

Nichtparametrische Modellierung von Zeitreihen mit Infektionshäufigkeiten

von Christoph Schürmann

Kurzfassung

Die Analyse der Häufigkeiten, mit der eine Infektionskrankheit auftritt, ist eine wesentliche Voraussetzung für die epidemiologische Bewertung und effiziente Überwachung dieser Krankheit innerhalb eines leistungsfähigen Gesundheitssystems. Zeitreihen mit wöchentlichen Meldefallhäufigkeiten sind dabei von besonderer Bedeutung. In dieser Arbeit wird dazu ein neues Verfahren basierend auf Dekompositionsmodellen vorgestellt, deren Komponenten mit nichtparametrischen Methoden geschätzt werden. Am Beispiel der drei in Nordrhein-Westfalen im Zeitraum 2001 bis 2006 am häufigsten gemeldeten Infektionskrankheiten, Campylobacteriose, Rotavirus-Infektion und Salmonellose, wird gezeigt, dass dieser Ansatz eine gute und sinnvolle Modellierung ermöglicht.

Grundlegend ist zunächst die Annahme, dass die Beobachtungen, ggf. nach geeigneter Transformation, als Summe eines Signals und eines zufälligen Fehlers in Form einer stationären Zeitreihe angesehen werden können. Das Signal wird dann in weitere deterministische Komponenten zerlegt, die einen Trend, saisonale, zyklische sowie kalenderbedingte Effekte beschreiben. Aus den inhaltlichen Eigenschaften dieser Bestandteile lassen sich strukturelle Bedingungen an Schätzer ableiten. Es wird untersucht, mit welchen nichtparametrischen Methoden wie Kernschätzern mit lokaler Bandbreite, Adaptive Weights Smoothing Splines, Straffe Saite oder Singular Spectrum Analysis Kandidatenfunktionen für die Schätzung berechenbar sind, die diese Anforderungen erfüllen. Anhand weiterer Kriterien, die für die Komponenten individuell formuliert werden, kann die Schätzung unter den jeweiligen Kandidaten bestimmt werden. Anschließend wird gezeigt, dass die nach Subtraktion der geschätzten Komponenten signalbereinigte Zeitreihe als Realisation eines ARMA-Prozesses niedriger Ordnung angesehen werden kann. Die geschätzten Modelle weisen insgesamt eine hohe Anpassungsgüte für die Beobachtungen und eine geeignete Glattheit auf. Zudem erlauben sie eine intuitive und aussagekräftige Interpretation der Daten.

Zum Vergleich mit einem häufig gewählten Ansatz wird die Modellierung durch ein parametrisches Poissonregressionsmodell herangezogen, in dem geeignete Kovariablen schrittweise durch ein Modellselektionskriterium in die Regressionsgleichung aufgenommen werden. Dieser Ansatz besitzt den Nachteil, dass die möglichen Kovariablen nicht ohne strukturelle Annahmen gewählt werden können und führt bei den betrachteten Beispielen zu einer Überanpassung der Schätzung an die Daten.

Möglichkeiten zur Vorhersage der Häufigkeiten werden auf Grundlage der nichtparametrischen Schätzung ebenfalls aufgezeigt.