Hidden-Markov Models for Ordinal Time Series

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

MATH STAT

Christian H. Weiß, Osama Swidan

Department of Mathematics & Statistics,

Helmut Schmidt University, Hamburg

Funded by DFG – Projektnr. 516522977

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

MATH
STAT

Hidden-Markov Models for Ordinal Time Series

Introduction

X **categorical r.v.** if X bounded qualitative range, say $\mathcal{S} = \{s_0, s_1, \ldots, s_d\}$ with some $d \in \mathbb{N} = \{1, 2, \ldots\}$.

Such r.v. called **ordinal** if $\mathcal{S}$ exhibits natural order among the categories, say $s_0 < \ldots < s_d$ (Agresti, 2010).

For modeling ordinal r.v., one may link X to some real-valued **latent variable** Q with cdf F_Q: (→ GLMs)

$$X = s_j \qquad \text{iff} \quad Q \in [\eta_{j-1}; \eta_j),$$

where $-\infty = \eta_{-1} < \eta_0 < \ldots < \eta_{d-1} < \eta_d = +\infty$ such that

$$p_j = P(X = s_j) \; = \; F_Q(\eta_j) - F_Q(\eta_{j-1}) \quad \text{and}$$

$$f_j = P(X \leq s_j) \; = \; F_Q(\eta_j) \qquad \text{hold for } j = 0, \ldots, d.$$

Hidden-Markov model (HMM) for (X_t, H_t),

where observations X_t have ordinal range $\mathcal{S}$, and

hidden states H_t have range $\mathcal{H} = \{0, \ldots, d_{\mathcal{H}}\}$ with $d_{\mathcal{H}} \in \mathbb{N}$:

(i) **observation equation**

$$P(X_t = x | X_{t-1}, \ldots, H_t = h, H_{t-1}, \ldots) = P(X_t = x | H_t = h) = p(x|h),$$

(ii) **state equation**

$$P(H_t | X_{t-1}, \ldots, H_{t-1}, \ldots) = P(H_t | H_{t-1}, \ldots),$$

(iii) **Markov assumption** (transition matrix $\mathbf{A} = (a_{h|i})_{h,i \in \mathcal{H}}$)

$$P(H_t = h | H_{t-1} = i, H_{t-2}, \ldots) = P(H_t = h | H_{t-1} = i) = a_{h|i}.$$

If $\boldsymbol{\pi}_0$ of H_0 as $\boldsymbol{\pi}_0 := \boldsymbol{\pi}$ satisfying $\mathbf{A}\boldsymbol{\pi} = \boldsymbol{\pi}$,
then HMM (X_t, H_t) becomes **stationary**.

Categorical HMMs discussed so far (Weiß, 2018)
are indeed HMMs for **nominal** (X_t),
i. e., assume **no relations among states** in $\mathcal{S}$.

$\Rightarrow$ the $d_{\mathcal{H}} + 1$ different state-dependent distributions $p(\cdot|h)$
have d parameters each, so altogether
$d_{\mathcal{H}}\,(d_{\mathcal{H}} + 1) + d\,(d_{\mathcal{H}} + 1) = (d_{\mathcal{H}} + d)\,(d_{\mathcal{H}} + 1)$ parameters.

Such fully non-parametric specification of discrete HMM
also considered by Turner (2022) in "`hmm.discnp`".

Here: ordinal range $\mathcal{S}$, so **order relation among states**.

To account for ordinal nature of X_t,

and reduce number of parameters at same time,

we combine categorical HMM with latent-variable approach.

Ordinal HMM:

Let $Q_t^{(h)}$ latent variable emitted at time t if $H_t = h$,

with corresponding state-*dependent* cdf $F_{Q,h}$.

Given state-*independent* parameters

$-\infty = \eta_{-1} < \eta_0 < \ldots < \eta_{d-1} < \eta_d = +\infty$, emission of X_t via

$$X_t = s_j \mid H_t = h \qquad \text{iff} \quad Q_t^{(h)} \in [\eta_{j-1}; \eta_j).$$

Specific examples: $(d_{\mathcal{H}}\,(d_{\mathcal{H}}+1)\ +\ d+d_{\mathcal{H}}$ parameters)

- **logit**(μ) **HMM:** $Q_t^{(h)} \sim \mathsf{L}(\mu_h, 1)$ with $\mu_0 := 0$ and
  $F_{Q,h}(x) = F_Q(x-\mu_h)$, where $F_Q(u) = 1/\big(1+\exp(-u)\big)$;

- **soft-clipping**(μ) **HMM:** $Q_t^{(h)} \sim \mu_h + \mathsf{U}(0,1) + \mathsf{L}(0,\delta)$
  with $\mu_0 := 0$ and $F_{Q,h}(x) = F_Q(x-\mu_h)$, where

$$F_Q(u) \ = \ \mathrm{sc}_\delta(u) := \min\big\{1, \max\{0,u\}\big\} \ + \ \delta\ \ln\!\left(\frac{1+\exp(-|\frac{u}{\delta}|)}{1+\exp(-|\frac{1-u}{\delta}|)}\right);$$

- **logit**(σ) **HMM:** $Q_t^{(h)} \sim \mathsf{L}(0, \sigma_h)$ with $\sigma_0 := 1$ and
  $F_{Q,h}(x) = F_Q(x/\sigma_h)$, where $F_Q(u) = 1/\big(1+\exp(-u)\big)$.

HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

MATH STAT

Hidden-Markov Models for Ordinal Time Series

Stochastic Properties

Stochastic Properties of Ordinal HMMs

Marginal and bivariate pmf via (Zucchini et al., 2016)

$$P(X_t = x) \;=\; \mathbf{1}^\top \mathbf{P}(x)\,\boldsymbol{\pi} \qquad \text{and}$$

$$P(X_t = x, X_{t-k} = y) \;=\; \mathbf{1}^\top \mathbf{P}(x)\,\mathbf{A}^k\,\mathbf{P}(y)\,\boldsymbol{\pi},$$

where $\mathbf{P}(x) := \text{diag}\big(p(x|0), \ldots, p(x|d_{\mathcal{H}})\big)$ for $x \in \mathcal{S}$.

$\Rightarrow$ Location via median, dispersion via IOV $= \frac{4}{d} \sum_{i=0}^{d-1} f_i(1-f_i)$,

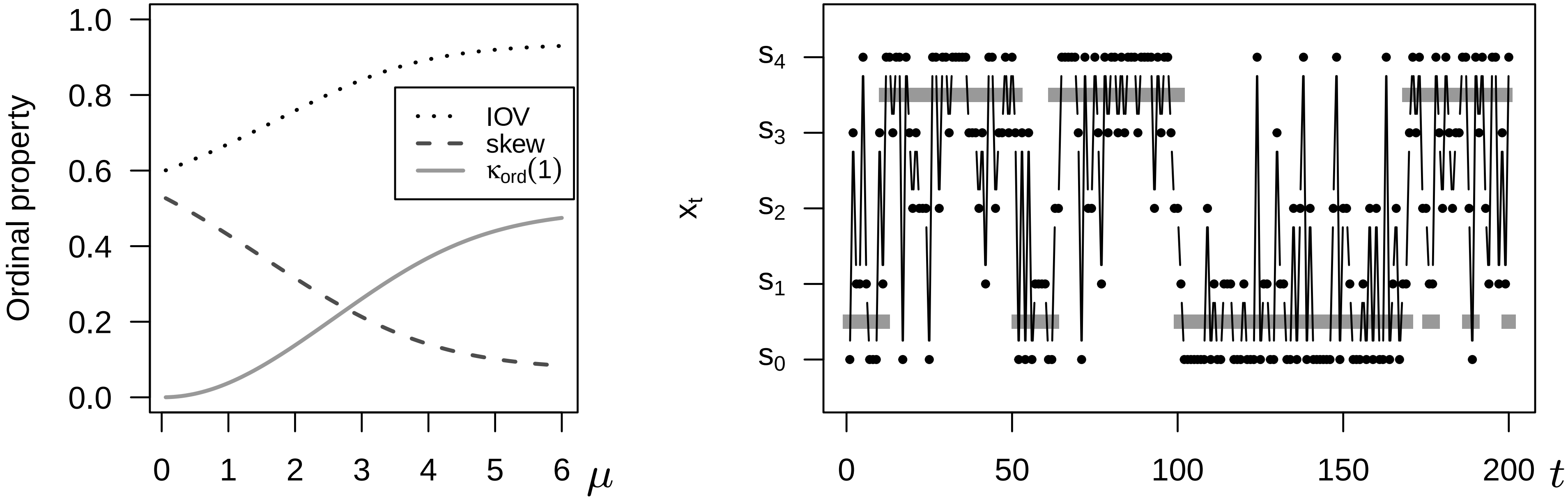
(a)symmetry via skew $= \frac{2}{d} \sum_{i=0}^{d-1} f_i \; - 1$,

and serial dependence via (Weiß, 2020)

$$\kappa_{\text{ord}}(k) \;=\; \frac{\sum_{j=0}^{d-1} \big(f_{jj}(k) - f_j^2\big)}{\sum_{i=0}^{d-1} f_i(1-f_i)}, \qquad f_{ij}(k) = P(X_t \leq s_i, X_{t-k} \leq s_j).$$

2-state logit(μ) **HMM** with range $\mathcal{S} = \{s_0, \ldots, s_4\}$,

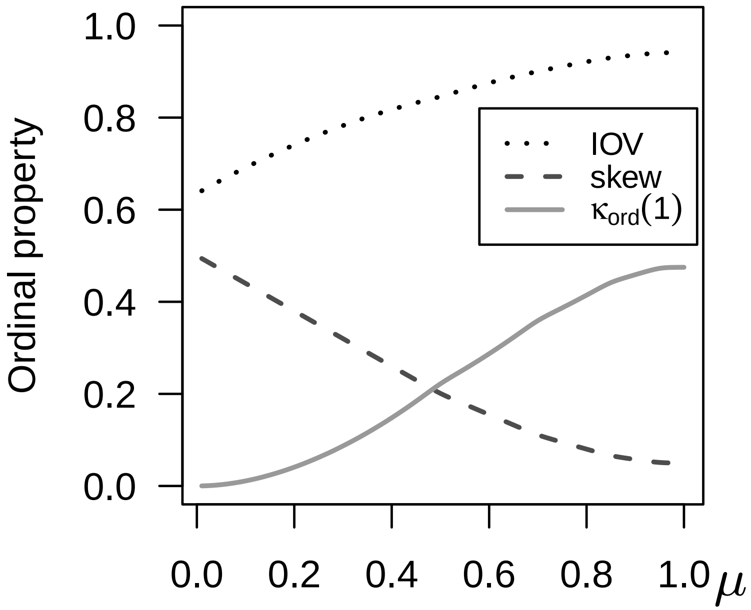
$\mathbf{A} = \begin{pmatrix} 0.97 & 0.07 \\ 0.03 & 0.93 \end{pmatrix}$, $\boldsymbol{\eta} = (0, 1, 2, 3)^\top$, $\mu_1 = \mu > 0$ ($\mu_1 = 3$):

