Time series 'memory' analysis

Mikhailov Dmitri
Supervisor of studies : Grafeyeva Natalya Genrikhovna
Saint-Petersburg State University

December 20, 2013

1 Introduction

An important issue in a number fields of science is prediction or forecasting of time series. It is widely applied from economics to speech recognition. In all cases the accurate prediction is required.

For effective putting together the forecast it is essential to analyse how previous values of time series influence future values:

- identify fitting model for different time series;
- set optimal set of data for prediction, etc.

One of similar subproblems is analysis of autocorrelation function. More precisely, time series 'memory' research using autocorrelation function.

2 Autocorrealtion function

The aim of this work is to extract any information about presence of periodicity, identify the missing fundamental frequency, based on values of autocorrelation function.

The autocorrelation function describes the correlation between values of the random processes of the same time series at different lags. Let X be some repeatable process, and i be some point in time after the start of that process. Then X_i is the value of the process at time i. Suppose that the process has known values for mean μ and variance σ^2. Then the definition of the autocorrelation of lag s for any value of t is

$$\rho(k) = \frac{E((X_t - \mu)(X_{t+k} - \mu))}{\sigma^2}$$

Obviously, $\rho(0) = 1$. It is, also, an even function ($\rho(k) = \rho(-k)$). Since that fact, autocorrelation function is symmetric about zero.

However, theoretical autocorrelation function cannot be applied in practice. In c̃itebox given next simple estimations of mean μ and autocorrelation. Suppose that we have time time series $z_1, z_2, \ldots, z_N$ of N observations. Then the mean μ is estimated as $\bar{z} = \sum_{i=1}^{n} z_t/N$. And estimation of the kth lag autocorrelation $\rho(k)$ is

$$r_k = \frac{c_k}{c_0}$$

where

$$c_k = \frac{1}{N} \sum_{t=1}^{N-k} (z_t - \bar{z})(z_{t+k} - \bar{z}) \qquad k = 0, 1, 2, \ldots, K$$

.

3 Data and methodology

Before applying any methods of identifying 'memory' to time series with unknown parameters it is necessary to test them on standard time series. In this work 3 types of standard time series were considered:

1. Base time series ('memoryless' Markov process). Time series values are obtained as $U(n) = \exp(a) * U(n-1)$, where a is normally distributed random variable with mean $\mu = 0$

2. Time series with 'short memory'. Values are obtained as $U(n) = \exp(b*a)*U(n-1)$, where a is the same and b is random variable taking one of two values: $+1$ and -1:

   (a) Persistent process: $P(b = +1) > P(b = -1)$

   (b) Antipersistent process: $P(b = +1) < P(b = -1)$

3. Time series with 'heavy memory'. Time series is built with the base time series formula in which, for example, first thousand values the normally distributed random variable with positive mean is taken and for the next thousand values — with negative mean. Then this process can be repeated infinitely.

Each of these time series was programmed with library <random> of c++11 standard. In each case, 50000 and 10000 subsequent values of time series and autocorrelation, respectively, were obtained 3 times. Actually, since $r_0 = 1$, only 9999 $(r_1, \ldots, r_{9999})$ values were considered. Used next parameters for programming:

1. Markov process: $\sigma(a)^2 = 3$

2. Persistent process: $\sigma(a)^2 = 3$ and $P(b = +1) = 0.7$

3. Antipersistent process: $\sigma(a)^2 = 3$ and $P(b = -1) = 0.2$

4. Time series with 'heavy memory': for each thousand of values normally distributed random variables with random mean and $\sigma^2 = 3$ were taken; mean μ takes values $\pm 1, \pm 2$

2(x3) charts for each of time series were built based on 9999 and 100 first values of autocorrelation function. All charts are provided at Appendix.

4 Interpretation of results

As can be seen from charts, there are features for each of standard time series. Most of values of autocorrelation function are negative and close to zero. At the same time, all peak values differs significantly from 0. However, any cycles, periods or fluctuations are not found.

5 Conclusion

This work has been aimed to find link between autocorrelation function and of time series. It had to get information about time series 'memory' from autocorrelation. Nevertheless, no positive results were obtained on 3 types of standard time series considered. Therefore, the further research for wider range of different data is necessary.

References

[1] Box, George E.P. Time Series analysis : forecasting and control / George E.P. Box, Gwilym M. Jenkins and Gregory C. Reinsel. –3rd ed. 1994 by Prentice-Hall, Inc.

[2] Ширяев А.Н. Основы стохастической финансовой математики. Том 1. Факты. Модели. Москва :ФАЗИС, 1998.

[3] William S. Hopwood. On the automation of the Box-Jenkins modeling procedures: an algorithm with an empirical test. Faculty working papers. College of Commerce and Business Administration. University of Illinois at Urbana-Champaign, 1978

[4] James C. McKeown, Kenneth S. Lorek. A comparative analysis of the predictive ability of adaptive forecasting, reestimation and reidentification using Box-Jenkins time series analysis. Faculty working papers. College of Commerce and Business Administration. University of Illinois at Urbana-Champaign, 1976

[5] Anil K. Bera, Sangkyu Lee. Interaction between autocorrelation and conditional heteroskedasticity: a random coefficient approach. Faculty working papers. College of Commerce and Business Administration. University of Illinois at Urbana-Champaign, 1989

[6] Tucker McElroy, Dimitris N. Politis. Distribution Theory for the Studentized Mean for Long, Short, and Negative Memory Time Series

[7] Oliver D. Anderson. Small-Sample Autocorrelation Structure for Long-Memory Time Series. The Journal of the Operational Research Society, Vol. 41, No. 8 (Aug., 1990), pp. 735-754

[8] Richard Finlay, Thomas Fung, and Eugene Seneta. Autocorrelation Functions. International Statistical Review (2011), 79, 2, 255–271

