Markov Chains and Buffering Strategies in Networks

Producing (and then suppressing) Correlations in Simulated Packet Data

Richard G. Clegg (richard@richardclegg.org)

Department of Mathematics,
University of York,
YO10 5DD

Queen Mary, University of London, March 2005

Introduction

- Looking for simple way of generating an ON-OFF series with a rich correlation structure — long-range dependence (LRD).
- Markov modulated processes are becoming a well-known tool for this.
- A simple method gives a known mean and asymptotic form of autocorrelation function.

Introduction

- Looking for simple way of generating an ON-OFF series with a rich correlation structure — long-range dependence (LRD).
- Markov modulated processes are becoming a well-known tool for this.
- A simple method gives a known mean and asymptotic form of autocorrelation function.
- This LRD stream is used as input to a network model.
- A simple method is used to suppress the ACF at given lags.
- Second part all experimental — I do not know how to do the mathematics.

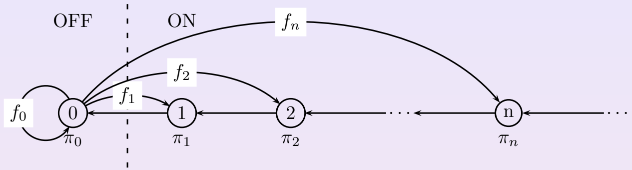
Model Requirements

- Generate a time series $\{Y_t : t \in \mathbb{N}\}$ of zeros and ones.
- The ones represent packets and the zeros represent interpacket gaps.
- The time series to be generated has a known mean μ.
- The autocorrelation function has the specific form

$$\rho(k) \sim Ck^{-\alpha},$$

  where $C \in \mathbb{R}_+$ and $\alpha \in (0, 1)$.
- The parameter α can be specified (related to LRD and Hurst parameter).

The Markov Model



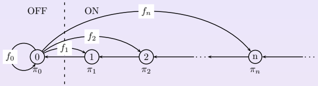
- If $\{X_t : t \in \mathbb{N}\}$ is a realisation of this chain then generate

$$Y_t = \begin{cases} 0 & X_t = 0 \\ 1 & \text{otherwise.} \end{cases}$$

- The f_i are trans. prob. and the π_i equilibrium densities. (Want simple values of f_i to work with.)
- Existing work on this chain: Wang[1989], Barenco[2003].

Setting the Transition Probabilities

- Two parms α and $\mu = 1 - \pi_0$ if ergodic.
- Find f_k such that $\sum_{i=k}^{\infty} \pi_i \sim Ck^{-\alpha}$.



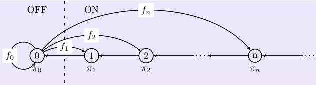
Transition Probabilities for this Markov model

$$f_k = \begin{cases} \frac{1-\pi_0}{\pi_0}\left[k^{-\alpha} - 2(k+1)^{-\alpha} + (k+2)^{-\alpha}\right] & k > 0 \\ 1 - \frac{1-\pi_0}{\pi_0}\left[1 - 2^{-\alpha}\right] & k = 0 \end{cases}$$

- From balance equations $\pi_k = \pi_{k+1} + f_k \pi_0$.
- Thus $\pi_k = \pi_0 \sum_{i=k}^{\infty} f_i$. (Note, if $k = 0$ this says $\pi_0 = \pi_0$).
- For $k > 0$ then $\pi_k = (1 - \pi_0)[k^{-\alpha} - (k+1)^{-\alpha}]$.
- Hence $\sum_{i=k}^{\infty} \pi_i = (1 - \pi_0)k^{-\alpha}$ for $k > 0$ as required.

Proving the Form of the ACF (1)

Since $\sum_{i=k}^{\infty} \pi_i \sim Ck^{-\alpha}$ suspect $\rho(k) \sim Ck^{-\alpha}$ but need proof.