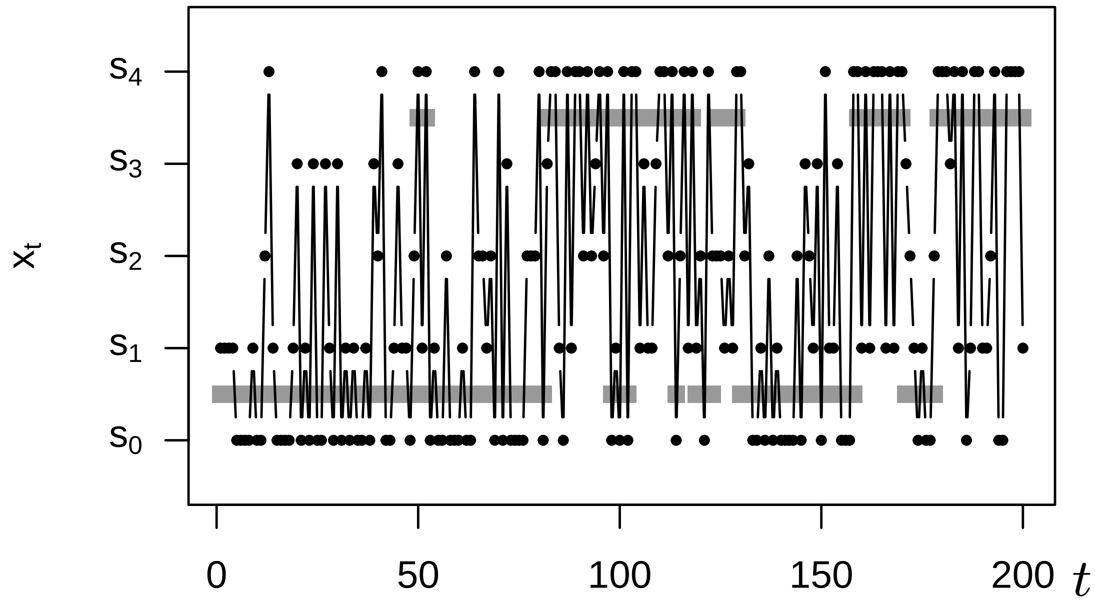


2-state soft-clipping(μ) **HMM** with $\mathcal{S} = \{s_0, \ldots, s_4\}$,

$\mathbf{A} = \binom{0.97\ 0.07}{0.03\ 0.93}$, $\boldsymbol{\eta} = (0.5, 0.7, 0.85, 0.95)^\top$,

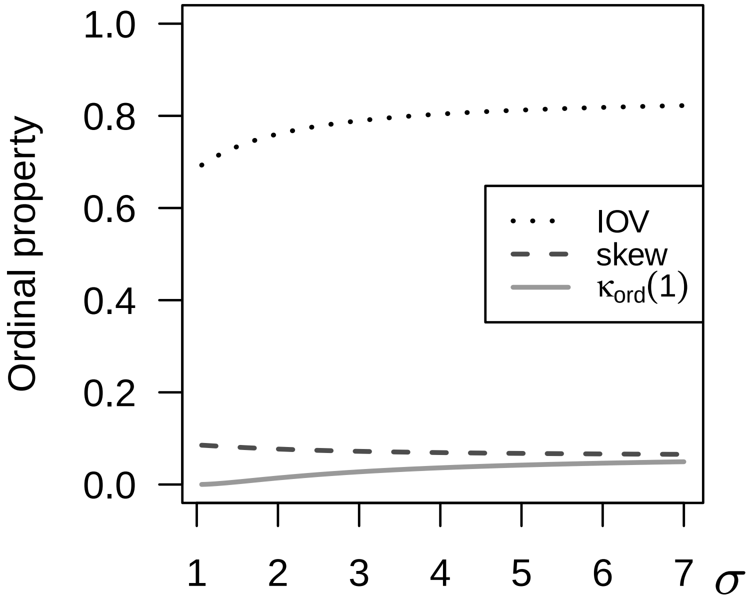
$\mu \in (0;1)$ ($\mu = 0.45$), and $\delta = 0.01$ ($\approx$ piecewise linear):