A Appendix. Charts

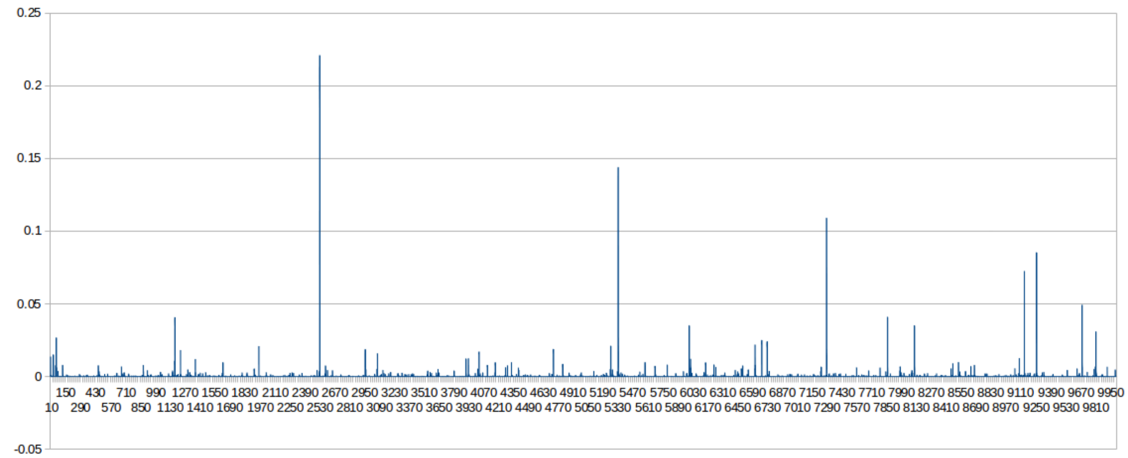

Figure 1: Autocorrelation of Markov process 1

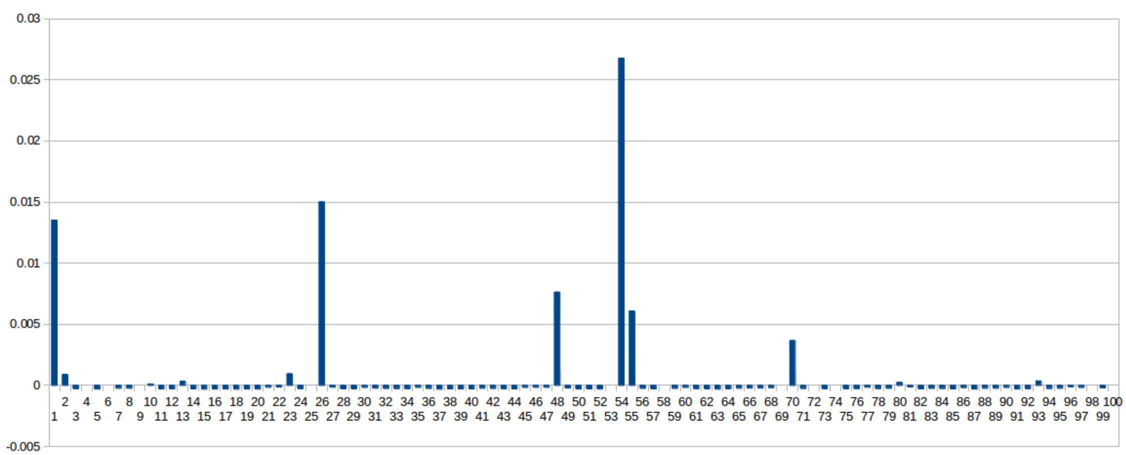

Figure 2: First 100 values of chart in Figure 1

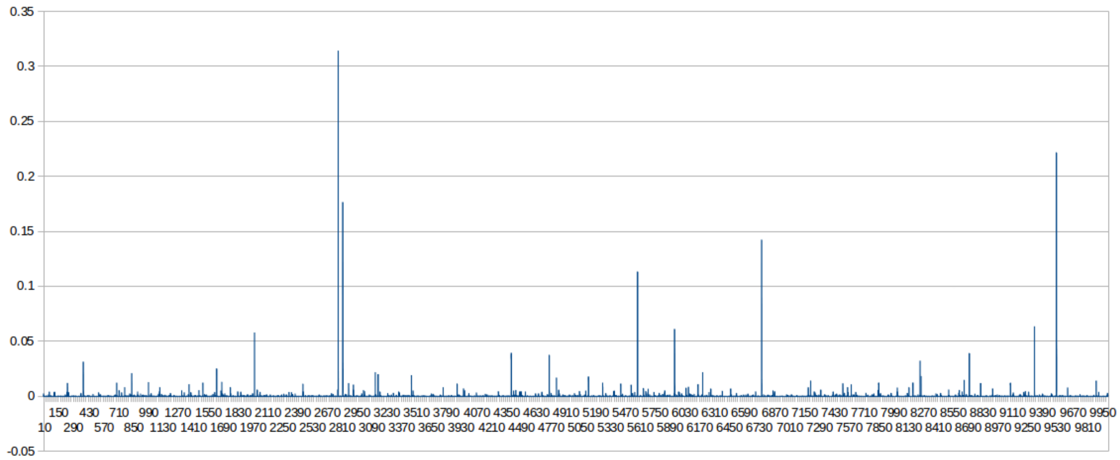

Figure 3: Autocorrelation for Markov process 2

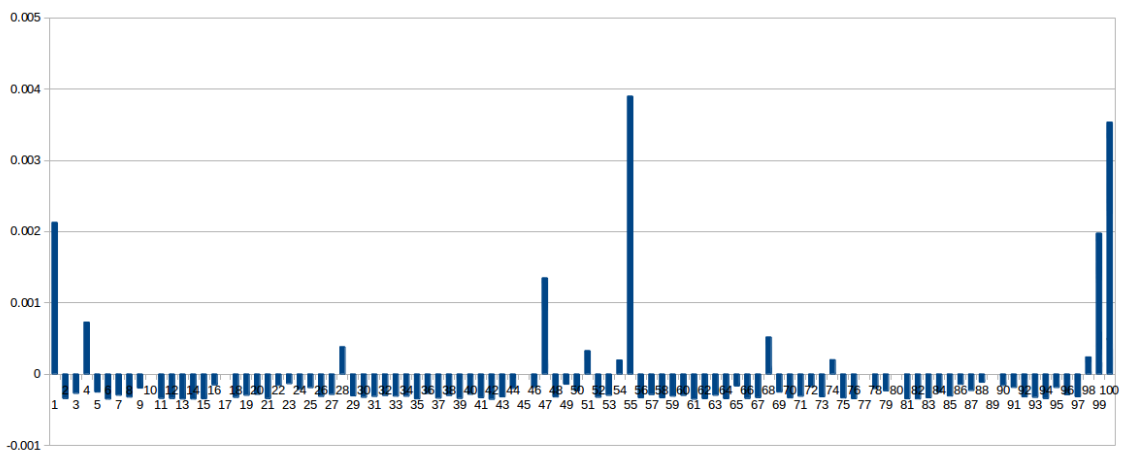

