Einführung in funktionale Daten

Institut für Statistik
Ludwig-Maximilians-Universität München

Sarah Brockhaus

10.10.2014

Funktionale Daten

Einfache Statistiken für funktionale Daten

Ausblick: Glättung und Registrierung

Funktionale Daten

- Daten, die Informationen über Funktionen enthalten
- meist Kurven, aber auch Oberflächen oder 3D-Daten möglich
- Funktionen haben einen stetigen Definitionsbereich, z.B. Zeit oder Raum
- Zentrale Annahme: zugrundeliegende Funktionen sind glatt

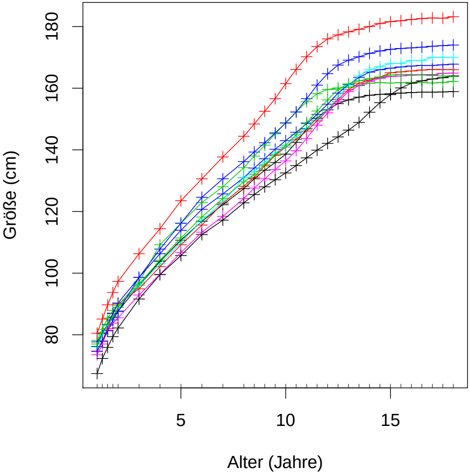
(Ramsay und Silverman 2006)

Handschriftliches fda

Cursive handwriting samples (Ramsay und Silverman 2006)

Wachstumskurven

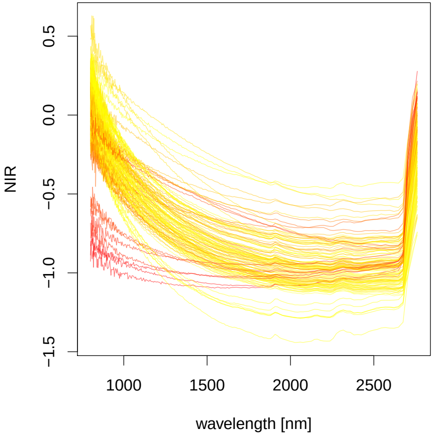
Wachstumskurven von 10 Mädchen

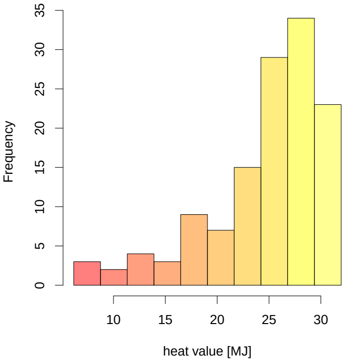


Berkeley Growth Study data (Ramsay und Silverman 2006)

Spektraldaten von fossilen Brennstoffen

Spektraldaten von fossilen Brennstoffen zur Vorhersage des Brennwertes

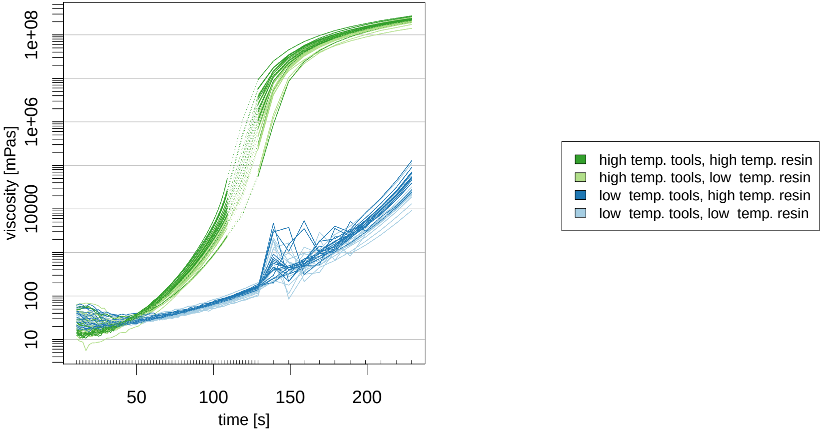




Spectral data of fossil fuels (Fuchs et al. 2015)