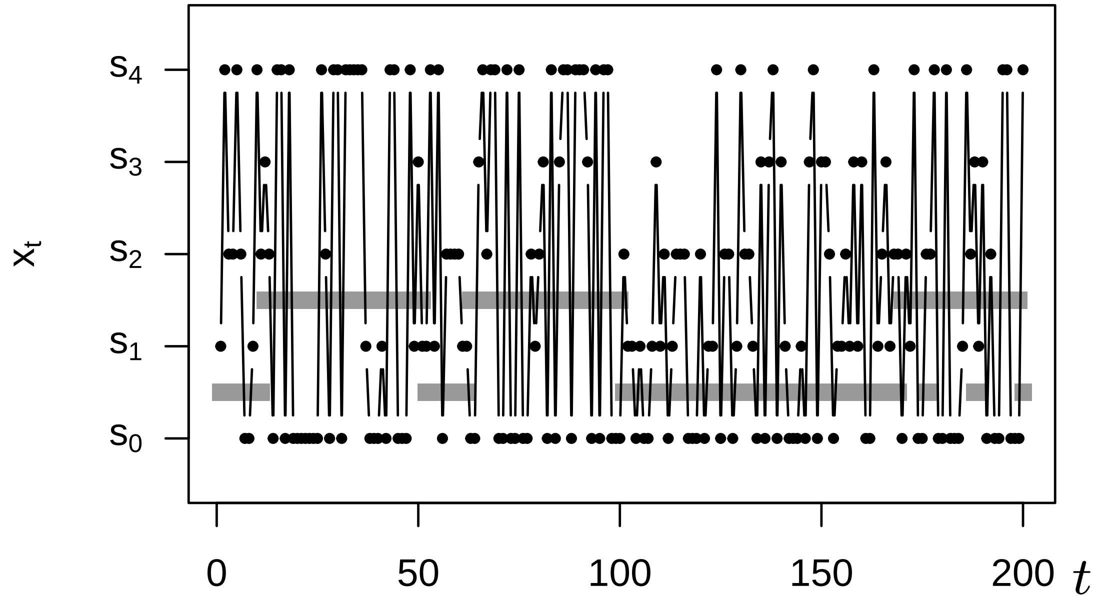




2-state logit(σ) **HMM** with range $\mathcal{S} = \{s_0, \ldots, s_4\}$,

$\mathbf{A} = \begin{pmatrix} 0.97 & 0.07 \\ 0.03 & 0.93 \end{pmatrix}$, $\boldsymbol{\eta} = (-1.5, -0.5, 1, 2)^\top$, $\sigma \in (1;7)$ $(\sigma = 3)$:





HELMUT SCHMIDT UNIVERSITÄT
Universität der Bundeswehr Hamburg

MATH STAT

Hidden-Markov Models for Ordinal Time Series

ML Estimation

Direct maximization of log-likelihood feasible, see Bulla & Berzel (2008), MacDonald (2014), Zucchini et al. (2016): compute **forward probabilities** $\boldsymbol{\alpha}_t = (\alpha_{t,0},\ldots,\alpha_{t,d_{\mathcal{H}}})^\top$ with