Figure 4: First 100 values of chart in Figure 3

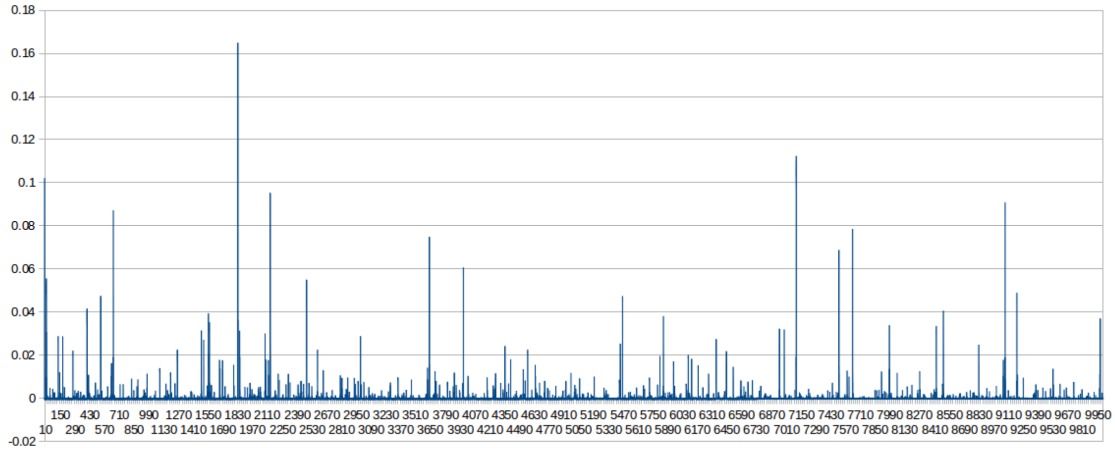

Figure 5: Autocorrelation of Markov process 3

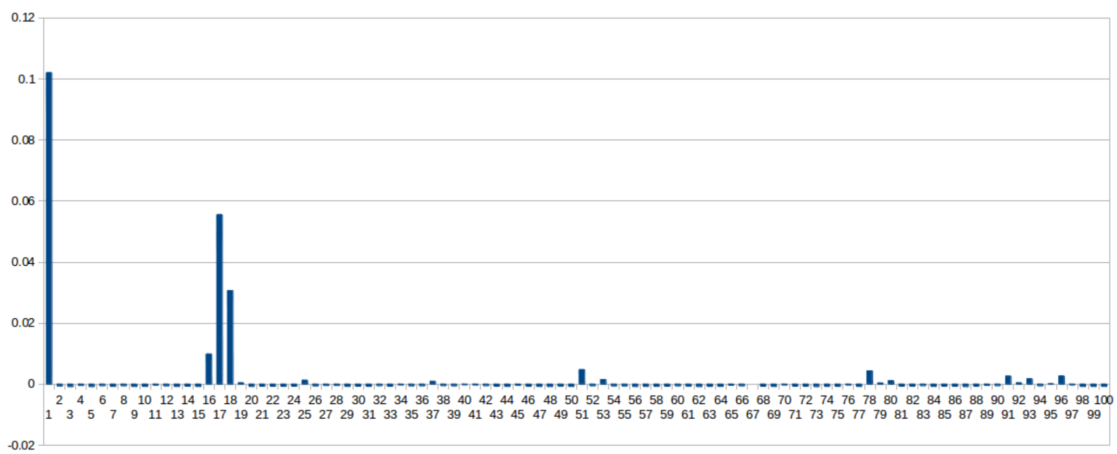

Figure 6: First 100 values of chart in Figure 5

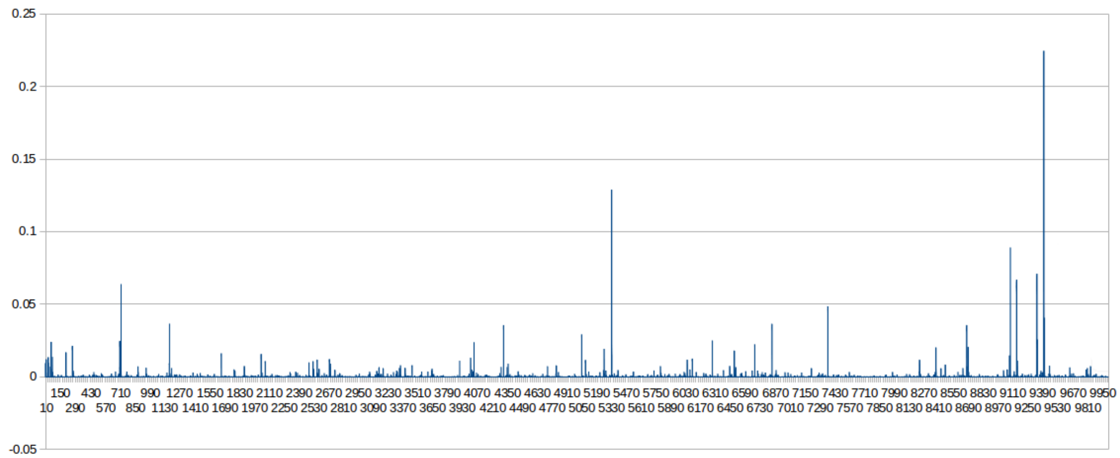

Figure 7: Autocorrelation of Persistent process 1

