MATH STAT

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

SPC Methods for Time-Dependent Processes of Counts

Christian H. Weiß

Department of Mathematics and Statistics, Helmut Schmidt University, Hamburg, Germany; **weissc@hsu-hh.de**.

Introduction

In many fields of application, we are concerned with *count data processes*. Typical examples are counts of defects per produced item in manufacturing industry, counts of new cases of an infection per time unit in health care monitoring, or counts of complaints by customers per time unit in service industry. Often, it is important to detect changes in the process as soon as possible to be able to start preventive actions or to avoid further damages. Methods of *statistical process control (SPC)* are a suitable tool for this purpose.

During the last years, there was increasing interest in SPC methods for *time-dependent processes of counts*. In the sequel, feasible models for autocorrelated counts processes are presented, approaches for corresponding control charts are considered, and process capability indices are discussed. More details and full references can be found in the **open access article**

Weiß CH (2015) SPC methods for time-dependent processes of counts–a literature review. *Cogent Mathematics* 2, 1111116.

Basic models for autocorrelated counts processes

One of the earliest approaches toward stationary count data processes $(X_t)_{\mathbb{N}}$ is the *INAR(1) model* by [McK85], defined by

$$X_t = \alpha \circ X_{t-1} + \epsilon_t,$$

where the innovations $(\epsilon_t)_{\mathbb{N}}$ are i.i.d. counts, and where "$\alpha \circ$" denotes the binomial thinning operator.

Several modifications to the basic INAR(1) model have been proposed, e.g., where the binomial thinning operator is replaced by another type of thinning [SWG15]. As an example, [RBN09] introduced the *NGINAR(1) model*, defined by

$$X_t = \alpha * X_{t-1} + \epsilon_t,$$

where "$\alpha *$" denotes the negative binomial thinning operator.

Another popular approach is the *INARCH(1) model*, where X_t is conditionally Poisson distributed in the following way:

$$X_t | X_{t-1}, \ldots \sim \mathrm{Poi}(\alpha \cdot X_{t-1} + \beta).$$

If systematic trend and seasonality have to incorporated, regression models like the *seasonal log-linear model* [HP08] can be used, with time-dependent mean M_t given by

$$\ln(M_t) = \gamma_0 + \gamma_1 t + \textstyle\sum_{s=1}^{S} (\gamma_{2s} \cos(s\,\omega t) + \gamma_{2s+1} \sin(s\,\omega t)).$$

Common SPC methods

Control charts

A basic approach is to plot the counts $(X_t)_{\mathbb{N}}$ directly on a chart with appropriately chosen control limits: *c chart* (if counts have full range $\mathbb{N}_0$) or *np chart* (if range $\{0, \ldots, n\}$). Applications of the c chart to INAR(1) processes were considered by [W07,W11b], to NGINAR(1) processes by [LWZ], and to INARCH(1) processes by [WT12]. The chart design is usually chosen based on ARL considerations, so it remains to ask how to compute the ARLs for the above types of process. Certainly, ARLs can always be approximated based on simulations with a sufficiently high number of replications. But since the INAR(1), NGINAR(1) and INARCH(1) model constitute a type of discrete Markov model, it is possible to adapt the *Markov chain approach* as first proposed by [BE72]. A detailed description together with corresponding software implementations is provided by [W11b].

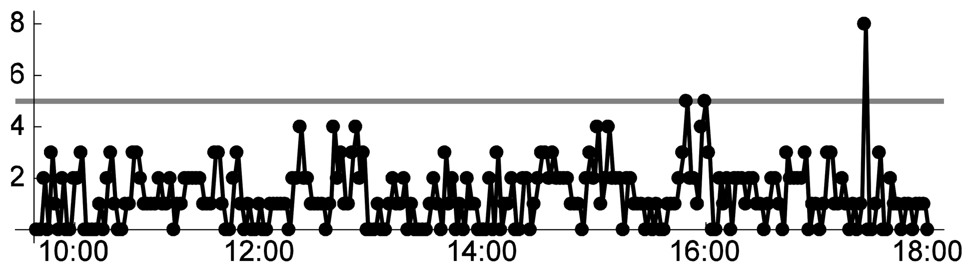

Fig. 1: c chart of users accessing a web server per 2-min interval [W07].

In practice, the true in-control model is hardly known, so one has to fit a model to a set of historic in-control data. Many articles considered the effect of estimated parameters on the charts' performance [JJCW06], where the properties of the used estimators or the sample size play an important role. In the context of autocorrelated count data processes, this topic was considered by, e.g., [ZNHH14,WT15] for the Poisson INAR(1) model and diverse types of control charts.

