# Residuals-based CUSUM Charts for Poisson INAR(1) Processes

TECHNISCHE UNIVERSITÄT DARMSTADT

**Christian H. Weiß**

Department of Mathematics,

Darmstadt University of Technology

*Fachbereich* **Mathematik**

**Murat Caner Testik**

Industrial Engineering Department,

Hacettepe University Ankara

# INAR(1) Model for Time Series of Counts

Motivation & Properties

Popular for **real-valued** stationary processes:

**AR(1) model** $X_t = \alpha \cdot X_{t-1} + \varepsilon_t$.

**Not** applicable to count data processes $N_1, N_2, \ldots$,

because generally, $\alpha \cdot N \notin \mathbb{N}_0$.

**Idea:** avoid "multiplication problem" by using

**binomial thinning** operator (Steutel & van Harn, 1979):

$$\alpha \circ N := \sum_{i=1}^{N} Y_i, \qquad \text{where } Y_i \text{ are i.i.d. Bin}(1, \alpha),$$

i. e., $\alpha \circ N \sim \text{Bin}(N, \alpha)$ and has range $\{0, \ldots, N\}$.

Let **innovations** $\epsilon_t$'s be i.i.d. with range $\mathbb{N}_0 = \{0, 1, \ldots\}$, let $\alpha \in (0; 1)$.

Process of $N_t$'s referred to as **INAR(1) process** if

$$N_t \;=\; \alpha \circ N_{t-1} \;+\; \epsilon_t \qquad \text{for } t = 1, 2, \ldots,$$

together with appropriate independence assumptions.

(McKenzie, 1985)

**Poisson INAR(1) model**:

$\epsilon_t \;\sim\; Pois\big(\mu(1-\alpha)\big)$ and $N_0 \sim Pois(\mu)$.

# Poisson INAR(1) Processes

- Stationary Markov chain with $Pois(\mu)$-marginals,
- transition probabilities

  $p(k|l) \ := \ P(N_t = k \mid N_{t-1} = l) \ =$
  $\sum_{j=0}^{\min(k,l)} \binom{l}{j} \alpha^j (1-\alpha)^{l-j} \cdot e^{-\mu(1-\alpha)} \frac{(\mu(1-\alpha))^{k-j}}{(k-j)!}$,

- autocorrelation $\rho(k) \ := \ Corr[N_t, N_{t-k}] \ = \ \alpha^k$,
- conditional mean $E[N_t \mid N_{t-1}] \ = \ \alpha \cdot N_{t-1} \ + \ \mu(1-\alpha)$.

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

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

**CUSUM Monitoring** of Poisson INAR(1) processes:

One-sided **CUSUM chart** by Weiß & Testik (2009,2011):

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

Reference value $k \geq \mu_0$, starting value $c_0 := 0$,

control limit $h > 0$ such that certain ARL performance,

ARLs computed with Markov chain approach.

Weiß & Testik (2009): Above CUSUM chart

particularly well-suited to detect mean increases.

**CUSUM Monitoring** of Poisson INAR(1) processes:

But how to detect changes in $\alpha$
or violation of Poissonian equidispersion?

Common approach: derive CUSUM from log-likelihood-ratio,
e. g., see Weiß & Testik (2012) for INARCH(1) model.

But: complex formulae for Poisson INAR(1) model.

$\Rightarrow$ **Idea:** CUSUM monitoring based on specialized residuals!

# Residuals for Poisson INAR(1) Models

Properties & Monitoring

**Specialized residuals** studied by

Freeland & McCabe (2004) and Park & Kim (2012):

$$R_{0;t} \;=\; N_t \;-\; E_0[N_t \mid N_{t-1}] \;=\; N_t - \alpha_0 \cdot N_{t-1} - \mu_0(1-\alpha_0),$$

$$\begin{aligned} R_{1;t} \;&=\; E_0[\alpha \circ N_{t-1} \mid N_t, N_{t-1}, \ldots] \;-\; \alpha_0 \cdot N_{t-1}, \\ R_{2;t} \;&=\; E_0[\epsilon_t \mid N_t, N_{t-1}, \ldots] \;-\; \mu_0(1-\alpha_0). \end{aligned}$$

