Continuously Monitoring
Categorical Processes

Christian H. Weiß
Department of Mathematics,
Darmstadt University of Technology
This talk is based on the paper

**Weiß (2010):**
Continuously Monitoring Categorical Processes.
Accepted for publication in
*Quality Technology & Quantitative Management.*

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

---

Christian H. Weiβ — Darmstadt University of Technology
Categorical Time Series Analysis

Brief review
Categorical process:

\((X_t)_{\mathbb{N}}\) with \(\mathbb{N} = \{1, 2, \ldots\}\), where each \(X_t\) takes one of finite number of unordered categories.

Categorical time series:

Realizations \((x_t)_{t=1,\ldots,T}\) from \((X_t)_{\mathbb{N}}\).

To simplify notations:

Range of \((X_t)_{\mathbb{N}}\) is coded as \(\mathcal{V} = \{1, \ldots, m, m+1\}\), i. e., \(P(X_t = m + 1) = 1 - \sum_{j=1}^{m} P(X_t = j)\).
Definitions according to Weiß (2009a), Section 7.1.1:

\((X_t)_{\mathbb{N}}\) said to be

- **marginally stationary**
  
  if marginal distribution invariant in \(t\);

- **bivariately stationary** or **weakly stationary**
  
  if pairwise joint distribution of \((X_t, X_{t-k})\) independent of \(t\) for each \(k \in \mathbb{N}_0\);

- **(strictly) stationary** if joint distribution of \((X_t, \ldots, X_{t+k})\) independent of \(t\) for all \(k \in \mathbb{N}_0\).
Notations for time-invariant probabilities:

- If \((X_t)_N\) marginally stationary:
  - marginal probabilities \(p_i := P(X_t = i) \in (0; 1)\).
  - \(p := (p_1, \ldots, p_{m+1})^\top\), and
  - \(s_k(p) := \sum_{j=1}^{m+1} p_j^k\) for \(k \in \mathbb{N}\); obviously \(s_1(p) = 1\).
- If \((X_t)_N\) weakly stationary:
  - bivariate probabilities \(p_{ij}(k) := P(X_t = i, X_{t-k} = j)\),
  - conditional probabilities \(p_{i|j}(k) := P(X_t = i \mid X_{t-k} = j)\).
Categorical Time Series Analysis

Let \((X_t)_N\) be marginally stationary.

**Measures of location:**
only mode of \(X_t\) in use, i. e., value \(i \in \mathcal{V}\) such that \(p_i \geq p_j\) for all \(j \in \mathcal{V}\).
Often not uniquely determined (e. g., uniform distribution).

**Measures of dispersion:**
dispersion \(\approx\) quantity of uncertainty, two extremes:
maximal dispersion if all \(p_j\) equal (uniform distribution),
minimal disp. if \(p_j = 1\) for one \(j \in \mathcal{V}\) (one-point distrib.).
Most simple measure of dispersion: **Gini index** of $X_t$,

$$\nu_G(X_t) := \frac{m+1}{m} \cdot (1 - \sum_{j=1}^{m+1} p_j^2) = \frac{m+1}{m} \cdot (1 - s_2(p)).$$

- continuous and symmetric function of $p_1, \ldots, p_{m+1}$,
- range $[0; 1]$,
- maximal value 1 iff uniform distribution,
- minimal value 0 iff one-point distribution.

Weiß & Göb (2008): **signed serial dependence**.

Weakly stationary categorical process \((X_t)_{\mathbb{N}}\) said to be

- **perfectly serially dependent** at lag \(k \in \mathbb{N}\) if for any \(j \in \mathcal{V}\), conditional distribution \(p_{i|j}(k)\) is one-point distribution;

- **serially independent** at lag \(k \in \mathbb{N}\) if \(p_{i|j}(k) = p_i\) (i. e., \(p_{ij}(k) = p_ip_j\)) for any \(i, j \in \mathcal{V}\).

\(\ldots\)

Christian H. Weiß — Darmstadt University of Technology
In case of perfect serial dependence at lag $k \in \mathbb{N}$:

- perfect **positive dependence**
  
  if $p_{i|j}(k) = 1$ iff $i = j$ for all $i, j \in \mathcal{V}$;

- perfect **negative dependence** if all $p_{i|i}(k) = 0$. 
Categorical Time Series Analysis

Weiß & Göb (2008): Cohen’s $\kappa$,

$$\kappa(k) = \frac{\sum_{j=1}^{m+1} (p_{jj}(k) - p_j^2)}{1 - \sum_{j=1}^{m+1} p_j^2} = 1 - \frac{1 - \sum_{j=1}^{m+1} p_{jj}(k)}{1 - s_2(p)} \approx \text{Gini index}.$$

- range $[-\frac{s_2(p)}{1-s_2(p)} ; 1]$,
- $\kappa(k) = 1$ iff perfect positive dependence at lag $k$,
- $\kappa(k) = 0$ if serial independence,
- $\kappa(k) = -\frac{s_2(p)}{1-s_2(p)}$ if perfect negative dependence.


