Group Inspection of Dependent Binary Processes.

Christian H. Weiß

University of Würzburg

Institute of Mathematics

Department of Statistics

This talk is based on the paper

Weiß, C.H. (2008). Group Inspection of Dependent Binary Processes. Accepted for publication in QREI, 2008.
See conference CD-ROM.

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

Dependent Binary Processes

Introduction

Statistical control of binary attribute processes $(X_t)_{\mathbb{N}}$ of particular relevance both in research and professional practice.

Typical example: Result of an inspection of an item.
$X_t = 1$ iff the item was *defective* or *nonconforming*.

Many control charts if $(X_t)_{\mathbb{N}}$ is monitored continuously, e. g.:

- run length (RL) charts of Bourke (1991), $(X_t)_{\mathbb{N}}$ serially independent.

- geometric CUSUM procedure of Bourke (1991), $(X_t)_{\mathbb{N}}$ serially independent.

- modified RL charts if $(X_t)_{\mathbb{N}}$ exhibits serial dependence of Markov type, see Blatterman & Champ (1992), Lai et al. (2000), Shepherd et al. (2007).

In practice, often impossible to monitor process continuously:

- Evaluation of items may be too expensive,
- evaluation of items may even destroy items,
- evaluation takes too much time (relative to process speed),
- evaluation complex (e. g., endurance or tolerance tests).

Examples of high-speed processes also outside manufacturing industry:

Lambert & Liu (2006) monitored and controlled a communication network by considering packets handled or dropped, processing errors of various types, and others.

Intrusion detection system described by Ye et al. (2001): controls sequences of counts (per unit time) of method requests or particular outcomes on the system.

In the following, assume that only segments of process $(X_t)_{\mathbb{N}}$ are analyzed, taken at times $t_1, t_2, \ldots$ with $t_k - t_{k-1}$ sufficiently large. Statistic: **segment sum** ${C_k}^{(n)} = X_{t_k} + \ldots + X_{t_k+n-1}$ (count of nonconforming items).

Many control charts if $(X_t)_{\mathbb{N}}$ serially independent, e. g.: p and np chart (Montgomery, 2005), Q chart (Quesenberry, 1991), EWMA chart (Gan, 1990), MA chart (Khoo, 2004), CUSUM chart (Gan, 1993).

Assumption of serial independence often violated in practice!

Assume that time distances $t_k - t_{k-1}$ sufficiently large such that segment sums ${C_k}^{(n)}$ are approximately i.i.d. Within samples, dependence structure of process has to be considered.

$\Rightarrow$ Marginal distribution of $({C_k}^{(n)})_{\mathbb{N}}$ in general not binomial.

Idea:

Use **Markov binomial distribution** $MB(n, p, \rho)$ to approximate true distribution of counts $C_k^{(n)}$.

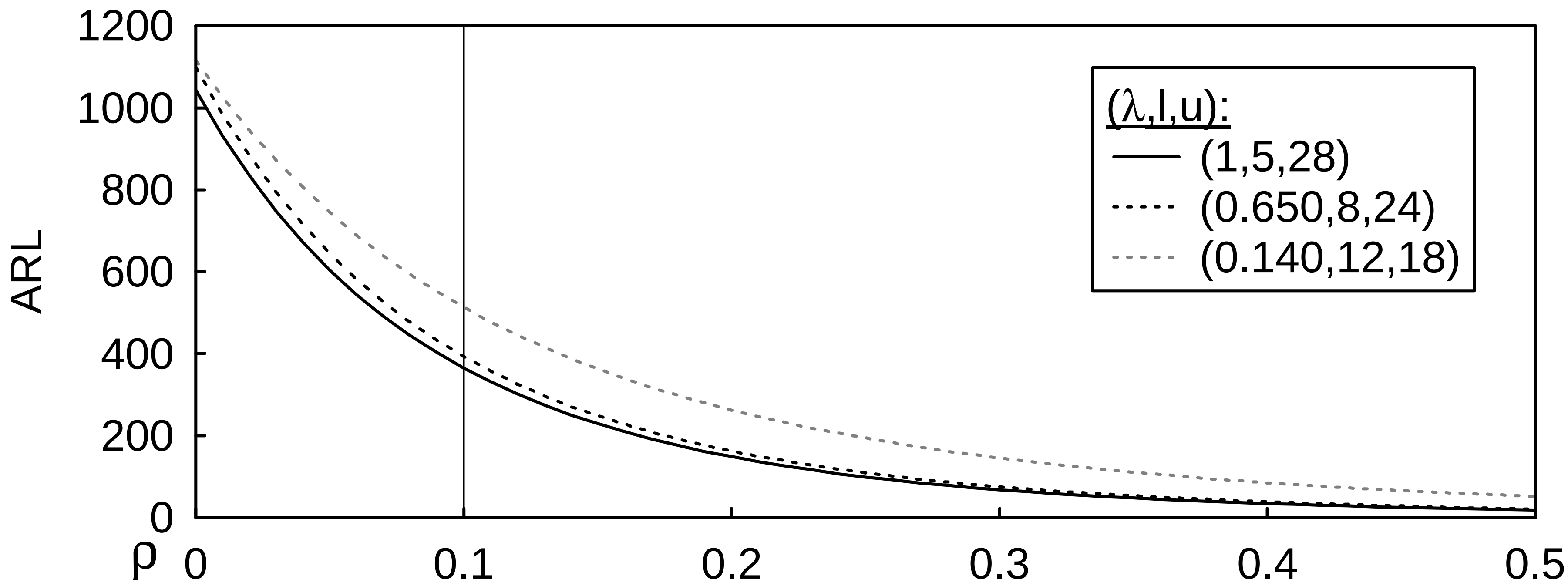
($MB(n, p, 0)$ distribution = binomial distribution $B(n, p)$.)