**Computation of residuals** (Freeland & McCabe, 2004):

$$\begin{aligned} R_{0;t} \;&=\; R_{1;t} \;+\; R_{2;t}, \\ R_{1;t} \;&=\; \left(\frac{p(N_t-1|N_{t-1}-1)}{p(N_t|N_{t-1})} \;-\; 1\right) \cdot \alpha_0 \cdot N_{t-1}, \\ R_{2;t} \;&=\; \left(\frac{p(N_t-1|N_{t-1})}{p(N_t|N_{t-1})} \;-\; 1\right) \cdot \mu_0(1-\alpha_0), \end{aligned}$$

**Properties of residuals**:

$R_{0;t}, R_{1;t}, R_{2;t}$ have mean 0 in in-control case.

$R_{0;t}, R_{1;t}, R_{2;t}$ are functions of $(N_{t-1}, N_t)$.

Now $\ldots, N_{t-1}, N_t$ from general stationary INAR(1) process with true mean $\mu$, true variance $\sigma^2$, true $\alpha$.

For residual $R_{0;t}$, stochastic properties easily derived:

$$\begin{aligned} E[R_{0;t}] &= (\mu - \mu_0)(1 - \alpha_0), \\ V[R_{0;t}] &= \sigma^2\big(1 - \alpha_0^2 \ - \ 2\alpha_0(\alpha - \alpha_0)\big). \end{aligned}$$

Mean of $R_{0;t}$ only affected by change of $\mu$,
while $\sigma^2 > \mu_0$ (overdispersion) increases only variance of $R_{0;t}$.

Simulated mean, standard deviation and correlation for the residuals statistics:

| $\mu_0 = 5$, $\alpha_0 = 0.5$ | in-control | out-of-control $\mu = 7.5 > \mu_0$ | $\alpha = 0.75 > \alpha_0$ | $\delta_o = 1$ |
|---|---|---|---|---|
| $E[R_{0;t}]$ | 0 | 1.25 | 0 | 0 |
| $E[R_{1;t}]$ | 0 | 0.46 | 0.09 | -0.05 |
| $E[R_{2;t}]$ | 0 | 0.79 | -0.09 | 0.05 |
| s.d.$(R_{0;t})$ | 1.94 | 2.37 | 1.58 | 2.50 |
| s.d.$(R_{1;t})$ | 0.69 | 0.91 | 0.56 | 0.81 |
| s.d.$(R_{2;t})$ | 1.31 | 1.55 | 1.06 | 1.82 |
| $Corr[R_{1;t}, R_{2;t}]$ | 0.85 | 0.85 | 0.90 | 0.77 |

Here: $\delta_o := \frac{\sigma_\epsilon^2 - \mu_\epsilon}{\mu_\epsilon}$ (relative overdispersion).

How did we generate the **INAR(1) overdispersion**?

$$\frac{\sigma_N^2}{\mu_N} = \frac{\frac{\sigma_\epsilon^2}{\mu_\epsilon} + \alpha}{1+\alpha},$$

i. e., observ. $N_t$ overdispersed **iff** innov. $\epsilon_t$ overdispersed.

**Idea:** $\epsilon \sim Pois(M)$, where $M$ lognormal $LN(a,b)$.

So $\epsilon$ **Poisson lognormal** $PLN(a,b)$ (Reid, 1981),

with relative overdispersion

$$\delta_{\mathrm{o}} := \frac{\sigma_\epsilon^2 - \mu_\epsilon}{\mu_\epsilon} = \exp\left(a + \frac{b^2}{2}\right) \cdot (e^{b^2} - 1).$$

Upper-sided **CUSUM chart for residuals** $R_{q;t}$ with $q = 0, 1, 2$, for $t = 1, 2, \ldots$:

$$C_{q;0} = c_0, \qquad C_{q;t} \;=\; \max\left(0;\; R_{q;t} - k + C_{q;t-1}\right).$$

Performance via simulations with 1,000,000 replications, **zero-state** in-control ARL metric: $\text{ARL}_0 \;\approx 500$, **steady-state** out-of-control ARL performance.

