

DER BRUCHPUNKT VON SCHÄTZERN

Wilhelm CASPARY

In: *BEINEKE, Dieter / HEUNECKE, Otto / HORST, Thomas / KLEIM, Uwe G. F. (Hrsg.) [2012]:*

Festschrift für Univ.-Prof. Dr.-Ing. Kurt Brunner

anlässlich des Ausscheidens aus dem aktiven Dienst

Schriftenreihe des Instituts für Geodäsie der Universität der Bundeswehr München, Heft 87, S. 39-46

ISSN: 0173-1009

Der Bruchpunkt von Schätzern

Zusammenfassung

Der Bruchpunkt gibt an, ab wann die Belastung eines Schätzers durch Ausreißer in den Daten so groß ist, dass den Schätzergebnissen nicht mehr getraut werden kann. Die übliche Berechnung von Mittelwert und Standardabweichung bricht schon zusammen, wenn nicht entdeckt wird, dass ein Messwert grob falsch ist. Die robuste Statistik hält eine Reihe von Schätzern bereit, die erst dann versagen, wenn 50% der Beobachtungen Ausreißer sind. Allerdings wird diese Eigenschaft durch zum Teil hohen Verlust an Effizienz erkaufte. Auch für den Fall, dass neben den Beobachtungen die Elemente der Designmatrix Ausreißer enthalten, sind Schätzer verfügbar, deren Bruchpunkt bei 50% liegt. Neben dem Effizienzverlust tritt bei diesen Schätzern das zusätzliche Problem auf, dass ihre Berechnung extrem rechenintensiv ist.

Summary

The breakdown point of an estimator is defined as the smallest ratio of contaminated observations, which leads to extremely aberrant results. The conventional estimators for the mean and the standard deviation of a sample fail already, when a single observation is an unnoticed outlier. Robust statistics offers a number of estimators, which can cope with close to 50% outliers in the observations. But this attractive property comes with a high loss of efficiency. There are also estimators for linear models with a breakdown point of 50% for the case, that the observations and the designmatrix are contaminated. This estimators suffer also from low efficiency and their computation requires extremely complex and extensive numerical methods.

1. Einleitung

Wenn eine Beobachtungsreihe, die zur Schätzung der Parameter eines Modells durchgeführt wurde, unerkannte grobe Fehler enthält, kann dies dazu führen, dass die Schätzergebnisse völlig unbrauchbar sind. Man sagt dann, der Schätzer ist zusammengebrochen. Dieses altbekannte Problem hat zu unzähligen Versuchen geführt, ein schlüssiges Entscheidungskriterium festzulegen, nach dem verdächtige Beobachtungen als noch akzeptabel oder aber als unbrauchbar gekennzeichnet werden kön-

nen. Eine allgemeine Lösung für dieses Problem kann es wohl nicht geben, da die Vielfalt der Aufgabenstellungen und der beobachteten Phänomene zu einer eben solchen Vielfalt unterschiedlicher Eigenschaften der Daten führt. Eine sachgerechte Entscheidung erfordert daher immer das Zusammenspiel von statistischen Entscheidungsregeln und fundiertem Fachwissen.

Wenn die richtige Entscheidung getroffen wurde, ergibt sich ein neues Problem, denn dann muss festgelegt werden, wie mit den zweifelhaften und den als Ausreißer markierten Daten weiter verfahren werden soll. In Anbetracht der Wichtigkeit, eine saubere Lösung zu finden, und der Schwierigkeit, das Problem in den Griff zu kriegen, wundert es nicht, dass es zu dieser Problematik eine schier unüberschaubare Zahl von Veröffentlichungen gibt. Als höchst informative Darstellungen der historischen Entwicklungen und des aktuellen Standes der Wissenschaft zum jeweiligen Veröffentlichungszeitpunkt seien die Arbeiten von *Hawkins [1980]*, *Beckman / Cook [1983]* und *Barnett / Lewis [1994]* genannt. In der Geodäsie wurde die Auseinandersetzung mit dieser Thematik insbesondere durch die Arbeiten von *Baarda [1968]* und *Pope [1976]* angestoßen, denen zahlreiche weitere Veröffentlichungen folgten, in denen die Ausreißersuche (data snooping) zu einem bewährten Verfahren ausgebaut wurde. Während diese Werke das Thema eher aus Sicht der Mathematischen Statistik behandeln und Ausreißertests in den Mittelpunkt stellen, findet man in *Schendera [2007]* eine breite Diskussion des Phänomens Ausreißer in Datenerhebungen der unterschiedlichsten Art, einschließlich ihrer Aufdeckung mit vorwiegend heuristischen graphischen Methoden, ihrer Interpretation und Regeln zur weiteren Behandlung.

Die Behandlung der Ausreißer folgt in der Regel einer der folgenden drei Strategien:

- (1) Für eine hypothetische Verteilung der Beobachtungen und eine gewählte Irrtumswahrscheinlichkeit wird eine Teststatistik berechnet. Beobachtungen, die diese Statistik überschreiten, gelten als Ausreißer. Dienen die Beobachtungen zur Schätzung von Modellparametern, so werden die Ausreißer gestrichen, um eine systematische Verfälschung der Schätzwerte zu verhindern. Gleichwohl enthalten sie oft wichtige Informationen und werden sorgfältig analysiert. Sie können auf Schwächen des Beobachtungsverfahrens oder instrumentelle Probleme hinweisen. Bei manchen Aufgabenstellungen wie

z. B. der Deformationsanalyse stellen sie sogar das eigentliche Messergebnis dar.