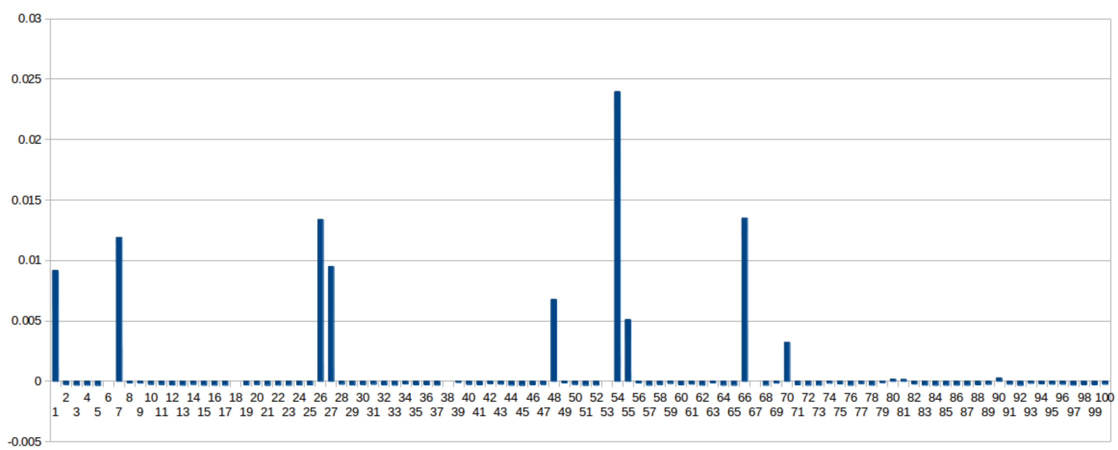

Figure 8: First 100 values of chart in Figure 7

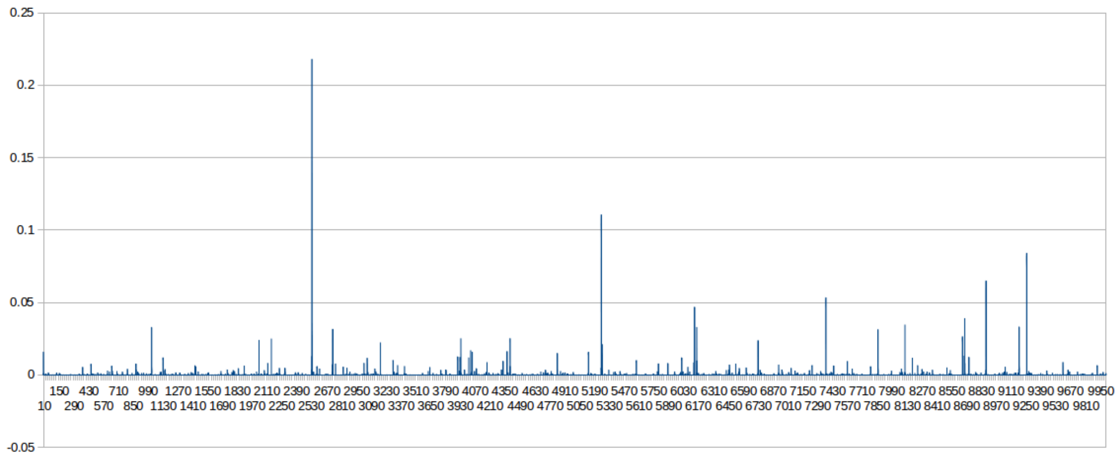

Figure 9: Autocorrelation of Persistent process 2

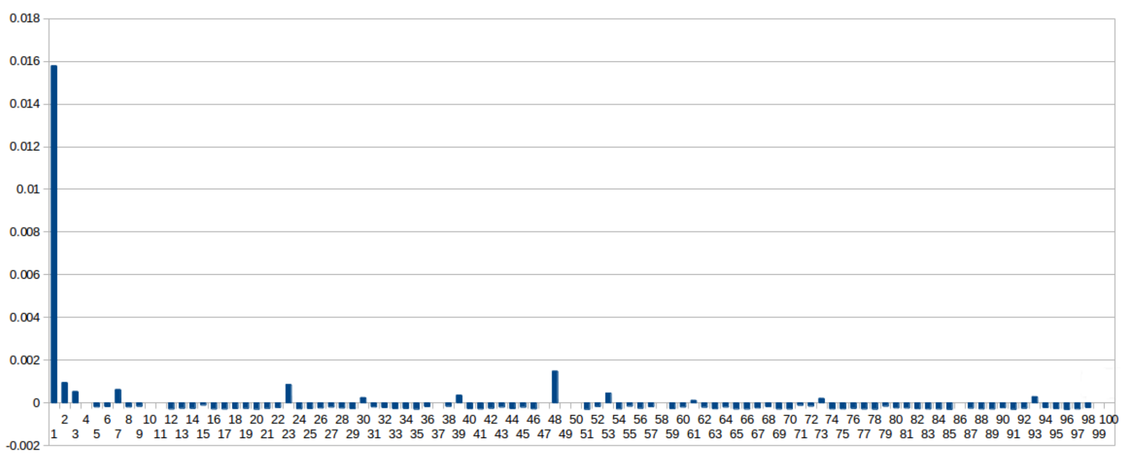

Figure 10: First 100 values of chart in Figure 9

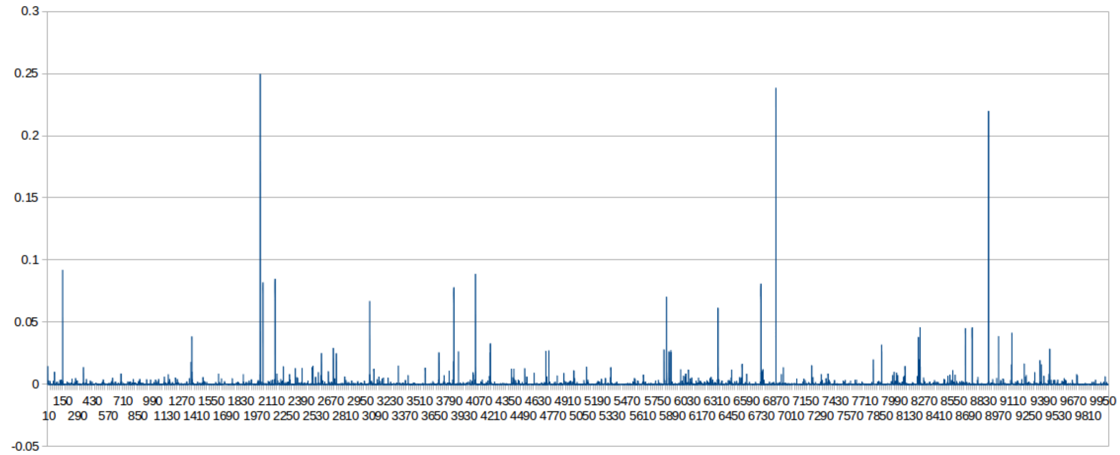

