B(a0) = 0$. Furthermore this process inherits stationary increments from $B(t)$:

$$\begin{aligned}\{V(t+s) - V(t)\} &= \left\{\frac{1}{\sqrt{a}}(B(at+as) - B(at))\right\} \\ &\overset{\mathcal{D}}{=} \left\{\frac{1}{\sqrt{a}}(B(as))\right\} \\ &= \{V(s)\},\end{aligned}$$

and independence of increments as well: pick $(t_k)_{k=1}^n$ with $t_k < t_{k+1}$ and observe that $\big(V(t_{k+1}) - V(t_k)\big)_{k=1}^n$ are mutually independent since $\big(B(s_{k+1}) - B(s_k)\big)_{k=1}^n$ for $s_k = at_k$ are. Path continuity follows from continuity of sample paths of $B(t)$ and the fact that $t \to at$ is a continuous map. Finally $\mathcal{N}(0, at) \sim \sqrt{a}\mathcal{N}(0, t)$ implies that $V(t)$ is Brownian motion as well.

Another example of an H-sssi process, which is frequently used as a reference for self-similarity and scale invariance studies is the **f**ractional **B**rownian **m**otion, or **fBm**. It is a generalisation of the $\frac{1}{2}$-ss Brownian motion to a general H-ss Gaussian process with $H \in (0, 1)$.

Formally, fractional Brownian motion introduced in [12] is a zero-mean Gaussian process $\{X(t)\}$, (its every finite-dimensional joint distributions are multivariate normal), with the covariance structure of a general zero-mean H-sssi stochastic process:

$$\mathbb{E}X(s)X(t) = \frac{1}{2}\big(t^{2H} + s^{2H} - |t-s|^{2H}\big).$$

Theorem 1.3.3 on p. 6 of [7] establishes that a fractional Brownian motion $\{B_h(t)\}_{t \geq 0}$ is an H-sssi process, for the following integral representation up to a multiplicative constant

$$\int_{-\infty}^{0} (t-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}} dB(s) + \int_{0}^{t} |t-s|^{H-\frac{1}{2}} dB(s).$$

If $H = 1$ then the fractional Brownian motion degenerates to $B_1(t) = tB(1)$ (almost surely). The class of all fractional Brownian motions coincides with the class of all Gaussian self-similar processes with stationary increments, see [7].

In fact, fractional Brownian motion is an example of a broader class of self-similar processes known as the **Hermite** processes. They exhibit non-Gaussianity and have strongly dependent increments.

3.3 Hermite processes

Hermite processes inherit their name from the stochastic integral kernel used in their definition, and are an extremely important example of processes, which have non-Gaussian finite-dimensional joint distributions.

A probabilistic Hermite polynomial of order $k \geq 0$ is defined as

$$H_k(x) = (-1)^k e^{-\frac{x^2}{2}} \frac{d^k}{dx^k} e^{-\frac{x^2}{2}},$$

and is a solution to the following differential equation

$$\frac{d}{dx}\left(e^{-\frac{x^2}{2}}\frac{d}{dx}f\right)+\lambda e^{-\frac{x^2}{2}}f=0\,.$$

These polynomials constitute an orthogonal basis of the Hilbert space $\mathcal{L}^2(\mathbb{R},\mu)$ with measure μ being the Lebesgue integral $\int e^{-\frac{x^2}{2}}dx$ and the inner product given by

$$\langle f,g\rangle=\int_{\mathbb{R}}fgd\mu=\int_{\mathbb{R}}fge^{-\frac{x^2}{2}}dx\,.$$

The Hermite process of order m with self-similarity parameter $H\in(\frac{1}{2},1)$, denoted by $\left\{Z_H^p(t)\right\}_{t\geq 0}$, is defined via s multiple stochastic integral

$$Z_H^p(t)=\int\underset{\mathbb{R}^m}{\cdots}\int\left(\int_0^t\prod_{k=1}^m(u-x_k)_+^{-\frac{1}{2}-\frac{1-H}{m}}du\right)dB(x_1)\dots dB(x_q)$$

over independent realizations of Gaussain white noise $dB(x_j)$. The Hermite process of order 1 is fractional Brownian motion, whereas higher order Hermite processes correspond to non-Gaussian H-sssi, [1].

3.4 Weierstrass function

Yet another example of non-Gaussian self-similar process is the random Weierstrass function defined as the limiting random function of the sum of randomly phase shifted and scaled cosines

$$f_H(t)=\sum_{k\in\mathbb{Z}}\lambda_0^{-kH}\left\{\cos(\phi_k)-\cos(2\pi\lambda_0^k t+\phi_k)\right\}\,,$$

for $(\phi_k)_{k\in\mathbb{Z}}\sim\mathcal{U}(0,2\pi)$ iid – the random phase shift of the k-th layer harmonics. The parameter λ_0 governs the base scale of the process and is knows as "the fundamental harmonic" (see [4]). The function $f_H(t)$ has continuous paths, yet is nowhere differentiable, even in the deterministic case when all phase shifts are identically zero. The Weierstrass function is not an H-sssi process, but it exhibits discrete self-similarity, which means that there exists $a_0>0$ such that self-similarity holds for values of a on a discrete grid $a_0\mathbb{N}$, that is for all $t\geq 0$

$$f_H(at)=a^H f(t)\,.$$

The H in the definition is in fact not the Hurst index, but a so called Hölder exponent, which means that for some $C\in[0,+\infty)$ it is true that for all $s,t\geq 0$

$$\left|f_H(s)-f_H(t)\right|\leq C|s-t|^H\,.$$

Functions, or random processes, for which there exists $H\geq 0$ such that this condition is satisfied are known as **Hölder continuous**. This Hölder continuity implies scale-invariance of the Weierstrass function. Indeed, for any $a\in a_0\mathbb{N}$ the condition implies that

$$a^{-H}\left|f_H(as)-f_H(at)\right|\leq C|s-t|^H\,,$$

for all $s,t\geq 0$. Furthermore, observe that by definition $f_H(0)=0$ and for all $t\geq 0$:

$$\begin{aligned}\left|f_H(at)-a^H f_H(t)\right|&\leq\left|f_H(at)-f_H(t)\right|+\left|a^H f_H(t)-f_H(t)\right|\\&\leq C|a-1|^H|t|^H+|a^H-1|\left|f_H(t)\right|\\&\leq C\big(|a-1|^H+|a^H-1|\big)|t|^H\,,\end{aligned}$$

which makes it obvious that the Weierstrass function has self-similarity.

4. the Crossing tree and H-sssi processes

Inspired by the simplicity of the statistical properties of the crossing tree for Brownian motions we formulate the following conjecture: whenever $\{X(t)\}$ is a continuous H-sssi process, then $X(0) = 0$ almost surely and the process is uniquely identified by the statistical properties of the crossing tree, namely the crossing durations, the excursion and the offspring distributions in the following way:
Conjecture

- The random variables Z_k^n are possibly correlated, but nevertheless identically distributed, with probability distribution
$$\mathbb{P}\big(Z_k^n = 2m\big) = 2^{1-H^{-1}}\big(1 - 2^{1-H^{-1}}\big)^{m-1},$$
for all $m \geq 1$;
- The distribution of excursion directions V_k^n (orientation patterns), conditional on the orientation of the parent crossing, is:
$$\mathbb{P}\big(+-|++\big) = \mathbb{P}\big(-+|--\big) = 2^{-2H^{-1}}.;$$
- The properly scaled crossing durations $4^{-n/2H}W_k^n$ are identically distributed with finite variance;

In order to gather empirical evidence either supporting or refuting this claim, extensive numerical study was in order. The Monte-Carlo simulation was performed on each of the processes mentioned in the previous chapter (p. 10). The software part of the experiment was implemented in Python in conjunction with Numpy and FFTW, a standalone open source library dedicated to efficiently computing Fast Fourier Transforms using .[1]

4.1 Generation of fBm, Hermite and Weierstrass processes

The basic building block of the fBm and the Hermite processes is a Gaussian noise with a particular correlation structure, as evidenced by the stochastic integral representations of these processes. All processes generated for this study were discrete versions of the corresponding continuous processes.

