Estimation methods of the Hurst parameter

George Mushore

February 1, 2018

Abstract:This is abstract

Contents

1 Introduction

2 General background on time series

A time series is a stochastic process where a set of variables are indexed by time. The time index could either refer to continuous time intervals or discrete time points. Examples of time series data , include hourly temperature measurements, daily stock prices, or the annual height of a river which we will look at later. In this thesis, I will consider time series in discrete time measured at regular time points noted as X_t where $t = 1, \ldots, n$ denotes equally spaced time points..

2.1 Concepts related to time series

2.1.1 Mean and variance

The mean μ_t of a time series is the expected value $E[X_t]$ if the mean does not change over time.This parameter is often estimated by the sum of values divided by the number of data values. The covariance between two random variables X_i and X_j for time points i and j is

$$\gamma(t_i, t_j) = E[(X_i - \mu_i)(X_j - \mu_j)] \tag{1}$$

When $t_i = t_j$ a we get

$$\gamma(t_j, t_i) = \sigma^2 \tag{2}$$

In time series analysis, an important question to ask is whether the mean and variance are the same throughout the series. When a time series is stationary, it means the the time series has constant mean and variance. Strictly stationary also says the joint distribution does not change over time. [Myrvoll-Nilsen, 2016]

$$F_{X_{t_1},\ldots,X_{t_n}}(x_1, \ldots, x_n) = F_{X_{t_{1+\tau}},\ldots,X_{t_{n+\tau}}}(x_{1+\tau}, \ldots, x_{n+\tau}) \tag{3}$$

τ is the difference between t_i and t_j

$$\tau = |t_i - t_j| \tag{4}$$

Wide-sense stationary processes do not require the latter. This is important when try to analyse data. Non-stationary data can sometimes be made stationary by mathematical transformations.We will only consider wide-sense stationary time series in this thesis.

2.1.2 Autocorrelation function and partial autocorrelation function

The autocorrelation function (acf) is the correlation of different times of a random process. Since we assume stationarity this means that the mean and the variance are time independent. The auto-covariance is now then

$$\gamma(t, t+\tau) = \gamma(\tau) \tag{5}$$

The acf is defined as the normalized auto-covariance

$$R(\tau) = \frac{\gamma(\tau)}{\sigma^2} = \frac{\gamma(\tau)}{\gamma(0)} \tag{6}$$

which gives the values between -1 and 1. Some important properties of the acf is that it is symmetric in a discrete time series and the continous time series $R(\tau) = R(-\tau)$ is an even function.
The acf has it's peak at $\tau = 0$. Another way to characterize the dependency structure of a time series is to consider the partial autocorrelation function(pacf) The partial autocorrelation function (pacf) is the conditional correlation function.

$$\rho(j) = \frac{\gamma(x_t, x_j | x_{t-1}, \dots x_{t-j-1})}{\sqrt{\gamma(x_t|x_{t-1}, \dots x_{t-j-1})\gamma(x_j|x_{t-1}, \dots x_{t-j-1})}} \tag{7}$$

It removes any indirect correlation to the time we are looking at. In other words it can be defined as the direct correlation between two random variables.

2.2 Some common time series models

2.2.1 White noise

White noise is a very time series model with a zero mean and has no correlation between variables at different points. Also the variance is time invariant. The most commonly used type of white noite noise is the gaussian noise defined as white noise $\{\epsilon_t\}$ with Gaussian or normal distribution $\epsilon_t \sim N(0, \sigma_t^2)$

