Distance-based Analysis of Ordinal Time Series

Christian H. Weiß

Department of Mathematics & Statistics, Helmut Schmidt University, Hamburg
Distance-based Analysis of Ordinal Time Series

Motivation & Outline
We analyze **ordinal time series** $X_1, \ldots, X_n$ having ordered categorical range $S = \{s_0, \ldots, s_m\}$ with $s_0 < \ldots < s_m$.

Ordinal r.v. $X$ closely related to **rank-count** variable $I$ with bounded range $\{0, \ldots, m\}$ defined by $X = s_I$.

Count process $(I_t)_\mathbb{Z}$ generates ordinal $(X_t)_\mathbb{Z}$ by $X_t = s_{I_t}$.

$(X_t)_\mathbb{Z}$ inherits distributional properties from $(I_t)_\mathbb{Z}$ like stationarity, mixing, etc. (Weiß, 2018). In particular,

$$P(X_t = s_k) = P(I_t = k), \quad P(X_t = s_k, X_{t-h} = s_l) = P(I_t = k, I_{t-h} = l).$$

**Rank-count duality** related to **latent-variable appr.** (Agresti, 2010), but neither continuous-valued nor unobservable.

---

Christian H. Weiß — Helmut Schmidt University, Hamburg
Here: distance functions on ordinal range $S$, measures of marginal or serial properties based on these distances:

- unified approach with obvious interpretation,
- great flexibility with respect to tailor-made distances,
- universal applicability, also beyond ordinal categories.

Distance-based approaches common in statistical analysis, e.g., Székely et al. (2007): serial dependence in real-valued processes as distances between characteristic functions, or Kvålseth (2011): ordinal variation via distance between CDFs.

Present work: distances defined on $S$ itself, not between CDFs or other stochastic properties.

Christian H. Weiß — Helmut Schmidt University, Hamburg
Outline:

- ordinal distances and their properties;
- distance-based analytic tools for marginal properties of ordinal t.s. like location, dispersion, or skewness;
- serial dependence of ordinal time series;
- sample versions and their asymptotics;
- application to time series of monthly credit ratings.
Ordinal Distances

Definition & Properties
Definition:

\( d : S \times S \to [0; \infty) \) distance measure on \( S \) if for all \( x, y \in S \):

(D1) If \( x = y \), then \( d(x, y) = 0 \). (positive semi-definite)

(D2) \( d(x, y) = d(y, x) \). (symmetry)

Distance \( d \) is pseudo-metric if it even satisfies

(D3) \( d(x, z) \leq d(x, y) + d(y, z) \) for all \( x, y, z \in S \). (triangular inequality)

Pseudo-metric \( d \) is metric if it even satisfies

(D1′) \( x = y \) iff \( d(x, y) = 0 \) for all \( x, y \in S \). (positive definite)
Possible Properties of Ordinal Distances:

Let $d : S \times S \to [0; \infty)$ be distance measure on ordinal range $S = \{s_0, \ldots, s_m\}$. It may satisfy:

(O1) $d(s_0, s_m) = \max_{x,y \in S} d(x,y)$.  (maximization)

(O2) $d$ is compatible with the ordering if

$$x < y < z \quad \text{implies that} \quad d(x, z) > d(x, y), d(y, z).$$

(O3) $d$ is additive if for given $d_1, \ldots, d_m > 0$, it holds that

$$d(s_i, s_{i+k}) = d_{i+1} + \ldots + d_{i+k} \quad \forall i=0,\ldots,m-1; \ k=1,\ldots,m-i.$$ 

(O4) $d(s_i, s_j) = d(s_{m-i}, s_{m-j}) \quad \forall 0 \leq i < j \leq m$.  (centrosymmetry)
We have (O3) ⇒ (O2) ⇒ (O1).

(O4) relevant for possible symmetry of ordinal distribution.

**Examples:**

**Hamming distance** $d_H(s_k, s_l) := 1 - \delta_{k,l}$

metric with (O1), (O4), but not (O2).

**Ordinal block distance** $d_{O,1}(s_k, s_l) := |k - l|$

so number of categories between $s_k$ and $s_l$; all properties above.

**Squared Euclidean distance** $d_{O,2}(s_k, s_l) := (k - l)^2$

(quadratic penalty) no metric (not (D3)), not (O3).
Tailor-made distances possible!

