

Kohärente Vorhersage und Risikoanalyse für Zählzeitenprozesse

Gefördert
durch die
IFF2016

Prognose von Zählzeitenzeitreihen

Zeitreihen aus Zählzeiten entstehen durch das Abzählen von Ereignissen in aufeinanderfolgenden Zeiträumen, siehe etwa die Abb. 1 bis 3. Die Vorhersage derartiger Daten wird angestrebt, um Planungen zu ermöglichen (etwa Bedarfsplanungen wie in [AW17]), oder um sich auf mögliche Extremsituationen vorzubereiten [G11]. Solche Extremsituationen sind mit einem bewertbaren Risiko verbunden, und die Vorhersage dieses Risikos, also seines möglichen Ausmaßes sowie seiner Eintretenswahrscheinlichkeit, sind von äußerster praktischer Bedeutung.

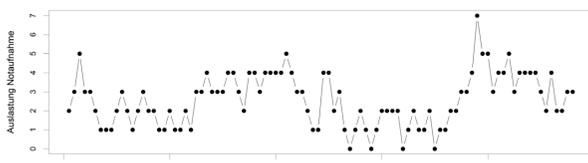


Abb. 1 Zahl belegter Betten (16.07.09, 08:00–23:59, 10-min) in Notaufnahme eines Kinderkrankenhauses [W13].

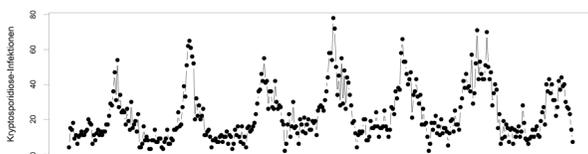


Abb. 2 Wöchentliche Zahl der Neuinfektionen mit Kryptosporidiose in Deutschland, 2002–2008 [W18].

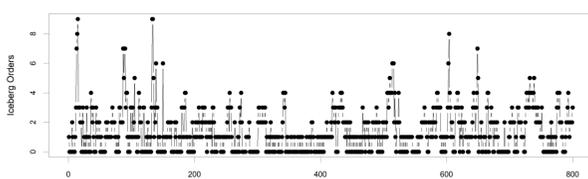


Abb. 3 Zahl der Iceberg Orders (1. Quartal 2004, 20-min) der Aktie der Deut. Telekom im XETRA-Handel [W15].

Zählzeitenzeitreihen haben zuletzt ein außerordentliches Forschungsinteresse auf sich gezogen [W18]. Sie können sehr unterschiedliche Charakteristika aufweisen (etwa einen stationären Verlauf mit autoregressiver Autokorrelationsstruktur wie in Abb. 1 und 3, oder einen nichtstationären Verlauf mit Trend und Saisonalität wie in Abb. 2), so dass mittlerweile eine große Zahl an entsprechenden Zeitreihenmodellen entwickelt wurde. Eines der grundlegendsten Modelle für Zählzeitenzeitreihen ist das INAR(1)-Modell von [McK85], definiert durch $X_t = \alpha \circ X_{t-1} + \varepsilon_t$ mit i.i.d. Zählzeitenzufallsvariablen $(\varepsilon_t)_{\mathbb{N}}$ und binomialer Ausdünnungsoperation „ $\alpha \circ$ “. Dieses Modell eignet sich für die Zeitreihe aus Abb. 1 [W13]. Für nichtstationäre Zeitreihen wie in Abb. 2 kann etwa das Poisson-Regressionsmodell von [ZQ88] verwendet werden, dessen bedingter Erwartungswert M_t definiert ist über

$$\ln M_t = \gamma^T \mathbf{Z}_t + \sum_{i=1}^p \alpha_i (\ln(X_{t-i} + 1) - \ln(\exp(\gamma^T \mathbf{Z}_{t-i}) + 1)),$$

wobei \mathbf{Z}_t die zur Zeit t verfügbare kovariante Information ist.

Prognosen von Zählzeitenzeitreihen wurden bisher oft mit Methoden ihres stetigen Pendantes approximiert, z.B. ein autoregressiver Zählzeitenprozess wie ein Gauß-AR-Prozess behandelt. Da eine auf diese Weise generierte Prognose i.Allg. jedoch keine nichtnegativen ganzzahligen Werte liefert, mussten zusätzlich noch Werte beschnitten und auf ganze Zahlen gerundet werden. Inwieweit die auf diese Weise gewonnenen Vorhersagen belastbar sind, ist jedoch unklar. [FM04] präsentieren ein kohärentes Prognoseverfahren, d.h. ein Verfahren, welches diskrete Vorhersagen generiert ohne runden zu müssen, indem die bedingte Wahrscheinlichkeitsverteilung $P(X_{t+h}|X_t, \dots)$ des Prozesses herangezogen und daraus z.B. der Median (Punktprognose) bzw. weitere Quantile (Bereichsprognose) berechnet wird.

IFF-Projektidee

Das Projekt befasst sich mit der kohärenten Prognose für Zählzeitenprozesse. Zur Generierung einer kohärenten Prognose werden jeweils ein modellbasierter und ein approximativ-prognoseansatz systematisch verglichen: „Liefert das modellbasierte Prognoseverfahren einen messbaren Mehrwert für die Vorhersage von Zählzeiten?“. Hierbei sind u.a. geeignete Kriterien zur Bemessung und zum effizienten Vergleich der Prognosegüte zu spezifizieren. Schließlich soll auch der Einfluss geschätzter Verteilungsparameter untersucht werden.