- (2) Die Ausreißer werden als Zeichen dafür gewertet, dass das Modell das beobachtete Phänomen nicht adäquat abbilden kann. Eventuell muss das Modell modifiziert oder erweitert werden. Oder es ist ein vollkommen anderes Modell erforderlich. Am Beispiel von Pegelmessungen an der Isar kann dies veranschaulicht werden. Aus den regelmäßigen Pegelablesungen soll ein Zeitreihenmodell für den Jahresverlauf der Wasserstände entwickelt werden. Nun tritt sporadisch Hochwasser auf. Die dann ermittelten Pegelablesungen sind im Vergleich mit den üblichen Werten Ausreißer. Sie sind aber natürlich keine groben Fehler, sondern eher von besonderem Interesse, obwohl sie das Zeitreihenmodell erheblich stören. Eine Lösung könnte in diesem Beispiel die Verwendung von zwei Modellen sein. Eines, das die normalen Wasserstände repräsentiert, und ein zweites, das das seltene Ereignis Hochwasser abbildet.
- (3) Man vermeidet es, die ohnehin fragwürdige und letztlich willkürliche Markierung von Beobachtungen als Ausreißer vorzunehmen. Stattdessen werden Schätzverfahren eingesetzt, die auch beim Vorhandensein von Ausreißern noch zuverlässige Schätzergebnisse liefern. Diese robusten Schätzverfahren, die in großer Zahl zur Verfügung stehen, haben allerdings einen Haken. Der Schutz vor Ausreißern reicht nur bis zum Bruchpunkt des Schätzers und muss mit einem Verlust an Effizienz bezahlt werden. Das bedeutet, wenn die Beobachtungen ausreißerfrei normalverteilt sind, weisen die Schätzergebnisse größere Varianzen auf als optimale Schätzungen nach der Methode der kleinsten Quadrate. Und je robuster der Schätzer ist, desto höher ist der Effizienzverlust.

2. Der Bruchpunkt bei einfachen Stichproben

Zunächst soll der Fall von einfachen Beobachtungsreihen betrachtet werden, die bei gleichgenauen Wiederholungsmessungen entstehen. Das Ziel der Messungen ist die Ermittlung einer möglichst genauen Schätzung der Messgröße. In der Terminologie der Statistik sind der Lage- und der Streuungsparameter der Stichprobe zu schätzen. Wenn die Beobachtungen normalverteilt sind, so ist bekanntlich die Methode der kleinsten Quadrate (MkQ) das optimale Schätzverfahren, das folgende Ergebnisse für den Lageparameter (\hat{x}) und den Streuungsparameter (s) einer Beobachtungsreihe (l) liefert:

$$l = (l_1, l_2, \dots, l_n)$$

$$\hat{x} = \frac{1}{n} \sum_{i=1}^n l_i \quad [1]$$

$$s = \sqrt{\frac{\sum_{i=1}^n (\hat{x} - l_i)^2}{n - 1}} \quad [2]$$

Die Schätzwerte sind erwartungstreu und gehen daher für $n \rightarrow \infty$ in die wahren Parameter ξ und σ der $N(\xi, \sigma^2)$ -verteilten Beobachtungen l_i über.

2.1 Überprüfung der Beobachtungen

Enthält die Beobachtungsreihe Werte, die den Erwartungen widersprechen, so ist dies Anlass, die Messumstände und die Modellvorstellungen zu überprüfen. Eine scharfe Grenze zwischen „normalen“ Beobachtungen und Ausreißern kann mit wissenschaftlichen Methoden nicht gezogen werden. Die Kennzeichnung eines Messwertes als Ausreißer setzt immer eine Hypothese über das erwartete statistische Verhalten der Stichprobe und eine Konvention über die akzeptable Wahrscheinlichkeit einer fehlerhaften Entscheidung voraus. Letztlich ist die Vorgehensweise damit immer subjektiv, auch wenn die Berechnung von Schwellenwerten mit strengen mathematischen Formeln erfolgt.

Wenn keine realistischen Annahmen über die Verteilung der Beobachtungen möglich sind, wird eine Überprüfungsmethode gewählt, die damit beginnt, dass die Werte nach ihrer Größe geordnet werden. Es entsteht dann die geordnete Stichprobe l_o (Ordnungsstatistik), bei der zur Unterscheidung von der ursprünglichen Stichprobe die Indizes in Klammern gesetzt werden:

$$l_o = (l_{(1)}, l_{(2)}, \dots, l_{(n)}), \quad l_{(i)} \leq l_{(i+1)}$$

Es wird nun lediglich vorausgesetzt, dass die Verteilung symmetrisch zum Lageparameter ist. Die geordnete Reihe wird in drei Abschnitte geteilt. Diese werden durch das untere Quartil Q_u und das obere Quartil Q_o begrenzt. Das sind die Grenzen, die von 25% der Werte unter- bzw. überschritten werden. Zwischen den Quartilen liegt daher die mittlere Hälfte der Beobachtungen. Als Ausreißer werden dann die Werte, die kleiner als $Q_u - k(Q_o - Q_u)$ und größer als $Q_o + k(Q_o - Q_u)$ sind, markiert. Für k wird der Wert 1,5 empfohlen, der sich in der Praxis bewährt hat. Neben diesem einfachen pragmatischen Vorgehen sind für diese Situation zahlreiche graphische Verfahren entwickelt worden.

In der Geodäsie wird in der Regel angenommen, dass die Beobachtungen zumindest näherungsweise eine Normalverteilung besitzen. Es folgt dann, dass die Verbesserungen $v_i = \hat{x} - l_i$ ebenfalls normalverteilt sind. Ihre Standardabweichung hat den Schätzwert:

$$s_v = s \sqrt{\frac{n - 1}{n}}$$

Wenn die Schätzungen \hat{x} und s statistisch unabhängig sind, weil sie beispielweise aus zwei unterschiedlichen Beobachtungsreihen berechnet wurden, dann folgt die sogenannte *extern studentisierte* Verbesserung

$$w_i = \frac{v_i}{s_v}$$

einer t -Verteilung mit $f = n - 1$ Freiheitsgraden. Aus den Tafeln der t -Verteilung können dann Grenzwerte t_α entnommen werden, die bei Richtigkeit der Normalverteilungshypothese nur mit $\alpha\%$ Wahrscheinlichkeit von den $|w_i|$ überschritten werden.