Figure 11: Autocorrelation of Persistent process 3

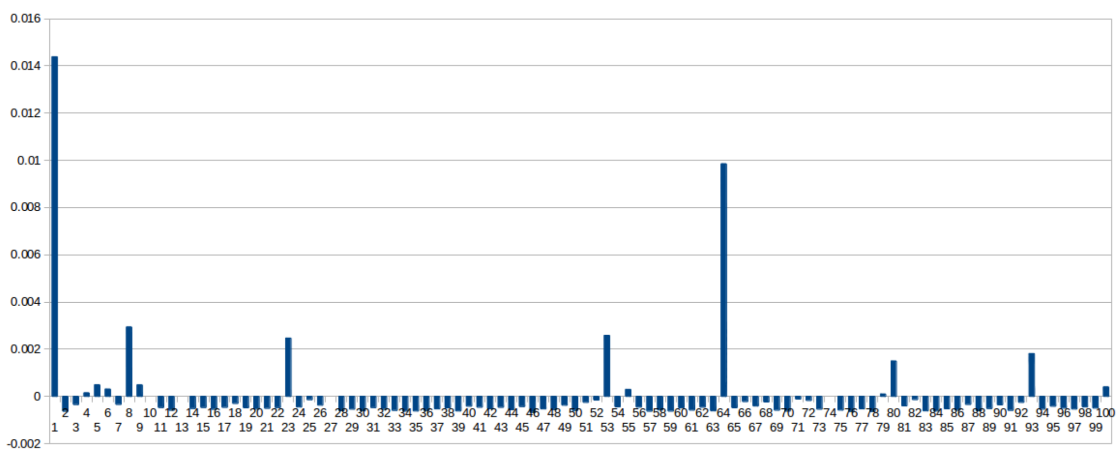

Figure 12: First 100 values of chart in Figure 11

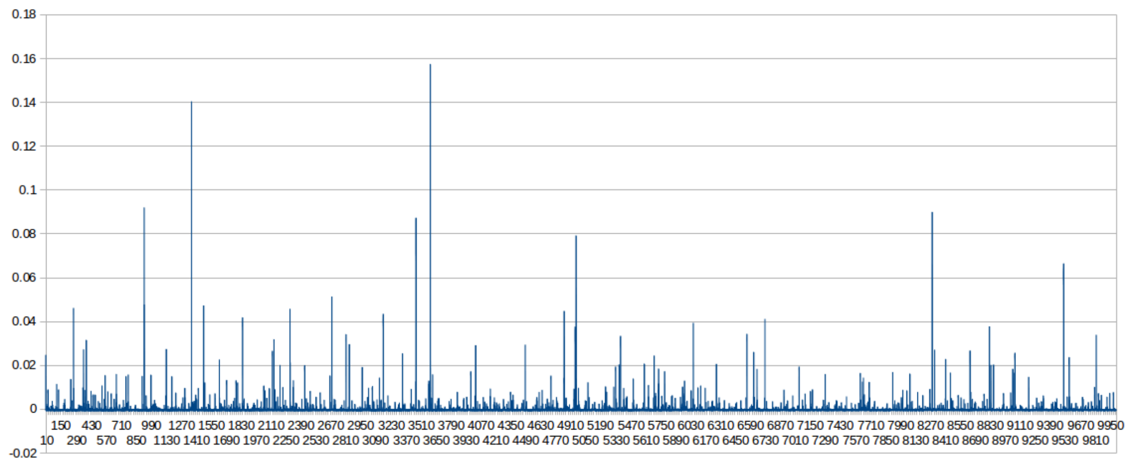

Figure 13: Autocorrelation of Antipersistent process 1

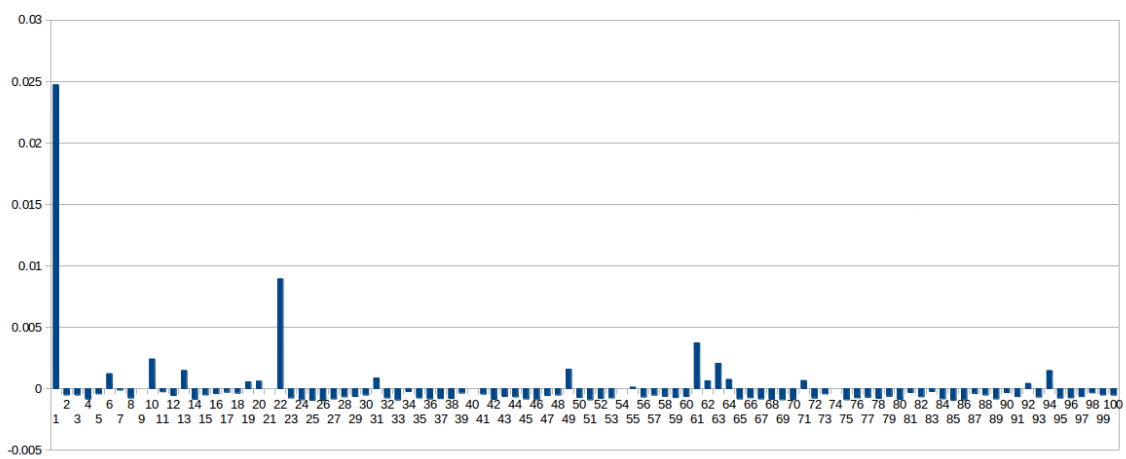

Figure 14: First 100 values of chart in Figure 13

