Omnibus Control Charts for Poisson Counts

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

MATH STAT

Christian H. Weiß

Department of Mathematics & Statistics,
Helmut Schmidt University, Hamburg

Majority of control charts for **Poisson counts** focus on mean changes, i. e., if $\mu = E[X_t]$ deviates from IC-mean μ_0, see Montgomery (2009), Alevizakos & Koukouvinos (2020).

For example,

ordinary **Poi-EWMA chart** (Borror et al., 1998) plots

$$Z_0 = \mu_0, \quad Z_t \;=\; \lambda \cdot X_t \;+\; (1-\lambda) \cdot Z_{t-1} \quad \text{for } t = 1, 2, \ldots$$

against specified LCL $< \mu_0$ and UCL $> \mu_0$,

where smoothing parameter $\lambda \in (0; 1]$ controls strength of memory ($\lambda = 1$: memory-less **c-chart**).

If not being concerned with *sole* mean changes
(i. e., if Poi-assumption preserved under OOC-conditions),
but if mean changes *together with distribution family*
(or if even pure changes in distribution family),
then performance of such charts may deteriorate severely.

For example, faulty measurement device might monitor some counts as zeros $\Rightarrow$ process zero-inflated with decreased mean.

Or if IC-distribution misspecified while μ_0 identified correctly, then pure change in distribution.

For such cases, Weiß (2023, 2024) proposed EWMA-type control charts relying on Stein identities, allow to flexibly adapt towards further distributional changes beyond mean.

Drawback: Stein-EWMA chart focuses on *certain* type of distributional change, e. g., underdispersion, overdispersion, zero inflation.
But if expected OOC-distribution not clear, one has to run several Stein-EWMA charts in parallel ("multiple testing").
User would prefer *single* chart type being powerful w.r.t. various OOC-scenarios at once ("**all-rounder**").

In context of goodness-of-fit (GoF) tests,

such broadly applicable tests called **omnibus tests**.

Several proposals of omnibus GoF-tests for Poi-null hypothesis, see Gürtler & Henze (2000). Here, consistency against various alternatives by utilizing distributional characterizations, e. g., statistics based on **probability generating function** (pgf), $\text{pgf}(u) = E[u^X]$, which equals $\text{pgf}_0(u) = \exp\big(\mu_0(u-1)\big)$ under Poi(μ_0)-assumption.

Idea: derive EWMA-type **omnibus control charts** for Poisson counts from common pgf-based Poi-GoF tests.

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

MATH STAT

Omnibus EWMA Charts for Poisson Counts

Integral GoF-Tests

Omnibus Poi-EWMA Charts: Integral

GoF-test by Rueda et al. (1991), Rueda & O'Reilly (1999):

$$\mathcal{R} \;=\; n\int_0^1 \Big(\widehat{\mathsf{pgf}}(u) - \mathsf{pgf}_0(u)\Big)^2\, u^a\, \mathrm{d}u.$$

Approach of Baringhaus & Henze (1992), utilizing

ODE $\mathsf{pgf}_0'(u) = \mu_0\, \mathsf{pgf}_0(u)$ under Poi(μ_0)-conditions:

$$\mathcal{B} \;=\; n\int_0^1 \Big(\mu_0\, \widehat{\mathsf{pgf}}(u) - \widehat{\mathsf{pgf}}'(u)\Big)^2\, u^a\, \mathrm{d}u.$$

Here, $\widehat{\mathsf{pgf}}(u) = \overline{u^X}$ and $\widehat{\mathsf{pgf}}'(u) = \overline{X\, u^{X-1}}$.

While original proposals without weighting ($a = 0$),

Baringhaus et al. (2000) propose additional weight function u^a,

where $a > 0$ puts weight near end of integration interval.

Integrals can be computed analytically, leading to

$$\mathcal{R} \;=\; n \sum_{k,l=0}^{\infty} \frac{(\widehat{p}_k - p_k^{(0)})(\widehat{p}_l - p_l^{(0)})}{k+l+a+1},$$

$$\mathcal{B} \;=\; n \sum_{k,l=0}^{\infty} \frac{\big(\mu_0\,\widehat{p}_k - (k+1)\,\widehat{p}_{k+1}\big)\big(\mu_0\,\widehat{p}_l - (l+1)\,\widehat{p}_{l+1}\big)}{k+l+a+1}.$$

To derive EWMA implementations, substitute frequencies $\widehat{p}_k$ by $P_{t,k} = \lambda \cdot \mathbb{1}(X_t = k) + (1-\lambda)\cdot P_{t-1,k}$ for $t = 1, 2, \ldots$

Crucial observation: if $M_t = \max\{X_1, \ldots, X_t\}$, then $\mathbb{1}(X_s = k) = 0$ for $k > M_t$ and all $s = 1, \ldots, t$.

