

Ordinale Zeitreihen: Modellierung, Vorhersage und Kontrolle

Gefördert
durch die
IFF2021

Ordinale Zeitreihen

Eine Zeitreihe x_1, \dots, x_T ist eine zeitlich geordnete Folge von Beobachtungen, von der man annimmt, dass sie durch einen zugrundeliegenden stochastischen Prozess $(X_t)_\mathbb{Z}$ generiert wurde. Für eine umfassende Behandlung von Zeitreihen orientiert man sich am (nach den Autoren des Buches [BJ70] benannten) *Box-Jenkins-Schema*, welches zwischen den Phasen der eigentlichen *Zeitreihenanalyse* (Anwendung statistischer Werkzeuge auf die Zeitreihe), der *Zeitreihenmodellierung* (Identifikation, Schätzung und Validierung von Modellen für den zugrundeliegenden Prozess) und der anschließenden Anwendung des gefundenen Modells unterscheidet. Letztere besteht üblicherweise entweder in der *Vorhersage* zukünftiger Werte oder in der *Kontrolle* („Monitoring“) des weiteren Verlaufs (etwa mittels sogenannter Kontrollkarten).

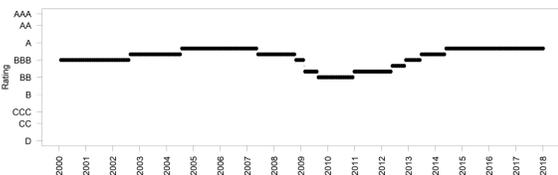


Abb. 1 Monatliches S&P Kredit-Rating von Lettland (2000–2017) [W20], von „D“ (Zahlungsverzug) bis „AAA“ (höchste Qualität).

Lange Zeit wurde dabei der Begriff „Zeitreihe“ automatisch mit „reellwertige Zeitreihe“ gleichgesetzt, d.h. man ging ganz selbstverständlich davon aus, dass der mögliche Wertebereich von Zeitreihen aus dem Kontinuum der reellen Zahlen (oder reeller Vektoren) besteht. Die in der Literatur verfügbaren Modelle und Methoden nutzten dabei die Eigenschaften reeller Zahlen redlich aus. Erst mit den 1980er Jahren begann sich die Erkenntnis durchzusetzen, dass es in der Praxis auch häufig diskretwertige Zeitreihen gibt, für welche die bis dahin entwickelten Ansätze kaum anwendbar sind. Vor allem im Bereich der Zählzeitreihen (bei denen die x_t Zählwerte sind) wurde seither ein umfassendes Werk an Beiträgen geschaffen, siehe [W18] für einen Überblick.

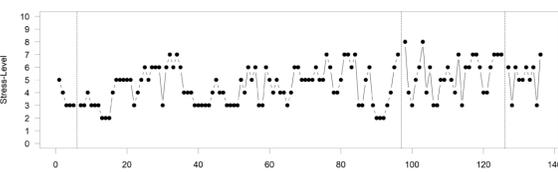


Abb. 2 Tägliches Ausmaß an Stress einer Migränepatientin (Likert-Skala 0–10; Fehldaten markiert) [W21].

Doch auch Zählzeitreihen stellen nur einen Teil der möglichen diskretwertigen Zeitreihen dar. Im vorliegenden Projekt geht es um eine sehr praxisrelevante Art von kategorialen Zeitreihen, nämlich um *ordinale Zeitreihen*. Diese besitzen einen qualitativen Wertebereich $\{s_0, \dots, s_m\}$, deren Ausprägungen eine natürliche Ordnung aufweisen, nämlich $s_0 < \dots < s_m$.

Beispiele ordinaler Zeitreihen sind in Abb. 1–2 zu sehen, bei welchen stets die Analyse und Modellierung von praktischer Relevanz wäre. Allerdings sind dafür weder existierende Methoden für nominale Zeitreihen noch für Zählzeitreihen geeignet, da diese entweder die Anordnung der Kategorien nicht berücksichtigen und z.B. in der Vorhersage zu unplausiblen „Sprüngen“ führen können, oder bei allen Teilen des Box-Jenkins-Schemas einen quantitativen Wertebereich erfordern (die Werte müssten z.B. additiv sein). Benötigt wird stattdessen ein „*ordinales Box-Jenkins-Schema*“, welches zwischen dem nominalen und dem Zählfall zu verorten ist. Doch trotz des großen Anwendungspotentials gibt es bis dato nur vereinzelte Beiträge, die sich zwar in ein solches Schema einfügen ließen, jedoch kein geschlossenes Rundumkonzept darstellen.

