CUSUM Monitoring of INAR(1) Processes of Poisson Counts

Christian H. Weiß

University of Würzburg, Germany

Murat C. Testik

Hacettepe University, Ankara, Turkey

For references in this talk, see

Weiß, C.H., Testik, M.C. (2009).
CUSUM Monitoring of First-Order Integer-Valued Autoregressive Processes of Poisson Counts.
Appears in Journal of Quality Technology.

Weiß, C.H. (2007).
Controlling correlated processes of Poisson counts.
QREI 23(6), 741-754.

Poisson INAR(1) Processes

Definition & Properties

Definition of Poisson INAR(1) process:

Let $(\epsilon_t)_{\mathbb{N}}$ be i.i.d. process with marginal distribution $Po(\lambda(1-\alpha))$, where $\lambda > 0$ and $\alpha \in (0;1)$. Let $N_0 \sim Po(\lambda)$.
If the process $(N_t)_{\mathbb{N}_0}$ satisfies

$$N_t \;=\; \alpha \circ N_{t-1} \;+\; \epsilon_t, \qquad\qquad t \geq 1,$$

plus sufficient independence conditions, then it follows a stationary *Poisson INAR(1) model* with marginal distribution $Po(\lambda)$.

McKenzie (1985), Al-Osh & Alzaid (1987, 1988)

Binomial thinning, due to Steutel & van Harn (1979):

N discrete random variable with range $\{0, \ldots, n\}$ or $\mathbb{N}_0$.

Binomial thinning

$$\alpha \circ N \ := \ \sum_{i=1}^{N} X_i,$$

where X_i are independent Bernoulli trials $\sim B(1, \alpha)$.

Guarantees that right-hand side always integer-valued:

$$N_t \ = \ \alpha \circ N_{t-1} \ + \ \epsilon_t.$$

Interpretation: $\alpha \circ N$ is number of survivors.

Basic properties of Poisson INAR(1) processes:

- Stationary Markov chain with $Po(\lambda)$-marginals and $p_{k|l} \ := \ P(N_t = k \mid N_{t-1} = l) \ = \ \sum_{j=0}^{\min(k,l)} \ \binom{l}{j}\alpha^j(1-\alpha)^{l-j} \cdot e^{-\lambda(1-\alpha)}\frac{(\lambda(1-\alpha))^{k-j}}{(k-j)!}$,
- autocorrelation $\rho(k) \ := \ Corr[N_t, N_{t-k}] \ = \ \alpha^k$.

Estimation from time series $N_1, \ldots, N_T$:

$$\hat{\lambda} \ := \ \frac{1}{T} \cdot \sum_{t=1}^{T} N_t, \qquad \hat{\alpha} \ = \ \frac{\sum_{t=2}^{T}(N_t - \bar{N}_T)(N_{t-1} - \bar{N}_T)}{\sum_{t=1}^{T}(N_t - \bar{N}_T)^2}.$$

Interpretation of INAR(1) process:

$$\underbrace{N_t}_{\text{Population at time } t} = \underbrace{\alpha \circ N_{t-1}}_{\text{Survivors of time } t-1} + \underbrace{\epsilon_t.}_{\text{Immigration}}$$

Interpretation applies well to many real-world problems, e. g.:

- N_t: number of users accessing web server, ϵ_t: number of new users, $\alpha \circ N_{t-1}$: number of previous users still active.
- N_t: number of faults, ϵ_t: number of new faults, $\alpha \circ N_{t-1}$: number of previous faults not rectified yet.

The Poisson INAR(1) model . . .

- is of simple structure,
- essential properties known explicitly,
- is easy to fit to data,
- is easy to interpret,
- applies well to real-world problems, . . .

In a nutshell: A simple model for autocorrelated counts, which is well-suited for SPC!

Controlling Poisson INAR(1) Processes

Control Concepts

Poisson INAR(1) model:

$(N_t)_{\mathbb{N}_0}$ is stationary Poisson INAR(1) process with innovations $(\epsilon_t)_{\mathbb{N}} \sim Po(\lambda(1-\alpha))$. So $N_t \sim Po(\lambda)$.

State of statistical control: $\lambda = \lambda_0$ and $\alpha = \alpha_0$.

Weiß (2007) proposed the following control charts:

- c-Chart for Poisson INAR(1),
- Residual control chart,
- Conditional control chart,
- Moving average control chart.

Simulation study for ARL performance.

Disadvantages of the charts proposed by Weiß (2007):

- Exact ARLs are extremely difficult to obtain
  $\Rightarrow$ design difficult;
- not very effective in detecting small to moderate shifts in process mean λ;
- completely insensitive to an increase in autocorrelation α if process mean λ does not change.

Therefore, . . .

Poisson INAR(1) CUSUM Chart

Definition & Properties

Poisson INAR(1) CUSUM – Definition

One-sided **CUSUM chart** for detecting positive shifts in λ:

$$C_0 = c_0,$$
$$C_t = \max\left(0;\ N_t - k + C_{t-1}\right), \quad t = 1, 2, \ldots$$

$c_0 \geq 0$: starting value, typically $c_0 = 0$.

