Continuously Monitoring Categorical Processes

Christian H. Weiß

Department of Mathematics,

Darmstadt University of Technology

Literature

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.

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 $(X_t)_{\mathbb{N}}$ marginally stationary:
  marginal probabilities $p_i := P(X_t = i) \in (0; 1)$.
  $\boldsymbol{p} := (p_1, \ldots, p_{m+1})^\top$, and
  $s_k(\boldsymbol{p}) := \sum_{j=1}^{m+1} p_j^k$ for $k \in \mathbb{N}$; obviously $s_1(\boldsymbol{p}) = 1$.

- If $(X_t)_{\mathbb{N}}$ weakly stationary:
  bivariate probabilities $p_{ij}(k) := P(X_t = i, X_{t-k} = j)$,
  conditional probabilities $p_{i|j}(k) := P(X_t = i \mid X_{t-k} = j)$.

Let $(X_t)_{\mathbb{N}}$ 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,

$$\nu_{\mathsf{G}}(X_t) \;:=\; \tfrac{m+1}{m}\cdot\Big(1-\textstyle\sum_{j=1}^{m+1}\; p_j^2\Big) \;=\; \tfrac{m+1}{m}\cdot\Big(1-s_2(\boldsymbol{p})\Big).$$

- 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.

Alternative measures of disp.: Section 6.1 of Weiß (2009a).

Empirical counterparts, asymptotic properties: Weiß (2009b).

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., $p_{ij}(k) = p_i p_j$) for any $i, j \in \mathcal{V}$.

(. . .)

(. . .)

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 $p_{i|i}(k) = 0$.

Weiß & Göb (2008): **Cohen's** κ,

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

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

Empirical counterparts, asymptotic properties: Weiß (2009b).

Alternative measure, modified κ: Weiß (2009a,b).

Basic models for categorical time series

Bernoulli process: i.i.d. random variables.

Stationary distribution determined by $\boldsymbol{p}$. (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 $\mathbf{P} = (p_{i|j})_{i,j=1,\ldots,m+1}$ denotes **transition matrix**, then stationary marginal distribution $\boldsymbol{p}$ has to satisfy $\mathbf{P}\boldsymbol{p} = \boldsymbol{p}$.

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

DAR(1) model (Jacobs & Lewis, 1983),

$p_{i|j} = p_i \cdot (1-\phi) + \delta_{ij} \cdot \phi$ with $\phi \in [0;1)$.($m+1$ parameters)

Stationary marginal distribution: $\boldsymbol{p}$. $\kappa(k) = \phi^k$.

Negative Markov model (Weiß, 2009b):

$\boldsymbol{\pi} \in (0;1)^{m+1}$ probability vector, $\alpha \in (0;1]$ and $\beta_j := \frac{1-\alpha\pi_j}{1-\pi_j}$.

Define

$$p_{i|j} \ := \ \left\{ \begin{array}{lll} \alpha \cdot \pi_j & \text{if } i = j, & < \pi_j \\ \beta_j \cdot \pi_i & \text{if } i \neq j. & > \pi_i \end{array} \right\} \Rightarrow \text{ negative dependence}$$

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.

Mukhopadhyay (2008): $m = 6$ paint defects of ceiling fan cover ('poor covering', 'bubbles', etc.).
Overall defect category = most predominant defect.

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

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

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

From now on, we assume $(X_t)_{\mathbb{N}}$ to be stationary.

In-control state: $(X_t)_{\mathbb{N}}$ serially independent, i. e., altogether i.i.d., where $\boldsymbol{p} = \boldsymbol{p}_0$ for known $\boldsymbol{p}_0$.

Aim: detect both violations of independence assumption and changes in $\boldsymbol{p}$ compared to $\boldsymbol{p}_0$.

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.

Control Charts based on a Comparative Statistic

Approaches

General strategy:

At each t, compute statistic, which summarizes characteristic properties of **true** marginal distribution $\boldsymbol{p}$.

Then compares these properties to ones expected from **in-control** marginal distribution $\boldsymbol{p}_0$.

Plot resulting real-valued process of **comparative statistics** on appropriately designed control chart.