We consider np **chart** and **EWMA chart** of Gan (1990):

Setting $Z_0 = z_0$, it controls the statistics $(Z_k)_{\mathbb{N}}$ defined by

$$Z_k \;=\; \text{round}\big(\lambda \cdot C_k^{(n)} \;+\; (1-\lambda) \cdot Z_{k-1}\big), \qquad \lambda \in (0; 1].$$

np and EWMA chart very sensitive towards serial dependence, $ARL(\rho)$:
(example of Gan (1990), where $n = 150$, $p = p_0 = 0.1$)



Therefore:

Consider degree of autocorrelation to design chart appropriately:

$\Rightarrow$ **Markov np chart**

$\Rightarrow$ **Markov binomial EWMA chart**

The
Markov np Chart

Design & Performance

Markov np chart with 3-σ limits:

Counts $C_k^{(n)}$ plotted on chart with center line np_0 and control limits

$$\begin{aligned} LCL &= \max\left(0,\ np_0 - 3\sqrt{np_0(1-p_0)\cdot\frac{1+\rho_0}{1-\rho_0}\cdot\left(1-\frac{2\rho_0(1-\rho_0^n)}{n(1-\rho_0^2)}\right)}\right), \\ UCL &= \min\left(n,\ np_0 + 3\sqrt{np_0(1-p_0)\cdot\frac{1+\rho_0}{1-\rho_0}\cdot\left(1-\frac{2\rho_0(1-\rho_0^n)}{n(1-\rho_0^2)}\right)}\right). \end{aligned}$$

ARL **computation:**

$$ARL(n,p,\rho) \;=\; \left(1 \;-\; \textstyle\sum_{j=LCL}^{UCL}\; P(C^{(n)}=j)\right)^{-1},$$

using the $MB(n,p,\rho)$ distribution

In-control performance

Choice of 3-σ limits: ARL_0s vary heavily for varying p_0, ρ_0, often control even impossible ($LCL = 0$ and $UCL = n$).

Recommendations for design (LCL, UCL) such that $ARL \approx 300$ $\rightarrow$ see Weiß (2008).

Out-of-control performance

See below, since the Markov np chart can be considered as a special EWMA chart with $\lambda = 1$.

The
Markov Binomial
EWMA Chart

Design & Performance

The process $(Z_k)_{\mathbb{N}_0}$, defined by

$$Z_k \;=\; \text{round}\big(\lambda \cdot C_k^{(n)} \;+\; (1-\lambda)\cdot Z_{k-1}\big), \qquad \lambda \in (0;1],$$

is a homogeneous Markov chain with range $\{0,\ldots,n\}$ and transition probabilities

$$p(x|y) \;:=\; P(Z_k = x \mid Z_{k-1} = y)$$
$$=\; P\big(\tfrac{1}{\lambda}(x-0.5-(1-\lambda)y) \leq C_k^{(n)} < \tfrac{1}{\lambda}(x+0.5-(1-\lambda)y)\big).$$

$\Rightarrow$ ARLs can be computed exactly by adapting the Markov chain approach of Brook & Evans (1972).