Generation of the fractional Brownian motion for this study is based on the Circulant Embedding method of Dietrich and Newsam for generating fractional Gaussian noise (fGn), [5]. In short, the method utilizes the structure of the correlation matrix of fGn to embed it into a larger circulant Toeplitz matrix, suitable for imposing the required covariance structure upon independent standard normal random variables via standard forward Discrete Fourier Transform. For details the reader is encouraged to refer to the original paper [5] and a more recent one [8].

As for the Hermite processes, synthesis of their sample path was based on the fundamental theorem of Lamperti (see [7]) which states that H-sssi processes are the only limiting law of normalized partial sum of a stationary random sequence. Formally, if $(X_i)_{i\geq 1}$ is stationary and for some regularly varying function $a_n \to \infty$ and the following limiting exists

$$\frac{1}{a_n}\sum_{i=1}^{[nt]} X_i \overset{\mathcal{D}}{\to} y(t),$$

where convergence is in finite-dimensional distributions, then the limiting process $(Y_t)_{t\geq 0}$ is H-sssi for some $H > 0$.

For instance, if $(X_i)_{i\geq 1}$ is an independent and identically distributed, sequence, then the limit of normed partial sums is the Brownian motion, which is $\frac{1}{2}$-sssi. In case if the stationary sequence X_i is **l**ong- **r**ange **d**ependent, i.e. with slowly decaying autocorrelation, the limit $Y(t)$ is often some H-sssi process with $H > \frac{1}{2}$ (see [7]). In particular if $X_i = \xi_i^2 - 1$ for some stationary Gaussian sequence with

[1]The source code of the developed toolkit for crossing tree construction and analysis is publicly available online at `https://github.com/ivannz/study_notes/tree/master/year_14_15/course_project/release`

$\xi_i \sim \mathcal{N}(0,1)$, but $\mathbb{E}(\xi_1 \xi_{n+1}) = n^{H-1}L(n)$ as $n \to \infty$ for some $H \in (\frac{1}{2}, 1)$ and some slowly varying function L, then the limiting process is a non-Gaussian H-ss process with strongly dependent, yet stationary increments corresponding to a Hermite process of order 2 (see [6]). The transformation used to compute $(X_i)_{i\geq 0}$ form $(\xi_i)_{i\geq 0}$ is the 2-nd order Hermite polynomial.

Sample paths of the Weierstrass process were simulated on a uniformly spaced grid $(t_k)_{i=0}^N \in [0,1]$ with $0 = t_0 < t_1 < \ldots < t_{N-1} < t_N = 1$. The spacing of the grid and the fundamental harmonic λ_0 were chosen so that the trigonometric functions are adequately sampled over the unit interval (see [4]).

4.2 Simulation results

In order to study the empirical support of the hypothesised statistical properties of the crossing tree as well as see how accurate in practice the crossing tree methodology confirms the theoretical result in [9], an extensive Monte-Carlo experiment was preformed.

The crossing trees were constructed for a base scale δ dependent on each particular sample realisation of the process, but in such a way as to enable meaningful comparisons of the crossing tree properties between different Monte-Carlo replications and between different classes of H-sssi processes. For a particular sample path $(x_j)_{j=0}^N$ of the process $\{X(t)\}$ the base scale was set to

$$\delta = \mathrm{med}\big(|\Delta x_j|\big)\,,$$

where $\Delta x_j = x_j - x_{j-1}$ and $\mathrm{med}(\cdot)$ is the median of the sample. The rationale behind the median of absolute increments of the sample path was to strike a balance between the biasedness of the crossing tree parameters, resulting from linear interpolation of the crossing times and the inaccuracy due to too coarse a resolution. Other choices for the base scale were considered, such as the standard deviation and the **i**nter**q**uartile **r**ange measure of statistical dispersion of the increments. Neither did produce any significantly different results from the median, except only that they tended to produce trees with fewer levels and thus cover a narrower range of resolutions of the process.

It seems natural to begin with the study of the fractional Brownian motion. To this end, 10^3 Monte-Carlo simulations of sample paths of the fBm process were simulated. The process was confined to the unit interval and discretized to have 2^{21} sample points.

Before proceeding to the evaluation of statistical properties of the crossing trees, it is necessary to determine the range of tree levels (grid scales, or resolutions) for which self-similarity is apparent.

Figure 4.1 suggests that scale-invariance properties for the studied fractional processes and the chosen method of computation of the base scale manifest their effects at across levels from 6 to 9. In fact the χ^2 test described in section 2.4 does not seem to support this finding and instead has the lowest empirical rejection rate exactly at levels from 7 to 8 (see table 4.1). Since the self-similarity of the simulated

Process	$6-7$	$7-8$	$8-9$	$6-8$	$7-9$	$6-9$
fBm-0.50	10.6	**8.2**	9.1	13.6	12.2	16.2
fBm-0.60	10.9	**9.1**	9.4	14.2	13.1	16.4
fBm-0.70	9.4	**7.0**	7.2	12.6	11.2	15.8
fBm-0.80	10.3	**6.1**	6.7	12.8	10.7	16.9
fBm-0.90	11.1	**7.8**	7.9	18.0	12.8	27.7

Table 4.1: The table of empirical rejection rate at significance level of $\alpha = 5\%$ of the χ^2 test for self-similarity between levels of the crossing tree.

fractional Brownian motion paths seems to become apparent at levels 7 and 8 on average, we pool the crossing tree data of these levels in order to obtain more reliable estimates.

Restricting the attention to the Brownian motion case (fBm with $H = \frac{1}{2}$) one can easily see that the numerical evidence is well aligned with the theoretical result of Jones and Rolls (fig. 4.2). Similarly, the theoretical probability of an up-down crossing conditional on an upcrossing is matched very closely by the numerical evidence (fig. 4.3). The results are similar in the down-up excursions case (fig. 4.4).

As for the conjectured distribution, the offspring distribution in the performed experiments seems to suggest that the levels of the tree, at least for the fractional Brownian motion, tend to be populated

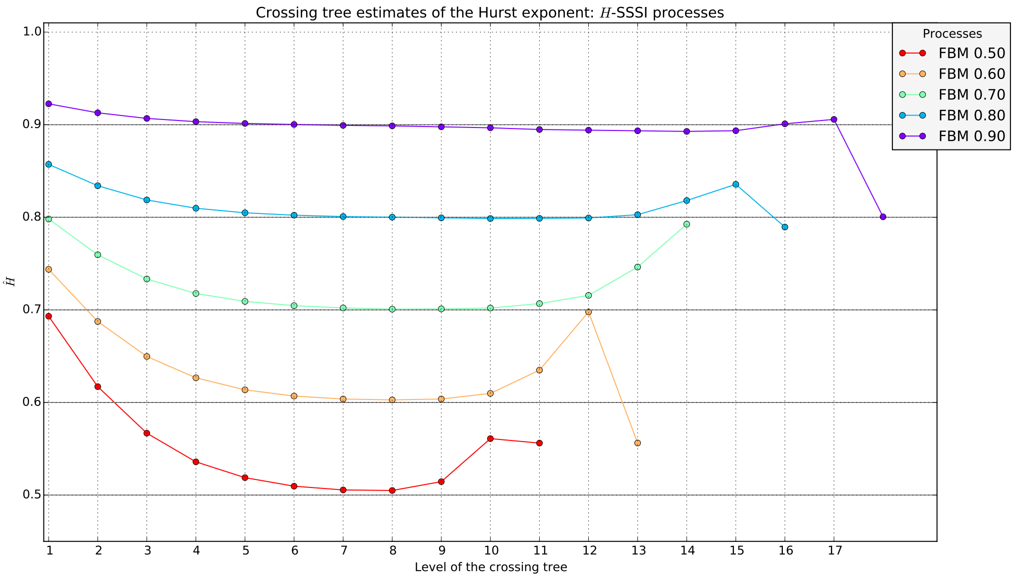

Figure 4.1: The plot of the estimates of the Hurst index H on the offspring data of a single level of the crossing tree based on 10^3 Monte-Carlo simulations of fBm.