Fast Initial Response (FIR) feature if $c_0 > 0$.

$k \geq \lambda_0$: reference value.

$h > 0$: upper control limit.

$(N_t)_{\mathbb{N}}$ considered in control unless alarm $C_t \geq h$ triggered.

Poisson INAR(1) CUSUM – Properties

$(N_t)_{\mathbb{N}}$ itself Markov chain $\Rightarrow$ $(C_t)_{\mathbb{N}}$ not Markovian.

But $(N_t, C_t)_{\mathbb{N}}$ Markov chain with transition probabilities

$$\begin{aligned} p(a,b|c,d) &= P(N_t = a, C_t = b \mid N_{t-1} = c, C_{t-1} = d) \\ &= \delta_{b,\max(0;\ a-k+d)} \cdot p_{a|c}, \\ p_1(a,b|c) &= P(N_1 = a, C_1 = b \mid C_0 = c) \\ &= \delta_{b,\max(0;\ a-k+c)} \cdot p_a. \end{aligned}$$

$\Rightarrow$ Adapt **Markov chain approach** of Brook & Evans (1972) for ARL computation.

ARL Computation of One-sided CUSUM Chart:

$\mathcal{I}(h,k)$: Set of reachable in-control values of (N_t, C_t).

Let $\mu_{m,a}$ be expected number of in-control signals before first alarm, given that $(N_1, C_1) = (m,a) \in \mathcal{I}(h,k)$. Define

$$\boldsymbol{\mu} := (\ldots, \mu_{n,i}, \ldots)^\top, \quad \mathbf{Q}^\top := \big(p(n,i \mid m,a)\big)_{(n,i),(m,a)\in\mathcal{I}(h,k)}.$$

Dimension of $\mathbf{Q}$ and $\boldsymbol{\mu}$ equals $|\mathcal{I}(h,k)|$.

Then $\boldsymbol{\mu}$ solution of linear equation $(\mathbf{I}-\mathbf{Q})\cdot\boldsymbol{\mu} = \mathbf{1}$, and

$$\text{ARL}(c_0) \ = \ 1 + \textstyle\sum_{(m,a)\in\mathcal{I}(h,k)} \ \mu_{m,a} \cdot p_1(m,a|c_0).$$

Important issue to speed up ARL computations:

Set $\mathcal{I}$ of reachable in-control values of (N_t, C_t). $|\mathcal{I}|$ determines dimension of matrix $\mathbf{Q}$ for Markov chain approach.

Consider case $h, k, c_0 \in \mathbb{N}_0$. First idea: $\mathcal{I} = \mathbb{N}_0 \times \{0, \ldots, h-1\}$?

But $C_t \geq h$ iff $N_t - k + C_{t-1} \geq h$.

So $N_t \geq k + h$ always leads to alarm.

Considering further restrictions leads to . . .

...

$\mathcal{I}(h,k)$:=

$\{(n,i) \mid 0 \le i \le h-1,\ \max(0;\ i+k-h+1) \le n \le i+k\}$,

which is of size

$$|\mathcal{I}(h,k)| \;=\; \tfrac{1}{2}(h-k)(h+k+1) \;+\; hk.$$

Above arguments can also be applied if h, k, c_0 take values from $\{\frac{r}{s} \mid r \in \mathbb{N}_0\}$, where common denominator $s \in \mathbb{N}$ larger than 1.

Poisson INAR(1) CUSUM Chart

Performance & Design

Poisson INAR(1) CUSUM – Performance

Implementation of MC approach in Matlab.