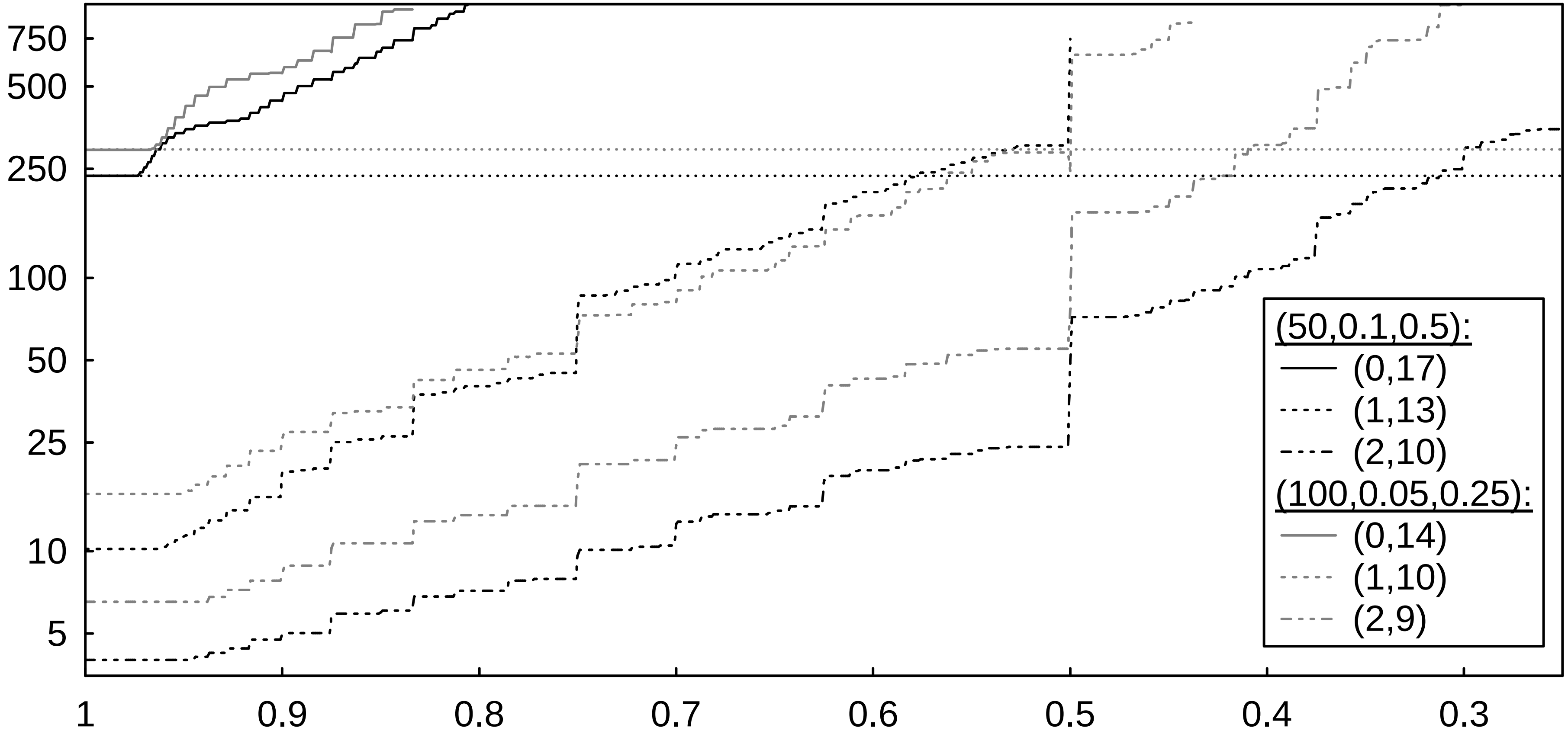
Using this approach, we recommend the following

Design strategy:

1. Start with Markov np chart (i. e.: $\lambda = 1$) and with sufficiently small ARL_0.

2. Once limits (l, u) have been fixed, then decreasing λ increases ARL_0.
   If desired ARL_0 reached, the design (λ, l, u) is complete.