by crossings with more than predicted number of subcrossings (see tables 4.3 for $H = 0.6$ and 4.4 for $H = 0.8$). However the failure of the conjectured crossing size distribution to match the empirical results closely, should not be considered as a serious evidence against it in this case. Indeed, the empirical offspring distribution for fBm with $H = 0.5, 0.6$ and 0.7 is aligned quite well with the conjecture, and since there is no theoretical reason as to why the dependence on H should break down for values of H closer to 1, we attribute this discrepancy to numerical issues of discretizing continuous processes and simulating fGn with long range dependence.

Finally, the mean crossing durations, averaged for each replication across all detected crossings of a single level indeed demonstrate plausibility of the conjectured scaling between levels of the crossing tree, see fig. 4.5 for fBm.

Now let's turn to the Hermite and Weierstrass processes. This time 10^4 random replications of sample paths of size 2^{17} were generated for each process. This size limitation was dictated by deteriorating numerical accuracy of the procedure, responsible for generating sample paths of Hermite processes. Thus in order to produce comparable results, sample paths of fBm and Weierstrass processes were limited to 2^{17} points as well.

Figure 4.6 suggests that the processes studied share common statistical properties of the crossing tree.

Table 4.2 summarizes the empirical rejection rates of the χ^2 test for self similarity across the range of levels indicated in the header of the table. There does not seem to be a common range of levels, at the corresponding resolutions of which all processes exhibit scale-invariance. This might be attributed to the fact that the path of the generated processes were insufficiently long to adequately populate the higher levels of the crossing tree. Nevertheless, it seems reasonable to expect a certain degree of self-similarity over levels from 7 to 8, which with due caution, could yield empirical evidence useful for analysing the conjecture.

Indeed, the offspring distribution plot (fig. 4.7) seems to suggest that even though the theoretical distribution seems to underestimate the real probability of a crossing of a particular size (in the number of subcrossings) as H increases to 1, there is still evidence for similarity of the statistical properties of the crossing tree across different self-similar processes.

The tables 4.3 and 4.4 compare the conjectured probabilities against the empirical ones. Though there is no strong evidence for the conjecture, one cannot say that it should be discarded. Indeed,

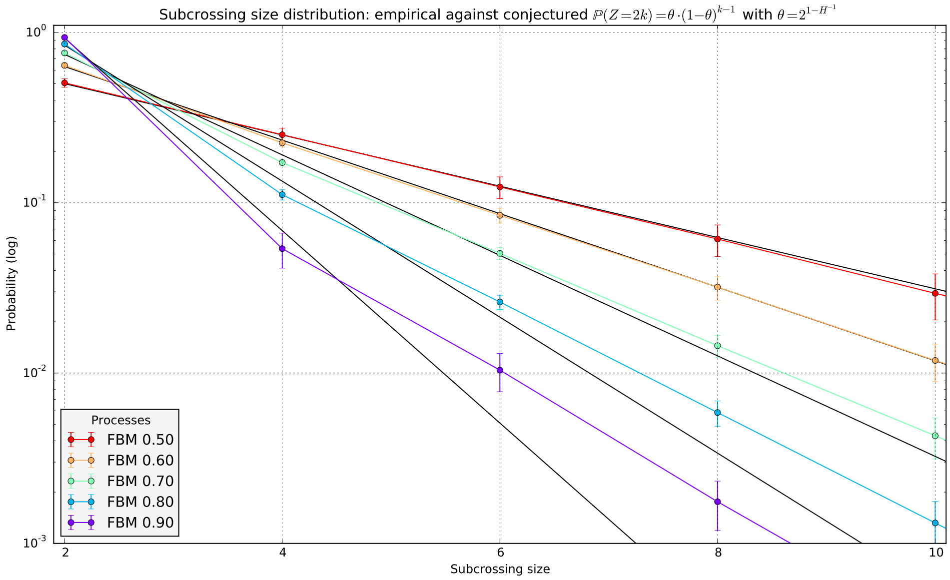

Figure 4.2: The log-plot of the offspring distributions estimated on 1000 sample discrete paths of the fBm process of length 2^{21} and Hurst exponents in the range from 0.5 to 0.9.

looking at the box-plots (4.9 and 4.8) of the conditional distribution of excursions, it is possible to see that even though there is significant margin of error there the probabilities seem to agree quite well with the hypothesised distribution.

Crossing duration also do not seem contradict the hypothesized scaling for H-sssi processes and the Weierstrass function (see fig. 4.10 - 4.13). Also we considered the properly scaled empirical quantiles (not presented here) of the crossing durations between the processes considered and across all sufficiently populated levels (from 1-st to 10-th) of the sample crossing tree. The conclusion is that there is indeed similarity (up to a multiplicative constant depending on the base scale δ) of the crossing durations' distribution among processes (see fig. 4.14, 4.15 and 4.16).

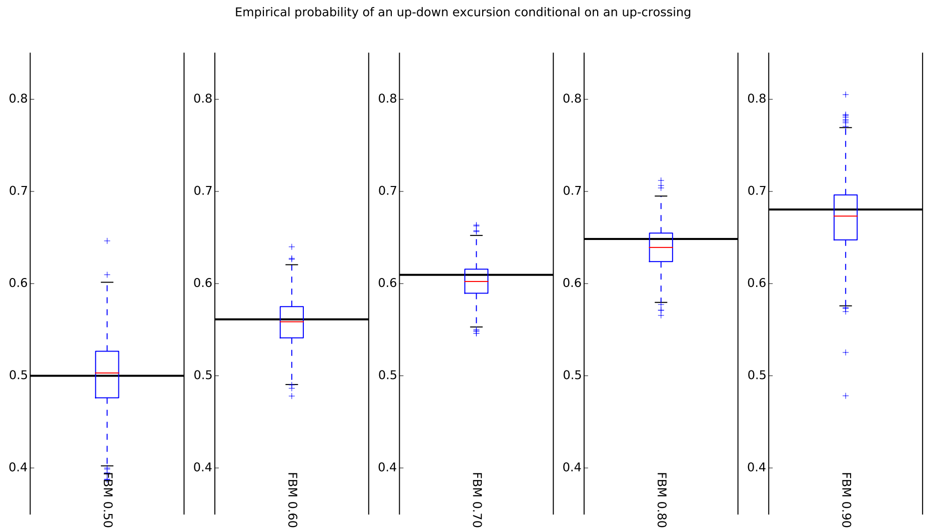

Figure 4.3: The estimated conditional probability of an up-down excursion given upward orientation of the parent crossing.