Tables for in-control ARL_0 about 500,

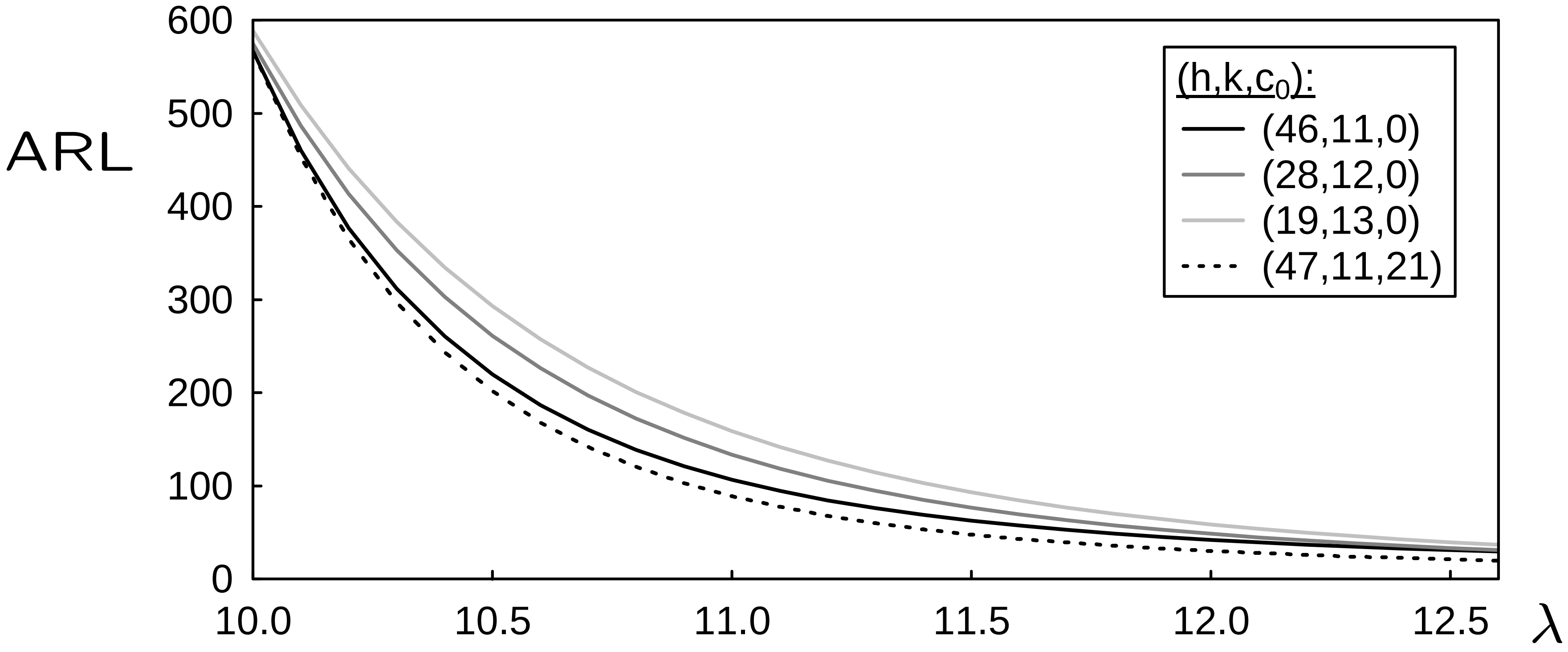
in-control mean values $\lambda_0 = 2.5$, 5, and 10,

in-control dependence values $\alpha_0 =$0.25, 0.50, and 0.75,

with and without FIR.

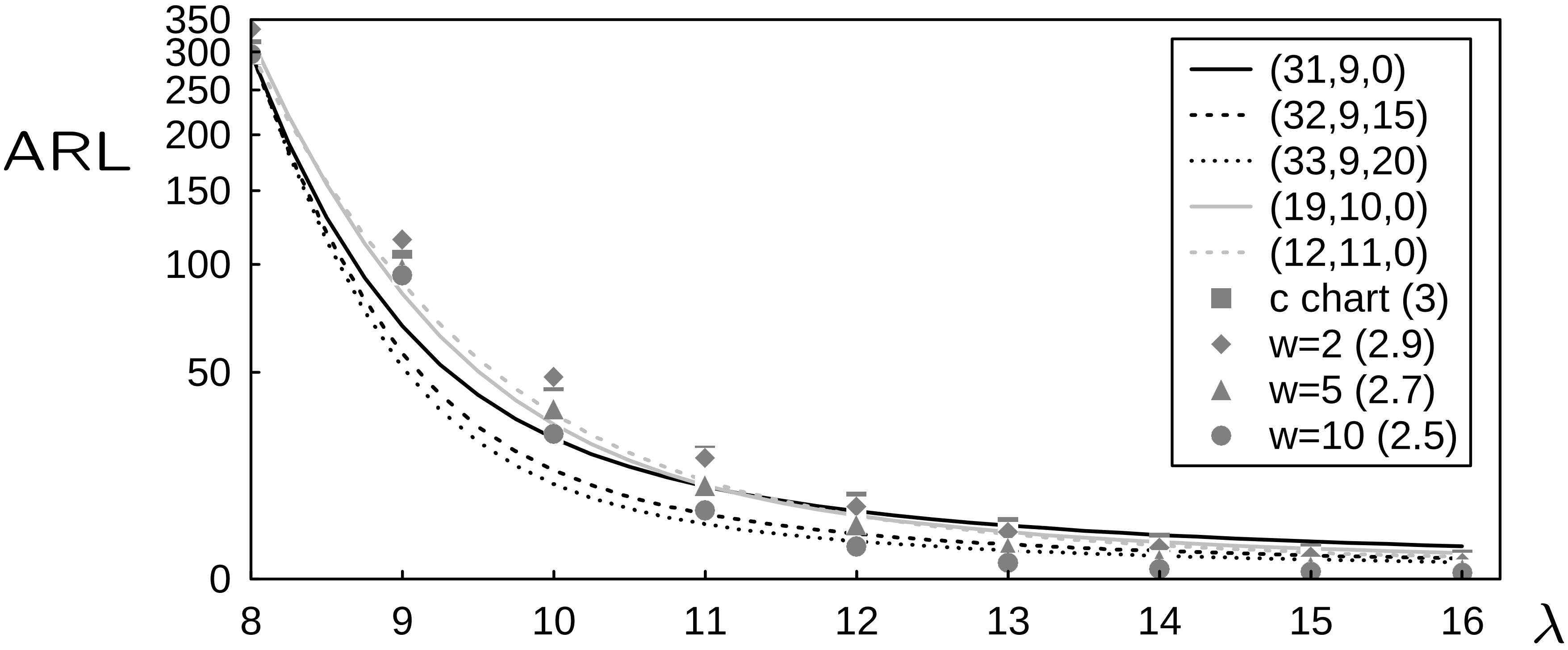
Positive shifts in both λ and α.