- Let $F(n)$ be the dist. fn. for the return time to $X_t = 0$.
- Let N_n be no. of $X_i = 0$ in $X_1, \ldots X_n$ if $X_0 = 0$.
- Feller [1949]: If $1 - F(n) \sim An^{-\gamma}$ where $A > 0$ and $1 < \gamma < 2$, then

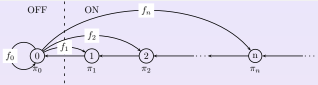
$$\mathrm{var}\,(N_n) \sim \frac{2A\pi_0^3 n^{3-\gamma}}{(2-\gamma)(3-\gamma)}.$$

- It can also be shown for this system that

$$\rho(n) \sim \frac{\mathrm{var}\,(N_{n+1}) - 2\mathrm{var}\,(N_n) + \mathrm{var}\,(N_{n-1})}{2\pi_0(1-\pi_0)}.$$

Proving the Form of the ACF (2)

Need to find form of $1 - F(n) = 1 - \sum_{i=0}^{n-1} f_i$ which in turn equals $\sum_{i=n}^{\infty} f_i$.



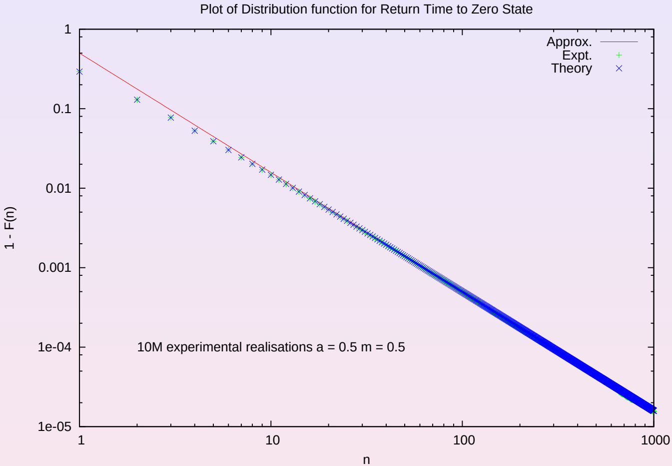
$$1 - F(n) = \left(\frac{1-\pi_0}{\pi_0}\right)\sum_{i=n}^{\infty}[i^{-\alpha} - 2(i+1)^{-\alpha} + (i+2)^{-\alpha}] \qquad n > 0$$
$$= \left(\frac{1-\pi_0}{\pi_0}\right)[n^{-\alpha} - (n+1)^{-\alpha}] \sim \left(\frac{1-\pi_0}{\pi_0}\right)\alpha n^{-(1+\alpha)},$$

which was the form needed. Hence $\rho(k) \sim \alpha\pi_0 k^{-\alpha}$.

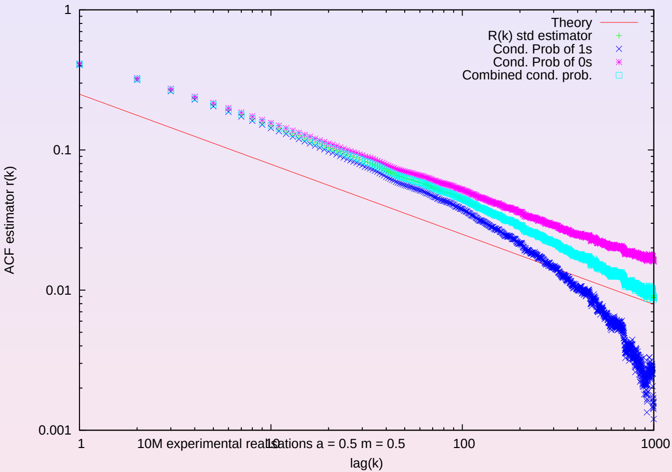
Interesting aside for **any** two-valued w.s. series:

$$\rho(k) = \frac{\mathbb{P}[Y_{t+k} = 1 | Y_t = 1] - \mathbb{P}[Y_t = 1]}{\mathbb{P}[Y_t = 0]}$$
$$= \frac{\mathbb{P}[Y_{t+k} = 0 | Y_t = 0] - \mathbb{P}[Y_t = 0]}{\mathbb{P}[Y_t = 1]}.$$