Choice of weights for dissimilarities depends on application. Successive ordinal states not necessarily “equally-spaced”:

- At HSU, grades are 1, 1.3, ..., 3.7, 4 (pass), and 4.3, 5 (fail). $d(4, 4.3)$ largest among successive differences?!

Distance-based Analysis of Ordinal Random Variables

Approaches & Examples
Proposal: quantify properties of ordinal r.v. (location, dispersion, skewness) by expected distances. Since distance between r.v. non-negative real number, expectation thereof well defined. Universal and flexible approach, covers already existing measures of marginal properties.

Notations:

- $p = (p_0, \ldots, p_m)^\top$, where $p_i = P(X = s_i)$. So $p$ summarizes PMF.
- CDF: $f_k = \sum_{l=0}^{k} p_l = P(X \leq s_k)$ and $f = (f_0, \ldots, f_{m-1})^\top$. 
Distance-based Analysis of Ordinal R.V.

**Location:**

Location as $x_{loc,d} := \arg \min_{x \in S} E[d(X, x)]$.

Then $x_{loc,d} \in S$.

Hamming distance $\mapsto$ mode(s),
block distance $\mapsto$ median.

Note that $d_{o,2} \mapsto$ mean $E[I]$, because

$E[d_{o,2}(X, s_i)] = E[(I - i)^2]$ minimized for $i \in \{0, \ldots, m\}$.

So $d_{o,2}$ leads to $s_{\text{round}}(E[I])$.

Christian H. Weiβ — Helmut Schmidt University, Hamburg
Dispersion (variability) of $X$ as expected distance between independent copies $X_1, X_2$ of $X$, i.e., $E[d(X_1, X_2)]$ → diversity coefficient (DIVC) by Rao (1982). So

$$\text{disp}_d(X) := E[d(X_1, X_2)] = \sum_{i,j=0}^{m} d(s_i, s_j) p_i p_j = p^\top D p,$$

If $d$ positive definite, then $\text{disp}_d = 0$ iff one-point distribution.
If property (O1) holds, $\text{disp}_d \leq d(s_0, s_m)$.

**Extreme scenarios** for ordinal data:
minimal for one-point distribution $p_{\text{one}}$ (maximal consensus),
maximal for extreme two-point $d$. $p_{\text{two}} = (0.5, 0, \ldots, 0, 0.5)^\top$
(polarized distribution, maximal dissent), see Kiesl (2003).
Distance-based Analysis of Ordinal R.V.

Examples:

- Block distance leads to **Gini’s mean difference**.
  
  In ordinal case,
  \[
  \text{disp}_{d_0,1} = \sum_{k,l=0}^{m} |k - l| p_k p_l = 2 \sum_{k=0}^{m-1} f_k (1 - f_k).
  \]
  
  To be normalized by factor \(2/d_{0,1}(s_0, s_m) = 2/m\), then **index of ordinal variation** (Kvålseth, 1995; Kiesl, 2003).

- Squared Euclidean distance leads to **variance**:
  \[
  \text{disp}_{d_0,2} = E[(I_1 - I_2)^2] = 2 V[I],
  \]
  
  normalized version given by \(\frac{2}{m^2} \text{disp}_{d_0,2} = \frac{4}{m^2} V[I]\).

Christian H. Weiß — Helmut Schmidt University, Hamburg
Symmetry (brief sketch!)

If $X$ ordinal r.v. with $P(X = s_i) = p_i$,
then $X^r$ reflected copy of $X$ if $P(X = s_i) = p_{m-i}$.

Symmetry reasonably defined only if $d$ centrosymmetric.

Symmetric if $X$ and $X^r$ are identically distributed, i.e.,
if $p_i = p_{m-i}$ or $f_i = 1 - f_{m-1-i}$ for all $i = 0, \ldots, m$.

$$\text{asym}_d(X) = E[d(X, X^r)] - \text{disp}_d(X) = \sum_{i,j=0}^m d(s_i, s_j) p_i (p_{m-j} - p_j).$$

Example:

$$\text{asym}_{d_0,1} = \sum_{j=0}^{m-1} (1 - f_j - f_{m-j-1})^2.$$
Also more refined distinction possible, \textbf{skewness} (Klein & Doll, 2018).

Maximal positive (negative) skewness iff $p_0 = 1$ ($p_m = 1$).