Viskosität von Harz

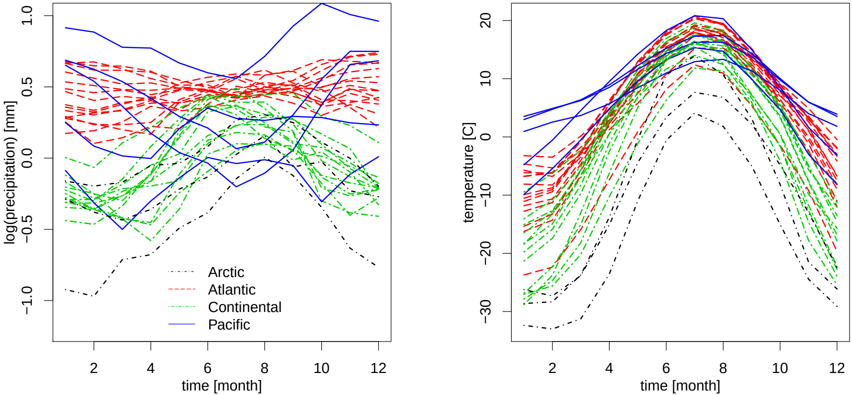
Viskosität von Harz abhängig von experimentellen Bedingungen



Viscosity of resin (Brockhaus et al. 2014)

Kanadische Wetterdaten

z.B. Betrachte klimatischen Zusammenhang zwischen Temperatur und Niederschlag



Canadian average annual weather cycle (Ramsay und Silverman 2006)

Mittelwert, Varianz und Kovarianz

- Betrachte funktionale Variable $X(t)$, mit $t \in \mathcal{T}$ und $\mathcal{T}$ Intervall in $\mathbb{R}$
- Stichprobe $x_i(t)$, $i = 1, \dots, n$

- funktionaler Mittelwert:

$$\hat{\mu}_X(t) = \bar{x}(t) = n^{-1} \sum_{i=1}^{n} x_i(t)$$

- funktionale Varianz:

$$\hat{\sigma}_X(t) = (n-1)^{-1} \sum_{i=1}^{n} [x_i(t) - \bar{x}(t)]^2$$

- funktionale Kovarianz:

$$\hat{\sigma}_X(t_1, t_2) = (n-1)^{-1} \sum_{i=1}^{n} [x_i(t_1) - \bar{x}(t_1)][x_i(t_2) - \bar{x}(t_2)]$$

Mittelwert, Varianz und Kovarianz

- Betrachte funktionale Variable $X(t)$, mit $t \in \mathcal{T}$ und $\mathcal{T}$ Intervall in $\mathbb{R}$
- Stichprobe $x_i(t)$, $i = 1, \ldots, n$
- funktionaler Mittelwert:

$$\hat{\mu}_X(t) = \bar{x}(t) = n^{-1} \sum_{i=1}^{n} x_i(t)$$

- funktionale Varianz:

$$\hat{\sigma}_X(t) = (n-1)^{-1} \sum_{i=1}^{n} [x_i(t) - \bar{x}(t)]^2$$

- funktionale Kovarianz:

$$\hat{\sigma}_X(t_1, t_2) = (n-1)^{-1} \sum_{i=1}^{n} [x_i(t_1) - \bar{x}(t_1)][x_i(t_2) - \bar{x}(t_2)]$$

Mittelwert, Varianz und Kovarianz

- Betrachte funktionale Variable $X(t)$,
  mit $t \in \mathcal{T}$ und $\mathcal{T}$ Intervall in $\mathbb{R}$
- Stichprobe $x_i(t)$, $i = 1, \ldots, n$
- funktionaler Mittelwert:

$$\hat{\mu}_X(t) = \bar{x}(t) = n^{-1} \sum_{i=1}^{n} x_i(t)$$

- funktionale Varianz:

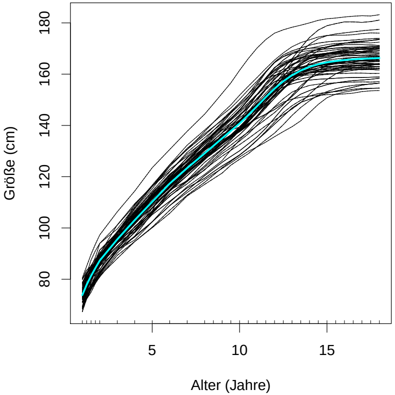
$$\hat{\sigma}_X(t) = (n-1)^{-1} \sum_{i=1}^{n} [x_i(t) - \bar{x}(t)]^2$$

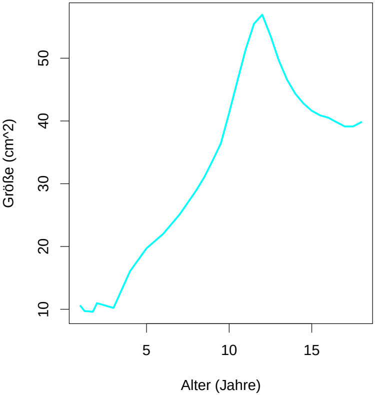
- funktionale Kovarianz:

$$\hat{\sigma}_X(t_1, t_2) = (n-1)^{-1} \sum_{i=1}^{n} [x_i(t_1) - \bar{x}(t_1)][x_i(t_2) - \bar{x}(t_2)]$$