**Out-of-control shift scenarios:**

increase in mean $\mu$, or in autocorr. $\alpha$, or overdispersion.

**Brief summary** of findings (details in article):

**Shifts in mean $\mu$:**

All charts sensitive, benchmark CUSUM $C_t$ best for small to moderate shifts, residuals CUSUM $C_{2;t}$ best for large shifts.

**Shifts in autocorrelation $\alpha$:**

Residuals CUSUM $C_{1;t}$ overally best, benchmark CUSUM $C_t$ second best, residuals CUSUM $C_{2;t}$ becomes ARL-biased.

**Overdispersion:**

Residuals CUSUM $C_{2;t}$ overally best,
residuals CUSUM $C_{0;t}$ second best.

# Data Example: Emergency Counts

# Data Example: Emergency Counts

Emergency department of children's hospital,
data from February 13, 2009, to August 13, 2009.

**Capacity utilization** of examination room per day
(between 08:00:00 to 23:59:59),
number of patients between call for examination
and first treatment (5-min intervals).

At day $d$: time series $n_{d;1}, n_{d;2}, \ldots, n_{d;192}$ of length 192.

Data from February, 2009, as **Phase I data**,
**in-control model:**
Poisson INAR(1) with $\mu_0 = 2.1$, $\alpha_0 = 0.78$.

All CUSUMs applied to remaining time series.
On each day, any chart started anew.

**Two unusual days for illustration:**

**March 28, 2009:**

$\bar{n} \approx 2.32$, $\hat{\rho}(1) \approx 0.82$ (close to in-control),
but variance-mean ratio increased to 1.31 (overdispersion)
$\Rightarrow$ CUSUM $C_{2;t}$ alarm at 50, other CUSUMs not before 170.

**June 25, 2009:**

increased mean ($\bar{n} \approx 3.47$), underdispersion ($s^2/\bar{n} \approx 0.65$)
$\Rightarrow$ signaled by $C_{1;t}$ ($t = 12$), by $C_t$ ($t = 17$), by $C_{0;t}$ ($t = 20$),
no alarm by $C_{2;t}$.

# Thank You

# for Your Interest!

TECHNISCHE UNIVERSITÄT DARMSTADT

*Fachbereich* **Mathematik**

Christian H. Weiß

Department of Mathematics

Darmstadt University of Technology

weiss@mathematik.tu-darmstadt.de

# Literature

Freeland & McCabe (2004): *Analysis of low count time series data by Poisson autoregression*. J. Time Series Analysis 25(5), 701-722.

McKenzie (1985): *Some simple models for discrete variate time series*. Water Resources Bulletin 21(4), 645-650.

Park & Kim (2012): *Diagnostic checks for integer-valued autoregressive models using expected residuals*. Stat. Papers 53(4), 951-970.

Reid (1981): *The Poisson lognormal distribution and its use as a model . . . .*
In: Statistical Distributions in Scientific Work, Reidel Publishing, 303-316.

Steutel & van Harn (1979): *Discrete analogues of self-decomposability and stability*. Ann. Prob. 7(5), 893-899.

Weiß (2008): *Thinning operations for modelling time series of counts – a survey*. Adv. in Stat. Analysis 92(3), 319-341.

Weiß & Testik (2009): *CUSUM monitoring of first-order integer-valued autoregressive processes of Poisson counts*. J. Quality Technology 41(4), 389-400.

Weiß & Testik (2011): *The Poisson INAR(1) CUSUM chart under overdispersion and estimation error*. IIE Transactions 43(11), 805-818.

Weiß & Testik (2012): *Detection of abrupt changes in count data time series: cumulative sum derivations for INARCH(1) models*. J. Quality Technology 44(3), 249-264.