$\Rightarrow$ $\mathcal{B}$-EWMA chart, where infinite sums become finite: (...)

$\mathcal{B}$-**EWMA chart** with specified $\lambda \in (0;1]$ and UCL > 0:

Initialize $M_0 = 0$ and $B_{k,0} = 0$, compute recursively:

$$M_t = \max\{X_t,\ M_{t-1}\},$$

$$B_{k,t} = \lambda \cdot \Big(\mu_0\, \mathbb{1}(X_t = k) - (k+1)\, \mathbb{1}(X_t = k+1)\Big) + (1-\lambda) \cdot B_{k,t-1}.$$

Then, $B_{k,s} = 0$ for all $k > M_t$ and $s \leq t$. At time t, plot

$$\mathcal{B}_t = \sum_{k,l=0}^{M_t} \frac{B_{k,t}\, B_{l,t}}{k+l+a+1}.$$

Unfortunately, for $\mathcal{R}$-statistic, reduction to finite sums not possible, because IC-Poi(μ_0) probabilities $p_k^{(0)} \neq 0$ for all $k \in \mathbb{N}_0$.

$\Rightarrow$ approximate truncation, where $p_k^{(0)} \approx 0$ for all $k > M$.

$\mathcal{R}$-**EWMA chart** with spec. $\lambda \in (0;1]$, $M \in \mathbb{N}_0$, and UCL > 0:

Initialize $R_{k,0} = 0$, compute recursively:

$$R_{k,t} \;=\; \lambda \cdot \Big(\mathbb{1}\big(\min\{M, X_t\} = k\big) - p_k^{(0)}\Big) \;+\; (1-\lambda) \cdot R_{k,t-1}.$$

At time t, plot

$$\mathcal{R}_t \;=\; \sum_{k,l=0}^{M} \frac{R_{k,t}\, R_{l,t}}{k+l+a+1}.$$

Choice of M: use $(1-10^{-d})$-quantile of Poi(μ_0).

For example, if target $\text{ARL}_0 = 370$, then

$$P(M_{370} > M) \;\approx\; 3.7 \cdot 10^{2-d}, \qquad \text{(e.g., with } d = 12\text{)}$$

i.e., first non-zero digit in $(d-2)^{\text{nd}}$ decimal place.

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

MATH STAT

Omnibus EWMA Charts for Poisson Counts

Discrete-Sum GoF-Tests

Inspired by **midpoint rule for integrals**, alternative implementation with fixed number $K \in \mathbb{N}$ of summands:

$$\mathcal{R}^{(K)} = n \sum_{k=1}^{K} \Big(\widehat{\mathrm{pgf}}(u_k) - \mathrm{pgf}_0(u_k)\Big)^2 u_k^a,$$

$$\mathcal{B}^{(K)} = n \sum_{k=1}^{K} \Big(\mu_0\, \widehat{\mathrm{pgf}}(u_k) - \widehat{\mathrm{pgf}}'(u_k)\Big)^2 u_k^a.$$

Here, grid $\boldsymbol{u} = (u_1, \ldots, u_K)^\top$ with equidistant midpoints $u_k := \frac{2k-1}{2K} \in (0;1)$ of sub-intervals $[\frac{k-1}{K}; \frac{k}{K})$.

