Stationary distribution of CUSUM with autoregressive process

March 20, 2015 | Antoine Tordeux[2] | Forschungszentrum Jülich, Germany

[2]a.tordeux@fz-juelich.de

CUSUM algorithm

We denote $S(X)$ the CUSUM statistic for the time series X

- In the CUSUM algorithm[3], a transition for X is detected if

$$S(X) > \theta(\gamma)$$

with $\theta(\gamma)$ the γ-quantile of S when X is in 'normal' situation ($\gamma \in]0,1[$)

- $\theta(\gamma)$ can be obtained :
  - **Empirically** if the data in normal situation are big enough
  - **By simulation** of X (and S) to approximate the distribution of S
  - **Analytically** by using mathematical tools of stochastic processes

[3] E.S. Page (1954) Biometrika 41:100

CUSUM with autoregressive process

- We assume that X in 'normal' situation is the autoregressive process

$$\begin{aligned} X(0) &= Z(0) \\ X(n) &= c\,X(n-1) + \sqrt{1-c^2}\,Z(n), \quad n \geq 1 \end{aligned} \tag{1}$$

with n the (discrete) time and $(Z(n), n \geq 0)$ independent normal random variables

- The (modified) CUSUM statistic is

$$\begin{aligned} S(0) &= 0 \\ S(n) &= \min\left\{L, \max\left\{0, S(n-1) + g(X(n))\right\}\right\}, \quad n \geq 1 \end{aligned} \tag{2}$$

with

$$g(X(n)) = \begin{cases} 1 & \text{if } |X(n)| > q(\alpha) \\ -1 & \text{otherwise} \end{cases} \tag{3}$$

$q(\alpha)$ being the α-quantile of the Gaussian distribution ($\alpha \in]1/2, 1[$)

JÜLICH FORSCHUNGSZENTRUM

Stationary distribution

The couple $(X(n), S(n))$ is a Markov chain with stationary distribution $\mu(x,s)\,\mathrm{d}x$ ($x \in \mathbb{R}$, $0 \le s \le L$) such that (global balance equation) :

$$
\begin{array}{ll}
\text{If } 0 < s < L \text{ and } |x| > q(\alpha): & \mu(x,s) = \int \mu(y,s-1)\,g(y,x)\,\mathrm{d}y \\
\text{If } 0 < s < L \text{ and } |x| \le q(\alpha): & \mu(x,s) = \int \mu(y,s+1)\,g(y,x)\,\mathrm{d}y \\
\text{If } |x| > q(\alpha): & \mu(x,L) = \int \big(\mu(y,L-1) + \mu(y,L)\big)\,g(y,x)\,\mathrm{d}y \\
 & \mu(x,0) = 0 \\
\text{If } |x| \le q(\alpha): & \mu(x,0) = \int \big(\mu(y,1) + \mu(y,0)\big)\,g(y,x)\,\mathrm{d}y \\
 & \mu(x,L) = 0
\end{array}
\qquad (4)
$$

with $g(y,x) = \frac{1}{\sqrt{2\pi(1-c^2)}} \exp\left(-\frac{(x-cy)^2}{2(1-c^2)}\right)$

The stationary distribution of S is $\mathbb{P}(S = s) = \int \mu(x,s)\,\mathrm{d}x$ for all $0 \le s \le L$

- Unfortunately, this equation doesn't seem to have explicit solution for $\mu(\cdot\,,\cdot)$

Numerical approximation

The system (4) is approximated by using the numerical scheme :

$$\mu(y,s) = \int \mu(x,s')g(x,y)\,\mathrm{d}x \quad \rightsquigarrow \quad a_{i,s} = \sum_k a_{i,s'}\, b_{i,k}$$
$$\text{with } \; a_{i,s} = \mu(x_i,s),\;\; b_{i,k} = \delta_K\, g(x_k,x_i),\;\; x_i = -x' + \delta_K i,\;\; \delta_K = 2x'/K \tag{5}$$

- The numerical approximation yields in the linear equation $M\mathbf{x} = 0$

where $\mathbf{x} = {}^T(a_{0,0}, a_{1,0}, \ldots a_{K,0}, a_{0,1}, \ldots a_{K,L})$ and $M = B - \mathrm{Id}$ with