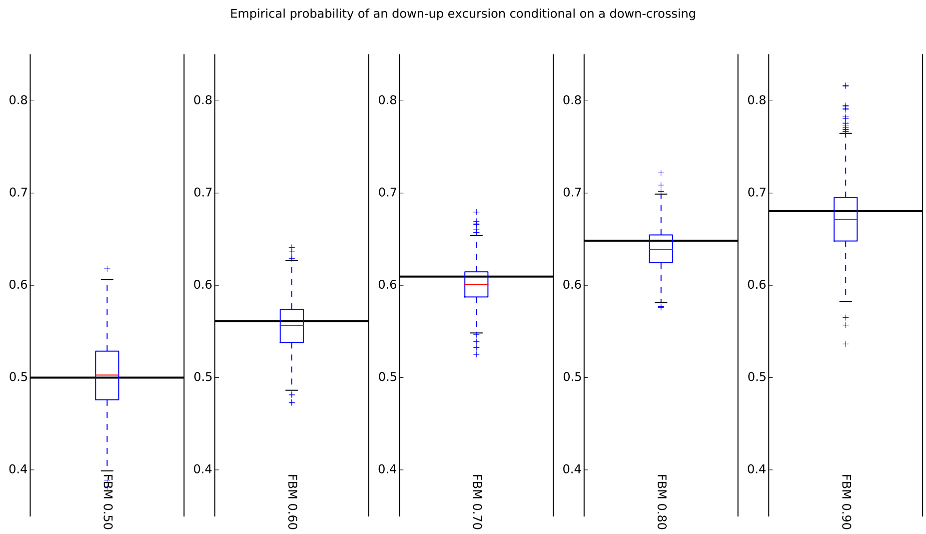

Figure 4.4: The box-plot of the estimates of the probability of an down-up excursion conditional on the orientation of the parent crossing being downward.

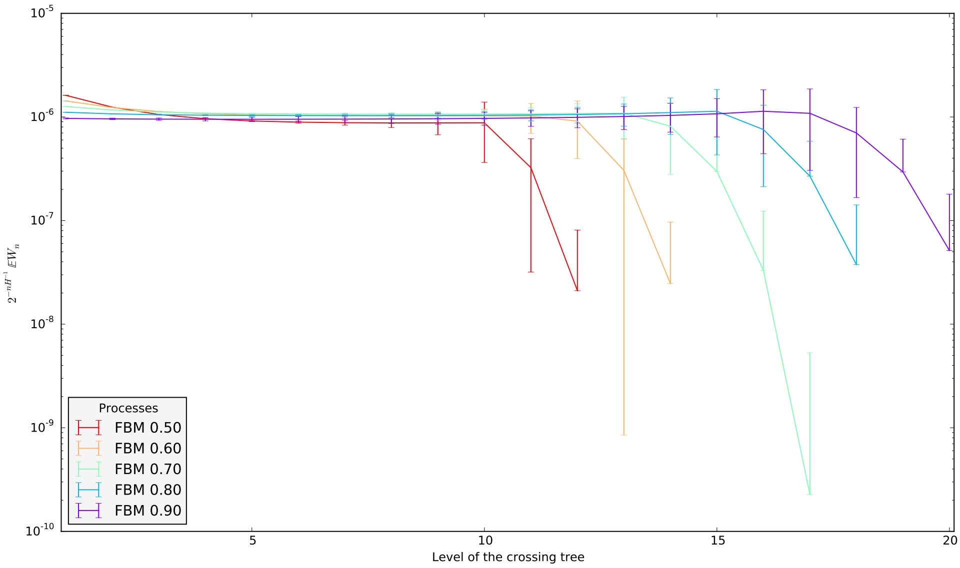

Figure 4.5: The average crossing duration at each level of the crossing tree built for fractional Brownian motion processes.

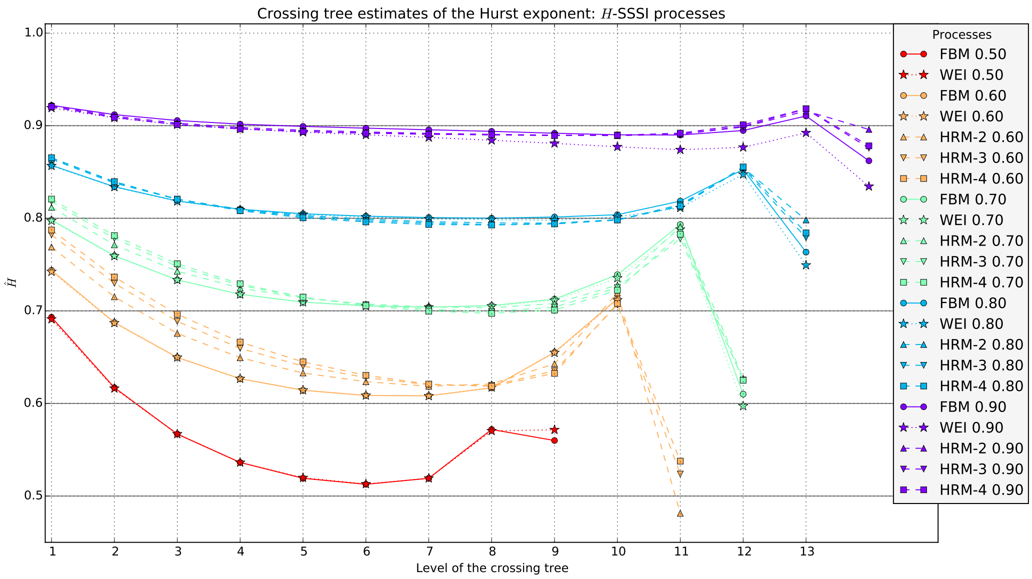

Figure 4.6: Estimates of the Hurst exponent H based on a single level of the crossing tree for the all studied H-sssi processes.

Process	6 − 7	7 − 8	8 − 9	6 − 8	7 − 9	6 − 9
fBm-0.50	9.7	15.7	–	13.8	**3.6**	5.4
fBm-0.60	8.8	9.3	**5.6**	14.2	14.3	16.5
fBm-0.70	7.6	**6.6**	10.5	13.4	15.7	17.8
fBm-0.80	6.5	4.8	**4.5**	11.5	13.1	16.5
fBm-0.90	**4.5**	5.1	6.9	11.7	13.6	19.8
WEI-0.50	9.6	12.4	–	13.9	**2.2**	5.1
WEI-0.60	**8.2**	9.6	11.1	13.2	14.8	15.5
WEI-0.70	7.0	**5.9**	**5.9**	12.9	14.6	15.8
WEI-0.80	6.2	5.5	**4.9**	13.1	13.2	17.0
WEI-0.90	5.0	**1.9**	8.7	12.2	9.7	21.0
HRM-2-0.60	9.6	9.2	**7.1**	15.2	15.9	19.6
HRM-2-0.70	8.5	**7.4**	11.3	13.9	15.1	18.2
HRM-2-0.80	6.5	**5.4**	5.6	12.5	12.3	17.8
HRM-2-0.90	6.4	**4.9**	0.0	12.7	14.5	18.7
HRM-3-0.60	9.6	**9.2**	13.7	15.6	17.9	20.3
HRM-3-0.70	**7.9**	8.1	**7.9**	14.5	16.4	19.2
HRM-3-0.80	5.5	**4.3**	5.4	11.8	13.1	16.7
HRM-3-0.90	**5.6**	7.7	12.5	13.4	14.1	19.8
HRM-4-0.60	9.7	**9.3**	12.3	16.5	17.4	21.4
HRM-4-0.70	8.1	7.2	**6.0**	15.4	16.8	20.4
HRM-4-0.80	5.7	7.4	**1.8**	11.7	13.7	16.8
HRM-4-0.90	5.1	**2.9**	8.3	9.8	12.9	19.0

Table 4.2: The table of empirical rejection rate at significance level of $\alpha = 5\%$ of the χ^2 test for self-similarity between levels of the crossing tree.