$$\alpha_{t,h} \;=\; P(X_t = x_t,\ldots,X_1 = x_1,\; H_t = h)$$

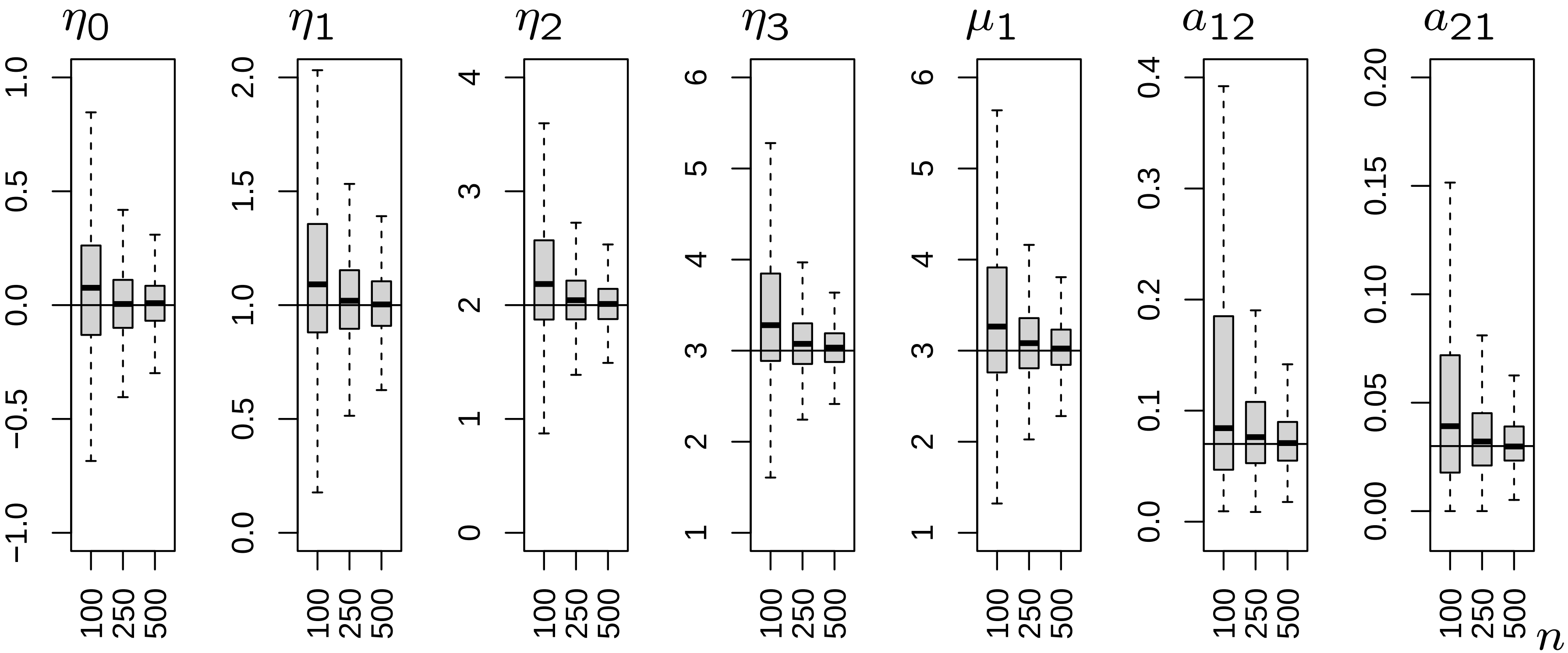
recursively via

$$\boldsymbol{\alpha}_1 \;=\; \mathbf{P}(x_1)\,\boldsymbol{\pi}, \quad \boldsymbol{\alpha}_t \;=\; \mathbf{P}(x_t)\,\mathbf{A}\,\boldsymbol{\alpha}_{t-1}, \qquad L(\boldsymbol{\theta}) = \mathbf{1}^\top\boldsymbol{\alpha}_n.$$

Also useful for **forecasting:**

$$P(X_{t+k} = x \mid x_t,\ldots,x_1) \;=\; \frac{\mathbf{1}^\top\,\mathbf{P}(x)\,\mathbf{A}^k\,\boldsymbol{\alpha}_t}{\mathbf{1}^\top\boldsymbol{\alpha}_t}.$$

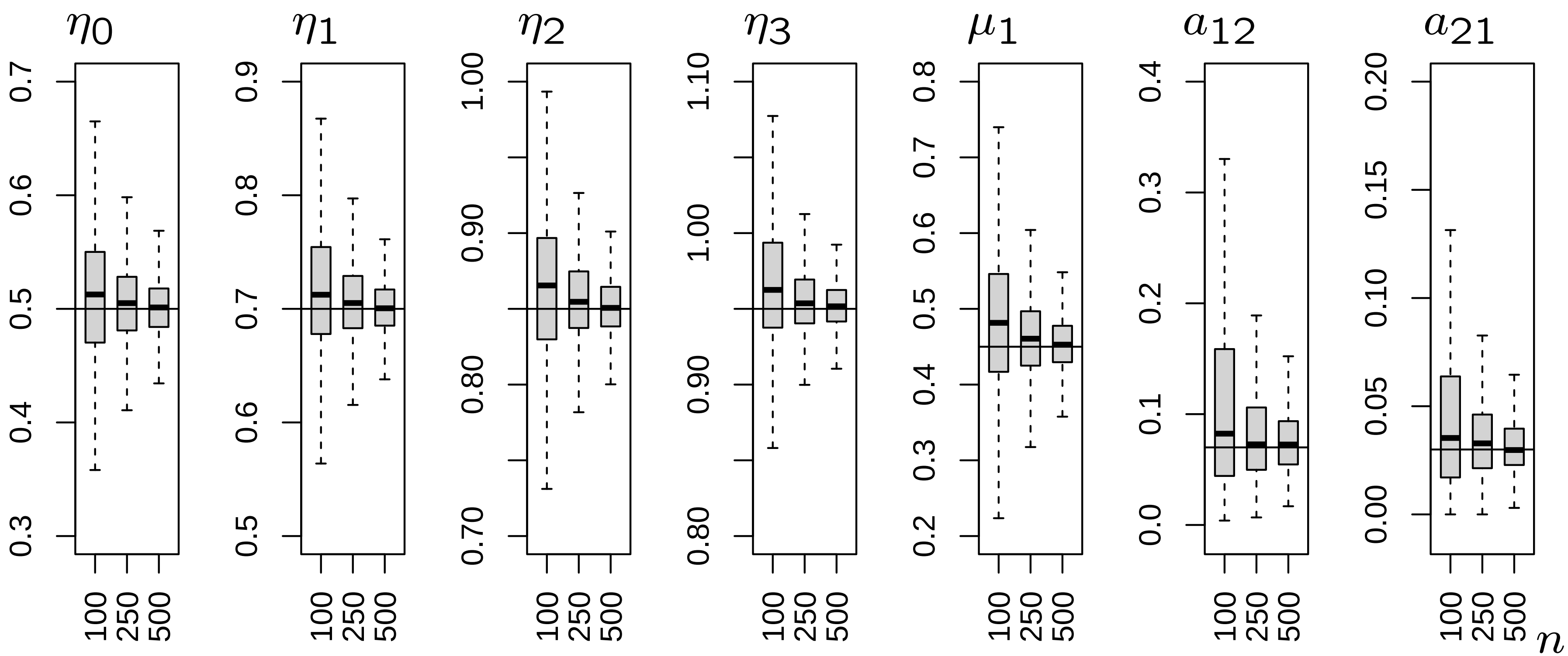
Performance analysis via simulations, illustrative boxplots:



2-state logit(μ) HMM with $a_{12} = 0.07$, $a_{21} = 0.03$,

and with $\boldsymbol{\eta} = (0, 1, 2, 3)^\top$ and $\mu_1 = 3$.

Performance analysis via simulations, illustrative boxplots:



2-state soft-clipping(μ) HMM with $a_{12} = 0.07$, $a_{21} = 0.03$,

and with $\boldsymbol{\eta} = (0.5, 0.7, 0.85, 0.95)^\top$ and $\mu_1 = 0.45$.