Normalerweise werden \hat{x} und s aber aus derselben Beobachtungsreihe geschätzt und sind daher stochastisch abhängig. Dann werden die *intern studentisierten* Verbesserungen

$$z_i = \frac{v_i}{s_v} = \frac{v_i \sqrt{n}}{\sqrt{\sum v_i^2}}$$

gebildet, die eine τ -Verteilung mit $f = n - 1$ Freiheitsgraden besitzen. Da Tabellen der τ -Verteilung wenig verbreitet sind, wird der Schwellenwert der τ -Verteilung meist über die Beziehung

$$\tau_f = t_{f-1} \sqrt{\frac{f}{f-1 + t_{f-1}^2}}$$

aus den vertafelten Werten der t -Verteilung berechnet.

Nach dem Streichen der als Ausreißer identifizierten Werte erfolgt die Parameterschätzung nach der Methode der kleinsten Quadrate (MkQ). Diese Vorgehensweise kann durchaus als robust bezeichnet werden, da sie, wenn nur einzelne Ausreißer vorhanden sind, den „Zusammenbruch“ des Schätzers verhindert.

2.2 Der Bruchpunkt

Wenn die Parameterschätzung in Echtzeit oder automatisch erfolgt oder wenn die Anzahl der Beobachtungen so groß ist, dass individuelle Ausreißersuchmethoden unpraktikabel sind, tritt die Frage auf, wie sich der Schätzer verhält, wenn die Ausreißer nicht eliminiert werden. Liefert er noch sinnvolle Ergebnisse oder bricht er zusammen?

Ähnlich wie in der Elastizitätslehre, wo der Bruchpunkt im Spannungs-Dehnungs-Diagramm angibt, bei welcher Zugbelastung der Prüfkörper reißt, hat jeder Schätzer einen Bruchpunkt, der anzeigt, ab welcher Ausreißerbelastung er versagt.

Nach *Donoho/Huber [1983]* ist der Bruchpunkt, grob gesagt, die kleinste Menge an Kontamination (Ausreißern), die dazu führen kann, dass der Schätzer beliebig große abnorme Werte annimmt. Der Begriff wurde nach *Donoho/Huber [1983]* von *Hampel [1968]* eingeführt und ist seither ein vieldiskutiertes Kriterium der robusten Statistik. Es gibt unterschiedliche Definitionen des Bruchpunktes, die sich vor allem in der Art, wie die Kontamination formalisiert wird, unterscheiden. Sie messen aber alle dieselbe Eigenschaft des Schätzers.

Die einfachste Definition ist das sogenannte Ersetzungsmodell für endliche Stichproben. Sei wieder $l = (l_1, l_2, \dots, l_n)$ die ursprüngliche Beobachtungsreihe.

Eine Untermenge von m Beobachtungen der Reihe wird durch willkürliche Zahlen $z = (z_1, z_2, \dots, z_m)$ ersetzt. Dies ergibt die kontaminierte Stichprobe l^* . Der Anteil kontaminierter Daten (Ausreißer) in der Beobachtungsreihe ist somit $\delta = m/n$. Sei nun $T(L_r)$ ein Schätzer, der für alle Stichproben vom Umfang $r = 1, 2, \dots$ einen Schätzwert liefert, und $T(l)$ das Ergebnis für die Beobachtungsreihe l . Als Bruchpunkt des Schätzers T für die Stichprobe l wird die Größe δ^* definiert, die der kleinste Wert für δ ist, bei dem der Schätzer, wenn er auf die kontaminierte Stichprobe l^* angewandt wird, ein beliebig stark von $T(l)$ abweichendes Ergebnis liefert.

Um diese Definition in Formeln auszudrücken, sei zunächst die maximale systematische Abweichung der Schätzung, die durch δ -Kontamination hervorgerufen werden kann, als

$$B(\delta; T, l) = \sup |T(l^*) - T(l)|$$

bezeichnet, wobei das Supremum über alle δ -kontaminierten Stichproben zu bilden ist. Das bedeutet, dass der ungünstigste Fall der Kontamination von m Beobachtungen anzunehmen ist. Der Bruchpunkt ist dann als

$$\delta^*(T, l) = \inf \left\{ \delta = \frac{m}{n} : B(\delta; T, l) = \infty \right\} \quad [3]$$

definiert.

2.3 Lage- und Streuungsparameter

Um den Bruchpunkt zu veranschaulichen, sei die Schätzung von Lage- und Streuungsparametern der einfachen Stichprobe betrachtet. Es ist sofort einsichtig, dass eine Beobachtung, die den Wert ∞ annimmt, dazu führt, dass sowohl \hat{x} als auch s nach [1] und [2] unendlich groß und damit zu unbrauchbaren Schätzergebnissen werden. Die Schätzer brechen also bereits bei $m = 1$ zusammen und haben daher den Bruchpunkt $\delta^* = 1/n$. Oft wird zur besseren Vergleichbarkeit der Schätzer der asymptotische Bruchpunkt angegeben, der sich auf $n = \infty$ bezieht und für beide Schätzer den Wert $\delta_a^* = 0$ annimmt.

Deutlich robuster sind der Median als Lage- und die Medianabweichung als Streuungsschätzer. Ihre Definition lautet:

$$\text{Median: } \tilde{x} = \text{med}\{l\}$$

$$\text{Medianabweichung: } \tilde{s} = \text{med}(|l_i - \text{med}\{l\}|).$$

Den Median findet man einfach durch Abzählen in der geordneten Stichprobe l_o bzw. $|l_i - \text{med}\{l\}|_o$. Wenn n eine ungerade Zahl ist ($n = 2k + 1$), so ist der Median der in der Mitte stehende Wert: $l_{(k+1)}$. Ist n gerade ($n = 2k$), so wird das Mittel der beiden mittleren Werte der geordneten Stichprobe genommen: $(l_{(k)} + l_{(k+1)})/2$. Sind die Beobachtungen $N(\xi, \sigma^2)$ -verteilt, so ist \tilde{x} ein erwartungstreuer Schätzer für ξ . Der Erwartungswert für \tilde{s} ergibt sich jedoch als $0,6745 \sigma$. Deshalb wird statt \tilde{s} der Schätzwert $\hat{s} = \tilde{s}/0,6745$ benutzt.