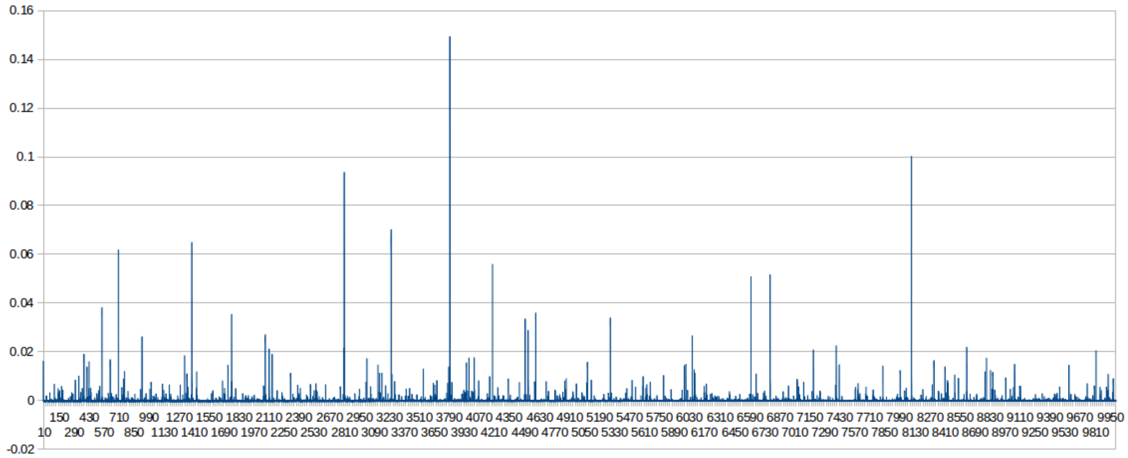

Figure 15: Autocorrelation of Antipersistent process 2

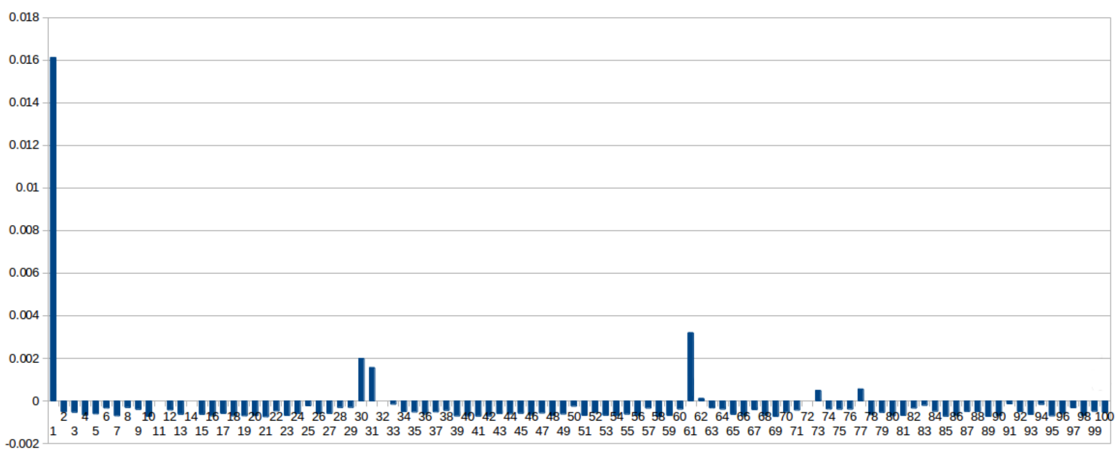

Figure 16: First 100 values of chart in Figure 15

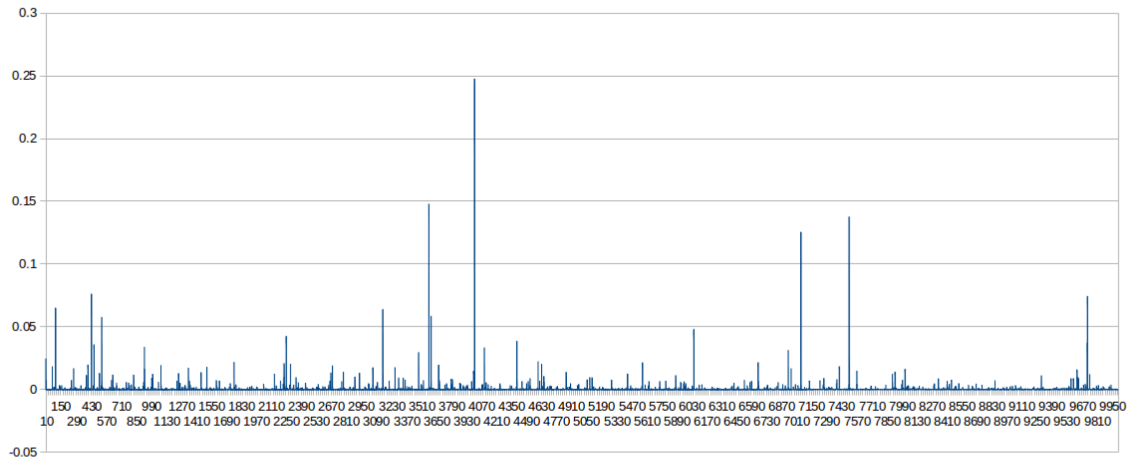

Figure 17: Autocorrelation of Antipersistent process 3

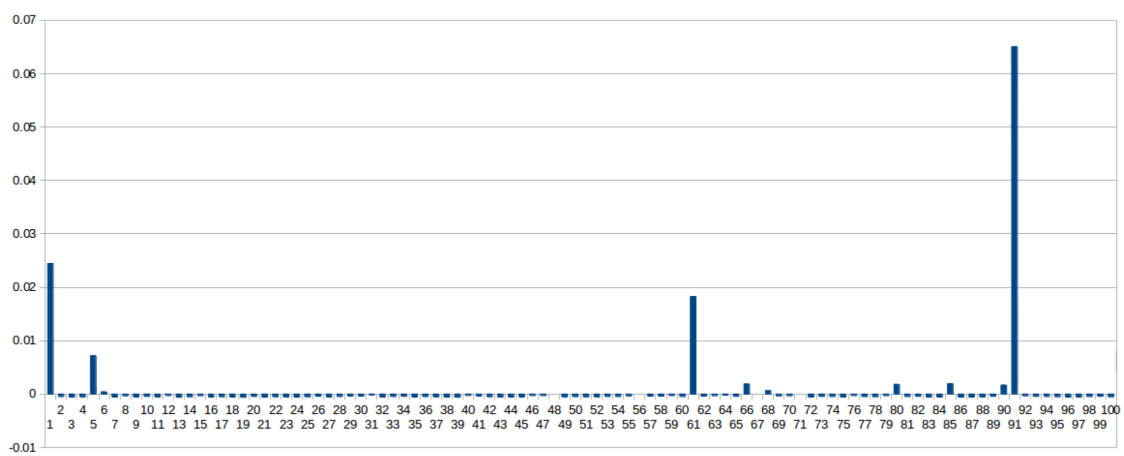

