Kapitel 1

Verteilungsschätzung

1.3 Kerndichteschätzung

Fragen? https://www.menti.com/alneo5om9sq5.

Dynamische Klassengrenzen — andere Schreibweise

Den Dichteschätzer mit dynamischen Klassen,

$$\hat{f}(x) = \frac{1}{nb} \sum_{i=1}^{n} \mathcal{I}(x_i \in [x - \frac{b}{2}, x + \frac{b}{2}]),$$

können wir mit einer Kernfunktion K alternativ schreiben als

$$\hat{f}(x) = \frac{1}{nb} \sum_{i=1}^{n} K\left(\frac{x - x_i}{b}\right),$$

wobei

$$K(y) = \begin{cases} 1 & \text{falls } -1/2 \leq y \leq 1/2; \\ 0 & \text{sonst.} \end{cases}$$

Fragen? https://www.menti.com/alneo5om9sq5.

Konstruktion des Kerndichteschätzers

$$\hat{f}(x) = \frac{1}{nb} \sum_{i=1}^{n} K\left(\frac{x - x_i}{b}\right)$$

Der Dichteschätzer ist somit eine **Summe von Rechtecksfunktionen**. Konkret erhält man $\hat{f}(x)$, indem man

1. über jeder Beobachtung x_i, $i = 1, \ldots, n$ ein Rechteck platziert[1]
2. an jedem Wert x die Höhen der x überdeckenden Rechtecke aufsummiert

Der resultierende Schätzer ist **nicht stetig**, da die Rechteckfunktionen $K(y)$ nicht stetig sind $\rightsquigarrow$ ersetze bisheriges $K(y)$ durch **eine stetige Funktion**!

[1] Breite b, Höhe $\frac{1}{nb}$, zentriert auf x_i

Fragen? https://www.menti.com/alneo5om9sq5.

Vom Rechteckkern zu allgemeinen Kernfunktionen

Eine Funktion $K(y)$ heißt Kernfunktion, falls

(i) $K(y) \geq 0$ für alle y
(ii) $K(y) = K(-y)$ für alle y
(iii) $\int_{-\infty}^{\infty} K(y)dy = 1$

Aus Eigenschaft (iii) ergibt sich, dass

$$\int_{-\infty}^{\infty} \hat{f}(x)dx = 1,$$

d.h. dass $\hat{f}(x)$ **eine Dichtefunktion ist**, wie gewünscht.

Fragen? https://www.menti.com/alneo5om9sq5.

Für beliebige **Kernfunktionen** $K(y)$ mit (i)-(iii) und fester Bandweite b heißt

$$\hat{f}(x) = \frac{1}{nb} \sum_{i=1}^{n} K\left(\frac{x - x_i}{b}\right)$$

Kerndichteschätzer (KDE).

Umsetzung in R:

```
hermannslauf_ergebnisse$Zeit_Minuten <- hermannslauf_ergebnisse$Zeit_Sekunden / 60
hermannslauf <- hermannslauf_ergebnisse$Zeit_Minuten
plot(density(hermannslauf)) # plottet den KDE
density(hermannslauf, kernel = "...", bw = "...") # Wahl von Kernfunktion/Bandweite
```

Im Folgenden diskutieren wir die Wahl von $K(y)$ sowie von b.

Fragen? https://www.menti.com/alneo5om9sq5.

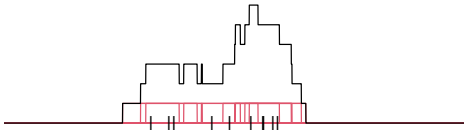
Kernfunktionen

Einige mögliche Kernfunktionen:

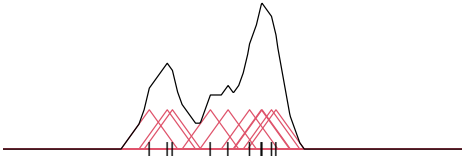
- **Rechteckkern**: $K(y) = \begin{cases} 1/2 & \text{falls } \mid y \mid \leq 1; \\ 0 & \text{sonst.} \end{cases}$
- **Dreieckkern**: $K(y) = \begin{cases} 1 - \mid y \mid & \text{falls } \mid y \mid \leq 1; \\ 0 & \text{sonst.} \end{cases}$
- **Gaußkern**: $K(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$
- **Epanechnikovkern**: $K(y) = \begin{cases} \frac{3}{4\sqrt{5}}\left(1 - \frac{1}{5}y^2\right) & \text{falls } \mid y \mid \leq \sqrt{5}; \\ 0 & \text{sonst.} \end{cases}$

Fragen? https://www.menti.com/alneo5om9sq5.