Werden k der $n = 2k + 1$ Beobachtungen durch willkürliche Werte ersetzt und wird der ungünstigste Fall angenommen, dass sich alle ersetzten Werte (Ausreißer)

auf derselben Seite des Medians befinden, so bleibt der Median trotzdem unverändert. Erst wenn $m = k + 1$ Werte ersetzt werden, kann der Median über alle Grenzen wachsen. Wir finden daher den Bruchpunkt

$$\delta^* = \frac{k + 1}{2k + 1} = \frac{1}{2} + \frac{1}{2n}$$

und $\delta_a^* = 0,5$. Ähnliche Überlegungen für n gerade führen zu

$$\delta^* = \frac{k}{2k} = 0,5 = \delta_a^*$$

Wenn keine weiteren Informationen genutzt werden können, z. B. für eine Bayes-Schätzung, oder wenn keine Zufallsverteilung der Vorzeichen der Ausreißer angenommen werden kann, kann der Bruchpunkt δ_a^* nur Werte im Intervall $[0; 0,5]$ annehmen. Dies ist auch plausibel. Denn wenn die Hälfte der Werte Ausreißer sind, kann der Schätzer nicht mehr zwischen brauchbaren und unbrauchbaren Werten unterscheiden. MkQ-Schätzer und Median weisen also die beiden möglichen Extremwerte des Bruchpunktes auf. Wie bereits angegeben wird der hohe Bruchpunkt der Medianschätzer durch geringe Effizienz erkauft. Als (relative) Effizienz wird üblicherweise der Quotient aus den asymptotischen Varianzen der zu vergleichenden Schätzer für $N(0, 1)$ -verteilte Beobachtungen bezeichnet. Hier ergibt der Vergleich von Mittel und Median

$$\begin{aligned} \text{Eff}(\hat{x}) &= \frac{\text{Var}(\hat{x})}{\text{Var}(\bar{x})} \\ &= \frac{1}{\frac{\pi}{2}} = \frac{2}{\pi} \approx 0,64 . \end{aligned}$$

Nun gibt es eine ganze Reihe von einfachen robusten Schätzern, deren Bruchpunkt zwischen den Extremwerten liegt und deren Effizienz akzeptabel ist. Als Beispiel sei das h -getrimmte Mittel genannt, das durch

$$\bar{x} = \frac{1}{n - 2h} \sum_{i=h+1}^{n-h} l_{(i)} \quad [4]$$

definiert ist. Es werden also h Werte am unteren und oberen Ende der geordneten Stichprobe gestrichen und von dem Rest das arithmetische Mittel genommen. Für $h = 0$ erhält man wieder das gewöhnliche Mittel und für $h = (n - 1)/2$ bzw. $(n - 2)/2$ den Median. Der Bruchpunkt beträgt $\delta^* = (h + 1)/n$. Durch die Wahl von h hat man es in der Hand, je nach Qualität der Daten, einen hohen Bruchpunkt oder eine hohe Effizienz zu erzielen.

2.4 M-Schätzer

Eine weitere Klasse von robusten Schätzern sind die M -Schätzer, deren Bezeichnung auf die Verwandtschaft mit den Maximum-Likelihood Schätzern hinweist. Wenn die Verteilung der Beobachtungen, die von den zu schätzenden Parametern $\alpha = (\alpha_1, \alpha_2, \dots)$ abhängen möge, kon-

tinuierlich und bekannt ist, hat man für jede Beobachtung l_i den Wert der Dichtefunktion $f(l_i; \alpha)$. Als Likelihoodfunktion für unabhängige Beobachtungen wird das Produkt $\mathcal{L} = \prod_{i=1}^n f(l_i; \alpha)$ bezeichnet. Die Parameter α werden nun so aus den Beobachtungen geschätzt, dass \mathcal{L} sein Maximum annimmt. Nähere Einzelheiten zu dieser Methode findet man unter anderem in *Caspary/Wichmann [2007]*.

Wenn die Beobachtungen eine $N(\xi, 1)$ -Verteilung besitzen, so lautet die Dichtefunktion

$$\varphi(l) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(l-\xi)^2}$$

und die Likelihoodfunktion

$$\mathcal{L} = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2} \sum_{i=1}^n (l_i - \xi)^2} .$$

Um diese Funktion bezüglich des zu schätzenden Parameters ξ zu maximieren, wird das Minimum der Funktion

$$-\ln \mathcal{L} = \frac{\sum (l_i - \xi)^2}{2} + K$$

berechnet. Dieses wird erreicht, wenn $\sum (l_i - \xi)^2 = \min$ bzw. $\sum (l_i - \xi) = 0$ ist. Die Summe

$$\frac{\sum (l_i - \xi)^2}{2} = \sum \rho(l_i, \alpha)$$

wird in der Statistik häufig als Verlust und $\rho(l_i, \alpha)$ als Verlustfunktion bezeichnet. Wenn die Beobachtungen keine Normalverteilung besitzen, so ändert sich die Verlustfunktion, die allgemeine Vorgehensweise bleibt aber unverändert. Nach einem Vorschlag von *Huber [1964]* wird eine Verteilung gewählt, die aus einer Stammverteilung und einer kontaminierenden Verteilung zusammengesetzt ist:

$$F_\varepsilon = (1 - \varepsilon) F + \varepsilon H , \quad 0 < \varepsilon < 1 , H \text{ symmetrisch.}$$

So kann z. B. angenommen werden, dass die Beobachtungen die Verteilung $F = N(\xi, \sigma^2)$ besitzen, die von einem kleinen Prozentsatz $\varepsilon \ll 1$ von Ausreißern überlagert ist. Um für eine solche Mischverteilung einen Schätzer mit minimaler asymptotischer Varianz zu erhalten, ist nach *Huber [1981]* für $u = (l - \xi)/\sigma$ folgende Verlustfunktion optimal:

$$\rho(u) = \begin{cases} \frac{1}{2} u^2 & \text{für } |u| < k \\ k|u| - \frac{1}{2} k^2 & \text{für } |u| \geq k \end{cases} ,$$

aus der nach Differentiation die Schätzgleichung

$$\sum \psi(u) = 0 \quad \text{mit} \quad \psi(u) = \frac{d\rho(u)}{du}$$

$$\psi(u) = \begin{cases} u & \text{für } |u| < k \\ k \text{ sign}(u) & \text{für } |u| \geq k \end{cases}$$

folgt. Für die Abstimmkonstante k ist ein Wert zwischen 1,5 und 2,0 empfehlenswert. Die Verlustfunktion wächst quadratisch für $-k \leq u \leq +k$ und außerhalb dieses Intervalls nur noch linear. Dieser Huber-Schätzer hat bei bekannter Varianz den maximalen Bruchpunkt $\delta_a^* = 0,5$.

