Vorlesung Mathematische Statistik

Wintersemester 2018/2019

Vorlesung: Prof. Dr. Dietmar Ferger
Mitschrift: Willi Sontopski

11. Oktober 2018

Inhaltsverzeichnis

1. Einführung

Viele Schätzer in der Statistik sind definiert als Minimal- oder Maximalstelle von bestimmten *Kriteriumsfunktionen*, z. B. der *Maximum-Likelihood-Schätzer (MLS)* oder *Minimum-Qudrat-Schätzer (MQS, KQS)* oder *Bayes-Schätzer*. Allgemein nennt man solche Schätzer **M-Schätzer**.
Ziel: Untersuchung des asymptotischen Verhaltens ($n \to \infty$) von M-Schätzern über einen *funktionalen Ansatz*. Als Beispiel:

1.1. Der Median

Sei $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$ eine reelle Zufallsvariable mit Verteilungsfunktion
$F_X : \mathbb{R} \to [0, 1],\ F_X(x) := \mathbb{P}[X \leq x]$, also $X \sim F_X$. Definiere

$$\begin{aligned}
Y(t) &:= \mathbb{E}\left(|X - t|\right) \qquad (1.0)\\
&= \int\limits_{\Omega} |X(\omega) - t| \ \mathrm{d}\mathbb{P}(\omega)\\
&\overset{\text{Trafo}}{=} \int\limits_{\mathbb{R}} |x - t| \Big(\mathbb{P} \circ X \ \mathrm{d}x\Big)\\
&= \int\limits_{\mathbb{R}} |x - t|(F \ \mathrm{d}x) \qquad \forall t \in \mathbb{R}\\
m &:= \arg\min_{t \in \mathbb{R}} Y(t) := \ \text{(irgendeine) Minimalstelle der Funktion}
\end{aligned}$$

Notation. $F(m-) := F(m - 0) := \lim\limits_{t \uparrow m} F(t)$

Charakterisierung der Menge aller Mediane in folgendem kleinen Lemma:

Lemma 1.1.1 Sei $X \sim F_X$ integrierbar und $m \in \mathbb{R}$. Dann äquivalent:

(a) $F(m-) \leq \frac{1}{2} \leq F(m)$

(b) $\mathbb{E}[|X - t|] \geq \mathbb{E}[|X - m|] \qquad \forall t \in \mathbb{R}$

(c) m ist Median

Beweis. Zeige (a) $\Rightarrow$ (b):

Setze $h(t) := \mathbb{E}[|X-t| - |X-m|] \overset{\text{Lin}}{=} Y(t) - Y(m)$. Dann ist 2. äquivalent zu $h(t) \geq 0 \ \forall t \in \mathbb{R}$. Dies ist noch zu zeigen.

Fall 1: $t < m$

$$\begin{aligned}
h(t) &\overset{\text{Trafo}}{=} \int\limits_{\mathbb{R}} |x-t| - |x-m| Q_F(\,\mathrm{d}x) \\
&= \int\limits_{(-\infty,t]} \underbrace{|x-t| - |x-m|}_{=t-x-(m-x)=-(m-t)} F(\,\mathrm{d}x) + \int\limits_{(t,m)} \underbrace{|x-t| - |x-m|}_{\underbrace{\underbrace{x-t}_{\geq 0} - \underbrace{(m-x)}_{\leq m-t}}_{\geq -(m-t)}} F(\,\mathrm{d}x) + \int\limits_{[m,\infty)} \underbrace{|x-t| - |x-m|}_{x-t-(x-m)=m-t} F(\,\mathrm{d}x) \\
&\geq -(m-t) \cdot \underbrace{Q((-\infty,t])}_{F(t)} + \Big(-(m-t) \cdot F(m-) - F(t)\Big) + (m-t) \cdot \underbrace{Q([m,\infty))}_{1-\underbrace{Q((-m,m)}_{F(m-)}} \\
&= -\underbrace{(m-t)}_{\geq 0} \cdot (\underbrace{1 - 2 \cdot F(m-)}_{\overset{1.}{\geq} 0}) \\
&\geq 0
\end{aligned}$$

Fall 2: $t > m$

$$\begin{aligned}
h(t) = \int\limits_{(-\infty,m]} &\ldots F(\,\mathrm{d}x) + \int\limits_{(m,t]} \ldots + \int\limits_{(t,\infty)} \ldots F(\,\mathrm{d}x) \\
&\ldots \\
&\geq (t-m) \cdot (\underbrace{2 \cdot F(m) - 1}_{\overset{1.}{\geq} 0}) \\
&\geq 0
\end{aligned}$$

Fall 3: $t = m$ ist trivial. #

Zeige (b) $\Rightarrow$ (a):
Nach Annahme ist $h(t) \geq 0 \ \forall t \in \mathbb{R}$.