ML Estimation of Ordinal HMMs

- Further simulation scenarios in Weiß & Swidan (2024).

- Certainly, worse estimation performance for 3-state HMMs, with 12 instead of 7 parameters.
  But reasonably well already for $n = 250$.

- Only problematic of a_{ij} very close to zero.
  But well known for HMMs "when one fits models with three or more states to relatively short series", namely "that the estimates of one or more of the transition probabilities turn out to be very close to zero" (Zucchini et al., 2016, p. 55).

Helmut Schmidt Universität
Universität der Bundeswehr Hamburg

MATH
STAT

Perceived Stress of Migraine Patients

Data Application

Data made available by Curelator Inc., who developed mobile app "N1-Headache™" for use by **migraine patients**. *Each day*, users asked "to log information about [their] headaches, migraine symptoms and medication use on days [they] have an attack, and track a range of factors (moods, weather, diet, etc.) on a daily basis that may influence [their] risk of attack" (`https://n1-headache.com/patients/faq/`).

The authors are very grateful to Curelator Inc. (`https://n1-headache.com/`) for providing the N1-Headache™ data, and to Dr. M. Vives-Mestres and Dr. A. Casanova for their support and for fruitful discussions about these data.

Ordinal HMMs: Data Application

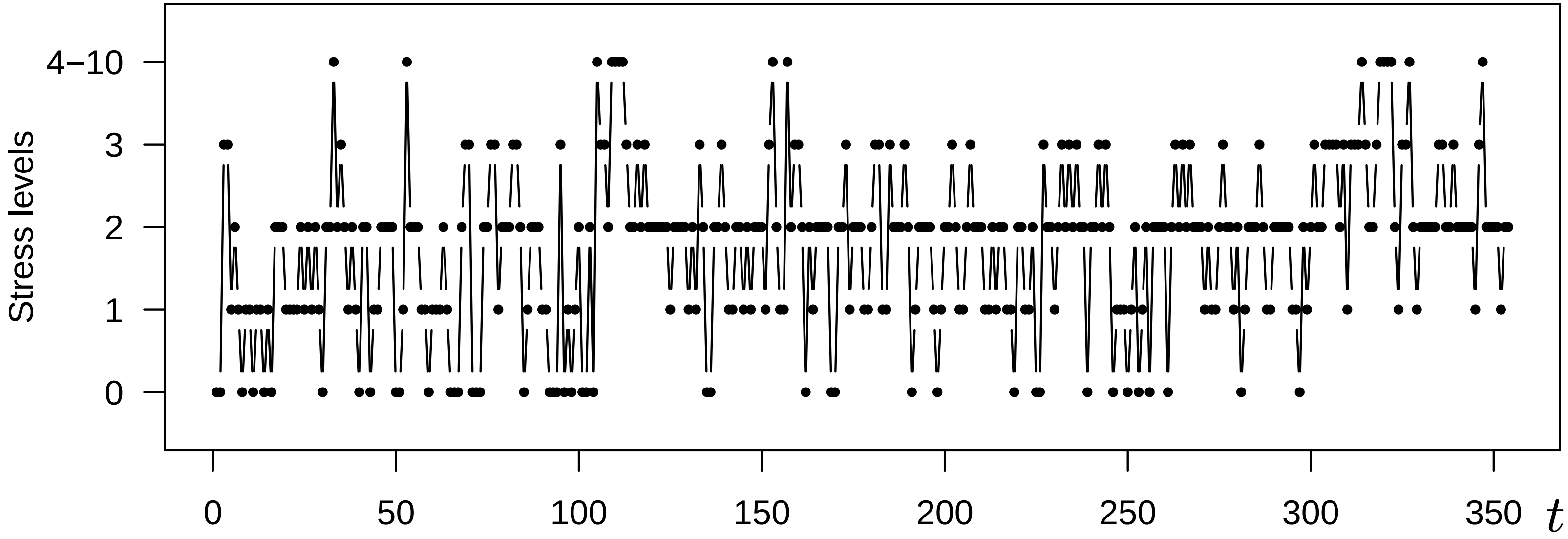
We focus on one of monitored emotional features:
level of perceived stress, also see Weiß (2021).

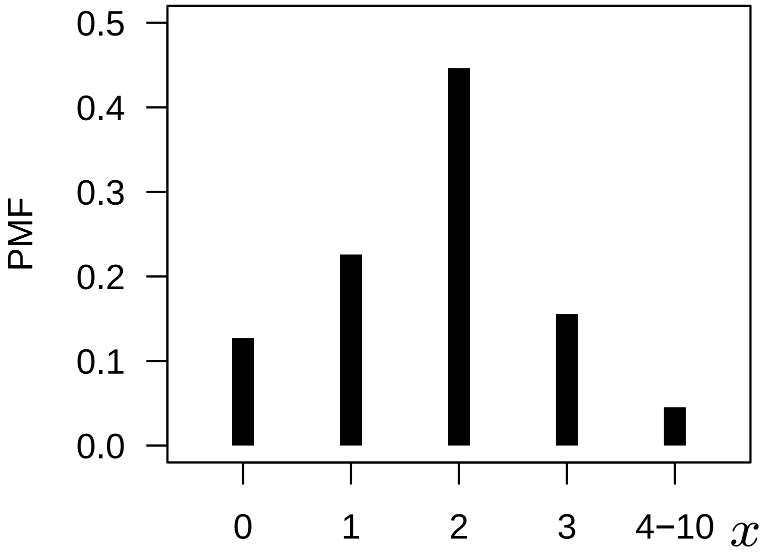
"How stressed have you felt today?" — expressed on 0–10 Likert scale, ranging from "not at all" to "a lot".

