

# Einführung in die Kerndichteschätzung

Yavuzâlp Dal

Stand: 14. Dezember 2025

- 1 Einleitung & Motivation
- 2 Der Rosenblatt-Schätzer
- 3 Das Histogramm
- 4 Kernschätzer (KDE)
- 5 Nichtparametrische Regression
- 6 Robuste Lineare Regression
- 7 Zusammenfassung
- 8 Literaturverzeichnis

# Parametrische vs. Nichtparametrische Dichteschätzung (Übersicht)

## **Parametrisch (kennen wir schon – > Stochastik)**

- Annahme: Verteilung ist bekannt!
- Wenige Parameter ( $\mu, \sigma^2$ )
- Schätzung durch Parameterschätzer wie bspw. Maximum-Likelihood

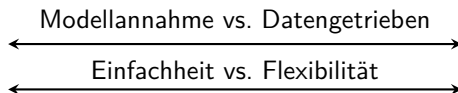
# Parametrische vs. Nichtparametrische Dichteschätzung (Übersicht)

## Parametrisch (kennen wir schon – > Stochastik)

- Annahme: Verteilung ist bekannt!
- Wenige Parameter ( $\mu, \sigma^2$ )
- Schätzung durch Parameterschätzer wie bspw. Maximum-Likelihood

## Nichtparametrisch (KDE)

- Keine feste Form; die Zufallsvariable  $X$  muss aber stetig sein, die Dichte  $f$  eine "glatte" Funktion
- Schätzung: Kern (wie bspw. Epanechnikov) sowie Bandbreite  $h$
- Erfolgt mittels des Kerndichteschätzers



## Warum nichtparametrisch? [BT94]

- Klassische parametrische Verfahren (z. B. unter Annahme einer Normalverteilung) sind oft zu starr.
  - Starke Kopplung an unsicheren Annahmen!
- Bei unbekannter Verteilung bietet die nichtparametrische Schätzung oft bessere Einblicke in die Datenstruktur.

## Warum nichtparametrisch? [BT94]

- Klassische parametrische Verfahren (z. B. unter Annahme einer Normalverteilung) sind oft zu starr.
  - Starke Kopplung an unsicheren Annahmen!
- Bei unbekannter Verteilung bietet die nichtparametrische Schätzung oft bessere Einblicke in die Datenstruktur.

## Zielsetzung:

- Annäherung an den Wert  $f(x)$  an einer Stelle  $x$ , ohne eine spezifische parametrische Familie vorauszusetzen.

- 1 Einleitung & Motivation
- 2 Der Rosenblatt-Schätzer
- 3 Das Histogramm
- 4 Kernschätzer (KDE)
- 5 Nichtparametrische Regression
- 6 Robuste Lineare Regression
- 7 Zusammenfassung
- 8 Literaturverzeichnis

- **Idee:** Da  $f(x) = F'(x)$ , nutzt man den Differenzenquotienten der empirischen Verteilungsfunktion  $F_n$ .



# Der Rosenblatt-Schätzer (Formel)

- Der Schätzer:

$$f_n(x) = \frac{F_n(x+h) - F_n(x-h)}{2h} \quad (1)$$

# Der Rosenblatt-Schätzer (Konsistenz)

- **Konsistenz:** Damit der Schätzer gegen die wahre Dichte konvergiert, muss für  $n \rightarrow \infty$  gelten:
  - Die Bandbreite  $h \rightarrow 0$  (Bias reduzieren).
  - $nh \rightarrow \infty$  (Varianz reduzieren).

- 1 Einleitung & Motivation
- 2 Der Rosenblatt-Schätzer
- 3 Das Histogramm**
- 4 Kernschätzer (KDE)
- 5 Nichtparametrische Regression
- 6 Robuste Lineare Regression
- 7 Zusammenfassung
- 8 Literaturverzeichnis

- **Funktionsweise:** Zerlegung des Datenraums in Boxen (Bins) der Breite  $h$ . Die Höhe der Box entspricht der relativen Häufigkeit  $\frac{n_k}{n}$  normiert durch  $h$ .