Beispiel Mittelwert und Varianz

Wachstumskurven von 54 Mädchen

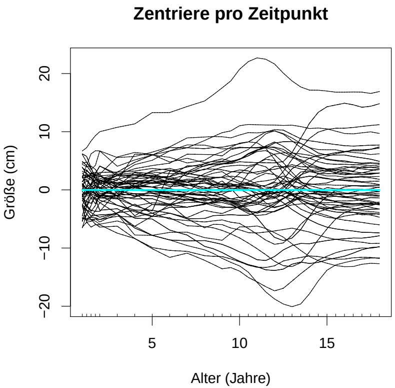




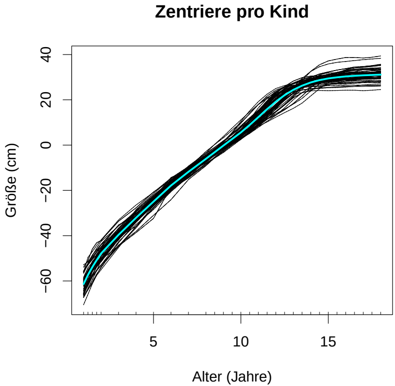
geschätzter Mittelwert

geschätzte Varianz

Beispiel Mittelwert: Zwei Arten der Zentrierung



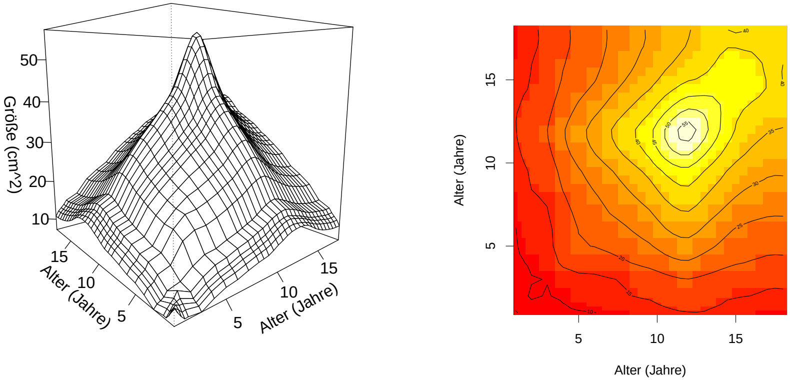
$$x_i^*(t) = x_i(t) - \bar{x}(t)$$



$$\tilde{x}_i(t) = x_i(t) - \int x_i(t)dt$$

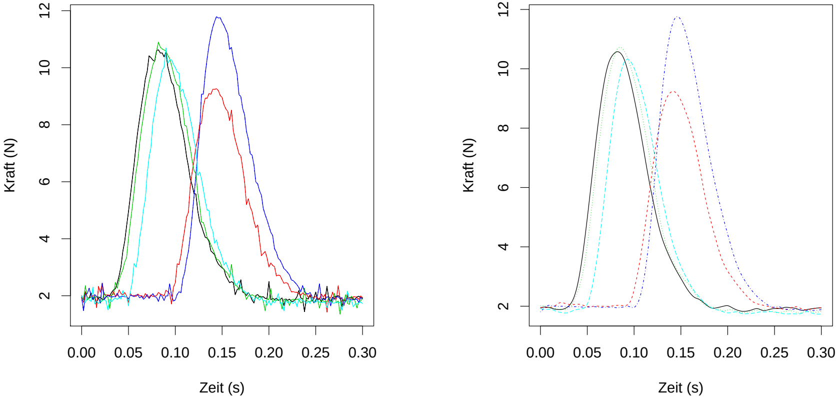
Beispiel Kovarianzfunktion

$$\hat{\sigma}_X(t_1, t_2) = (n-1)^{-1} \sum_{i=1}^{n} [x_i(t_1) - \bar{x}(t_1)][x_i(t_2) - \bar{x}(t_2)]$$



Glättung

- funktional Variable $x(t)$ mit Fehler beobachtet: $w_i(t) = x_i(t) + \epsilon_i(t)$
- Ziel: Finde die zugrundeliegenden glatten Funktionen $x_i(t)$



Pinch force (Ramsay und Silverman 2006)

Glättung

- Vielzahl an Methoden zur Glättung, zum Beispiel
  - Splines (Polynom-Splines, B-Splines, P-Splines)
  - lokale Glätter (Nächste-Nachbarn, Running Mean/Median, Loess)
  - funktionale Hauptkomponentenanalyse (fPCA)

