Detection of Abrupt Changes in Count Data Time Series: Cumulative Sum Derivations for INARCH(1) Models

TECHNISCHE UNIVERSITÄT DARMSTADT

Christian H. Weiß

Department of Mathematics,

Darmstadt University of Technology

Fachbereich **Mathematik**

Murat C. Testik

Industrial Engineering Department,

Hacettepe University, Beytepe-Ankara

This talk is based on the article

Weiß, C.H., Testik, M.C. (2012).
Detection of Abrupt Changes in Count Data Time Series: Cumulative Sum Derivations for INARCH(1) Models.
Journal of Quality Technology **44**(3), 249–264.

All references mentioned in this talk correspond to the references in this article.

INARCH(1) Model for Serially Dependent Counts

Motivation & Properties

INARCH(1) Model for Dependent Counts

Process monitoring often concerned with **attributes data**.

We consider process $\{X_t: \ t = 1, 2, 3, \ldots\}$ of **counts**,

i. e., where each X_t has range $\mathbb{N}_0 = \{0, 1, \ldots\}$.

Examples:

- number of nonconformities in production unit,
- number of service errors in given time interval.

Such data often exhibit serial dependence, e. g.,

Knoth & Schmid (2004): "typical pattern of data"

in medical, environmental, or financial statistics.

Variables data $\rightarrow$ **Gaussian AR(1) model:**

$Z_{t+1} \;=\; \rho \cdot Z_t + \epsilon_t, \qquad$ where (ϵ_t) i.i.d. $\mathrm{N}(\mu_\epsilon, \sigma_\epsilon^2)$.

Not applicable to counts data.

Possible discrete-valued counterpart:

INARCH(1) model, where

$$X_t\Big|_{X_{t-1},X_{t-2},\ldots} \sim\; Pois\big(\beta \;+\; \alpha \cdot X_{t-1}\big),$$

i. e., Markov chain (X_t) with transition probabilities

$$P(X_t = i \mid X_{t-1} = j) \;=\; \exp\left(-\beta - \alpha \cdot j\right) \cdot \frac{(\beta + \alpha \cdot j)^i}{i!}.$$

INARCH(1) model: $X_t \sim Pois\big(\beta + \alpha \cdot X_{t-1}\big)$.

Properties:

- Stationary and ergodic Markov chain (Ferland et al., 2006; Zhu and Wang, 2009);
- all moments exist (Ferland et al., 2006), in particular
- $\mu = \dfrac{\beta}{1-\alpha}, \quad \sigma^2 = \dfrac{\beta}{(1-\alpha)(1-\alpha^2)}$ (overdispersion);
- AR(1)-like autocorrelation function $\rho(k) = \alpha^k$.

INARCH(1) model: $X_t \sim Pois\big(\beta + \alpha \cdot X_{t-1}\big)$.

Parameter estimation:

- Method of moments via $\bar{X}_T$, $\hat{\rho}(1)$.
- Conditional maximum likelihood via

$$L(\beta,\alpha) := P(X_T = x_T, \ldots, X_2 = x_2 \mid X_1 = x_1)$$
$$\sim e^{-(T-1)\beta} \cdot e^{-\alpha \cdot \sum_{t=2}^T x_{t-1}} \cdot \prod_{t=2}^T (\beta + \alpha \cdot x_{t-1})^{x_t}.$$

Note: Simple expression for log-likelihood ratio,

$$\ell R(\beta_0, \alpha_0, \beta_1, \alpha_1) := \ln \tfrac{L(\beta_1,\alpha_1)}{L(\beta_0,\alpha_0)}$$
$$= \sum_{t=2}^T \Big(-(\beta_1 - \beta_0) - (\alpha_1 - \alpha_0) x_{t-1} + x_t \cdot \ln\big(\tfrac{\beta_1 + \alpha_1 x_{t-1}}{\beta_0 + \alpha_0 x_{t-1}}\big)\Big).$$

INARCH(1) Model for Dependent Counts

INARCH(1) model: $X_t \sim Pois\big(\beta + \alpha \cdot X_{t-1}\big)$.

Interpretation:

$$\begin{aligned} X_t &= \textit{total} \text{ number of ``events'' at time } t \\ &= \text{events that newly start at time } t \\ &\quad \textit{plus} \underbrace{\text{events that have started earlier}}_{\substack{\text{also included in } X_{t-1}, \\ \text{``survivors'' from time } t-1}}. \end{aligned}$$

$\alpha \approx$ mean rate of survivors from previous period,

$\beta \approx$ mean number of new events starting at each time.