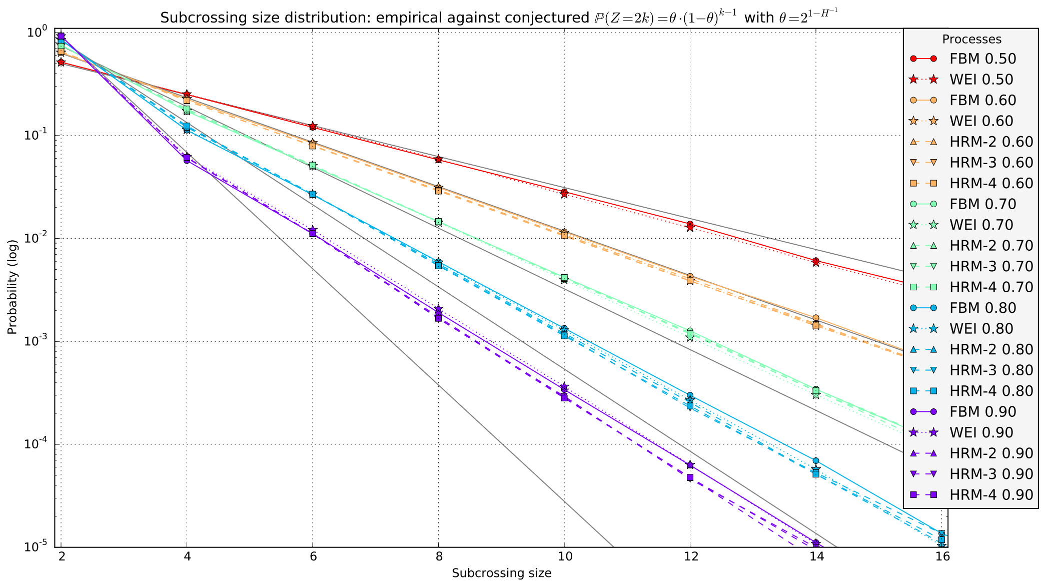

Figure 4.7: Empirical probabilities of the crossing sizes against the hypothesized theoretical probabilities for 10^4 sample paths of length 2^{17} points.

$H = 0.6$	$Z_k = 2$	$Z_k = 4$	$Z_k = 6$	$Z_k = 8$
fBm	0.630	0.233	0.086	0.032
	0.644 ± 0.033	0.224 ± 0.027	0.083 ± 0.017	0.031 ± 0.011
WEI	0.630	0.233	0.086	0.032
	0.642 ± 0.027	0.226 ± 0.026	0.084 ± 0.017	0.031 ± 0.010
HRM-2	0.630	0.233	0.086	0.032
	0.663 ± 0.034	0.215 ± 0.026	0.078 ± 0.015	0.028 ± 0.009
HRM-3	0.630	0.233	0.086	0.032
	0.666 ± 0.038	0.214 ± 0.026	0.076 ± 0.015	0.027 ± 0.009
HRM-4	0.630	0.233	0.086	0.032
	0.667 ± 0.039	0.215 ± 0.025	0.076 ± 0.016	0.027 ± 0.009

Table 4.3: The table of empirical probabilities of the first four values of the number of subcrossings in a parent crossing for H-sssi processes with $H = 0.6$. Levels from 6 to 8 were pooled to get the estimates.

$H = 0.8$	$Z_k = 2$	$Z_k = 4$	$Z_k = 6$	$Z_k = 8$
fBm	0.841	0.134	0.021	0.003
	0.855 ± 0.023	0.112 ± 0.017	0.026 ± 0.006	0.006 ± 0.002
HRM-2	0.841	0.134	0.021	0.003
	0.848 ± 0.027	0.118 ± 0.022	0.026 ± 0.006	0.005 ± 0.002
HRM-3	0.841	0.134	0.021	0.003
	0.846 ± 0.024	0.121 ± 0.019	0.026 ± 0.006	0.005 ± 0.002
HRM-4	0.841	0.134	0.021	0.003
	0.844 ± 0.021	0.122 ± 0.016	0.026 ± 0.006	0.005 ± 0.002
WEI	0.841	0.134	0.021	0.003
	0.854 ± 0.019	0.113 ± 0.015	0.026 ± 0.005	0.006 ± 0.002

Table 4.4: The table of empirical probabilities of the first four values of the number of subcrossings in a parent crossing for H-sssi processes with $H = 0.8$. Levels from 6 to 8 were pooled to get the estimates.

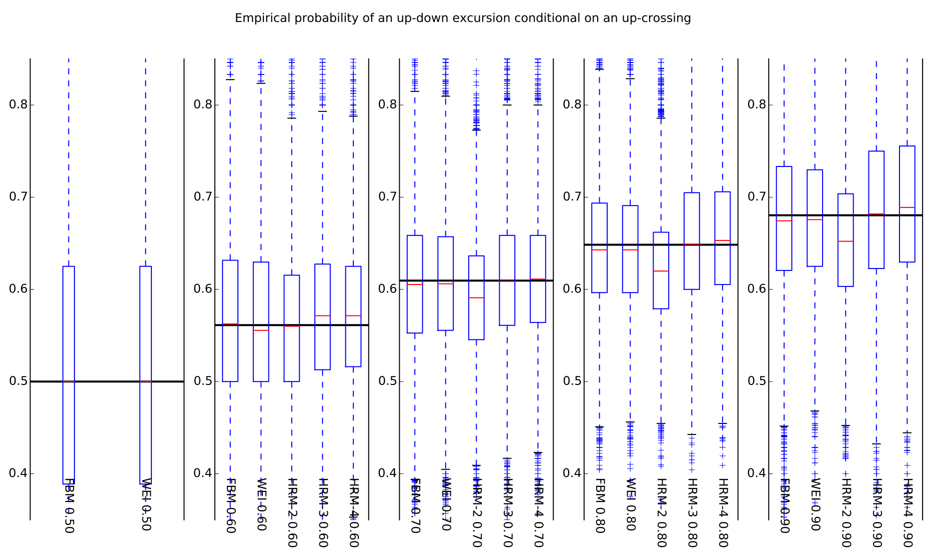

Figure 4.8: The empirical estimates of the probability of an up-down excursion conditional on the upward orientation of the parent crossing.

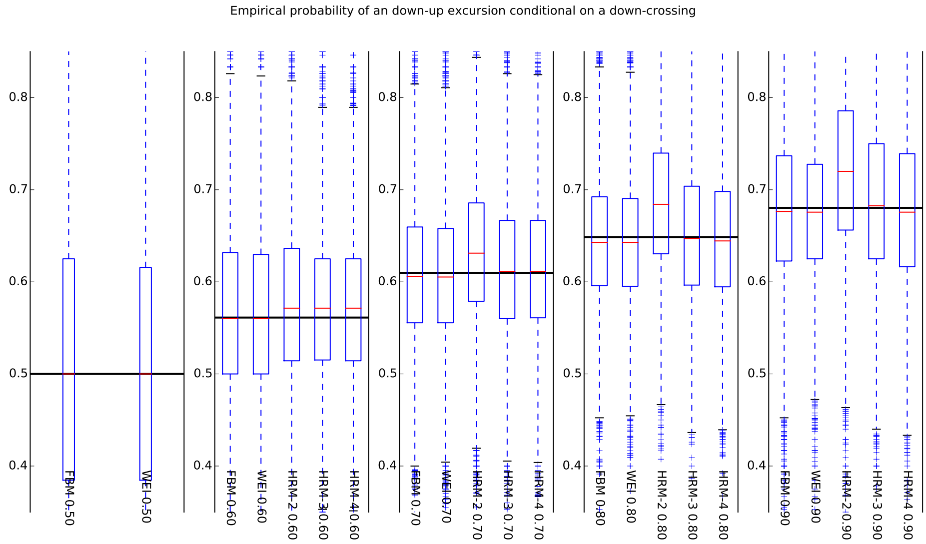

Figure 4.9: The same figure as 4.8 but for down-up excursions conditional on the orientation of the parent crossing being downward.