- Bias-Varianz-Trade-Off / Konflikt zwischen Datentreue und Glattheit

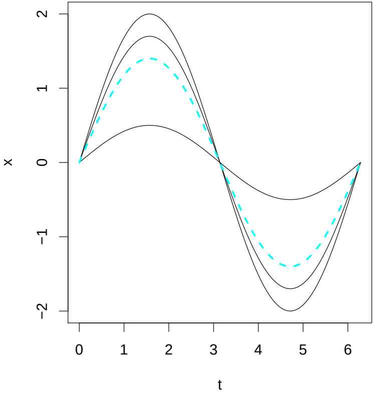
Glättung

- Vielzahl an Methoden zur Glättung, zum Beispiel
  - Splines (Polynom-Splines, B-Splines, P-Splines)
  - lokale Glätter (Nächste-Nachbarn, Running Mean/Median, Loess)
  - funktionale Hauptkomponentenanalyse (fPCA)
- Bias-Varianz-Trade-Off / Konflikt zwischen Datentreue und Glattheit

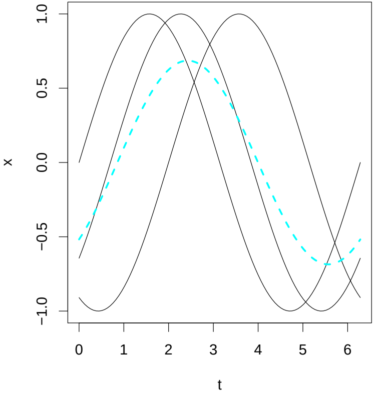
Motivation für Registrierung

zwei Arten von Variation

- Amplitudenvariation, 'Variation in y-Richtung'



- Phasenvariation, 'Variation in x-Richtung'



- meist Interesse an Amplitudenvariation

Registrierung

Vielzahl an Methoden zur Registrierung, zum Beispiel

- Shift Registrierung:
  - einfache lineare Transformation der t-Variable
  - gesucht $x_i^*(t) = x_i(t + \delta_i)$
  - finde δ_i, sodass die vertikale Abweichung aller Kurven von der geschätzten Mittelwertkurve minimal ist
- Landmark Registrierung:
  - Landmarks sind spezielle Charakteristika von Kurven, z.B. Extrema, Nullstellen
  - Kurven werden an Landmarks ausgerichtet
  - verwende streng monotone Transformationsfunktion $h_i(t)$

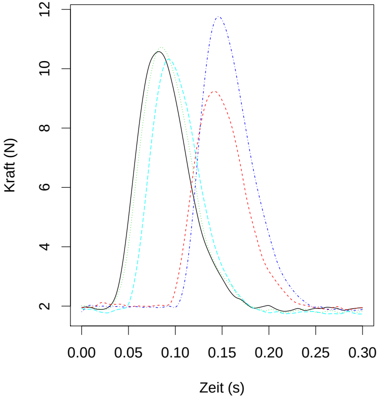
Registrierung

Vielzahl an Methoden zur Registrierung, zum Beispiel

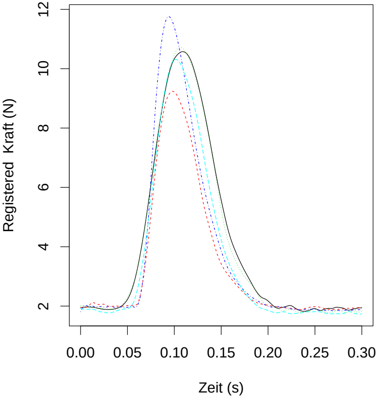
- Shift Registrierung:
  - einfache lineare Transformation der t-Variable
  - gesucht $x_i^*(t) = x_i(t + \delta_i)$
  - finde δ_i, sodass die vertikale Abweichung aller Kurven von der geschätzten Mittelwertkurve minimal ist
- Landmark Registrierung:
  - Landmarks sind spezielle Charakteristika von Kurven, z.B. Extrema, Nullstellen
  - Kurven werden an Landmarks ausgerichtet
  - verwende streng monotone Transformationsfunktion $h_i(t)$

Beispiel Shift Registrierung

Pinch force (Ramsay und Silverman 2006)



Geglättete Daten



Registrierte Daten

Literatur

- Brockhaus, S., Scheipl, F., Hothorn, T., & Greven, S. (2014). 'The functional linear array model.' under review.

- Fuchs, K., Scheipl, F., & Greven, S. (2015). 'Penalized scalar-on-functions regression with interaction term.' *Computational Statistics & Data Analysis*, **81**, 38–51.

- Scheipl, F., Staicu, A. M., & Greven, S. (2014). 'Functional additive mixed models.' *Journal of Computational and Graphical Statistics*, in print.

- Ramsay, J. O. & Silverman, B. W. (2006), *Functional data analysis,* Springer, New York.