Signed \textbf{skewness measure:}

$$\text{skew}_d(X) = E\left[d(X, s_m)\right] - E\left[d(X, s_0)\right].$$

\textbf{Example:}

$$\text{skew}_{d_{0,1}} = 2 \sum_{i=0}^{m-1} f_i - m.$$
Distance-based Analysis of Ordinal Time Series

Serial Dependence
Now \( x_1, \ldots, x_n \) time series from stationary ordinal process \((X_t)_{\mathbb{Z}}\). DGP assumed to satisfy appropriate mixing condition such that application of CLT possible, e.g., \( \alpha \)-mixing with exponentially decreasing weights.

We use bivariate probabilities (for \( i, j = 0, \ldots, m - 1 \))

\[
 f_{ij}^{(h)} = P(X_t \leq s_i, X_{t-h} \leq s_j) = P(I_t \leq i, I_{t-h} \leq j)
\]

and corresponding sample frequencies \( \hat{f}_{ij}^{(h)} \).

We assume that \((X_t)_{\mathbb{Z}}\) some kind of “short memory”:

\[
 \sum_{h=1}^{\infty} \left| f_{ij}^{(h)} - f_i f_j \right| < \infty \text{ has to hold for all } i, j = 0, \ldots, m - 1.
\]
Measure of **serial dependence** at lag $h$:

$$
\kappa_d(h) = \frac{\text{disp}_d(X) - E[d(X_t, X_{t-h})]}{\text{disp}_d(X)} \quad \text{for } h \geq 1.
$$

Type of **weighted Cohen's $\kappa$** (Cohen, 1968).

$\text{disp}_d(X)$ as $E[d(X_t, X_{t-h})]$ under null of independence, i.e., $\kappa_d(h)$ relative deviation of DIVCs for dependent and indep. r.v.

**Examples:**

$$
\kappa_{dH}(h) = \frac{\sum_{i=0}^{m} p_{ii}^{(h)} - s_2(p)}{1 - s_2(p)},
$$

so serial Cohen's $\kappa(h)$ for nominal t.s. (Weiß & Göb, 2008).
Examples:

$$\kappa_{d_0,2}(h) = \rho_I(h), \text{ i.e., lag-} h \text{ ACF of rank counts } (I_t)_{\mathbb{Z}}.$$

Particularly relevant is $d = d_{0,1}$:

$$\kappa_{d_0,1}(h) = \frac{\sum_{i=0}^{m-1} (f_{ii}^{(h)} - f_i^2)}{\sum_{i=0}^{m-1} f_i(1 - f_i)} \in \left[ -\frac{\sum_{i=0}^{m-1} f_i^2}{\sum_{i=0}^{m-1} f_i(1 - f_i)} ; 1 \right].$$

Strongest negative dependence if all $f_{ii}^{(h)} = 0$, i.e., $X_{t-h} \leq s_i$ necessarily followed by $X_t > s_i$.

Maximal positive dependence if $X_{t-h} \leq s_i$ necessarily followed by $X_t \leq s_i$.
Distance-based Analysis of Ordinal Time Series

Sample measures
Distance-based Analysis of Ordinal Time Series

Sample versions easily obtained by replacing expected distances by sample means of distances, e. g., sample version of $x_{\text{loc},d}$ as $\hat{x}_{\text{loc},d} := \arg \min_x \frac{1}{n} \sum_{t=1}^{n} d(x_t, x)$.

DIVC-related measures lead to
\[
\hat{\text{disp}}_d = \hat{p}^\top D \hat{p},
\hat{\text{asym}}_d = \hat{p}^\top (J - I)D \hat{p},
\hat{\text{skew}}_d = \sum_{i=0}^{m} \left( d(s_i, s_m) - d(s_i, s_0) \right) \hat{p}_i.
\]

For asymptotics, we focus on ordinal block distance $d_{o,1}$:
\[
\hat{\text{disp}}_{d_{o,1}} = 2 \sum_{i=0}^{m-1} \hat{f}_i (1 - \hat{f}_i),
\hat{\text{asym}}_{d_{o,1}} = \sum_{j=0}^{m-1} (1 - \hat{f}_j - \hat{f}_{m-j-1})^2,
\hat{\text{skew}}_{d_{o,1}} = 2 \sum_{i=0}^{m-1} \hat{f}_i - m,
\hat{\kappa}_{d_{o,1}}(h) = \frac{\sum_{i=0}^{m-1} (\hat{f}^{(h)}_{ii} - \hat{f}_i^2)}{\sum_{i=0}^{m-1} \hat{f}_i (1 - \hat{f}_i)}.
\]

