Numerische Methoden in der Physik

DR. BJÖRN SCHELTER

Aufgabenblatt Nr. 5

Übung 6
Suboptimalität des Least-Squares Schätzers bei nicht Gauß-verteilten Daten
Heute machen wir eine Simulationsstudie, in der klar wird, inwiefern der Least-Squares Schätzer suboptimal ist, wenn die zugrunde liegenden Daten nicht Gauß-verteilt sind.

Wiederhole folgende Schritte M=200 Mal:

- Simuliere N Daten aus dem Modell
$$y_i = a + bx_i + \epsilon_i, \quad p(\epsilon_i) = \frac{1}{2}e^{-|\epsilon_i|}, \quad i = 1, \dots, N$$
  - Tipp:
    Generiere ϵ_i durch exponentiell-verteilte Zufallszahlen mit `exprnd.m`. Das zufällige Vorzeichen der double exponential Verteilung kann basierend auf Gaußschen Zufallszahlen gezogen werden.
- Schätze die Parameter a und b mit einem Least-Squares (LS) Fit:
$$\hat{a}_{LS} = \frac{S_{xx}S_y - S_x S_{xy}}{SS_{xx} - S_x^2}; \quad \hat{b}_{LS} = \frac{SS_{xy} - S_x S_y}{SS_{xx} - S_x^2};$$
  mit
$$S = N; \quad S_x = \sum_{i=1}^{N} x_i; \quad S_y = \sum_{i=1}^{N} y_i; \quad S_{xy} = \sum_{i=1}^{N} x_i y_i; \quad S_{xx} = \sum_{i=1}^{N} x_i x_i;$$
- Schätze die Parameter a, b mit dem robusten Medianfit aus der Vorlesung:
  - Wähle Startbedingungen $(a_0 = \hat{a}_{LS}, b_0 = \hat{b}_{LS})$ aus dem LS Fit:
  - Iteriere für $j = 1, \dots$
    * $a_{j+1} = \text{median}(y_i - b_j x_i)$
    * Nullstellensuche für b_{j+1} : $0 = \sum_{i=1}^{N} x_i \text{sign}(y_i - a_{j+1} - b_{j+1} x_i)$
      Was ist der theoretische Hintergrund dieser Nullstellensuche?

  bis sich die Parameter nur wenig verändern (z.B. $|a_{j+1} - a_j| < 0.0001$). Setze dann $\hat{a}_{\text{rob}} = a_{j+1}$ und $\hat{b}_{\text{rob}} = b_{j+1}$

Vergleiche die kumulative Verteilungen aller M LS Schätzungen ($\hat{a}_{LS}$,$\hat{b}_{LS}$) mit denen aus der robusten Schätzung $(\hat{a}_{\text{rob}}, \hat{b}_{\text{rob}})$. Was fällt auf?

- Tipp: Verwende für die kumulative Verteilung die routine `ecdf.m`.