INARCH(1) model: $X_t \sim Pois\big(\beta + \alpha \cdot X_{t-1}\big)$.

Real-data example: (Weiß, 2010)

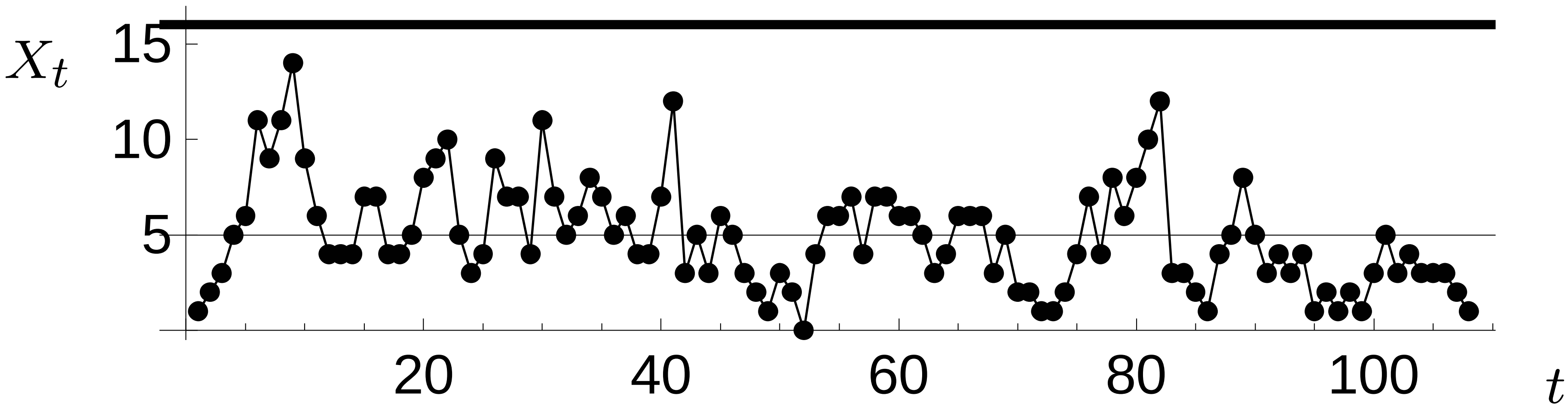
Monthly strike counts, U.S. Bureau of Labor Statistics, Jan. 1994 to Dec. 2002 (108 observations).

x_t = total number of work stoppages leading to $\geq$1000 workers being idle in effect in period t.

Mean 4.94 but variance 7.92 $\Rightarrow$ overdispersed.

AR(1)-like autocorrelation structure with $\hat{\rho}(1) \approx 0.57$.

c chart of **strike count data**:



ML estimates $\hat{\alpha} \approx 0.64$ and $\hat{\beta} \approx 1.81$,

i. e., model mean ≈ 5.0 and model variance ≈ 8.5.

Interpretation: mean number of new strikes is 1.81,

about 64 % of strikes in a month continue in next month.

CUSUM Monitoring of an INARCH(1) Process

Approaches

Situation:

Process $(X_t)_{\mathbb{Z}}$, monitoring starts at time $t = 1$.

Before change point $\tau \in \mathbb{Z}$:

$\ldots, X_{\tau-2}, X_{\tau-1}$ follow *in-control model*,

i. e., INARCH(1) with $\alpha = \alpha_0$, $\beta = \beta_0$.

For $t \geq \tau$: parameters shifted to *out-of-control* values α_1, β_1.

Benchmark chart: conventional (upper-sided) CUSUM

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

where upper limit h adjusted for serial dependence and over-dispersion. (Weiß & Testik, 2009)

CUSUM Charts based on Likelihood Ratio:

Log-likelihood ratio $\ell R(\beta_0, \alpha_0, \beta_1, \alpha_1) \;=\; \sum_{t=1}^{T} \ \ell R_t$, where

$$\ell R_t \;:=\; -(\beta_1 - \beta_0) - (\alpha_1 - \alpha_0)x_{t-1} + x_t \cdot \ln(\tfrac{\beta_1+\alpha_1 x_{t-1}}{\beta_0+\alpha_0 x_{t-1}}),$$

with $\ell R_1 := 0$ (see above).