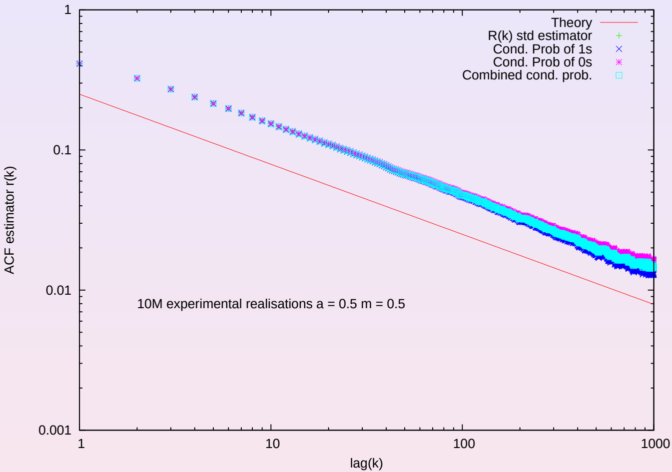
Distribution Function of Return Times



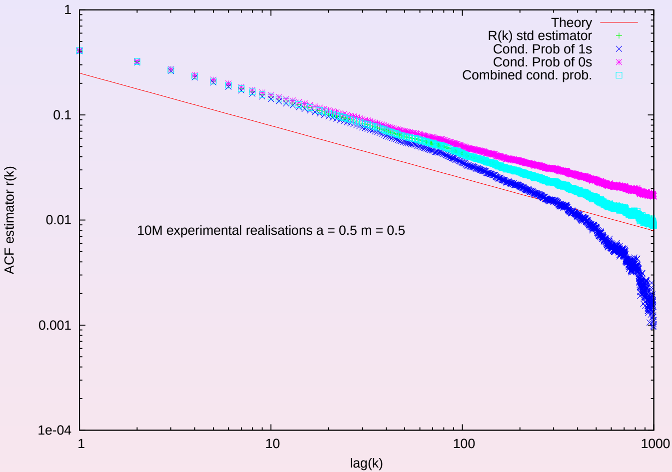
Autocorrelation Function versus lag



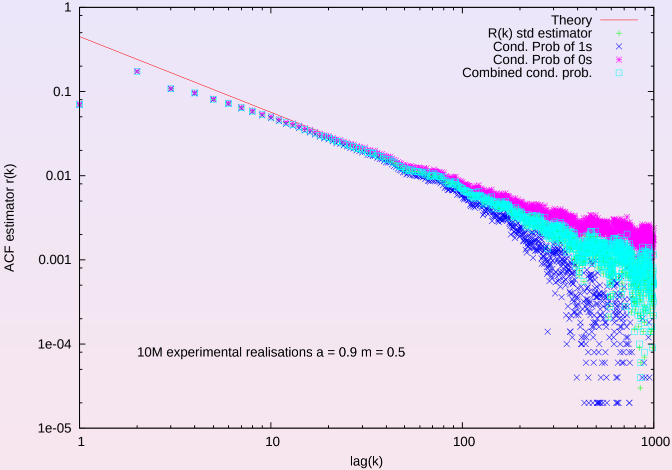
Autocorrelation Function versus lag



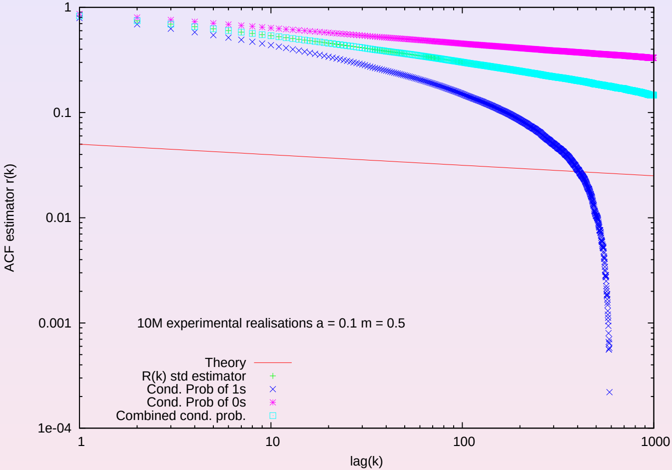
Autocorrelation Function versus lag



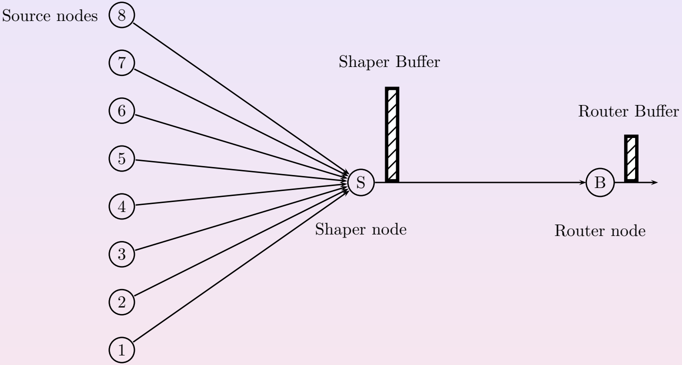
Autocorrelation Function versus lag



Autocorrelation Function versus lag



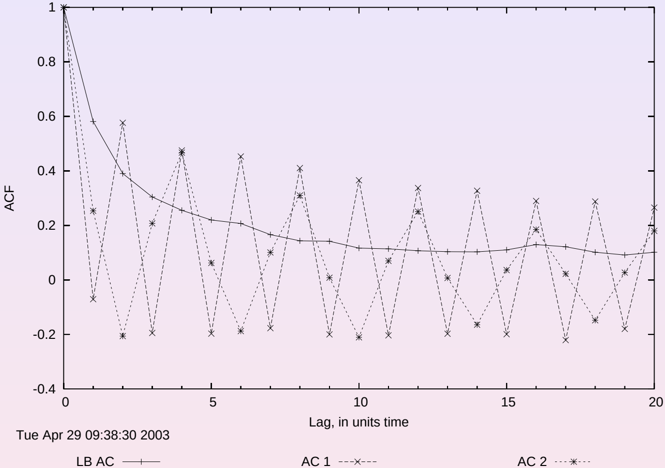
Buffer Topology



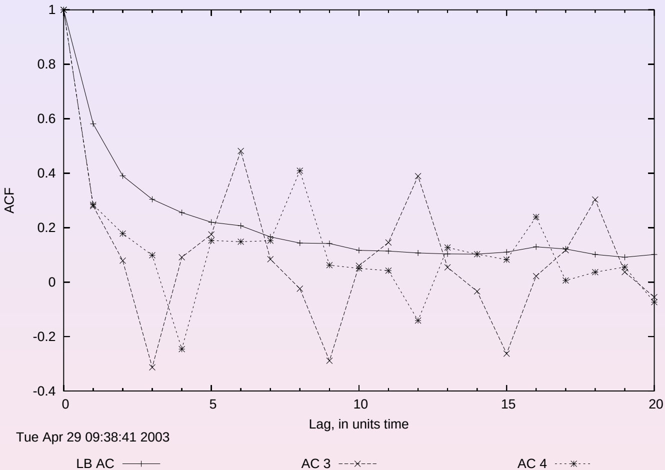
Buffering Algorithm

1. Output traffic at its normal rate for T time periods to get the series Y_T. Let $\hat{k}(Y_T)$ be an estimator for some stat. to reduce.
2. Estimate the three output samples Y_T^H, Y_T^M and Y_T^L assuming that the buffer rate is set to high, medium or low respectively and with the assumption that the buffer input continues as before.
3. Calculate the values of $\hat{k}(Y_T^H)$, $\hat{k}(Y_T^M)$ and $\hat{k}(Y_T^L)$ for three estimated samples.
4. Set the buffer output rate to high, medium or low according to which of the three statistics is lowest.
5. Wait one time step. Move the time series along by one time step so all the T length samples are moved one step to the right. Go to step two.