Christian H. Weiß — Darmstadt University of Technology
Basic models for categorical time series

**Bernoulli process:** i.i.d. random variables. Stationary distribution determined by \( p \). \((m\) parameters\)

**Markov chain:** first order Markov process. Stationary distribution determined by transition probabilities

\[
p_{i|j} = P(X_t = i \mid X_{t-1} = j).
\]

\((m(m+1))\) parameters

If \( P = (p_{i|j})_{i,j=1,...,m+1} \) denotes **transition matrix**, then stationary marginal distribution \( p \) has to satisfy \( Pp = p \).
Categorical Time Series Analysis

Reduce number of parameters, \( m(m + 1) \):

**DAR(1) model** (Jacobs & Lewis, 1983),
\[
p_{i|j} = p_i \cdot (1 - \phi) + \delta_{ij} \cdot \phi \quad \text{with} \quad \phi \in [0; 1). (m + 1 \text{ parameters})
\]
Stationary marginal distribution: \( p. \kappa(k) = \phi^k \).

**Negative Markov model** (Weiß, 2009b):
\[\pi \in (0; 1)^{m+1}\] probability vector, \( \alpha \in (0; 1] \) and \( \beta_j := \frac{1 - \alpha \pi_j}{1 - \pi_j} \).
Define
\[
p_{i|j} := \begin{cases} 
\alpha \cdot \pi_j & \text{if } i = j, \\
\beta_j \cdot \pi_i & \text{if } i \neq j. 
\end{cases}
\]
\( \Rightarrow \) negative dependence

Ergodic Markov chain, invariant distribution \( p_j = \frac{\pi_j/\beta_j}{\sum_{i=1}^{m+1} \pi_i/\beta_i} \).
Continuously Monitoring Categorical Processes

Motivation & Scope
Examples for categorical time series from diverse fields: biological sequence analysis, speech recognition, part-of-speech tagging, network monitoring, . . . (Weiß, 2009a).

SPC: $X_t =$ result of inspection of item, with $X_t = i$ for $i = 1, \ldots, m$ iff item has nonconformity type $i$, $X_t = m + 1$ iff conforming.


Ye et al. (2002) monitor network traffic data (284 different types of audit events) for intrusion detection.
Continuously Monitoring Categorical Processes

Already some work about monitoring of ‘multi-attribute processes’ or ‘multinomial processes’, **but** not always . . .

- about **mutually exclusive** categories (→ multivariate approaches),
- about **unordered** categories (not ordinal!),
- about **probabilistic** approaches (e. g., fuzzy theory),
- about **continuously** monitoring (i. e., 100 % inspection, not samples).

Christian H. Weiß — Darmstadt University of Technology
Continuously Monitoring Categorical Processes

From now on, we assume \((X_t)_N\) to be stationary.

**In-control state:** \((X_t)_N\) serially independent, i.e., altogether i.i.d., where \(p = p_0\) for known \(p_0\).

**Aim:** detect both violations of independence assumption and changes in \(p\) compared to \(p_0\).

**Example of categorical quality characteristics:**
states 1, \ldots, \(m\) as different nonconformity categories, \(m + 1\) represents conforming item.
We expect that \(p_{0,m+1} \gg p_{0,1}, \ldots, p_{0,m}\), relation between \(p_{m+1}\) and \(p_1, \ldots, p_m\) particularly relevant.

Christian H. Weiß — Darmstadt University of Technology
Control Charts based on a Comparative Statistic Approaches
Control Charts – Comparative Statistic

General strategy:

At each $t$, compute statistic, which summarizes characteristic properties of true marginal distribution $p$.

Then compares these properties to ones expected from in-control marginal distribution $p_0$.

Plot resulting real-valued process of comparative statistics on appropriately designed control chart.
Control Charts – Comparative Statistic

We restrict to moving average (MA) estimator $\hat{p}_t(w)$.

Define binarization $(X_t)_N$ via $X_{t,i} = 1$ iff $X_t = i$ and 0 otherwise.

We consider $\hat{p}_t(w) := \frac{1}{w} \cdot \sum_{r=0}^{w-1} X_{t-r}$ for $t \geq w$.