Given a certain out-of-control parametrization (β_1, α_1), we can derive related CUSUM chart for this purpose.

Plotting always starts with time $t = 2$.

Approach exemplified for three cases: . . .

Case 1: $\beta_1 = (1+\delta)\beta_0$ and $\alpha_1 = (1+\delta)\alpha_0$, i. e., simultaneous change by same relative amount.

Then

$$\ell R_t \;=\; -\delta\cdot(\beta_0+\alpha_0 x_{t-1}) + \ln(1+\delta)\cdot x_t$$

leads to CUSUM

$$C_t^{(1)} \;=\; \max\big(0;\ C_{t-1}^{(1)} - k\cdot(\beta_0+\alpha_0 X_{t-1}) + \ln(1+k)\cdot X_t\big).$$

Note: For $\delta \to 0$, we obtain residuals CUSUM,

$$\ell R_t \;\approx\; \delta\cdot(x_t - \beta_0 - \alpha_0 x_{t-1}).$$

Case 2: $\beta_1 = (1+\delta)\beta_0$ and $\alpha_1 = \alpha_0$, i. e., change only in β.

Then

$$\ell R_t \;=\; -\delta\beta_0 + x_t \cdot \ln(1 + \tfrac{\delta\beta_0}{\beta_0+\alpha_0 x_{t-1}})$$

leads to CUSUM

$$C_t^{(2)} \;=\; \max\big(0;\; C_{t-1}^{(2)} - k\beta_0 + X_t \cdot \ln(1 + \tfrac{k\beta_0}{\beta_0+\alpha_0 X_{t-1}})\big).$$

Case 3: $\beta_1 = \beta_0$ and $\alpha_1 = (1+\delta)\alpha_0$, i. e., change only in α.

Then

$$\ell R_t \;=\; -\delta\alpha_0 x_{t-1} + x_t \cdot \ln(1 + \tfrac{\delta\alpha_0 x_{t-1}}{\beta_0+\alpha_0 x_{t-1}})$$

leads to CUSUM

$$C_t^{(3)} \;=\; \max\big(0;\ C_{t-1}^{(3)} - k\alpha_0 \cdot X_{t-1} + X_t \cdot \ln(1 + \tfrac{k\alpha_0 X_{t-1}}{\beta_0+\alpha_0 X_{t-1}})\big).$$

CUSUM Monitoring of an INARCH(1) Process

Performance

Performance metrics:

zero-state ARL, i. e., $ARL^{(1)} := E[L \mid \tau = 1]$,

steady-state ARL, i. e., $ARL^{(\infty)} := \lim_{\tau\to\infty} ARL^{(\tau)}$,

where $ARL^{(\tau)} := E[L - \tau + 1 \mid L \geq \tau,\ \tau]$ for $\tau \geq 1$.

Chart design always such that $ARL_0^{(1)} \approx 370$.

Out-of-control behavior in terms of $ARL^{(1)}$ and $ARL^{(\infty)}$.

Computation of ARLs:

For benchmark CUSUM exactly via MC approach,

for LR-CUSUMs via simulations using $\widehat{ARL}^{(200)}$

(1 mio. replications; details in Appendix of article).

Performance evaluation: Detailed tables in article.

In-control par.: $\beta_0 = 2$, $\alpha_0 = 0.6$ and $\beta_0 = 3.5$, $\alpha_0 = 0.3$.

Considered δ values 0, 0.025, 0.05, 0.1, 0.25, 0.5, 1, 1.5.

Chart design for LR-CUSUMs with $k = 0.1, 0.25, 0.5, 1.0$.

Besides above out-of-control cases (used for design),

also situation considered, where

β, α shifted by different relative amounts ($\rightarrow$ robustness).

Performance evaluation: Detailed tables in article.

Main findings:

- Benchmark CUSUM is good choice if relative shifts $\geq 50\,\%$ in α or β to be detected (advantage: simple, exact ARL computation).
- Otherwise: LR-CUSUM chart recommended, with k close to relevant shift level δ. Note: Choice of k has large effect on performance.
- Appropriate choice of k with regard to α more important than with regard to β.

Future Work

- **Work in progress:**

  Poisson INAR(1) model: complex likelihood,

  but effective process monitoring

  by considering specialized residuals?

- **Further research issues:**

  Monitoring of higher order or

  non-stationary INGARCH processes.

  Extend INARCH(1) model and related charts

  to situation of varying sample size.

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