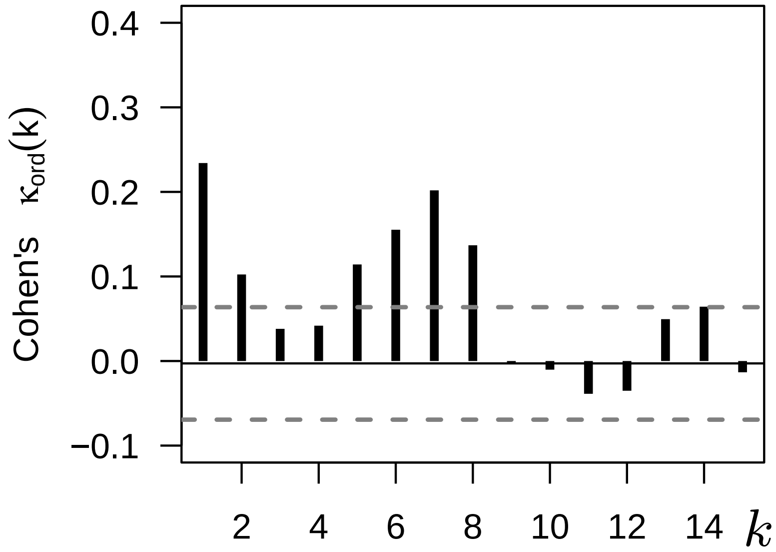
Stress is known trigger of migraine attacks and also associates with severity of attack (Vives-Mestres et al., 2021).

Illustrative example: daily stress level of a migraine patient on $n = 354$ successive days,
levels 4–10 aggregated into (maximal) category s_4.

Ordinal HMMs: Data Application







Competitors for model fitting:

discrete autoregressive (DAR) models (Weiß, 2018, Sec. 7.2),

fully non-parametric HMMs from "`hmm.discnp`" (Turner, 2022).

Fitted 2-state logit(μ) HMM:

$\hat{\mathbf{A}} \approx \left(\begin{smallmatrix} 0.828 & 0.283 \\ 0.172 & 0.717 \end{smallmatrix}\right)$ $\hat{\boldsymbol{\eta}} \approx (-1.356, 0.279, 9.393, 11.507)^\top$,

$\hat{\mu}_1 \approx 9.505$, leading to state-dependent distributions

$$\begin{array}{lcl} p(x|0) & \approx & 0.205, 0.364, 0.431, 0.000, 0.000, \\ p(x|1) & \approx & 0.000, 0.000, 0.472, 0.409, 0.119, \end{array} \quad \text{for } x = s_0, \ldots, s_4.$$

Stationary marginal distribution of H_t: $(0.622, 0.378)^\top$.

AIC ≈ 954.0 and BIC ≈ 981.1 clearly outperform DAR models.

Ordinal HMMs: Data Application

Full 2-state HMM has nearly same stochastic properties, but 10 instead of 7 parameters and hence worse AIC ≈ 959.7 and BIC ≈ 998.4.

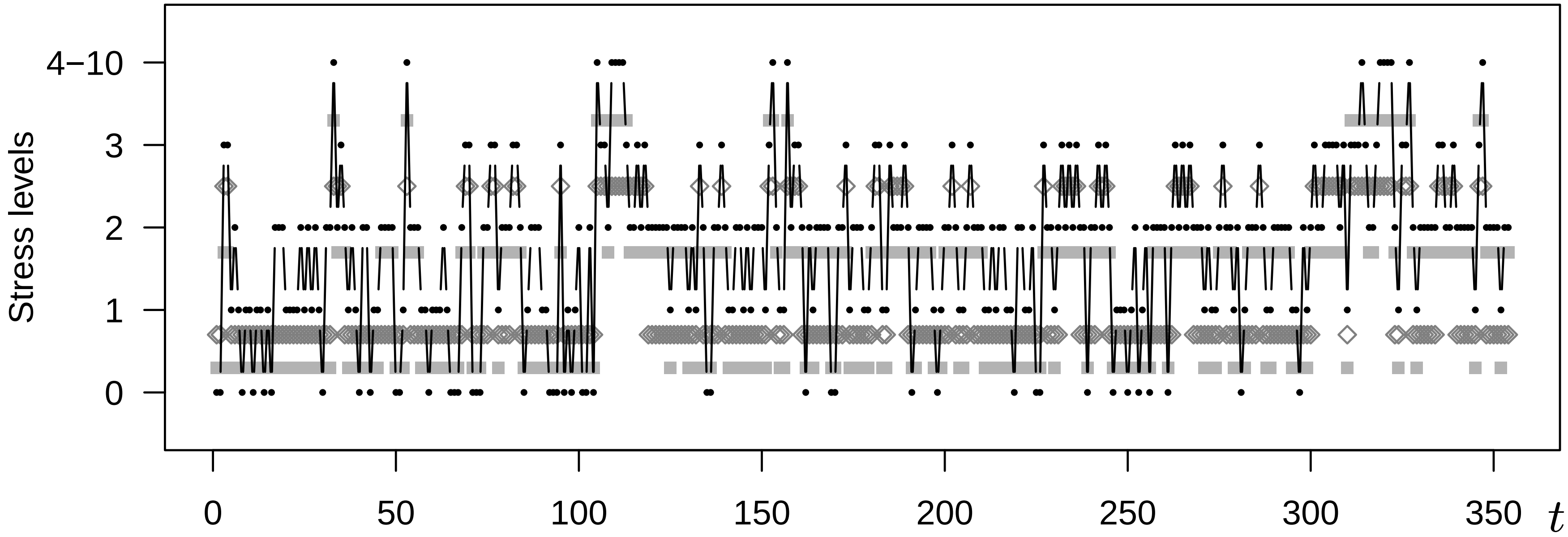
Fitting **2-state soft-clipping**(μ) **HMM** with $\delta = 0.01$ leads to virtually identical model, so user may decide by practical aspects.