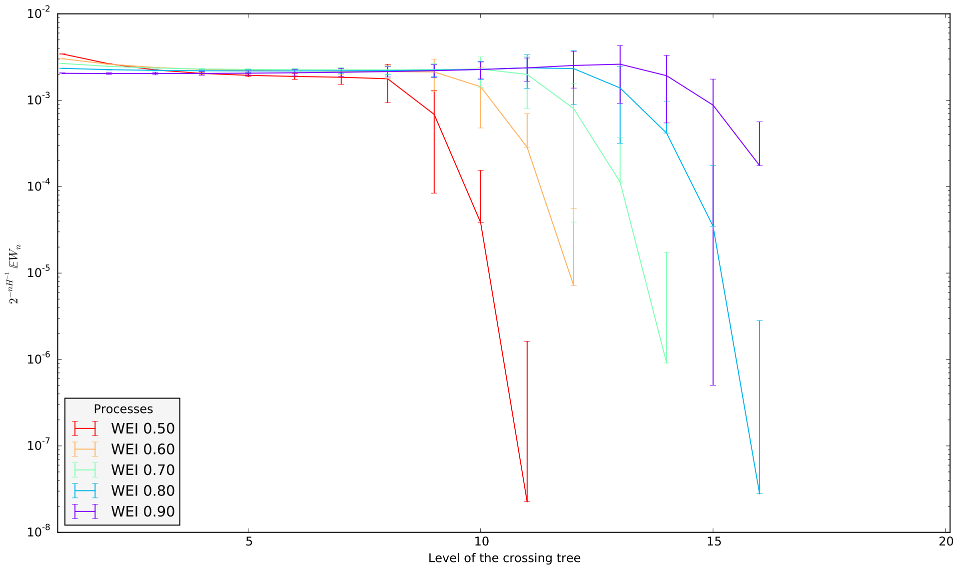

Figure 4.10: The average crossing duration at each level of the crossing tree built for the Weierstrass processes.

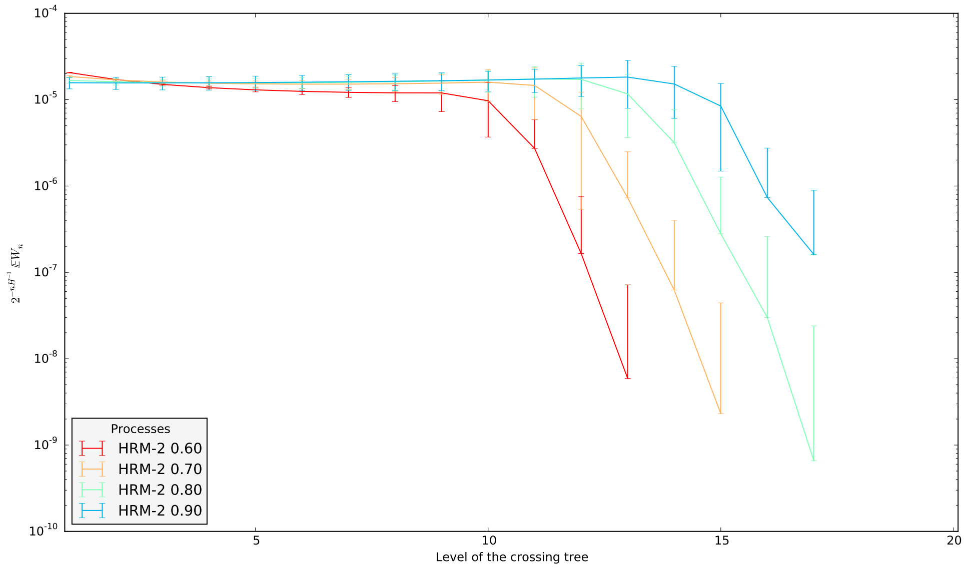

Figure 4.11: The average crossing duration at each level of the crossing tree built for the Hermite processes of order 2.

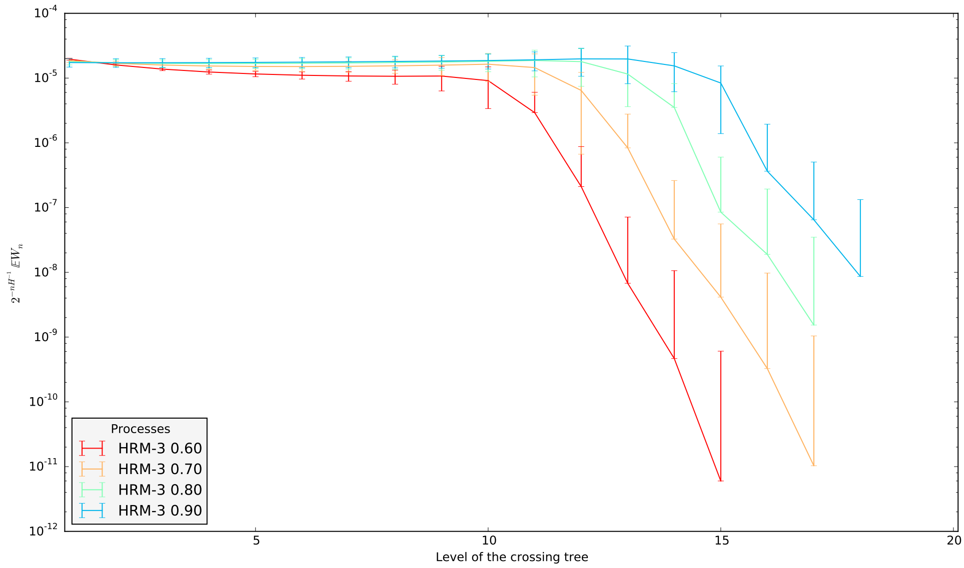

Figure 4.12: The average crossing duration at each level of the crossing tree built for the Hermite processes of order 3.

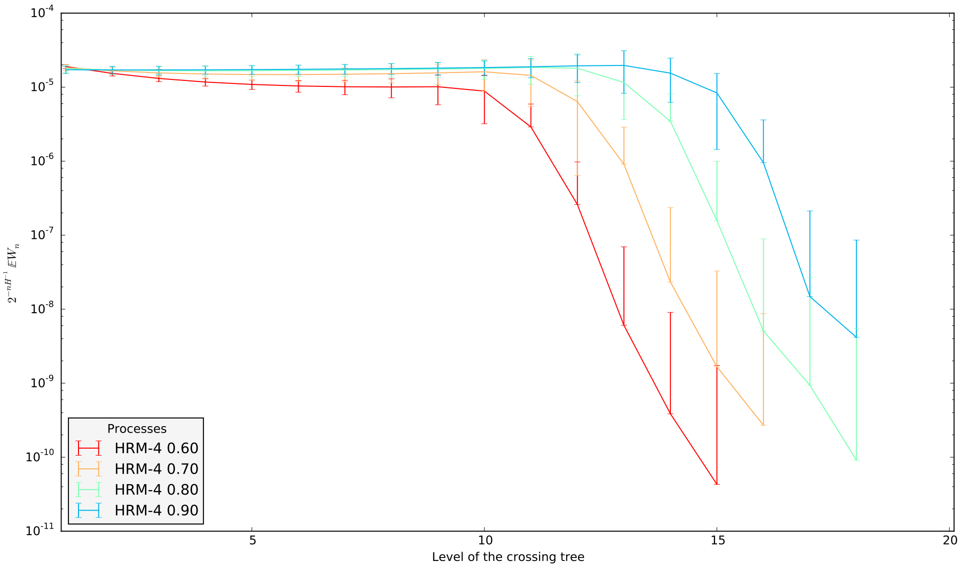

Figure 4.13: The average crossing duration at each level of the crossing tree built for the Hermite processes of order 4.

