



Angewandte Stochastik 2 - Übungsblatt 5

Besprechung: 21. November im **R**-Tutorium.

Aufgabe 1 (8 Punkte)

In dieser Aufgabe untersuchen wir den Einfluss der Bandbreite h auf den Kerndichteschätzer.

- Schreibe eine Funktion in **R**, die den Funktionswert des Kerndichteschätzers für einen Vektor x an der Stelle \mathbf{t} mit Bandbreite h mithilfe des Epanechnikov Kerns bestimmt.
- Teste deine Funktion für eine Realisierung von $n = 200$ standardnormalverteilten Zufallsvariablen und Bandbreiten $h \in \{0.1, 0.2, 0.5, 1, 2, 4, 5\}$. Plote die Kerndichteschätzer inklusive der Dichte der Standardnormalverteilung in verschiedenen Farben in ein Koordinatensystem. Füge dem Plot eine passende Legende hinzu.
- Für die Wahl einer geeigneten Bandbreite h kann man wie folgt vorgehen. Sei \hat{f}_{ni}^h der Kerndichteschätzer mit Bandbreite h für den Vektor $x^{(i)} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$, d.h. den Vektor x ohne den i ten Eintrag. Als Bandbreite h_{opt} wählen wir den Wert h der

$$\prod_{i=1}^n \hat{f}_{ni}^h(x_i)$$

maximiert. Implementiere dieses Verfahren.

- Finde mithilfe deiner Funktion aus (c) die optimale Bandbreite aus $h \in \{0.7, 0.75, \dots, 2\}$ und plote den Kerndichteschätzer mit optimaler Bandbreite zusammen mit der richtigen Dichte der Standardnormalverteilung in ein Fenster.

Aufgabe 2 (8 Punkte)

Bekanntermaßen kann man die Lebensdauer von bestimmten technischen Geräten gut durch die Exponentialverteilung modellieren. Ein Computerhersteller möchte dies ausnutzen, um die durchschnittliche Lebensdauer von USB-Sticks im Dauerbetrieb zu testen. Dazu hat er von 5 Herstellern jeweils 20 USB-Sticks gekauft. Um die Lebensdauer möglichst genau zu schätzen, möchte er vorab ein paar Schätzer vergleichen. Aus seiner Studienzeit erinnert er sich an die Schätzer \bar{X}_n , S_n , und $n \cdot \min_{i=1, \dots, n} X_i$ für $1/\lambda$ bei $X_1, \dots, X_n \sim \text{Exp}(\lambda)$ i.i.d.

- Hilf dem Hersteller, die Schätzer zu vergleichen, indem Du
 - 100 mal 20 unabhängige und identisch exponentialverteilte Zufallsvariablen mit $\lambda = 5$ simulierst,
 - für die 3 Schätzer jeweils für jeden der 100 Datensätze die quadratische Abweichung¹ berechnest,
 - diese Abweichungen für die 3 Schätzer mit einer geeigneten Grafik vergleichst.
- Wende den laut Teil (a) (iii) besten Schätzer auf die Daten in `lebensdauer.txt` an. Welchen Hersteller würdest Du vorziehen?

¹unter der quadratischen Abweichung ist hier die mittlere quadratische Abweichung zwischen Schätzer und wahren Parameter zu verstehen (gemittelt über alle Stichproben, d.h. hier speziell gemittelt über die 100 Datensätze.).