Fall 1: $t < m$ Die obige Rechnung im Fall 1 bei 1. $\Rightarrow$ 2. zeigt:

$$\begin{aligned}
0 \le h(t) &= -(m-t)\cdot F(t) + \int_{(t,m)} \underbrace{\underbrace{x}_{<m} - t - (m-x)}_{=2x-t-m\le m-t} F(\,\mathrm{d}x) + (m-t)\cdot(1-F(m-)) \\
&\le -(m-t)\cdot\Big(F(t) - 1 \underbrace{+F(m-) - F(m-)}_{=0} + F(t)\Big) \\
&= \underbrace{(m-t)}_{>0}\cdot(1-2\cdot F(t)) \\
&\Longrightarrow \forall t < m : 0 \le (m-t)\cdot(1-2\cdot F(t)) \\
&\Longrightarrow \forall t < m : 0 \le 1 - 2\cdot F(t) \\
&\Longrightarrow \forall t < m : F(t) \le \frac{1}{2} \\
&\overset{t\uparrow m}{\Longrightarrow} F(m-) \le \frac{1}{2}
\end{aligned}$$

Fall 2: $t > m$ Siehe 2. Fall, analog.

Zeige (a) $\Leftrightarrow$ (c): (b) ist offensichtlich äquivalent zur Definition des Medians. $\square$

Bemerkung 1.1.2

1. Lemma 1.1.1 (a) besagt, dass $\{m \in \mathbb{R} : m \text{ erfüllt } 1.\}$ die Menge aller Mediane von F ist.

2. Im Allgemeinen gibt es mehrere Mediane. Üblicherweise wählt man $m := F^{-1}(\frac{1}{2})$, wobei

$$F^{-1}(u) := \inf\{x \in \mathbb{R} : F(x) \ge u\} \quad \forall u \in (0,1)$$

die **Quantilfuntion / die verallgemeinerte Inverse** ist. Da

$$F\left(F^{-1}(u)-\right) \le u \le F\left(F^{-1}(u)\right) \qquad \forall u \in (0,1),$$

erfüllt $m = F^{-1}\left(\frac{1}{2}\right)$ die Bedingung (a) in Lemma 1.1.1 und ist somit ein Median, nämlich der kleinste.

3. Die obige Funktion (1.0), also

$$Y : \mathbb{R} \to \mathbb{R}, \qquad Y(t) = \int |x-t|\; F(\,\mathrm{d}x) \qquad \forall t \in \mathbb{R},$$

ist stetig (nutze Folgenkriterium + dominierte Konvergenz bzw. Satz von Lebesgue), aber im Allgemeinen nicht differenzierbar, z. B. falls $F \sim X$ eine diskrete

Zufallsvariable ist. In diesem Fall ist somit die Minimierung über Differentiation nicht möglich!

Zur Schätzung von m seien $X_1, \dots, X_n$ i.i.d.$\sim F$ mit zugehöriger **empirischer Verteilungsfunktion**

$$F_n(x) := \frac{1}{n} \cdot \sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}} \qquad \forall x \in \mathbb{R}.$$

Tatsächlich ist F_n die Verteilungsfunktion zum **empirischen Maß**

$$Q_n := \frac{1}{n} \cdot \sum_{i=1}^{n} \delta_{X_i} \text{ wobei } \delta_x \text{das Dirac-Maß in} x \in \mathbb{R}$$

Gemäß dem Satz von Gliwenko-Cantelli gilt:

$$\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \overset{n\to\infty}{\longrightarrow} 0 \text{ konvergiert } \mathbb{P}\text{-fast sicher für alle Vereteilungsfunktionen } F$$