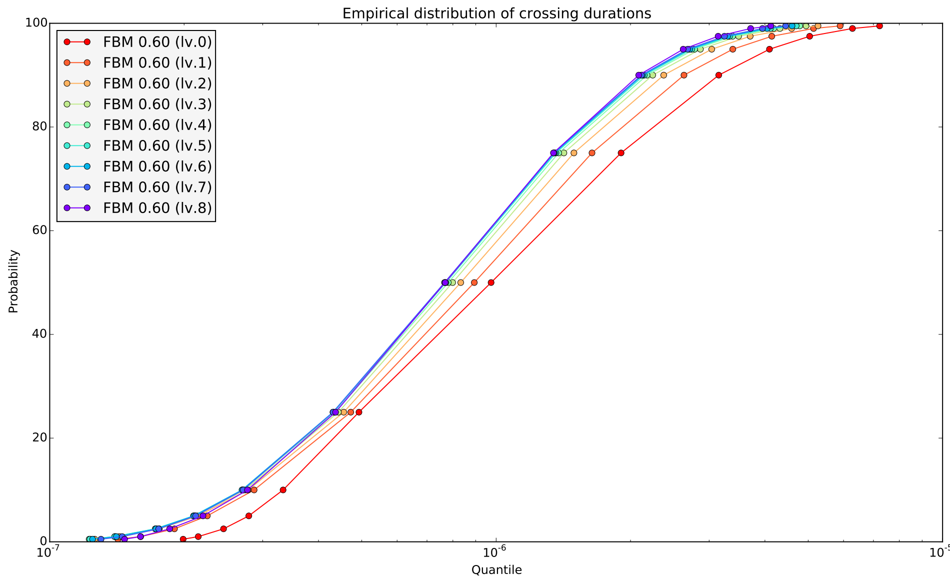

Figure 4.14: The empirical distribution of crossing durations at each level of the crossing tree built for the fractional Brownian motion (2^{21} datapoints) with $H = 0.6$. Averaged across all Monte-Carlo realisations.

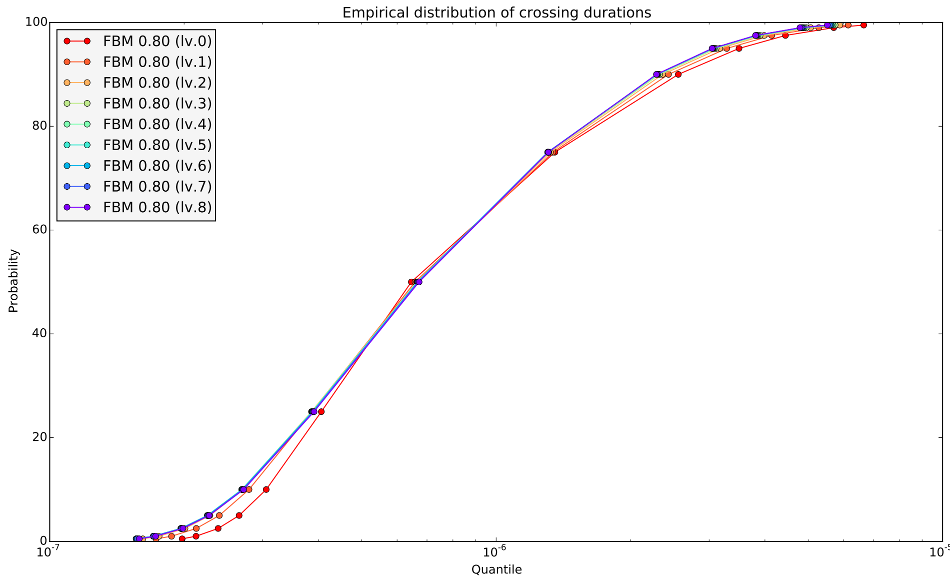

Figure 4.15: Similarly to figure 4.14 but for fBm with $H = 0.8$.

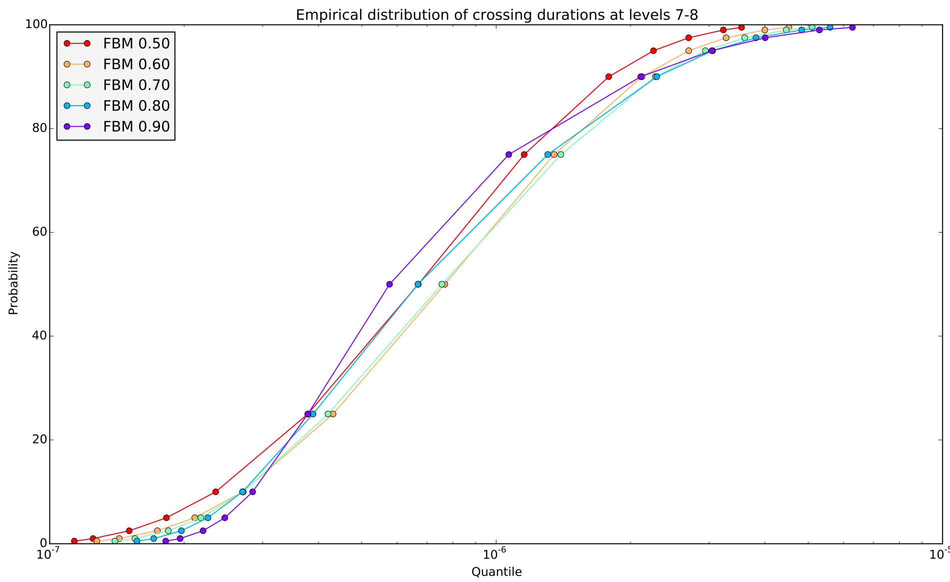

Figure 4.16: Empirical distributions of crossing durations estimated on levels 7-8 for the fBm (2^{21} datapoints). Averaged across all Monte-Carlo realisations.

5. Conclusion and further work

This preformed extensive Monte-Carlo simulation suggests, that the considered classes of H-SSSI stochastic processes indeed shares common statistical properties of the crossing tree. To findings are summarized below:

1. The crossing tree seems capable of correctly identifying the Hurst exponent of the studied series, provided correct tree levels are chosen to produce the estimate. The tree levels should not be too close to the leaves of the tree, for the sake of avoiding excessive bias due to linear interpolation, and be such that there is no statistical evidence to significant variations in the offspring distribution at each level;

2. Processes seem to share similar shape of the empirical distribution of the crossing sizes (the number of offspring), though the hypothesized relationship between the offspring distribution and the Hurst exponent does not seem to hold exactly, (see figures 4.2 and 4.7);

3. The conditional distributions of excursion within each crossing seems to agree with the hypothesised values of $2^{-1/2H}$.

4. The collected empirical evidence seems to confirm similarity of the scaling properties of crossing durations across all studied processe.

To summarize, the overall evidence is in favour of the conjecture, it is absolutely necessary to address the issue of the upward bias in the empirical probabilities of crossings of size higher than 2.