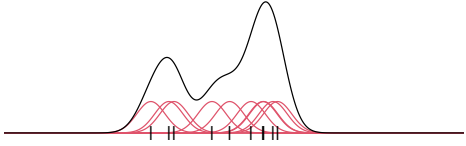
KDE−Konstruktion mit Rechteckkern



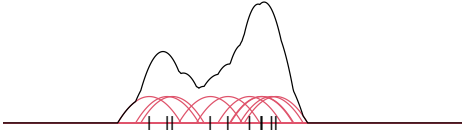
KDE−Konstruktion mit Dreieckkern



KDE−Konstruktion mit Gaußkern



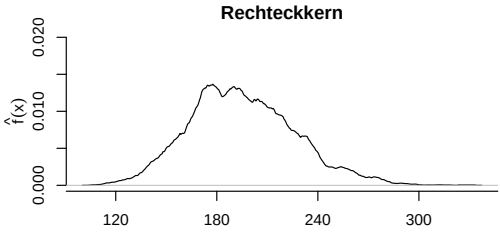
KDE−Konstruktion mit Epanechnikovkern

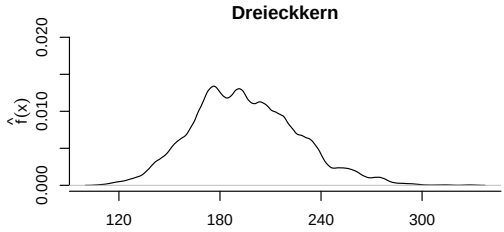


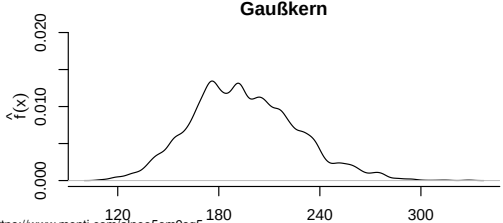
Dreieckkern, Gaußkern und Epanechnikovkern führen zu stetigem $\hat{f}(x)$ — bei Nutzung des Gaußkerns ist $\hat{f}(x)$ sogar differenzierbar.

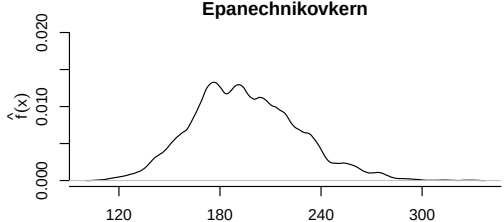
Fragen? https://www.menti.com/alneo5om9sq5.

- KDE für alle Zeiten beim 50. Hermannslauf (jeweils mit $b = 3$)
- ähnliche Verläufe, aber bzgl. Stetigkeit/Differenzierbarkeit gibt es nennenswerte Unterschiede









Fragen? https://www.menti.com/alneo5om9sq5.

Zwischenstand zur Wahl von $K(y)$ und b

Illustration von voriger Slide zeigt: Wahl von $K(y)$ ist nicht wirklich wichtig — nur der Rechteckkern sollte vermieden werden.

Viel wichtiger ist die Wahl der Bandweite b, und das führt uns wieder zum Bias-Varianz-Trade-off $\rightsquigarrow$ daher bestimmen wir jetzt zunächst Bias und Varianz.

Fragen? https://www.menti.com/alneo5om9sq5.

Erwartungswert des Kerndichteschätzers:

$$E\left(\hat{f}(x)\right) = E\left(\frac{1}{nb}\sum_{i=1}^{n} K\left(\frac{x - x_i}{b}\right)\right)$$

$$= \frac{1}{nb}\sum_{i=1}^{n}\int_{-\infty}^{\infty} K\Big(\underbrace{\frac{z-x}{b}}\Big) f(z)dz$$