We restrict to **moving average (MA) estimator** $\widehat{\boldsymbol{p}}_t^{(w)}$:

Define **binarization** $(\boldsymbol{X}_t)_{\mathbb{N}}$ via $X_{t,i} = 1$ iff $X_t = i$ and 0 otherwise.

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

Let $T_{\boldsymbol{p}_0}: [0;1]^{m+1} \to \mathbb{R}$ be **comparative function** with respect to in-control marginal distribution $\boldsymbol{p}_0$.

Compute **comparative statistics** $T_t := T_{\boldsymbol{p}_0}(\widehat{\boldsymbol{p}}_t^{(w)})$ 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.

$\Rightarrow$ misleading, since first w original observations $X_1, \ldots, X_w$ necessary before first statistic T_w.

$\Rightarrow$ 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 $\widehat{p}_{t,i}^{(w)}$ multiples of $\frac{1}{w} \Rightarrow w$ sufficiently large to express smallest probability among $p_{0,i}$.

1st approach: Pearson's χ^2-statistic for goodness of fit.

$$T_{\boldsymbol{p}_0}(\widehat{\boldsymbol{p}}_t^{(w)}) \;:=\; \sum_{j=0}^{m} \frac{(\widehat{p}_{t,j}^{(w)} - p_{0,j})^2}{p_{0,j}} \qquad \text{for } t \geq w.$$

Plot $T_t := T_{\boldsymbol{p}_0}(\widehat{\boldsymbol{p}}_t^{(w)})$ on one-sided chart with $UCL > 0$.

If $(X_t)_{\mathbb{N}}$ i.i.d. with $\boldsymbol{p}_0$, then $w \cdot T_t \underset{\text{approx}}{\sim} \chi^2_m$. $\Rightarrow UCL$

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 $\boldsymbol{p}$ with respect to $\boldsymbol{p}_0$.
Already mean $E[T_t]$ sensitive to (positive) serial dependence.

2nd approach: empirical Gini index.

$$T_{\boldsymbol{p}_0}(\widehat{\boldsymbol{p}}_t^{(w)}) \; := \; \frac{1 - s_2(\widehat{\boldsymbol{p}}_t{}^{(w)})}{1 - s_2(\boldsymbol{p}_0)} - 1.$$

Plot $T_t := T_{\boldsymbol{p}_0}(\widehat{\boldsymbol{p}}_t{}^{(w)})$ on two-sided chart, $LCL < 0 < UCL$.

Detect those changes in $\boldsymbol{p}$, which result in changed dispersion with respect to $\boldsymbol{p}_0$ (counterexample: permutation).

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

$$E[T_t] \; = \; -\frac{1}{w}, \qquad V[T_t] \; \approx \; \frac{4}{w} \cdot \frac{s_3(\boldsymbol{p}_0) - s_2^2(\boldsymbol{p}_0)}{(1 - s_2(\boldsymbol{p}_0))^2}.$$

T_t discrete with a finite range.

Control Charts based on Runs

Approaches

TECHNISCHE UNIVERSITÄT DARMSTADT

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

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

$$\begin{aligned} Y_1^{(k,r)} &:= \text{No. obs. until first occurrence of } k\text{-tuple of '1's or ... '}r\text{'s,} \\ Y_n^{(k,r)} &:= \text{No. obs. after } (n-1)^{\text{th}} \text{ occurrence of } k\text{-tuple of '1's or ... '}r\text{'s} \\ &\quad\ \text{until } n^{\text{th}} \text{ occurrence of } k\text{-tuple of '1's or ... '}r\text{'s, for } n \geq 2. \end{aligned}$$

Example: $m = 3$ (i. e., $\mathcal{V} = \{1,2,3,4\}$)

and $(k,r) = (2,3)$, (fictive) time series:

$$\underbrace{1\ 2\ 4\ 4\ 4\ 3\ 4\ \mathbf{2}\ \mathbf{2}}_{9}\ \underbrace{4\ 3\ 4\ 4\ 1\ 4\ \mathbf{1}\ \mathbf{1}}_{8}\ \underbrace{\mathbf{1}\ \mathbf{1}}_{2}\ \underbrace{3\ 4\ 2\ 4\ \mathbf{3}\ \mathbf{3}}_{6}\ 2\ 3\ \ldots$$

$(Y_n^{(k,r)})_{\mathbb{N}}$ 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

$(Y_n^{(k,r)})_{\mathbb{N}}$ 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

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

probab. gen. funct. (pgf) $p_{Y^{(k,r)}}(z) \;=\; \frac{c_{k,r}(z)}{1-z+c_{k,r}(z)}$.