$\mathcal{R}^{(K)}$-**EWMA chart** with spec. $\lambda \in (0;1]$, $K \in \mathbb{N}$, and UCL > 0:

Initialize $R_{k,0}^{(K)} = 0$, compute recursively:

$$R_{k,t}^{(K)} \;=\; \lambda \cdot \Big(u_k^{X_t} - \text{pgf}_0(u_k)\Big) \;+\; (1-\lambda)\cdot R_{k,t-1}^{(K)}.$$

At time t, plot
$$\mathcal{R}_t^{(K)} \;=\; \sum_{k=1}^{K} \big(R_{k,t}^{(K)}\big)^2\, u_k^a.$$

$\mathcal{B}^{(K)}$-**EWMA chart** (. . .)

$$B_{k,t}^{(K)} \;=\; \lambda \cdot \Big(\mu_0\, u_k^{X_t} - X_t\, u_k^{X_t-1}\Big) \;+\; (1-\lambda)\cdot B_{k,t-1}^{(K)}.$$

At time t, plot
$$\mathcal{B}_t^{(K)} \;=\; \sum_{k=1}^{K} \big(B_{k,t}^{(K)}\big)^2\, u_k^a.$$

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

MATH
STAT

Omnibus EWMA Charts for Poisson Counts

Performance Analyses

Comprehensive **simulation study** with lot of tables.

Overdispersion and zero-inflation OOC-scenarios (NB and ZIP, respectively),			
Interpre-tation:	sole mean change	pure distrib. change	zero inflation overdispersion
Underdispersion OOC-scenarios (Good),			
Interpre-tation:	sole mean change	pure distrib. change	increasing underdispersion

Illustrative example:

μ_0	$\mu =$	$\mu_0 - 0.25$	μ_0	$\mu_0 + 0.25$	$\mu_0 - 0.25$	μ_0	$\mu_0 + 0.25$	$\mu_0 - 0.25$	μ_0	$\mu_0 + 0.25$
	$a = 0$	E, $\lambda = 0.05$, *0.562*			$\mathcal{B}$-E, $\lambda = 0.005$, *0.0043*			$\mathcal{B}^{(25)}$-E, $\lambda = 0.005$, *0.1075*		
2	ZIP	78.7	107.5	62.6	32.6	40.9	46.5	32.6	40.9	46.5
	Poi	141.8	**368.4**	101.3	116.7	**369.9**	131.9	116.7	**370.0**	132.1
	NB	80.5	113.3	66.4	48.5	80.6	112.3	48.5	80.6	112.7
2	Poi	141.8	**368.4**	101.3	116.7	**369.9**	131.9	116.7	**370.0**	132.1
	Good	212.5	905.2	131.9	127.7	115.0	68.0	127.8	115.0	68.0
	Good	427.7	5231.3	214.4	73.7	62.5	48.7	73.7	62.5	48.6

Upper block:

Zero inflation or overdispersion, dispersion index $I = 5/3$.

Lower block:

Increasing underdispersion, $I = 3/4$ or $I = 1/2$, respectively.

Main findings:

- In most cases, $\mathcal{B}$-EWMA improves over $\mathcal{R}$-EWMA chart.
- $\mathcal{B}^{(25)}$-EWMA chart (with $K = 25$ summands) virtually equivalent to $\mathcal{B}$-EWMA without discretization.
- "all-rounder" if no weighting ($a = 0$).
- "all-rounder" if $\lambda = 0.005$ (both over- and underdispersion). If only overdispersion expected, then better $\lambda = 0.01$.

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

MATH STAT

Omnibus EWMA Charts for Poisson Counts

Data Applications

Three real-world **data examples:**

- particle counts in semiconductor manufacturing,
- counts of killed persons in road traffic crashes,
- weekly counts of lottery winners in Germany.