Christian H. Weiß — Helmut Schmidt University, Hamburg
Theorem:

(i) If \((f_0, \ldots, f_{m-1}) \neq (1/2, \ldots, 1/2)\) (extreme two-point),
then \(\hat{\text{disp}}_{d_o,1}\) is asymptotically normally distributed with mean

\[
(1 - \frac{1}{n}) \text{disp}_{d_o,1} - \frac{4}{n} \sum_{h=1}^{n-1} (1 - \frac{h}{n}) \sum_{i=0}^{m-1} (f_{ii}^{(h)} - f_i^2)
\]

and variance

\[
\frac{4}{n} \sum_{i,j=0}^{m-1} (1 - 2f_i)(1 - 2f_j) (f_{\min \{i,j\}} - f_i f_j)
\]

\[
+ \frac{8}{n} \sum_{h=1}^{n-1} (1 - \frac{h}{n}) \sum_{i,j=0}^{m-1} (1 - 2f_i)(1 - 2f_j) (f_{ij}^{(h)} - f_i f_j).
\]
Theorem:

(ii) If distribution of $X_i$ not symmetric, then $\widehat{\text{asym}}_{d_{o,1}}$ asymptotically normally distributed with mean

$$\text{asym}_{d_{o,1}} + \frac{1}{n} \text{disp}_{d_{o,1}} + \frac{2}{n} \sum_{j=0}^{m-1} \left( f_{\min \{m-j-1,j\}} - f_j f_{m-j-1} \right)$$

$$+ \frac{4}{n} \sum_{h=1}^{n-1} \left( 1 - \frac{h}{n} \right) \sum_{j=0}^{m-1} \left( f_{jj}^{(h)} - f_j^2 + f_{j,m-j-1} - f_j f_{m-j-1} \right)$$

and variance

$$\frac{16}{n} \sum_{i,j=0}^{m-1} (1 - f_i - f_{m-i-1})(1 - f_j - f_{m-j-1}) \left( f_{\min \{i,j\}} - f_i f_j \right) +$$

$$\frac{32}{n} \sum_{h=1}^{n-1} \left( 1 - \frac{h}{n} \right) \sum_{i,j=0}^{m-1} (1 - f_i - f_{m-i-1})(1 - f_j - f_{m-j-1}) \left( f_{ij}^{(h)} - f_i f_j \right).$$
Theorem:

(iii) \( \hat{\text{skew}}_{d_0,1} \) asympt. normal with mean \( \text{skew}_{d_0,1} \) and variance

\[
\frac{4}{n} \sum_{i,j=0}^{m-1} (f_{\min \{i,j\}} - f_i f_j) + \frac{8}{n} \sum_{h=1}^{n-1} (1 - \frac{h}{n}) \sum_{i,j=0}^{m-1} (f_{ij}(h) - f_i f_j).
\]

Modifications necessary for **boundary cases**, leads to asymptotic quadratic-form distributions.

Finite-sample performance confirmed with simulations.
Testing for significant dependence:

Let \( X_1, \ldots, X_n \) be time series from i.i.d. ordinal process \((X_t)_Z\), let \( h \geq 1 \).

Then asymptotic approximation for \( \hat{\kappa}_{d_0,1}(h) \) is normal distribution with mean \(-1/n\) and variance

\[
\frac{1}{n} \frac{4}{\text{disp}^2_{d_0,1}} \sum_{k,l=0}^{m-1} (f_{\min \{k,l\}} - f_k f_l)^2.
\]

Finite-sample performance confirmed with simulations.

Christian H. Weiβ — Helmut Schmidt University, Hamburg
Application: Credit Ratings

EU countries

Motivated by Trading Economics rating, we treat \( s_0, \ldots, s_{22} \) as equidistant, so \( d_{0,1} \).

Unique time range for all EU countries: \( n = 216 \) months from January 2000 to December 2017.
Monthly S&P credit rating of Latvia:
time series plot (left), $\hat{\kappa}_{d_0,1}(k)$ (right).