$({Y_n}^{(k,r)})_{\mathbb{N}}$ i.i.d. $\Rightarrow$ ARLs of **plotted statistics** ${Y_n}^{(k,r)}$:

$$ARL \;=\; \big(1-\textstyle\sum_{y=LCL}^{UCL} P(Y^{(k,r)}=y)\big)^{-1}.$$

Misleading: ${Y_n}^{(k,r)}$ represents many observations from $(X_t)_{\mathbb{N}}$.

$\Rightarrow$ consider ANEs. Weiß (2010):

If $(X_t)_{\mathbb{N}}$ be a stationary Markov chain, then

exact ANE **computation** with Markov chain approach.

Performance
of
Control Charts

ANE Comparison

ANE comparison, simulations or MC approach.

Out-of-control situations:

$$\begin{aligned} p_i &= \beta \cdot p_{0,i} \quad \text{for } i = 1, \ldots, m, \\ p_{m+1} &= 1 - \beta \cdot (1 - p_{0,m+1}), \end{aligned} \qquad \beta \in [0; (1-p_{0,m+1})^{-1}].$$

$\beta = 1$: in-control situation,

$\beta > 1$: p_{m+1} decreased, other probabilities increased.

Violations of the independence assumption:

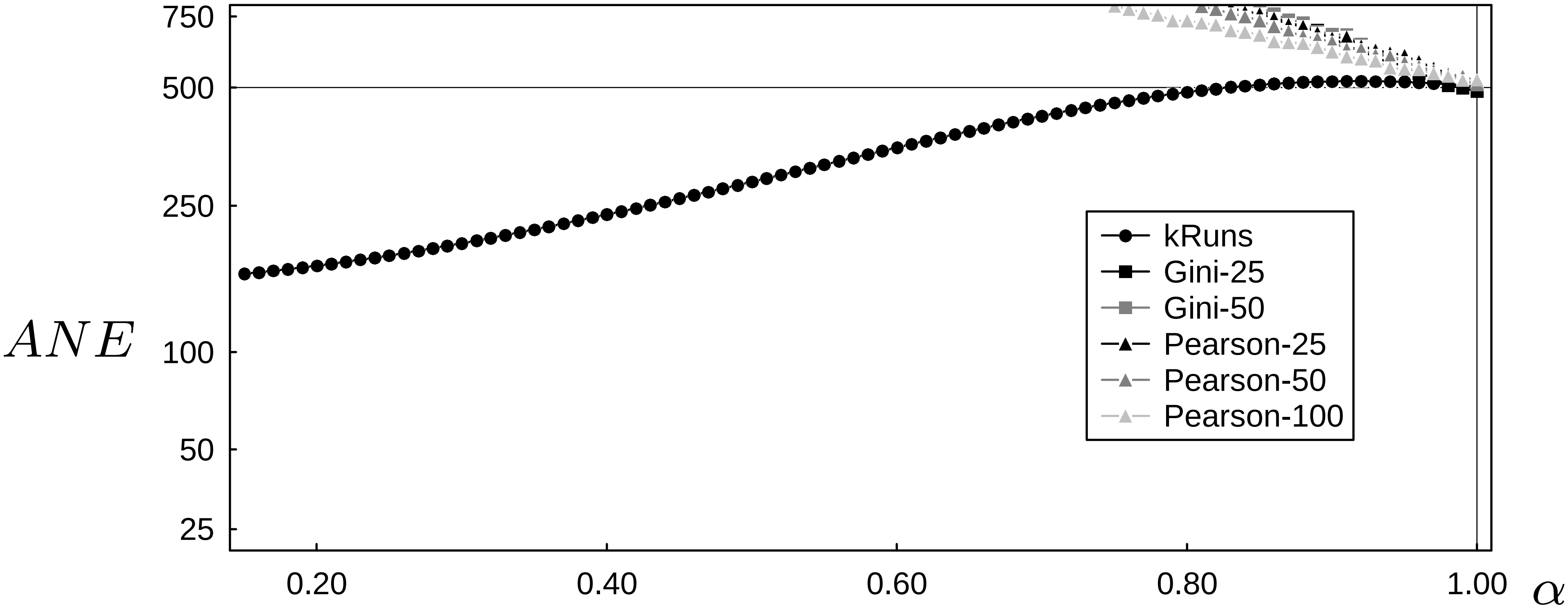
DAR(1) model with dependence parameter $\phi \in [0; 1)$