ARL performance of CUSUM charts for $(\lambda_0, \alpha_0) = (10, 0.5)$ to detect an increase in λ:

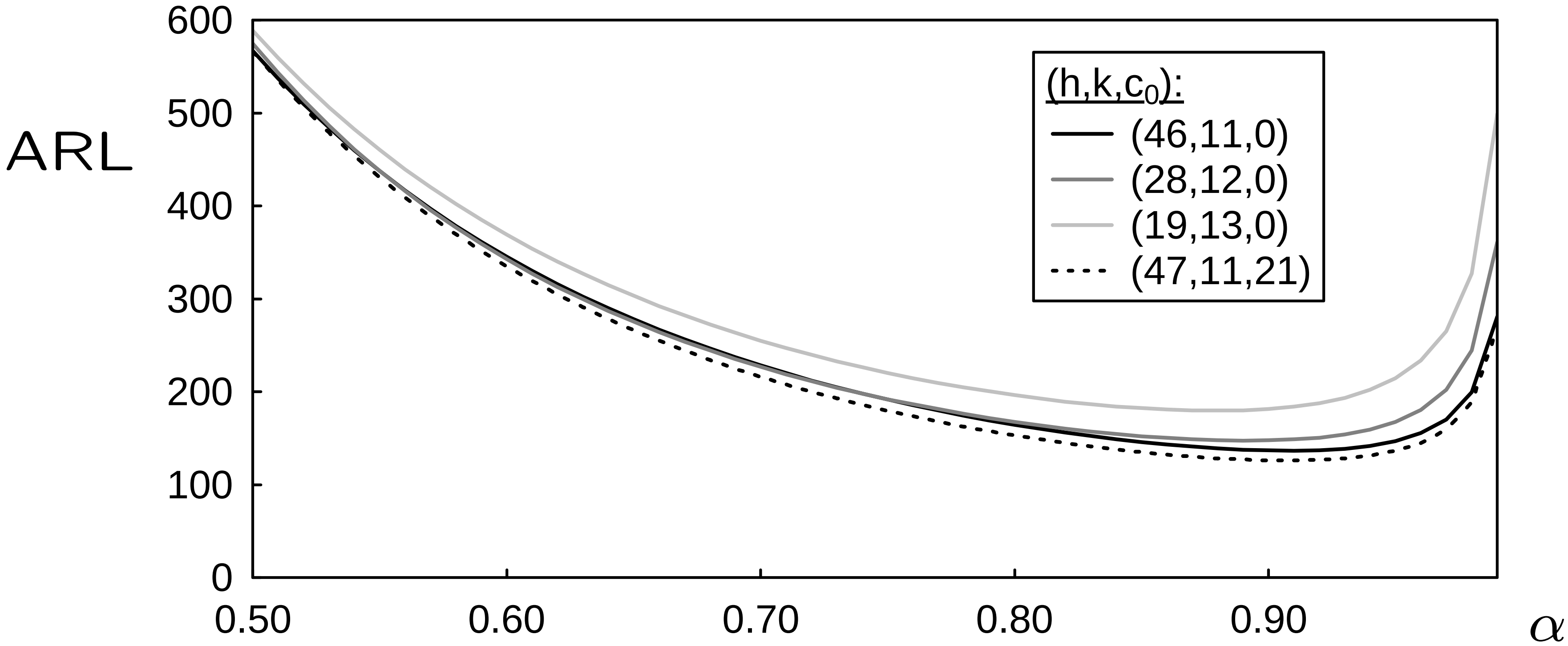


Poisson INAR(1) CUSUM – Performance

ARL performance of c chart, moving average charts with window length w (Weiß, 2007) and CUSUM charts with design triples (h, k, c_0) for $(\lambda_0, \alpha_0) = (8, 0.5)$:



ARL performance of CUSUM charts for $(\lambda_0, \alpha_0) = (10, 0.5)$ to detect an increase in α:



Summary:

CUSUM very effective for small to moderate shifts in λ.

Also sensitive to shifts in α.

In addition:

Better sensitivity than CUSUM based on residuals.

Design recommendations:

Choose k as $\lfloor \lambda_0 + 1 \rfloor$.

Additional FIR feature further improves out-of-control performance.

Poisson INAR(1) CUSUM Chart

Real-Data Example

Weiß (2007): counts of accesses to Statistics web server. Each count represents number of different IP addresses (≈ different users) registered within periods of 2-min length.

