Übung 1

1. a) Sei $\hat{\mu}_n$ das α-gestutzte Mittel von $(X_1, \ldots, X_n)$. Berechne seine Veränderung bei Hinzufügen einer zusätzlichen Beobachtung bei x, d. h.

$$\hat{\mu}_{n+1}(x_1, \ldots, x_n, x) - \hat{\mu}_n(x_1, \ldots, x_n)$$

im Fall $[n\alpha] = [(n+1)\alpha]$. Gib auch die Sensitivität des α-gestutzten Mittels an.

b) Wir betrachten das Modell $Y_i = \beta X_i + \epsilon_i,\ i = 1, \ldots, n$ mit $X_i > 0$ zufällig. Das heisst, dass Y bis auf einen zufälligen Fehler proportional ist zu X. Berechne den empirischen Bruchpunkt des Schätzers

$$\hat{\beta} = \text{argmin}_\beta \sum_{i=1}^{n} |y_i - \beta x_i| \,,$$

d. h. den minimalen Anteil von Ausreissern unter den Paaren (y_i, x_i), der zum Zusammenbruch von $\hat{\beta}$ führt.
Hinweis: Schreibe $\hat{\beta} = \text{argmin}_\beta \sum_{i=1}^n w_i |z_i - \beta|$, wobei $z_i = y_i / x_i$, $w_i = \frac{x_i}{x_1 + \ldots + x_n}$. Wenn wir die Zufallsvariable Z einführen, die den Wert z_i mit Wahrscheinlichkeit w_i annimmt, dann können wir $\hat{\beta}$ berechnen mithilfe von Satz A.8.

c) Wir betrachten das Modell $Y_i = \beta x_i + \epsilon_i,\ i = 1, \ldots, n$ mit $x_i = \frac{i}{n}$ fest. Berechne den empirischen Bruchpunkt des oben angegebenen Schätzers für $n \to \infty$.

2. Sei $X_i = \mu + \epsilon_i\ i = 1, 2, \ldots$, wobei die ϵ_i i.i.d. sind mit Dichte

$$f_\lambda(x) = \frac{1}{\pi} \frac{\lambda}{\lambda^2 + x^2}$$

(λ ist ein Parameter, der die Streuung der ϵ_i beschreibt). Diese Verteilung heisst Cauchyverteilung.

a) (freiwillig) Es gilt, dass $\sum_{i=1}^n \epsilon_i$ Cauchy(nλ)-verteilt ist. Zeige dies im Falle, wo $\lambda = 1$, $n = 2$, d.h. berechne die Dichte von $\epsilon_1 + \epsilon_2$ mit der Faltungsformel.
Hinweis: Benütze die folgende Partialbruchzerlegung

$$\frac{1}{1+y^2} \; \frac{1}{1+(x-y)^2} = \frac{1}{x(4+x^2)} \left[\frac{2y+x}{1+y^2} + \frac{2(x-y)+x}{1+(x-y)^2} \right].$$

b) Folgere aus a), dass $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ die gleiche Verteilung hat wie X_1. Ist das ein Widerspruch zum Gesetz der grossen Zahlen?

c) Sei $\widehat{\mu}_n = \text{median}(X_1, \ldots, X_n)$. Berechne $\lim_{n \to \infty} P[|\widehat{\mu}_{2n+1} - \mu| > c]$ $(c > 0)$.
Hinweis: Überlege, dass

$$P[\widehat{\mu}_{2n+1} > \mu + c] = P[\sum_{i=1}^{2n+1} 1_{\{X_i > \mu + c\}} > n]$$

gilt, und wende das Gesetz der grossen Zahlen an.

d) Welcher der beiden Schätzer, arithmetisches Mittel oder Median, ist also in diesem Fall vorzuziehen, wenn man $P[|\widehat{\mu}_n - \mu| \leq c]$ als Genauigkeitsmass verwendet?

3. Wir definieren (mit $\tilde{X} = (X_1, \ldots, X_n)$):

$$\mathrm{MAD}(\tilde{X}) = \mathrm{median}(|X_1 - \mathrm{median}(\tilde{X})|, \ldots, |X_n - \mathrm{median}(\tilde{X})|),$$

$$S_n = \sqrt{\tfrac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2}.$$

Die ersten 10 Werte des Lichtgeschwindigkeits-Datensatzes aus der Vorlesung sind
28 26 33 24 34 -44 27 16 40 -2.
Berechne für diesen Teildatensatz

a) arithmetisches Mittel,

b) arithmetisches Mittel der nicht verworfenen Daten unter folgender Verwerfungsregel: Verwerfe alle Daten, die weiter als $5\mathrm{MAD}(\tilde{X})$ entfernt vom Median sind,

c) arithmetisches Mittel der nicht verworfenen Daten unter folgender Verwerfungsregel: Verwerfe alle Daten, die weiter als $2.29\sqrt{S_n}$ entfernt vom arithmetischen Mittel aller Daten sind,

d) 0.1- und 0.2-gestutztes Mittel,

e) (freiwillig) eine sinnvolle Näherung des Huber-Schätzers mit dem im Skript angegebenen Wert von c,

f) den Vertrauensbereich zum Niveau 0.95 für μ basierend auf dem t-verteilten Pivot auf Skript S. 5 unten (die benötigten oberen und unteren Quantile der t-Verteilung mit 9 Freiheitsgraden sind ± 2.26),

g) den Vertrauensbereich zum Niveau 0.95 für μ basierend auf dem Binomial-verteilten Pivot auf Skript S. 6 oben ($k_{0.025} = 2$).

Trage den Datensatz und alle berechneten Schätzer und Bereiche auf eine Zahlengerade auf. Diskutiere die Schätzer im Lichte des gesamten Lichtgeschwindigkeits-Datensatzes.

Abgabe: Montag, 12. November