$\frac{z-x}{b}=t \Leftrightarrow z=x+tb \Rightarrow dz=bdt$ (Substitution)

$$= \frac{1}{b}\int_{-\infty}^{\infty} K(t)\ \underbrace{f(x+tb)}\ bdt$$

$\approx f(x)+tbf'(x)+\frac{(tb)^2}{2}f''(x)$ (Taylor-Approximation 2. Ordnung)

$$\approx f(x)\underbrace{\int_{-\infty}^{\infty} K(t)dt}_{=1 \text{ (Dichte)}} + bf'(x)\underbrace{\int_{-\infty}^{\infty} tK(t)dt}_{=0 \text{ (Symmetrie)}} + \frac{b^2}{2}f''(x)\int_{-\infty}^{\infty} t^2K(t)dt$$

$$= f(x) + \frac{b^2}{2}f''(x)\int_{-\infty}^{\infty} t^2K(t)dt$$

Fragen? https://www.menti.com/alneo5om9sq5.

Bias des Kerndichteschätzers

Der **Bias** des Kerndichteschätzers, für gegebenes x, ist somit

$$E\left(\hat{f}(x) - f(x)\right) \approx b^2 \frac{f''(x)}{2} k_2,$$

wobei $k_2 = \int_{-\infty}^{\infty} t^2 K(t) dt$.

Wir stellen fest:

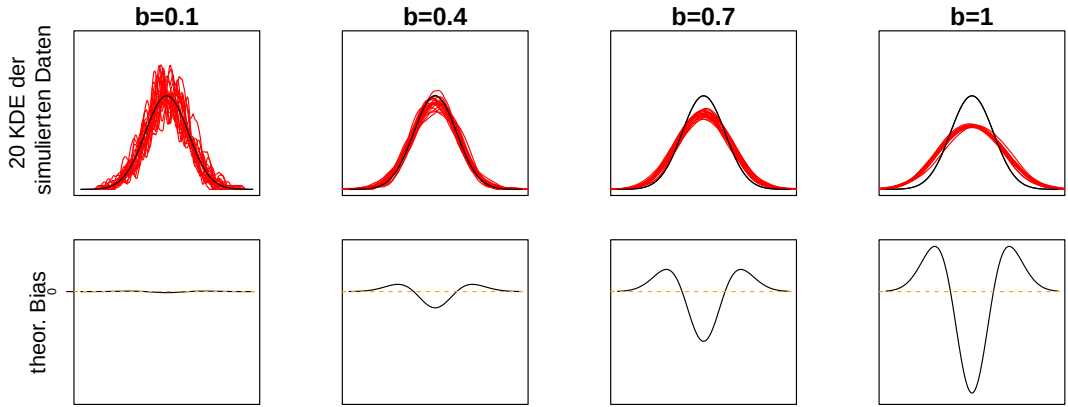
- der Bias ist hoch, wenn die Krümmung von $f(x)$ hoch ist
- wenn $f(x)$ in x linear ist, ist der Bias gleich Null
- **der Bias wird umso kleiner, je kleiner die Bandweite** b ist

Ein niedriger Bias allein macht aber noch keinen guten Schätzer $\rightsquigarrow$ wir wollen auch einen, der stabil ist!

Fragen? https://www.menti.com/alneo5om9sq5.

Illustration des Bias anhand aus $N(0,1)$-Verteilung simulierter Daten

- Dichte der $N(0,1)$-Verteilung in schwarz, Kerndichteschätzer für 20 simulierte Beobachtungen aus $N(0,1)$



- keine systematische Verzerrung bei kleinem b (gut), systematische Unterschätzung des Gipfels bei großem b (schlecht)

Fragen? https://www.menti.com/alneo5om9sq5.

Varianz des Kerndichteschätzers

Die **Varianz** des Kerndichteschätzers, für gegebenes x, ergibt sich wie folgt[2]:

$$\mathrm{Var}\left(\hat{f}(x)\right) \approx \frac{1}{nb} f(x) j_2,$$

wobei $j_2 = \int_{-\infty}^{\infty} K(t)^2 dt$.