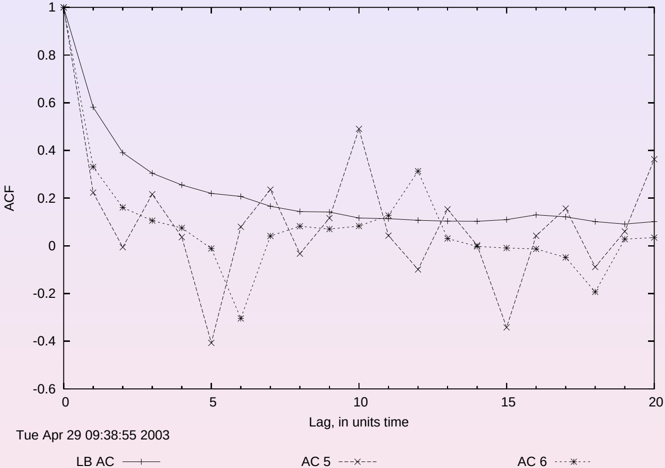
Suppressing ACF at Lags One and Two



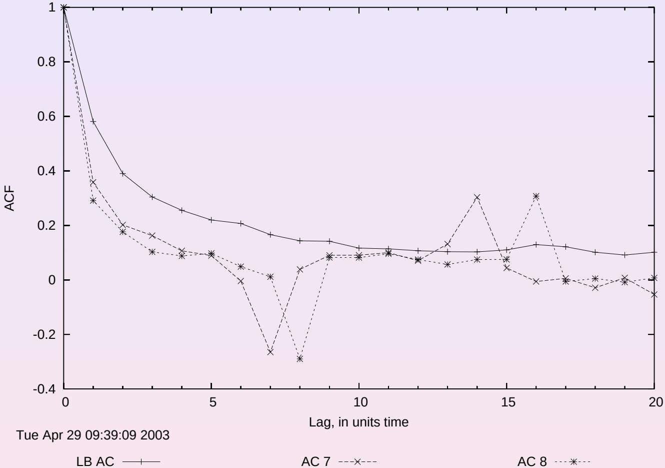
Suppressing ACF at Lags Three and Four



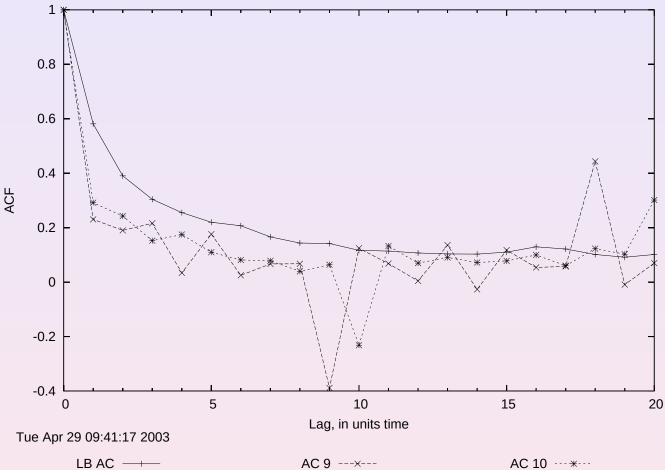
Suppressing ACF at Lags Five and Six



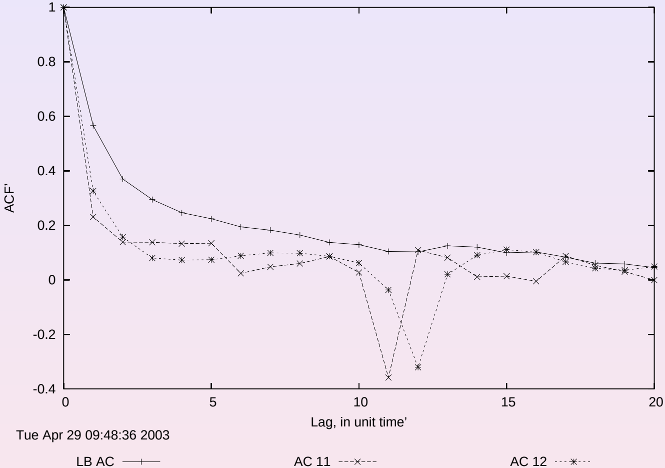
Suppressing ACF at Lags Seven and Eight



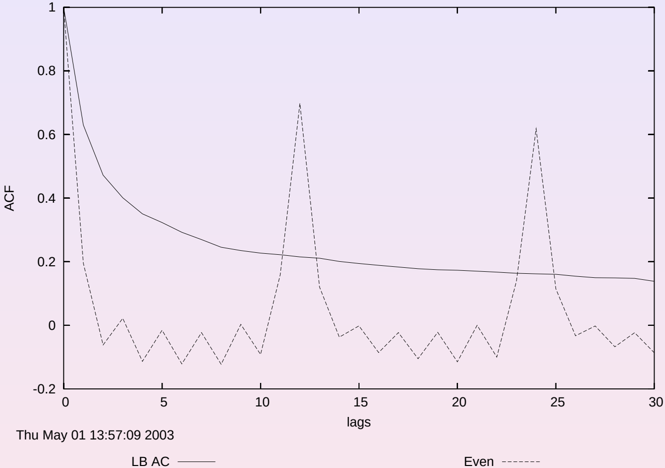
Suppressing ACF at Lags Nine and Ten



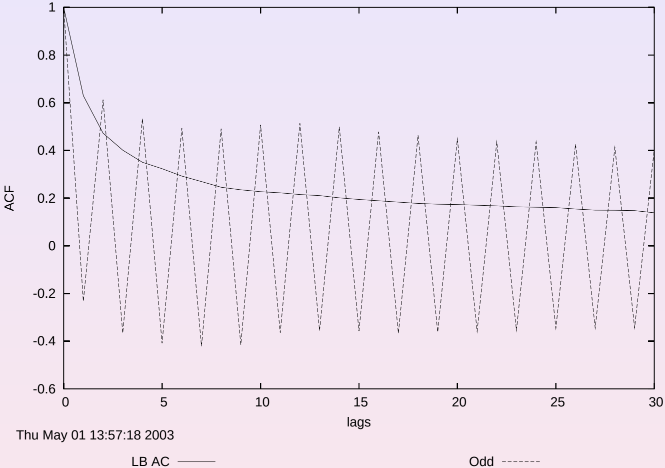
Suppressing ACF at Lags Eleven and Twelve



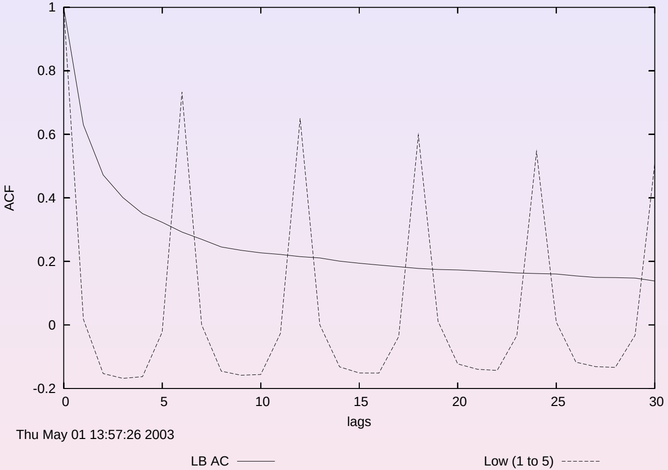
Suppressing ACF at even lags



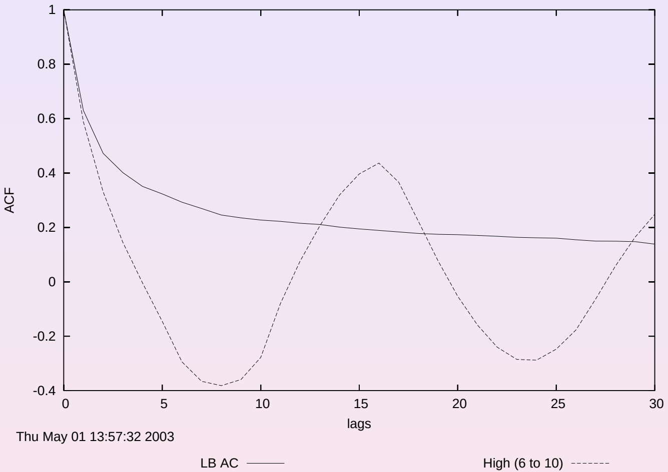
Suppressing ACF at odd lags



Suppressing ACF at Lags One to Five



Suppressing ACF at Lags Six to Ten



Conclusions

- The presented Markov Chain is simple and easy to do mathematics with, computationally fast and produces LRD with a known mean and Hurst parameter.
- While it is not flexible enough to represent real traffic it could be a mathematical lever to gain insight into the situation.
- Estimating the ACF is a very curious thing indeed.
- By careful control of a buffer we may be able to gain control of selected statistical measures on the output traffic.
- What statistical measures do we WANT to control? (LRD is unreachable by this method.)
- How can queuing theory be applied when the queue output rate varies under our control?

Bibliograpy

1) This talk online at www.richardclegg.org/pubs/.
2) Clegg(2004) "The Statistics of Dynamic Networks" PhD Thesis. Dept of Maths. University of York. Online at: www.richardclegg.org/pubs/thesis.pdf
3) X. J. Wang(1989), "Statistical Physics of Temporal Intermittency", Phys. Rev. A, 40 (11), pp6647–6661
4) M. Barenco and D. K Arrowsmith (2004), "The Autocorelation of Double Intermittency Maps...", Dynamical Systems, 19(1), pp61–74.
5) W. Feller (1949), "Fluctuation Theory of Recurrent Events", Trans, Amer. Math. Soc., 67(1), pp94–119