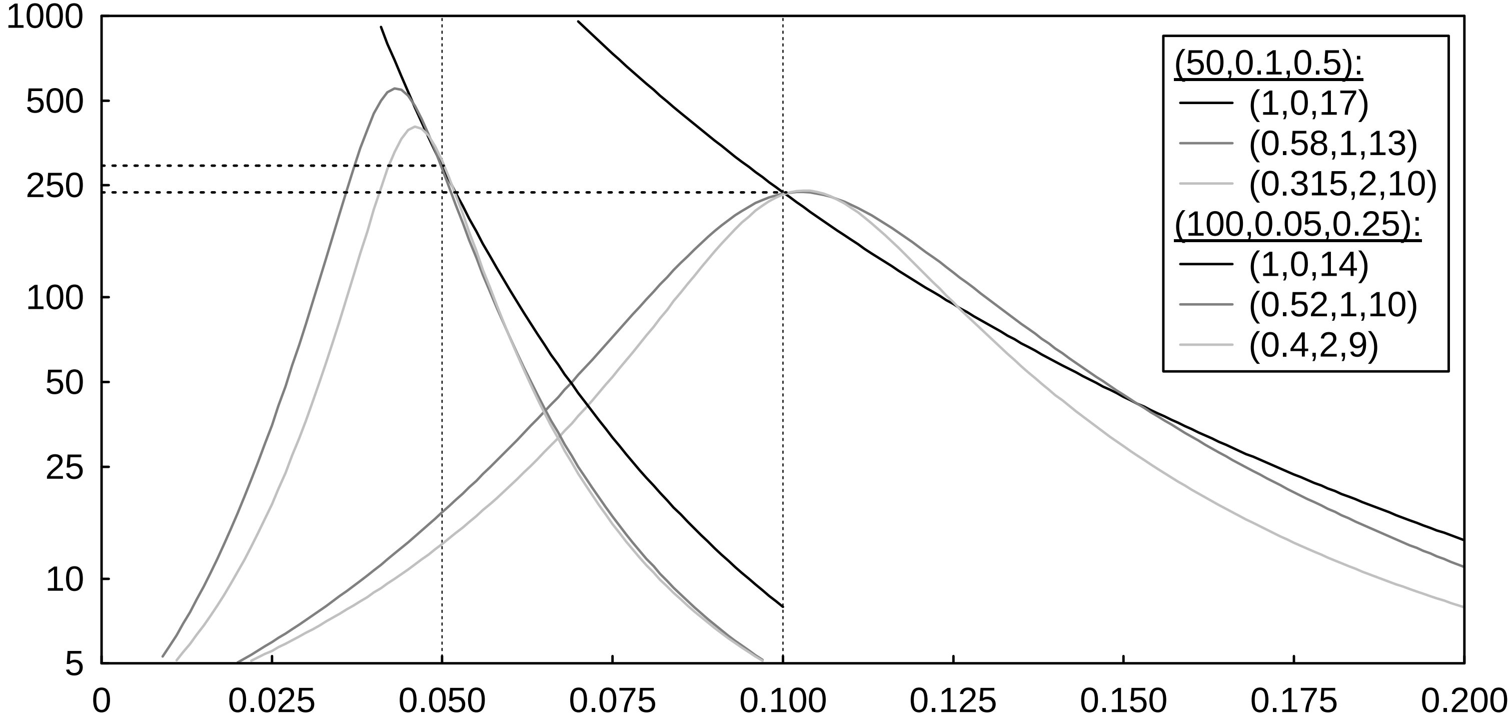
Markov EWMA Charts: Design & Performance



Out-of-control performance:

$(n, p_0, \rho_0) = (50, 0.1, 0.5)$ and $(100, 0.05, 0.25)$,

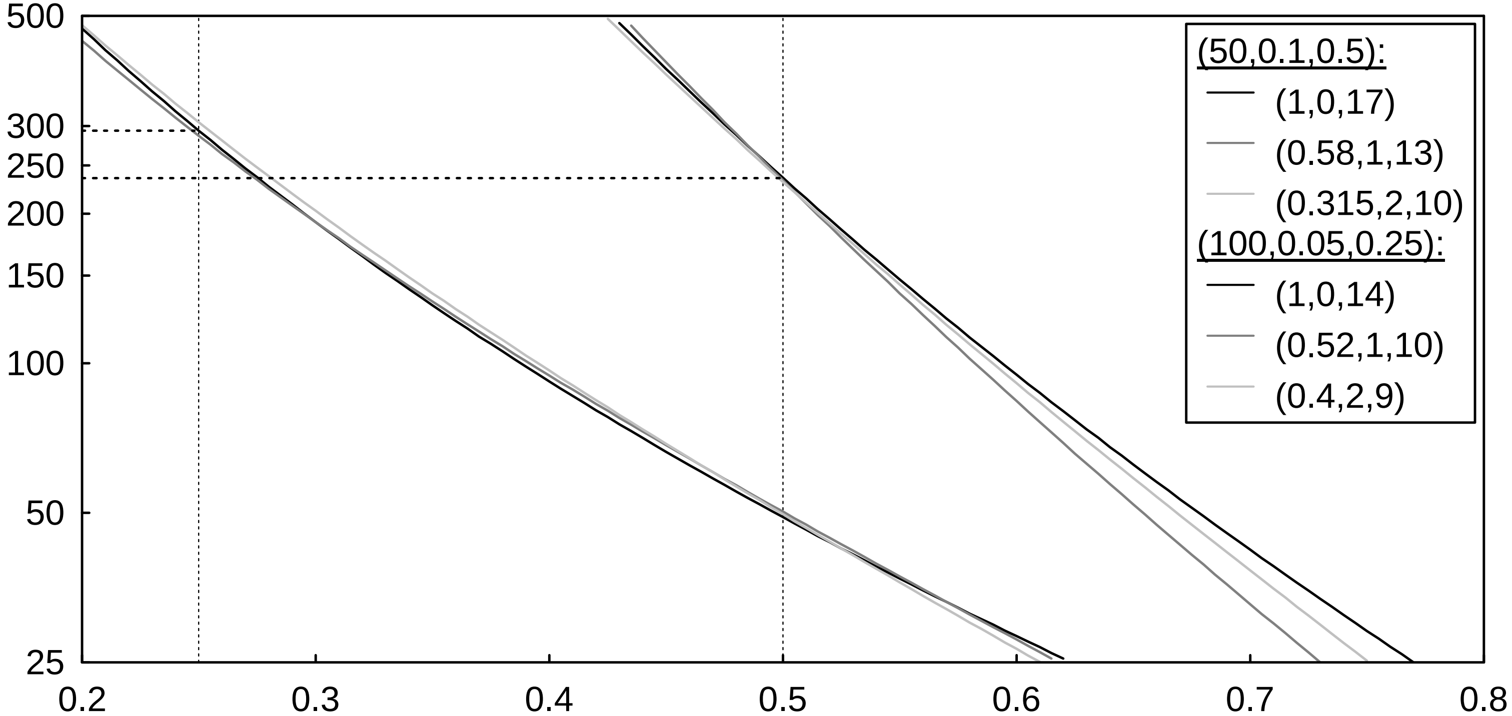
ARL_0 of about 236 and 294.