Process capability indices

Only few of the works about capability indices refer to attributes data processes. [PX05] picked up the idea of considering the actual "proportion of conformance": if the upper specification limit USL describes, e.g., the maximal acceptable number of non-conformities per produced item, then the probability $P(X > USL)$ is compared to a prespecified acceptable probability level $1 - p_0$. [PX05] considered an index defined by the quotient

$$C_{\mathrm{PX}} := \frac{1-p_0}{P(X>USL)} \in [1 - p_0; \infty).$$

A closely related approach was proposed by [BH01]. For practice, a relevant question is how to estimate the index from given in-control data. While [PX05] considered this task for an underlying i.i.d. process of Poisson counts, [W12] extended this work to an underlying Poisson INAR(1) process, distinguishing between the process capability for observations or innovations, respectively.

Advanced control charts

The basic c chart allows for a continuous monitoring of a serially dependent count data process, but the statistic plotted on the chart at time t (= t^{th} observation) does not comprise information about past values of the process (not beyond the mere effect of autocorrelation). Therefore, the c chart (any Shewhart-type chart) is not particularly sensitive to small or moderate changes in the process. For this reason, several types of advanced control charts have been proposed, where the plotted statistic at time t also uses past observations of the process and hence accumulates information about the process for a longer period of time.

CUSUM charts

The traditional *cumulative sum (CUSUM)* control chart [P54], applied directly to the observations X_t, is perhaps the most natural advanced candidate for monitoring autocorrelated processes of counts, because it preserves the discrete nature of the process by only using additions (but no multiplications). The *upper-sided CUSUM* is defined by

$$C_t^+ = \max(0;\ X_t - k^+ + C_{t-1}^+) \quad \text{with} \quad C_0^+ := c_0^+ \geq 0,$$

where an alarm is triggered if $C_t^+ > h^+$ (control limit). While the upper-sided CUSUM is mainly designed to detect increases in the process mean, the *lower-sided CUSUM* aims at uncovering decreases in the mean:

$$C_t^- = \max(0;\ k^- - X_t + C_{t-1}^-) \quad \text{with} \quad C_0^- := c_0^- \geq 0.$$

If (C_t^+, C_t^-) are monitored simultaneously, then this chart combination is referred to as a *two-sided CUSUM chart*.

In the context of monitoring autocorrelated counts processes, the upper-sided CUSUM was applied to INAR(1) processes by [WT09], to NGINAR(1) processes by [LWZ], and to INARCH(1) processes by [WT12]. The lower-sided and the two-sided version were applied to INAR(1) processes by [YWTB13]. For performance evaluation, it is important that the CUSUM preserves the discrete range. Therefore, exact run length computations are possible with a type of MC approach [W11b]: the one-sided CUSUM requires to consider the bivariate Markov chain $(X_t, C_t^{\pm})$ [WT09], the two-sided CUSUM the trivariate Markov chain (X_t, C_t^+, C_t^-) [YWTB13].

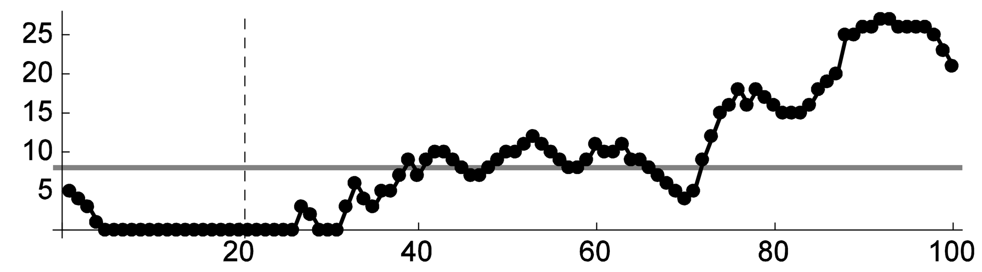

Fig. 2: Upper-sided CUSUM chart with $(h^+, k^+, c_0^+) = (8,2,7)$ applied to simulated INAR(1) data with change point at time 21 (mean shift).

Besides applying the standard CUSUM scheme, one may also look at the *log-likelihood ratio (log-LR)* related to the model for $(X_t)_{\mathbb{N}}$; the statistic plotted at time t is defined as the contribution to the log-LR by the t^{th} observation. For i.i.d. Poisson counts, we obtain again the standard CUSUM chart [HO98]. For a Markov model, this approach will lead to a useful LR-CUSUM scheme as long as the transition probabilities for $(X_t)_{\mathbb{N}}$ are of a feasible form, as exemplified by [WT12] for the case of the INARCH(1) model. The log-LR approach can also be used for non-Markovian types of count data processes. As an example, [HP08] derived such a log-LR CUSUM chart for counts stemming from the seasonal log-linear model, which proved to be useful for the surveillance of epidemic counts.