Norm. dispersion 0.155, asymmetry 0.240 and skewness $-0.318$. 

Christian H. Weiß — Helmut Schmidt University, Hamburg
Marginal properties of S&P time series of 28 EU countries:

Zero dispersion for Germany and Luxembourg ($s_{22} = \text{AAA}$).

Christian H. Weiß — Helmut Schmidt University, Hamburg
Marginal properties of S&P time series of 28 EU countries:

\[ \frac{1}{m} \text{asym}_{d_{0,1}} \leq 1 - \text{IOV}. \]

Christian H. Weiß — Helmut Schmidt University, Hamburg
Would also be reasonable to consider reduced range with \( \tilde{m} + 1 = 10 \) states \( \tilde{s}_0, \ldots, \tilde{s}_9 \) given by ‘D’, ‘R’, ‘CC’, ‘CCC’, ‘B’, ‘BB’, ‘BBB’, ‘A’, ‘AA’, ‘AAA’.

To make results comparable, equidistance not appropriate.

Following (additive) distance matrix used for reduced range:

\[
D = \begin{pmatrix}
0 & 2 & 3 & 5 & 8 & 11 & 14 & 17 & 20 & 22 \\
2 & 0 & 1 & 3 & 6 & 9 & 12 & 15 & 18 & 20 \\
3 & 1 & 0 & 2 & 5 & 8 & 11 & 14 & 17 & 19 \\
5 & 3 & 2 & 0 & 3 & 6 & 9 & 12 & 15 & 17 \\
8 & 6 & 5 & 3 & 0 & 3 & 6 & 9 & 12 & 14 \\
11 & 9 & 8 & 6 & 3 & 0 & 3 & 6 & 9 & 11 \\
14 & 12 & 11 & 9 & 6 & 3 & 0 & 3 & 6 & 8 \\
17 & 15 & 14 & 12 & 9 & 6 & 3 & 0 & 3 & 5 \\
20 & 18 & 17 & 15 & 12 & 9 & 6 & 3 & 0 & 2 \\
22 & 20 & 19 & 17 & 14 & 11 & 8 & 5 & 2 & 0 \\
\end{pmatrix}
\]

Not centrosymmetric (e.g., \( d(\tilde{s}_0, \tilde{s}_2) = 3 \neq 5 = d(\tilde{s}_9, \tilde{s}_7) \)).
Application: Credit Ratings

Location, dispersion and serial dependence computed w.r.t. $d$:

\[
\hat{x}_{\text{loc},d} = \arg \min_x \frac{1}{n} \sum_{t=1}^n d(x_t, x) \quad \text{instead of median } \hat{x}_{\text{loc},d_{o,1}},
\]

similarly

\[
\hat{\text{disp}}_d = \hat{p}^\top D \hat{p} \quad \text{(normalizing factor } \frac{2}{d(\tilde{s}_0, \tilde{s}_9)} = 1/11) \]

and

\[
\hat{\kappa}_d(h) = \left( \hat{\text{disp}}_d - \frac{1}{n-h} \sum_{t=h+1}^n d(x_t, x_{t-h}) \right) / \hat{\text{disp}}_d.
\]

**Crucial question:** how do location, dispersion and serial dependence change if using reduced range instead of full one? (...)

Location and dispersion robust against this change in range.

Serial dependence stronger affected, mainly larger values:

reduction of states causes less variation,

i.e., longer runs and thus increased positive dependence.

---

Christian H. Weiß — Helmut Schmidt University, Hamburg
Comparison of properties of S&P time series of 28 EU countries if using either full range or reduced range.

Not centrosymmetric, so dispersion of original t.s. and reflected copies thereof might differ. Latvia: 0.207 vs. 0.165.

Christian H. Weiβ — Helmut Schmidt University, Hamburg
Conclusions

- Location, dispersion, symmetry and serial dependence for ordinal time series based on expected distances. Well interpretable, can deal with many types of distance function, asymptotics available.

- Distance-based approach also for r.v. and processes having other types of range, e.g., compositional data. Thus allows for novel analytic tools for such data types, and offers opportunity to treat different data types with unique approach and unique interpretation.

Christian H. Weiß — Helmut Schmidt University, Hamburg
Thank You
for Your Interest!

Christian H. Weiß
Department of Mathematics & Statistics
Helmut Schmidt University, Hamburg
weissc@hsu-hh.de