Out-of-control performance: (continued)

$(n, p_0, \rho_0) = (50, 0.1, 0.5)$ and $(100, 0.05, 0.25)$,

ARL_0 of about 236 and 294.



The
Markov Binomial
EWMA Chart

Real-data Example

High-speed attribute process: communication networks.

Statistics server accessed with high frequency, great number of different features per access $\Rightarrow$ log files from October to December 2005 required disk space $\approx$ 99.3 mb.

Complete information cannot be monitored in real-time $\Rightarrow$ confine oneself to few features measured within distant time segments.

Example: Data collected between October 17th and December 31th, 2005.

Accesses to home directory of particular member of Department of Statistics: $X_t = 1$ iff access at minute t. Segments between 9:00 and 9:59 a.m. and between 6:00 and 6:59 p.m.

$\Rightarrow$ segment sums ${C_k}^{(60)} = X_{t_k} + \ldots + X_{t_k+59}$ with $k = 1, \ldots, 110$, where each ${C_k}^{(60)}$ has range $\{0, \ldots, 60\}$.

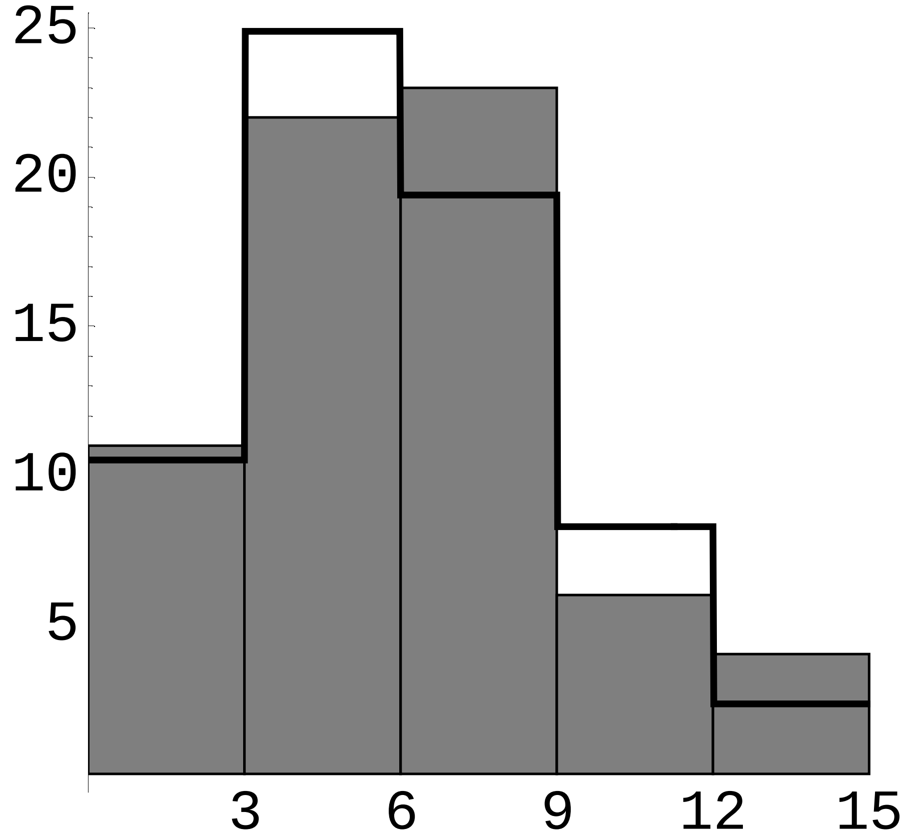
First, consider $C_1^{(60)}, \ldots, C_{66}^{(60)}$ (October 17$^{\text{th}}$ to November 30$^{\text{th}}$, 2005) $\Rightarrow$ identify in-control model!

Empirical mean and variance: 5.56061 and 9.69627

Fit Markov binomial distribution $MB(60, p, \rho)$ with $\widehat{p} \approx 0.0926768$ and $\widehat{\rho} \approx 0.320894$

Histogram with $MB(60, 0.093, 0.32)$ distribution.

Pearson's χ^2 test: p-value ≈ 0.59.



$\Rightarrow$ *In-control model*: $C_k^{(60)}$ i.i.d. $MB(60, 0.093, 0.32)$.

