Multivariate Methoden: Übungsblatt 3

Aufgabe 1: Asymptotische Eigenschaften der Hauptkomponenten

Theorem Sei $\Sigma > 0$ mit voneinander verschiedenen Eigenwerten mit der Eigenwertzerlegung $\Sigma = \Gamma \Lambda \Gamma'$. Dann gilt

$$(a) \sqrt{n-1}(\hat{\lambda} - \lambda) \xrightarrow{\mathcal{L}} N_p(0, 2\Lambda^2),$$

$$(b) \sqrt{n-1}(\hat{\gamma}_j - \gamma_j) \xrightarrow{\mathcal{L}} N_p(0, 2\mathbf{V}_j),$$

mit $\mathbf{V}_j = \lambda_j \sum_{k \neq j} \frac{\lambda_k}{(\lambda_k - \lambda_j)^2} \gamma_k \gamma_k'$,

(c) die Elemente in $\hat{\lambda}$ sind von den Elementen von $\hat{\Gamma}$ asymptotisch unabhängig.

Leiten Sie die unter Verwendung des obigen Theorems die asymptotische Verteilung von $\log \ell_j$ her und konstruieren Sie daraus ein Konfidenzintervall für λ_j .

Aufgabe 2

2.1. Führen Sie eine Hauptkomponentenanalyse der Variablen des Datensatzes BANK.DTA durch (d.h. für alle Variablen außer derjenigen, die angibt, ob die betreffende Banknote echt ist). Gehen sie von der Hauptkomponentenanalyse mit der Kovarianzmatrix aus.

2.1,a) Wie groß ist die Bedeutung der "Länge"-Variable für die 1. Hauptkomponente? (Tipp: Berechnen Sie den relevanten Korrelationskoeffizient). Was impliziert es für die eventuelle Wichtigkeit dieser Variable?

2.2,b) Betrachten Sie die Anteile an erklärten Varianz und erstellen Sie ein scree-plot der Ergebnisse der Hauptkomponententransformation. Wieviele Hauptkomponenten würden Sie zur adäquaten Beschreibung der Daten für notwendig erachten?

2.3,c) Führen Sie rückwärts von $k = 2$ einen Selektionstest der Hypothese

$$H_{0k} : \text{Var}(Y_{p-k+1}) = \dots = \text{Var}(Y_p)$$

aus, und bestimmen Sie formal die Anzahl an relevanten Eigenwerten.

2.2. Führen Sie nun die gleiche Hauptkomponentenanalyse mit der Korrelationsmatrix durch.

2.2,a) Wie unterscheidet sich der 1. Eigenwert in dieser Hauptkomponententransformation von dem 1. Eigenwert aus der Transformation mit der Kovarianzmatrix?

2.2,b) Wieviel relevante Hauptkomponenten ergeben sich durch die Analyse mit der Korrelationsmatrix?

2.2,3) Sagen Sie die ersten zwei Hauptkomponenten vorher (Tipp: *predict var1 var2*).