2.2.2 Autoregressive models

An autoregressive process (AR) is a time series where future values depend on it's past values and a random white noise. The values have a linear correlation. The simplest AR model is the first order autoregressive process AR(1)

$$X_t = c + \phi X_{t-1} + \epsilon_t \tag{8}$$

The ϕ is referred to as the first-lag autocorrelation function. When $|\phi| < 1$, the process is wide-sense stationary. The mean is

$$E[X_t] = \frac{c}{1 - \phi_1}$$

And the variance is

$$Var(X_t) = \frac{\sigma^2}{1 - \phi_1^2}$$

And the autocorrelation function

$$\gamma(\tau) = \phi_1^\tau$$

Where τ is the lags

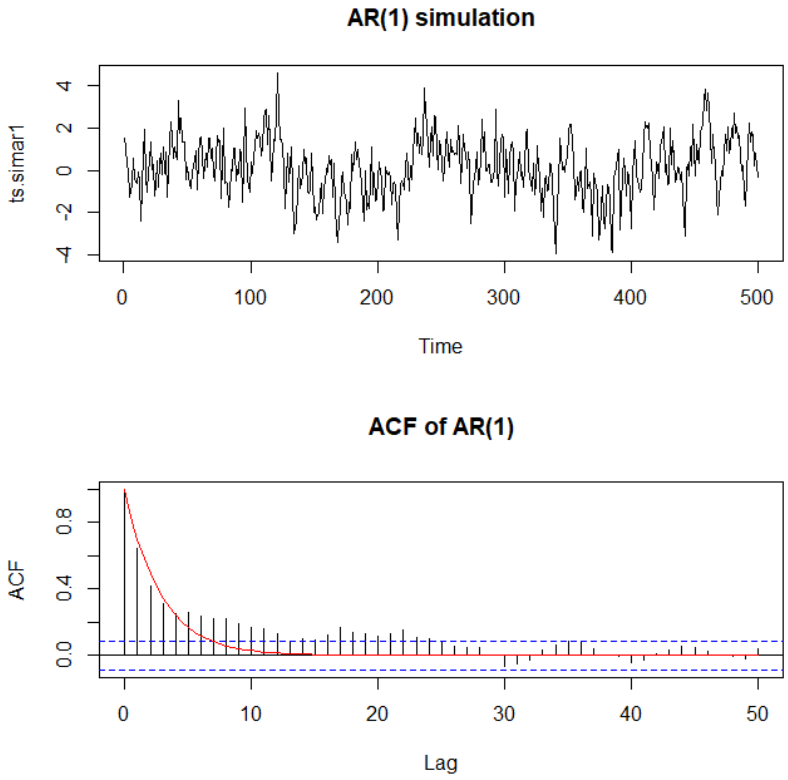

Figure 1: A Ar(1) model and its autocorrelation function

The general case is AR(p)

$$X_t = c + \epsilon_t + \sum_{i=1}^{p} \phi_i X_{t-i} \tag{9}$$

2.2.3 Moving average models

The moving average (MA) is a process determined by the current white noise process and previous white noise processes. The simplest MA(1) depends on one previous term. and can be defined

$$X_t = c + \epsilon_t + \theta_1 \epsilon_{t-1} \tag{10}$$

θ is a coefficient. c is a constant. The mean is

$$E[X_t] = c$$

And the variance is

$$Var(X_t) = \sigma^2(1 + \theta_1^2)$$

And the acf

$$\gamma(1) = \frac{\theta_1}{1 + \theta_1^2},$$

$$\gamma(\tau) = 0 \text{ for } \tau \geq 2$$

Where τ is the lags
In general MA(q) where q is the order of the moving average.

$$X_t = \mu + \epsilon_t + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} \tag{11}$$

2.2.4 Autoregressive moving average

The arma is described as

$$X_t = c + \epsilon_t + \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} \tag{12}$$

The first sum represents the autoregressive processes and the second process represents the move average processes. These are the most common time series. ϵ_t is the white noise error term. The best way of describing these processes is looking at the acf. For MA processes it can be seen that the it is a sum of white noise processes. And the acf of white noise processes as mentioned earlier has a peak at $\tau = 0$ thus q determines the number of peaks the ma process has.
For autoregressive(ar) processes the p is the order of the ar process. p gives at the relation of data p lags apart.

