

Robuste Statistik mit Kompositionsdaten

P. Filzmoser¹

¹ Technische Universität Wien, Institut für Statistik und Wahrscheinlichkeitstheorie, Wiedner Hauptstr. 8-10, A-1040 Wien, Österreich

Schlüsselwörter: Kompositionsdaten, Ausreißer, Robuste Statistik.

Kurzfassung

Kompositionsdaten kommen in vielen Gebieten vor, wie z.B. in der Geochemie, wenn Stichproben nach der Konzentration von chemischen Elementen analysiert werden (siehe z.B. Reimann et al., 2008). Wenn das Stichprobenmaterial zur Gänze analysiert wird, summieren die Datenwerte auf 100% auf (wenn sie in Gewichts-% gemessen wurden), oder auf einen konstanten Wert.

Allgemein kann eine Beobachtung von Kompositionsdaten, bestehend aus D Anteilen, definiert werden als

$$\mathbf{x} = (x_1, \dots, x_D)^t, x_i > 0, i = 1, \dots, D, \sum_{i=1}^D x_i = 1.$$

Die Bedingung von konstanter Summe hat den Effekt, dass die Daten nicht mehr in \mathbf{R}^D , sondern in einem Unterraum von \mathbf{R}^D liegen, dem sogenannten Simplex. Das hat bedeutende Konsequenzen für die statistische Analyse, und ein möglicher Lösungsweg ist, vor Anwendung einer statistischen Methode die Daten geeignet zu transformieren (Aitchison, 1986). Eine Schwierigkeit für den Einsatz robuster statistischer Methoden ist, dass einige dieser Transformationen zu singulären Daten führen.

Im Vortrag werden wir Probleme von robusten multivariaten Methoden für Kompositionsdaten diskutieren und Lösungswege aufzeigen (Filzmoser and Hron, 2008b; Filzmoser et al., 2008c, 2008d). Weiters wird eine Methode zur Ausreißerererkennung bei Kompositionsdaten vorgestellt (Filzmoser and Hron, 2008a).

Literatur

- J. Aitchison (1986). *The Statistical Analysis of Compositional Data*. Chapman & Hall, London.
- P. Filzmoser and K. Hron (2008a). Outlier detection for compositional data using robust methods. *Mathematical Geosciences*, 40(3), 233–248.
- P. Filzmoser and K. Hron (2008b). Correlation analysis for compositional data. *Mathematical Geosciences*. To appear.
- P. Filzmoser, K. Hron, and C. Reimann (2008c). Principal component analysis for compositional data with outliers. *Environmetrics*. To appear.
- P. Filzmoser, K. Hron, C. Reimann, and R.G. Garrett (2008d). Robust factor analysis for compositional data. Under revision.
- C. Reimann, P. Filzmoser, R.G. Garrett, and R. Dutter (2008). *Statistical Data Analysis Explained. Applied Environmental Statistics with R*. John Wiley & Sons, Chichester.