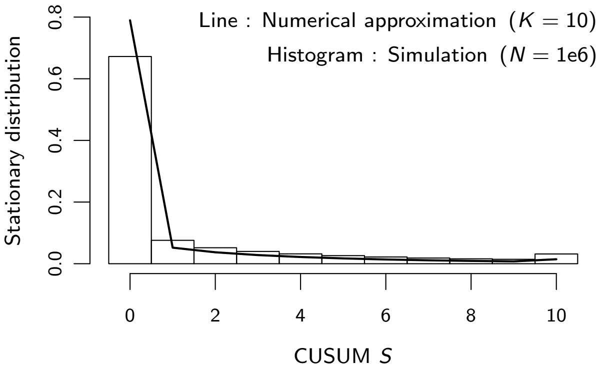
$$B = \begin{bmatrix} B_2 & B_2 & & & \\ B_1 & & B_2 & & \\ & & & & \\ & & B_1 & & B_2 \\ & & & B_1 & B_1 \end{bmatrix} \quad B_1 = \begin{bmatrix} b_{0,0} \ldots & b_{0,K} \\ b_{i_a,0} \ldots & b_{i_a,K} \\ & \\ b_{i_b,0} \ldots & b_{i_b,K} \\ b_{K,0} \ldots & b_{K,K} \end{bmatrix} \quad B_2 = \begin{bmatrix} & \\ b_{i_a+1,0} \ldots & b_{i_a+1,K} \\ b_{i_b-1,0} \ldots & b_{i_b-1,K} \\ & \end{bmatrix}$$

$i_a = \arg\max_i\{x_i < -q(\alpha)\}$ and $i_b = \arg\min_i\{x_i > q(\alpha)\}$

$\rightsquigarrow$ Approximation of the stationary distribution by solving a linear system with $(K+1)(L+1)$ equations