IP data between 10 a.m. and 6 p.m. on 29.11.2005: Poisson INAR(1) model with $\lambda = 1.28$ and $\alpha = 0.29$.

$\Rightarrow$ Now 241 counts from 6.12.2005, 10 a.m. to 6 p.m.: In-control model with $\lambda_0 = 1.28$ and $\alpha_0 = 0.29$.

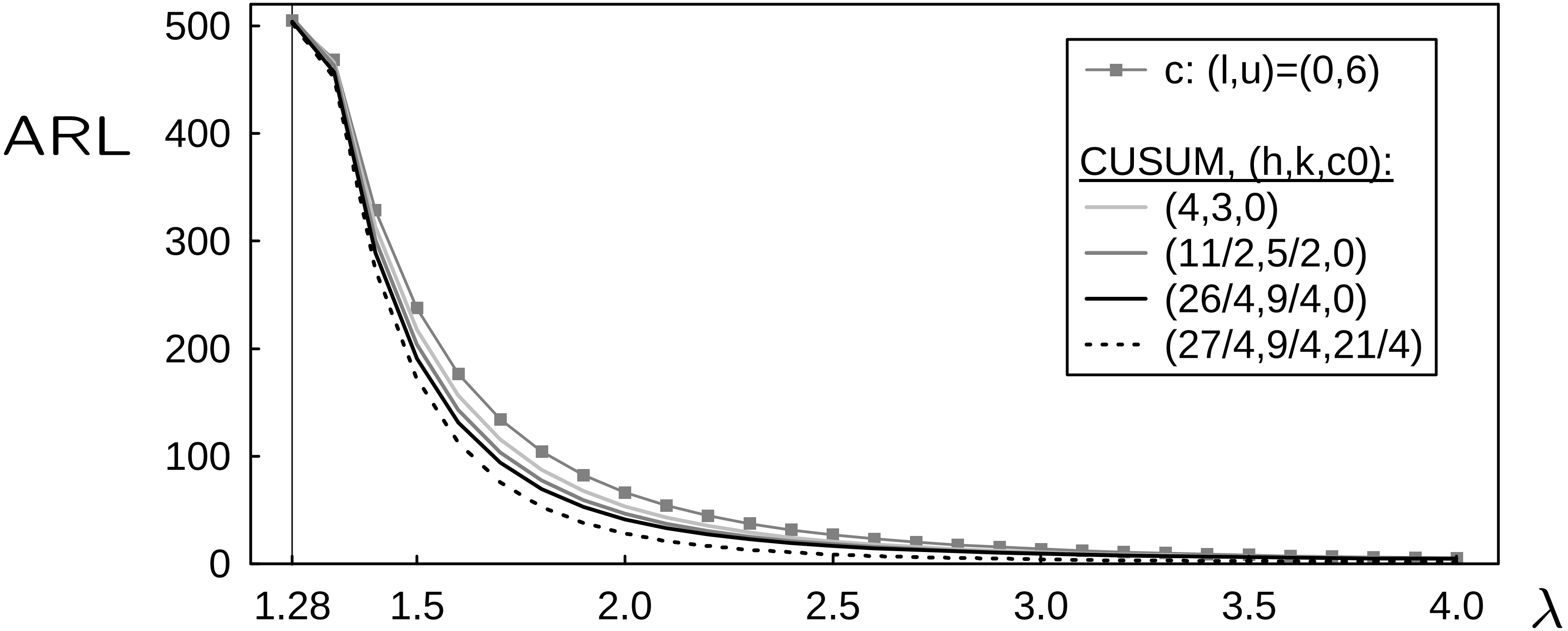
Poisson INAR(1) CUSUM – Example

Considered control charts:

- c-chart, LCL 0 and UCL 6, with $\text{ARL}_0 = 504.949$,
- $(h, k, c_0) = (4, 3, 0)$ with $\text{ARL}_0 = 506.915$,
- $(h, k, c_0) = (\frac{11}{2}, \frac{5}{2}, 0)$ with $\text{ARL}_0 = 507.447$,
- $(h, k, c_0) = (\frac{26}{4}, \frac{9}{4}, 0)$ with $\text{ARL}_0 = 503.867$, and
- $(h, k, c_0) = (\frac{27}{4}, \frac{9}{4}, \frac{21}{4})$ (FIR feature) with $\text{ARL}_0 = 502.586$.