Markov EWMA Charts: Real-data Example

Control charts, first applied in phase I to $C_1{}^{(60)}, \ldots, C_{66}{}^{(60)}$, then in phase II to $C_{67}{}^{(60)}, \ldots, C_{110}{}^{(60)}$.

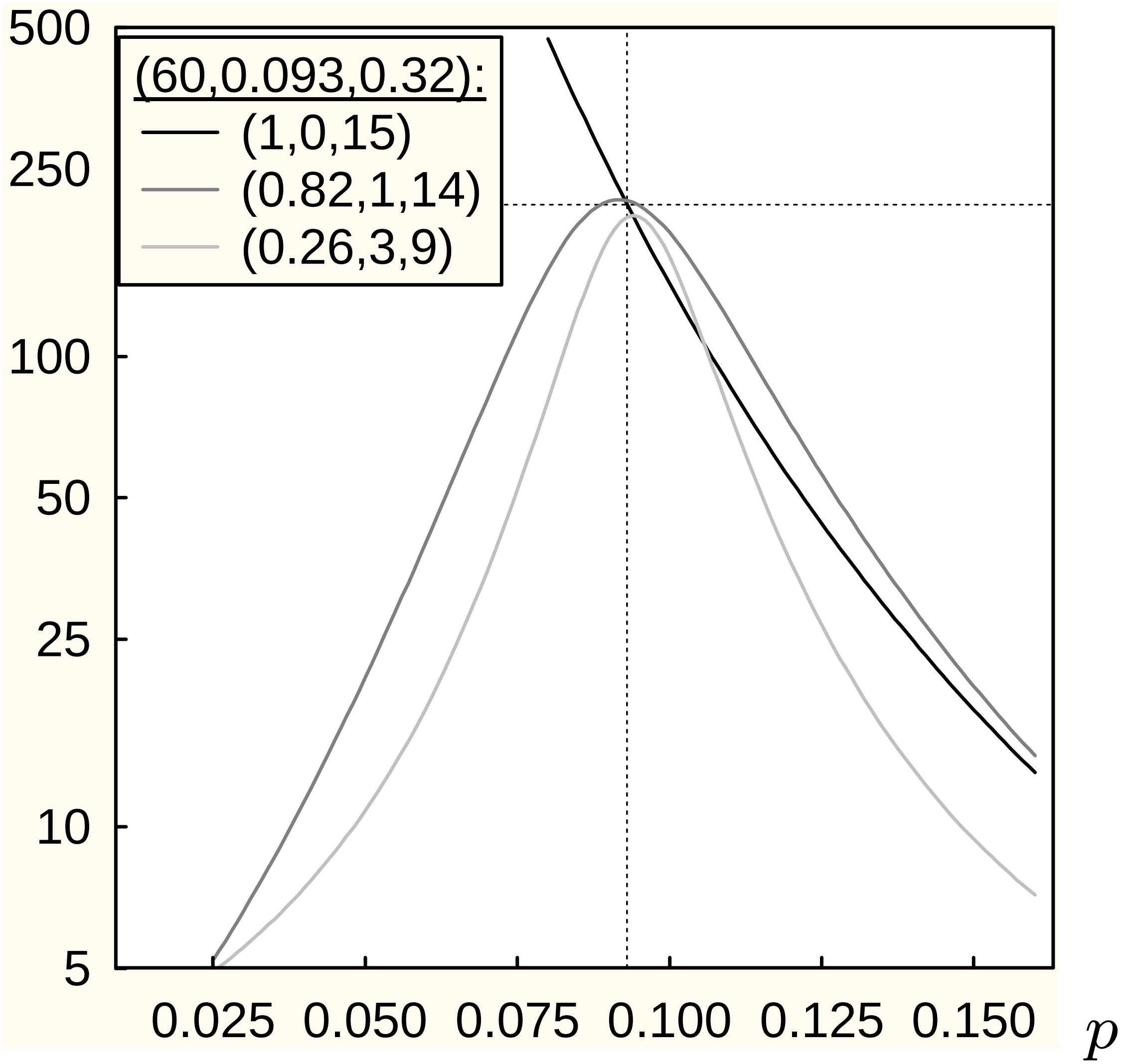
- One-sided Markov np chart with $(\lambda, l, u) = (1, 0, 15)$;
  $\rightarrow$ $ARL_0 \approx 210.634$
- Markov EWMA chart with $(\lambda, l, u) = (0.82, 1, 14)$;
  $\rightarrow$ $ARL_0 \approx 214.358$
- Markov EWMA chart with $(\lambda, l, u) = (0.26, 3, 9)$.
  $\rightarrow$ $ARL_0 \approx 197.392$