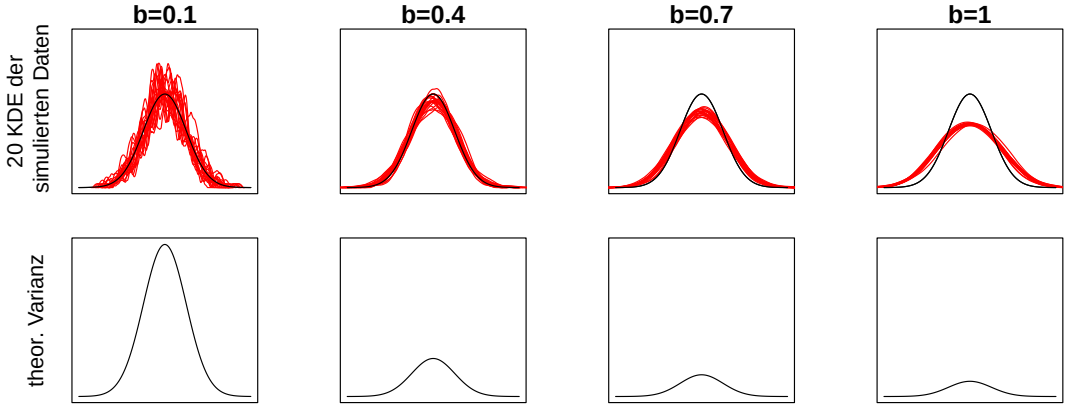
Wir stellen fest:

- die Varianz wird kleiner wenn der Stichprobenumfang n wächst (logisch!)
- die **Varianz wird umso kleiner, je größer die Bandweite** b

[2] wir sparen uns hier die Rechnung — ist aber im Prinzip analog wie beim Erwartungswert

Fragen? https://www.menti.com/alneo5om9sq5.

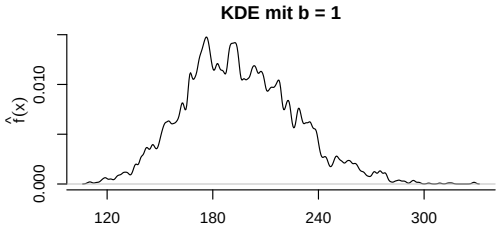
Illustration der Varianz anhand aus N(0, 1) — Verteilung simulierter Daten

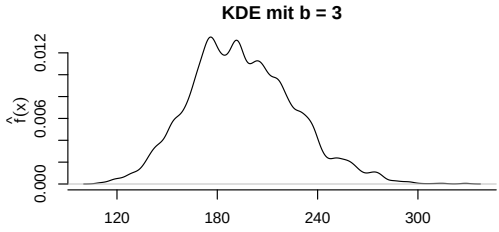


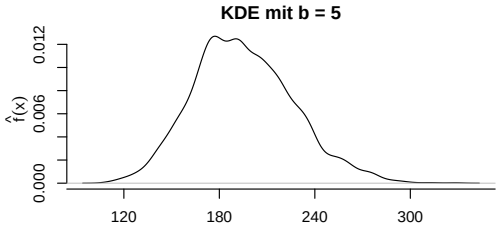
- hohe Varianz/unruhiger Verlauf bei kleinem *b* (schlecht), schöner glatter Verlauf bei großem *b* (gut)
- Die Bandweite *b* soll also einerseits klein (für niedrigen Bias), andererseits groß (für niedrige Varianz) sein ⇝ der optimale Wert liegt irgendwo in der Mitte

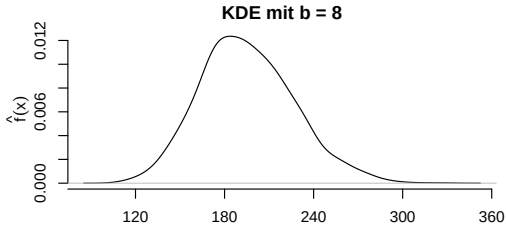
Fragen? https://www.menti.com/alneo5om9sq5.

KDE im Hermannslauf-Beispiel









Die “Extreme” $b = 1$ und $b = 8$ ergeben die schlechtesten Ergebnisse.

Fragen? https://www.menti.com/alneo5om9sq5.

MSE und MISE des Kerndichteschätzers

Der “mean squared error” (MSE) kombiniert Bias und Varianz in einem Fehlermaß:

$$\begin{aligned}\mathrm{MSE}\big(\hat{f}(x)\big) &= E\big(\hat{f}(x) - f(x)\big)^2 = \mathrm{Var}\big(\hat{f}(x)\big) + \mathrm{Bias}\big(\hat{f}(x)\big)^2 \\ &\approx \frac{1}{nb} f(x) j_2 + b^4 \frac{f''(x)^2}{4} k_2^2\end{aligned}$$

Der MSE hängt von dem Wert x ab, an dem die Dichte $f(x)$ geschätzt wird $\rightsquigarrow$ es handelt sich um ein lokales Fehlermaß.

