Statistical analysis of non-standard time series models

Many time series of practical interest show strong dependence (or long memory). This phenomenon appears visually as fractal behaviour and is connected to heavy tails of a distribution of the series. This aspect occurs for example in computer networks. It is empirically observed that network data follow fractal behaviour.

What is the goal ?
The goal of this project is to describe statistical behaviour of network traffic models and estimate relevant parameters such as the memory parameter

Model description

Traffic Process from one source : $W(t) = \sum_{j=-\infty}^{\infty} 1(\tau_j < t < \tau_j + X_j), \qquad t \in [0, \infty)$
-Tj points from Poisson Process = beginning of a transmission
-Xj random variables = duration times of each transmission

Assumption 1 :

Duration follow heavy-tailed distribution:

$$P(X > x) \sim \text{const. } x^{-\alpha}, \qquad \alpha \in (1, 2).$$

The parameter α is related to the size

Assumption 2 :

Sequences Tj and Xj are independent form each other.

Long memory behaviour

$$\text{Var}\left(\int_0^t W(u)du\right) \sim \text{const. } t^{3-\alpha},$$

With α in (1,2) which indicates long range dependent behaviour.

Then $d = 1 - \alpha/2$ is called the **memory parameter**

Central limit theorem not valid anymore for the long memory traffic models

Traffic models that exhibit long memory .

Rescaled Range method

The re-scaled range method R/S was introduced by hydrologists Hurst who studied water level in Nile river. For a given time series X_t, $t = 1, \ldots, n$ (in our case, X_t is a number of packets in consecutive time intervals) we define

$$R_n = \max_{1 \le k \le n} \sum_{t=1}^{k} (X_t - \bar{X}_n) - \min_{1 \le k \le n} \sum_{t=1}^{k} (X_t - \bar{X}_n),$$

where $\bar{X}_n$ is the sample mean. Furthermore, we define the sample variance

$$S_n^2 = \frac{1}{n-1} \sum_{t=1}^{n} (X_t - \bar{X}_n)^2.$$

Then, the R/S statistics is defined as

$$\text{R/S-statistics} = \frac{R_n}{S_n}.$$

It is calculated from dividing the range R_n of the values exhibited in a portion of the time series by the sample standard deviation S_n of the values over the same portion of the time series.
Intuitively, we have

$$\log(R_n/S_n) \approx (d + 1/2) \log n + \text{error}.$$

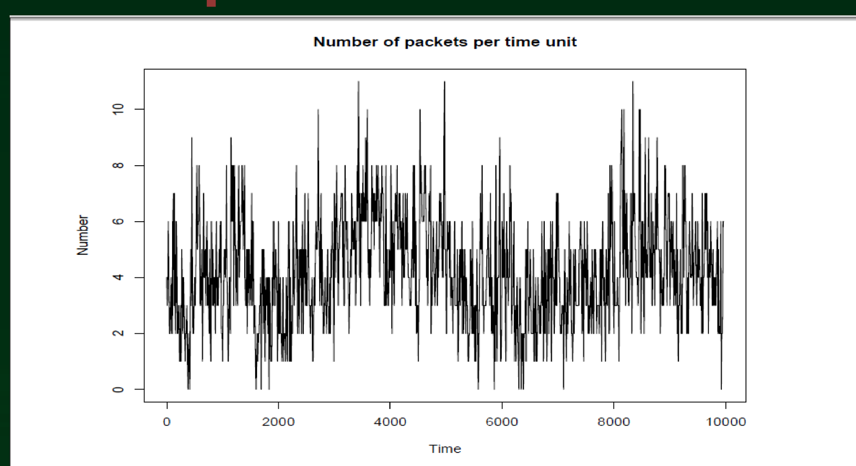

Figure 1: Simulation of network traffic model

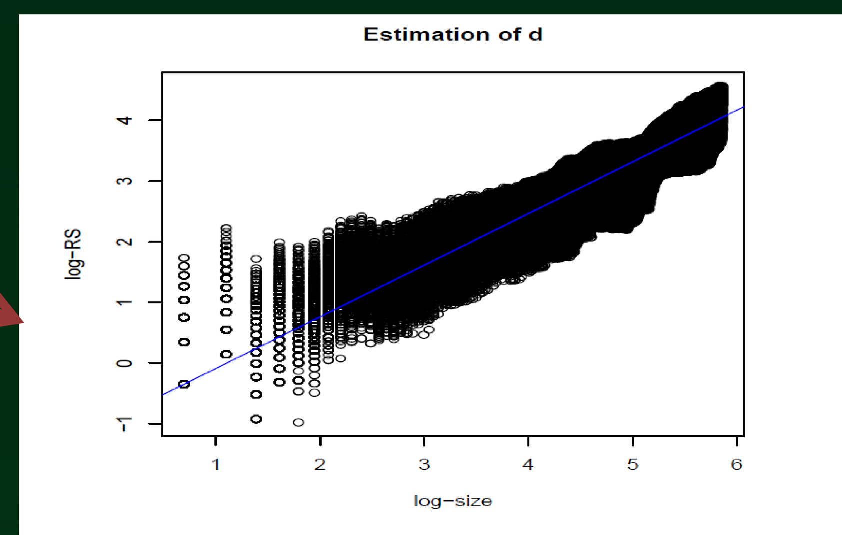

Figure 2: R/S method

Estimation of the memory parameter :

d+1/2 can be interpreted as the slope of a regression line of log(Rn/Sn) against log n.

$$\hat{d}_{R/S} = \frac{\log(R_n/S_n)}{\log n} - \frac{1}{2}$$

is the R/S estimator of the memory parameter d which is needed needed to construct a network in an optimal way. Intuitively, the bigger d is, the bigger the capacity of the network will be.

Few definitions :

Long memory: A time series is said to have a long memory if values from the distant past still have an effect on present values.
Fractal behaviour: It is a type of pattern used in technical analysis to predict a reversal in the current trend.
Computer network: It can be defined as a group of computers and other devices connected in some ways in order to exchange data. It usually consists of many sources.
Heavy-tailed distribution: It is a distribution that assigns relatively high probabilities to regions far from the mean or the median.
Long-range dependency: It is a notion that can be defined as a slow decay of correlations .It is a phenomenon that may arise in the analysis of time series data.

RESULTS

We analysed a simulated network traffic model. We computed the estimator of the memory parameter for that simulated traffic model. The results indicate a good performance of the R/S-estimator. Our simulations studies motivate further research on theoretical properties of the estimator and its precise relation to the optimal capacity of the network.
We also simulated a traffic model from M = 100 sources. We chose alpha = 1.3. The simulated model is displayed on figure 1. The result of the regression procedure is shown on figure 2. The estimated parameter d is d = 0.350714, whereas the formula d = 1− alpha/2 gives 0.35. Hence, our estimation is very accurate.

Did you know ?

In October 2011, the Blackberry network crashed, leaving its users unable to browse or use BBM. This was due to a data backlog. Blackberry's network capacity was not big enough to support all the data that was sent and received.

Thank you to Professor Rafal Kulik for his supervision and to the UROP project for this opportunity

Dioulde Mariko
dmari060@uottawa.ca