# Das Histogramm (Probleme)

- **Probleme:**
  - Unstetigkeit: Die geschätzte Dichte ist eine Treppenfunktion.
  - Subjektivität: Abhängig von  $x_0$  und  $h$ .

# Das Histogramm (Optimale Bandbreite)

- **Optimale Bandbreite ( $h$ ):**

- Scott:  $h \approx 3.49sn^{-1/3}$
- Freedman-Diaconis:  $h^* = 2(x_{0.75} - x_{0.25})n^{-1/3}$

- 1 Einleitung & Motivation
- 2 Der Rosenblatt-Schätzer
- 3 Das Histogramm
- 4 Kernschätzer (KDE)**
- 5 Nichtparametrische Regression
- 6 Robuste Lineare Regression
- 7 Zusammenfassung
- 8 Literaturverzeichnis

- **Der Schätzer:**

$$\hat{f}_n(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \quad (2)$$



- **Kernwahl:** Der Kern muss zu 1 integrieren ( $\int K(x)dx = 1$ ).
- Gängige Kerne:
  - Rechteck-Kern
  - Gauß-Kern
  - Epanechnikov-Kern

- **Bandbreitenwahl (Silverman's Rule):**

$$h_{opt} \approx 1.06\sigma n^{-1/5} \quad (3)$$

- 1 Einleitung & Motivation
- 2 Der Rosenblatt-Schätzer
- 3 Das Histogramm
- 4 Kernschätzer (KDE)
- 5 Nichtparametrische Regression**
- 6 Robuste Lineare Regression
- 7 Zusammenfassung
- 8 Literaturverzeichnis

## Watson-Nadaraya Schätzer:

$$\hat{m}_{WN}(x) = \frac{\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) Y_i}{\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)} \quad (4)$$

- Schätzung von  $m(x)$  als gewichteter Mittelwert der umliegenden  $Y$ -Werte.

- 1 Einleitung & Motivation
- 2 Der Rosenblatt-Schätzer
- 3 Das Histogramm
- 4 Kernschätzer (KDE)
- 5 Nichtparametrische Regression
- 6 Robuste Lineare Regression**
- 7 Zusammenfassung
- 8 Literaturverzeichnis

- **Theil-Methode I:** Steigung zwischen Paaren  $(i, i + n/2)$ , Median als Schätzer.
- **Theil-Sen-Schätzer (Methode II):** Median der Steigungen zwischen allen Paaren  $i < j$ :

$$H_{ij} = \frac{Y_j - Y_i}{X_j - X_i} \quad (5)$$

- Der Schätzer für den Anstieg  $\beta$  ist der Median dieser Steigungen.

# Outline

- 1 Einleitung & Motivation
- 2 Der Rosenblatt-Schätzer
- 3 Das Histogramm
- 4 Kernschätzer (KDE)
- 5 Nichtparametrische Regression
- 6 Robuste Lineare Regression
- 7 Zusammenfassung**
- 8 Literaturverzeichnis



- **Flexibilität:** Nichtparametrische Methoden passen sich den Daten an.
- **Parameter  $h$ :** Kritischer Schritt, Balance zwischen Bias und Varianz.

- **Robustheit:** Verfahren wie der Theil-Sen-Schätzer bieten robuste Alternativen zur klassischen Regression, besonders bei Ausreißern.

# Outline

- 1 Einleitung & Motivation
- 2 Der Rosenblatt-Schätzer
- 3 Das Histogramm
- 4 Kernschätzer (KDE)
- 5 Nichtparametrische Regression
- 6 Robuste Lineare Regression
- 7 Zusammenfassung
- 8 Literaturverzeichnis**

- [BT94] Herbert Büning und Götz Trenkler. *Nichtparametrische statistische Methoden*. Berlin, Boston: De Gruyter, 1994. ISBN: 9783110902990. DOI: [doi:10.1515/9783110902990](https://doi.org/10.1515/9783110902990). URL: <https://doi.org/10.1515/9783110902990>.