

Ein semi-bayes'scher Anpassungstest für das logistische Regressionsmodell mit schwach besetzten Zellen

Oliver Kuß

*Institut für Medizinische Epidemiologie, Biometrie und Informatik,
Martin-Luther-Universität Halle-Wittenberg, Halle/Saale

Das logistische Regressionsmodell ist inzwischen die Standardmethode in der Biometrie und Epidemiologie, wenn es um die Auswertung von binären Zielgrößen geht. Methoden zur Evaluierung der Anpassungsgüte (Goodness-of-Fit) im logistischen Regressionsmodell sind jedoch, im Vergleich zum z.B. dem linearen Regressionsmodell, immer noch unterentwickelt. Als globale Anpassungstests werden im allgemeinen die Devianz oder die Pearson-Statistik empfohlen, die auf bewährten Testprinzipien beruhen und auch im Rahmen anderer statistische Modelle Verwendung finden. Es ist jedoch bekannt [siehe z.B. McCullagh u. Nelder, 1986], dass diese beiden Tests in Situationen mit schwach besetzten Zellen, also z.B. bei stetigen Kovariablen oder einer großen Anzahl von Kovariablen, eher die Regel als die Ausnahme in realen Datensätzen, nicht zu verlässlichen Ergebnissen führen. Zur Lösung dieses Problems wird u.a. auch vorgeschlagen, die Prüfverteilung zu modifizieren und McCullagh, 1985, und Osius/Rojek, 1992 leiten solche Prüfverteilungen unter der Annahme von schwach besetzten Zellen und Maximum Likelihood-Schätzung her.

Wählt man zur Parameterschätzung jedoch nicht die Maximum-Likelihood-Methode, sondern einen bayes'schen Ansatz, z.B. den Metropolis-Hastings-Algorithmus von Gamerman, 1997, so kann man die Posterior-Verteilungen der beiden Test-Statistiken mitschätzen. Die Idee ist nun, die ersten asymptotischen Momente unter der Null-Hypothese aus der Standard-Theorie zu benutzen, deren Werte als Maß für eine gute Modellanpassung anzunehmen und diese als kritische Werte für das Quantil der Posterior-Verteilung zu benutzen.

Im Vortrag wird das Verfahren anhand eines Datensatzes zu berufsbedingten Handekzemen bei Friseuren dargestellt, eventuell liegen bis dahin auch schon erste Simulationsergebnisse für die Güte des Verfahrens unter Nullhypothese und Alternative vor.

Literatur:

- Gamerman D. Efficient sampling from the posterior distribution in generalized linear mixed models. *Statistics and Computing*, 7: 57-68, 1997.
- McCullagh P. On the Asymptotic Distribution of Pearson's Statistic in Linear Exponential-Family Models. *International Statistical Review*, 53:61-67, 1985.
- McCullagh P, Nelder JA. *Generalized Linear Models*. Chapman & Hall, 1989.
- Osius G, Rojek D. Normal Goodness-of-Fit Tests for Multinomial Models With Large Degrees of Freedom. *JASA*, 87:1145-1152, 1992.