Prof. Dr. U. Stadtmüller
M. Zampiceni

Sommer 2013
03.06.2013
Blatt 06

Funktionale Datenanalyse

(Abgabe: Mo., 10.06.2013, vor 13:15)

1. Wir betrachten den Datensatz `heptathlon` aus dem R-Packet `HSAUR`. Er enthält die Ergebnisse des Siebenkampffinales in Seoul 1988. Man möchte durch Hauptkomponentenanalyse die Leistungen der 25 Teilnehmerinnen 'kompakter' beschreiben. Bilde einen neuen Datensatz, der die Ergebnisse der einzelnen Wettbewerbe beinhaltet, aber nicht die erreichte Punktzahl, und dann plotte diesen Datensatz.

   (a) Führe zweimal eine Hauptkomponentenanalyse durch, einmal ohne zusätzliche Optionen und einmal mit `cor=T`. Im zweiten Fall basiert das Verfahren auf der (empirischen) Korrelationsmatrix anstatt auf der (empirischen) Kovarianzmatrix: Hiermit werden die Variablen im Prinzip skaliert (so, dass die Stichprobenvarianz immer 1 ist). Warum ist das zweite Modell hier sinnvoller?

   (b) Wie viele Hauptkomponenten sollte man betrachten?

   (c) Betrachte die Gewichte (`loadings`) der ersten Hauptkomponente. Welche Variablen (Disziplinen) haben den größten Einfluss, warum haben manche ein positives und manche ein negatives Vorzeichen? Gibt es Variablen, die die erste Komponente wenig beeinflussen? Wie kann man die 'Leistung' der 25 Athletinnen bezüglich der ersten Hauptkomponente bekommen?
   Betrachte nun die Gewichte der zweiten Hauptkomponente. Welche Variablen sind hier am wichtigsten?

   (d) Plotte die ersten zwei Komponenten in einem `biplot`: In einem gemeinsamen Schaubild werden die ersten zwei HKn vom Datensatz (mit einer Skala) zusammen mit der Projektion der ursprünglichen Variablen auf der HK1-HK2-Ebene (mit einer anderen Skala) dargestellt. Stimmt das Bild mit den Überlegungen aus (c) überein?

   (e) Überprüfe graphisch, ob die erste oder die zweite Komponente das Endergebnis beeinflussen.

(4 Punkte)

2. **(Vgl Vorlesung)** Sei $\{W(t), t \in [0,1]\}$ ein Wienerprozess. Insbesondere, für $s, t \in [0,1]$ sind dann $\mathbb{E}(W(t)) = 0$ und $C(s,t) := \mathrm{Cov}(W(s), W(t)) = \min(s,t)$. Der Operator $\mathcal{C}$ sei wie in der Vorlesung definiert:

$$\begin{aligned} \mathcal{C} : L_2([0,1]) &\to L_2([0,1]) \\ f(\cdot) &\mapsto \int_0^1 C(\cdot, t) f(t) dt \end{aligned}$$

   Man möchte die Eigenwerte λ und die Eigenfunktionen $\varphi(t)$ von $\mathcal{C}$ explizit bestimmen.

   (a) Leite aus $(\mathcal{C} \circ \varphi)(s) = \lambda \varphi(s)$ die Differentialgleichung:

$$-\varphi(s) = \lambda \varphi''(s) \qquad s \in [0,1] \tag{1}$$

und die Bedingungen:

$$\varphi(0) = 0 \tag{2}$$
$$\varphi'(1) = 0 \tag{3}$$

her.

(b) Löse die Differentialgleichung (1) mit dem Ansatz

$$\varphi(s) = c_1 \sin\left(\frac{s}{\sqrt{\lambda}}\right) + c_2 \cos\left(\frac{s}{\sqrt{\lambda}}\right).$$

Benutze hierfür die Bedingungen (2), (3), und $\|\varphi\|_2 = 1$. Ist die Lösung eideutig? (Man muss die drei Parameter c_1, c_2 und λ bestimmen.)

(c) Seien φ_k und λ_k die Eigenfunktionen und die Eigenwerte von $\mathcal{C}$. Wir wollen zeigen, dass

$$\xi_k = \langle W, \varphi_k \rangle_{L_2} \sim \mathcal{N}(0, \lambda_k) :$$

Für $n \in \mathbb{N}$, sei $X_n^{(k)} := \frac{1}{n} \sum_{j=1}^n W\left(\frac{j}{n}\right) \varphi_k\left(\frac{j}{n}\right)$.

- Zeige, dass $X_n^{(k)} \sim \mathcal{N}(0, \sigma_k^2(n))$
- Zeige, dass $\sigma_k^2(n) \to \lambda_k$, für $n \to \infty$
- Zeige, dass $X_n^{(k)} \overset{f.s.}{\to} \xi_k$ für $n \to \infty$.

(2+3,5+4,5 Punkte)