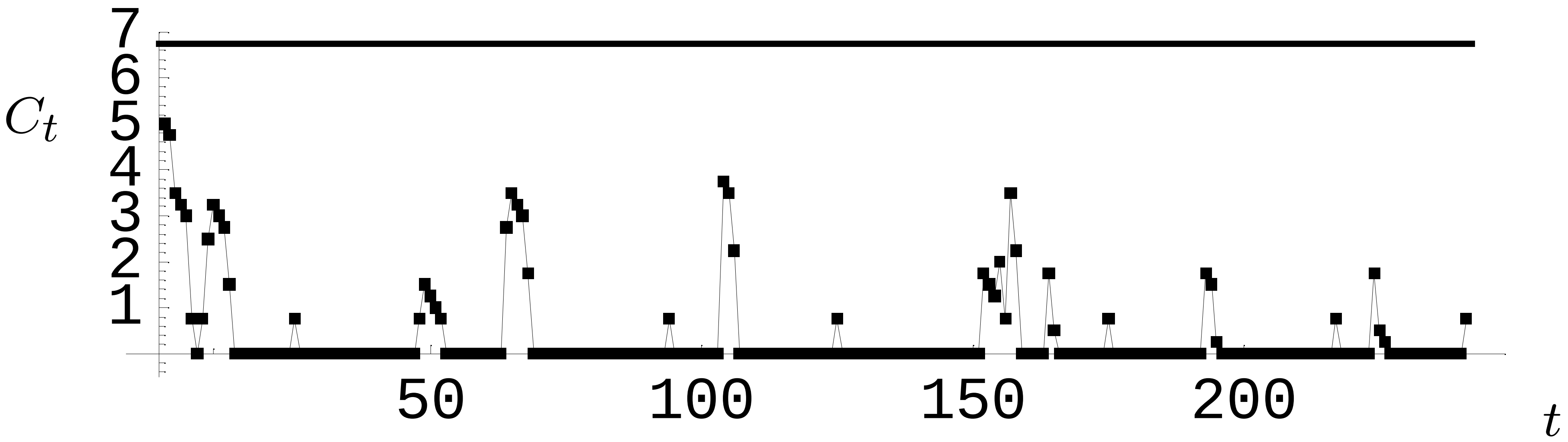
Poisson INAR(1) CUSUM – Example

ARL performance of c and CUSUM charts for $(\lambda_0, \alpha_0) = (1.28, 0.29)$ to detect an increase in λ:



Poisson INAR(1) CUSUM – Example

After removing an outlier, data seems in control. E. g., CUSUM chart with design $(h, k, c_0) = (\frac{27}{4}, \frac{9}{4}, \frac{21}{4})$:

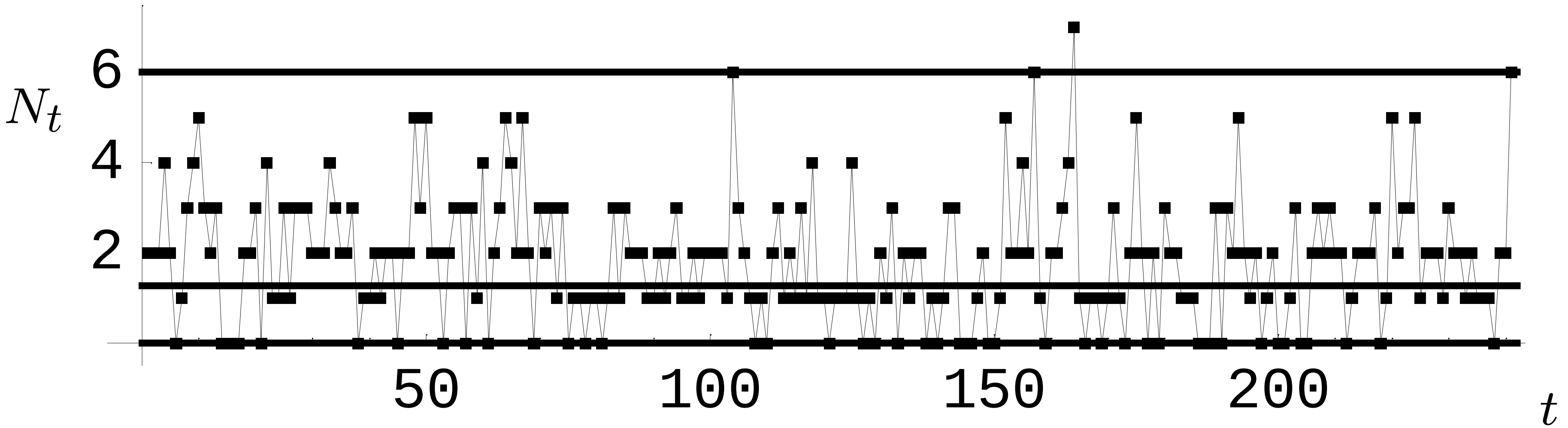


Next, we generated i.i.d. Poisson errors with mean 0.72 and added these counts to the corrected IP data

$\Rightarrow$ shifted IP data, increased Poisson mean 2.

We applied above control charts to the shifted IP data: . . .