Erinnerung. Für das Dirac-Maß $\delta_x : \mathcal{A} \to \mathbb{R}_+, \qquad \delta_x(A) := \mathbb{1}_A(x)$ gilt:

$$\int f(t)\ \delta_x(\,\mathrm{d}t) = f(x)$$

Ein vages Stetigkeitsargument motiviert folgenden Schätzer für m:

$$\begin{aligned}
\hat{m}_n &:= \arg\min_{t\in\mathbb{R}} Y_n(t) := \text{ (irgendeine) Minimalstelle der Funktion} \\
Y_n(t) &:= \int_\Omega |x-t| F_n(\,\mathrm{d}x) \\
&= \int_\Omega |x-t| Q_n(\,\mathrm{d}x) \\
&= \frac{1}{n} \cdot \sum_{i=1}^{n} \int |x-t|\ \delta_{X_i}(\,\mathrm{d}x) \\
&= \frac{1}{n} \cdot \sum_{i=1}^{n} |X_i - t|
\end{aligned}$$

$\hat{m}_n$ heißt **empirischer Median** von $X_1, \dots, X_n$ mit üblicher Auswahl $\hat{m}_n = F_n^{-1}\left(\frac{1}{2}\right)$ gemäß Lemma 1.1.1 (da empirische Verteilungsfunktion eie Verteilungsfunktion ist).

Bemerkung.

- Wenn man eine ungerade Anzahl von Daten hat, ist der Median der mittlere Wert, nachdem man die Daten der Größe nach geordnet hat.
- Hat man hingegen eine gerade Anzahl an Daten, dann ist der Median der kleinere der beiden mittleren Werte.

Mit dem starken Gesetz der großen Zahlen (SGGZ) gilt

$$\forall t \in \mathbb{R} : \left(Y_n(t) \overset{n\to\infty}{\longrightarrow} \mathbb{E}[|X_1 - t|] = Y(t) \text{ fast sicher} \right) \tag{1.1}$$

Problem: Folgt aus (1.1) bereits, dass

$$\arg\min_{t\in\mathbb{R}} Y_n(t) \overset{n\to\infty}{\longrightarrow} \arg\min_{t\in\mathbb{R}} Y(t) \text{ fast sicher?}$$

Dann folgte:

$$\hat{m}_n \overset{n\to\infty}{\longrightarrow} m \text{ fast sicher } (\textbf{starke Konvergenz})$$

Wir formalisieren und verallgemeinern:

$$\begin{aligned} &X_i : (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})) \text{ messbar}, \qquad \omega \mapsto X_i(\omega) \\ &\Longrightarrow Y_n(t) := Y_n(t, \omega) = \frac{1}{n} \cdot \sum_{i=1}^{n} |X_i(\omega) - t| \\ &\Longrightarrow Y_n(t, \cdot) : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})) \text{ messbar } \forall t \in \mathbb{R} \end{aligned}$$

Definition. Die **Kollektion**

$$Y_n := \{Y_n(t, \cdot) : t \in \mathbb{R}\} = \{Y_n(t) : t \in \mathbb{R}\}$$

heißt **stochastischer Prozess (SP)**. Die Abbildung

$$X_n(\cdot, \omega) : \mathbb{R} \to \mathbb{R}, \qquad t \mapsto Y_n(t, \omega)$$

heißt **Trajektorie / Pfad** des SP Y_n zu festem $\omega \in \Omega$.

In unserem Beispiel sind für <u>alle</u> $\omega \in \Omega$ die Pfade stetig auf $\mathbb{R}$. Die Abbildung

$$Y_n : \Omega \to XC(\mathbb{R}, \mathbb{R}), \qquad \omega \mapsto Y_n(\cdot, \omega)$$

heißt **Pfadabbildung** des SP Y_n. Wir identifizieren also den SP Y_n mit seiner Pfadabbildung. Damit ist Y_n eine Abbildung von Ω in den Funktionenraum

$$C(\mathbb{R}) := C(\mathbb{R}, \mathbb{R}) := \{f : \mathbb{R} \to \mathbb{R} : f \text{ ist stetig } \}.$$