For illustration, **counts of particles** on produced wafers,

originally published by Nishina (2007),

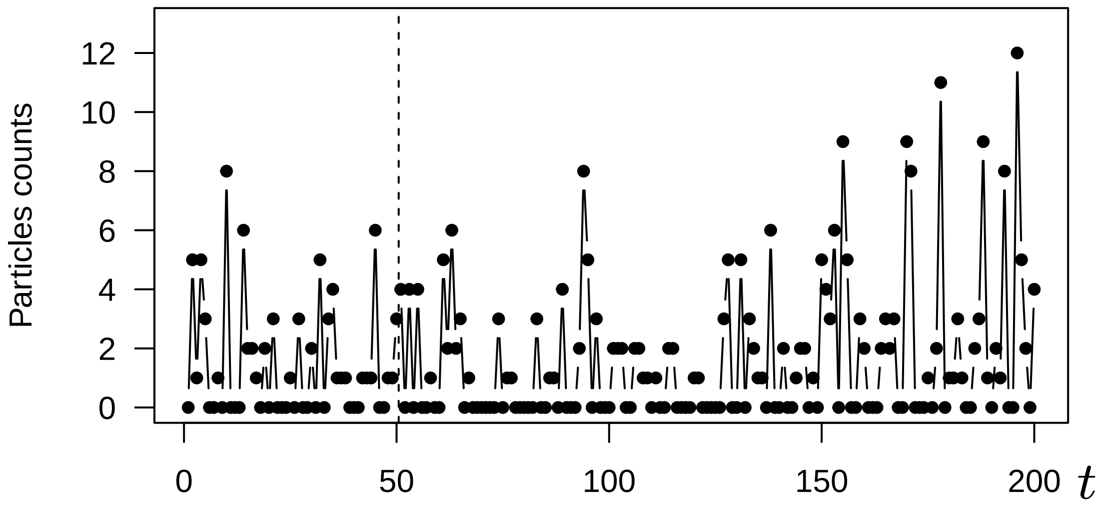
further analyzed by Weiß (2023):

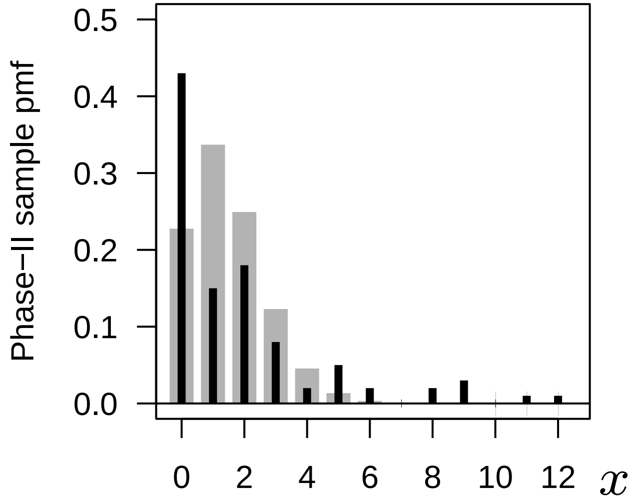
First 50 counts as Phase-I sample

$\Rightarrow$ IC-model i. i. d.-Poi(1.48) for particle counts.

150 remaining counts as Phase-II data: (. . .)

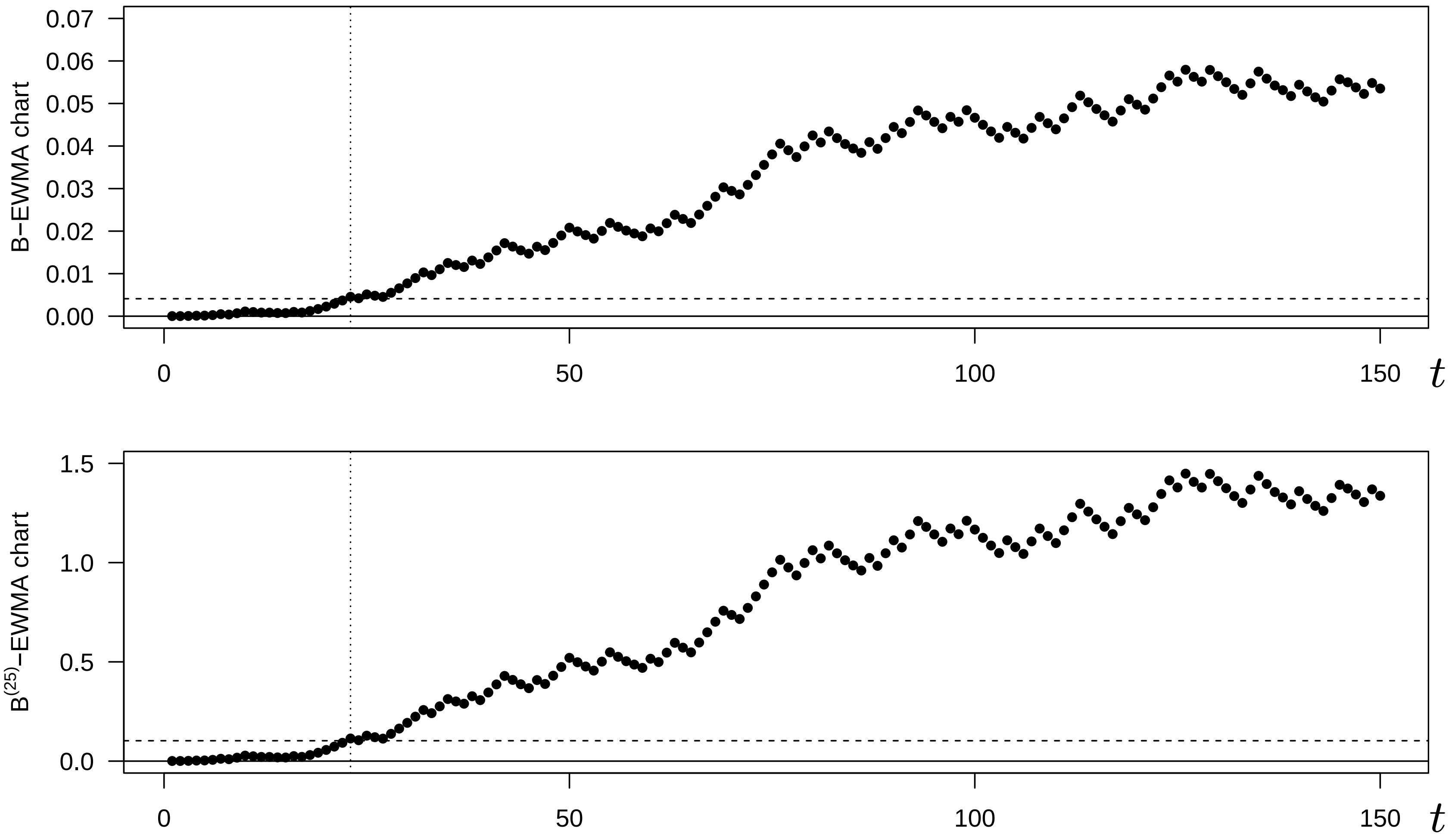
Plot and Phase-II sample pmf (black) against IC pmf (grey):