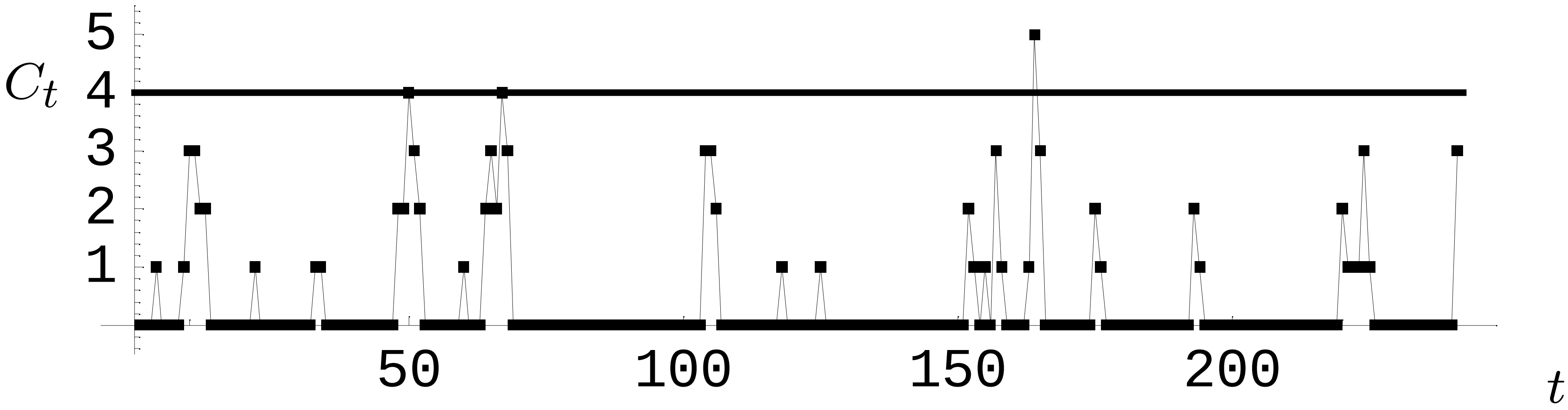
c chart of shifted IP data:



Alarm at time $t = 104$ ($n_{104} = 6$).

Poisson INAR(1) CUSUM – Example

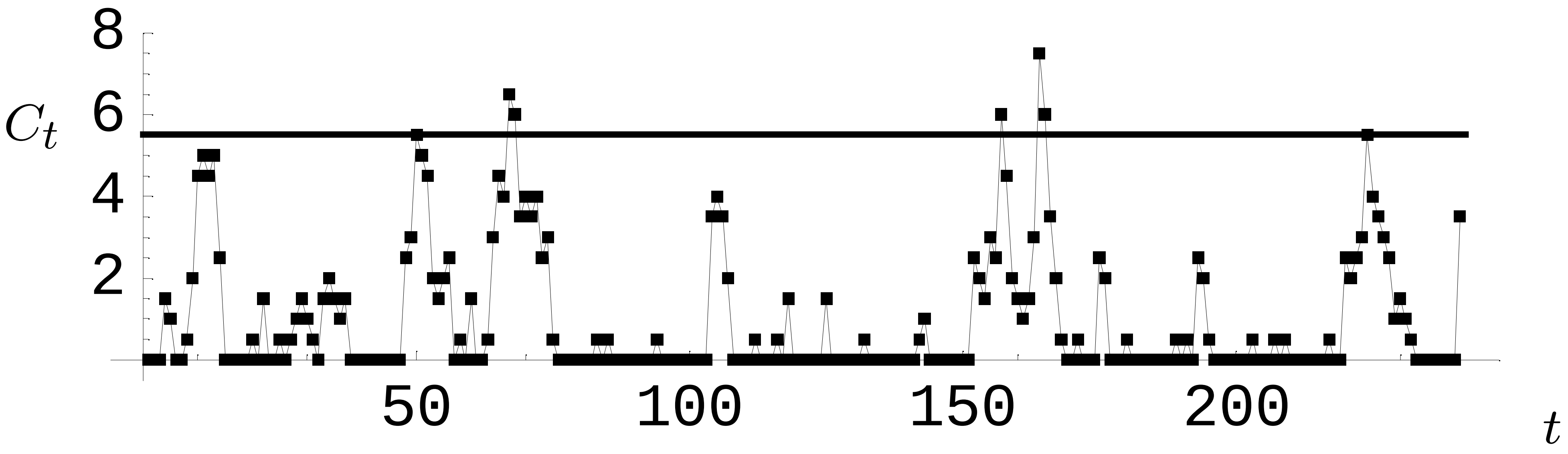
CUSUM chart of shifted IP data with $(h, k, c_0) = (4, 3, 0)$:



Alarm at time $t = 50$ ($c_{50} = 4$).