negative Markov model with dependence par. $\alpha \in (0; 1]$

Few illustrative results for $\boldsymbol{p}_0 = (0.09, 0.12, 0.25, 0.54)^\top$:

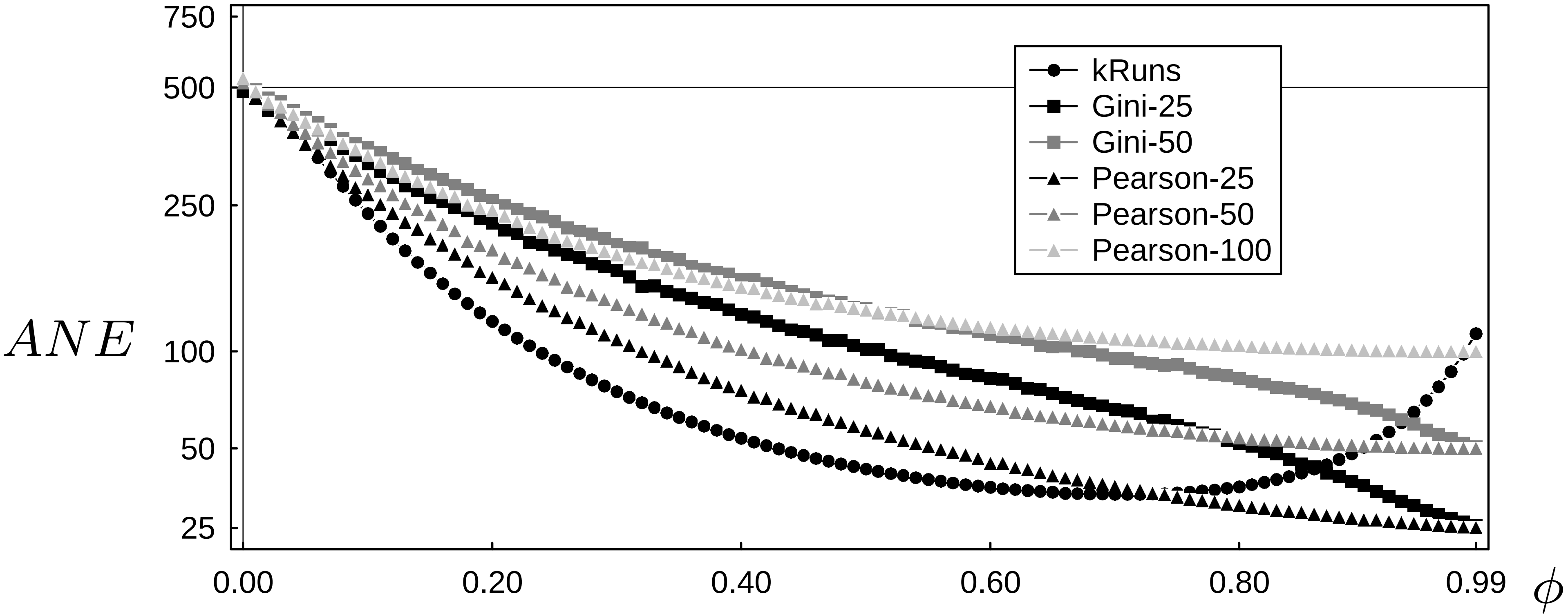
- runs: $(k, r; LCL, UCL) = (2, 2; 5, 165)$, $ANE_0 \approx 503$,
- Gini: $(w; LCL, UCL) = (25; -0.45, 0.175)$, $ANE_0 \approx 484$,
- Gini: $(w; LCL, UCL) = (50; -0.255, 0.135)$, $ANE_0 \approx 509$,
- Pearson: $(w; UCL) = (25; 0.4725)$, $ANE_0 \approx 521$,
- Pearson: $(w; UCL) = (50; 0.2000)$, $ANE_0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