While ordinary EWMA chart triggers first alarm at $t = 31$, $\mathcal{B}$-EWMA and $\mathcal{B}^{(25)}$-EWMA chart with $a = 0$ and $\lambda = 0.005$ much faster, $t = 23$. **Possible explanation** for alarms: model misspecification, overdispersion and zero inflation.

$\mathcal{B}$-EWMA and $\mathcal{B}^{(25)}$-EWMA chart with $a = 0$ and $\lambda = 0.005$:



Conclusions

- If monitoring count data, not solely focus on mean, but also account for additional changes in distribution.
- Instead multiple specialized control charts in parallel, use single "all-rounder": **omnibus control chart**.
- Different **omnibus EWMA charts** derived from pgf-based GoF-statistics.
- $\mathcal{B}$-EWMA and $\mathcal{B}^{(25)}$-EWMA most promising (and virtually equivalent), where no weighting ($a = 0$) and $\lambda \leq 0.01$ to achieve omnibus property.

Thank You

for Your Interest!

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

MATH STAT

Christian H. Weiß

Department of Mathematics & Statistics

Helmut Schmidt University, Hamburg

weissc@hsu-hh.de

Literature

Weiß (2024) Omnibus control charts for Poisson counts. *Manuscript*, submitted.

Alevizakos & Koukouvinos (2020) A comparative ... *QTQM* **17**, 354–382.

Baringhaus & Henze (1992) A goodness of fit test ... *SPL* **13**, 269–274.

Baringhaus et al. (2000) Weighted integral test ... *ANZJS* **42**, 179–192.

Borror et al. (1998) Poisson EWMA control charts. *JQT* **30**, 352–361.

Gürtler & Henze (2000) Recent and classical good... *JSPI* **90**, 207–225.

Montgomery (2009) *Introduction to Statistical Quality ...* 6th ed., Wiley.

Nishina (2007) Sampling in implem... *Encyc Stat Qual Reliab*, Wiley.

Rueda et al. (1991) Goodness of fit for ... *CSTM* **20**, 3093–3110.

Rueda & O'Reilly (1999) Tests of fit for ... *CSSC* **28**, 259–274.

Weiß (2023) Control charts for Poisson ... *Adv Stat Meth ...*, Springer.

Weiß (2024) Stein EWMA control ... *Meth Appl Sys Ass Qual*, CRC.