Let $T_{p_0} : [0; 1]^{m+1} \rightarrow \mathbb{R}$ be comparative function with respect to in-control marginal distribution $p_0$.

Compute comparative statistics $T_t := T_{p_0}(\hat{p}_t(w))$ based on estimated marginal distribution.

Plot statistics $T_t$ on control chart with appropriately chosen control limits $LCL, UCL$.
Performance evaluation: Average run length (ARL) counts plotted statistics $T_t$ until alarm.
⇒ misleading, since first $w$ original observations $X_1, \ldots, X_w$ necessary before first statistic $T_w$.
⇒ Average number of events (ANE) counts the original observations, in our case: $ANE = ARL + w - 1$.

Obviously, always $ANE \geq w$, i.e., large $w$ avoids quick detection of process changes.
But components $\hat{p}_{t,i}^{(w)}$ multiples of $\frac{1}{w} \Rightarrow w$ sufficiently large to express smallest probability among $p_{0,i}$.

Christian H. Weiß — Darmstadt University of Technology
1\textsuperscript{st} approach: Pearson\textquotesingle s $\chi^2$-statistic for goodness of fit.

$$T_{p_0}(\hat{p}_t^{(w)}) := \sum_{j=0}^{m} \frac{(\hat{p}_{t,j}^{(w)} - p_{0,j})^2}{p_{0,j}}$$ for $t \geq w$.

Plot $T_t := T_{p_0}(\hat{p}_t^{(w)})$ on one-sided chart with $UCL > 0$.

If $(X_t)_N$ i.i.d. with $p_0$, then $w \cdot T_t \sim \chi^2_m$. $\Rightarrow UCL$

Rough approx. for small $w$, $(T_t)_{t \geq w}$ of MA($w-1$) type.

$T_t$ discrete, finite range $\Rightarrow$ ANE target not met exactly.

$T_t$ measures any type of change in $p$ with respect to $p_0$.

Already mean $E[T_t]$ sensitive to (positive) serial dependence.
2\textsuperscript{nd} approach: empirical Gini index.

\[ T_{p_0}(\hat{p}_t(w)) := \frac{1 - s_2(\hat{p}_t(w))}{1 - s_2(p_0)} - 1. \]

Plot \( T_t := T_{p_0}(\hat{p}_t(w)) \) on two-sided chart, \( LCL < 0 < UCL \).

Detect those changes in \( p \), which result in changed dispersion with respect to \( p_0 \) (counterexample: permutation).

Useful for finding \( LCL, UCL \), see Weiß (2009b):

\[ E[T_t] = -\frac{1}{w}, \quad V[T_t] \approx \frac{4}{w} \cdot \frac{s_3(p_0) - s_2^2(p_0)}{(1 - s_2(p_0))^2}. \]

\( T_t \) discrete with a finite range.
Control Charts based on Runs

Approaches
Control Charts – Runs

\((k, r)\)-run: finished after \(k\) successive observations of either ‘1’ or ‘2’ or \(\ldots\) or ‘\(r\)’, i. e., after observing one of \((1, \ldots, 1), (2, \ldots, 2), \ldots, (r, \ldots, r)\) of length \(k\) each.

\((k, r)^{th}\) run lengths \((Y_n^{(k,r)})_\mathbb{N}\) determined as

\[ Y_1^{(k,r)} := \text{No. obs. until first occurrence of } k\text{-tuple of ‘1’s or } \ldots \text{ ‘r’s,} \]
\[ Y_n^{(k,r)} := \text{No. obs. after } (n-1)^{th} \text{ occurrence of } k\text{-tuple of ‘1’s or } \ldots \text{ ‘r’s, until } n^{th} \text{ occurrence of } k\text{-tuple of ‘1’s or } \ldots \text{ ‘r’s, for } n \geq 2. \]

Example: \(m = 3\) (i. e., \(V = \{1, 2, 3, 4\}\))
and \((k, r) = (2, 3)\), (fictive) time series:

\[
\begin{array}{cccccccccccc}
1 & 2 & 4 & 4 & 4 & 3 & 4 & 2 & 2 & 4 & 3 & 4 & 4 & 1 & 4 & 1 & 1 & 1 & 1 & 3 & 4 & 2 & 4 & 3 & 3 & 2 & 3 & \ldots \\
9 & | & 8 & | & 2 & | & 6 &
\end{array}
\]

Christian H. Weiß — Darmstadt University of Technology
Control Charts – Runs

\((Y_n^{(k,r)})_N\) plotted on chart with \(k \leq LCL < UCL\).

**Properties:**

- increase in \(p_1, \ldots, p_r \Rightarrow\) reduced run lengths and vice versa,
  i. e.,
- violation of lower limit indicates increase in \(p_1, \ldots, p_r\).