Figure 18: First 100 values of chart in Figure 17

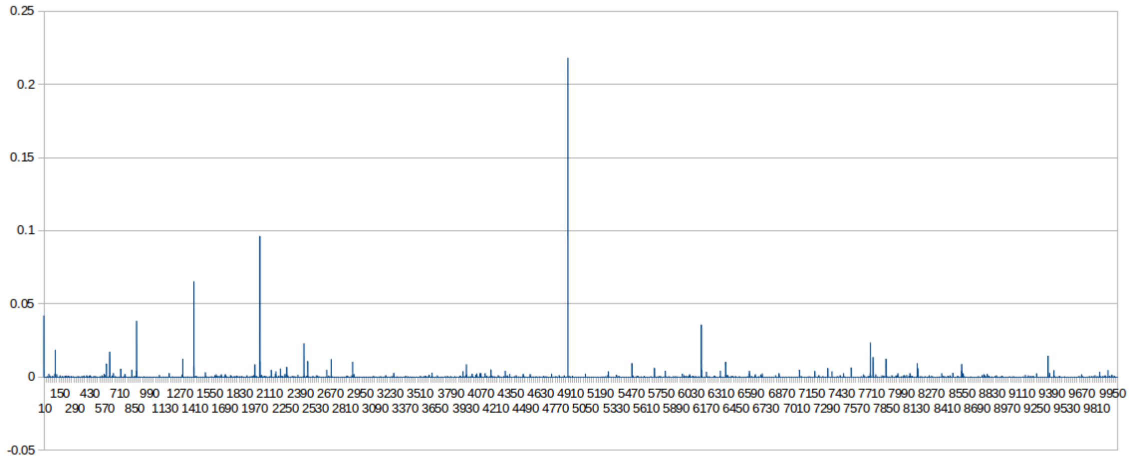

Figure 19: Autocorrelation of process with 'heavy memory' 1

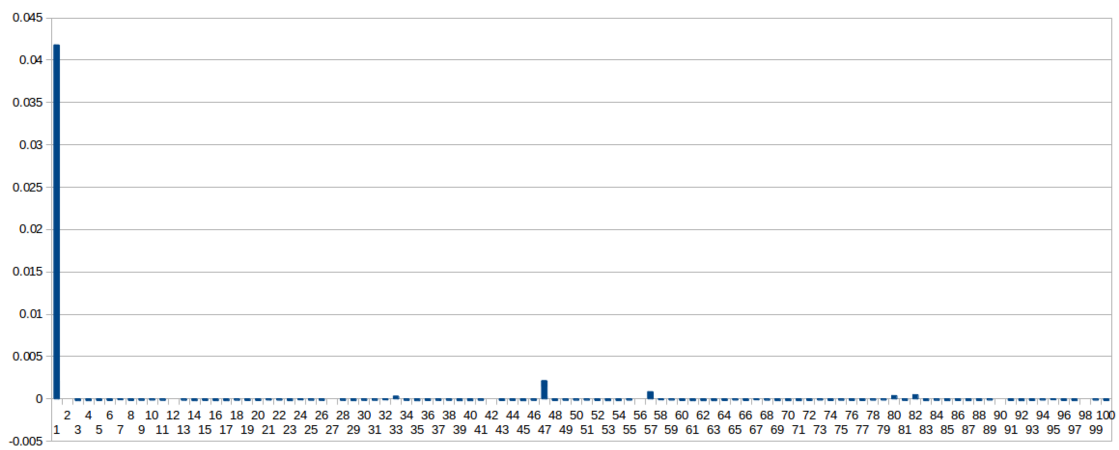

Figure 20: First 100 values of chart in Figure 19

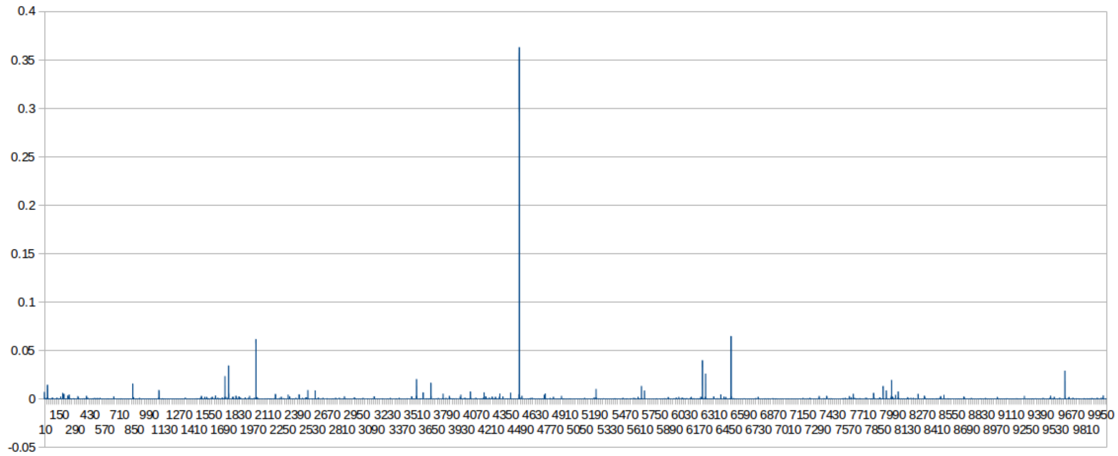