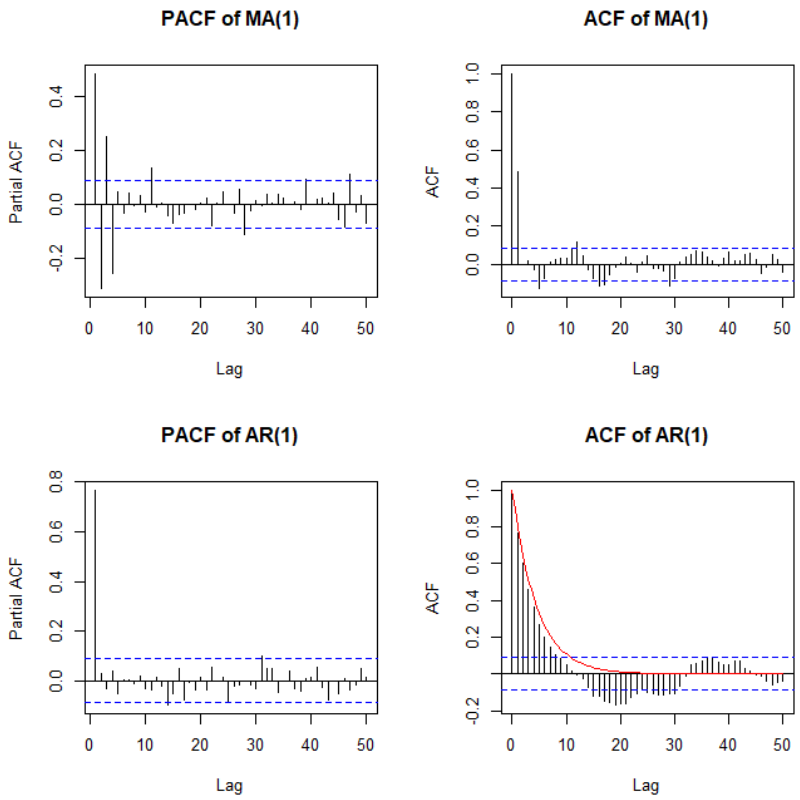

Figure 2: Simulation of Ar(1) and Ma(1) process and their autocorrelation function

2.3 Fractional Gaussian noise

2.3.1 Fractional Brownian motion

Brownian motion is a stochastic process that is an incremental Gaussian process. A Brownian process B_t is normal distributed, with a zero mean.

$$B_t \sim N(0, \sigma^2 t) \quad , t > 0 \text{ ,and continous} \tag{13}$$

The process B_t has independent increments where $W_{t+\tau} - W_t \sim N(0, \tau)$. Fractional Brownian motion is defined as having its increments to have dependence described by the Hurst exponent H. The distribution of a fractional Brownian motion process is

$$B_{H_t} \sim N(0, \sigma^2 |t| 2^H) \tag{14}$$

where H is a self-similarity parameter. A fractional Brownian motion process B_{H_t} is self-similar if for all constants c we have

$$B_{H_{ct}} \doteq c^{2H} B_{H_t} \tag{15}$$

where $\doteq$ means that the distribution is the same and scale invariant. And the acf is

$$\gamma(t) = \frac{\sigma^2]}{2}(|t+1|^{2H} - 2|t|^{2H} + |t-1|^{2H}) \tag{16}$$

When $H < 0.5$ the fractional Brownian motion is negatively correlated. and positive when $H > 0.5$. When $H = 0.5$ the process in a pure Brownian motion(13). The increments of a B_{H_t} are stationary.

$$B_{H_t} - B_{H_s} \sim B_{H_{(t-s)}} \tag{17}$$

2.3.2 Definition of fractional Gaussian noise

Fractional Gaussian noise (FGN) derives from the increment of fractional Brownian motion. These increments as seen in (17) are stationary. We define FGN as a discrete time process

$$X_t = B_{H_{t+1}} - B_{H_t} \tag{18}$$

The FGN process is defined for $0 > H > 1$

$$\gamma(\tau) = \frac{\sigma^2}{2}(|\tau+1|^{2H} - 2|\tau|^{2H} + |\tau-1|^{2H}) \quad , 0 < H < 1 \text{ and } \tau \geq 1 \tag{19}$$

When H is between 0.5 and 1 we can see that for all tau, the acf is not summable. Unlike for ARMA(p,q) processes the acf converges to a finite number. The FGN has a correlation matrix

$$\Sigma_N(H) = [\gamma_{|i-j|}] \tag{20}$$

where N is the length on the time series and $\gamma_0 = 1$.