Die Veröffentlichung von Huber [1964] hat eine Flut von Arbeiten auf dem damit initiierten Gebiet der robusten Schätzung ausgelöst. Insbesondere wurde das Konzept der M -Schätzung von zahlreichen Autoren verallgemeinert, indem gezeigt wurde, dass die Verlustfunktion $\rho(l_i, \alpha)$ nicht mit einer Zufallsverteilung in Verbindung stehen muss, sondern nur bestimmte Eigenschaften aufweisen muss, um zu robusten Schätzfunktionen zu führen. Sie sollte differenzierbar, symmetrisch, konvex und nicht abnehmend sein. Von den vielen Vorschlägen für geeignete ρ -Funktionen sei noch die sehr populäre biquadratische Funktion nach TUKEY angegeben, die zu den robusten M -Schätzern mit Verwerfungspunkt gehört. Alle standardisierten Residuen u , die größer als die Abstimmkonstante c ausfallen, werden verworfen. Die Schätzgleichung wird mit

$$\psi(u) = \begin{cases} u \left[1 - \left(\frac{u}{c} \right)^2 \right]^2 & \text{für } |u| \leq c \\ 0 & \text{für } |u| > c \end{cases}$$

gebildet. Auch dieser Schätzer hat den maximalen Bruchpunkt von 0,5.

Die Berechnung der M -Schätzung muss iterativ erfolgen, da sie von den Verbesserungen abhängt, die erst nach der Schätzung bekannt sind. Das einfachste Rechenverfahren ist die iterativ nachgewichtete MkQ. Dazu werden mit

$$w(u) = \begin{cases} \frac{\psi(u)}{u} & \text{für } u \neq 0 \\ \psi'(u) & \text{für } u = 0 \end{cases}$$

Gewichte definiert, mit denen die Schätzgleichung formuliert wird und die Form

$$\sum_{i=1}^n w(u_i) u_i = 0 \quad \text{mit} \quad u_i = \frac{l_i - x}{\sigma}$$

annimmt oder nach Einsetzen von u_i die Darstellung als gewogenes Mittel

$$x = \frac{\sum w_i l_i}{\sum w_i}$$

erhält. Als Startwert der Median oder der Mittelwert gewählt werden.

3. Lineare Modelle

Das lineare (linearisierte) Model hat in der Geodäsie gewöhnlich die Form

$$\begin{aligned} \mathbf{l} &= \mathbf{A}\mathbf{x} + \boldsymbol{\varepsilon}, \text{ bzw. } \mathbf{l} + \mathbf{v} = \mathbf{A}\hat{\mathbf{x}}, \mathbf{P}, \\ E(\boldsymbol{\varepsilon}) &= \mathbf{0}, \text{ Var}(\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma}_\varepsilon = \boldsymbol{\Sigma}_l = \sigma_0^2 \mathbf{Q}, \\ \mathbf{Q}^{-1} &= \mathbf{P}. \end{aligned} \quad [5]$$

Es sei angenommen, dass die Beobachtungen unabhängig sind und damit \mathbf{P} diagonal ist. Um die folgenden Darstellungen zu vereinfachen, denken wir uns das Modell [5] von links mit $\mathbf{P}^{1/2}$ multipliziert. Für die so homogenisierten Beobachtungen gilt nun $\mathbf{P} = \mathbf{Q} = \mathbf{I}$. Da in linearen Modellen neben den Ausreißern in den Beobachtungen auch grob fehlerhafte Werte in der Modellmatrix auftreten können, wird eine Kontamination der Zeilenvektoren $\mathbf{w}_i^t = (\mathbf{a}_i^t, l_i)$ bei der Definition des Bruchpunktes angenommen.

3.1 Der Bruchpunkt

Wir wollen uns hier an der Terminologie von Rousseeuw/Leroy [1987] orientieren und die einfache Definition für Stichprobenmodelle übernehmen: Sei das lineare Modell durch n Datenpunkte

$$\mathbf{w}_i^t = (a_{i1}, a_{i2}, \dots, a_{iu}, l_i) \in \mathcal{R}^{u+1}, \quad i = 1, 2, \dots, n$$

$$\mathbf{W} = \begin{pmatrix} \mathbf{A} & | & \mathbf{l} \\ \hline n \times u & & n \times 1 \end{pmatrix}$$

gegeben und sei das Funktional $\mathbf{T}(\mathbf{W}) = \hat{\mathbf{x}}$ der Vektor der geschätzten Modellparameter. Wird nun die Stichprobe \mathbf{W} kontaminiert, indem in m der Datenpunkte beliebige Werte eingesetzt werden, so entsteht eine neue Stichprobe \mathbf{W}^* . Der Vektor $\mathbf{T}(\mathbf{W}^*)$ der mit dieser kontaminierten Stichprobe geschätzten Parameter besitzt die als euklidische Norm definierte systematische Abweichung (Verzerrung)

$$b(\mathbf{T}, \mathbf{W}) = \|\mathbf{T}(\mathbf{W}^*) - \mathbf{T}(\mathbf{W})\|.$$

Ziehen wir nun alle möglichen auf diesem Wege kontaminierbaren Stichproben \mathbf{W}^* in Betracht und bezeichnen mit

$$B(m; \mathbf{T}, \mathbf{W}) = \sup_{\mathbf{W}_m^*} \|\mathbf{T}(\mathbf{W}_m^*) - \mathbf{T}(\mathbf{W})\|$$

die maximal mögliche Verzerrung, die durch beliebige Verfälschung von m Datenpunkten erzeugt werden kann. Nimmt diese Verzerrung den Wert unendlich an, so haben m Ausreißer den Schätzer \mathbf{T} am Modell \mathbf{W} zusammenbrechen lassen. Als Stichprobenversion δ^* des Bruchpunktes wird nun der kleinste Anteil $m/n = \delta$ der Datenpunkte definiert, der im ungünstigsten Fall den Zusammenbruch des Schätzers verursachen kann. Dies führt in Analogie zu Gleichung [3] zur Darstellung des Bruchpunktes für Schätzer in linearen Modellen:

$$\delta_n^*(\mathbf{T}, \mathbf{W}) = \inf \left\{ \frac{m}{n} = \delta \geq 0 ; B(m; \mathbf{T}, \mathbf{W}) = \infty \right\}.$$

