

Ulrike Schach

Vector-Autoregressive Time Series Models with Spatial Dependence

Kurzfassung der Dissertation

Das Thema der Dissertation ist die Modellierung räumlich und zeitlich abhängiger Daten. Der Arbeit zugrunde liegen Lungen- und Magenkrebs-Mortalitätszahlen über einen Zeitraum von 15 Jahren für die Landkreise der Bundesrepublik Deutschland. Diese Daten werden nach einer geeigneten Standardisierung zunächst als unabhängige multivariate Zeitreihen angesehen und analysiert. Weiterhin werden die Daten räumlich unter Anwendung von *Markov Random Field* Modellen geglättet. Diese räumlichen Modelle können auch als Modelle mit zufälligen Effekten für die jeweiligen räumlichen Einheiten des Untersuchungsgebiets angesehen werden. Üblich in diesem Zusammenhang sind Gauß- und Poisson-Modelle für die Modellierung der altersstandardisierten Raten bzw. der Mortalitätsanzahlen. Beide Modellierungsansätze werden verwendet, um anschließend die Ergebnisse zu vergleichen und zu diskutieren.

Da klassische statistische Schätzverfahren bei einer solch großen Anzahl von unbekanntem Parametern ohne geeignete Restriktionen schnell an ihre numerischen Grenzen stoßen, werden die Gebiets-spezifischen Effekte Bayesianisch modelliert und mit Hilfe von Markov Chain Monte Carlo (MCMC) Simulation geschätzt. In diesem Zusammenhang werden Grundlagen der Bayesschen Inferenz beschrieben und geeignete Computer-intensive Algorithmen dargestellt, um die a posteriori-Verteilungen der unbekanntem Parameter zu evaluieren. Dabei wird ein erst kürzlich veröffentlichtes Kriterium zur Modellwahl in Bayes Modellen, das *Deviance Information Criterion*, eingeführt, mit dem die räumlichen Modelle mit und ohne Kovariablen verglichen werden.

Nach einer separaten räumlichen und zeitlichen Analyse der Mortalitätsdaten wird mittels Vektor-autoregressiver Zeitreihenansätze eine kombinierte Raum-zeitliche Modellierung vorgestellt. Dabei wird für die jeweiligen Jahresscheiben innerhalb des autoregressiven Zeitreihenansatzes auch die räumliche Korrelation erfasst. Die Parameter dieses Raum-Zeit Modells werden mit einem klassischen Schätzverfahren nach der Maximum-Likelihood-Methode, und wiederum mit Bayesscher Inferenz geschätzt. Während das klassische Schätzverfahren schon bei einer relativ geringen Anzahl an Parametern an seine Grenzen stößt, können mit Hilfe von MCMC Methoden ohne weiteres auch zusätzliche Kovariablen in das Modell aufgenommen werden.

Aufgrund der Tatsache, dass die Krebs-Mortalitätsraten in Deutschland aus Datenschutzgründen nicht jährlich und auf der Basis von Landkreisen vorliegen, sondern nur entweder zeitlich oder räumlich aggregiert veröffentlicht werden, wird eine Modell-basierte Schätzung der fehlenden Daten mit der höchsten räumlichen und zeitlichen Auflösung präsentiert. In diesem Zusammenhang wird eine Methode vorgestellt, die unter Einhaltung der gegebenen zeitlichen und räumlichen Randsummen die fehlenden Daten als unbekanntem Parameter mit Bayes-Verfahren schätzt.