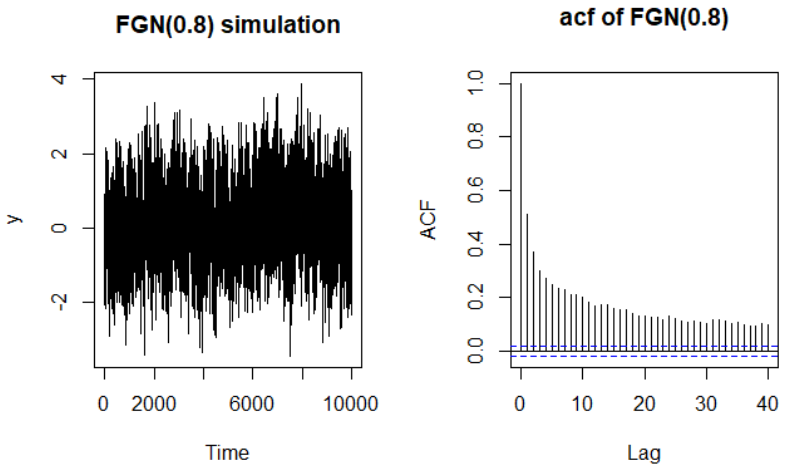

Figure 3: test

2.3.3 Long-range memory

Long-ranged dependences processes can be characterized by having their auto-covariance functions diverge.
An important property of long-range memory is self similarity. The Hurst parameter H can be used to check for long range dependency. The hurst parameter ranges from 0 to 1 and if it is 0.5 and below, it means the time series lacks long dependency.

3 Estimation of the Hurst Exponent

The R/S statistic was introduced by Hurst when he was looking and data from the height of a river. Thus long-ranged time series are widely studied in hydrology.
We will now compare different estimations of H using We will compare the HurstK, FitFGN and the one we implemented ourselves from the Rescaled Range.

3.1 The rescaled range

The first statistical observation of the long range dependency were done by Harold Edwin Hurst by using the rescaled range. He measured the annual height of the Nile River over many years in order to optimize the dam size in regards to seasonal wet and dry periods.The rescaled range measures the maximum and the minimum values is a given time period and divides by

the standard deviation. If a logarithmic plot of the rescaled range is done against the logarithm of the number observation we can find the slope. This slope will estimate the Hurst exponent.
To find the rescaled range [Bo Qian, 2005] we need to find the range and standard devitation.To find the range (R) of a time series $X = X_1, X_2..X_n$

$$R(n) = max(Z_1, Z_2, .., Z_t) - min(Z_1, Z_2, ..Z_t) \tag{21}$$

Where Z is the cumulative sum.

$$Z_t = \sum_{i=1}^{t}(X_t - mean(X)) \tag{22}$$

To find the Standard deviation series (S):

$$S_t = \sqrt{\frac{1}{t}\sum_{i=t}^{t}(X_i - \frac{1}{t}\sum_{i=1}^{t}X_i)^2} \quad for\ t=1,2,..n$$

So the rescaled range is the relationship between R and S.

$$(R/S)_t = \frac{R_t}{S_t} for\ t=1,2,..,n \tag{23}$$

Here we then make different segments of length n, n/2, n/3,n/4,n/16,n/32 and n/64. We find the rescaled range of each segment. The hurst parameter is found by

$$E(\frac{R_n}{S_n}) = Cn^H \tag{24}$$

where C is a constant.Here we can use find log of the Rescaled range of each segment and the log of the segment sizes and use linear regression to find a fit of the slope will be our H.