Der “mean integrated squared error” (MISE) ist ein globales Fehlermaß:

$$\mathrm{MISE}\big(\hat{f}(x)\big) = \int_{-\infty}^{\infty} \mathrm{MSE}\big(\hat{f}(x)\big)\, dx \approx \frac{1}{nb} j_2 + b^4 \frac{\int_{-\infty}^{\infty} f''(x)^2 dx}{4} k_2^2$$

Fragen? https://www.menti.com/alneo5om9sq5.

Minimierung des MISE

$$\mathrm{MISE}\left(\hat{f}(x)\right) \approx \frac{1}{nb} j_2 + b^4 \frac{\int_{-\infty}^{\infty} f''(x)^2 dx}{4} k_2^2$$

Wir haben zwei Stellschrauben: Wahl von b und Wahl von $K(y)$.

Man kann zeigen:

- der Epanechnikovkern ist optimal im Sinne des MISE — insgesamt sind die (MISE-) **Unterschiede zwischen den Kernen aber eher marginal**
- die **Bandweitenwahl ist wichtiger**, und die (MISE-)optimale Bandweite b ist

$$b_{\text{opt.}} = \left(\frac{j_2}{n k_2^2 \int_{-\infty}^{\infty} f''(x)^2 dx} \right)^{1/5}$$

Fragen? https://www.menti.com/alneo5om9sq5.

Bandweitenwahl

$$b_{\text{opt.}} = \left(\frac{j_2}{n k_2^2 \int_{-\infty}^{\infty} f''(x)^2 dx} \right)^{1/5}$$

Problem: hängt von der unbekannten wahren Dichte $f(x)$ ab!

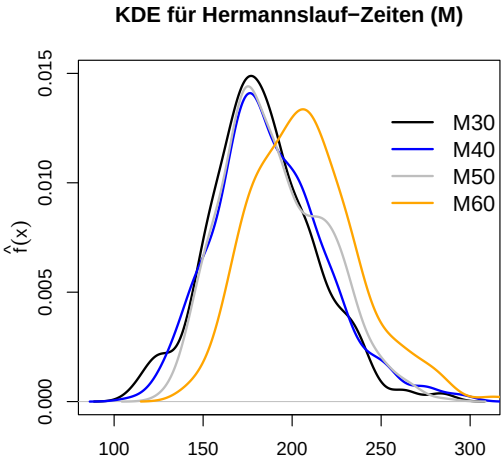
Einige Methoden zur Wahl von b:

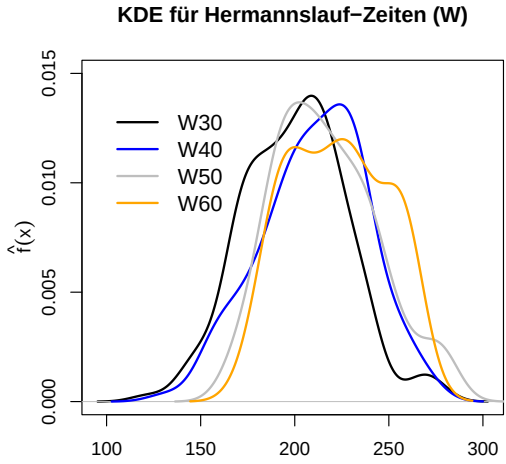
- subjektiv entscheiden, für welches b der Schätzer sinnvoll aussieht
- Silverman's rule I — nehme an, dass $f(x)$ die Dichte einer Normalverteilung ist $\rightsquigarrow$ $b = 1.06 s n^{-1/5}$, wobei s die empirische Standardabweichung ist
- Silverman's rule II — wie zuvor, aber mit (willkürlichem) Faktor 0.9 anstelle von 1.06 (denn letzterer führt meist zu zu glatten Funktionen)

Fragen? https://www.menti.com/alneo5om9sq5.

KDE für Gruppenvergleiche

- Kerndichteschätzer kann auch verwendet werden, um Unterschiede zwischen verschiedenen Schichten im Datensatz zu untersuchen (hier: Altersklassen):





Fragen? https://www.menti.com/alneo5om9sq5.