3.2 M -Schätzer

Die Methode der kleinsten Quadrate hat natürlich auch in linearen Modellen den Bruchpunkt 0, wie man sich

leicht klarmachen kann. Wenn ein Element eines beliebigen Vektors \mathbf{w}_i den Wert ∞ annimmt, erhält man beliebig abnorme Schätzwerte. Etwas komplizierter ist die Situation bei den robusten M -Schätzern.

Als M -Schätzer werden alle Schätzer für lineare Modelle bezeichnet, die durch die Zielfunktion

$$\sum_{i=1}^n \rho \left(\frac{l_i - \mathbf{a}_i^t \mathbf{x}}{\hat{\sigma}} \right) = \min$$

und die Schätzgleichung

$$\sum_{i=1}^n \psi \left(\frac{l_i - \mathbf{a}_i^t \hat{\mathbf{x}}}{\hat{\sigma}} \right) \mathbf{a}_i = \mathbf{0}, \quad \psi = \rho'$$

definiert sind. Ihre Eigenschaften hängen von der Form der ρ -Funktion ab, für die wieder vorausgesetzt wird, dass sie differenzierbar, symmetrisch, konvex und nicht abnehmend ist. Wenn Ausreißer nur in den Beobachtungen auftreten können, behalten die M -Schätzer auch in linearen Modellen ihren hohen Bruchpunkt. Allerdings schneiden die Schätzer mit Verwerfungspunkt unter Robustheitsgesichtspunkten etwas besser ab, da Ausreißer nicht herabgewichtet, sondern komplett gestrichen werden. Gegen Ausreißer in Elementen der Designmatrix \mathbf{A} bieten M -Schätzer jedoch keinen Schutz. Wenn diese in Betracht gezogen werden müssen, so erhält man den Bruchpunkt 0. Um eine Verbesserung dieser Situation zu erreichen, sind die M -Schätzer weiterentwickelt worden.

Ein Maß für den Einfluss einer Zeile \mathbf{a}_i^t der Designmatrix auf die Parameterschätzung ist ihr sogenannter Hebelwert h_{ii} . Die Hebelwerte sind die Diagonalelemente der Projektionsmatrix $\mathbf{H} = \mathbf{A}(\mathbf{A}^t \mathbf{A})^{-1} \mathbf{A}^t$, die alle n -Vektoren orthogonal aus dem Spaltenraum von \mathbf{A} abbildet. Da für alle orthogonalen Projektionsmatrizen Spur und Rang gleich sind, erhält man

$$r(\mathbf{H}) = sp(\mathbf{H}) = \sum_{i=1}^n h_{ii} = u.$$

Die Hebelwerte h_{ii} , oder besser die Redundanzbeiträge $f_i = 1 - h_{ii}$ und $\sum f_i = f = n - u$, spielen in der Geodäsie als Kenngrößen der Zuverlässigkeit seit Langem eine wichtige Rolle. Sie geben an, wie gut eine Beobachtung durch die anderen Beobachtungen kontrolliert ist.

Um die M -Schätzer gegen Ausreißer in den Vektoren \mathbf{a}_i unempfindlich zu machen, werden zusätzliche, von f_i abhängende Gewichte eingeführt. Die Funktion

$$\psi \left(\frac{l_i - \mathbf{a}_i^t \hat{\mathbf{x}}}{\hat{\sigma}} \right) = \psi(u)$$

wird dazu durch die Funktion

$$\phi(\mathbf{a}, u) = g_1(\mathbf{a}) \psi(u, g_2(\mathbf{a}))$$

ersetzt, in der g_1 und g_2 mit den Vektoren \mathbf{a} gebildete Gewichte sind. Von den vielen Möglichkeiten seien die zwei bekanntesten angegeben. Man erhält verallgemei-

nernte M -Schätzer (GM -Schätzer) vom „Mellows-Typ“, wenn $g_2 = 1$ gesetzt und für g_1 beispielsweise eine Funktion von f_i^{-1} gewählt wird. Da dieser Ansatz auch Beobachtungen mit kleinen Verbesserungen herabgewichtet, was in der Regel nicht erwünscht ist, wird meist dem GM -Schätzer vom „Schweppe-Typ“ der Vorzug gegeben. Bei diesem wird $g_1 = g_2^{-1}$ gesetzt und mit $\phi(\mathbf{a}, u) = g^{-1} \psi(u, g)$ die Schätzfunktion gebildet. Von den vielen Vorschlägen zur Wahl von g sei die in *Huber [1981]* als für moderate Hebelpunkte optimal abgeleitete Gewichtsfunktion $g = \sqrt{f_i}$ genannt.

In der Literatur findet man zahlreiche weitere Varianten der GM -Schätzer, die sich vor allem in der Wahl der Gewichtsfunktionen unterscheiden. Sie besitzen allerdings einen niedrigen Bruchpunkt, der mit wachsender Zahl der Parameter gegen Null strebt, so dass sie, wenn Robustheit gegen extreme Ausreißer erforderlich ist, nicht zu empfehlen sind. In nur geringfügig kontaminierten Modellen, wie sie in der Geodäsie eher zu erwarten sind, können sie allerdings, da sie durch die nachgewichtete MkQ einfach zu berechnen sind, mit Gewinn eingesetzt werden.

3.3 Schätzer mit hohem Bruchpunkt

Als Nächstes sei die Frage behandelt, welche Schätzer erreichen in linearen Modellen einen hohen Bruchpunkt, und gibt es für den Bruchpunkt auch hier eine Obergrenze?