3.2 Simple non-parametric method.

The non-parametric estimate method is similar to the rescaled ranged with the difference being that it ffind the range of the whole time series X_t instead of doing it sequential as in (21). The standard deviation of the simple non-parametric method also is not empirical as in (3.1). To find H we divide the range by the standard deviation.

$$\hat{H} = \frac{log(R)}{log(S)} \tag{25}$$

3.3 Maximum likelihood estimation

A common and efficient method for estimating paremeters is the maximumlikehood method. For a fgn time series X_t the log likehood of μ, σ and H is

$$logL(\mu, \sigma, H) = -\frac{1}{2}log|\Sigma_n(H)| - (2\sigma)^{-1}(\mathbf{x} - \mu)^T[\Sigma_N(H)]^{-1}(\mathbf{x} - \mu) \quad (26)$$

where $\Sigma_n(H)$ is the NXN correlation matrix (20) and $\mathbf{x} - \mu$ is X_t represented as a 1XN vector subracted by the 1XN mean vector.To find the optimal H.

$$\frac{\delta logL}{\delta H} = 0 \quad (27)$$

The maximum log likelihood of H [Keith W. Hipel, 1994] then becomes

$$maxlogL(H) = \frac{1}{2}log|\Sigma_n(H)| - \frac{N}{2}log(\frac{(\mathbf{x} - \mu)^T[\Sigma_N(H)]^{-1}(\mathbf{x} - \mu}{N}) \quad (28)$$

3.4 Estimation using the INLA methodology

bayesian inference? [Sørbye SH, 2017] To understand the INLA methodology we first need to introduce a few concepts.

3.4.1 Bayesian inference

It uses Bayesian inference meaning that our goal is to find the posterior distribution $p(\theta|\mathbf{y})$ of a parameter and this is equal to the prior distribution $p(\theta)$ times the likelihood $p(\mathbf{y}|\theta)$.

$$p(\theta|\mathbf{y}) = \frac{p(\mathbf{y}|\theta)p(\theta)}{\int p(\mathbf{y}|\theta)p(\theta)d\theta} \propto p(\mathbf{y}|\theta)p(\theta) \quad (29)$$

It is called Bayesian inference because it derives from the Bayes rule and it can be used to find confindence intervals and other statistical properties of a parameter.Parameter in Bayesian inference are treated as random variables. Conjugate priors can be used to find the closed form of the posterior distribution because they make the integrals easy.

3.4.2 Latent gaussian models

Latent Gaussian models are used as conjugate priors. They have three characteristics. Given observations $\mathbf{y} = (y_1, \ldots, y_m)^T$ and the latent field of unobserved variables $\mathbf{x} = (x_1, \ldots, x_n)^T$ the likelihood function is assumed to be conditionally independent

$$p(\mathbf{y}|\mathbf{x}, \theta_\mathbf{1}) = \prod_{i=1}^{n} p(y_i|(\mathbf{x}, \theta) \quad (30)$$

where $\theta = (\theta_1, \ldots, \theta_m)$ is the hyper-parameter vector.The other characteristic is that the latent field of unobserved variables $\mathbf{x}$ given the the hyperparameters θ are multivariate-normal.

$$\mathbf{x}|\theta \sim N(\mu(\theta), \mathbf{\Sigma}(\theta))) \tag{31}$$

The last characterstic say the prior distribution is formed by the hyper-parameter.

$$\theta \sim p(\theta) \tag{32}$$

where μ is the mean vector and Σ is the correlation matrix.