Negative Markov model, $\alpha_0 = 1$:



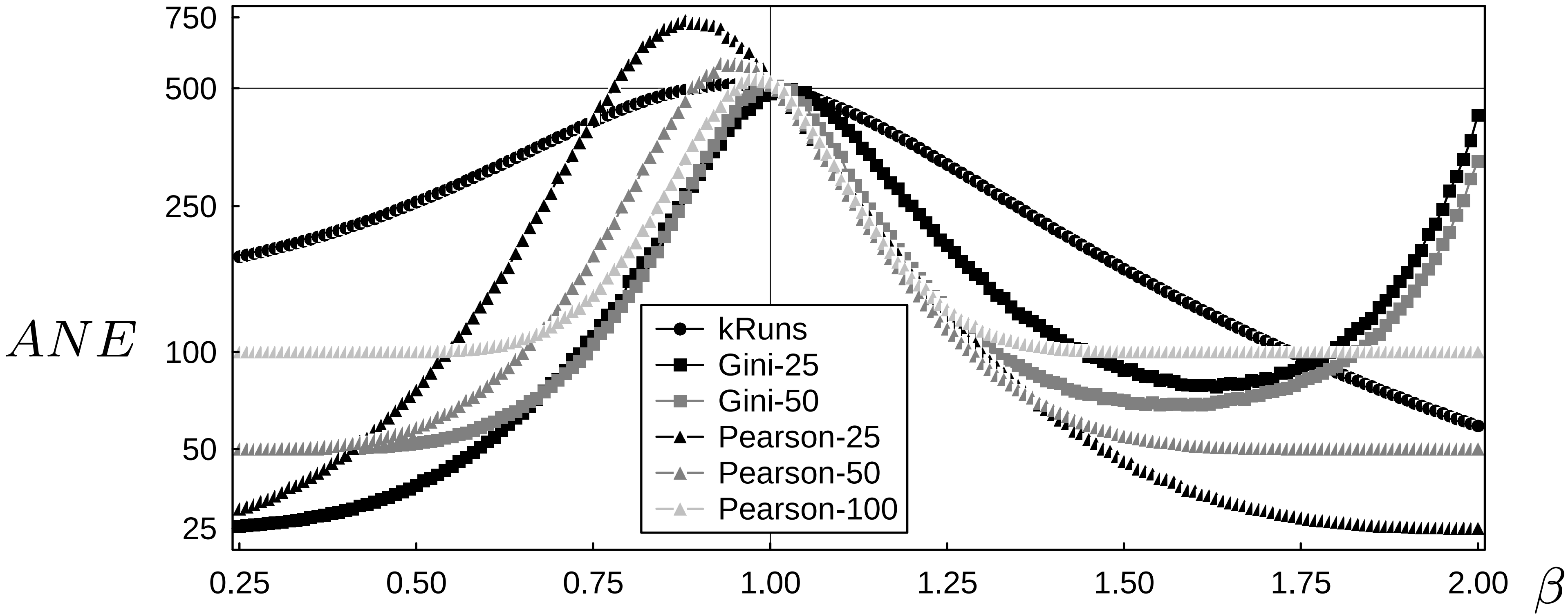
$\Rightarrow$ only (k, r)-runs chart sensitive

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



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

i.i.d. but changed $\boldsymbol{p}$, $\beta_0 = 1$:



$\Rightarrow$ Gini chart (two-sided!) nearly ANE-unbiased,

but sometimes insensitive, e. g.:

$\beta = 2 \Rightarrow \boldsymbol{p} = (0.18, 0.24, 0.50, 0.08) \approx$ permutation of $\boldsymbol{p}_0$.

Conclusions

- Gini chart, Pearson chart:
  chart design and evaluation based on simulations,
  superior concerning changes in $\boldsymbol{p}$,
  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 $\boldsymbol{p}$ except:
  low dispersion (e. g., high-quality process)
  $\Rightarrow$ more sensitive concerning negative shifts in β.

Thank You

for Your Interest!

TECHNISCHE
UNIVERSITÄT
DARMSTADT

Christian H. Weiß

Department of Mathematics

Darmstadt University of Technology

weiss@mathematik.tu-darmstadt.de