Die Herleitung der Obergrenze für den erreichbaren Bruchpunkt in linearen Modellen findet man in *Rousseeuw / Leroy [1987]*. Wie in der Statistik meist üblich, wird angenommen, dass die Elemente der Designmatrix Realisationen von Zufallsvariablen sind. Man hat es also mit Zufallsvektoren $\mathbf{w}_i^t = (a_{i1}, a_{i2}, \dots, a_{iu}, l_i) \in \mathcal{R}^{u+1}$, $i = 1, 2, \dots, n$ zu tun. Unter dieser Annahme kann man mit einer Wahrscheinlichkeit nahe eins annehmen, dass die Designmatrix in *regulärer Anordnung* (general position) ist. Das bedeutet, dass jede $u \times u$ -Untermatrix, die mit den Zeilen von \mathbf{A} gebildet werden kann, vollen Rang besitzt. Unter dieser Voraussetzung erhält man den maximalen Bruchpunkt von

$$\delta^* = \frac{[n/2] - u + 2}{n}$$

wobei $[n/2]$ für die größte ganze Zahl steht, die kleiner als $n/2$ ist.

Bei Anwendungen der Parameterschätzung im Bereich der technischen Wissenschaften hat man es meist mit geplanten Experimenten oder mit geometrischen Größen zu tun, bei denen Modellmatrizen mit festen Koeffizienten auftreten. Man kann bei diesem Design nicht damit rechnen, dass \mathbf{A} die reguläre Anordnung besitzt. Insbesondere wenn es um Höhen- oder Positionsschätzungen in der Geodäsie geht, enthält die Koeffizientenmatrix eine große Zahl von Nullen. Die Designmatrix ist dann meist in *singulärer Anordnung* und be-

sitzt zahlreiche $u \times u$ -Untermatrizen, die singular sind. In diesen Modellen beträgt nach Mizera / Müller [1999] der maximale Bruchpunkt

$$\delta^* = \frac{\left[\frac{n - \mathcal{N}(\mathbf{A}) + 1}{2} \right]}{n}.$$

Hierbei ist $\mathcal{N}(\mathbf{A})$ die maximale Anzahl von Vektoren \mathbf{a}_i , die in einem Unterraum des \mathcal{R}^u liegen:

$$\mathcal{N}(\mathbf{A}) = \max_{x \neq 0} \text{card}\{m : \mathbf{a}_m^t \mathbf{x} = 0\}.$$

Wenn $\mathcal{N}(\mathbf{A}) = u - 1$ ist, ist das Modell in regulärer Anordnung und man erhält denselben Bruchpunkt wie oben. Leider ist keine direkte Methode zur Ermittlung von $\mathcal{N}(\mathbf{A})$ bekannt.

Es gibt mehrere robuste Schätzer für lineare Modelle, die den maximalen Bruchpunkt erreichen und deren Zielfunktionen leicht angegeben werden können. Am bekanntesten ist der *LMS*-Schätzer (least median of squares), der den Median der Quadrate der Verbesserungen minimiert:

$$\text{med}_i(\mathbf{a}_i^t \mathbf{x} - l_i)^2 \rightarrow \min_x \Rightarrow \hat{\mathbf{x}}.$$

Ein weiterer Schätzer ist eine Verallgemeinerung des getrimmten Mittels für einfache Stichproben [4] und beruht auf folgender Minimierung:

$$\sum_{i=1}^h (\mathbf{a}^t \mathbf{x} - l)_{(i)}^2 \rightarrow \min_x \Rightarrow \hat{\mathbf{x}}.$$

Als *S*-Schätzer wird ein Schätzer bezeichnet, der einen aus den Verbesserungen zu schätzenden Streuungsparameter minimiert:

$$s\{v_1(\mathbf{x}), v_2(\mathbf{x}), \dots, v_n(\mathbf{x})\} \rightarrow \min_x \Rightarrow \hat{\mathbf{x}}.$$

Der Streuungsparameter wird in der Regel als *M*-Schätzer indirekt durch die Beziehung

$$\frac{1}{n} \sum_{i=1}^n \rho\left(\frac{v_i}{s}\right) = K$$

definiert. Hierin ist $\rho(t)$ eine gerade, zweimal stetig differenzierbare Funktion mit $\rho(0) = 0$, die im Intervall $[0, c]$ streng monoton steigend und in $[c, \infty]$ konstant ist. c ist der Verwerfungspunkt und K eine geeignete Konstante, die meist als $K = E_\Phi(\rho(t))$ gewählt wird, um zu erreichen, dass bei normalverteiltem t für s die Standardabweichung erhalten wird.

Der maximale Bruchpunkt, den diese Schätzer erreichen können, wird durch geringe Effizienz und zum Teil durch extremen Rechenaufwand erkauft. Da die Zielfunktionen keine direkten Funktionen der zu schätzenden Parameter sind, kommen zur Berechnung nur rechenintensive Methoden der nichtlinearen Optimierung oder der statistischen Versuche infrage. Mit der mathematischen Optimierung können zur Zeit nur relativ kleine Modelle bewältigt werden. Wenn mit statistischen Versuchen z. B. der *LMS*-Schätzer exakt berech-

net werde soll, so müssen $\binom{n}{u}$ Gleichungssysteme der Ordnung $u \times u$ gelöst werden, um das Minimum zu finden. Bei $n = 100$ und $u = 4$ ergibt dies bereits 3.921.225 Gleichungssysteme, und wenn u auf 10 erhöht wird, erhält man die gigantische Zahl $17,3 \times 10^{12}$. Daraus folgt, dass statt der exakten Lösung nur eine mit akzeptabler Wahrscheinlichkeit richtige (gute) Lösung berechenbar ist. Der Einsatz von Segmentierungsmethoden und gewisse Vereinfachungen haben schnelle Algorithmen ermöglicht, mit denen heute diese Schätzungen bewältigt werden können.

4. Die Bedeutung der Schätzer in der Geodäsie

Der in der Literatur über robuste Schätzverfahren ausführlich behandelte Bruchpunkt, ist immer dann bedeutsam, wenn mit zahlreichen extremen Ausreißern in den Beobachtungen gerechnet werden muss.