Figure 21: Autocorrelation of process with 'heavy memory' 2

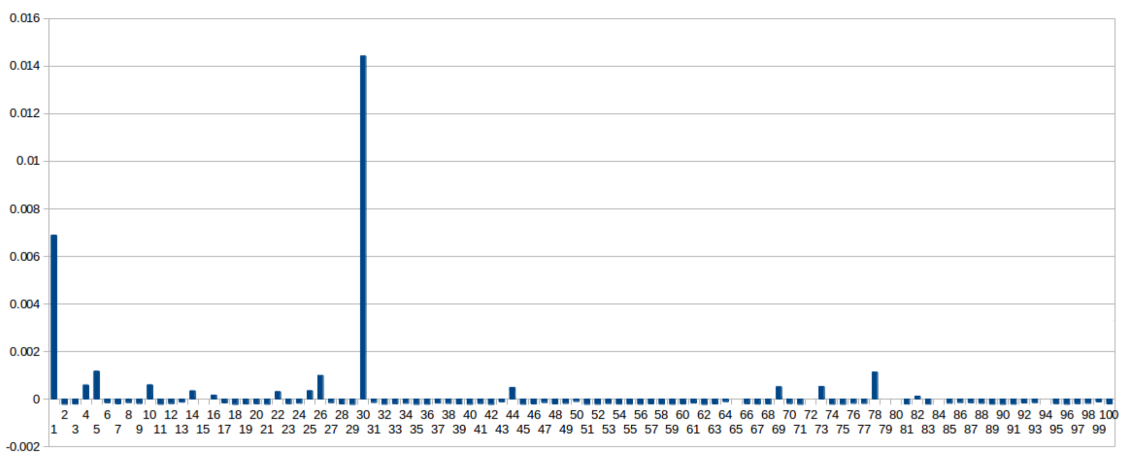

Figure 22: First 100 values of chart in Figure 21

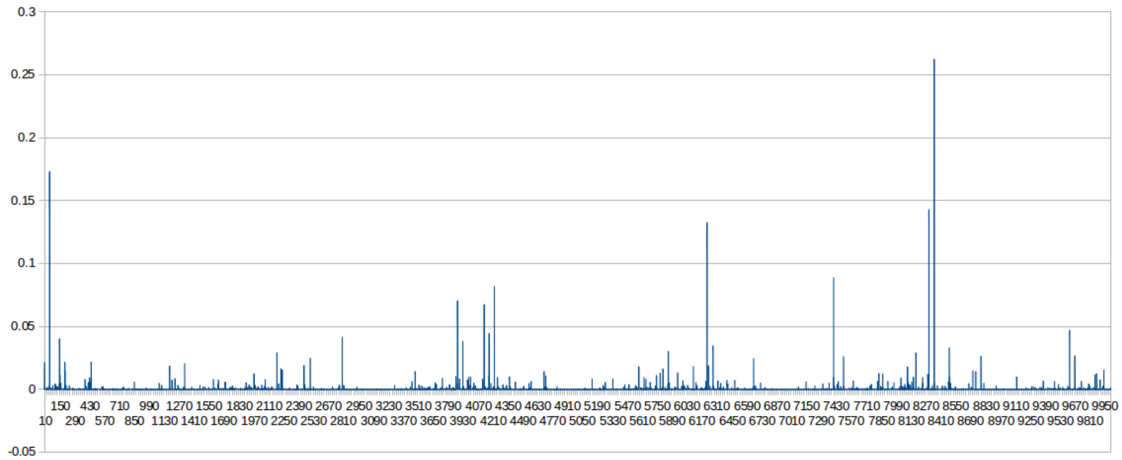

Figure 23: Autocorrelation of process with 'heavy memory' 3

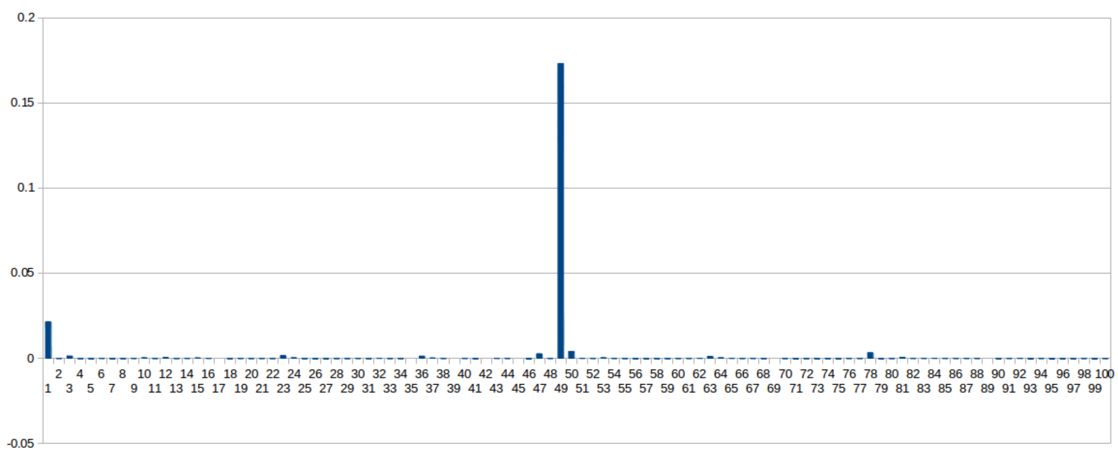

Figure 24: First 100 values of chart in Figure 23