3.4.3 Gaussian Markov Random Field

When LGM have the Markov property they become Gaussian Markov random field(GMRF). The markov property of the field means that there is conditional independence between pairs.

$$p(x_i, x_j|\mathbf{x}_{-ij}) = p(x_i|\mathbf{x}_{-ij})p(x_j|\mathbf{x}_{-ij}) \tag{33}$$

GMRFs become useful because it can be shown that the precision matrix , that is $\mathbf{Q} = \Sigma^{-1}$, is equal to 0

$$p(x_i \perp x_j|\mathbf{x_{-ij}}) \Leftrightarrow \mathbf{Q}_{ij} = 0 \tag{34}$$

When the precision matrix of GMRF is sparse, the computional cost becomes very low. FGN models are not GMRFs but in the paper by [**?**] they can be approximated to GMRF . The idea is to aproximate FGN with a weighted sum of independent first order AR1 proccesses

$$z_{j,t} = \phi_j z_{j,t-1} + \nu_{j,t} \quad j = 1, \ldots, m \quad t = 1, \ldots, n \tag{35}$$

and choose parameter that fit acf of the FGN. Here j is the number of AR(1) processes,ϕ is the autocorrelation coefficient and $\nu_j \sim N(0, \sigma^2/(1-\phi_j^2))$

alternative to markov chain monte carlo simulation hvor numerisk beregnet aposteriori fordeling

4 Results

4.1 Simulation results

Now to check for the accuracy of the estimation methods, we can apply the methods on simulated FGNs. Since the FGNs are simulated we know the true H and and see which method does best. An expected result is that the maximum likelihood method and the INLA method should give best results, followed by the rescaled range and lastly the non-parametric

Data	Real H	Rescaled Range	HurstK	FitFGN	INLA
SimulatFGN	H=0.6	0.6447	0.6538	0.5913	0.5921
Standard Deviation		0.0645	0.0442	0.0291	0.0275
SimulatFGN	H=0.7	0.7209	0.7114	0.6929	0.692
Standard Deviation		0.0713	0.0474	0.0297	0.0299
SimulatFGN	H=0.8	0.7922	0.7677	0.7906	0.7907
Standard Deviation		0.0764	0.0509	0.0309	0.0308
SimulatFGN	H=0.9	0.8578	0.8177	0.8891	0.8865
Standard Deviation		0.0780	0.0510	0.0301	0.0284

Table 1: Estimation of H from simulataed FGNs and their standard deviation

method. A 1000 simulations will reflect how stable the estimation methods are by look at the standard deviation and it should not be high. Looking at the table 1 INLA gives the best results and followed by the maximum likelihood which was expected. The Rescaled range does better than the HurstK even though it has higher standard deviation. We can argue that the results make sense, since the non-parametric method naively Range and standard deviation of the series as whole.

4.2 Real data results

Now that we have looked simulations of fgn processes, the next step is to see how good the estimation methods are on real data. The historical Nile river measurements are a good example of long ranged data. The measurements where taken from the year 622 for 1284. In ([S. Benmehdi, 2011]) their estimated Hurst parameter was 8.3 Another data set we can look is annual temperature measure of England. [C.K Folland, 1992]

References

[Bo Qian, 2005] Bo Qian, K. R. (2005). Hurst exponent and financial market predictability.

[C.K Folland, 1992] C.K Folland, D. P. T. L. (1992). Volume contents and author index, volume 12, 1992. *International Journal of Educational Development*, 12(4):317–342.

[Keith W. Hipel, 1994] Keith W. Hipel, A. I. M. (1994). *Time Series Modelling of Water Resources and Environmental Systems*. Elsevier, Amsterdam,The Netherlands. Ch.10.

[Myrvoll-Nilsen, 2016] Myrvoll-Nilsen, E. (2016). Computationally efficient bayesian approximation of fractional gaussian noise using ar1 processes.

[S. Benmehdi, 2011] S. Benmehdi, N. Makarava, N. B. (2011). Bayesian estimation of the self-similarity exponent of the nile river fluctuation. *Nonlinear Processes in Geophysics*, 18.

[Sørbye SH, 2017] Sørbye SH, R. H. (2017). Fractional gaussian noise: Prior specification and model comparison. S.