- segment length \(k\) increases \(\Rightarrow\) increasing run lengths
- segment number \(r\) increases \(\Rightarrow\) decreasing run lengths

Christian H. Weiß — Darmstadt University of Technology
Control Charts – Runs

\((Y_n^{(k,r)})_N\) i.i.d. process, range \(\mathbb{N}_k := \{k, k + 1, \ldots\}\).

**Properties:** (Chryssaphinou et al., 1994, Theorem 2.1 and Corollary 2.2)

Let \(c_{k,r}(z) := \sum_{i=1}^{r} \frac{(1 - p_i z)(p_i z)^k}{1 - (p_i z)^k}\), then

\[
E[Y^{(k,r)}] = \frac{1}{c_{k,r}(1)}, \quad V[Y^{(k,r)}] = \frac{1 + c_{k,r}(1) - 2c'_{k,r}(1)}{c_{k,r}(1)^2},
\]

probab. gen. funct. (pgf) \(p_{Y^{(k,r)}}(z) = \frac{c_{k,r}(z)}{1 - z + c_{k,r}(z)}\).
(\(Y_n^{(k,r)}\))_N \text{ i.i.d.} \Rightarrow \textbf{ARLs of plotted statistics } Y_n^{(k,r)}:\n
\[ ARL = \left(1 - \sum_{y=LCL}^{UCL} P(Y^{(k,r)} = y)\right)^{-1}.\]

Misleading: \(Y_n^{(k,r)}\) represents many observations from \((X_t)_N\).

\(\Rightarrow\) consider \textit{ANE}s. Weiß (2010):

If \((X_t)_N\) be a stationary Markov chain, then

\textbf{exact ANE computation} with Markov chain approach.

Christian H. Weiß — Darmstadt University of Technology
Performance of Control Charts

ANE Comparison
Comparison, simulations or MC approach.

**Out-of-control situations:**

\[ p_i = \beta \cdot p_{0,i} \quad \text{for } i = 1, \ldots, m, \]
\[ p_{m+1} = 1 - \beta \cdot (1 - p_{0,m+1}), \]
\[ \beta = 1: \text{in-control situation}, \]
\[ \beta > 1: p_{m+1} \text{ decreased, other probabilities increased}. \]

**Violations of the independence assumption:**

DAR(1) model with dependence parameter \( \phi \in [0; 1) \)

negative Markov model with dependence par. \( \alpha \in (0; 1] \)
Few illustrative results for $p_0 = (0.09, 0.12, 0.25, 0.54)^T$:

- **runs:** $(k, r; LCL, UCL) = (2, 2; 5, 165)$, $ANE_0 \approx 503$,
- **Gini:** $(w; LCL, UCL) = (25; -0.45, 0.175)$, $ANE_0 \approx 484$,
- **Gini:** $(w; LCL, UCL) = (50; -0.255, 0.135)$, $ANE_0 \approx 509$,
- **Pearson:** $(w; UCL) = (25; 0.4725)$, $ANE_0 \approx 521$,
- **Pearson:** $(w; UCL) = (50; 0.2000)$, $ANE_0 \approx 508$,
- **Pearson:** $(w; UCL) = (100; 0.0825)$, $ANE_0 \approx 519$.
Negative Markov model, $\alpha_0 = 1$:

$\Rightarrow$ only $(k, r)$-runs chart sensitive
Control Charts – Performance

DAR(1) model, $\phi_0 = 0$:

$\Rightarrow (k, r)$-runs chart best except large $\phi$

Christian H. Weiß — Darmstadt University of Technology
i.i.d. but changed $p$, $\beta_0 = 1$:

$\Rightarrow$ Gini chart (two-sided!) nearly $ANE$-unbiased, but sometimes insensitive, e. g.:

$\beta = 2 \Rightarrow p = (0.18, 0.24, 0.50, 0.08) \approx$ permutation of $p_0$. 

Christian H. Weiß — Darmstadt University of Technology
Conclusions

- Gini chart, Pearson chart: chart design and evaluation based on simulations, superior concerning changes in $p$, Gini often preferable (exception: permutations).

- $(k, r)$-runs chart: $ANE$ computation by Markov chain approach, superior concerning violations of serial independence, worst concerning changes in $p$ except: low dispersion (e. g., high-quality process) ⇒ more sensitive concerning negative shifts in $\beta$. 

Christian H. Weiß — Darmstadt University of Technology
Thank You
for Your Interest!

Christian H. Weiß
Department of Mathematics
Darmstadt University of Technology
weiss@mathematik.tu-darmstadt.de