Poisson INAR(1) CUSUM – Example

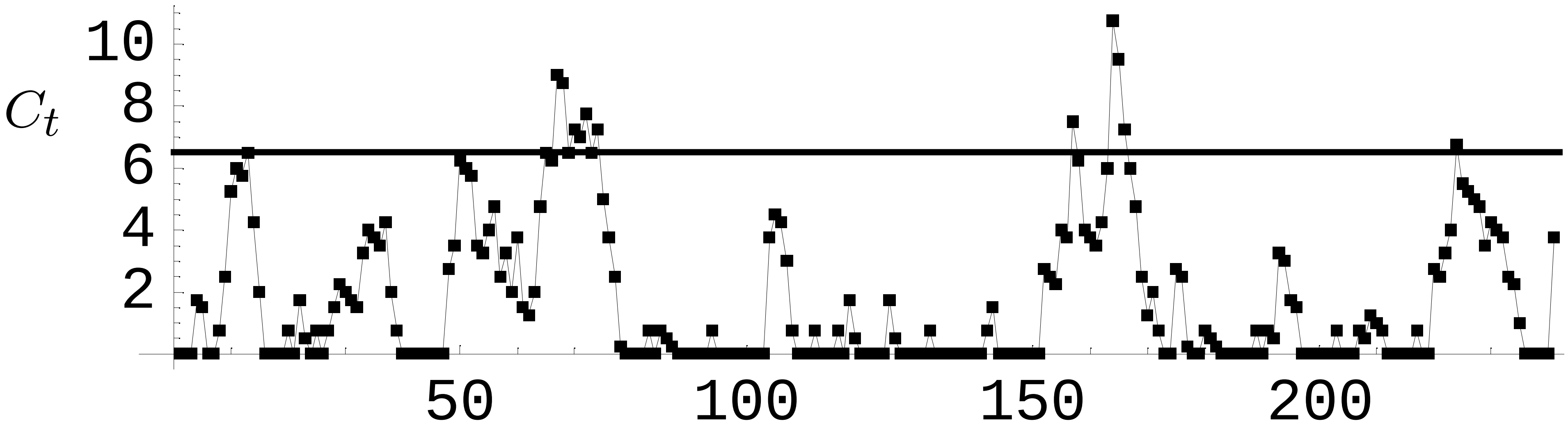
CUSUM chart of shifted IP data with $(h, k, c_0) = (\frac{11}{2}, \frac{5}{2}, 0)$:



Alarm at time $t = 50$ $(c_{50} = \frac{11}{2})$.

Poisson INAR(1) CUSUM – Example

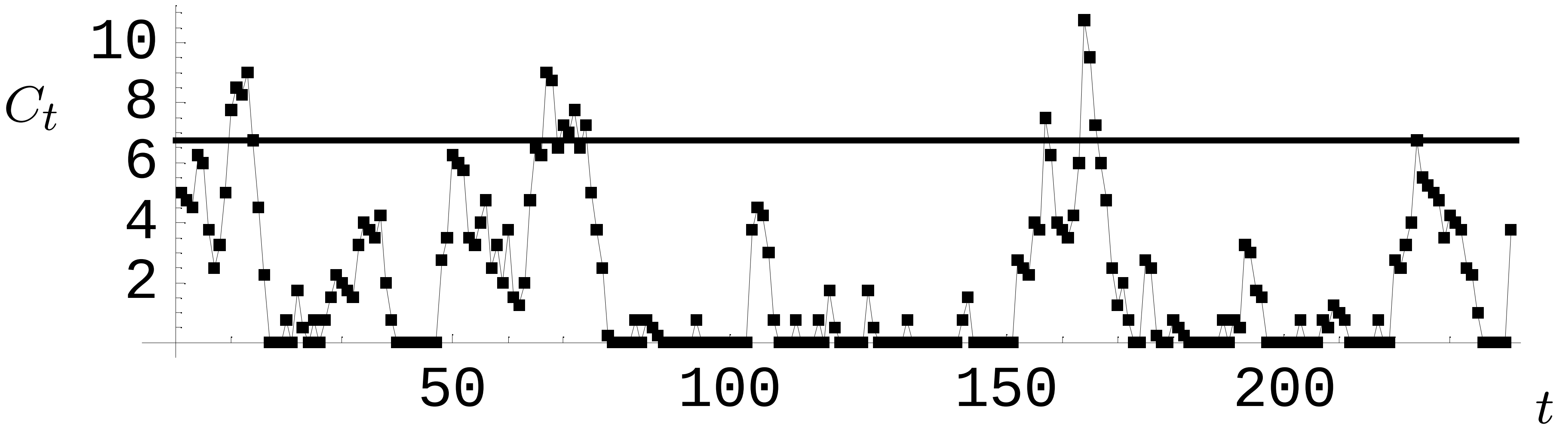
CUSUM chart of shifted IP data with $(h, k, c_0) = (\frac{26}{4}, \frac{9}{4}, 0)$:



Alarm at time $t = 13$ ($c_{13} = \frac{26}{4}$).

Poisson INAR(1) CUSUM – Example

CUSUM chart of shifted IP data with $(h, k, c_0) = (\frac{27}{4}, \frac{9}{4}, \frac{21}{4})$:



Alarm at time $t = 10$ ($c_{10} = \frac{31}{4}$).

- **INAR(1) model:**

  Simple, easily interpretable model, well-suited for real-world problems from SPC.

- **CUSUM Chart:**

  Exact ARL computation with Markov chain approach,
  easy to design (only three design parameters, small k, further improvement through FIR),
  very sensitive to small to moderate shifts in λ,
  sensitive to shifts in α.

Thank You
for Your Interest!

Christian H. Weiß

University of Würzburg

Institute of Mathematics

Department of Statistics