Extreme Quantile werden auch im Bereich der Risikoanalyse benötigt. Ziel ist die Vorhersage eines Risikos mittels prognostizierter Quantile und daraus abgeleiteter Risikomaße (auch unter Schätzunsicherheit), ferner die Bemessung der „Güte“ der Risikoprognose („Wie genau sind Prognose und prognostiziertes Risiko?“).

Erzielte Resultate

Die einjährige IFF-Förderung des Projektes ermöglichte die Zwischenfinanzierung einer Stelle, welche mit einer Nachwuchswissenschaftlerin besetzt wurde. Im Rahmen dieser Förderung wurden exemplarisch für Poisson-INAR(1)-Prozesse Konzepte entwickelt und Codes in der Programmiersprache R implementiert, mit denen die exakte h -Schritt-Vorhersageverteilung und diverse Gauß-AR(1)-Approximationen berechnet werden können. Insbesondere wurden diverse Kenngrößen und grafische Werkzeuge zur Bewertung der Güte der Approximation implementiert. Ein Teil der Resultate wurde von der geförderten Nachwuchswissenschaftlerin in dem Arbeitspapier [H17] zusammengefasst. Auch konnten diverse reale Datensätze gesammelt werden, welche die praktische Bedeutung der Vorhersageproblematik illustrieren.

Nächste Schritte

Mithilfe der IFF-Stelle konnte ein inhaltlich wesentlich erweiterter Projektantrag unter dem gleichlautenden Titel „Kohärente Vorhersage und Risikoanalyse für Zählzeitenprozesse“ erarbeitet werden, welcher die kohärente Prognose und Risikoanalyse in einer bisher nicht dagewesenen Breite beleuchten soll. Für den modellbasierten Ansatz soll eine Vielzahl an Modellen herangezogen werden, die weit über den Poisson-INAR(1)-Fall hinausgehen und somit die kohärente Prognose in einer bisher nicht dagewesenen Breite beleuchten. Neben INAR(p)-Modellen mit allgemeineren Randverteilungen (etwa zur Berücksichtigung der in der Realität weit verbreiteten Phänomene der Überdispersion und Nullinflation, vgl. auch [WHP16]) und Modellen für den Fall eines endlichen Trägers der Form $\{0, \dots, n\}$ (z.B. dem (beta-)binomialen AR(1)-Modell [McK85, WK14]) sollen insbesondere Modelle einbezogen werden, die jenseits ARMA-artiger Autokorrelationsstrukturen angesiedelt sind. Hierbei ist u.a. an Regressionsansätze [ZQ88] gedacht, mit denen beispielsweise Trend und Saisonalität wie in Abb. 2 berücksichtigt werden können.

Der Projektantrag (Sachbeihilfe) wurde am 31.05.2017 bei der Deutschen Forschungsgemeinschaft (DFG) eingereicht, und am 05.04.2018 von der DFG bewilligt. Die von der DFG bewilligte Sachbeihilfe umfasst u.a. die Finanzierung einer Doktorand(inn)enstelle auf die Dauer von 36 Monaten.

Projektpartner:

- Prof. Layth C. Alwan, Univ. Wisconsin, USA.
- Prof. Gabriel Frahm, HSU Hamburg,
- Prof. Rainer Göb, Universität Würzburg.

Literatur:

- [AW17] Alwan LC, Weiß CH (2017) INAR implementation of newsvendor model ... *Int. J. Prod. Res.* 55, 1085–1099.
- [FM04] Freeland RK, McCabe BPM (2004) Forecasting discrete valued low count time series. *Int. J. Forecasting* 20, 427–434.
- [G11] Göb R (2011) Estimating value at risk and conditional value at risk for count variables. *Qual. Reliab. Engng. Int.* 27, 659–672.
- [H17] Homburg A (2017) Criteria for evaluating approximations of count distributions. *Arbeitspapier*.
- [McK85] McKenzie E (1985) Some simple models for discrete variate time series. *Water Resour. Bull.* 21, 645–650.
- [W13] Weiß CH (2013) Integer-valued autoregressive models for counts showing underdispersion. *J. Appl. Statist.* 40, 1931–1948.
- [W15] Weiß CH (2015) A Poisson INAR(1) model with serially dependent innovations. *Metrika* 78, 829–851.
- [W18] Weiß CH (2018) *An Introduction to Discrete-Valued Time Series*. John Wiley & Sons, Inc., Chichester.
- [WHP16] Weiß CH, Homburg A, Puig P (2016) Testing for zero inflation and overdispersion in INAR(1) models. *Statist. Papers*.
- [WK14] Weiß CH, Kim HY (2014) Diagnosing and modelling extra-bin. variation ... *Appl. Stoch. Mod. Bus. Ind.* 30, 588–608.
- [ZQ88] Zeger SL, Qaqish B (1988) Markov regression models for time series: a q.-l. approach. *Biometrics* 44, 1019–1031.