Markov EWMA Charts: Real-data Example

np chart biased,

EWMA charts

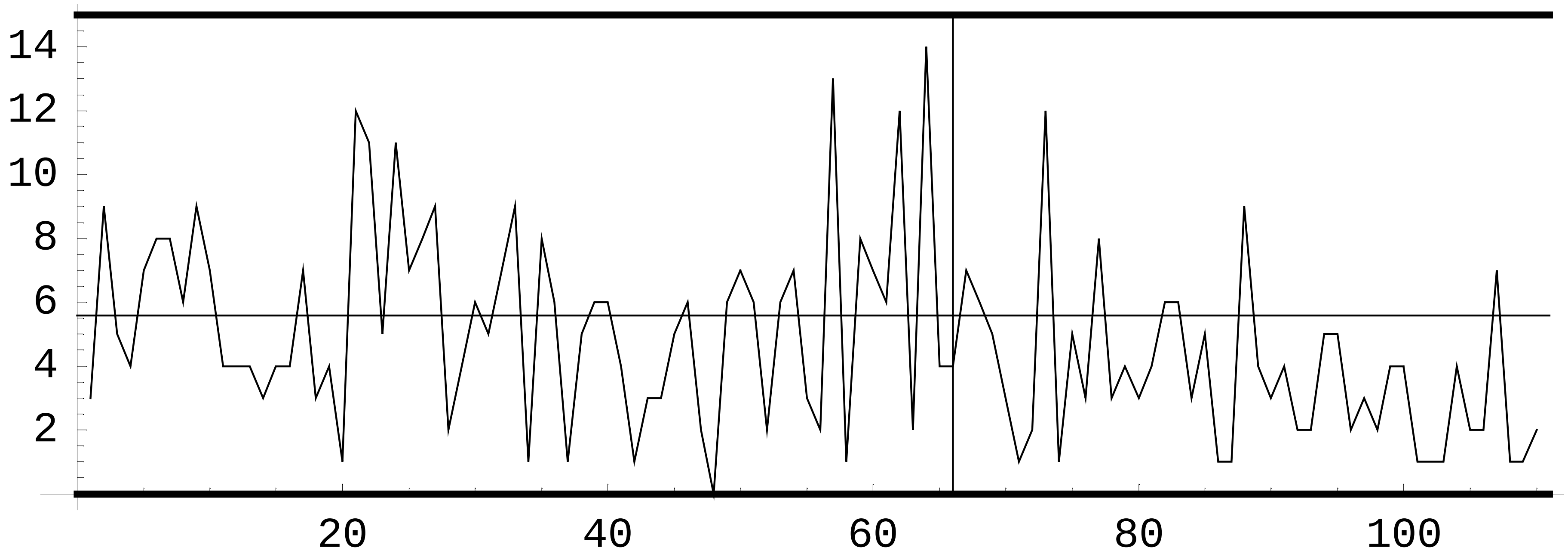
nearly unbiased.



Markov EWMA Charts: Real-data Example

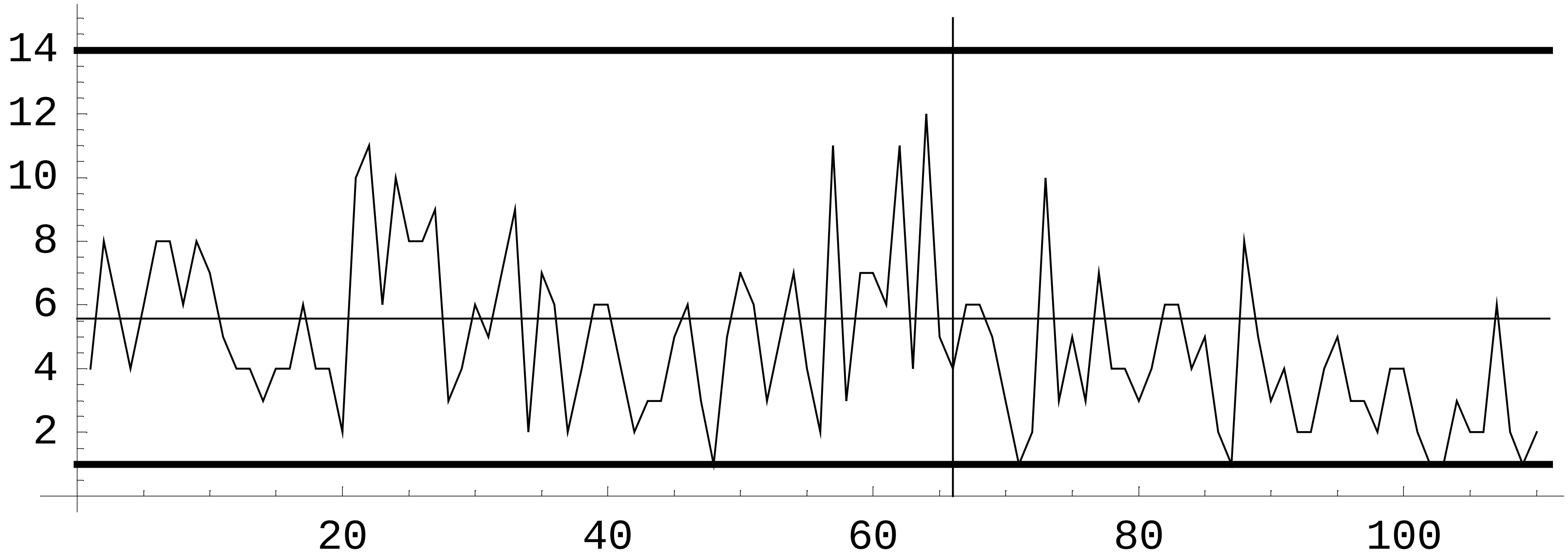
One-sided Markov np chart with $(\lambda, l, u) = (1, 0, 15)$;