Fernziel: DFG-Projekt

In eigenen Vorarbeiten konnte bereits einer der vier Teilbereiche des angestrebten ordinalen Box-Jenkins-Schemas ausgearbeitet werden, die *Analyse ordinaler Zeitreihen*. In [W19] und [W20] wurden hierbei die nötigen Werkzeuge entwickelt, um Streuung, Gestalt und serielle Abhängigkeit von ordinalen Zeitreihen eingehend zu analysieren. Das angestrebte DFG-Projekt soll daher aus den folgenden drei Teilprojekten bestehen:

1. Es soll ein Baukasten aus *Modellen* für ordinale Zeitreihen aufgestellt werden, der sowohl bestehende wie auch neu zu entwickelnde Modelle umfasst.
2. Es sollen Methoden zur *Vorhersage* ordinaler Zeitreihen entwickelt und hinsichtlich ihrer Performanz untersucht werden.
3. Es sollen neue *Kontrollkarten* für ordinale Prozesse entwickelt und ihre Performanz hinsichtlich gängiger Kriterien analysiert und mit schon vorhandenen Kontrollkarten verglichen werden.

IFF-Projektidee

Ziel des IFF-Projektes ist es, den DFG-Antrag gründlich vorzubereiten und die Antragsfähigkeit herzustellen. Neben umfassenden Literaturrecherchen zu den o.g. drei Teilprojekten ist das Hauptziel, eine Vergleichsstudie bereits vorhandener Konzepte zur Kontrolle seriell unabhängiger Ordinaldaten durchzuführen, ergänzt um erste eigene Vorschläge basierend auf Statistiken aus [W19, W20]. Die gewonnenen Erkenntnisse sollen als weitere repräsentative Vorarbeit für den DFG-Antrag dienen.

Erzielte Resultate

Die einjährige IFF-Förderung ermöglichte die Finanzierung einer Stelle, welche mit einem Nachwuchswissenschaftler besetzt wurde.

Nach der Entwicklung von Kriterien zu Klassifikation und Vergleich ordinaler Kontrollkarten wurde eine umfassende vergleichende Simulationsstudie mit teils überraschenden Resultaten durchgeführt (welche auch erste neue Vorschläge für Kontrollkarten umfasste). Auch wurden die betrachteten Karten auf ein reales Datenbeispiel aus der Fertigungsindustrie angewandt, siehe Abb. 3. Alle Ergebnisse wurden gemeinsam mit dem Kooperationspartner, Prof. Dr. Murat C. Testik (Ankara), in einem Arbeitspapier zusammengefasst, welches zwischenzeitlich im international sehr renommierten *Journal of Quality Technology* zur Publikation angenommen wurde [OWT23].

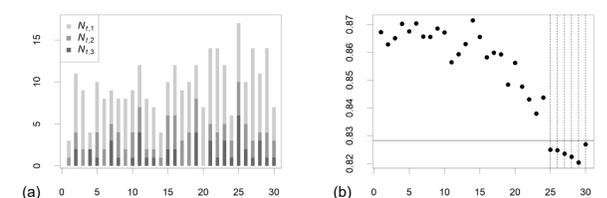


Abb. 3 (a) Häufigkeiten zur Schweißnaht bei je $n = 64$ gefertigten elektrischen Zahnbürsten (1=klein, 2=mittel, 3=groß; Rest ohne); (b) skew-EWMA-Karte mit Alarmzeitpunkten [OWT23].

Nächste Schritte

Mithilfe der IFF-Stelle konnte ein inhaltlich wesentlich erweiterter Projektantrag unter dem gleichlautenden Titel „*Ordinale Zeitreihen: Modellierung, Vorhersage und Kontrolle*“ erarbeitet werden, in dessen Rahmen die zuvor skizzierten drei Teilprojekte umfassend bearbeitet werden sollen. Zudem sollen zukünftige Forschungsarbeiten vorbereitet werden, etwa die Modellierung von zeitlich-räumlichen Ordinaldaten analog zu [JWK23]. Der Projektantrag (Sachbeihilfe) wurde am 15.06.2022 bei der DFG eingereicht und am 20.10.2022 bewilligt. Die Sachbeihilfe umfasst u.a. die Finanzierung einer Doktorand*innenstelle für 36 Monate. Der geplante Projektbeginn ist der 01.04.2023.

Projektpartner:

- Prof. Layth C. Alwan, Univ. Wisconsin, USA.
- Prof. Murat C. Testik, Univ. Ankara, Türkei.

Literatur:

- [BJ70] Box GEP, Jenkins GM (1970) *Time Series Analysis: Forecasting and Control*. 1st edition, Holden-Day, San Francisco.
- [JWK23] Jahn M, Weiß CH, Kim HY (2023) Approximately linear INGARCH models for spatio-temporal counts. *Journal of the Royal Statistical Society (Ser. C)*, DOI: 10.1093/jrsssc/qlad018.
- [OWT23] Ottenstreuer S, Weiß CH, Testik MC (2023) A review and comparison of control charts for ordinal samples. *Journal of Quality Technology*, DOI: 10.1080/00224065.2023.2170839.
- [W18] Weiß CH (2018) *An Introduction to Discrete-Valued Time Series*. John Wiley & Sons, Inc., Chichester.
- [W19] Weiß CH (2019) On some measures of ordinal variation. *Journal of Applied Statistics* **46**(16), 2905–2926.
- [W20] Weiß CH (2020) Distance-based analysis of ordinal data and ordinal time series. *Journal of the American Statistical Association* **115**(531), 1189–1200.
- [W21] Weiß CH (2021) Analyzing categorical time series in the presence of missing observations. *Statistics in Medicine* **40**(21), 4675–4690.