EWMA charts

Another advanced approach for process monitoring is the *exponentially weighted moving average (EWMA)* control chart dating back to [R59]. The standard EWMA recursion defined by

$$Z_t = \lambda \cdot X_t + (1 - \lambda) \cdot Z_{t-1} \quad \text{with} \quad \lambda \in (0;1],$$

however, has an important drawback if applied to count data processes: it does not preserve the discrete range. Therefore, [G90] suggests to plot rounded values of the EWMA statistic:

$$Q_t = \mathrm{round}(\lambda \cdot X_t + (1 - \lambda) \cdot Q_{t-1}) \quad \text{with} \quad \lambda \in (0;1].$$

A possible disadvantage of the *rounded EWMA approach* was presented in [W11a]: especially for small values of λ, which are generally recommended if small mean shifts are to be detected, one may observe some kind of "oversmoothing", i.e., Q_t becomes piecewise constant in time t and rather insensitive to process changes. Therefore, [W11a] proposed a modification, where a refined rounding operation is used: For $s \in \mathbb{N}$, the operation s-round maps x onto the nearest fraction with denominator s. ARLs can be computed again exactly by adapting the MC approach.

References:

[BH11] Borges W, Ho LL (2001) A fraction defective based capability index. *Qual. Reliab. Engin. Intern.* 17, 447–458.

[BE72] Brook D, Evans DA (1972) An approach to the probability distribution of CUSUM run length. *Biometrika* 59, 539–549.

[G90] Gan FF (1990) Monitoring Poisson observations using modified expon. weighted moving average … *Comm. Stat. Sim. Comp.* 19, 103–124.

[HO98] Hawkins DM, Olwell DH (1998) *Cumulative sum charts and charting for quality improvement*, Springer-Verlag, Inc., New York.

[HP08] Höhle M, Paul M (2008) Count data regression charts for the monitoring of surveillance time series. *Comp. Stat. Data Anal.* 52, 4357–4368.

[JJCW06] Jensen WA, Jones-Farmer LA, Champ CW, Woodall WH (2006) Effects of param. estim. on control chart prop.: a lit. review. *J. Qual. Tech.* 32, 395–409.

[LWZ] Li C, Wang D, Zhu F (in press) Effective control charts for monitoring the NGINAR(1) process. *Qual. Reliab. Engin. Intern.*

[McK85] McKenzie E (1985) Some simple models for discrete variate time series. *Water Resources Bulletin* 21, 645–650.

[P54] Page E (1954) Continuous inspection schemes. *Biometrika* 41, 100–115.

[PX05] Perakis M, Xekalaki E (2005) A process capability index for discrete processes. *J. Stat. Comp. Sim.* 75, 175–187.

[RBN09] Ristić MM, Bakouch HS, Nastić AS (2009) A new geometric first-order integer-valued autoregressive process. *J. Stat. Plann. Infer.* 139, 2218–2226.

[R59] Roberts, S. W. (1959) Control chart tests based on geometric moving averages. *Technometrics* 1, 239–250.

[SWG15] Scotto MG, Weiß CH, Gouveia S (2015) Thinning-based models in the analysis of integer-valued time series: a review. *Stat. Modelling* 15, 590–618.

[W07] Weiß CH (2007) Controlling correlated processes of Poisson counts. *Qual. Reliab. Engin. Intern.* 23, 741–754.

[W11a] Weiß CH (2011a) Detecting mean increases in Poisson INAR(1) processes with EWMA control charts. *J. Appl. Stat.* 38, 383–398.

[W11b] Weiß CH (2011b) The Markov chain approach for perf. eval. of control charts. In Werther (Ed.), *Process control: Problems, …*, Nova Science, 205–228.

[W12] Weiß CH (2012) Process capability analysis for serially dependent processes of Poisson counts. *J. Stat. Comp. Sim.* 82, 383–404.

[WT09] Weiß CH, Testik MC (2009) CUSUM monitoring of first-order integer-valued autoregressive processes of Poisson counts. *J. Qual. Tech.* 41, 389–400.

[WT12] Weiß CH, Testik MC (2012) Detection of abrupt changes in count data t. ser.: Cum. sum derivations for INARCH(1) models. *J. Qual. Tech.* 44, 249–264.

[WT15] Weiß CH, Testik MC (2015) On the phase I analysis for monitoring time-dependent count processes. *IIE Transactions* 47, 294–306.

[YWTB13] Yontay P, Weiß CH, Testik MC, Bayindir ZP (2013) A two-sided CUSUM chart for first-order integer-valued autoregressive … *Qual. Reliab. Engin. Intern.* 29, 33–42.

[ZNHH14] Zhang M, Nie G, He Z, Hou X (2014) The Poisson INAR(1) one-sided EWMA chart with estimated parameters. *Intern. J. Prod. Res.* 52, 5415–5431.