3-state logit(μ) **HMM** with refined states:

$$
\begin{array}{lcll}
p(x|0) & \approx & 0.243, 0.435, 0.322, 0.000, 0.000 & \text{with median } s_1, \\
p(x|1) & \approx & 0.000, 0.000, 0.694, 0.304, 0.002 & \text{with median } s_2, \\
p(x|2) & \approx & 0.000, 0.000, 0.003, 0.423, 0.575 & \text{with median } s_4.
\end{array}
$$

Improved AIC ≈ 946.7, worse BIC ≈ 993.2.

Global decoding using 2 (diamonds) or 3 states (squares):



Interpretation of hidden states as propensity towards low, medium (only 3-state HMM), or high stress levels.

Checks of **model adequacy** slightly prefer 3-state HMM.

Ordinal HMMs: Data Application

Only deficiency: data exhibit (mild) weekly seasonal pattern, also see dominant lag 7 in plot of $\kappa_{\text{ord}}(k)$.

Weekly seasonality for perceived stress plausible, e. g., with tendency to higher stress levels during working days.

Feasible solution (Zucchini et al., 2016, Section 10.2.1): include covariates into state-dependent distributions ($\rightarrow$ GLM).
ML estimation still possible by forward approach,
just use time-dependent $\mathbf{P}_t(\cdot)$ where

$$f_t(s_j|h) = P(X_t \leq s_j \mid H_t = h, \boldsymbol{z}_t) \quad = F_{Q,h}(\eta_j + \boldsymbol{\gamma}^\top \boldsymbol{z}_t).$$

For perceived stress data, we use **harmonic component**

$$f_t(s_j|h) = F_{Q,h}\Big(\eta_j + a\,\cos(\tfrac{2\pi}{7}\,t) + b\,\sin(\tfrac{2\pi}{7}\,t)\Big).$$

Similar ML estimates for $\mathbf{A}$, $\boldsymbol{\eta}$, and $\boldsymbol{\mu}$ as before,

but $(\widehat{a},\widehat{b}) \approx (-0.013, -1.200)$ for 2-state HMM,

and $(\widehat{a},\widehat{b}) \approx (0.043, -1.277)$ for 3-state HMM.

Considerable improvement of information criteria:

AIC ≈ 917.8 and BIC ≈ 952.7 for 2-state HMM,

and AIC ≈ 908.7 and BIC ≈ 962.9 for 3-state HMM.

Also very good adequacy checks (PIT histogram, etc.).

- Novel class of HMMs for ordinal time series,
  help to understand their structure and behavior,
  more parsimonious than ordinary categorical HMMs.
- Stochastic properties, ML estimation,
  various special cases for incorporated GLM structure.
- Successfully applied to perceived stress time series,
  extension to non-stationary ordinal HMM.

Future research:

- Forecasting ordinal HMMs, effect of parameter estimation;
- control charts for monitoring ordinal HMMs.

Thank You for Your Interest!

Christian H. Weiß

Department of Mathematics & Statistics

Helmut Schmidt University, Hamburg

`weissc@hsu-hh.de`

HELMUT SCHMIDT UNIVERSITÄT

Universität der Bundeswehr Hamburg

MATH STAT

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), Projektnummer 516522977.

Weiß & Swidan (2024) Hidden-Markov models for ordinal time series. *AStA Advances in Statistical Analysis*, under review.

Agresti (2010) *Analysis of Ordinal Categorical Data*. 2nd ed., Wiley.

Bulla & Berzel (2008) Computational issues in . . . *Comp Stat* **23**, 1–18.

MacDonald (2014) Numerical maximisation . . . *Int Stat Rev* **82**, 296–308.

Turner (2022) `hmm.discnp`: Hidden Markov . . . *R package*, version 3.0-9.

Vives-Mestres et al. (2021) Perceived stress . . . *Headache* **61**, 1245–1254.

Weiß (2018) *An Introduction to Discrete-Valued Time Series*. Wiley.

Weiß (2020) Distance-based analysis . . . *JASA* **115**, 1189–1200.

Weiß (2021) Analyzing categorical time series . . . *Stat Med* **40**, 4675–4690.

Zucchini et al. (2016) *Hidden Markov Models for T.S.* 2nd ed., Chapman.