One of the way in which this study can be improved is the generation of samples paths of Hermite stochastic processes. Another line of inquiry in the topic of crossing trees if their application to machine learning and generalization to higher dimensions, which would find application in image processing. The third possible extension is the utilization of the crossing tree, or its modification, in the problem of early detection of structural breaks in univariate or multivariate time series data.

Bibliography

[1] Bai Shuyang, Taqqu Murad S. Generalized Hermite processes, discrete chaos and limit theorems // Stochastic Processes and their Applications. — 2014. — Vol. 124, no. 4. — P. 1710 – 1739. — URL: http://www.sciencedirect.com/science/article/pii/S0304414913003086.

[2] Barlow Martin T., Perkins Edwin A. Brownian motion on the Sierpinski gasket // Probability Theory and Related Fields. — 1988. — Vol. 79, no. 4. — P. 543–623. — URL: http://dx.doi.org/10.1007/BF00318785.

[3] Self-similarity parameter estimation and reproduction property for non-Gaussian Hermite processes : Rep. : arXiv:0807.1208 ; Executor: Alexandra Chronopoulou, Ciprian Tudor, Frederi Viens : 2008. — Jul. — URL: http://cds.cern.ch/record/1114288.

[4] Decrouez Geoffrey, Amblard Pierre-Olivier, Jones OD. Estimation of the multifractal spectrum using the crossing tree // XXIVème colloque GRETSI (GRETSI 2013). — 2013. — P. cdrom.

[5] Dietrich C. R., Newsam G. N. A fast and exact method for multidimensional gaussian stochastic simulations // Water Resources Research. — 1993. — Vol. 29, no. 8. — P. 2861–2869. — URL: http://dx.doi.org/10.1029/93WR01070.

[6] Embrechts Paul, Maejima Makoto. An introduction to the theory of self-similar stochastic processes // International Journal of Modern Physics B. — 2000. — Vol. 14, no. 12n13. — P. 1399–1420. — http://www.worldscientific.com/doi/pdf/10.1142/S0217979200001047.

[7] Embrechts Paul, Maejima Makoto. Selfsimilar Processes. — Princeton : Princeton University Press, 2002. — ISBN: 9780691096278. — URL: http://press.princeton.edu/titles/7319.html.

[8] Fast and exact synthesis for 1-D fractional Brownian motion and fractional Gaussian noises / E. Perrin, R. Harba, R. Jennane, I. Iribarren // Signal Processing Letters, IEEE. — 2002. — Nov. — Vol. 9, no. 11. — P. 382–384.

[9] Jones Owen, Rolls David. A characterisation of, and hypothesis test for, continuous local martingales // Electron. Commun. Probab. — 2011. — Vol. 16. — P. no. 56, 638–651. — URL: http://ecp.ejpecp.org/article/view/1673.

[10] Jones Owen Dafydd, Shen Yuan. Estimating the Hurst index of a self-similar process via the crossing tree // Signal Processing Letters, IEEE. — 2004. — April. — Vol. 11, no. 4. — P. 416–419.

[11] Jones Owen Dafydd, Shen Yuan. A non-parametric test for self-similarity and stationarity in network traffic // Fractals in Engineering / Ed. by Jacques Lévy-Véhel, Evelyne Lutton. — Springer London, 2005. — P. 219–234. — URL: http://dx.doi.org/10.1007/1-84628-048-6_14.

[12] Mandelbrot Benoit B., Ness John W. Van. Fractional Brownian Motions, Fractional Noises and Applications // SIAM Review. — 1968. — Vol. 10, no. 4. — P. 422–437. — http://dx.doi.org/10.1137/1010093.

[13] Булинский А.В., Ширяев А.Н. Теория случайных процессов. — ФИЗМАТЛИТ Москва, 2005. — С. 400. — ISBN: 5-9221-0335-0.

Appendix

Details on the practical procedure

We briefly describe of the procedure for constructing the crossing tree on time series data $(t_i, x_i)_{i=0}^N$.

Recall that the tree is constructed in two passes: initial detection passages over levels of $\delta\mathbb{Z}$ grid, which also discards within-band movements, and the pruning phase, where re-crossing events are eliminated.

Before the first pass through the data, the time series (x_i) is shifted and scaled to series $z_i = \frac{1}{\delta}\bigl(x_i - x_0\bigr)$, so that the constructed tree is rooted at the crossing time $t = 0$, since the process sets off from the origin.

The first pass sweeps through consecutive increments of the series given by pairs (t_i, z_i) and (t_{i+1}, z_{i+1}) for $i = 1, \ldots, N-1$. For each increment, its direction is determined by the sign of the difference $\Delta_i = z_{i+1} - z_i$, and, depending on it, the range of grid levels traversed by it is computed. The range of the i-th increment a subset of integers $R_i = [a_i, b_i] \cup \mathbb{Z}$, where the boundaries a_i and b_i are determined using the following logic:

Upcrossing if $\Delta_i > 0$ then $a_i = \lceil z_i \rceil$ and $b_i = \lfloor z_{i+1} \rfloor$;

Downcrossing for $\Delta_i < 0$ the range given by $b_i = \lceil z_{i+1} \rceil$ and $a_i = \lfloor z_i \rfloor$.

In a rare case of sideways movement $\Delta_i = 0$, the range is set to be $\emptyset$. Also note that $[a, b] = \emptyset$ whenever $b < a$.

Movements of the normalised process (z_i) that wiggle strictly within a band between levels of the grid $\delta\mathbb{Z}$ and do not do not pass through or touch a level of the grid have empty range, and are discarded.

This sequential elimination of between grid lines movements crucially depends on the presumption that the process is continuous and linear interpolation between the endpoints of each increment is valid.

The second pass checks if $b_{j-} = a_j$ for any crossing $j \in J$, where $J = \{j : R_j \neq \emptyset\}$, and the index $j-$ for any $j \in J$ is defined as $j- = \max\{i \in J : i < j\}$, or $-\infty$ if $j \in J$ is the very first apparent crossing of the grid $\delta\mathbb{Z}$, and a_{j-} is taken to be $-\infty$.

If the first level passed by $j \in J$ coincides with the very last level crossed by $j-$ then the very first crossing event in the j-th increment is a re-crossing event, and thus should be eliminated. This is done by adjusting the a_i in the direction of the j-th increment. The ranges are updated accordingly and empty ones are discarded.

These passes compute the levels the sample path of the process appears to have crossed, and the crossing times are estimated by linearly interpolating between times t_i and t_{i+1} (endpoints of the i-th increment) with weights determined by passed levels of the grid. The formula of the crossing time of some level $l \in R_i$ during the i-th increment is

$$\tau_{il} = t_i + \bigl(t_{i+1} - t_i\bigr)\frac{l - z_i}{z_{i+1} - z_i}\,.$$

The crossing tree is constructed from the finest resolution up by gradually making the resolution coarser. The crossing times at some grid are computed based on the times and levels provided by the least finest resolution, which is still finer than the this one, in the same way the second pass functions: since the resolution is halved between successive tree levels, the crossing times estimated at resolution $\delta 2^{n+1}$ are by construction a subset of crossing times of a finer grid $\delta 2^n$. Indeed, if an increment j passed a line $m\delta 2^{n+1}$ then this very same passed a line $2m\delta 2^n$ of the finer grid and the interpolation coefficient used to compute the crossing time is unchanged:

$$\left(m - \frac{x_j - x_0}{\delta 2^{n+1}}\right)\frac{x_{j+1} - x_j}{\delta 2^{n+1}} = \left(2m - \frac{x_j - x_0}{\delta 2^n}\right)\frac{x_{j+1} - x_j}{\delta 2^n}\,.$$

Thus, since the same increment crossed the grid line, its time remains unchanged.