$\rightarrow \; ARL_0 \approx 210.634$



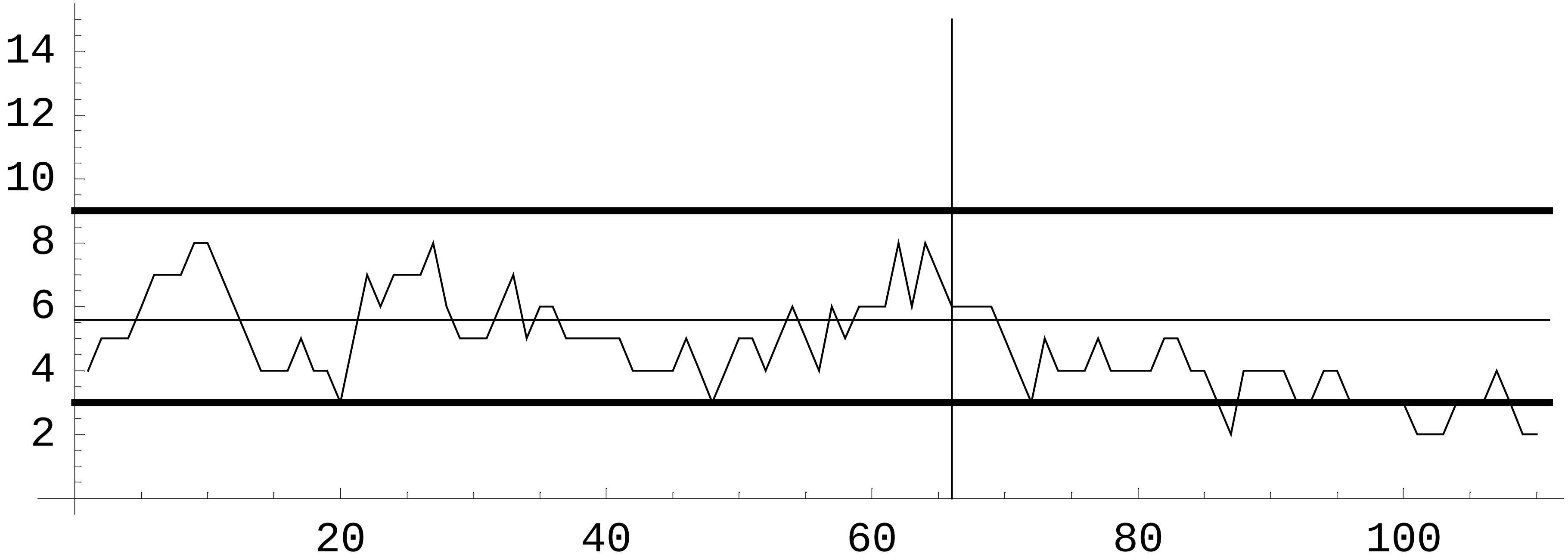
Markov EWMA chart with $(\lambda, l, u) = (0.82, 1, 14)$;

$\longrightarrow \quad ARL_0 \approx 214.358$



Markov EWMA chart with $(\lambda, l, u) = (0.26, 3, 9)$.

$\rightarrow \quad ARL_0 \approx 197.392$



Markov EWMA Charts: Real-data Example

- np chart $(1, 0, 15)$: no alarm, but decreasing trend. But several runs rules violated, e. g., $C_{89}{}^{(60)}, \ldots, C_{97}{}^{(60)}$ plotted below center line.
- EWMA chart $(0.82, 1, 14)$: no alarm, decreasing trend.
- EWMA chart $(0.26, 3, 9)$: six points below lower limit $\Rightarrow$ out-of-control.

$\Rightarrow$ Christmas season starts in December, so people busy with preparations.

Conclusions

- Monitoring distant segments from binary process;
- Markov binomial distribution to approximate distribution of the segment sums;
- Markov np chart and a Markov EWMA chart;
- exact ARL computations, performance towards several types of out-of-control situation:
  Properly designed Markov EWMA chart able to control any type of change in p and at least an increase in ρ.

Thank You
for Your Interest!

Christian H. Weiß

University of Würzburg

Institute of Mathematics

Department of Statistics