Sei $d : C(\mathbb{R}) \times C(\mathbb{R} \to \mathbb{R}$ die Metrik der gleichmäßigen Konvergenz der Kompakta auf $\mathbb{C}(\mathbb{R})$ (formale Definition kommt später) und sei

$$\mathcal{B}(C(\mathbb{R})) := \mathcal{B}_d\big(C(\mathbb{R})\big) = \sigma\big(\{G \subseteq C(\mathbb{R}) : G \text{ ist offen bzgl. } d\}\big)$$

die von d induzierte **Borel-σ-Algebra**.
Wir werden sehen, dass die Abbildung

$$Y_n : (\Omega, \mathcal{A}, \mathbb{P}) \to \Big(C(\mathbb{R}), \mathcal{B}\big(C(\mathbb{R})\big)\Big)$$

messbar ist. Y_n ist also eine Zufallsvariable mit Werten im metrischen Raum $\big(C(\mathbb{R}), d\big)$.
Formulierung des Problems im allgemeinen Rahmen: Seien $Y_n,\ n \in \mathbb{N}$ mit Y SP mit stetigen Pfaden (stetige SP). Was lässt sich sagen über die Gültigkeit der folgenden Implikationen?

$$Y \overset{n\to\infty}{\longrightarrow} Y \text{ fast sicher } \Longrightarrow \arg\min_{t\in\mathbb{R}} Y(t) \overset{n\to\infty}{\longrightarrow} \arg\min_{t\in\mathbb{R}} Y(t) \text{ fast sicher} \tag{1.2}$$

Ziel: Welche Art der Konvergenz $Y_n \overset{n\to\infty}{\longrightarrow} Y$ reicht für obige Implikation aus? Gleichmäßige Konvergenz, gleichmäßige Konvergenz auf Kompakta punktweise konvergenz oder sogar nur (1.1)?
Y besitzt womöglich (unter positiven Wahrscheinlichkeiten) keine eindeutige Minimalstelle. Und dann?

Für die Konstruktion von (asymptotischen) Konfidenzintervallen für m benötigt man **Verteilungskonvergenz**:

$$a_n(\hat{m}_n - m) \xrightarrow{\mathcal{L}} \xi \text{ in } \mathbb{R} \tag{1.3}$$

wobei $a_n \to \infty$ in ξ Grenzvariable, die es zu identifizieren gilt. Für die Herleitung von (1.3) favorisiere wieder einen *funktionalen Ansatz*. Sei

$$Z_n(t) := \beta_n \cdot \left(Y_n\left(m + \frac{t}{a_n}\right) - Y_n(m)\right) \qquad t \in \mathbb{R}$$

der sogenannte **reskalierte Prozess zu** Y_n, wobei β_n deine geeignete positive Folge ist.
Damit folgt

$$a_n(\hat{m}_n - m) = \arg\min_{t\in\mathbb{R}} Z_n(t) \tag{1.4}$$

Klar: Z_n ist wieder ein stetiger stochastischer Prozess und damit $(Z_n)_{n\in\mathbb{N}}$ eine Folge von Zufallsvariablen in $\big(C(\mathbb{R}), d\big)$. Wünschenswert auch hier wäre die Gültigkeit folgender Implikation:

$$Z_n \xrightarrow{\mathcal{L}} Z \text{ in } \big(C(\mathbb{R}), d\big) \Longrightarrow \arg\min_{t\in\mathbb{R}} Z_n(t) \xrightarrow{\mathcal{L}} \arg\min_{t\in\mathbb{R}} Z(t) \tag{$*$}$$

Dazu erforderlich ist das Konzept der **Verteilungskonvergenz** von Zufallsvariablen in metrischen Räumen, damit $(*)$ eine wohldefinierte Bedeutung erhält. Dies folgt später. Natürlich auch hier wieder das Problem: Z besitzt mit positiver Wahrscheinlichkeit mindestens 2. Minimalstellen. Und dann?

A. Anhang

A.1. Grundlagen, die man kennen sollen

TODO

Satz A.1.1 (Maßkorrespondenzsatz) Sei X eine Zufallsvariable. Dann bestimmt das Maß eindeutig die Verteilung F von X und die Verteilung bestimmt eindeutig das Maß.

Notation. $F(\,\mathrm{d}x)$ bedeutet man integriert bzgl. Maß Q, was durch F eindeutig festgelegt ist.

Satz A.1.2 (Transformationsformel)

Satz A.1.3 (Lebesgue / dominierte Konvergenz)

Satz A.1.4 (Monotone Konvergenz)

Satz A.1.5 (Gliwenko-Cantelli)