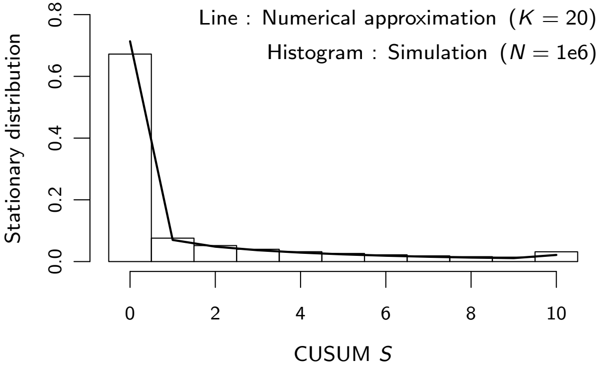
Simulation/Analytic comparison

$L = 10 \quad \alpha = 0.9 \quad c = 0.9$



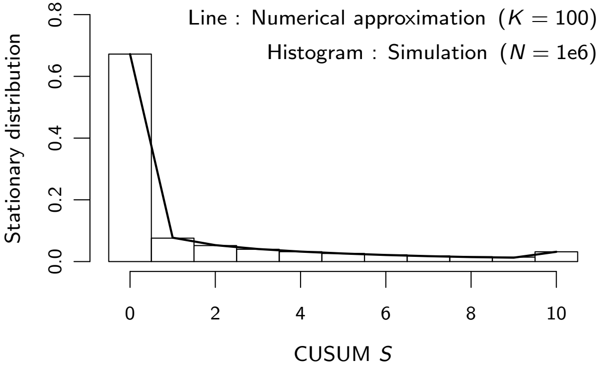
Simulation/Analytic comparison

$L = 10 \quad \alpha = 0.9 \quad c = 0.9$



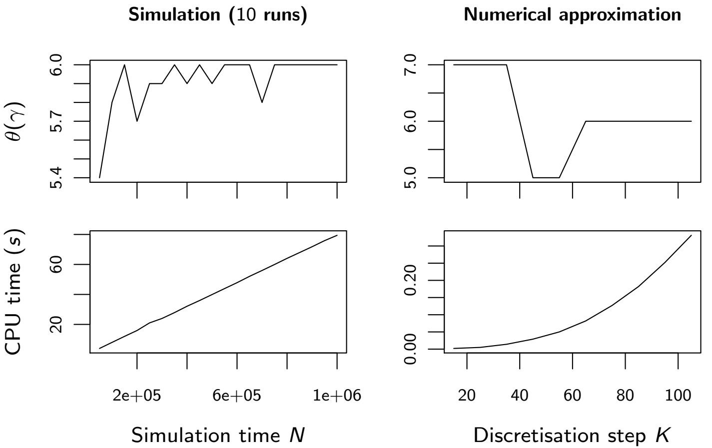
Simulation/Analytic comparison

$L = 10 \quad \alpha = 0.9 \quad c = 0.9$



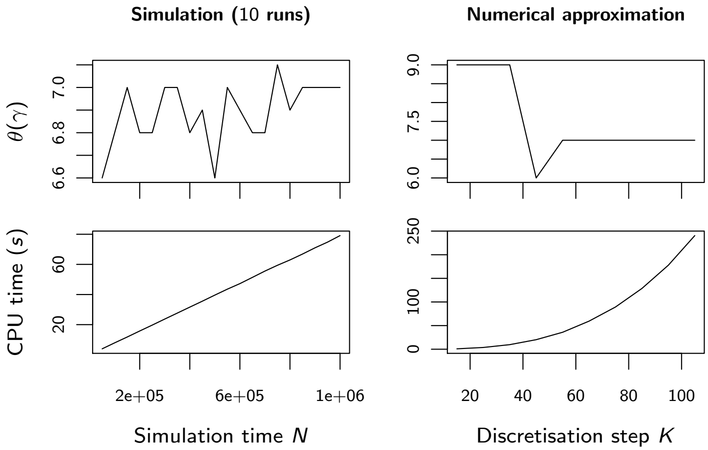
Simulation/Analytic comparison

$L = 10 \quad \alpha = 0.9 \quad c = 0.9 \quad \gamma = 0.9$



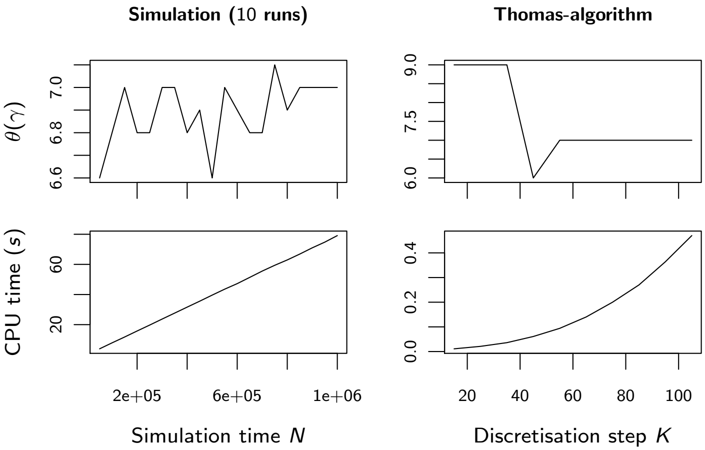
Simulation/Analytic comparison

$L = 100 \quad \alpha = 0.9 \quad c = 0.9 \quad \gamma = 0.9$



Simulation/Analytic comparison

$L = 100 \quad \alpha = 0.9 \quad c = 0.9 \quad \gamma = 0.9$



Conclusion

Threshold θ depends on c, α, L and γ

- Stationary distribution of the Markovian couple (X, S) – and marginal distribution of S – is not explicit

$\rightarrow$ Numerical approximation with discretisation step $K > 1$

$\rightarrow$ Computational effort in $O\big((KL)^3\big)$ – Optimization leads to $O(K^3L)$

- Simulation of X (and S) with random number generator until time N

$\rightarrow$ Random estimation : Realization of $m > 1$ runs to control the precision

$\rightarrow$ Computational effort in $O(mN)$

CCL : No explicit formula for θ / Numerical calculus faster than the simulations
Parameter phase to propose explicit relations

Parameter phase for c

$L = 100 \quad \alpha = 0.99 \quad \gamma = 0.99$