Bei den klassischen geodätischen Messungen zur Höhen- und Positionsbestimmung sind Ausreißer eher selten, da schon im Feld bewährte Methoden der Qualitätskontrolle zum Zuge kommen. Und sollten diese doch einmal versagen, so werden die Ausreißer zuverlässig durch statistische Tests der Verbesserungen aufgedeckt. Dasselbe gilt für Kalibriermessungen im Feld oder im Labor. Die Methode der kleinsten Quadrate, flankiert von Messkontrollen und statistischen Tests, ist für diesen Aufgabenbereich nach wie vor empfehlenswert.

Anders ist die Situation bei der Auswertung von Messungen zur Aufdeckung von Deformationen. Wenn man nicht mit Sicherheit zwischen stabilen und deformationsgefährdeten Punkten unterscheiden kann, bringt der Einsatz robuster Schätzverfahren große Vorteile. Mit ihrer Hilfe sind Punktbewegungen leichter aufzudecken und nachzuweisen. Da die Deformationen selten extrem sind, sind *M*- oder *GM*-Schätzer mit oder ohne Verwerfungspunkt für diesen Zweck sehr geeignet. Ein hoher Bruchpunkt hat hier kaum Bedeutung. Allerdings geht die Entwicklung heute zu permanenten Messverfahren mit automatischer Auswertung. Dies erfordert eine intensive Untersuchung möglicher Fehlerquellen und eventuell darauf basierend eine Neubewertung der Anforderungen an das Schätzverfahren.

Ein weiteres, relativ junges Gebiet der Datenauswertung ist durch die Scannertechnologie, durch Satelliten zur Erdbeobachtung und durch das maschinelle Sehen entstanden. Kennzeichnend sind riesige Datenmengen, aus denen die interessierende Information zu gewinnen ist. Die Auswertemethoden in diesem Bereich sind sehr unterschiedlich, da sie in dem jeweiligen Arbeitsgebiet entwickelt wurden und den Besonderheiten der Aufgabenstellung angepasst sind. Für serielle Daten wurden robuste Filter zur Elimination von Ausreißern entwickelt. Die Parameterschätzung mit den bereinigten Da-

ten erfolgt mit robusten Schätzern. Bei Punktwolken geht es oft darum, Objekte zu erkennen oder parametrisierte Strukturen zu erfassen. Dazu müssen Objektpunkte identifiziert und Parameter geschätzt werden. Wegen der meist großen Zahl von nicht verwertbaren Punkten (Ausreißern) werden Methoden mit hohem Bruchpunkt benötigt. Als Beispiel einer solchen Schätzmethode sei RANSAC genannt. In diesem Algorithmus wird eine Strategie realisiert, die mit den numerischen Methoden der LMS-Schätzung vergleichbar ist.

5. Literatur

- Baarda, Willem [1968]:* A testing procedure for use in geodetic networks. Publications on Geodesy, New Series, Vol. 2, No. 5, Rijkscommissie voor Geodesie, Delft.
ISSN 0165-1706
- Barnett, Vic / Lewis, Toby [1994]:* Outliers in Statistical Data. Third Edition. Wiley series in probability and mathematical statistics. Wiley, Chichester.
ISBN 0-471-93094-6
- Beckman, Richard J. / Cook, R. Dennis [1983]:* Outlier.....s: discussion paper with comments and response. In: Technometrics, Vol. 25, No. 2, Alexandria, S. 119-163.
ISSN 0040-1706
- Caspary Wilhelm / Wichmann, Klaus [2007]:* Auswertung von Messdaten. Statistische Methoden für Geo- und Ingenieurwissenschaften. Oldenbourg, München/Wien.
ISBN 3-486-58351-4
ISBN 978-3-486-58351-9
- Donoho, David L. / Huber, Peter J. [1983]:* The notion of breakdown point. In: *Bickel, Peter J. / Doksum, Kjell A. / Hodges, Joseph L. (Hrsg.):* A Festschrift for Erich L. Lehmann in honor of his sixty-fifth birthday. Wadsworth International Group, Belmont, S. 157-184.
ISBN 0-534-98044-9
- Hampel Frank R. [1968]:* Contribution to the Theory of Robust Estimation. Ph.D. Thesis, University of California, Berkeley.
- Hawkins, Douglas M. [1980]:* Identification of Outliers. Monographs on Applied Probability and Statistics. Chapman and Hall, London / New York.
ISBN 0-412-21900-X
- Huber, Peter J. [1964]:* Robust estimation of a location parameter. In: The Annals of Mathematical Statistics, Vol. 35, No. 1, Baltimore, S. 73-101.
ISSN 0003-4851
- Huber, Peter J. [1981]:* Robust Statistics. Wiley Series in Probability and Statistics. 1. Auflage, Wiley & Son, New York u. a., (2. Auflage, 2004).
ISBN 0-471-41805-6 (1. Auflage)
ISBN 0-471-65072-2 (2. Auflage)
- Mizera, Ivan / Müller, Christine H. [1999]:* Breakdown Points and Variation Exponents of Robust M -Estimators in Linear Models. In: The Annals of Statistics, Vol. 27, No. 4, Cleveland, S. 1164-1177.
ISSN 0090-5364
- Pope, Allen J. [1976]:* The Statistics of Residuals and the Detection of Outliers. U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Ocean Survey, Geodetic Research and Development Laboratory, NOAA Technical Report, NOS 65, NGS 1, Rockville. Springfield.
- Rousseeuw, Peter J. / Leroy, Annick M. [1987]:* Robust Regression and Outlier Detection. Wiley Series in Probability and Statistics. 1. Auflage, Wiley & Sons, New York u. a. (2. Auflage, 2003).
ISBN 0-471-85233-3 (1. Auflage)
ISBN 0-471-48855-0 (2. Auflage)
- Schendera, Christian F. G. [2007]:* Datenqualität in SPSS. Oldenbourg, München.
ISBN 3-486-58214-3
ISBN 978-3-486-58214-7

Anschrift des Verfassers:

Univ.-Prof. Dr.-Ing. Wilhelm Caspary
Grillparzerstr. 51, D-81675 München
E-Mail: wilhelm.caspary@unibw.de