

Entwicklung statistischer Tests mit computergestützter Algebra

Kurzfassung der Dissertation

vorgelegt von

Anne Krampe

Fakultät Statistik

Technische Universität Dortmund

Im Fokus dieser Dissertation steht die Anwendung der computergestützten Algebra in der Statistik. Aus diesem Teilgebiet der Mathematik werden Methoden verwendet, um statistische Tests für eine Analyse struktureller Abhängigkeiten zwischen Komponenten eines multivariaten Zufallsvektors zu entwickeln.

Zur Beschreibung von Abhängigkeits- bzw. Unabhängigkeitsbeziehungen werden log-lineare sowie graphische Modelle herangezogen. Für die Beurteilung der Anpassungsgüte eines Modells werden Anpassungstests wie der χ^2 -Test und der Likelihood-Quotienten-Test verwendet. In der Praxis basieren diese traditionell auf asymptotischen Ergebnissen bzw. exakten Berechnungen, wobei die Approximation beispielsweise aufgrund eines zu geringen Stichprobenumfangs nicht gerechtfertigt bzw. ein entsprechender exakter Tests zu aufwändig sein kann. Diaconis und Sturmfels (1998) verbinden Methoden der computergestützten Algebra mit Markov Chain Monte Carlo-Verfahren, um aus der bedingten Verteilung einer diskreten Exponentialfamilie mit beobachteter suffizienter Statistik zu simulieren. Der Ansatz von Diaconis und Sturmfels wird in dieser Arbeit für die Entwicklung neuer algebraischer Tests genutzt.

Liegen, wie z. B. bei Gutachterverlässlichkeitsstudien, der Untersuchung so genannte gepaarte Beobachtungen zugrunde, so sind häufig strukturelle Übereinstimmungen der Realisationen von besonderem Interesse. Da das perfekte Symmetriemodell restriktiv ist und selten von einem Datensatz unterstützt wird, werden zusätzlich alternative Symmetriemodelle (bedingtes, diagonales, ordinales Quasi-Symmetriemodell) betrachtet. Für diese vier Modelle werden neue algebraische Anpassungstests vorgeschlagen.

Häufig ist für eine Analyse kategorialer Daten nicht nur von Bedeutung, ob Abhängigkeiten zwischen Komponenten eines Zufallsvektors bestehen, sondern es interessiert zudem, wie stark diese Abhängigkeiten tatsächlich sind. Das Odds Ratio und ein zugehöriges Konfidenzintervall geben hierzu Aufschluss. Ausgehend von dem Spezialfall einer 2×2 -Kontingenztafel wird ein algebraisches Konfidenzintervall für das Odds Ratio entwickelt, wobei das vorgestellte Konstruktionsprinzip beispielsweise auch für Konfidenzintervalle anderer Parameter wie das Relative Risiko oder das Odds Ratio einer $2 \times 2 \times K$ -Tafel, $K \in \mathbb{N}$, anwendbar ist.

Für die Analyse eines Datensatzes werden oftmals verschiedene Modelle geprüft. Algebraische Tests nach Diaconis und Sturmfels basieren auf eine separate Simulation für jedes betrachtete Modell, weshalb sie in der Regel zu aufwändig und daher für solche Fragestellungen ungeeignet sind. Weisen die interessierenden Modelle eine hierarchische Struktur auf, so kann der nötige Simulationsaufwand für algebraische Tests z. T. erheblich reduziert werden. Dazu wird bewiesen, dass die Simulation gemäß einem Modell ausreicht, um algebraische Tests für alle darauf aufbauenden Modelle durchzuführen. Das Vorgehen wird am Beispiel der Modellselektion für graphische Modelle sowie für die betrachteten Symmetriemodelle vorgestellt.

Liegen der Analyse kategoriale Daten zugrunde, so haben sich algebraische Tests als wichtige Ergänzung zu den traditionellen asymptotischen und exakten Tests erwiesen. Es wird weiter gezeigt, dass algebraische Anpassungstests für gemischt stetige-diskrete Daten prinzipiell entwickelt werden können. Der nötige Simulationsaufwand scheint jedoch sehr groß, so dass alternative Verfahren für eine Anwendung in der Praxis geeigneter erscheinen.